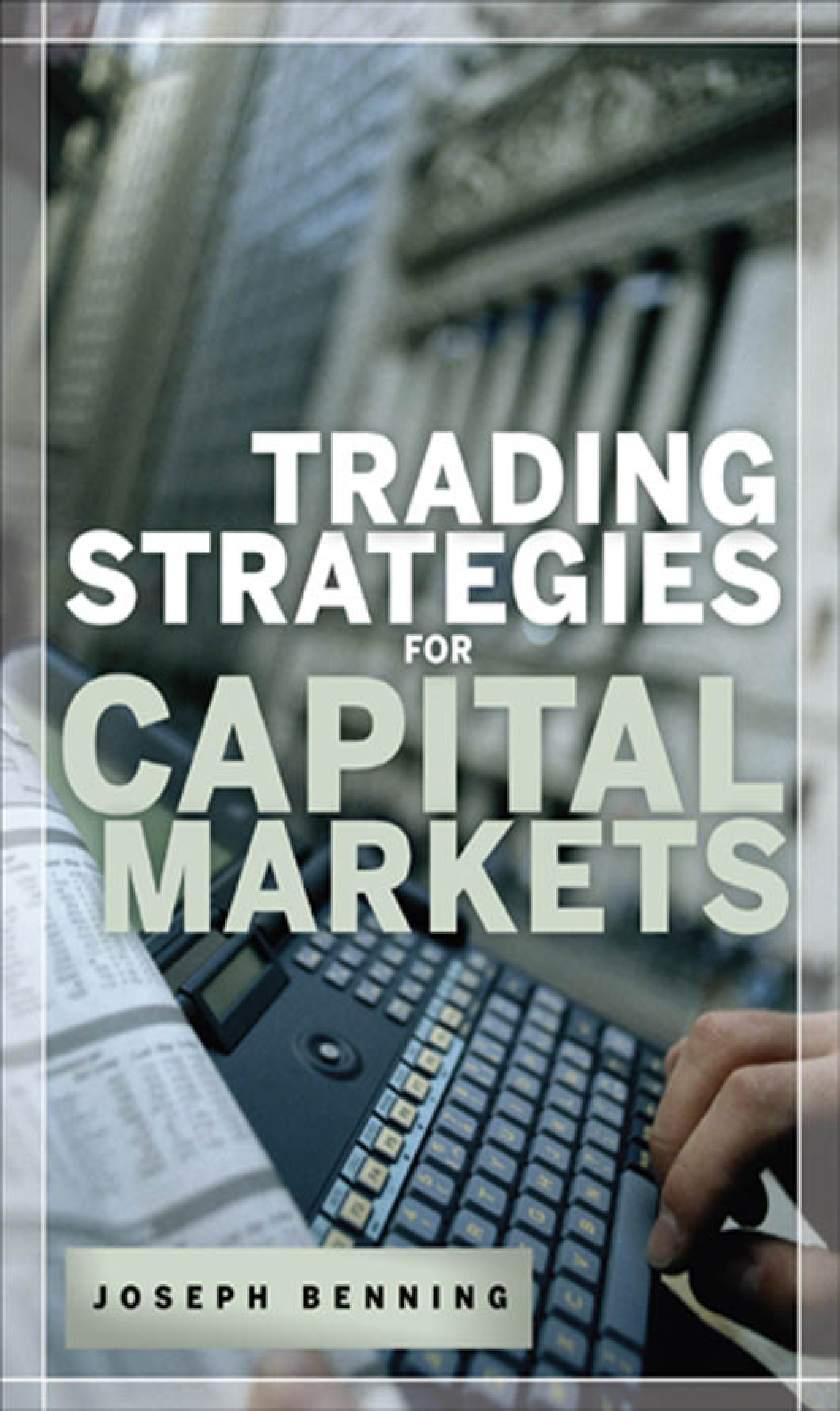


TRADING STRATEGIES FOR CAPITAL MARKETS

JOSEPH BENNING

TRADING STRATEGIES FOR CAPITAL MARKETS

TRADING STRATEGIES FOR CAPITAL MARKETS

JOSEPH F. BENNING

McGraw-Hill

New York Chicago San Francisco
Lisbon London Madrid Mexico City
Milan New Delhi San Juan Seoul
Singapore Sydney Toronto

0-07-159340-3

The material in this eBook also appears in the print version of this title: 0-07-146496-4.

DOI: 10.1036/0071464964

To my wife
Mary Anne McDonald

This page intentionally left blank

Contents

TRADING STRATEGIES FOR CAPITAL MARKETS

This page intentionally left blank

Introduction

In his superbly entertaining and informative book, *Hedgehogging*, investment guru Barton Biggs quotes Charlie Munger, alter ego of Warren Buffett and vice chairman of Berkshire Hathaway: “I have known no wise person over a broad subject matter who didn’t read all the time—none, zero. Now I know all kinds of shrewd people who by staying within a narrow area can do very well without reading. But investment is a broad area. So if you think you’re going to be good at it and not read all the time, you have a different idea than I do . . . . You’d be amazed at how much Warren [Buffett] reads. You’d be amazed at how much I read.” In its 2005 listing of the world’s 400 richest people, *Forbes* described Munger as a self-made man with a net worth at $1.7 billion, making him number 387.

This is a book about trading and financial markets that takes Munger’s advice seriously. It borrows insights from many disciplines and casts a wide net in the search for perspective. It uses financial history to develop an understanding of market institutions and to provide a filter for viewing current market practice. It uses finance theory as a framework for analysis and to build a toolbox that traders can use as they battle in the trenches, which is where it really counts. In this respect, the book seeks to place the markets and trading strategies in the bigger picture of the global political economy. It describes how the capital markets work in practice; what the drivers are, how they can be recognized, and how sensible trading strategies can be developed and implemented. Above all, the book understands markets to be dynamic real-time reflectors of the world in which we live, rather than mathematical abstractions.

Financial markets are rational (but not perfect) discounters of events, information, and trends. However, the book presents an argument (along with evidence) that some bets in the market are better than others. That is, there are some bets that have a demonstrably better chance of success, with less risk. The market is not a strictly random walk. While some may be *fooled by randomness,* in Nicholas Taleb's memorable phrase, this book argues that the opposite can be the case as well: Nonrandom events are sometimes incorrectly attributed to chance. The result is that valuable opportunities are missed. The underlying thrust of strategy should be to take a broad enough view of events to be able to recognize favorable conditions and capitalize on them, a modus operandi similar to card counting at blackjack, but infinitely more complex.

The book is divided into three sections. The first presents a conceptual framework for market analysis. It can be thought of as a road map to the discussion that follows. The second section explores core elements of the capital markets. It describes market organization; key elements of the securities and derivatives that account for the bulk of the instruments traded, and key price drivers. With the liberal use of examples, it goes on to elucidate different trading strategies and how to implement them, depending on market view. Finally, the last section considers risk management, reviews some recent developments in behavioral finance research, offers a view of how markets function, and sums up.

THE CONCEPTUAL FRAMEWORK

The conceptual framework for the book sees financial markets not as mathematical abstractions but as engines of discovery that reflect the economic, political, and social forces that shape—and are shaped by—the societies in which they reside. Accordingly, the model presented to organize this discussion is a qualitative one. It portrays the markets in terms of cycles, big ones and little ones. The little ones are business cycles. The big ones can be sparked by major events, or they can gather force as trends in apparently unrelated areas coalesce and produce discontinuities that result in rapid and unexpected change. These discontinuities can produce effects that ripple through financial markets for years or even decades to come.

The idea that ripple effects can affect prices down the road years later in a predictable way is controversial. Many, if not most, finance theorists (though not practitioners) subscribe to the idea that markets quickly, almost instantaneously, absorb and discount new information. Market prices are a combination of random noise and reaction to new information. The new information is thought to be quickly reflected in market prices, which are a reasonably accurate assessment of fundamental values, on average, over time. If prices do fall temporarily out of line, arbitrageurs will buy the cheap securities and sell the expensive ones until their prices revert back to true fundamental values. Sometimes price are high; sometimes they are low, but over time and on average they are priced about right. Markets are mean-reverting, or at least sufficiently so, so that no one is able to beat the average on a risk-adjusted basis over time. That, in a nutshell, describes the core of the efficient market hypothesis (EMH).

This book has no serious quarrel with the general thrust of that argument. Considerable research has shown that most new information is in fact discounted fairly quickly. Among the more compelling evidence offered is the fact that no one has shown a way in which the market can be beaten systematically over time, after adjusting for risk. The view developed here is from a different perspective. It is that there are market tipping points where the odds of generating trading profits are better than average. In that sense, a successful trader is like a blackjack player who is adept at counting the cards. But there is an important difference. Successful card counters are routinely escorted out of the casino. In the financial markets the players are typically given broad hints about policies likely to affect market prices. Knowing how to interpret those hints and act on them is critical to successful trading. But that requires understanding the historical forces that drive markets and the economic, political, and social institutions that guide them. Note the idea that there are historical forces in play. The Danish philosopher Kierkegaard is credited with making the observation that life must be lived forward, but can only be understood backwards. So we must understand history. As history unfolds, so do markets. The pace is uneven, filled with discontinuities. But understanding history provides both a framework and a lens for viewing and understanding market behavior.

This book suggests that while most information is rapidly incorporated into market prices, not all of it is, at least not immediately. Sometimes it takes the market rather a while to fully incorporate new information into the price structure. Recognizing and acting on situations like this is the trader's equivalent of card counting. But that is difficult for many traders to do. There is always the uncomfortable feeling of the train having left the station. And it is difficult to act without putting the market's behavior in some kind of recognizable context that gives structure to decision making. Accordingly, the market model presented in the first chapter is designed to help structure decision making in such a way that the odds of success are better than they ordinarily would be.

After the model is presented, the first section goes on to detail economic, political, social, and institutional developments that have shaped (and continue to shape) the capital markets. An important part of the story, often forgotten, is that history, institutions, and culture matter. Financial markets do not exist in a vacuum. They function in the rough and tumble of the real world; they have long memories, *pace* Harry Markowitz, and they exhibit recurrent patterns of behavior. It is temptingly easy to be seduced by the four most dangerous words in the English language: "This time it's different." Trying to trade without putting market behavior in some sort of political, institutional, and historical context is simply asking for trouble.

INSTRUMENTS, INSTITUTIONS, AND RISK

The second section of the book concerns the instruments and institutions of the marketplace. It discusses key characteristics of the major instruments in the major market sectors. Treasury notes, bills, bonds, and repurchase agreements are covered on the debt side, as are Chicago Board of Trade Treasury futures and federal funds. Eurodollar futures contracts are briefly discussed. The mechanics of swaps and options are considered. On the equity side indexes such as the Dow Jones 30 Industrials, the S&P 500, and the Russell 2000 are analyzed, as well as the companion futures contracts that trade against them. Gold, though not a capital market instrument, is followed by many market professionals and is regarded by many as the canary in the coal mine. It is also central to the story of the formation of modern capital markets. Accordingly, some notes on gold trading are included as well.

On the debt side of the equation, considerable time is spent on Fed policy, the yield curve, and Treasury futures contracts. Pricing models based on the yield curve and Fed policy are developed. Market structure and the mechanics of Treasury auctions come in for some discussion. Some rationales for adopting trading strategies based on monetary policy are explored. Methods of strategy implementation are examined as well. The chapter on basis trading reviews the traditional cash-and-carry model using Chicago Board of Trade (CBOT) Treasury bond and note futures contracts. Specifics of the delivery process and its relevance to basis trading are considered using numerous examples.

With respect to equities, the book reviews some of the more prominent stock pricing models and criticisms of them. It examines three prominent U.S. equity indexes and the methodology used to construct them and their usefulness as benchmarks. It discusses important structural features of equity markets, requirements for exchange listing, valuation theories, and theories of market timing. It also takes a look at exchange traded funds (ETFs), their pricing and uses. In addition to cash market indexes and ETFs this section devotes a chapter to equity index futures contracts. Particular attention is paid to the cash-and-carry model. Several examples are provided for calculating the fair value of equity index futures contracts, given the cash index price. Strategies discussed for equity indexes include program trading, growth versus value, trading sectors against broad-based indexes, and correlation trading against other indexes and ETFs.

The third and final section of the book surveys some recent developments, touches on value at risk (VaR) as a framework for risk management, discusses gold markets, considers market volatility, and examines some of the more important criticisms behavioral finance has leveled at the neoclassical model. Finally, it offers a summary and suggests that there is an art to trading that combines analysis and interpretation.

SECTION I

The Development of Modern Capital Markets

This page intentionally left blank

CHAPTER 1

Mapping the Market

> Models are to be used, not believed.
>
> –*H. Theil, Principles of Econometrics*

Models are devices we use to simplify reality into bite-sized pieces in order to make sense of it. They provide a way of filtering out background noise so we can focus on the main event. Consider the famous map of the London underground. It depicts the system as an oblong grid, and over the years it has efficiently guided millions of people to their destinations around the city. But it bears almost no geographical relationship to the actual physical layout of the system. That is not its purpose. Its purpose is to show you how to get around London by using the underground. Using it as an above-ground map will only succeed it getting you hopelessly lost.

This book begins with a model designed to help traders navigate financial markets. It is different from many conventional models that use advanced mathematics and the physical sciences (particularly physics) to describe market behavior, although it makes liberal use of the tools and theories that have guided much of their development. An important difference in this model is the use of qualitative variables and feedback loops that chart flows of information, market response, and subsequent reevaluation. Using these types of qualitative data can add nuance to descriptions of human behavior that mathematically sophisticated models can easily overlook.

To develop this idea, first imagine a market as a physical system of impulses and reactions. Flip the switch on, and the light goes on. Turn off the switch, and the light goes off. That's deterministic, and no one argues that market behavior can be turned on and off like that. But change the example just a bit. Imagine a diver dragging a balloon full of air down below the surface of a deep lake. Eventually the balloon will burst, allowing the air to escape. Released from its prison, the air will form hundreds if not thousands of bubbles that will make their way to the surface. But they won't all take exactly the same path and they won't make it to the surface at exactly the same time. The random swirls, eddies, and currents in the water will see to that.

Nevertheless, physicists can statistically estimate the approximate path the collection of bubbles will take on its ride to the surface, and the time it will take to get there. Think of the lake as a metaphor for the marketplace and think of the bubbles as individual securities. Suddenly the mathematics that physicists use to estimate the path the bubble collection takes can be used to estimate the path that securities prices, on average, are likely to take if certain conditions are met. And the individual bubbles can be handicapped, as in a horse race, once enough is known about them. This type of modeling is growing rapidly enough that its adherents refer to it (only half tongue-in-cheek) as *econophysics.*

There is a problem here. Bubbles rise to the surface because of the laws of physics. The bubbles don't have a lot of say in the matter. Securities prices move for a very different reason. People choose to buy and sell them. And now it gets complicated because human behavior is infinitely more complex than any physical process. Economists deal with the complexity of human behavior by making some powerful simplifying assumptions: namely, that humans are rational, self-serving beings who seek to maximize pleasure and avoid pain. In the parlance of economists, humans are "subjective utility-maximizers."

With that idea in mind the French mathematician Bachelier observed the workings of the bond market on the French bourse in 1900 and asked the question, why do bond prices move up and down? Bachelier researched the question for his dissertation (which went undiscovered for 50 years) and asserted that sellers wished to sell at the highest possible price while buyers wanted to buy at the lowest possible price. A seller who thought the price was too high would sell immediately at the bid price in the market. Likewise, a buyer who thought prices were too low would buy immediately at the price offered in the market. Therefore, Bachelier reasoned, the bid/offer spread in the market had to represent the market's collective assessment of the value of the securities. Since rational investors buy securities to earn profits, expected future returns are embedded in the price.

Bachelier's insight led to the neoclassical revolution in finance 50 years later that still dominates thinking about financial markets today. The market's price represents the collective wisdom of all its participants. Since the market price represents all known information, it takes new information to change the price. Well, not exactly. Some people trade on the basis of misinformation, rumor, or faulty reasoning. They are noise traders. The smart traders will, in theory, outwit the noise traders and in the end, prices will wind up pretty much where they are supposed to be. The market represents fair value. That gives us a model:

Market Prices = Expected Returns + Noise

The model can be easily diagramed and analogized as a physical process in which prices react to changes in expected returns due to a combination of the arrival of new information and noise trading. New information can be thought of as an impulse that causes a reaction. Noise trading is fluff or static that eventually dissipates. When the information is clearly, quickly, and widely disseminated the market reprices quickly and accurately. In such a case the signal-to-noise ratio is high. See Figure 1.1.

Models of financial markets (like the simple one above) tend to depict markets as information processing machines. It is a powerful metaphor, and it describes a core function that financial markets do perform. But there is more to it than that. It has to do with what we mean by information. *Information* is one of the more abused words in the English language. *Merriam-Webster's* dictionary defines it as "the communication or reception of knowledge or intelligence." But that doesn't really tell us much. Context is all important. Suppose a friend calls you from his mobile phone and tells you that it's cold outside. What information do you have that you didn't have before? Not much. You don't know what the temperature is. You just know your friend thinks it is cold. Suppose he is calling from Singapore, from right on the equator where it's usually about 95 degrees with virtually no seasonal change. Maybe he means that the temperature dipped to 89 degrees. An Eskimo would probably think that was kind of toasty. If your friend had told you it was 89 degrees, you would have a different kind of information. What does that have to do with the capital markets? A lot. Information, even when it is presented in factual form, is contingent, subject to revision and interpretation. Unfortunately, your profit and loss (P&L) isn't.

Suppose you know without a doubt that tomorrow the Fed is going to raise overnight interest rates. What will long-term Treasury bonds do after the Fed makes its move? Well, that depends. It is likely that Treasury bond rates will rise and prices fall, because that is what they have tended to do in the past in the wake of a tightening in monetary policy. But it is by no means certain. It depends on context. Was the move expected, or

FIGURE 1.1

Basic Impulse-Response Model of a Financial Market

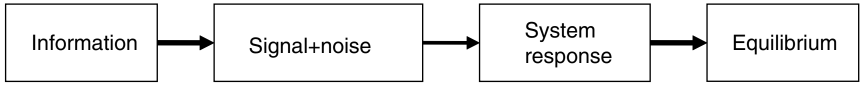
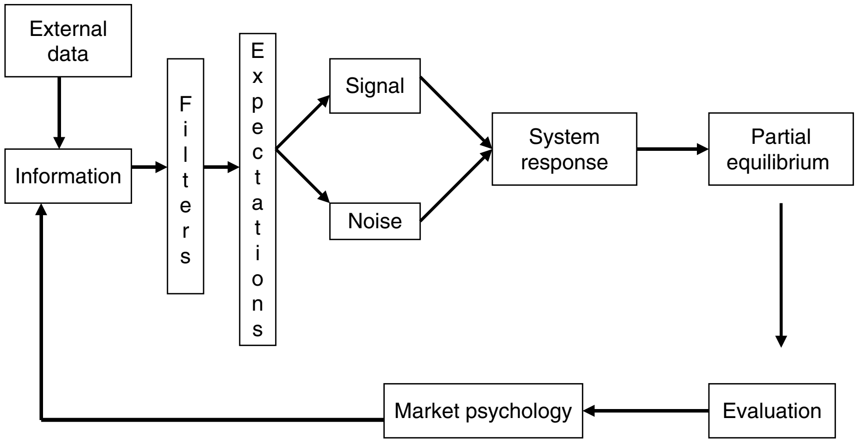


FIGURE 1.2

Complex Model of a Financial Market



was it a surprise? Is it a change in the direction of policy, or possibly the last of a long series? And how do we know?

Context is all important in the interpretation of facts. So are feedback loops, which are one way to put events in context. Add feedback loops and interpretative filters to the simple impulse-response model diagramed in Figure 1.1 and it becomes a lot richer—and a lot more complicated. Consider the model as it is redrawn in Figure 1.2.

Contrast Model 2 with Model 1. The world portrayed in Figure 1.1 is one that consists of discrete events and responses. Light bulb on, light bulb off. Now consider Model 2. New information comes into the market. The information is interpreted through the filters used by market participants; and it provokes an initial response. Some of the initial reaction is noise, some of it signal. How do we know which is which? Was the new information good news or bad news? If the market goes up, we might be inclined to interpret the news as being good. But maybe the news was actually bad, except not as bad as market expectations, so the market responds by going up instead of down, at least in the short term. Two pieces of information actually wind up finding their way into the market. The first is the original piece of data; the second is the market's reaction to it.

The model depicted in Figure 1.1 gives us nothing but the short term, because it posits that prices rapidly, accurately, and mechanistically reflect all known information. Further, it is a closed system at rest, lying inert until new stimuli cause the system to react. The form of that stimulus is new information that comes from outside the system. Moreover, Model 1 tacitly assumes that new information comes at the market in random fashion, that each piece of information is independent of the others. No one really believes that. Information arrives along a time line that represents trends in the business cycle. Good news tends to be followed by good news and vice versa until the cycle turns. Good and bad news tend to be autocorrelated in the parlance of statistical theory.

On the other hand, the market model sketched out in Figure 1.2 is dynamic, not static. It is like a whirling dervish, constantly being buffeted by new information, constantly reacting to it, constantly reevaluating it, constantly learning from it. Further, market reactions are contingent, subject to revision. The market is adaptive; people learn from it and adjust their behavior. Altogether this seems a far more realistic description of the world than the one described in Figure 1.1.

MARKETS ARE CHANGE AGENTS

The idea that traders learn from the market's response to events is a critical and probably an underappreciated fact. For one, it implies that the market's response to the same initial information changes with context. More to the point, it implies that everyone learns (or can learn) from market behavior, including policy makers, consumers, and anyone with a stake in the game. This in turn implies that people of all stripes—civilians, traders, and policy makers alike—change their behavior in response to markets.

To the extent that people's expectations and behavior change along with financial markets, it is clear that there are spillover effects that extend beyond the narrow confines of organized markets. Just as the law is a teacher, price signals from financial markets are a form of communication. Policy makers, portfolio managers, traders, consumers, and businesses respond to those signals, and their responses feed back into the system. For instance, housing markets respond to changes in interest rates. Consumer spending responds to the wealth effect of a rising stock market.

Correctly anticipating reactions to market price signals requires understanding the context in which they are embedded and interpreted. Describing and interpreting context, not a traditional strength of economics, depends in large measure on the filters one looks through. For example, economists tend to look at the world through the filter of self-centered individual behavior. Preferences are accepted as given. Behavior changes as incentives are altered.

Other disciplines have a different take on this. Anthropology and sociology concern themselves with how preferences are formed. Cultural anthropologists in particular are prone to argue that men are born as blank slates whose tastes, preferences, and values are largely determined by the culture into which they are born. Sociologists are likely to see preference formation strongly influenced by a society's institutional arrangements. Recent developments in neuroscience seem to indicate that the cognitive development, perhaps including risk-taking propensity, may be surprisingly susceptible to environmental factors. The point is that there are many filters to look through; no one filter is always dominant.

No matter how you slice it, it seems clear that traders need to be able to size up the market environment, or market psychology, to successfully forecast market behavior. But market environments are hardly static. They change as institutions, theory, and practice adapt to new developments. Sometimes changes are gradual, almost imperceptible; sometimes they are sudden and dramatic. But they are not likely to be random events; they are more likely than not to be predictable consequences of economic, political, and social decision making. They have causes. Understanding those causes is a key to successful trading.

THE ROAD TO MODERNITY

Today when traders casually pull out their mobile phone to check e-mail and the latest quote in a stock or a bond, it probably doesn't occur to them that the telephone was only invented 130 years ago. Financial markets as we know them today are the product of revolutions in science, technology, politics, and economics. They have been transformed by the way societies are organized and governed, the way risk and efficiency are understood, and the way transactions are organized. That series of transformations is the conceptual framework this book uses to analyze how modern capital markets function.

The story of modern capital markets is a complicated evolution of ideas and institutions. The thread that runs through the story is the sometimes gradual and sometimes rapid dismantling of the instruments of command and control in favor of the institutions of freedom and choice. It begins with classical Smithean economics, local markets, family farms, and the gold standard. At the beginning of the 21st century, the neoclassical economics of the Chicago School dominates market thinking. Markets are global and national economies are giving way to regional ones. The most successful firms are increasingly likely to be knowledge-based organizations that provide professional services. Perhaps the most important part of the story is the demise of the gold standard, ultimately to be replaced by global capital markets that trade freely across national borders and provide a measure of external discipline on the system.

The road to modernity, which is another way of describing market liberalism, has not been an easy path. It is strewn with the wreckage of two world wars, several midsize ones, the costs of a cold war, bouts of

FIGURE 1.3

Financial markets and the road to modernity: changing filters

Time line	1860–1914	Transition	1944–1971	Transition	1980+
U.S. economy	Agricultural		Industrial		Service
Output	Food		Manufactured goods		Knowledge
Firm organization	Family		Corporation		Professional services
Authority structure	Tradition		Bureaucracy		Network
Social control	Culture / community		Government		Bargaining / cooperation
Economic paradigm	Classical		Keynesian		Neoclassical
Financial markets	Local	→	National	→	Global
Regulatory regime	None / Laissez-faire		Strict / Interventionist		Light / market oriented
Monetary policy	Gold standard		Fiat money /fixed rates		Floating exchange rates
Prices	Deflation		Inflation		Disinflation

depression, deflation, recession, and inflation. And yet it emerged anyway as free societies and liberal institutions, primarily in the West and later Japan, adapted to changing circumstances. There is still a long road to travel; we have not reached the end history, *pace* Francis Fukayama. Nevertheless, it is worth noting that financial markets have more often than not led the charge, forcing difficult but necessary reforms. One way to see this is to look at a multilayered map of the journey thus far, beginning with the first era of global trade before World War I, up to the present, as displayed in Figure 1.3.

Figure 1.3 is a matrix designed to present a bird's-eye view of the sweeping economic, social, and political changes that transformed the developed world from the agrarian economies of the 19th century to the knowledge-based economies of the 21st century. Note that the matrix can be read either across the rows or down the columns. Read across the rows and it depicts changes over time in key areas of social organization and economic output. Read down the columns and it highlights institutions and structures that tended to mold and inform the way people of the era thought. Going forward that will be important, because the way people think about things—the filters they look through—powerfully influence the way they respond to markets. Usury laws, for example, rested on the misconception of capital as a static asset as opposed to a wealth-producing asset. Another example: For a long time financial analysts thought that dividend yields should exceed bond yields since stocks were riskier. It is a view that sees the world in static rather than dynamic terms.

The matrix displayed in Figure 1.3 is a road map to the modern era. It is divided into three distinct eras, separated by two transition periods. The first transition period is the time encompassed by two world wars, the Great Depression, and the establishment of the modern administrative welfare state. By the time it is over, the United States has become an urban, industrial society. It is also bureaucratic, hierarchical, dominated by interventionist government, Keynesian economics, large oligopolistic corporations, and economies that, by design, mostly stay within national borders. There are for instance, three major network television stations, three dominant automobile firms. AT&T *is* the telephone business. The seven sisters dominate the oil patch; RJ Reynolds and Phillip Morris control most of the tobacco market. IBM monopolized computing by making it impossible to hook its mainframes up with other machines. Wearing the corporate uniform of gray suits and white shirts and with THINK signs on their desks, IBM's middle managers are the prototype for William Whyte's *The Organization Man*.

That era came unglued in the 1970s. The 1970s served as a transition period to the global liberalism that dominates finance, financial markets, and trading today. The most profitable firms sell professional services rather than things. Hardware is just as likely to have to conform to software specifications as the other way around. Networking and connectivity turned out to be the key to success. Light regulation, flat organizations, strategic partnerships, bargaining, and cooperation supplanted the ideology of command and control. What accounts for this stunning transition that turned the game on its head? Mostly it is a consequence of what Josef Schumpeter called the "gale force winds of creative destruction." Postwar political and economic structures designed for a simpler era could not cope with the evolution of Western nations into a complex of knowledge and information-based societies. Earlier economic and political structures, agreed to at Breton Woods in 1944, were designed to permit Western governments to manage domestic finances without pressure from international markets. The United States agreed to be the guarantor of the system through an elaborate system of pegged currency and interest rates, which were tied to a U.S. pledge to convert the dollar holdings of foreign governments into gold upon request at the fixed price of $35 an ounce.

Price fixing inevitably fails, and the Breton Woods agreement is no exception. It finally collapsed under its own weight in August of 1971. The subsequent policy disasters of the 1970s had the benefit of teaching us what not to do and eventually led to corrective measures such as the freeing up of currency and interest rate markets as well as substantial deregulation of commercial and investment banking. Out of the rubble emerged a new respect for the free capital markets that are the basis of today's global trade regime.

CHAPTER 2

The Politics of Finance

Thou shalt not crucify mankind on a cross of gold.

–William Jennings Bryan at the 1896 Democratic Convention in New York

With the publication of *The Wealth of Nations* in 1776, Adam Smith launched a blistering attack on mercantilism and laid the foundation for Ricardo's subsequent development of the theory of comparative advantage.[1] Smith, moreover, did not shrink from the political implications of his argument for free trade. As America is declaring its independence, Smith asserts in *The Wealth of Nations,* British imperialism is not worth the cost. He writes:

> A great empire has been established for the sole purpose of raising up a nation of customers who should be obliged to buy from the shops of our different producers all the goods with which these could supply them. For the sake of that little enhancement of price which this monopoly might afford our producers, the home-consumers have been burdened with the whole expense of maintaining and defending that empire. For this purpose, and for this purpose only, in the two last wars, more than a hundred and seventy millions has been contracted over and above all that had been expended for the same purpose in former wars. The interest of this debt alone is not only greater than the whole extraordinary profit, which, it ever could be pretended, was made by the monopoly of the colony trade, but than the whole value of that trade, or than the whole value of the goods, which at an average have been annually exported to the colonies.[2]

And so from the beginning, political and economic thought, institutions, and practice have been inextricably intertwined. Well-functioning markets depend on the rule of law, political stability, sensible regulation, and competent execution of monetary policy. These functions require institutions that are stable yet adaptive, possess technical capacity, and have political legitimacy. They are the framework within which markets function. A proper analysis of the capital markets therefore depends on understanding the structure and evolution of their practices and institutions. Market structures, both formal (like the regulatory regime) and informal (tradition and culture), shape the incentives faced by the players. Accordingly, they exert a powerful influence on market behavior. Understanding how these structures evolved sheds light on both the system's embedded values and its adaptability. Sometimes adaptation is smooth; sometimes it is disruptive.

Financial market tipping points occur when the system can no longer reconcile either its embedded values or its incentive structures (or both) with the reality of the facts on the ground. In this regard modern capital markets, borne of necessity, are the embodiment of a clash between the logic and power of market liberalism and what Hayek called "the fatal conceit: the urge for central planning and control." The supreme irony is that the emergence of a global market liberalism evolved from frantic efforts to resist its inexorable logic. It is the story of how global trade evolved from informal local markets to highly regulated national markets, and from there to lightly regulated global markets.

Along the way disruptive events have torn apart the conventional wisdom, sometimes with cataclysmic financial market reactions, ultimately leading to structural change. Events of this type, some of which will be discussed in detail later, are often described as random shocks to the system. This book disagrees. Systemic shocks have causes; they don't just fall out of the sky. They are the result of pent-up imbalances, that once unleashed, throw the system's equilibrium out of balance. In this respect, financial market upheavals are, on occasion, a way the system reacts to rapidly changing circumstances. Sometimes these events lead change; sometimes they force change. Some of the truly great traders have made their fortunes anticipating these events and positioning themselves accordingly. George Soros's breaking the Bank of England comes to mind.

For several reasons, the focal point for analyzing financial markets is government economic and financial policy. Governments are dominant players in the capital markets. They are enormous borrowers, using sales of government bonds to finance budget deficits. They also play the role of umpire, promulgating and enforcing financial market regulatory regimes. They strongly influence interest rates and inflation through the conduct of monetary policy. Private sector borrowers and lenders operate in governments' wake. They compete for funds only after government has claimed its share of the savings pool. But governments are neither unitary nor disinterested parties. They contain varied and competing interests. And they can enhance welfare by encouraging market efficiency, or they can harm citizen well-being with heavy-handed intervention in the marketplace.

Policy matters, and it matters a lot. It can attract or repel capital and investment.

In financial markets the coin of the realm is returns; in politics it is power. Modern capital markets mediate competing claims for money and power among governments and businesses, savers and borrowers. Just as it did in Smith's day, the ongoing struggle to establish and maintain free markets pits the logic of returns against the logic of power. Market liberalism is based on free choice, property rights, and the rule of law. Rival systems are instinctively coercive, depend on centralized planning, and tend toward authoritarianism. Unlike Smith's day, capital mobility now serves as a constraint on government power just as market competition disciplines businesses. Poorly conceived policy can send capital fleeing across borders at the click of a mouse button. The upshot is that capital is priced globally, not locally. That is a relatively new development. Unfortunately it took a string of catastrophic policy failures, the worst of which occurred in the 1930s and 1970s, to set the stage for the ascendance of market liberalism over the economics of command and control. Mostly the battle was fought out in the realm of monetary policy.

MONEY, MARKETS, AND MODERNITY

At the beginning of the 20th century the United States was an agrarian nation with about 60% of the population living in rural areas. Agrarian societies tend to be relatively static compared to industrial ones which tend to be dynamic and adaptive. The United States was no exception. Moreover, it was relatively isolated, protected by two oceans. But by the end of the century the United States had emerged from relative isolation to become the world's greatest industrial power. The transformation from an agrarian to an industrial economy produced profound social and structural changes. Douglas North cites unprecedented population growth, large gains in standards of living, rapid urbanization, increasing specialization and interdependence, with technological change becoming the norm.[3]

As the United States transitioned from an agricultural economy dominated by crop cycles to an industrial economy dominated by product development cycles, financial markets transitioned as well. Banks began to concentrate on providing financial services to meet the demands of the emerging industrial base. Government became increasingly interventionist and developed a voracious appetite for financing, both to fight wars and later to administer a vast new regulatory state. This required a broad and deep bond market through which the government could borrow in lieu of directly raising taxes.

To facilitate its funding needs, the U.S. government took active steps to develop a national bond market, placing itself first on line for financing.

The result is the dealer market we know today, which developed so successfully that it is now widely emulated around the world, often with technical assistance provided by the U.S. Treasury and its agencies. Partly as a consequence, by the end of the 20th century the bond markets had become truly global, with something close to standardized trading and settlement practices. They can be easily tapped by national governments, supranational agencies, quasi-governmental organizations, and transnational corporations to meet financing requirements. They bear little resemblance to the markets of yesterday.

Today's modern trader would scarcely recognize the capital markets as they existed at the turn of the last century. On a typical day in 1900, the New York Stock Exchange (NYSE) traded about 500,000 shares of stock and $2 million worth of bonds. By way of comparison, by the fourth quarter of 2006 over a trillion dollars' worth of Treasury, corporate, agency, and mortgage securities could easily change hands, not to mention about 1.5 billion shares of stock on the Big Board worth over $65 billion.[4]

In 1900 the Federal Reserve had not yet been created, nor was there an income tax. Like Western Europe, the United States was on the gold standard, where it remained until the Great Depression of the 1930s. The financial system was not very stable, and it had to cope with periodic bouts of deflation, rather than the inflation that proved so troublesome in the post-World War II period. That may explain the inflationary bias of post-World War II policy making that proved to be the undoing of Keynesian economics.

The post-Civil War era was one of falling prices, a dynamic that tended to favor Eastern commercial and financial interests over the agrarian Midwest. Populist William Jennings Bryan placed the blame for declining farm prices on The Coinage Act—the so-called Crime of 1873—which had put the United States back onto the gold standard from gold/silver bimetallism.[5] Seeking to increase the money supply to inflate farm prices, Bryan called for the free coinage of silver at a ratio of 16:1.

At the 1896 Democratic convention in New York Bryan delivered his famous cross of gold speech, directing his wrath at hard-money Eastern bankers, whom he portrayed as enemies of the common man. The speech, which is still regarded as a classic of the genre, captivated the convention delegates and secured for Bryan the Democratic party's nomination for president. But he was crushed by the Republican's William McKinley, who was backed by Wall Street, in the general election.

The pivotal election of 1896 was a cultural as well as economic contest between rural-agrarian interests and values and urban-commercial ones, ironically memorialized by L. Frank Baum in his classic children's tale, *The Wonderful Wizard of Oz.* In Oz (the abbreviation for ounce is oz) Dorothy follows the hazardous yellow (read gold) brick road to the

Emerald City (Washington) naively seeking help. She is tormented by the Wicked Witch of the East, a reference to Wall Street bankers and industrialists who exert control over the common people (the Munchkins). The straw man is the helpless farmer, the tin man the dehumanized factory worker, the cowardly lion William Jennings Bryan. The Wizard is perhaps Mark Hanna, McKinley's campaign manager. Originally, Dorothy makes her way home in silver slippers—although Judy Garland wore red ones for the movie.[6]

Munchkins aside, sectional interests dominated politics. The Republicans were the party of the urban, manufacturing North, and the Democrats were the party of the rural, agricultural South. The Republicans relied on high tariffs, which protected their manufacturing political base from foreign competition and provided revenues that could be used to finance public works projects designed to buy off the opposition. Not surprisingly, spending outpaced revenues, leading to demands for higher tariffs to cover the resulting deficits, a familiar sounding refrain today. Under McKinley, the tariff rate ultimately reached a level of 50%.[7]

The charade could only go on so long, and the internal contradictions of the high-tariff regime ultimately led to its collapse. The tariff could never keep up with the built-in spending demands it generated. Moreover, the progressive movement, which was gaining power, came to the realization that the tariff amounted to a tax on the agrarian poor to subsidize the urban rich. Progressives began to campaign against high tariffs, demanding a national income tax as a substitute, a move that was not exactly popular on Wall Street.

In the end the progressives could not finish off the power brokers of the Eastern banking establishment; nor did they manage to restructure the financial system. That honor belonged to the Banking Panic of 1907, which eventually led to the creation of the Federal Reserve and the beginning of a national financial system regulated by Washington. Banking panics had come and gone before, but they had mostly affected regional banks. The panic of 1907 erupted in the heart of Wall Street. After that it was time to change the rules of the game.

THE BANKING PANIC OF 1907

In 1907 F. Augustus Heinz attempted to corner the stock of United Copper. He failed. In the implosion that followed, hidden financial arrangements between prominent New York City banks, brokerage houses, and trust companies were revealed. For depositors, this was particularly disconcerting because banks had the reputation of being solid, unlike their roguish cousins, the brokerage houses, which were thought to be shaky. There was no such thing as deposit insurance, so when the tie between banks and brokerages came to light, depositors raced to get their money out of banks thought to be involved with Heinz. Inevitably rumors began over which banks were associated with Heinz. As the rumors spread, depositors ran from bank to bank to be first on line to get their money out, hence the term *bank run.*

This was a devastating development. Financial markets and institutions depend on confidence for their survival. And confidence was collapsing. Moreover this was no run-of-the-mill panic. It was the first ever to take place in New York City, the very heart of the financial system. J. P. Morgan eventually decided out of necessity to arrange a rescue; the system had been brought right to the edge. Before it was over Morgan Bank, Kuhn Loeb, J. D. Rockefeller and the U.S. Treasury each had to kick in deposits to diffuse the panic and keep the system going. But it was a close call. The New York banks were sufficiently unnerved by the experience that they decided that, for the future, they would abandon their traditional role as lender of last resort. The absence of a de facto lender of last resort made plain the need for a de jure one, paving the way for the creation of the Federal Reserve System.

Bank panic or no, structural reform of the financial system still had to wait for the elections of 1912. Woodrow Wilson won the presidency for the Democrats with 41% of the vote after William Howard Taft (the incumbent Republican) and Teddy Roosevelt (running as the Bull Moose Party candidate) split the Republican vote. The Democrats also captured the Senate and greatly expanded their majority in the House. In the aftermath, Congress passed the Federal Reserve Act on December 23, 1913, creating the central bank; it passed the Underwood tariff on October 3, 1913, replacing the high protective tariffs of the Republican era, and on February 3, 1913, the 16th amendment to the Constitution was ratified allowing for the imposition of a progressive income tax.

TRADE AND WILSONIAN LIBERALISM

Woodrow Wilson, one of America's most consequential presidents, is in many ways the unrecognized architect of today's global trading system. Wilson brought the United States out of its isolationist tradition and into the heart of Europe (and its wars). But Wilson rejected the power politics of the European Realist School. He instead asserted universal law and trustworthiness to be the foundation of international order.[8] The language of law, trustworthiness, and contracts is the language of market liberalism. It stands in opposition to the zero-sum language of mercantilism that sees trade in terms of winners and losers, as power gained and power lost.

The Wilsonian vision was (and is still today) breathtaking in its sweep. A strong proponent of American exceptionalism, Wilson maintained that American values were universal ones and that America served as a beacon of freedom to peoples around the world. He rejected the idea that the State had an independent claim on morality and asserted the inherent superiority of the foreign policies of democratic states. He argued that America's security was dependent on the security of all mankind, and he implied it was America's obligation to oppose aggression everywhere.

When Wilson presented his 14 points to the Congress, he made clear his conviction that the United States had entered World War I (the Great War) not for the acquisition of power and position, but to secure a new liberal order. He claimed a clear principle ran through his program. It was "the principle of justice to all peoples and nationalities, and their right to live on equal terms of liberty and safety with one another, whether they be strong or weak."[9] Free trade was a critical component of Wilson's new world order. The second of his 14 points called for freedom of the seas. The third called for "removal, so far as was possible, of all economic barriers and the establishment of an equality of trade conditions among all nations consenting to the peace and associating themselves for its maintenance."[10]

Not surprisingly, Wilson's language was anti-imperialist. Nation's should be free to determine their own institutions and destinies; the will of the people must be given equal weight when questions of sovereignty were to be decided. Rhetorically, he placed the power and prestige of the United States in opposition to Europe's balance-of-power mercantilist politics. In rectifying wrongs and asserting right, the United States would be "intimate partners of all the governments and peoples associated together against the Imperialists."[11]

The Wilsonian internationalist vision of democracy and freedom, dependent on national institutions and the will of the people, had its roots in his analysis of the structure of domestic politics. Wilson had little use for Madisonian checks and balances and the separation of powers. His concern was not with the concentration of power; it was with power used irresponsibly. Further, he thought that when power was distributed widely, it masked accountability, and so invited irresponsible use.

For Wilson, politics and administration were separable. The will of the people would be found in the Congress. The purpose of the administration was to enforce the law, carrying out Congressional intent. Wilson wished for the government to be efficient. The science of administration he said should be "to straighten the paths of government, to make its business less unbusinesslike and to strengthen and purify its organization."[12]

The Wilsonian quest for efficiency in administration laid the philosophical groundwork for the 20th century's administrative state and its centralizing tendencies. The emphasis would be on management rather than checks and balances. Freedom would be served by accountability rather than the separation of powers and the pull and haul of politics. Government would be unleashed rather then constrained. For Wilson, the public administration was much more than technical expertise, more than the mere machinery of government. The public administration's greater principles directly connected it "with the lasting maxims of political wisdom, the permanent truths of political progress." The object of administrative study, he said, "is to rescue the executive from the confusion and costliness of empirical experiment and set them upon foundations laid deep in stable principle."[13]

And so here we have the two Wilsons who would ultimately crash against each other in the last quarter of the 20th century when the markets laid bare the contradiction between the top-down centralization of the administrative state and the animal spirits of global capitalism. The inherent contradiction between the dynamism of market liberalism and the innate conservatism of the administrative state was submerged, for a time, by World War II and the postwar financial architecture put in place in the war's aftermath. A rise in the price of gold ultimately brought it to the surface—compliments of the Federal Reserve.

THE FEDERAL RESERVE

The Federal Reserve—the Fed as it is known on the Street—is one of the most powerful institutions in the American (and hence global) political economy. Any real understanding of U.S. capital markets begins with the Fed, its structure and its role in the American economic system. When the Fed acts (or as the case may be, fails to act), the effects reverberate throughout the system. Banks, capital markets, political markets, and, ultimately, the real economy are all affected. But power should not be confused with omniscience. The Fed was instrumental in producing two policy disasters whose effects are with us today, decades later. The first was the deflation of the 1930s that turned a mild recession into to the Great Depression. The second was the great inflation of the 1970s.

When it comes to policy making, the Fed has often been described as knowing only two speeds: fast forward and reverse. Surprisingly, it turns out that the success or failure of Fed policy making can depend not only on policy maker's economic views but also on the chairman's leadership abilities—or lack thereof. As it happens, a series of catastrophic Fed policy errors was responsible for turning a mild cyclical downturn into the Great Depression of the 1930s, setting off a chain of events that remain important forces shaping today's political economy and capital markets.

The devastation of the Great Depression was qualitatively different from previous downturns. Its depth and breadth changed the dominant view of the proper role of government and the efficacy of markets. It created the conditions that allowed Franklin D. Roosevelt to build the administrative state envisioned by Woodrow Wilson almost a half-century earlier. FDR's New Deal was built around an expansive government and an alphabet soup of executive agencies. Power was vested in the executive branch as never before. Constraints on government power were greatly loosened, and the bureaucracy began to actively intervene in markets. The Wilsonian vision of executive power and accountability through the chain of command, financed by progressive income taxation, was put into play.

A powerful myth, that capitalism had to be saved from the capitalists, grew out of the New Deal. It was (and is still) taken as an article of faith that the New Deal was primarily responsible for getting the country back on its feet; proof positive that markets need an elaborate regulatory apparatus to function efficiently and that government can "manage" the economy. Central planning and bureaucratic rationality came into vogue. Fiscal policy reigned supreme. Pump priming, deficit spending, jobs programs, and the like became the policy tools of choice. Monetary policy, it was thought, didn't matter, or at least it didn't matter very much.

That idea turned out to be spectacularly mistaken. But it took a long time for scholars to realize that the severity of the Great Depression was directly attributable to Federal Reserve policy. And it was not until the 1970s that the regulatory apparatus of the New Deal began to collapse, paving the way for free markets to allocate capital to its highest and best use.

The thesis that the Fed was largely responsible for the Great Depression was first advanced by Milton Friedman and Anna Schwartz, buttressed later by Christina Romer. Friedman and Schwartz argued that the severity of the Great Depression was due to the Fed's overly restrictive monetary policies. As they put it in their authoritative *Monetary History of the United States,* "No other contraction before or since has been preceded by such a long period over which the money stock failed to rise. Monetary behavior during the contraction itself is even more striking. From the cyclical peak in August 1929 to the cyclical trough in 1933, the stock of money fell by over a third."[14]

The Depression was marked by a series of banking crises and failures that had the effect of reducing the money supply. But even as banks were failing all around it, the Fed failed to take action to offset the outflow of reserves from the system. Depositors, fearful for the safety of their bank accounts, began to line up to withdraw their funds. Banks in turn had to scramble for liquidity to convert deposits into cash. When banks could not generate the liquidity necessary to satisfy depositors' claims, they simply failed.

The process fed on itself. As banks failed, confidence in the system plummeted. The public continued to withdraw funds from the banks and to quite literally put the money under the mattress. Currency in circulation increased, and demand deposits shrank. The banking system was based on fractional reserves, so when currency demand increased at the expense of demand deposits, reserves were depleted, and the money supply fell.

Friedman and Schwartz note that by today's standards the Fed's behavior is almost inconceivable. The Fed failed to inject sufficient reserves into the system to replace those destroyed when depositors took their money out of the banks. With banks collapsing around them, depositors figured that their money was safer under the mattress than in the bank. Note that confidence was being destroyed in the banks, not in the value of the currency. Even so, the stronger banks in the private banking system failed to backstop the system, a function they had previously performed. They assumed that the Federal Reserve would take confidence-building steps. In this they were mistaken.

The disaster was not confined to the United States. Runs began on sterling and forced England off the gold standard on September 21, 1931. Private investors began to worry that the United States would similarly abandon gold. They began to sell their dollar-denominated money market holdings. They were not alone. Foreign central banks began to sell huge quantities of banker's acceptances to the Fed. Upon sale they converted their dollars to gold and immediately exported it. This maneuver further depleted reserves from the banking system, depressing the money supply even more.

To offset the outflow of gold, the Fed tightened policy quickly and decisively, raising the discount rate from 1.5% to 3.5% over a few months' time. Raising the discount rate succeeded in stabilizing the gold supply, but at the cost of intensified strains on the banking system. The banks were being hit from all sides. Reserves were being depleted both by the outflow of gold and by the conversion of demand deposits into cash. They had to move quickly to generate enough liquidity to meet depositors' demands. But depositors, concerned about the safety of their funds, continued to demand currency, further depleting liquidity in the system.

The banks needed to generate liquidity, but they only had had two options. They could borrow from the Fed at the discount window, or they could sell assets to raise cash.

Borrowing at the discount window presented a dilemma. Depositors viewed heavy discount window borrowing as a sign of weakness. Consequently, they were afraid to leave deposits in banks that visited the window once too often. Furthermore, to borrow from the Fed, banks had to pledge assets which the Fed discounted at progressively higher rates. On the other hand, selling assets was a costly proposition. Yields were rising (and prices falling) on the assets they owned because of Fed tightening to stem the gold flow.

In the end the banks did whatever they could to keep their doors open, but a financial tsunami rolled over them. Failed banks littered the financial landscape like abandoned cars do in dicey neighborhoods; between August 1931 and January 1932 over 1,800 banks reportedly shut their doors, some temporarily, some for good.[15]

Why did the Fed fail to take corrective action as the banking system was collapsing around it? After all, one reason the Federal Reserve Act was passed was to avoid banking panics to begin with. Here the Fed just made matters worse by tightening policy rather than easing it.

Friedman and Schwartz say there are three principal explanations for the Fed's behavior. First, operational control of monetary policy shifted from the New York Fed to the Board of Governors in Washington. The New York Fed, with technical expertise in banking, was focused on the financial system. But the Board of Governors in Washington was focused on politics. Second, personnel is policy. After the 1928 death of the New York Fed Governor Benjamin Strong, there was a dearth of leadership, and policy was allowed to drift. Third, few people really understood what was happening. (A not-dissimilar pattern of events would play out again decades later in the 1970s, with weak Federal Reserve leadership and economic policy dominated by political calculations).

When it first came into being, the structure of the Fed was fairly decentralized. It consisted of the Board of Governors in Washington and 12 regional Federal Reserve banks, with the New York Fed being the most important. The 12 regional reserve banks were charged with regulating the member banks and the terms of credit provision in their districts. The Board in Washington oversaw the regional banks. At the outset, the functions of the Federal Reserve were couched largely in terms of technical efficiency. The Federal Reserve would provide elasticity that the current system could not provide. The preamble to the act establishing the Federal Reserve directed it ". . . to furnish an elastic currency, to afford the means of rediscounting commercial paper, to establish a more effective supervision of banking in the United States, and for other purposes."

The framers of the act assumed that credit extended by the Fed would rise and fall both with short-term seasonal (read agricultural) factors and with longer-term variations in demand related to changes in business activity. They were unconcerned with the potential inflationary consequences of the central bank's ability to accommodate increased credit demand. They reasoned that gold inflows and outflows would make the system self-correcting.

Even though the Fed was established in 1913, by 1929 it still had not developed standard operating procedures for dealing with financial shocks. It was unprepared for the October 29 market crash. One reason suggested by Friedman is that it probably did not fully understand the power it had to respond to cyclical economic events; for that matter, it was divided on whether it should be in the business of adopting countercyclical policies in the first place. Besides which, there was a power struggle under way between the regional Federal Reserve Banks and the Board of Governors in Washington.

Friedman and Schwartz argue that by 1930 structural changes at the Fed shifted decision-making power from the regional Banks (and in particular the New York Fed) to the Board of Governors in Washington. A new committee, the Open Market Policy Conference (OMPC), took over responsibility for policy making. The OMPC consisted of the members of the Federal Reserve Board and all 12 Federal Reserve Bank governors. The power to call meetings was taken away from the New York Fed and given to the Board of Governors in Washington. An executive committee of the OMPC met more frequently than the Board and coordinated with the New York Fed's trading desk for carrying out open market operations.

The Fed's new OMPC viewed the contraction of money and credit with relative equanimity. It considered weak money demand to be a natural market response to weak business conditions. The Fed's analysis made the crucial mistake of conflating the economic contraction on the real side of the economy with the liquidity and banking crisis on the financial side—two separable events. The real problem was that the money supply—over which the Fed had control—declined even faster than money demand, thus squeezing liquidity, causing rates to rise. As Friedman and Schwartz point out, this was uncharacteristic: The Fed had reacted to earlier banking panics in the 1920s by supplying liquidity to the system.

The difference was that New York Fed Governor Benjamin Strong had died in 1928. Strong had gained practical experience in the financial markets early in his career working as a commercial banker. He had also been an active player in the 1907 banking system rescue. That experience had taught him the danger of financial panics and the importance of nipping them in the bud to avoid contagion. The Fed under Strong knew what to do, and in the 1920s it moved rapidly to contain banking panics.

But by the time Strong had died, the Fed's policy-making apparatus had been radically altered. The New York Fed no longer dominated policy making; Washington did. Financial expertise gave way to politics and bureaucratic inertia. Policy simply drifted. The Fed's failure to act allowed the money supply to collapse, and with it the banking system. Unfortunately, the same combination of weak leadership and policy ineptitude would again lead the Fed into a series of policy disasters decades later.

That is a point that needs to be emphasized. As big, complex, and powerful as it is, any banking system retains some inherent fragility. It is based on trust and the leadership and policy-making skills of a relatively small number of people. It matters who the Fed chairman is; who the Treasury secretary is; what they think, and what their leadership abilities are. As Friedman and Schwartz put it, "The detailed story of every banking crisis in our history shows how much depends on the presence of one or more outstanding individuals willing to assume responsibility and leadership."[16]

REMAKING THE BANKING SYSTEM

The period immediately preceding the Depression was characterized by change across many facets of American life. The roaring twenties produced rapid economic growth, industrialization, and urbanization. Automobiles gave citizens new mobility. Agriculture diminished in importance. Small country banks discovered that the competitive landscape had shifted underneath them. New large industrial firms required sophisticated financial services beyond the capabilities of small country banks. So they turned to big city banks, which began to underwrite and distribute securities to finance the new firms. Commercial and investment banking still resided under one roof.

The Great Depression changed that. Congress adopted a wholesale transformation of the regulatory regime. The structure of the commercial banking system was changed; the powers of the Fed were changed, and monetary standards were changed as well. Additionally, financial institutions were changed (or created) as a result of the experience of the Depression. All the changes were aimed at promoting stability.

Friedman and Schwartz point to three developments as being particularly important. The first was the insuring of bank deposits through the creation of the Federal Deposit Insurance Corporation (FDIC). The second was the separation of commercial and investment banking mandated by the Glass-Steagall Act. The third was suspension of the gold standard, later partially replaced.

The FDIC was created as a corrective to a system flaw exposed by the Fed's failure to act as lender of last resort. As a result of the Fed's failure, banks had to scramble for liquidity on their own. As a practical matter they were forced to sell off their bond holdings as prices fell, or go to the window, or most likely, do both. Certainly not the type of scenario that is likely to assuage the nervousness of depositors concerned about the safety of their deposits. By assuring customers that their deposits were safe, the FDIC deposit guarantee was designed to maintain system stability.

Note that the FDIC did not guarantee the bank; it merely guaranteed the banks' deposits (up to certain amounts). With deposits insured, the reason for bank runs vanished. So did the reason for depositors to assess the quality of a bank. That job fell to the regulatory agencies charged with supervising the banking system.

An important part of the new regulatory regime turned out to be separating commercial and investment banking. It was (and still is) generally believed that the stock market crash of October 1929 caused the Great Depression. That belief was (and still is) bolstered by the image of the roaring twenties, complete with its speakeasies, flappers, and wildly extravagant financial market speculation, just as the eighties were mischaracterized as "the decade of greed."

In any event, the preferred solution was to put a ring fence around commercial banks to separate them from investment banks. One set of agencies was designated to regulate the commercial banks; another got the investment banks and brokers. Neither would be permitted to invade the other's turf. The FDIC, the Fed, the comptroller of the currency, the Treasury, and other agencies got various aspects of the banking business to regulate. The commercial banks were restricted to taking in deposits and making loans. On the investment banking side, the Securities and Exchange Commission was created to clean up and regulate the always suspect securities business. Probably on the theory that it takes one to know one, Joe Kennedy became the first chairman of the SEC.

The FDIC and the SEC were designed to cordon off two sources of potential instability in the banking system: contagion and speculation. The FDIC deposit guarantee was designed to prevent individual bank failures from spreading panic through the system. In addition, separating commercial and investment banking would stop banks from speculating in the securities markets with insured depositor's funds. In the meantime the SEC began to tear away at the veil of secrecy over securities markets by enforcing disclosure requirements. That left a third source of potential instability that wasn't so easy isolate: namely, international money and capital markets, anchored by the price of gold.

AS GOOD AS GOLD?

Before the Great Depression, accounts were settled in gold. Gold and currency were freely convertible at the official price. Citizens could demand gold for their dollars, and so could nation-states. International finance did not exist as we know it today because transactions were implicitly settled in gold. As long as national currencies were pegged to gold, the relative prices of currencies remained static. To attract capital, a nation simply paid higher interest rates and gold would flow into the country. With fractional reserve banking, the money supply would expand when the new reserves (in the form of gold) came into the system to accommodate increased borrowing demand.

Free convertibility of currency into gold posed a problem for the new administrative state, however. First, it severely limited the government's ability to inflate. Second, because the gold market was international in scope, people and institutions outside the jurisdiction of the United States could affect the money supply by importing or exporting gold into or out of their U.S.-based accounts. This was an anathema to the architects of the new administrative state. A free market in gold combined with currency convertibility placed a check on government power.

FDR's New Dealer's decided to address that issue head-on by abolishing the rights of citizens, banks, and other businesses to hold gold for monetary purposes. Under the authority vested in him by the Gold Reserve Act of 1934, President Franklin D. Roosevelt nationalized gold ownership. Banks were directed to turn over the gold they held in return for paper money. Roosevelt's justification was enunciated in one of his fireside chats. He claimed that, "Since there was not enough gold to pay all holders of gold obligations, the Government should in the interest of justice allow none to be paid in gold."[17]

But without currency convertibility there was no way to settle international accounts. And without a payments system there could be no international trading system. No one would accept pieces of paper as legal tender in the international system, at least without explicit backing in gold or silver. Wilson's vision of world peace based on a liberal order of freedom of the seas, free trade, anticolonialism, and multilateral institutions could not take root.

By the time the United States got around to outlawing the monetary possession of gold, the international trading system already lay in tatters, a casualty of the First World War and the Great Depression. Before the Great War trade flowed freely as did capital, and there were relatively few restrictions on immigration. The stunning destructiveness of World War I ended all that. To make matters worse, postwar protectionism and the tidal wave of bank failures that occurred during the Great Depression made it impossible to resurrect the world trading system, leading to yet another world war.

By the end of the Second World War, Europe and Japan lay in ruins. Economic liberalism was under attack, buffeted by war and depression. The supposed superiority of rational central planning over the chaos of markets gained currency among intellectuals. State socialism seemed to be ascendant in the East; Russia was firmly in Stalin's grip; China was torn by civil war between Mao's communists and Chiang's nationalists.

In the interwar period John Maynard Keynes had published *The Economic Consequences of the Peace,* in which he warned the victorious Allies that the harsh WW I peace terms they had imposed on Germany set the stage for more war. His prophecy having been borne out, the Western Allies of World War II vowed not to make the same mistake again. They set out to remake the postwar order by building a set of international institutions that would organize collective security, finance global development, and oversee international finance. The United Nations, the International Bank for Reconstruction and Development (the World Bank), and the International Monetary Fund (IMF) were born.

The architects of the postwar order needed to construct a payments system that allowed international trade and finance to flourish. At the same time they were determined not to allow themselves to once again become slaves of the gold standard—"the barbarous metal," as Keynes put it. That posed a problem. To work, the payments system had to be credible; the terms of trade could not be undermined after the fact by currency devaluations. The dilemma the architects faced (or thought that they faced) was to design a system that was credible without lashing domestic monetary policy to the external discipline of the gold standard.

In 1944 they gathered at Breton Woods to hammer out a new regime to regulate international financial markets. They tried to square the circle by creating a structure that imposed a quasi-external discipline on governments in their dealings with each other, but at the same time freed governments to follow domestic policies unconstrained by international financial considerations. The only way they could do that was to restrict the flow of financial capital across borders. And so the policy regime was set. Exchange rates were pegged within narrow bands; citizens were no longer permitted to own gold for monetary purposes, nor were citizens allowed to demand redemption of currency into gold. The United States would be the guarantor of the system and would exchange gold for currency with other governments at the official rate, then pegged at $35 per ounce.

The Breton Woods accords, as they became known, became the foundation of the postwar trade and finance regime that survived essentially intact until August of 1971, when President Nixon brought it crashing down by slamming the gold window shut. That event set off the chain of policy catastrophes that dominated the 1970s and marked the beginning of the end of postwar embedded liberalism. In his desperate attempt to dodge market discipline by severing gold and the dollar, Nixon unwittingly created the impetus for the global liberalism that dominates today's hypermobile capital markets.

THE BEGINNING OF THE END

Nixon's suspension of dollar/gold convertibility was accompanied by the decision to let the dollar float freely in foreign exchange markets. It also included a dollar devaluation expressed as an increase in the official price of gold. Normally, currency devaluations spark an instant rise in inflation. But the Nixon administration sought to put off that day of reckoning by simultaneously imposing wage and price controls. Fed chairman Arthur Burns, a Nixon loyalist, obligingly stepped on the monetary gas and aggressively eased policy. The overnight federal funds rate, which had been trading at around 5.5%, plunged to 3.5% over the next six months. The toxic combination of easy money, dollar devaluation, and wage and price controls set the stage for the great inflation of the 1970s, along with its commodity shortages, gas lines, and supposedly exogenous shocks, prominent among them, rapidly rising oil prices.

Nixon's decision to suspend (and later end) gold convertibility and to float the dollar had the effect of freeing domestic monetary policy from the constraints imposed on it by foreign exchange markets. Governments can choose between a fixed exchange rate policy and an independent monetary policy, but they cannot have both. If capital is mobile and exchange rates are fixed, the costs of finance will adjust until they are equalized across all countries. It is the law of one price. On the other hand, if capital is mobile and exchange rates are free to float, governments (or their central banks) can set short-term interest rates, but they will lose control over the value of their currency in world markets.

Declaring that "we are all Keynesians now," Nixon signed on to an ambitious and hopelessly doomed macroeconomic policy management scheme. At the time most policy makers mistakenly believed that the macroeconomy could be managed partly by paying attention to the Phillips curve, named after A.W. Phillips, a New Zealander who observed an inverse relationship between employment and inflation rates in 1950s Britain. Policy makers then proceeded onward to make the rather heroic assumption that the observed historical relationship between the unemployment and inflation rate could be used to fine-tune the economy. The inflation rate could be brought down at the cost of higher unemployment rates and vice versa. The policy question is which spot to choose along the curve.

In retrospect it seems hard to believe that this idea was taken seriously. But it was. More to the point, the whole idea is almost irresistible to politicians because it gives them—or at least appears to give them—actionable policy levers that can be used to secure reelection. In Nixon's case that goal was the clear policy driver. Nixon had lost an extraordinarily tight race against John F. Kennedy in his first run for the presidency in 1960. As Yergin and Stanislaw tell the story in *Commanding Heights,* Nixon was convinced that economic officials of the Eisenhower administration (whom he referred to as "financial types") had mismanaged the economy.[18] They had made the 1959 recession deeper and longer than necessary by putting too much emphasis on reining in inflation and too little on reducing employment. Nixon was determined to avoid a repeat performance; the "financial types" were not going to jeopardize his reelection.

Nor was previous opposition to increased government interference in the economy going to jeopardize his reelection prospects. Power was what mattered, not theoretical niceties about market efficiency. The Nixon administration came to see the economy "as organized by relations of power, status, rivalry and emulation."[19] Government intervention was necessary to balance the scales between increasingly powerful labor unions and large corporations. Cutting off dollar convertibility, floating the currency, and imposing wage and price controls allowed Nixon to argue that he was protecting the common man against price-gouging corporations, greedy labor unions, and foreign currency speculators.

In an all too familiar refrain, politicians and policy makers frequently attack markets in an attempt to shift the blame for their own bad policy choices. Traders, especially currency traders, make particularly tempting targets because so few people understand what they do. And so 75 years after William Jennings Bryan's famous cross of gold speech, Nixon picked up the baton and vowed to crush currency speculators by slamming the gold window shut.

By severing dollar/gold convertibility and allowing the dollar to float, the Nixon administration gained temporary domestic monetary independence. Without the constraints imposed by dollar convertibility, the Fed was free to ease policy. It did this with gusto, giving the economy a short-term boost and virtually ensuring the reelection of President Nixon to a second term. At the same time, the inflationary effects of too-easy monetary policy were disguised in the short term by the imposition of wage and price controls.

Combining wage and price controls with monetary ease creates a seductive elixir that is irresistible for the opportunistic politician facing reelection. In the short run it places more cash in the hands of the voting public while keeping a lid on prices. Longer term it is a surefire recipe for massive inflation, product shortages, rationing, financial instability, and fleeing capital. Which is precisely what the Nixon program produced.

There is some historical debate about whether Federal Reserve Chairman Arthur Burns cynically advocated an "incomes policy" (a euphemism for wage and price controls) in order to pursue an easy-money policy designed to reelect President Nixon. It is entirely possible that he was merely incompetent. Alan Reynolds of the Heartland Institute has made the point that terrible economic theories were ubiquitous at the time and that brand-name economists and policy makers, among them John Kenneth Galbraith and Treasury Secretary Connelly, were enthusiastic about wage and price controls.

In any event, the controls were implemented on August 15, 1971, and phased out toward the end of 1974. At first the Nixon program was well received, at least on the surface. Real GDP, which had grown at a measly 0.02% in 1970 and then a moderate 3.4% in 1971, began to grow full throttle. Over the course of 1972 and 1973 real GDP grew by 5.3%

and 5.8%, respectively. The unemployment rate, which had hovered around 6%, fell steadily to 5.3% by Election Day 1972.[20]

The financial markets responded positively at first. The Dow Jones Industrials gained 3.8% the day following the Nixon announcement and proceeded to continue to trade up from just under 900, hitting 1,000 shortly after Nixon was decisively reelected in 1972. Bond yields initially dropped in a sign of confidence in the new policy regime. The week before Nixon's speech outlining the new economic policy, 10-year Treasury notes yielded 6.88%, AAA corporate bonds traded to yield 7.69%, and Baa corporate bonds traded to yield 8.85%. A few months later, Treasury 10-year yields had dropped by almost a full percentage point to yield 5.89%. Corporate bonds rallied as well. AAA-rated bond yields dropped 47 basis points to yield 7.22%; Baa bond yields followed suit, dropping 49 basis points to 8.36%.[21]

Under the surface the corrosive effects of wage and price controls coupled with too-loose monetary policy were working their way through the system, undermining the faux prosperity. The predictable (and predicted) result soon emerged: skyrocketing inflation, slumping capital markets, distress in the banking system, a deep recession, and the beginning of capital flight from the United States. Inflation, measured by the Consumer Price Index, which had been running at about 4.5% in 1971, clocked in at 11.4% by the end of 1974. Short-term banker's acceptances, which had been trading at around 5.5% in August 1971, soared to over 12% by August of 1974. Yields of triple A-rated corporate bonds rose from 7.5% in August 1971 to over 9% by August of 1974. By mid-1974 the unemployment rate was on its way back up, peaking at 9% in 1975. The price of gold, which was trading at around $40 and ounce when Nixon shut the window, quadrupled to reach an average price of over $160 an ounce in 1975.[22]

The country fell into recession. GDP growth fell into negative territory for both 1974 (–0.5%) and 1975 (–0.2%). Recession and inflation took their combined toll on the Dow Jones Industrials, sending the average tumbling by 42% from its 1972 high before finding a bottom at 577 in December of 1974. As gold prices soared and stock and bond prices slumped, investors began to look for places to put their money where it would be safe from the ravages of inflation.

That was more easily said than done. The 1930s-era regulatory regime had walled off the savings of individuals from market rates of interest. The Fed, wishing to prevent banks from engaging in bidding wars with each other to attract deposits, had implemented Regulation Q which forbade the paying of interest on demand deposits and set ceilings on interest rates for time deposits. Under Reg Q, bank time deposit rates were capped out at 5% and thrift time deposit rates at 5.25%. The differential was supposed to provide an indirect subsidy to the housing industry by giving the thrifts, a key source of mortgage financing, a slight advantage in bidding for funds.

But while savings accounts had maximum legal rates capped out at around 5% to 5.25%, Treasury bills carried yields approaching 10% by the summer of 1974. The banking system began to bleed deposits as individuals emptied out their savings accounts to buy Treasury paper. At the same time, business loan growth exploded. Commercial and industrial (C&I) loans, which had been growing at around 7% in 1970–1971 suddenly began to surge. By September of 1973 year-over-year growth in commercial and industrial loans topped 20%.

Commercial loan growth continued to soar despite rising rates because, as traders say, the Fed was behind the curve. Actual as well as expected inflation was rising faster than nominal interest rates. So real (inflation-adjusted) rates were falling. Naturally enough, business scrambled to borrow dollars that got cheaper by the day (in real terms) to finance inventories whose prices continued to rise faster than the cost of funds. The banks were now once again being hit from all sides. Depositors were draining their low-rate savings accounts to buy higher-yielding Treasury bills; commercial loan demand continued to swell, and the Fed was clamping down on the money supply. The banking system needed a new source of funding for its loan portfolios.

Ironically enough, it came from overseas.

THE EURODOLLAR MARKET: BORN OF NECESSITY

Eurodollars are dollar deposits held in banks outside the United States. Often these deposits are not held in Europe at all; today large quantities of dollar-denominated deposits are held in the Cayman Islands, Hong Kong, Japan, Canada, Singapore, and South Korea. But by convention they are referred to as Eurodollars if they reside outside the United States.

The market for Eurodollars originally developed in the 1950s by virtue of the fact that banks from communist countries were reluctant to leave their dollar deposits in U.S. banking institutions. They feared the deposits could be easily seized by the U.S. government in times of crisis. The fear was not without merit: The United States seized some Iranian assets during the Carter-era hostage crisis. In any event British banks turned to the developing Eurodollar market to finance trade outside the commonwealth after the U.K. government imposed capital controls following the 1956 Suez War. Similarly, U.S. banks began to look to the Eurodollar market to finance loans made to U.S. companies operating overseas after the government began to impose capital controls through the Interest Equalization Tax (1963) and the Voluntary Foreign Credit Restraint Program (1965).

Eurodollars had several advantages over domestic money markets. For one, the market was essentially unregulated. Although the market was based in London, British authorities limited their regulatory activities to sterling-denominated assets and liabilities. Moreover, Regulation Q interest rate ceilings did not apply; nor did the Fed (at first) impose reserve requirements on Eurodollar deposits. Since the deposits were not insured, banks were not required to pay insurance premiums.

The Fed showed itself to be remarkably sensitive to the U.S. banking system's need to be competitive in the Eurodollar markets. Among other measures, the Fed made it easy for U.S. banks to set up "brass plate" operations in which shell branches were established for bookkeeping purposes, while in reality the work was actually done at U.S. headquarters. As a result, large international banks could turn to the Eurodollar markets for financing while avoiding the costly regulatory apparatus associated with domestic markets.

The Fed also eased the way for banks to compete for funds by gradually lifting, and then eliminating, Reg Q interest rate ceilings on certain domestic deposits. They began by lifting the ceilings on large negotiable certificates of deposit in amounts of $100,000 and up. Eventually, large retail brokerage houses like Merrill Lynch began to buy and pool these money market instruments, which they then reoffered to investors as money market mutual funds. By pooling and reoffering money market instruments as shares in a fund, Merrill Lynch and other brokers and mutual fund managers enabled small investors to dodge the $100,000 requirement needed to earn market rates of interest on negotiable certificates of deposit and other money market instruments. As market rates rose over savings accounts capped by Reg Q, retail investors piled into money market funds (and out of conventional bank accounts) with a vengeance. In November of 1973 money fund assets stood at only $63 million; two years later they had grown by 37 times to equal $2.3 billion. Today they approach the trillion dollar mark.

Access to the unregulated Eurodollar markets combined with the weakening and ultimate demise of Reg Q allowed banks to secure the funding they needed. But the price was steep. The formerly secure base of low-cost deposits was gone forever as retail investors were lured by the high returns offered by money funds. Fierce competition broke out among banks, brokerage firms, and mutual funds to attract the investment dollars of retail investors.

The competition was not limited to the liability side of the balance sheet. Borrowers, retail and wholesale, were becoming increasingly savvy about how they managed their financing needs. And so the banks were getting squeezed on both sides of the ledger. On the liability side they had to compete against brokers and their money funds to secure their deposit base. On the asset side, borrowers were getting increasingly picky about who their lenders were and how the loans were priced. The public markets offered an alternative source of funds. Decent quality credits could borrow in the commercial paper market instead of going to the bank for a loan. For longer-term borrowing, the corporate bond market was available as well.

THE EMERGENCE OF MODERN CAPITAL MARKETS

Before the 1970s the bond business was a backwater. Government deficits were relatively small, interest rates didn't move very much, and there wasn't much competition. A small number of firms, generally unknown to the public, dominated trading in the U.S. government debt. Many of the major investment banking houses had only token corporate bond trading desks, and many had no government bond desks at all.

The great inflation of the 1970s changed that. One date, October 6, 1979, stands out as pivotal. For one thing it was a Saturday and the Federal Reserve had an announcement to make. Weekend Fed announcements are never happy occasions. That Saturday evening the Federal Reserve issued a press release in which the Board of Governors finally acknowledged what had long been obvious: the Fed needed to take drastic action to stem the out-of-control inflationary spiral that had engulfed the financial system. And take drastic action it did. The Fed announced that it would no longer peg short-term interest rates. It would leave that task to the market. The Fed would concentrate on reducing growth of the money supply. Rates could go where they may. Overnight—actually over the weekend—the Fed turned the cozy world of debt finance into a bare-knuckled free-for-all.

The inflation that was initially touched off by Nixon and helped along by the Carter administration had inevitably spun out of control and forced the Fed's hand. The bond market was a shambles. The price of gold was soaring on world markets. Rationing had been imposed on gasoline. Shortages—the natural result of price controls—emerged in basic commodities like sugar. Command and control had once again supplied what it produces best: lines.

The initial reaction to the Fed's October 6, 1979, Saturday night massacre was swift and severe. Bond prices plummeted and interest rates soared. The financial markets were shaken to their very foundations. It wasn't just that interest rates ratcheted upward; it was the unprecedented volatility that was so stunning. Before the Fed's announcement the daily federal funds rate had been trading "comfortably" at slightly over 11%. A mere two weeks later the funds rate was trading at 15% and headed higher. Over the next two years the funds rate would rise and fall between

9% and 20% several times. By November of 1981, the U.S. Treasury was reduced to putting a 15.75% coupon on 20-year bonds to get the deal done.

It is impossible to manage a business when the cost of financing for long-term capital projects gyrates wildly minute to minute. But corporate treasurers were faced with just that unenviable task—finding and managing financing while interest rates careened about violently on a daily basis. Securing long-term financing could mean locking in horrific rates. But relying on short-term financing was dangerous. Liquidity could easily dry up. And to boot, short-term rates were higher than long rates—but for how long nobody knew. Similarly, the banking system found new sources of funds through the Eurodollar markets and money market funds. But that meant paying upwards of 15% for short-term money to fund assets on the books with single-digit yields.

Many of those assets were fixed-rate mortgages held by thrift institutions like savings banks and savings and loans. The inflation of the 1970s had touched off a frenzy of real estate speculation. Money began to pour into single-family housing well above trend. It is easy to see why. The inflation rate (measured by year-over-year changes in the CPI) had begun to take off, but mortgage rates were rising at a relatively moderate pace. Taking into account the fact that mortgage interest payments were tax deductible, the after-tax, after-inflation cost of borrowing had turned substantially negative. By the end of the 1970s, a tax payer in the 30% marginal rate bracket could take out a mortgage at a rate of slightly over 11%, paying an effective after-tax rate under 8% while the inflation rate galloped along at over 11%. And if rates turned down, refinancing was always an option.

INVESTMENT BANKING AS TRANSACTIONAL FINANCE

Corporate borrowers responded similarly. They could deduct interest payments against income while at the same time rising inflation made inventory stockpiling a profitable strategy. Estimated corporate bond issuance rose 28% from 1978 to 1979 and another 55% from 1979 to 1980.[23] But the decision to issue bonds was becoming an increasingly difficult one. For one thing the bonds had to be registered with, and then approved by, the regulatory authorities. It was a process that could take months, even when the markets were madly gyrating about on a day-to-day basis. For another, there was the inherent danger of being locked in to paying high borrowing rates for upwards of 30 years.

The second problem was fairly easily dealt with. Corporations simply resorted to issuing callable bonds. Often the bonds had nominal

30-year maturities, but they could be called-in (or redeemed) as early as 5 years after issuance. This was akin to a mortgagee's ability to refinance in a falling rate environment. However, the first problem required a regulatory change that reflected the new volatility of the financial markets. That came about in 1982 when the Securities and Exchange Commission adopted Rule 415 allowing firms to file so-called shelf registrations for new securities.

Through shelf registrations, companies meeting certain criteria were allowed to register (in one statement) all the securities they reasonably expected to sell over a two-year period. That way the company could keep the securities "on the shelf" and ready for sale depending on market conditions. Rule 415 represented an important procedural, but not substantive, change for the SEC. Disclosure is at the core of the SEC's mission, and Rule 415 represented a form of bulk disclosure. The SEC's adoption of the rule provided companies with a less costly and more flexible mechanism for issuing new securities.[24]

The adoption of Rule 415 had far-reaching implications for the capital-raising process. Investment banking firms had been in the business of providing corporate clients with advice on market conditions and financing options. Investment banking relationships tended to be stable and long term. If a corporation wished to raise money in the bond market, the firm's investment banker would check the pulse of the market and advise the client what his cost of funds would look like. The banker's advice would be based on gauging market demand shown by the investment bank's investment clients. Rule 415 changed all that. A corporate treasurer could sell preregistered bonds by putting them up for competitive bid whenever he thought market conditions were ripe. The investment banker was just the other side of the trade. Transactional finance was born.

The "credit" in credit markets comes from the Latin *credere,* meaning to believe or to trust. What really mattered in credit markets was the trust between lender and borrower. The borrower had access to credit because the lender trusted the borrower to pay back his debts; the borrower was good on his word. It is true that lenders did and do expect to get their money back. But the great inflation of the 1970s redefined what getting your money back really meant. Before Nixon slammed the gold window shut, foreign governments that owned dollars thought that they could redeem them for $35 an ounce in gold. Holders of U.S. Treasury paper would get their money back in the sense that they would receive the correct quantity of paper money at coupon payment time and maturity, but the value of the dollars received was cause for hollow laughter.

Moreover the Fed's October 6, 1979, decision to allow rates to go where they may, had injected extreme volatility into the bond markets. Now bondholders had to contend with market volatility as a significant source of portfolio risk. Bond market volatility placed a premium on market timing and trading skills. Shelf registrations turned treasurers into market timers. Investment bankers and underwriting clients, bond investors and traders, increasingly began to regard each other as nothing more than the other side of the trade. Naturally enough, the emphasis in the business began to migrate from managing credit risk to managing market risk. That is the nature of transactional finance during periods of market volatility.

Market risk and credit risk are two very different propositions. Managing them requires different tools and skill sets. So Wall Street went about acquiring the tools and skill sets it needed for the new environment. Wall Street is a brutally competitive place—because it is so adaptive. Firms adapt, compete, and prosper. Or they die. As the Street adapted to the rapid transformation of the financial markets, it began to create new instruments for hedging, managing, and gaining exposure to the possibilities presented by the new financial landscape. The instruments would include exchange-traded financial futures and options, zero coupon bonds, Treasury strips, floating-rate notes, inverse floaters, various types of mortgage bonds and junk bonds as well as over-the-counter derivatives such as swaps, options, and swaptions.

By the early 1980s America's domestic capital markets were on their way to being liberated from the constraints imposed on them by the 1930s regulatory regime. Global finance was likewise on the cusp of modernization, led by the American model. It is no small irony that America was able to lead the restructuring of the global financial system because the American government borrowed so much money, much of it from overseas. This created a new and powerful constituency: lenders to the U.S. government not under its political jurisdiction. The slightest hint of inflationary Federal Reserve policies would send the bond vigilantes scurrying for cover, selling their bonds along the way, pushing up interest rates, and threatening U.S. economic growth. In so doing the bond vigilantes imposed an external discipline on the system, constraining (but not eliminating) the ability of politicians and policy makers to adopt economically foolish but politically popular choices.

SUMMARY

In the blink of an eye the world political economy turned upside down. The gold standard went out the window and for the first time in history paper fiat money became the monetary standard. The deflation of the previous half century or so gave way to recurrent inflation. The advanced economies of the West rapidly evolved from agricultural to industrial and then to information- and service-based economies. Governments began to run large structural budget deficits. The need for new sources of finance, rampant inflation, and disintermediation contributed to the collapse of an archaic regulatory structure designed in and for the 1930s. Old-fashioned investment banking gave way to transactional finance. Finally, financial markets began to go global, paradoxically making more untapped savings available to borrowers, while at the same time imposing more stringent conditions on their use. The globalization of the capital markets, the emergence of the "bond vigilantes," and the external discipline they imposed on policy makers has had a profound impact in shaping how today's markets are priced.

NOTES

[1] See the *Concise Encyclopedia of Economics*: http://www.econlib.org/library/enc/bios/Smith.html.

[2] Adam Smith, *The Wealth of Nations,* Book 4.

[3] D. C. North, *Structure and Change in EconomicHistory,* 1981.

[4] Historical trading volume data can be downloaded from the Web sites of the Fed (http://www.newyorkfed.org/) and the NYSE (http://www.nyse.com).

[5] See Milton Friedman, *Money Mischief: Episodes in Monetary History,* 1992.

[6] See David B. Parker, "The Rise and Fall of the Wonderful Wizard of Oz as a Parable of Populism," *Journal of the Georgia Association of Historians* 15, 1994, pp. 49–63; and Henry M. Littlefield, "The Wizard of Oz: A Parable on Populism," *American Quarterly* 16/1, Spring 1964, pp. 47–58.

[7] For a full discussion see John H. Mackin and Norman J. Ornstein, *Debt and Taxes,* 1994.

[8] Henry Kissinger, *Diplomacy,* 1994.

[9] Woodrow Wilson, "President Wilson's Fourteen Points, the Basis of the New World Order," 1918.

[10] See Wilson.

[11] See Wilson.

[12] J. M. Shafritz and A. C. Hyde, *Classics of Public Administration,* 1997.

[13] See Shafritz and Hyde.

[14] Milton Friedman and Anna Jacobson Schwartz, *A Monetary History of the United States, 1867–1960,* 1963.

[15] See Friedman and Schwartz.

[16] See Friedman and Schwartz, p. 418.

[17] James D. Paris, *Monetary Policies of the United States,* Columbia University Press, 1938.

[18] Daniel Yergin and Joseph Stanislaw, *The Commanding Heights,* Simon & Schuster, 1998.

[19] See Yergin and Stanislaw, pp. 60–64.

[20] For historical data see *The Economic Report of the President,* http://www.gpoaccess.gov/eop/index.html

[21] These data can be downloaded from the St. Louis Fed: http://research.stlouisfed.org/fred2/

[22] Gold price data can be downloaded from The World Gold Council: http://www.gold.org/

[23] Corporate bond data supplied by Bear Stearns.

[24] See Marlin Jensen, R. H. Hudson, Carl D. Sullivan, et al., "Should Managers Register Shelf Offerings?" *Quarterly Journal of Business and Economics,* March 22, 1995.

CHAPTER 3

Going Mobile: Globalization and the Bond Market

I used to think that if there was reincarnation, I wanted to come back as the president or the pope or a .400 baseball hitter. But now I want to be the bond market: you can intimidate everybody.

–James Carville, quoted in *The Wall Street Journal* (February 25, 1993, p. A1)

In his own inimitable fashion, James Carville pretty much got it right. Politicians *are* intimidated by the bond market. They ought to be. Bond markets, no longer limited by geography, set interest rates in world markets rather than national ones. Financial capital is increasingly free to cross national borders to seek out the best risk-adjusted returns. In so doing this capital provides a modicum of discipline on otherwise free-spending politicians.

Of necessity, default-free government bonds are the backbone of the world's financial system—and they vote every day. The voting mechanism is transactions, millions of them that reflect changing expectations about the future. The markets are huge. In the United States over $500 billion worth of Treasury paper changes hands in the marketplace daily. Mortgage-backed bonds, federal agency bonds, corporate bonds, and municipal bonds kick in another $500 billion or so. At the Chicago Board of Trade (CBOT) another $200 billion of notional bond futures trade every day, mostly electronically, where the trading day lasts for 22 hours and is accessible from servers in Chicago, London, Paris, Amsterdam, and Singapore.

And that doesn't count equities, foreign exchange (FX), or the money markets. The major U.S. stock exchanges trade about $80 billion worth of corporate stocks on the average day. Over $2 trillion in foreign exchange trades daily. And the Chicago exchanges trade trillions of notional value daily in Eurodollar and federal funds futures and options. Eurex, the Swiss-German futures exchange trades millions of contracts on German Bunds, Bobls, and Schatz. Euronext LIFFE trades millions of STIRS—short-term interest rate futures denominated in Euros.[1]

The sheer magnitude of the trade has discouraged most governments from attempting to set prices, realizing that in the end, the market sets the terms. This refreshing fact is regarded with dismay by much of the political establishment, which is not all that surprising considering that governments are by far and away the world's largest borrowers. Nor should the magnitude and persistence of deficit finance be thought of as accidental; it is part and parcel of the modern welfare state that no discussion of capital markets can ignore.

FINANCING THE WELFARE STATE

Government bonds are deferred tax collections. Because they are default-free, government bonds are benchmarks for all other financial market instruments denominated in the same currency. Other financial instruments necessarily trade at a discount to them. Put another way, they carry a risk premium. Government bonds sold to finance deficit spending therefore rest at the financial foundation of the modern welfare state.

They also lie at the fault lines of the capital markets because a demographic time bomb is ticking away in the Western democracies and Japan. The present value of unfunded future liabilities for retirement and healthcare programs (like Social Security and Medicare) is massive and looming as the first wave of baby boomers prepares to retire. Conservative estimates put the total at well over $50 trillion in the United States alone. The numbers for Europe are generally thought to be even more onerous. Reforming these programs is going to require massive financing, much of it coming from external sources, even more so than today's demands.

The U.S. government has been addicted to deficit finance for quite some time, tapping the bond markets for new cash more or less continually since 1970. The surpluses produced by divided government in the late 1990s were an exception to the rule, the result of a massive and unexpected wave of capital gains receipts stemming from the dot-com bubble. At first the two major political parties were so perplexed by the surpluses that neither quite knew what to do. The Democrats were afraid that the Republicans would use the surpluses to reduce taxes for the wealthy; Republicans, on the other hand, were afraid that the Democrats would spend it on new social programs aimed at building up Democratic constituencies. Senator Phil Gram (R,TX) remarked that the majority party Republicans finally found a solution to the Democrats' spending plans. "We'll spend it first," he is reported to have said.

MARKETS AND MOBILITY

Government bonds are sold at the nexus of politics, markets, and mobility. Borrowing demand is produced in the political process; the bonds are priced in the capital markets and lie on the knife's edge of capital flight. Government bonds are only as safe as the entity that issues them, which partly explains why bonds of the same maturity from different jurisdictions do not always converge to trade at the same yield. Other reasons include liquidity, the business cycle, and inflation expectations. For example, bonds of the same maturity issued by the United States, Japan, Britain, and Germany, representing four different currencies, may trade at different yields because they are at different points of the business cycle. In theory, over time risk-adjusted yields on national government bonds should move toward convergence.

Some of the spread that exists between government bonds of different countries has to do with the moral hazard that has been built into the financial system in the service of politics. Before Keynes, fiscal prudence was considered to be virtuous. Profligacy was abhorred; budget balancing was the order of the day. Markets cleared, or were at least self-correcting. After all, Adam Smith had said, "What is prudence in the conduct of every family can scarcely be folly in that of a great kingdom."[2] That was then.

Keynes argued that budget balancing *was* folly. Markets, he said, were not self-correcting and could not be counted on to clear. The economy was perpetually threatened by either the prospect of overheating or a slowdown. At any moment it faced downturns marked by falling prices, insufficient demand, overcapacity, and high rates of unemployment. Wages were sticky, unable to fall to market-clearing levels. When the cycle turned, the economy faced inflation, overheating, and a lack of productive capacity, owing in part to businesses' reluctance to invest in plant and equipment, the memory of previous downturns fresh in their minds. The solution to the inherent instability of the system, according to Keynes, was for government to use its budget to fine-tune the economy and mitigate the effects of the business cycle. Government should run budget deficits during slack times and surpluses during boom years.

Wittingly or not, John Maynard Keynes removed the constraints that previously had held politicians back. They were now free to overspend, overborrow, and overinflate. In fact, they were scientifically obligated to. They took up the call with gusto. From 1960 to 2004, the U.S. government produced 39 deficits but only 6 surpluses. The surplus years totaled $562 billion; the deficit years, $4.4 *trillion*. In practice the Keynesian system took no account of the real world of politics. Nor did it acknowledge the asymmetry of the incentives it put into play. It assumed, or pretended to assume, that politicians would produce surpluses as readily as deficits. Or, as Nobel laureate James Buchanan puts it, "They forgot the elementary rule that politicians enjoy spending and do not like to tax."[3]

As long as the incentives to do so remain intact, political man, no less a rational utility maximizer than economic man, will borrow and spend freely. The mountain of debt that governments have issued over the last 40 years stands as testament to the power of that simple fact. Consider, for example, the issuance of government bonds by the Organization for Economic Cooperation and Development (OECD) countries. Between 1980 and 1998 central government debt in OECD countries rose from 27% of GDP to 49% of GDP.[4]

The year 1980 is crucial because it is the year in which the United States elected Ronald Reagan to be its 39th president. In short order the Regan administration would reduce tax rates substantially, increase defense spending dramatically, and give Fed Chairman Paul Volker the political backing needed to wring inflation out of the financial system. The effect was to create a series of enormous budget deficits "as far as the eye could see" in the words of the Office of Management and Budget (OMB) chief David Stockman.

The deficit had to be financed, and the Reagan administration went shopping for the money overseas. In so doing, it planted the seeds of global financial market integration. But this era of free trade would be very different from the preceding one. The last period of global free trade that ended with the onset of World War I used gold to settle accounts and was primarily concerned with trading in commodities. The globalization of financial markets that began in the 1980s settled accounts more or less in dollars. What mattered was mobility of financial assets. Trading across borders in financial instruments dwarfed trading in commodities. The perfect instrument for this was U.S. Treasury bonds.

For a number of reasons, Treasury bonds were the perfect instrument for financial globalization. First, the United States offered political stability. Second, world commodities (especially oil) were (and still are) mostly priced in dollars. Third, U.S. capital markets were (and are) large enough to accommodate enormous flows of funds. Most importantly, the Reagan administration reversed the policy mix adopted by the Nixon, Ford, and Carter administrations. In order to break the back of inflation, the Reagan administration relied on tight monetary policy, allowing (even encouraging) the Fed to keep real interest rates high. To promote growth, it loosened fiscal policy by reducing tax rates. The result was a soaring dollar that attracted global investors who wanted to put their money into high-yielding Treasury bonds.

The markets that were developed to sell all that debt soon began to call the shots. After all, the lender calls the tune, and governments were now competing with each other to borrow money in world markets. But not all of the borrowers fell under their respective jurisdictions; nor were the borrowers necessarily constrained by tight national regulatory regimes. Politicians now had to pay obeisance to the markets, or at least pay attention, which circumscribed their room for maneuver (hence the quote from James Carville at the beginning of the chapter).

In the short run, highly mobile capital can be politically and economically disruptive; far more so than mobile populations. Capital is far easier to move. Transaction costs are minimal compared to the costs of picking up sticks and moving to a new country. Once unleashed, capital simply seeks out the highest returns for the least risk. Unlike people it cannot be assimilated into a culture or a political system. Partly as a consequence, highly mobile capital markets blur the formerly bright lines that used to separate domestic and world markets.

MOBILITY AND THE TIEBOUT MODEL

Ironically enough, the internal logic of international capital mobility was first described in the context of local public finance by Charles Tiebout, a professor of economics at the University of Washington, when he published his famous article, "A Pure Theory of Local Public Expenditure."[5] Tiebout sought to address a central problem in public finance theory posed by Paul Samuelson in 1954. The public finance problem posited by Samuelson was that decentralized markets were inherently incapable of producing an optimally efficient quantity of public goods.

Samuelson argued that public and private goods were inherently different. Public goods are consumed jointly, in stark contrast to private goods. The classic example of a jointly consumed public good is defense. Everyone in the society benefits from national defense; the protection it affords one citizen is conferred on all. Clean air is another example. But, Samuelson argues, private markets will not produce the economically efficient quantity of public goods. Some citizens will mask their true demand and act as free riders. Absent the proper price signals, supply will not be sufficient to meet all potential demand. Since left to its own devices, the market will fail to produce a sufficient quantity of public goods, the decision-making process should be transferred from economic markets to political markets, where it is presumed that citizens will reveal their true preferences at the ballot box.

After examining the Samuelson thesis, Tiebout agreed, but only up to a point. In "A Pure Theory of Local Expenditure" published in the *Journal of Political Economy* in 1956, he argued that given sufficient mobility, citizen-voters would implicitly reveal their true demand for public goods by moving to the communities that best mirrored their preferences. To draw out the implications of mobility, he described an extreme model of local expenditure that included the following seven assumptions:

1. Consumer-voters are fully mobile and live in the communities that best satisfy their preferences.
2. They have perfect knowledge of local revenue and expenditure patterns.
3. There are many communities from which to choose.
4. All communities possess the same employment possibilities.
5. There are no spillovers with respect to public services.
6. Every community has an optimal size, which is essentially at the point where the marginal and average costs of public goods provision are equal.
7. Communities seek to attain optimal size. Smaller than optimal communities seek immigrants to reduce the average cost of service provision. Larger than optimal communities seek to shed residents.

The Tiebout model of mobility is located within the confines of a single nation-state. Within the state people are mobile; they can vote with their feet. They go to live in communities that best reflect their demand for public goods, whether it be good schools, good roads, or good police protection. On the other hand, public goods suppliers (local governments) are stationary, limited by geopolitical boundaries. Their continued existence depends on their ability to attract people to live within their boundaries. To do so (given the assumptions of the Tiebout model) local governments must efficiently supply the quantity of public goods that existing (or potential) residents demand.

The Tiebout model is often criticized as unrealistic. For instance, he ignores the importance of community ties, heterogeneous consumption preferences, and changing demographics. Also, transaction costs associated with moving are large. The assumption of perfect knowledge with respect to local fiscal affairs by consumer-voters is a bit hard to swallow. Not to put too fine a point on it, employment possibilities clearly vary across place—and always have. Some migration is clearly motivated by those opportunities. Moreover, spillovers in the provision of public goods are rampant. Anyone who doubts this should go to a municipal recycling center and note the number of recyclers without town stickers on their cars—or even with out-of-state license plates.

Grant that the assumptions of the Tiebout model are unrealistic. The real world has always had the annoying habit of being a lot more complicated than the model builders' descriptions. But the point of a model is not to be realistic in every detail; it is to simplify reality to its essential qualities. Make two changes in the Tiebout model, and suddenly it doesn't seem so unrealistic after all. Change the focus of mobility from people to funds for portfolio investment. Then change the policy arena from municipal government to nation-states. And then the model begins to look pretty realistic.

GOING MOBILE

Consider the Tiebout model's assumptions when mobility is defined in terms of actual 21st century capital mobility. Is financial capital "fully mobile"? It is close enough; well over a trillion dollars in foreign exchange trades daily. Hit the "send" button for your Internet bank account, and your money can be instantly on its way to an account in pretty much any major financial institution anywhere in the developed world—and a good deal of the developing world. There may not be perfect knowledge about receipt and expenditure patterns for the jurisdiction in question, but there is plenty more information (and more transparency) about how OECD countries tax, spend, and keep their books than there is for a typical local government in the United States.

It is demonstrably easy for a citizen or a commercial enterprise of a free country to send financial capital almost anywhere (and back) around the globe at very little cost. Moreover, the preferences that have to be satisfied are easily defined. They are reward maximization and risk minimization.

Easy capital mobility within a nation-state induces efficiencies by pushing local governments to compete to keep citizens satisfied with the production of public goods and the tax-price they pay. But capital mobility across nation-states is an entirely separate matter. The movement of financial capital across national borders changes the domain of the game from domestic politics to world politics. Through jurisdiction shopping and regulatory arbitrage, globalized finance challenges the power and authority of governments, redistributes power between and within nation-states, changes political incentives, redistributes global savings and investment, gives rise to powerful nonstate actors, and subverts traditional societies.

FINANCE IS GLOBAL; SAVINGS ARE LOCAL

The function of financial markets is to pair up savings with investment. Businesses and governments seek the use of savers' capital by offering returns or at least potential returns. Savers accept those returns as compensation for deferring consumption. Before finance went global, savings pools tended to be local, as were the capital investments they financed. As better and more efficient communications technologies developed, financial markets became bigger and less fragmented. Local markets grew into national markets, and rates of return across political subdivisions converged. Differences in rates of return were arbitraged away.

International markets are a whole lot trickier. Markets, financial or otherwise, do not exist in vacuum. For markets to function properly, property rights need to be protected, contracts have to be enforced, and regulatory regimes established. These are vital functions, typically performed by national governments. But international capital flows exist outside the legal jurisdiction of any one national government. At the same time, global markets can reach deep inside national borders in search of savings that can be extracted and redirected to locales offering the highest risk-adjusted returns. Moreover the structure of world politics makes it easy for traders (illegally or otherwise) to seek out tax havens with weak regulatory regimes to shield their activities from the reach of the law. And the structural borrowing needs of governments are so great that they can ill afford to challenge the institutional arrangements that provide them with the capital they so desperately need.

Governments need to borrow almost continually to fund massive deficit spending. Actually they run two types of deficits: cyclical and structural. Cyclical deficits result from subnormal growth. When growth slows, tax revenues fall off and expenditures rise as automatic stabilizers (like unemployment insurance) kick in. When growth picks up, so do tax revenues, and deficits are reduced, all else equal. Structural deficits are different. They are the result of expenditure in excess of tax receipts even when the economy is growing at full capacity. For instance, if transfer payments are pegged at 20% of GDP and taxes remain at 18% of GDP, no rate of growth will make up the difference.

The structural foundation of deficit finance is woven inextricably into the fabric of the welfare state. Fertility rates for Western Europe and Japan are below replacement levels, resulting in rapidly aging populations that, along with the United States, will shortly begin to see record numbers of retirees in proportion to workers. Unfunded liabilities of social insurance programs for retirement and health care are massive and will require commensurately massive intergenerational transfers and borrowing. High-spending societies will seek financing from high-savings societies for years to come. That necessitates going across borders.

POLITICALLY DISRUPTIVE CROSS-BORDER BORROWING

Cross-border borrowing creates a whole new set of incentives because it is inextricably intertwined with foreign exchange markets and the conception of national sovereignty. By the 1980s, non-U.S. nationals had begun to acquire significant quantities of Treasury debt. As foreign buyers bought more and more U.S. Treasury bonds, the terms of trade began to change, however subtly. The result was that domestic political choices were subject to review by people who were not U.S. citizens, who were not under U.S. jurisdiction, and who could withdraw their money at any time by selling their dollars and Treasury bonds. The borrower could no longer set the terms; now there were conditions. The conditions are, as always, set in markets and expressed as prices. The prices are rates: short-term interest rates or currency rates.

Governments can set either short-term domestic interest rates or their currency's foreign exchange rate, but they can't set both. It is a truism that sets lots of politicians and central bankers on edge. Naturally enough, they shoot the messenger by attacking the markets, sometimes to the point where they begin to sound a bit unhinged. No less an eminence than former Fed Chairman Paul Volcker castigated the FX markets when in 1992 he wrote that, "The [foreign exchange] market is essentially made up of a few hundred traders with open telephone lines. . . . They care little about basic economic trends, they do not pretend to be thoughtful economists, and in fact they do not especially want to associate with them."[6] David DeRosa relates that French President Jacques Chirac has described foreign exchange speculation as "the AIDS of the world economy."[7] But it would be hard to outdo Mahathir Mohamad, former prime minister of Malaysia. Paul Krugman quotes Mahathir as saying that, "Currency fluctuations are caused by hostile elements bent on . . . unholy actions that constitute 'villainous acts of sabotage' and 'the height of international criminality.'"[8]

The hostility directed at currency traders by politicians is really directed at markets that fall outside the control of policy makers, forcing them to make choices they would rather not have to make. Consider the wisdom of Franz Muntefering, chairman of Germany's Social Democratic Party, and a key ally of the German Chancellor Gerhard Schroder. He is quoted by Wolfgang Munchau in the May 2, 2005, *Financial Times* as saying, "Financial investors remain anonymous. They have no face, they descend upon companies like locusts, destroy everything and move on." The rhetoric, comparing human beings to insects is a calculated political ploy, not an isolated instance of getting carried away by the moment. One of his deputies, Ute Vogt, subsequently went on to urge the party to boycott companies that had recently dismissed employees.

As Mr. Muntefering uttered those words, the German unemployment rate hovered around a postwar high, upwards of 11%. Similarly, the French unemployment rate was 12.5%. But the more market-oriented economies of Britain and the United States had far lower unemployment rates: about 4.7% in Britain and 5.5% in the United States. The liberal economies of the United States and Britain are flexible and adaptable; they stand in contrast to the rigid and bureaucratic "social market" economies of France and Germany. Liberal economies attract investment capital from around the globe far more easily than do social market economies, tying the hands of enthusiasts for market intervention.

The upshot is that governments will continue to borrow across borders, but the process will be politically disruptive. The more liberal economies will be high attractors of capital, putting pressure on the more rigid social market economies to liberalize in order to compete for investment funds. To attract capital they will need to pay premium rates. Bond and currency market traders will arbitrage away attempts to peg rates.

It would be a mistake to simply assert that today's flexible market economies will be the main attractors of capital and leave it at that. To maintain their attractiveness to global investors, nations need to maintain the quality of their institutions and manage their finances well. Failure to do so can undermine the business environment. In this respect it is worth noting that in its 2006 survey of global competitiveness, the World Economic Forum dropped the rank of the United States from first place down to sixth, largely as a result of the mismanagement of its public finances. Huge budget deficits, according to the Switzerland-based institute, threaten U.S. competitiveness both by increasing businesses' borrowing costs and by reducing funds available for productive investment.[9]

SUMMARY

The logic of savings and investment; of profit-maximizing behavior; and technological capacity; combined with high-spending, low-saving social welfare states implies continuing structural deficits that will be financed across borders. Highly developed liberal economies will continue to import capital and develop the financial architecture necessary to facilitate global borrowing.

The law of one price will push prices toward convergence, adjusted for risk, as arbitrageurs and traders speculate across borders on interest and exchange rates. But the process is likely to be disruptive, bringing political tensions to the surface. Trading stocks, bonds, and commodities across borders will encompass all sorts of risk. And it will open up all kinds of opportunities. But market risk, political risk, currency risk, and credit risk will all have to be managed, and new instruments for doing so developed.

NOTES

[1] These data are available on the Web sites of the various exchanges and the NY Fed Web site.

[2] Adam Smith, *The Wealth of Nations*.

[3] James M. Buchanan, *Constitutional Economics,* Basil Blackwell, 1991, p. 42.

[4] OECD data are available at: http://www.oecd.org/home/

[5] Charles Tiebout, "A Pure Theory of Local Public Expenditure," *Journal of Political Economy,* 1956.

[6] Paul Volcker and Toyoo Gyohten, *Changing Fortunes,* Times Books, 1992, pp. 230–231.

[7] David F. DeRosa, *In Defense of Free Capital Markets,* Bloomberg Press, 2001.

[8] See Krugman at http://web.mit.edu/krugman/www/baht.html

[9] See the World Economic Forum: http://www.weforum.org/en/index.htm

This page intentionally left blank

CHAPTER 4

Electronic Markets

> Dave, I believe you're taking this a little too personally.
>
> *–Hal, 2001: A Space Odyssey*

THE TRANSFORMATIONAL POWER OF ELECTRONIC MARKETS

It is hard to exaggerate the importance of the market's migration from the trading floor to the screen. It is changing the way markets are organized, analyzed, quoted, and priced. Transactions costs are a fraction of what they once were, thereby contributing to an explosion in trading volume. Price quotes are transparent and much easier to obtain. Markets react with increasing speed to new information (and pseudo information). But the demand for speedy news reporting has reduced the filters that information travels through. Accuracy can easily suffer in the rush to report first. The line between news, gossip, and rumor has become blurry. Theory would suggest that greater speed, lower transaction costs, and increased transparency have made markets more efficient.

Electronic trading changes the configuration of the competitive landscape faced by both securities firms and organized exchanges. Rigid and hierarchical business models that depend on layering and channeling order flow simply cannot compete with flexible network models that absorb and redistribute information rapidly and seamlessly. To remain as venues where large numbers of buyers and sellers can transact, exchanges need to compete to attract global pools of liquidity. Consequently, they have pushed aside barriers to entry, opening up opportunities for entrepreneurs to establish new trading firms. Not surprisingly, markets have become more liquid and trading volumes have rocketed ahead. Not only that but formerly bright distinctions between exchange and OTC trading have gotten blurrier by the day. Times have changed.

FROM GEOGRAPHIC PLACE TO FUNCTIONAL SPACE

Trading (or exchange) is deeply rooted in the human experience, so much so that we simply take it for granted. For most of man's existence trading has taken place face-to-face. Transactions were typically executed between people who either knew each other, or if not, knew the middleman. Trade took place in physical space (whether an exchange, the village market, bazaar, or the modern grocery store) and transactions were executed according to local custom. But electronic markets operate in functional space; geographic place no longer dominates.

Trade unlocks the efficiencies created by the division of labor; without it the advanced economies of liberal states would not be able to function as they do. But trade does not exist in vacuum: It is embedded in norms, customs, and shared cultural understandings. In liberal societies the organization of trade into formal market arrangements rests on a foundation of private property rights, the right of contract, and the rule of law. Market institutions are the product of centuries of experimentation through trial and error. In liberal societies market institutions are adaptive, constantly evolving in response to changing conditions. They are not dictated from the top down; they are built from the bottom up.

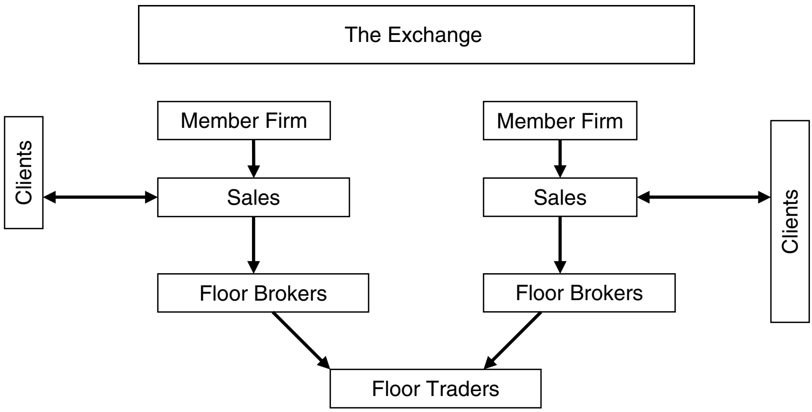
Capitalism is by its nature adaptive, experimental, and innovative, always driving change. Its agents of change are entrepreneurs. Joseph Schumpeter described them as being innovators in products, organizational forms, or production methods that earn them high profits, attract imitators, and disturb the system. The fundamental impulse that keeps the capitalist engine in motion is innovation. It "incessantly revolutionizes the economic structure from within, incessantly destroying the old one, incessantly creating a new one." Schumpeter goes on to argue that "the gale force winds of creative destruction" are at the heart of the capitalist enterprise.[1] Marginal increases in technical efficiency are beside the point. It is the power of new ideas and technologies, ever present, ever in the making, that drive and change the system.

THE CREATIVE DESTRUCTION OF ELECTRONIC MARKETS

Electronic exchanges are a key piece of the infrastructure of global capitalism. They operate across borders. Their price signals are important inputs to the capital allocation processes. They are capable of achieving massive

FIGURE 4.1

Flowchart of Traditional Floor-Based Trading



economies of scale, and their trading rules apply across cultures and customs. They are profoundly threatening to central planners and illiberal regimes everywhere. But are electronic capital markets truly Schumpeterian goods? Have they unleashed the gale force winds of creative destruction?

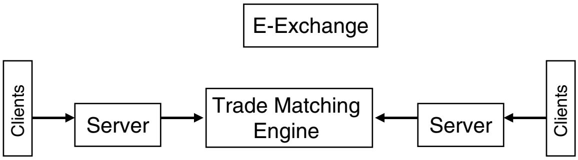
To examine the question let us begin by first considering a simple model of traditional floor-based trading, as diagrammed in Figure 4.1.

Figure 4.1 presents a flow chart of client orders as they work their way through the marketplace. The prospective client first gets a price quote from a salesperson who is an employee of an exchange member firm. After getting the quote, the client may decide to place a buy or sell order with the salesperson, who then routes the order to a broker's broker for execution on the floor of the exchange. The broker's broker may (or may not be) an employee of the same firm as the salesperson. The broker then takes the order to the post (or pit) where the security or commodity is traded. At that point the broker looks for the other side of the trade. The other side may be another broker representing a different customer. Or it may be the specialist (in stocks) or a local trader (in the commodities markets). Once the transaction is executed, the report of the execution or "fill" travels back up the same chain, but in reverse order. The specialist confirms with the broker, who reports the trade to the salesperson who in turn reports the price of the fill to the customer, after which the member firm sends out a written confirmation of the trade to the customer.

Note the hierarchy of the system. It funnels the order flow through various layers (each of which takes a cut) and ultimately channels it through the local floor traders (or specialists), providing them with valuable market information. The information is particularly valuable because it is real-time

FIGURE 4.2

Flowchart of an Electronic Exchange (E-Exchange)



and actionable, giving the local trader (or specialist as the case may be) an informational advantage in the marketplace. (In return for this NYSE specialists are supposed to maintain an orderly market in the securities for which they are designated as market makers). The hierarchical structure of the system makes it relatively easy to impose a regulatory regime on it. There are clear layers of authority along with functional differentiation. There is little question of who has legal jurisdiction over the players. Now compare the model of the floor-trading system with a model of an electronic exchange, displayed in Figure 4.2.

As Figure 4.2 illustrates, the organizational structure of the e-exchange (as we shall call it) is considerably flatter than the traditional floor-based exchange. Straight-through processing dominates the order flow, from quote acquisition to instantaneous electronic confirmation and profit-and-loss statement. As a result of technology, information flows faster and more freely. Price quotations along with the depth of orders on the book are no longer the sole province of local traders. That information is freely displayed on computer screens. There is no longer any need to call for a price quote. In fact, calling for a quote would be a waste of time. The market may have changed during the time it took to make the call, absorb the information, and perhaps place an order.

FROM CHANNELS TO NETWORKS

Marketers like to think in terms of distribution channels, a metaphor that comes from the physical world that is used to describe how customers are guided (herded might be a better word) to consume what the organization has for sale. Think, for instance, of grocery stores. High-margin impulse items are mostly found at the checkout counter. Brand names are placed on the shelves at adult eye level. Candy, a tad lower, fine for enticing kids.

In e-markets though, the dynamic is different. The customers are in charge of product placement: They have to decide what information they want taking up valuable space on the computer screens they constantly monitor. Further, the technology is interactive. Traders enter their own orders into the system. It's throughput that matters. In the flattened structure of the e-market, traders put orders into the system through their own computer "front ends" that are linked to the exchange's computers. Exchange computers have trade-matching engines that do precisely that. The exchange's computers are programmed to match buyers' and sellers' orders when they (electronically) agree on a price. After transactions are consummated, they are electronically reported back to the transacting parties. The technology is such that execution reports and a running tally of positions, risk profile, value-at-risk (VaR), and P&L are provided and updated virtually instantaneously and continuously. Trading rules that establish order execution priorities, acceptable order sizes, price change increments, or other conditions are programmed into the exchange's computer software algorithms to prevent illegal trades from taking place.

The current state of the art gives traders the capability to put their computers on autopilot, so to speak. To execute their preferred strategies, traders need only program their computer front ends to search out and automatically execute transactions if and when prices (or spreads) happen to hit predetermined limits. The technology is sophisticated enough that these electronic trade execution strategies are not confined to any one particular exchange; they can be stretched across many exchanges, for instance by buying a security or financial instrument on one exchange and automatically selling a different, offsetting instrument on another. Electronic exchanges are thus becoming subsumed into de facto distributed networks.

FIRM ORGANIZATION

Consider the organizational implications of global electronic exchanges on member firms. Traditionally, firm departments, personnel, and functions are organized around securing, executing, and clearing customer orders to buy and sell securities. For instance, stock trading, bond trading, and commodities trading departments are (or were) separate. Salespeople would get orders by talking to customers; those orders would be sent to a broker to be executed, and the firm's back office would process confirmations, deliveries, and cash settlements.

However, at electronic exchanges many of the functions that were traditionally performed by brokers and salespeople are performed by computers and trade-matching engines. Traders put their own orders into the system via their computer front ends, which are linked to the exchange's trade-matching engines. Traders don't need salespeople for market information and ideas; traders simply want fast connectivity and trade execution. Marketing efforts are just as likely to revolve around technology.

Many of the services provided by investment banking and brokerage firms can be delivered directly to customers electronically, bypassing the need for (and expense of) armies of salespeople. Execution services now take the form of providing electronic access to exchange markets. Firms compete for this business by providing fast connectivity, research services, and minimal transaction fees. This leads brokerage firms to seek competitive advantage by achieving economies of scale.

Research services can be easily provided electronically as well. The real question is how firms arrange to be paid for it. Some research firms have been able to persuade clients to pay directly for the research. More often than not, the buy side pays for research with "soft dollars." Soft dollars have been mildly controversial on Wall Street for a long time. Basically what it boils down to is that buy-side firms guarantee brokerage firms a certain amount of brokerage business in return for research. As a result, brokers can attach an expected revenue stream to the research services they provide, which allows them to budget how much effort they will expend for research to begin with. But as commission rates steadily drop, research budgets are cut and fewer securities are covered by Wall Street analysts. Recent severing of investment banking revenues from research departments has put further pressure on research budgets.

With the value of the customer execution business seemingly on the wane, investment banking and brokerage firms have increasingly begun to trade for their own accounts. New types of trading firms have sprung up that eschew the customer execution business altogether, relying exclusively on in-house trading for their profits. Prime brokerage services for hedge funds have become increasingly important as well. Some proprietary trading firms have come to rely on a combination of trading skills and brokerage commissions. Quite a few trading firms have set up shop in Gibraltar and other tax havens.

A TRIP TO THE ROCK

Gibraltar—the Rock—is a tiny bit of land (about 2.5 square miles) that sits at the entrance to the narrow strait that is the only exit from the Mediterranean Sea to the Atlantic Ocean. And so it has been a strategically important naval gateway for thousands of years. The British grabbed it during the War of Spanish Succession (1701–1714); declared it to be a colony in 1830, and exercise sovereignty over it to this day. Not surprisingly, the Spanish have never quite gotten over this, and so Gibraltar is the focal point of the occasional diplomatic flare-up. The Spanish have been known to occasionally shut off electric power to the island; especially after having lost a close football match to the Brits. And for their part, the Brits insist on periodically reenacting battles fought and won by Admiral Nelson, something the Spanish find less than amusing. While the Rock has always been valued for its significance as a naval gateway, it now possesses financial significance: It is a tax haven.

Gibraltar is one of the most densely populated pieces of real estate on the planet at 10,800 residents per square mile, even though it has no natural resources to speak of and limited supplies of its own fresh water. It is a member of the European Union (EU) having joined the European Economic Community along with Britain in 1973. A special treaty, negotiated by Britain, applies to Gibraltar's relationship with the EU, exempting it from certain EU laws. Specifically, Gibraltar legally falls outside the EU's customs territory; it is excluded from EU agricultural policies; it is excluded from the requirements of VAT harmonization; and no part of its customs revenues go to the EU.

Gibraltar makes an ideal setting for companies dealing in stocks and bonds, holding portfolios, and trading in futures and other derivatives markets mainly because interest and capital gains are entirely exempt from tax. Company directors may be of any nationality; there is no residency requirement, and board meetings may be held anywhere. The company name must end with either "Limited" or "Ltd" which, all else being equal, doesn't seem terribly burdensome. Gibraltar encourages high-net-worth individuals to set up tax residency on the Rock. For such people the rules stipulate a minimum tax of ≤10,000 and a maximum of ≤20,000. As an added bonus, Gibraltar does not tax earnings from other jurisdictions. Consequently, residents pay a small fraction of what they would ordinarily be required to pay in Europe.

But the terrain of Gibraltar is kind of forbidding; it is just a big rock, there is nothing idyllic about it. So lots of traders make their homes across the border in Spain and declare tax residency in Gibraltar. During weekdays they hop on their mopeds and stream across the border to go to work trading in the global capital markets. They can do this because their firms have electronic access to the markets through the Internet, as well as hubs and servers that have been installed by exchanges to provide traders with superfast connectivity to their respective markets.

The traders who work on the Rock come from all over. It is easy to find British, Australian, Spanish, Italian, German, and (occasionally) American traders at the bar of the Elliot hotel. And they all speak the same evolving language—trader talk—which is nearly indistinguishable from the language spoken by traders in London, New York, Paris, Rome, Amsterdam, and Dublin. There are minor variations; for instance, the British *take* decisions, and the Americans *make* them. But to a man (and they are almost all men) the traders watch and debate the moves of the major central banks: the Fed, the Bank of England, the European Central Bank, the Bank of Japan, and the Bank of China. There is intense speculation about currency movements and the effects on inflation and bond prices. Careful attention is paid to the pronouncements of finance ministers and other economic spokespeople; "Fed conspiracy" theories, in particular, spread through the markets like wildfire.

Traders are skeptical of economic releases, often voicing the opinion that the numbers are "doctored" or manipulated to serve those in power.

The traders in Gibraltar, like many of their contemporaries in London's Canary Wharf or Dublin or Chicago or Amsterdam, mostly work for proprietary trading firms. The "prop shops" or "trading arcades," as they are often called, are in the business of speculating in futures contracts, options on futures, and sometimes cash securities. For all the bravado normally associated with trading—made famous by Tom Wolfe's *Masters of the Universe*—the prop traders appear to be almost innately conservative. In part it is because they have profit-sharing deals with their firms, so in a sense they have their own money on the line. Overwhelmingly they are day traders, seeking to scalp a few ticks before calling it quits for the day. Rarely do they carry a position overnight. That privilege is reserved for the most senior traders.

Despite outward appearances it would be a mistake to think of traders as a homogenous group. In a clinical study of day traders, Andrew Lo, Dmitry Repin, and Brett Steenbarger found no evidence of a "trader personality type," using a standard battery of psychological personality tests.[2] They did find that traders with the most intense emotional reactions to profit and loss had significantly worse trading performance. In another study of professional traders at investment banks, Fenten-O'Creevey et al. found that the more successful traders tended to be emotionally stable introverts who had open minds.[3] No surprise there.

There is one trait that is overwhelmingly shared by day traders. It is a near obsession with charting. Technical analysis, the art of trying to predict future prices on the basis of past price patterns, is for many the holy grail of trading. Not surprisingly there are plenty of technical analysts around to sell brightly colored charts of head-and-shoulders patterns, Fibonacci retracements, Kondratief cycles, and the other esoterica of the craft (or witchcraft if you prefer).

The use of technical analysis is almost irresistible for most prop traders. Go into the trading room of almost any prop shop and you will see an ocean of computer terminals with flat screens displaying the latest exchange rates in the currency markets, real-time stock market indexes, bond market prices, and money market rates. The data come from the major markets in the United States, Britain, Germany, France, and Japan among others. Naturally there are charts of all this frenetic activity being displayed on the traders' computer terminals. What better way to get a quick visual update on trends (or what appear to be trends) on market prices all around the world, all at once?

Checking the screens for quick visual updates inevitably leads to formalizing the process and finding patterns in the data—whether the patterns are really there or not. Just as inevitably traders will begin to follow "systems" based on their readings of the charts. Some patterns will be identified as buy signals; others as sell signals. Eventually an avid chart reader will go on a sustained money-making streak. The laws of chance demand it. But the streak will not be read as a random event; on the contrary, it will be seen as compelling evidence confirming the prescience of the trading system. Perhaps that is because in the human psyche there is a fundamental need to believe in order, stability, and predictability.

Charts are a way of mapping the unfolding order that traders see, or profess to see, in the day-to-day movements of the markets. But there is a big difference between mapping where the market has been and predicting where it is going. More about this in later chapters.

EXPECTATIONS, RATIONALITY, AND THE LIMITS TO ARBITRAGE

Economists tend to scoff at chartists, believing for the most part that market prices do not follow predictable patterns. The great majority of economists believe that short-term price movements of financial assets are randomly distributed. Market prices reflect all that is known about a security at any given time. Prices change when traders buy and sell in response to new information. But it is not simply the new information that matters. What matters is the degree to which the new information conforms to earlier expectations. To be successful, traders need to know what everybody else thinks, so they can assess market expectations—and vice versa.

This is the phenomenon that Keynes had in mind when he compared stock markets to beauty contests. Around the time that Keynes wrote his general theory, the newspapers of Fleet Street would run contests in which readers were asked to select the 6 prettiest girls out of 100 pictured in the newspaper.[4] The beauty contest winner was the girl who got the most votes. Everyone who voted for the winner had his name included in a lottery drawing, giving that person a chance to win a cash prize. Since the winner of the beauty contest was determined by voting, the economically rational trader would place his (or her) bet on the woman who he thought would garner the most votes; not necessarily the one he thought was the most beautiful. So too Keynes thought with stocks. The trick was to guess which stocks other people would want to buy rather than determine their true value. It is a game of infinite regress: I will buy what I think you will buy, and you will buy what you think I will buy.

Efficient market theorists argue that, on the contrary, the short-term randomness of the market signals efficient pricing. Short-term randomness implies that, adjusting for risk, no one can systematically outguess what everyone else in the market will do. Risk, in this context, refers to the magnitude of potential price fluctuation. Stocks that fluctuate more are riskier because the potential for loss is greater. Furthermore they argue, over time prices will tend to reflect the true underlying value of a stock, which is the present value of expected future cash flows. If and when prices get too far out of line with intrinsic values, canny arbitrageurs will step in and buy the cheap stocks and sell the expensive ones thereby forcing prices back into line. When prices are widely known and trading access is easy, each new bit of information is quickly incorporated into market prices as traders buy and sell on the news.

Widespread electronic trading is likely to create an evidentiary tug of war between the beauty contest and the efficient market theorists. In addition to easy market access, the revolution in communications technology has made news dissemination virtually instantaneous across the major (and not so major) financial centers. As a result, news can be seen, analyzed, and discounted in the marketplace by an extraordinarily diverse set of actors. A broader and more diverse set of players would seem to be the perfect recipe for improving market breadth, depth, and efficiency.

But there are two assumptions here that need to be explored. The first is that the new technology is in fact communicating real information. The second is that the diverse set of players is in fact diverse in ways that matter. The technology that has made instant communications possible has also made possible the instantaneous transmission of vast quantities of misinformation, market-moving rumor, and outright falsehoods to be consumed by the gullible. Anyone who doubts this has merely to log on to Google and type in the phrase "Federal Reserve Greenspan heart attack rumor" and push the send button. In less than one-half of one second with a reasonably good broadband connection, Google will come up with over 5,000 Web hits cataloguing various stories, rumors, and denials that former Federal Reserve Chairman Greenspan has had a heart attack. There are thousands of trader message boards with postings from people who have handles—monikers like Rodger Dodger, Chart Man, and Million Dollar Baby—instead of names. They argue passionately about support and resistance levels, Elliot Waves, gold, and the effects of geomagnetic space storms on equity markets. Not surprisingly, the message boards are full of wild rumors, innuendo, and conspiracy theories.

Efficient market theorists argue that over time the smart, truly professional traders will arbitrage away the distortions caused by the geomagnetic space storm traders. But recent research from behavioral finance suggests that it's a quite bit more complicated than that. Behavioral finance economists base their case on two assertions. First, there are limits to arbitrage. Second, psychology prevents traders from behaving in ways that are fully rational. Are these assertions plausible? And if so, are the effects sufficiently large to be a source of concern? Finally, is electronic trading likely to be an exacerbating or mitigating factor?

The ability of arbitrage trading to serve as a corrective to market distortions is a critical component of the efficient market hypothesis (EMH). According the EMH, market prices reflect a security's fundamental value over the long run, where fundamental value is defined as the present value of expected future cash flows. In the short run a security's price may deviate significantly from its true fundamental value. At that point arbitrageurs possessing superior information about the security's intrinsic value will buy undervalued securities and sell overvalued securities, eventually forcing prices back to their true fundamental values.

The problem is that arbitrage is not a riskless enterprise. Once the possibility of price distortions is allowed into the equation, it does not follow that prices will necessarily find their way back to intrinsic value. The forces that pushed prices out of line to begin with may be persistent and powerful enough to continue to push prices further out of line, causing arbitrageurs to suffer substantial trading losses after positions have been established. Or as Keynes is reputed to have said, "The market can stay irrational longer than you can stay solvent." Behavioral economists go a step beyond demonstrations of persistent irrationality in pricing. Nicholas Barberis and Richard Thaler of the University of Chicago have published an extensive literature review discussing the work done by psychologists designed to explain whether and why people systematically make financial decisions that most economists would call irrational, or at least less than 100% rational.[5] A summary of their review follows.

PSYCHOLOGY AND TRADING

A critical factor in decision making concerns expected outcomes. It turns out that people tend to be overconfident of the quality of their decision-making abilities. A closely related phenomenon is that people are unduly optimistic. For instance, when asked to estimate sample quantities in lab experiments, people assumed they would guess correctly about 98% of the time, but they actually guessed accurately only about 60% of the time.[6] Events they thought virtually certain actually occurred only 80% of the time; events thought virtually impossible actually happened 20% of time.[7] Over 90% of people participating in surveys thought they were above average in driving abilities, sense of humor, and ability to get along with others.[8] People systematically underestimate the time it will take them to accomplish tasks.[9] They read more into the data than is there. A classic example is the sports fan's belief in a player's hot hand even when there is no evidence of it.[10] People tend to persevere in their beliefs. They avoid contradictory evidence, and when confronted with it, tend to overdiscount it.[11] When people form estimates starting with an arbitrarily determined initial value, that value often becomes a critical point of reference for no good reason. The estimation process can simply become psychologically anchored in an arbitrary reference point.[12]

The economics profession is skeptical that evidence garnered from psychological experimentation is applicable to financial markets. Economists argue that people learn from mistakes and adapt their behavior to avoid similar mistakes in the future. Further, professional traders are likely to be systematically different from the rest of the population; in any event competitive forces are likely to ensure the survival of successful traders and winnow out the unsuccessful ones. Expert traders at investment banking firms will make fewer errors than most, and trading firms can structure incentives to minimize mistakes.

Barberis and Thaler, scourges of the EMH, concede that the biases that people bring to financial decision making can be "unlearned." But, they say, bias cannot be eliminated altogether. Moreover, there is evidence that expertise can be more of a hindrance than a help: It turns out that financial market professionals sometimes exhibit more overconfidence than laymen. Finally, while getting incentive structures right may reduce bias in decision making, a literature review by Camerer and Hogarth could find no replicated study that eliminated bias just by changing incentives.[13]

If psychological foibles can lead to irrational trading decisions that overwhelm arbitrage, is electronic trading likely to be an exacerbating or mitigating factor? On the one hand, it would seem that electronic trading ought to reduce the incidence and duration of mispricing. Expanding the scope, depth, and breadth of the market (as electronic trading does) brings more information, more agents, and more perspective to events, which ought to reduce the impact of outliers. On the other hand, electronic networks may exacerbate herding effects, facilitate information cascades, and increase the danger of financial gridlock. The question centers on the role played by "noise traders."

NOISE TRADERS

Financial market economists draw a distinction between "noise traders" and arbitrageurs. Arbitrageurs can be thought of as fundamentalists. They assess a security's worth in terms of the present value of its expected future cash flows, adjusting for risk. For instance, take two different Treasury bonds with similar coupons and maturities. Rational traders would expect the two bonds to have the same yield. If the yields of the two bonds were to diverge significantly, arbitrageurs would be expected to sell the low-yielding bond short, buy high-yielding ones, and wait for their yields to converge.

Noise traders on the other hand can be thought of as uninformed traders. Their trading is uninformed in that positions are taken without new information, with erroneous information, or as a result of misinterpretations of factually correct information. For all intents and purposes, noise traders simply follow the tape. They are trend followers who are trying to guess the likely future behavior of their fellow traders, much as Keynes described the reasoning of beauty contest players.

For a long time the question of whether noise traders had much impact on market prices was dominated by Milton Friedman's framing of the argument. In an article discussing the merits of floating exchange rates, he argued that speculators (i.e., uninformed traders) could be destabilizing only to the extent that they sold at low prices and bought at high prices (otherwise they would have a stabilizing effect). But to the extent that speculators sold low and bought high, they would eventually put themselves out of business. They are self-defeating

Friedman allows that his argument does not settle the question. He acknowledges that professional speculators might make money even as a constantly replaceable body of amateurs regularly loses. But, he says, we should presume that is not the case. It is an admonition that may have been a discerning one in 1953 when Friedman wrote the article, but there are reasons to doubt its applicability today. They include cultural, institutional, and technological factors.

In 1953, memories of the 1929 stock market crash and the Great Depression were less than 25 years old. It is hard to imagine that in the 1950s a substantial body of replaceable amateurs would be interested in financial market speculation as a way to earn a living. Leaving that aside, institutional arrangements would have precluded all but a few amateurs from participating in the financial markets as speculators, at least as we know them today. Commissions on stock exchanges were high (and fixed), quotes on bonds were hard to come by, and capital markets were heavily regulated. Communications technology was primitive. It could take hours (and a good deal of money) to get a telephone call patched through from New York to London, an arrangement hardly conducive to amateurs punting in the foreign exchange markets.

Fast-forward to 2005. In matters of finance, conservatism has gone out the window. Forty-eight states now have some form of legal gambling, compared to the 48 that *outlawed* it two decades ago.[14] Legal gambling through casinos, horse races, and state-sponsored lotteries now generates more than $60 billion in gross revenues annually.[15] Las Vegas markets itself as "family friendly." Hotel casinos will gladly arrange babysitting services while you are busy losing the college tuition money. Thousands of people show up daily to play games of chance where the present value of the payoff is far removed from the possibility of winning. These are Friedman's amateurs. They will be welcome at the tables until their money has evaporated, at which time they (and their families) will be politely but firmly shown the door.

In some ways electronic trading arcades bear an uncanny resemblance to casinos. Like casinos, some arcades consider the traders to be customers of the firm and charge commissions and fees for support services and connectivity. Others consider the traders to be firm employees and negotiate a sharing agreement for profits and losses, but profits are net of commissions and support services. The traders themselves tend to be young (many in their twenties), single, male, recent college graduates, often with degrees in business, finance, and economics. That rough profile holds true in Chicago, New York, London, Dublin, and Gibraltar.

Like more traditional investment banking firms, some trading arcades have fairly elaborate selection and training programs, believing that they can pick winning traders and trading methodologies. Some have had trained psychologists on staff to try to profile psychological traits of successful traders. Others put prospective employees and/or trainees through rituals that include playing competitive rounds of poker to test their mettle under fire. In an air of studied casualness, many have game rooms, pool tables, and video games set up just off the trading areas where the traders can take a break and blow off some steam.

Not surprisingly, the arcades experience high trader turnover. Traders who lose their initial stake or who don't meet profit targets need to pony up, or they are shown the door. They are the cannon fodder in the Darwinian trading game; they are Friedman's improbable floating supply of punters. The arcades, needing to replace those who have recently fallen in battle, are constantly on recruiting campaigns.

What explains the apparently never-ending supply of beginning traders willing to take on such risk despite the long odds? And what difference does it make to the markets?

NOISE TRADERS, MARKET BUBBLES, AND INFORMATION CASCADES

Brad De Long, an economist from U.C. Berkeley, has studied noise traders extensively. He has found fascinating evidence suggesting that, contrary to Friedman's hypothesis, noise traders can stick around for a long time and exert considerable influence on stock prices, even when the noise trader's market assumptions are incorrect. Ironically, *the very fact of noise trader irrationality keeps them in business.* This counterintuitive finding requires some explanation.

The argument made by Friedman and later Fama was that rational traders—arbitrageurs—would take the other side of the noise trader transactions. In so doing the arbitrageurs would make money while contributing to market efficiency by pushing prices back in line. But a paper written by De Long, Schleifer, Summers, and Waldmann explains why arbitrageurs may fail to aggressively take advantage of noise traders' market misperceptions.[16]

Unlike buy-and-hold investors, arbitrageurs have relatively short time horizons. (Actually, so do lots of portfolio managers who are subject to quarterly mark-to-market evaluations.) Moreover, being rational, they are risk-averse. The arbitrageurs' short-term time horizon precludes them from taking positions opposite noise traders if they fear that the noise traders will persist in their irrational behavior for extended periods of time. Since noise trader feedback loops tend to be self-reenforcing, it is entirely possible for stocks (or, for that matter, bonds) to be priced away from their fundamental values for considerable periods of time. Not only can securities be priced incorrectly for extended periods of time, but the limited ability of arbitrageurs to take the other side of noise trader transactions increases market volatility. Increased market volatility implies that returns are higher than they would otherwise be to compensate investors who rationally trade off risk and reward.

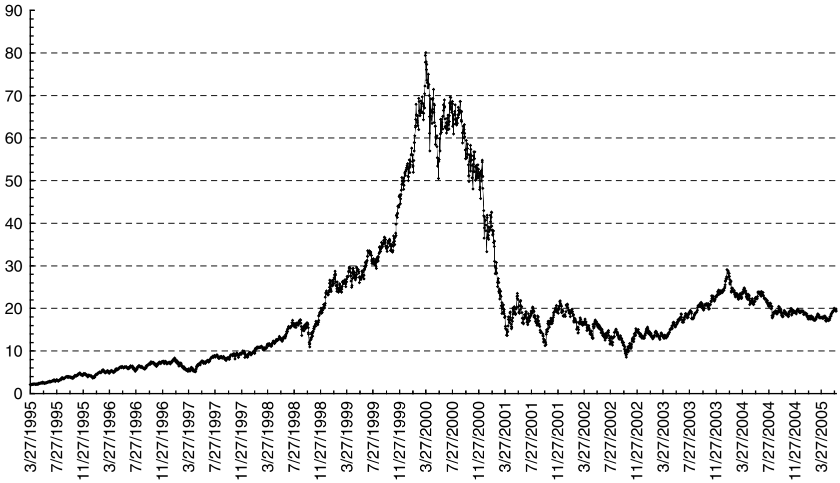
On the other hand, noise traders are not risk averters—they simply go where the action is, or at least where they think it is. As a result some noise traders earn superior returns because they hold assets that are riskier than average. And they are riskier than average because of noise trader–induced volatility. In effect, the noise traders have created their own high-risk space in which some earn superior returns (and survive) while the rest lose their shirts. But there is more to it than that: Noise traders can be significant contributors to market bubbles. And electronic trading can either reduce or increase the likelihood and severity of market bubbles, depending on institutional arrangements.

The metaphor of a market bubble is a useful one, even granting that market bubbles are (generally) recognized only after they have been punctured. Consider what a market bubble really is (pricing far out of line with fundamental values, generally with significant leverage) and how it forms (by aggregating misinformation and pricing it as fact). The formation of market bubbles is driven by a phenomenon known to economists as an *information cascade,* in which the sequence of decision making determines (or at least greatly influences) the outcome. What makes information cascades particularly problematic is that they are self-reinforcing, the way crowd behavior tends to be. As the German philosopher Francis Schiller once put it, "Anyone taken as an individual is tolerably sensible and reasonable—as a member of a crowd, he at once becomes a blockhead." On the other hand, former Treasury Secretary Robert Rubin has pointed out that heading for the exit when someone yells fire isn't necessarily irrational either.

Take for example the dot-com bubble that finally began to deflate during the first quarter of 2000. The price of Cisco Systems stock eventually came to exemplify the crowd mentality that dominated the stock market. As has been amply documented elsewhere, the stock market was on

FIGURE 4.3

Cisco Systems Daily Closing Prices (March 1995–June 2005)



Data source: Yahoo! Finance

a tear; the latest dot-com stocks were a staple of conversation at cocktail parties; MSNBC, Bloomberg television, and others were on 24/7 discussing every wiggle in the market. People who could barely figure out how to log in to an AOL dial-up account (without their kids' help) were buying Cisco stock at 150 times earnings without having the slightest idea what either P/E ratios or routers were. But they were quite certain the stock was going up; it was the new economy, and stocks were meant to go up every day. Figure 4.3, a chart of Cisco's daily closing stock price, sums up the mania.

How is it possible for not one, but scores of companies to have their stock prices explode to the upside and then turn around and collapse in a very short period of time? Information cascades. The best way to see how information cascades can take hold in financial markets is to first consider an example from another realm. Picture a vacationing couple touring in a city they have never been to before. They are window-shopping restaurants deciding where to go for dinner. Suppose they come to an intersection where there are two restaurants on opposite corners. The establishments have similar menus and prices. But glancing through the windows they can see that although both are open for business, one restaurant has no patrons inside and the other is three-quarters full. Which restaurant will the couple choose for dinner? They will choose the one that has already attracted patrons. Why? They will assume something is wrong with the empty restaurant.

Now suppose instead that our couple comes equipped with some information. They have been told by a neighbor back home that restaurant

A is an excellent value. But restaurant A turns out to be the empty one. Where will they go? They will still probably go to the one with the patrons, making the assumption that something has changed or that their neighbor's information was faulty to begin with.

Now conduct a thought experiment. Imagine we take four couples and decide by the flip of a coin which restaurant each couple will go to. If the coin comes up heads, they go to restaurant A; if it comes up tails, they go to restaurant B. The odds of the first couple going to either restaurant are even. The likelihood of the first two couples going to the same restaurant is $.5 \times .5 = 25\%$. The probability of all three couples winding up in the same restaurant is $.5 \times .5 \times .5 = 12.5\%$. The probability of all four couples dining in the same place is only 6.25%. Now suppose a fifth couple, unconnected with the experiment, comes by. Almost certainly the fifth couple will choose to enter the restaurant that already has the four other couples dining in it, even though a random selection process would have predicted only about a 3% chance that all five couples would wind up in the same restaurant.

Even with private contrary information (e.g., the neighbor's opinion that restaurant A is excellent value), there is a very reasonable possibility that the fifth couple will simply choose to head into the restaurant that already has patrons in it. The couple will take its cue from what they regard as the rational behavior of others. The restaurant without patrons must be inferior. Why else would it be empty? But it turns out that the one restaurant was empty because four sequential flips of a coin produced four heads, slightly more than a 6% probability. For all the outward appearance of rationality, the fifth couple's decision does not reflect fundamental information about the relative desirability of the restaurants. Instead it reflects a choice made by people who thought they were rationally evaluating relevant information, but who were actually fooled by random coin tosses.

Now let's switch the choice: Instead of which restaurant, make it a choice to buy or sell a stock or bond. And make the decision maker a "momentum trader." Momentum traders (a particular type of noise trader) like to say that they "go with the flow" by which they mean that they buy when they see (or think they see) buyers and sell when they see (or think they see) sellers. Of course, this is self-reenforcing, just like the restaurant selection process. It is a game of follow-the-leader. What matters is following the person in front. It assumes (without evidence) that the person in the front of the line knows where he or she is going. The more people buy, the higher the price goes. The higher the price goes, the more people buy. Around and around it goes until possibly, a market bubble forms. At that point the market no longer represents the independent assessments of many players. It is more like a herd of elephants racing through the tall grass trampling everything in its path. Which is why the arbitrageurs have already headed for the hills.

Is there a reason to suspect that electronic trading contributes to herd behavior? Quite possibly yes. Before electronic trading began to take hold at the major futures exchanges, order flow was transmitted to a central physical location where it could be observed by local traders. For instance CBOT 10-year Treasury note contracts all traded in one trading pit, populated by local floor traders. No more: Electronic trading has for all intents and purposes eliminated the filter of local floor traders who assess demands for liquidity and fill orders. Consequently, the impact of local floor traders on market practice (and prices) has dissipated.

Some research suggests that the result is increased volatility. In studying the impact of trader type, Daigler and Wiley found that floor traders who were able to observe order flow dampened volatility with their trading.[17] This study is consistent with previous work that found evidence of increased volatility related to noise traders who could not differentiate between liquidity demand and fundamental values. And it is consistent with the idea that some noise traders make outsized returns due to their willingness to bear outsized risks (ironically of their own making).

The apparent willingness of noise traders to bear outsized risks (wittingly or not) addresses the conundrum raised by Friedman. Remember, he argued that it was possible (but not presumptively likely) that a few professional speculators could make large sums, while waves of amateurs would lose their money only to be replaced by the next wave of money-losing amateurs. Electronic exchanges provide the infrastructure that makes this possible. The exchanges provide the playing fields; the trading arcades recruit the waves of new traders. Occasionally a trader will hit it big; the laws of chance demand it. And it makes for fabulous advertising. Without the possibility of the big jackpot (however remote), the casinos would be empty. That is why casinos (and lottery ticket sellers) publicize the occasional big winner who happened to come by their store.

Easy electronic access may create conditions conducive to short-term herding behavior in financial markets, but it is hard to argue that sustained herding behavior derives from electronic trading. Laboratory experiments have shown information cascades to be largely ephemeral. Under controlled experimental conditions, market bubbles have shown a tendency to form, collapse, and then quickly form again. In experiments, these cycles occurred repeatedly, and more often than not, one cycle would reverse the other, which suggests that the cycles served as market equilibrating mechanisms. In addition sophisticated traders showed themselves to be aware that other traders were less-sophisticated traders, and altered their own trading to take advantage of that fact. Finally, over longer periods of time there were more temporary and self-correcting cascades and fewer permanent ones.[18]

One way in which electronic trading can dampen herding behavior is by market linkages. Markets can be linked so that arbitrageurs can buy in one market and sell in another. For instance, cash Treasury bonds can be easily traded against Treasury bond futures contracts and options on futures contracts. Individual stocks, and groups of stocks, can be traded against stock indexes—for instance, the Dow Jones Industrials, the S&P 500, or the Russell 2000, to name a few. And stock indexes can be linked and arbitraged against each other. For instance, a trader can sell the S&P 500 short while buying, or going long, the Russell 2000.

It is widely (and mistakenly) believed that cross-market linkages exacerbate rather than dampen market volatility. The truth is rather the opposite. Electronic linkages between and among markets serve to create a complex network of markets that diffuses rather than concentrates risk. Think of modern financial markets as webs of prices for various financial assets, some of which are highly correlated, some of which are lightly correlated, some of which are not at all correlated (or are negatively correlated) at statistically significant levels.

Webs are flexible, made of threads that bend before they break. They transmit information rapidly and widely to the rest of the network along those very flexible threads. This mode of information transmission acts as a prophylactic against information cascades. Think of information cascades as a having a vertical structure with the first decision determining subsequent decisions as in the game follow the leader. The leader is the only information point in the game. Given the game's vertical structure, the players are isolated from outside information; they have no choice but to follow the person in front, all the while hoping he is headed in the right direction.

Now imagine the line of players from a different perspective, not from the top down, but from the side. Seen horizontally rather than vertically, it is easy to see that information can be communicated to each of the players simultaneously so that they can make independent assessments and can act independently. That is what properly linked electronic markets achieve. They disseminate market price information instantaneously and broadly. As a result, faulty decisions due to bad sequencing are minimized. Multitudes of decision makers can act independently and competitively, collapsing and correcting small market bubbles before they become big ones. Properly linked electronic markets can be "smart markets" that serve to cushion shocks, spread out risk in manageable pieces, and provide transparency and immediacy, so that markets can rapidly equilibrate.

SUMMARY

Electronic markets have had an important impact on how financial markets trade. Transaction costs are lower, transparency is greater, and access is easier than for markets that traded primarily on exchange floors. Electronic access has flattened the organizational structure of the marketplace and has allowed the flow of information to be seen and discounted by more people more quickly than ever before. It is possible that noise traders are more likely to thrive in electronic environments, but it is too early to tell. On the other hand, the evidence from experimental economics suggests that electronic trading may reduce rather than foment market bubbles and information cascades.

NOTES

[1] Joseph Schumpeter, *Capitalism, Socialism, and Democracy*, 1984.

[2] Andrew Lo, Dmitry Repin, and Brett Steenbarger, "Fear and Greed in Financial Markets: A Clinical Study of Day Traders," *NBER,* March 2005.

[3] M. Fenton-O'Creevey, N. Nicholson, E. Sloane, and P. William, *Traders: Risks, Decisions, and Management in Financial Markets,* Oxford University Press, 2004.

[4] See Ross M. Miller, *Experimental Economics,* John Wiley and Sons, 2002, especially pp. 87–88.

[5] Nicholas Barberis and Richard Thaler, "A Survey of Behavioral Finance," NBER Working Paper No. 9222. Available from SSRN online at: http://ssrn.com/abstract=327887); see also Richard Thaler, *The Winner's Curse,* The Free Press, 1992.

[6] M. Alpert and H. Raiffa, "A Progress Report on the Training of Probability Assessors" (Harvard University, 1968); reprinted in D. Kahneman, P. Slovic, and A. Tversky, eds., *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, 1982, pp. 294–305.

[7] Baruch Fischhoff, Paul Slovic, and Sarah Lichtenstein, "Knowing with Certainty: The Appropriateness of Extreme Confidence," *Journal of Experimental Psychology*, Human Perception and Performance Issue 3, 1977, pp. 552–564.

[8] A. Weinstein, "Unrealistic Optimism about Future Life Events," *Journal of Personality and Social Psychology* 39, 1980, pp. 806–820.

[9] R. Buehler, D. Griffin, and M. Ross, "Exploring the Planning Fallacy: Why People Underestimate Their Task Completion Times," *Journal of Personality and Social Psychology,* 67, 1994, pp. 366–381.

[10] T. Gilovich, R. Vallone, and A. Tversky, "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology* 17, 1985, pp. 295–314.

[11] C. Lord, L. Ross, and M. Lepper, "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37, 1979, pp. 2098–2109.

[12] D. Kahneman and A. Tversky, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185, 1974, pp. 1124–1131.

[13] C. Camerer and R. Hogarth, "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor Production Framework." *Journal of Risk and Uncertainty*, 19, 1999, pp. 7–42.

[14] See Jason White, "States Seek Gambling Fix," Stateline.org, November 2002.

[15] Encarta: http://encarta.msn.com/encyclopedia_761560772/Gambling.html#endads.

[16] De Long, et al., "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, University of Chicago Press vol 98 (4), pages 703–738.

[17] Robert T. Daigler and Marilyn K.Wiley, "The Impact of Trader Type on the Futures Volatility-Volume Relation," *The Journal of Finance* 54/6, December 1999.

[18] Jacob K Goeree, Thomas R. Palfrey, Brian W. Rogers, and Richard D. McKelvey, "Self-Correcting Information Cascades," 2004. This is a paper, available online from the UCLA Department of Economics at: http://ideas.repec.org/p/cla/levrem/122247000000000153.html

CHAPTER 5

Are Markets Really Efficient, Mr. Markowitz?

A surprising problem is that a number of economists are not able to distinguish between the economic models we construct and the real world.

–Alan Greenspan, as quoted in The New York Times, *August 26, 2005, p. C2*

Markets look a lot more efficient from the banks of the Charles than from the banks of the Hudson.

–Fischer Black

Sometime in the early 1970s a freshly minted graduate of the famous University of Chicago MBA finance program landed a job as a bond salesman at Goldman Sachs. Shortly thereafter he had the occasion to take one of his formers professors out to lunch. Toward the end of the lunch, as told to the author, the then-young salesman looked at his guest and said, "Mr. Markowitz, do you really think that markets are efficient?" To which Markowitz reportedly replied, "What, are you crazy?"

Harry Markowitz's central insight, that risk and reward can be systematically traded off to optimize portfolio performance, is an intellectual tour de force. It represents a staggering leap forward in our understanding of finance and markets. It is elegant, subtle, and descriptive. Unfortunately, money being money and people being people, elegance and subtlety often go out the window when nuance and theory are introduced to the real world of financial markets.

The efficient market hypothesis (EMH) dominates modern finance. An entire edifice of financial market practice has been constructed around its core assumptions. Portfolio management, risk management, and trading strategies increasingly rely on complex computer models based on the mathematics of the EMH. This inevitably leads to the question: Do the markets actually behave (within reason) as the theory predicts? And to what degree can the theory be used?

According to the EMH, capital markets are efficient because they aggregate the collective wisdom of rational investors. These rational investors evaluate financial assets by estimating the present value of expected future cash flows, adjusted for risk. Risky assets (like stocks) are discounted more deeply than are less risky assets (like government bonds) to compensate for the relative uncertainty of their future cash flows. Increased risk requires increased expected reward. More precisely, rational investors only buy risky assets to the extent that they expect the potential for increased returns to outweigh their increased risk.

Prices change when new information comes into the market. The underlying assumption is that anyone who possesses market-moving information will act on it, causing prices to adjust to a new equilibrium price. But it can only be new and unexpected information that changes market prices because the market already reflects all the previous information known by the players, in the aggregate.

However, not all traders are informed traders. Some are "noise traders." They trade on misinformation, rumors, or bits of information that are factually correct but are either of little relevance or badly interpreted. In the end, the assumption of the EMH is that informed traders will "outtrade" misinformed traders, ensuring that, on average, prices are approximately correct. The proof of the pudding is that it is difficult to show that anyone has systematically produced superior risk-adjusted returns over time, not attributable to chance. The implication is that the market is a fair game.

The argument is inevitably made that so-and-so beat the market on a risk-adjusted basis for years, proof positive that that the market is not a random walk. It can be, and has been, beaten. But Burton Malkiel, author of *A Random Walk Down Wall Street,* easily disposes of that argument with the analogy of a coin-flipping contest. Suppose that 1,000 people enter a contest in which the object is to guess how coin flips come out. Assume that half guess tails and half guess heads, each time. Before each coin toss, each of the contestants guesses either heads or tails. After each flip, the contestants who guessed incorrectly are removed from the game. Then the process is repeated.

After the first flip there are only about 500 people left in the game; after the second flip, 125. The third flip winnows out 63. After the fourth flip the game is down to 32. After the sixth and seventh flips, only 8 quarterfinalists are left standing. The eighth flip reduces the pool by half-again to 4 semifinalists. The ninth flip narrows it down to 2 finalists and the tenth flip produces a winner. The winner is, of course, chosen by random flips of the coin. It is simply a matter of luck. Somebody had to be left standing at the end of the contest. But that doesn't mean the winner was good at guessing coin tosses.

MARKET EFFICIENCY: IS IT STRONG, MEDIUM, OR WEAK?

Back in the stock market, price changes only arise because of new information; the past is not determinative. Everything that happened in the past is already priced into the stock; it bears no influence on what will happen in the future. Just as in a coin toss, past flips of a coin have no bearing on future flips.

Proponents of the weak form EMH argue that since price changes are independent (like sequential coin flips), the market has no memory. The next price can just as easily be up or down; it is a random walk. The proffered evidence for the weak form is that thus far no one has demonstrated the ability to consistently beat the market—that is to say, consistently earn higher-than-market returns, *after adjusting for risk*.

The strength of the market efficiency assumption is important because ultimately it is likely to influence the choices about which trading strategies to use. The strong form case for the EMH argues that financial market research is essentially useless. It asserts that all that is known, *and all that is knowable*, is already in the price. Why bother to spend time and money searching for information that has already been discounted? Research analysts counter that the market is efficient *because* their collective research produces a consensus estimate of fair value. That argument leads to the formation of the semistrong form of the EMH. The market is efficient because, in the aggregate, professional analysts know how to properly value companies and lead investors to buy cheap and sell dear, forcing prices to an approximation of their true value. There may be random short-term fluctuations in prices, but they revolve around true value because of the work of professional analysts.

Changes in valuations are the result of a host of factors that affect a company's well-being. Things like management, product launches, industry trends, business cycles, and so on have a bearing on how well companies do. Professional analysts know this. They also know how to collect data, analyze it, and interpret it with respect to stock (or bond) prices. The competition among analysts to do this and publish results leads to new information being discovered, assessed, and priced into the market reasonably quickly.

THE EFFICIENT FRONTIER

It isn't quite enough to simply say that short-term changes in market prices are random events and just leave it at that. The EMH argues that rational investors trade off risk and reward. That is to say, for any given level of risk an investor will choose the investment that offers the most potential reward. Conversely, for any given reward potential, an investor will select the stock that carries the least risk. This leads to the idea of the efficient frontier, which can be thought of as the set of risk/return trade-offs that represent the best possible choices an investor can make.

If investors trade off risk for return, then it ought to be possible to represent the trade-off with a graph that plots risk coordinates on one axis and return coordinates on the other. The resulting plot should be upward sloping—the higher the risk, the higher the reward. If a trader's risk/reward profile lies somewhere off the frontier, the trader's position can be adjusted to move it closer to the efficient frontier either by increasing return potential or by reducing risk. That requires operational definitions of risk and reward, which is where the trouble begins. How do we quantify risk? How do we quantify reward? What common metric do we use for comparison purposes? What constitutes a fair bet?

A FAIR BET

A good operational definition of reward potential is the expected payoff. Risk can be thought of as the chance something will go wrong and a loss will result. If traders are rational, the magnitude of the potential payoff (reward) has to be at least equal to, if not greater than, the potential for loss (risk). A common denominator is needed to compare the two. Otherwise there is no way for a rational trader to identify potential opportunities, much less gauge their relative attractiveness.

The framework for calculation begins with probability theory and the idea of a fair bet. In a fair bet, the payoff is equal to the odds of winning. For instance, the flip of an evenly weighted coin produces one of two possible outcomes, each with the same probability of occurrence, namely 50%. Since the odds are even (50:50), a fair bet is one in which the payoff for winning is the same as the penalty for losing. On the other hand, if the bet were to be decided by whether the flip of 2 coins produced 2 heads, the probability would be $.5 \times .5 = .25$, so a fair bet would be one in which the payoff reflected a 1 in 4 chance of winning.

The odds would be defined as the chance of winning (p) divided by the chance of losing ($1 - p$), which in this case is equal to:

$$Odds = \frac{\frac{1}{p}}{\frac{1}{1-p}} = \frac{\frac{1}{4}}{\frac{3}{4}} = 3{:}1$$

Therefore, a fair bet would be in the ratio of 3:1, say a \$25 bet against \$75.

Using the same logic, a probability table for the distribution of heads and tails can be drawn up for a repeated sequence of coin flips. Over an infinite period of time the distribution of realized outcomes will resemble a normal Gaussian or bell-shaped curve.

The normal distribution represented by the bell curve is very popular not only because it occurs in nature with amazing regularity, but because it can be used to unlock powerful statistical tools to predict the likelihood of an event happening, providing certain conditions are met. One of the more important is the requirement that the observations are independently and identically distributed, which can be problematic in financial markets. Leaving that aside for the time being, the mean of the distribution is the most likely outcome, so it represents market expectations, or expected returns. Risk is the chance something can go wrong, which is to say, the outcome fails to meet expectations.

The difference between risk and uncertainty is that risk can be quantified. Quantifying financial market risk is a matter of measuring how much variation there is around the mean (i.e., expected) return. A good measure of variation around the mean is the standard deviation σ. The larger the standard deviation, the more variation there is around the mean, so the riskier the investment is.

Variation from expectations can occur on either side of the distribution, which is to say that the results can be either better or worse than expected. But if the returns distribution is perfectly normal, the chance of better-than-average returns is just as likely as the chance of worse-than-average returns. Chances that the returns will be better or worse than expected can be quantified using the standard deviation. According to the central limit theorem, about 68% of all occurrences will reliably fall within 1 standard deviation of the mean; 95% come within 2 standard deviations, and virtually all (99.73%) will fall within 3 standard deviations.

These statistical properties of the bell curve are the basis for estimating reward potential and risk exposure, using historical data. For instance, suppose that the mean historical return of the stock market is 10% per annum with a standard deviation of 20%. In any given year, we would say with a 95% degree of confidence (2 standard deviations) that the expected return from investing in stocks is 10% with a downside risk of 30% and upside potential of 50%.

But there are two problems here. One is the embedded assumption that the future will look like the past. The other is different time frames for risk and reward.

REWARD, RISK, AND TIME

Reward is typically thought of in terms of long time horizons or holding periods, often months, quarters, years, or even longer. On the other hand, risk has more immediacy. It is typically thought of in terms of days or weeks. Therefore risk (or volatility) is typically expressed as the variance of returns over short time periods like days or weeks. The standard deviation of returns is simply the square root of the variance. Since it is much easier to work with, the standard deviation of returns for the chosen time frame serves as the basis for expressing risk. But for comparability purposes, it needs to be put on a common time scale with reward.

The time scale used for risk estimation will vary depending on the trader's perspective. Some traders may prefer to assess volatility daily; others may prefer one-week periods, two weeks, a month, or longer. However, that leaves risk measured in days or weeks while reward is measured in years. In order to set up risk/reward trade-offs, it is necessary to approximate a common time denominator. This can be done by annualizing the volatility measure.

Annualizing volatility is a matter of multiplying the standard deviation of sample period returns by the square root of the number of time units in the trading year. For example, monthly volatility has 12 annual time units, weekly volatility has 52 time units, and so on. By convention a day count of 252 trading days a year is generally used for daily volatility. On the other hand, average annual holding period returns are conventionally calculated by raising the mean return to the power of the number of annual time units. For instance, consider a sample of daily price changes with a mean μ of 0.0003155 and a standard deviation σ of 0.015.

Mean annualized returns (denoted R) would be:

$$R = (1 + \mu^t) - 1 = (1+0.0003155)^{252} - 1 = 8.27\%$$

Annualized daily volatility would be:

$$Vol = \sigma\sqrt{t} = 0.015 \times \sqrt{252} = 23.8\%$$

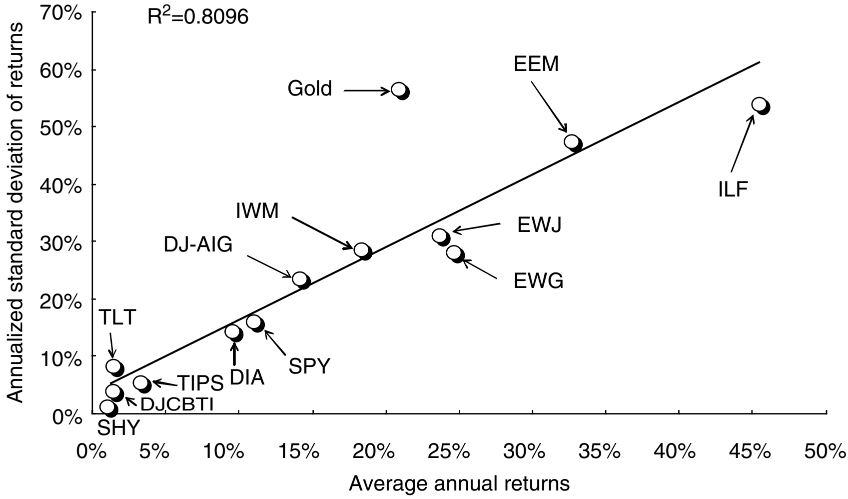
BACK TO THE FRONTIER

With operational definitions of risk and reward in hand, it is possible to use historical data to see how well theory comports with actual results. One way to do this is to collect historical risk and return data for different asset classes and then plot a risk/return scatter plot, as shown in Figure 5.1.

Figure 5.1 plots 3-year average asset class returns against their respective 30-day volatilities. The asset classes in the graph are Treasury securities, U.S. stocks, foreign equities, and spot gold. U.S. Treasuries are represented by exchange traded funds (ETFs) of various maturities based on Lehman indexes that trade on the American Stock Exchange. They are the Lehman 20-year bond fund; intermediate note fund; short-term note fund, and inflation-protected ETF fund identified by their respective ticker symbols: TLT, IEF, SHY, and TIPS. Big cap U.S. stocks

FIGURE 5.1

Risk-Return Scatter Plot: The Efficient Frontier



are represented by Spiders (SPY), the ETF that replicates the S&P 500. The benchmark for small cap stocks is the Russell 2000 (IWM). German stocks, Japanese stocks, Latin American stocks, and emerging market stocks are represented by iShares. Their respective ticker symbols are: EWG, EWJ, ILF, EEM. Gold calculations are based on spot prices.

Figure 5.1 seems to confirm the theory. For the most part, the securities with historically greater returns are also the ones with greater risk, measured by the annualized standard deviation of returns. Moreover, there is very little deviation from the regression line, an approximation of the efficient frontier, that runs through diagram. The exception is gold, whose riskiness seems to far outweigh its returns potential, based on historical patterns. But gold may have special properties that can account for this seeming anomaly, among them its perceived value as insurance against disaster. Suffice it to say that the scatter plot provides some evidence that risk and reward are traded off, moving prices toward the efficient frontier.

VOLATILITY AND STABILITY

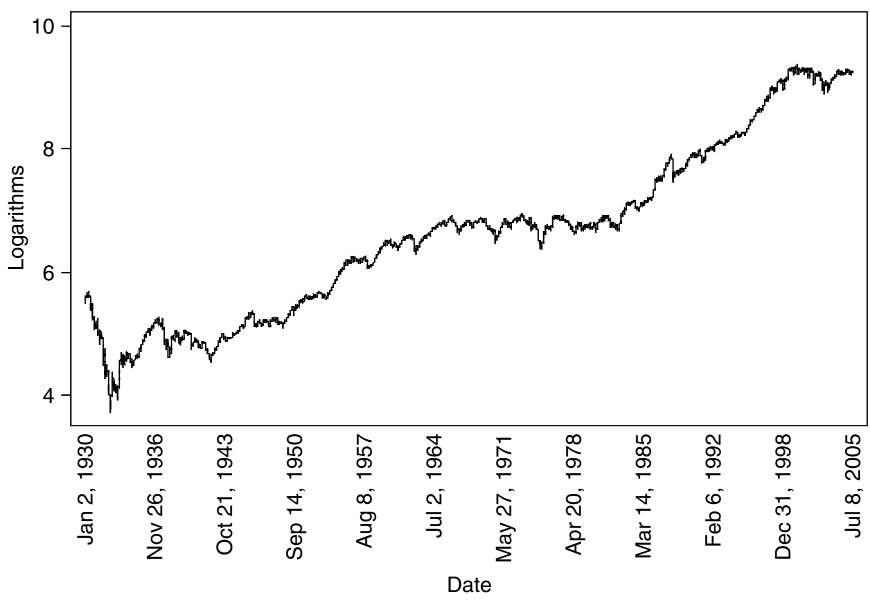
Risk/reward trade-offs are based are implicit forecasts of the future based on past market behavior. If it were not so, there would be no point in calculating historical returns, variances, and standard deviations. How reliable is history as a guide to future volatility and returns? Fortunately, there is a mountain of historical financial market data that is available to analyze this question. These data can then be used to examine whether in fact financial markets have operated over time the way they are supposed to in theory.

In that spirit, consider a very long time series of the Dow Jones Industrial Average using the period beginning January 2, 1930, and ending July 8, 2005. These data, containing 18,965 observations, can be downloaded free from Yahoo! Finance. For simplicity's sake, only price changes will be considered, leaving out dividends. Note that the data set goes back far enough in time to cover all kinds of market conditions. It encompasses bull and bear markets; war (hot and cold); peace; prosperity, depression, and recession; inflation and disinflation; conservative, moderate, and liberal governments; as well as 13 presidents and 6 Federal Reserve chairmen.

Over the very long term, stocks have produced excellent returns, averaging about 11% annually. This performance is summed up in Figure 5.2, a graph of the Dow Jones Industrial Average in a logarithmic scale covering almost 75 years. Using logarithms puts the data in percentage terms, so the graph displays what the market would have "felt like" to participants at the time. But from a trading standpoint, what the market felt like is better expressed in much shorter time units—days instead of months and years.

FIGURE 5.2

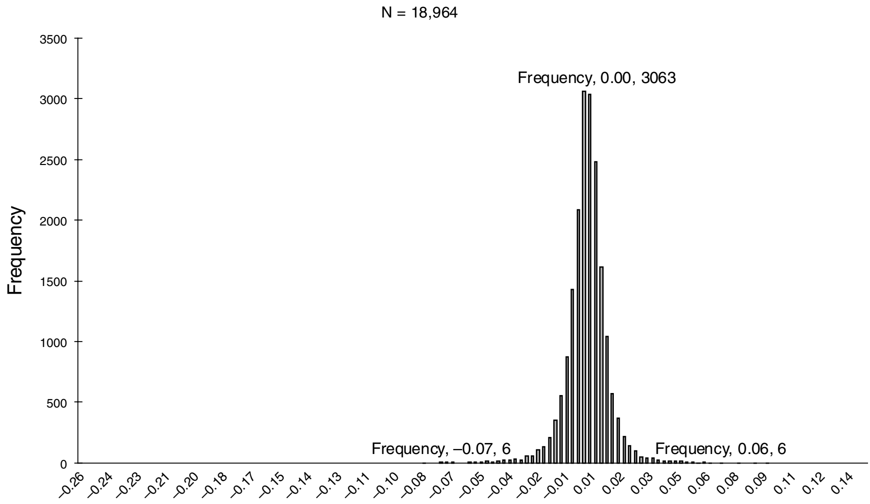
Daily Logged Closing Prices: Dow Jones Industrials (January 1930–July 2005)



Data source: Yahoo! Finance

F I G U R E 5.3

Histogram: Dow Jones Daily Returns (January 2, 1930–July 8, 2005)



When the data are displayed in daily terms, a whole new picture emerges. The market looks a lot less stable and a lot more volatile than it does over the long run. One way to see this is to examine the distribution of daily returns for the sample period, as displayed in Figure 5.3.

The shape of the distribution on display in Figure 5.3 is *not* a good representation of the normal bell curve that forms the basis for so much of the statistical work done on financial markets. It has a high peak and long, low tails that flare out toward the extremes of the distribution. Fat-tailed or leptokurtic distributions like this one have statistical properties that are different in important ways from normal or Gaussian-shaped distributions. These differences are nontrivial; they go to the heart of describing how financial markets behave.

As discussed earlier, in a normal Gaussian distribution about 68% of all observations are expected to occur within 1 standard deviation of the mean; 95% fall within 2 standard deviations and virtually all (99.73%) are expected to fall within 3 standard deviations. How well a curve conforms to this expectation is measured by its kurtosis, which comes to 3 in a perfectly shaped bell curve. In addition, a perfectly shaped bell curve is symmetric. Deviations from the mean are distributed evenly, without tilting in either direction.

An examination of Figure 5.3 and simple descriptive statistics clearly show that the distribution of daily returns for the Dow Industrials does not fall into a normal bell-shaped curve. Instead it is fat-tailed with the tails of the distribution extending far from the mean. Kurtosis, which would normally measure 3, actually measures 26 for this sample. Moreover it is not symmetric; it has a negative skew that measures –0.54. Under these circumstances, the use of standard statistical tools to calculate probabilities is suspect. Events that would normally be considered to be extremely unlikely turn out to be far more likely—and problematic—than would otherwise be expected.

Just how far away financial markets deviate from normal distributions can be seen by expressing daily returns of the Dow Industrials in terms of *z*-scores, a procedure that normalizes the distribution around a mean μ of zero and a standard deviation σ of 1. This provides an easy way to compare actual market behavior with predicted market behavior, which gets to the nub of the question of how markets work in the real world. To this end Figure 5.4 presents a histogram of daily logged returns in the Dow Industrials, expressed in terms of *z*-scores. It also displays a count of the number of returns occurrences that fall between ±7 standard deviations from the mean μ in increments of 1. In theory, 99.73% of the observations should occur within ±3 standard deviations.

There are 18,963 observations of daily returns represented in Figure 5.4. If the returns were normally distributed, returns should be expected to fall outside ±3 standard deviations no more than 0.26% of the time, or about 49 days out of the 18,963 days observed. But when expectations are compared to actual events, it is a different story altogether: Daily returns exceeded ±3 standard deviations on 573 occasions, not 49. That is one time for every 33 trading days, more than 11 times what would be expected by chance.

Financial market time-series data are very tricky, so it is important to take care when analyzing them. There is the danger of biasing the analysis by choice of start date and end date. Circumstances can change over time, data collection methods can change as can data quality. It may be the case that outliers unduly influence the results. In the present case of the Dow Jones Industrials, these objections can be dealt with by examining the data further.

Circumstances do change, and the composition of the Dow Industrials has changed markedly over time. But it remains a benchmark measure of the performance of big cap stocks in the United States. The data sample used in this analysis covers a 75-year period; it can (and will) be sliced into sections to compare how the index fared over different time periods. The most important question has to do with outliers. Are the results of the analysis biased because of a few bad data points?

First there is the question of the definition of an outlier. There are numerous days when the market was exceptionally volatile, with the volatility attributable to major events. Should those days or weeks be

FIGURE 5.4

Histogram of Dow Jones Daily Returns in *z* Scores

Standard deviations	Frequency	
–7	7	
–6	15	
–5	15	
–4	43	
–3	96	*
–2	290	**
–1	1,354	*********
0	7,475	**
1	7,917	***
2	1,354	*********
3	248	**
4	83	*
5	38	
6	11	
7	17	
Total	18,963	

considered outliers? Are they, as is often suggested, once-in-a-lifetime events that can be safely ignored? To consider this question, it is worth looking at a list of some of these market-shaking events. This type of qualitative look at the data can provide some additional perspective that quantitative analyses can overlook. Doing so suggests that market-shaking events are not all that rare.

Even a truncated list of pivotal events would have to include: the Great Depression of the 1930s along with the collapse of the banking system. Then there is the attack on Pearl Harbor, the declaration of war against Japan and Germany, the death of President Roosevelt, President Truman's use of the atomic bomb on Hiroshima and Nagasaki, the Korean War. From the 1960s onward are included the Vietnam War, the assassination of President Kennedy, the assassinations of Senator Robert Kennedy and Martin Luther King, race rioting in major cities across the United States. The Cuban missile crisis brought the world to the brink of nuclear war. There were two assassination attempts on President Ford, one on President Reagan, and one on Pope John Paul XXIII. Anwar Sadat, president of Eqypt, was assassinated by Muslim extremists; the Shah of Iran was overthrown, ushering in the Ayatollah Khomeini and the Islamic Revolutionary Republic. President Nixon, who slammed the gold window shut and imposed wage and price controls, was forced to resign. His attorney general went to jail. There were at least three oil embargoes and several wars in the Middle East.

The Fed eventually adopted monetary targeting for a short while and drove overnight rates to 20%. The Berlin Wall came down, the Soviet Union collapsed, and so did the S&L industry. Iraq invaded Kuwait; the United States invaded Iraq twice. Orange County California defaulted on its debt. President Clinton was impeached then found not guilty at his Senate trial. The United States was attacked by Islamic terrorists on September 11, 2001.

The list goes on. The point being that what appear to be exceptionally rare events turn out to be not so rare after all. As it happens, psychologists have found quite a bit of evidence to suggest that people systematically underestimate the probability of "unlikely" events, a subject that will be discussed in later chapters dealing with behavioral finance. Notwithstanding the statistical evidence to the contrary, it is interesting to note that these large $> 3\sigma$ market moves are commonly described as "once in a lifetime events." Except that since 1930, on average they burst onto the scene every 33 trading days or so. Under ideal laboratory conditions the expected lifetime of the *Drosophila melanogaster*, otherwise known as the red-eyed pomace fly, is about 36 days. And so it turns out that "once in a lifetime" may be an accurate descriptor of these periodic financial market conflagrations if you happen to be a red-eyed pomace fly.

The problem with extreme bouts of market volatility is not simply that they are dangerous. The real problem is that the likelihood that they will occur in the first place exceeds the expectations built into many, if not most, financial market models. That is a problem because the underlying idea is that risk (volatility) can be traded off against reward (return). But if high-impact/high-volatility events are actually far more likely to occur than the models suggest, it is also likely that market prices are less efficient than advertised. By virtue of the risk/reward trade-off, expected volatility is embedded in the price. But if the volatility estimate is suspect because the distribution has fat tails, what does that say about market efficiency? And there is another problem. Volatility is not stable. It tends to cluster.

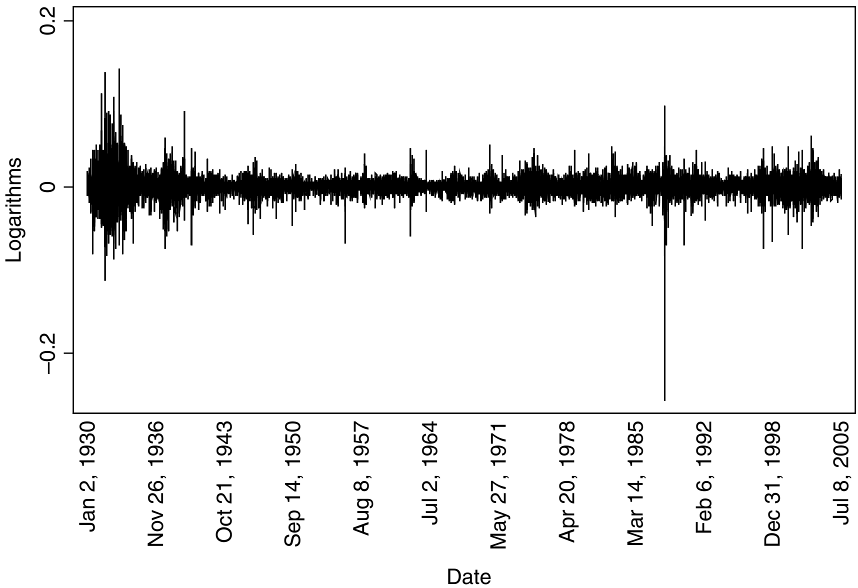
VOLATILITY CLUSTERING

Volatility, it seems, begets volatility, a phenomenon known as *volatility clustering*. That seems only reasonable. When there is a good deal of uncertainty, people pull in their horns and head for the hills. Liquidity can begin to dry up as traders and market makers go into a defensive crouch. Prices move more than they normally would for a given level of trading volume.

Volatility clustering is evident in Figure 5.5, a graphical display of average five-day annualized volatility of the Dow Industrials from January 1930 up to July 2005. The horizontal axis is time. The vertical

FIGURE 5.5

Daily Logged Price Changes: Dow Jones Industrial (January 1930–July 2005)



Data source: Yahoo! Finance

axis is percent.The black grassy lines that make up the body of the graph are volatility observations. It is clear that there are periods of high volatility that cluster together. Then they die out, only to later explode onto the scene again, apparently without warning. The graph shows that volatility is clearly unstable. Periods of relative calm are routinely interrupted by bursts of wild volatility that fade as rapidly as they arrived. This pattern has continued for decades.

Volatility clustering presents a problem. If markets are efficient because rational investors trade off risk and reward, and the risk measure (i.e., volatility) is unstable, subject to clustering, and liable to arrive on the scene without notice, how is it possible to properly take it into account? A partially satisfactory solution may be to use ARCH/GARCH models.

ARCH and GARCH (auto-regressive and generalized auto-regressive conditional heteroskedasticity) models, based on the work of Nobel laureate Robert F. Engle, are designed to fix the problem of volatility clustering by making the variance conditional on past information, which is then incorporated into current estimates. Then the model's parameters are more or less continuously reestimated with advanced statistical software in order to keep volatility estimates current.[1] But models of this type necessarily remain black box configurations. They fail to address why volatility varies; the model's parameters have no intrinsic meaning, and like all time-series models they have difficulty dealing with longer-term phenomena.

This suggests another very recent development worth pondering. As computing power grows along the path of Moore's law and calculation and database software get ever more sophisticated, these tools are increasingly being used to build algorithms that automatically drive trading systems. But there is a tendency for the models to be based on data-mining techniques rather than a priori causal theories. The natural tendency is for the model builders to replicate each other. This can have at least two self-defeating effects. The first is to weaken the independence of the variables that go into the models to begin with. The second is to run the danger of having a lot of people trying to go through the same doors at the same time, setting off low-probability, high-impact events.

SUMMARY

Are the markets efficient? They appear to meet the criteria of weak-form efficiency. It is difficult to show that anyone has come up with a set of decision rules that systematically produces superior risk-adjusted stock market returns. And plotting historical data provides clear evidence of the predicted trade-off between risk and reward. Having said that, it is difficult to square the assumptions (and predictions) of stronger forms of market efficiency with the way markets have actually behaved over long periods of time. Market volatility is far greater than theory would suggest. High-impact (but low-probability) events are actually far more likely to occur than conventional risk models assume. Moreover volatility clusters can (and do) arrive apparently without warning. This suggests that risk/return trade-offs may not fully capture the risks embedded in securities prices. It also suggests that the models may be too narrowly focused.

In this respect it is worth noting that the EMH was developed largely through the study of equity markets. But there are substantial differences in equity and debt markets. For instance, equity markets are mostly auction markets in which exchanges dominate. Bond trading takes place mostly in over-the-counter dealer markets. Bonds are far more influenced by macroeconomic trends than stocks, which are liable to respond more forcefully to firm-specific developments. Bonds are priced from the top down; stocks from the bottom up. That is an important reason why there is far more variation in the returns of individual stocks than there is across individual bonds. In the same vein, equities as an asset class are far more volatile than bonds. Investors take this into account, evidenced by the risk/return scatter plot displayed in Figure 5.1 at the beginning of this chapter.

Stocks and bonds (as well as commodities, real estate, collectibles, etc.) have different risk/return characteristics. They sometimes respond differently to the same data. There are also significant structural and institutional differences in debt and equity markets. But for the capital markets to be priced efficiently, both stocks and bonds have to be priced correctly with respect to each other. That implies that arbitrage money ought to flow between the stock and bond markets to exploit any pricing anomalies between them as new information becomes available. After all, stock and bond markets compete for the same investment dollars. So what drives the market? This question is discussed in the next chapter.

NOTE

[1] See B. Mandelbrot and R. Hudson, *The Misbehavior of Markets*, 2004, especially pp. 247–249.

This page intentionally left blank

CHAPTER 6

What Drives the Market?

Events, dear boy, events.

–Harold Macmillan

The unexpected always happens.

–Margaret Thatcher

Markets are constantly jolted by events seemingly out of the blue (the unexpected always happens). Why?

Economists have often tended to view big market-moving events as external shocks to the system, whose occurrence (and market impact) are inherently unpredictable. That is to say, they are random. This book takes a different view, offering several alternative propositions. First, new information that causes markets to move does not necessarily arrive in random fashion, although it may. Second, new pieces of information are not independent of past information. Data tend to reflect economic trends rather than isolated events. Third, economic trends do not just happen. They have a cause: namely, economic policy. Good policy produces good outcomes and vice versa. Fourth, existing information is not always instantly digested and discounted by the market. Fifth, the market does not always calibrate risk and reward accurately.

Taken together these propositions form the basis for developing an approach to trading. Accordingly, this chapter examines the propositions outlined above. After that it suggests a structure for developing and implementing trading strategies. At heart, trading strategies are designed to profit by anticipating future changes in market prices. But before discussing why market prices change, it makes sense to examine a model that explains why prices are where they are to begin with. The best model for that is the capital asset pricing model (CAPM).

THE CAPITAL ASSET PRICING MODEL

Investing can be a risky business, and investors are generally thought to be risk averse. Given the choice between a risky and risk-free investment with the same expected return, investors will choose the risk-free investment. By implication, risk-averse investors need some kind of inducement to make risky investments. The inducement takes the form of higher expected returns. Faced with a choice of risky and risk-free alternatives, investors will tend to trade off risk against expected return until the desire for more expected return is matched evenly with a desire to avoid risk. The stock price therefore represents the market's assessment of expected return and risk.

But that implies a question: How much additional potential return is needed to induce investors to swap out of risk-free assets into more risky ones? The CAPM provides a framework for thinking about this. Originally developed by William Sharpe, John Lintner, and Jan Mossin, the CAPM offered a way to quantify risk/return trade-offs and optimize performance by diversifying portfolio holdings. Sharp was awarded the Nobel Prize in economic science in 1990 for his work on the CAPM.

In CAPM there are two types of assets: risk-free and risky. And there are two sources of risk: systematic and idiosyncratic. Idiosyncratic risk is specific to a particular company; systematic risk is pervasive. Idiosyncratic risk can be diversified away by not putting all your eggs in one basket. Systematic risk goes with the territory; by definition it cannot be diversified away. The extent to which the risk in a company's stock is idiosyncratic as opposed to systematic can be measured and controlled with mathematical models.

The starting point is the risk-free rate of interest. All other returns represent payment for the assumption of risk. The difference between the risk-free rate and expected additional returns on the margin is the risk premium. A good proxy for the risk-free rate of return is the three-month Treasury bill rate. There is no threat of default and price movements resulting from interest rate changes are minimal. The risk portion of return is related to asset price volatility. Why asset price volatility? Asset price volatility comes from the introduction of new information. Returns expectations adjust to take the new information into account. The more uncertainty there is, the greater the market's discount. If the rate of return were known with certainty, it wouldn't be a risky asset.

There are many sources of potential uncertainty that affect the earnings of a firm. Capital structure, the regulatory regime, industry position, competitiveness, innovation, and management are only a few of the more obvious ones. The efficient market hypothesis coupled with CAPM rolls them all up into how volatile a company's stock is. To the extent that all relevant information is known and priced into the market, variation in the price of the stock, relative to all other stocks, represents the level of uncertainty associated with the stock. The operational measure of risk is variability, or more precisely, variance. Stocks with predictable earnings are less risky than stocks with highly variable earnings.

The degree to which a company's stock varies with the market as a whole therefore is a measure of how risky it is in relative terms. That measure is called the stock's *beta*. For instance, if on average, the stock of Ron's Flawless Jeans typically rises 1.1% when the stock market rises by 1%, its beta would be 1.1. On the other hand suppose the stock of another company, Rich's Quilt Factory, typically rises only by 0.9% when the market rises by 1%. It would have a beta of 0.9. Companies whose stocks move in the opposite direction of the overall market are negatively correlated. Stock betas can be estimated by using regression analysis. Regression studies cover different time frames and various market environments, and they typically compare stock price changes against the major benchmark indexes. Stock betas are routinely posted on finance Web sites. Yahoo! Finance, for instance, publishes estimates of stock betas under key statistics when a ticker symbol is keyed in.

Once a stock's beta is estimated, it can be used in conjunction with the risk-free rate and the expected market rate of return to calculate a stock's required rate of return. The required (or expected) rate is the one needed to induce an investor to substitute a risky for a risk-free investment. It is expressed in the CAPM equation below,

where:

$$\text{Exp}\ (R) = R_{\text{RF}} + \beta[\text{Exp}(R_M) - R_{\text{RF}}]$$

Exp = Expected
R = Rate of Return
RF = Risk-Free
M = Market
β = Beta

The equation can be easily solved to find the required rate. Taking the example above, assume a risk-free rate of 5%, historical long-term stock market returns of 11%, and a beta (β) for Ron's Flawless Jeans of 1.1%. Substituting those values into the equation, the required rate of return is 11.6%.

$$\text{Exp}\ (R) = 5\% + 1.1\% \times (11\% - 5\%) = 11.6\%$$

The difference between the risk-free rate and the expected market rate of return, in this example 6%, can be thought of as the rate of inducement for investing in a risky security. In other words it is the risk premium. Note that the risk premium needs to be greater than zero.

Otherwise risky assets would wind up being overpriced—the model would imply the same or deeper discounts for risk-free assets, a result that makes no sense.

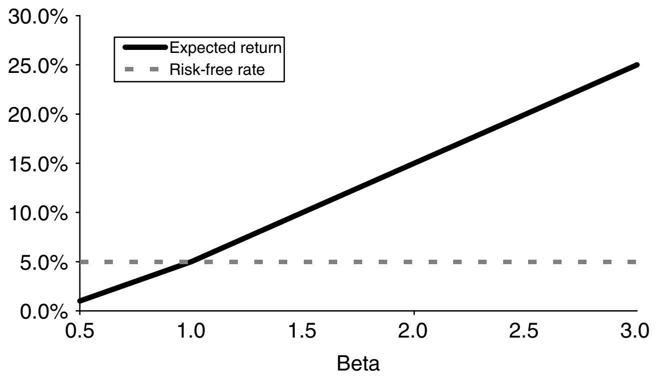
Another (closely related) way to estimate the required return is to use the Treasury yield for the relevant time horizon as the baseline return, and then add in a risk premium. The Treasury yield is readily ascertainable, but it is difficult to quantify risk factors that affect earnings. Variables that affect earnings volatility include sensitivity to the business cycle, capital structure, industry domain, and competitive position. A good estimate of the proper risk premium needs to take all these factors into account, plus others.

In theory, as investors trade off risk for reward, stock by stock, the market as a whole begins to be efficiently priced in the sense that it reflects the preferences of all the participants, based on all known information. If a stock becomes overpriced relative to the others, it will be sold until its price decreases enough to increase its expected return. Conversely, if a stock is priced too cheaply compared to the rest, it will be bought until its price rises enough so that its expected (risk-adjusted) return falls. Again, in theory, each stock has some equilibrium value compared to all the others, along a risk-return continuum. Once expected returns and betas have been calculated for all the securities in the marketplace, a hypothetical portfolio (the market portfolio) can be constructed using the entire universe of stocks. Then the pricing of each stock, adjusted for risk, can be compared to the market portfolio and a determination can be made if the stock is priced too high, too low, or about right.

Figure 6.1 graphically illustrates the trade-off between risk and reward. As the risk (or beta) increases along the horizontal axis, expected

FIGURE 6.1

The Security Market Line



returns increase as well. The line depicting this is the security market line (SML). Stocks priced in the zone over the line are too cheap by the CAPM model. The discount rate (the expected return) is too steep for the risk. Conversely, the stocks priced below the SML are too expensive. Their expected returns don't compensate owners enough for the risk they carry.

The process extends across asset classes. Not only are individual stocks compared with each other, but entire asset classes (stocks, bonds, commodities) are evaluated with respect to each other. Labeling a stock or a bond as too cheap, too rich, or just about right implies some valuation process. And the point of the valuation process is to profit by buying the assets that are too cheap while selling or avoiding the ones that are too expensive. Generally speaking, valuation processes tend to fall into one of two categories: bottom up and top down. Bottom-up models are based on individual company level data. Top-down models are those that depend on macrolevel economywide variables, rather than on company specific information. As new information becomes available, valuations change to reflect that information. Anyway, that's the theory. Now let's see how it works in the real world.

THE FULCRUM OF FED POLICY

In theory, the price of a security reflects the present value of expected future cash flows, adjusted for risk. The CAPM provides a framework for risk adjusting the discount rate. The key variables are the risk-free rate, long-run expected returns, and volatility. For all intents and purposes the risk-free rate is set by the Federal Reserve in its conduct of monetary policy. Similarly the risk premium is heavily influenced by the Fed, since it is the difference between the Fed-administered rate and long-run expected returns. (Other variations of the risk premium use intermediate government bond rates as the measure, but they too are highly correlated with Federal Reserve operations). Finally, volatility is itself volatile, partly because it is influenced by the level of interest rates. No matter how you slice it, interest rate levels, and in particular the Fed, are key driving forces in the capital markets.

All this suggests that information flows into the marketplace should be seen through the filter of how that information is liable to influence both the risk-free rate and the risk premium. First, it necessitates a thorough understanding of the Fed (how it conducts policy, its structure, operating procedures, and instincts) as well as an understanding of the government bond market and its place in the American political economy. Second, it means that information should be evaluated with a view to how it may affect risk preferences, not just in the United States, but globally. In short: How does new information affect market expectations and policy making?

NEW INFORMATION

Does new information come to the market at randomly generated intervals? Of course not. Some of the most important market-moving information is scheduled well in advance of its release. Virtually all macroeconomic data are released on a regularly scheduled basis. For example, with rare exceptions, the monthly employment report is released the first Friday of every month. GDP is released quarterly. Companies announce ahead of time the day and time (often after the close) when they will release earnings reports or guidance. Similarly, traders set up or flatten out positions ahead of macroeconomic releases depending on their market views.

Leave aside for a moment the likelihood that positions and data releases cluster around tightly defined time frames. Conduct a thought experiment and assume that macroeconomic data releases and earnings reports are not scheduled for specific release times. Imagine instead that the numbers are released as soon as they are calculated, at whatever time of day. Then the releases would hit the market essentially at random with respect to market positions, but they would still form a time series. However, the release dates would still be sequentially ordered. February's employment data would be reported sometime in March. March's employment data would be reported sometime in April and so on.

The sequential ordering implies that the pieces of data being released to the market are not independent of each other unless economic activity is conducted randomly, hardly a reasonable proposition. To the extent that there are economic cycles and multiplier effects, each piece of economic data that is released may provide a hint about what the state of the economy might be with respect to important economic variables like GDP, inflation, employment trends, capital investment, retail spending, and so on. These are data that can have an impact on markets, policy makers, and expectations generally. But these data are not isolated bits of information. They are likely to be correlated with other macrodata. As such they can provide evidence that tends to confirm, or cast doubt on, a trend.

For instance, employment trends are correlated with consumer spending, construction spending is correlated with interest rates, and interest rates are (negatively) correlated with inflation. In turn, both inflation and inflation expectations are correlated with the Federal Reserve's conduct of monetary policy, which in turn is partly presaged on employment trends and productivity growth. These are factors that are deeply embedded in the business cycle, market structure, and institutions of the economy. They don't happen all at once. They take time to unfold, and they take time to reverse.

INFORMATION PROCESSING

The length of time it takes for information to be processed and discounted in the marketplace is a key point of contention. To the extent that information—however defined—is digested and discounted rapidly and accurately, the marketplace can be said to be *informationally* efficient. Superior analysis is unlikely to produce trading profits in excess of normal market returns. The evidence usually proffered is (1) the absence of traders who have shown superior returns resulting from better (and legal) collection and analysis of information, or (2) the apparent absence of a set of straightforward and unambiguous trading rules that would have produced superior returns based on past history and that are likely to be applicable today. On the other hand, no less a theorist than Fischer Black has argued that investor risk preferences can change without being fully reflected in market prices until after a considerable time lag. Following this line of reasoning, he went on to argue that the stock market crash of 1987 reflected belated recognition of a yearslong shift in investors taste for risk.[1] Moreover the range of values in which the market could trade without being fundamentally out of equilibrium is fairly wide in the world of Fischer Black. If so, there is plenty of room for trenchant analysis to yield superior results.

This book is sympathetic to the argument that keen insight can yield superior trading results. Some circumstances are more conducive than others to profit making. There are other times when enormous, but largely unrecognized, risk is embedded in the system. And there are trading guidelines that, when followed, are more likely to produce profits than losses. The trick is in understanding how to adapt trading rules of thumb to developing market dynamics. Oftentimes the key to the puzzle is monetary policy.

FED POLICY, MONEY MARKETS, AND COIN FLIPS

A fundamental tenet of the efficient market hypothesis is that short-term market movement is random. As a result the rational trader should have an expected return of zero; the same as a gambler who takes even money bets on the flip of an evenly weighted coin. The difference between gambling and investing (not the same as trading) is that an investor can rationally expect positive returns over time. With securities, the underlying trend (or drift rate) dominates over the long term as wealth is created and value added to productive enterprises. On the other hand, gamblers can flip coins ad infinitum, and their returns expectations remain firmly lodged at zero. Value is neither created nor destroyed. But if a gambler could make an infinite or at least very large series of bets on the flip of a coin weighted to come up heads 55% of the time, that gambler would expect to emerge a winner. Is there an analogy to a weighted coin in the securities markets?

Perhaps there is. On any given day movements in the price of a stock may very well be random and therefore unpredictable. But short-term interest rates are another matter entirely. Short rates are where they are because the Fed puts them there. That is how it conducts monetary policy. Monetary policy is not decided by coin flips. It is a deliberate process. Moreover Federal Open Market Committee (FOMC) decisions tend to be telegraphed well in advance. And there are active markets in fed funds futures and options that can be used to gauge market expectations of Fed policy.

Fed policy dominates Treasury market pricing, particularly at the short end of the yield curve, which is where the bulk of the trading takes place. Short-term Treasury bills and coupon bearing notes with maturities under three years account for about $200 billion or more a day. Bid-ask spreads are especially narrow, typically about 1/128th of 1/32nd of 1 point, or about $78 per million in par value. Quotes are freely available, published daily by major newspapers. Electronic quote vendors publish live quotes as well, pretty much 24 hours a day.

The Treasury market is extraordinarily important. It serves as the baseline for U.S. (and possibly) world capital markets, and for good reason. First, the Fed effectively sets the risk-free rate when it conducts monetary policy using Treasuries. Second, Treasuries are free of default risk. Third, the United States is by far the biggest economy on earth, and U.S. dollars are readily accepted around most of the planet. Fourth, foreign central banks are large holders of Treasuries, which they use to manage their own macropolicies. In short, Treasuries are a textbook example of a market that should be efficiently priced.

Paradoxically, the short end of the Treasury market may fail one of the tests of efficient pricing. Consider the import of the Fed's use of the Treasury market to conduct monetary policy. When the Fed wants to ease policy by increasing the money supply, it buys Treasury securities until it drives rates down sufficiently and vice versa. The problem is that market efficiency demands randomness in short-term price/rate movements. If Fed policy is to drive short-term interest rates in a particular direction, day-to-day movements in short rates are likely to be systematically biased in the direction of Fed policy, which is to say, nonrandom. Are they?

Beginning in 1994 the Fed adopted a policy of announcing its target for the federal funds rate, the principal benchmark for monetary policy. This presents a situation in which virtually all information about policy is known. There is little or no ambiguity about Fed intentions, and there is a highly liquid, easily accessible market directly affected by the Fed's policy stance. Moreover, Fed policy meeting days are regularly scheduled and announced via the Fed's Web site well in advance, sometimes by months and years. Virtually every utterance of Fed policy makers is followed in minute detail by an army of economists, analysts, and traders. Any arbitrage opportunities that exist ought to be fleeting. It amounts to a natural experiment that can be used to test the hypothesis that daily price changes on short-term Treasury notes are randomly distributed.

If the market reflects all known information and price changes in short-term notes are randomly distributed, there should be no systematic difference in the average overnight change of two-year Treasury note yields during periods of Fed ease or tightening. Otherwise traders would be able to beat the market by systematically adjusting their trading strategies in accordance with the Federal Reserve's policy stance. To test the proposition, we compare market changes in Treasury two-year notes during periods of Fed tightening and Fed easing. The time periods are easy to cordon off using official FOMC policy statements. The announcement dates, the direction and magnitude of the rate change, and the targeted funds rate are all posted on the Web site of the Fed's Board of Governors.[2]

Accordingly, that information can be used to categorize the Fed's policy stance as tight, easy, or neutral. Tight periods are ones in which the last two interest rate decisions moved the funds rate higher; periods of ease are ones in which two successive rate decisions moved the funds rate lower. Intervening periods are categorized as neutral. For instance, suppose the Fed announced an increase in the federal funds rate on March 10, May 10, and June 10, but no change on July 10 and a reduction on August 10 and September 10. The period from March 10 to June 10 would be classified as "tight," the June 10 to August 9 would be "neutral," and August 10 through September 10 would be "easy."

The market's response to these policy changes can be gauged by referencing the St. Louis Fed's publication of daily closing yields on constant maturity two-year Treasury notes. Methodologically the use of the St. Louis Fed data has several notable benefits attached to it. The first is that the base of comparison is constant maturity Treasuries, so there is no bias from maturity drift. The second is that comparing overnight yield changes with constant maturity notes leaves out cost-of-carry complications. The third is that the data are collected and calculated on a consistent basis.

In addition, the classification scheme has the advantage of being unambiguous. Policy is as defined by the Fed. It also clearly defines policy with respect to actions taken rather than anticipation. For instance, markets sometimes anticipate and discount policy shifts, in whole or in part. This taxonomy clearly separates time periods by actual policy shifts. Periods in between are considered neutral. As a result, two questions can be investigated for the sample period. First, to what extent is daily directional change associated with the Fed's policy stance. Second, what is the magnitude of

FIGURE 6.2

Relationship between Fed Policy and Daily Directional Rate Change for Period from February 4,1994 to July 15, 2002

Market Direction	Fed Policy			Total
	Ease	Neutral	Tighten	
Lower Rate	403 50.63	414 43.22	450 40.58	1,267 44.25
Unchanged	71 8.92	101 10.54	149 13.44	321 11.21
Higher Rate	322 40.45	443 46.24	510 45.99	1,275 44.53
Total	796 100.00	958 100.00	1,109 100.00	2,863 100.00

Pearson chi2(4) = 24.0651 Pr = 0.000

Key
Frequency Column Percentage

change? The sample period (February 4, 1994–July 15, 2005) covers slightly over 10 years, during which there were about 2,800 trading days. Days categorized as falling in "easing" or "tightening" periods were split about evenly, each accounting for about 44% of the available trading days. Neutral periods accounted for about 11% of the trading days.

If price (rate) changes are distributed randomly, up days and down days should be distributed approximately evenly across Fed easing, tightening, and neutral periods. But they are not. During periods of policy ease, two-year rates fell on 50% of the trading days, but rose only 40% of the time. Similarly during tightening periods, rates rose 46% of the time, but fell on only 41% of the trading days. See Figure 6.2. Standard statistical tests indicate less than a 1% probability that the association of market direction and policy stance is the result of chance.

In addition, rate changes can be compared across easing and tightening periods to see if, on average, they are different. To do this, the easing and tightening periods are compared directly, with the neutral periods excluded. This comparison shows that during periods of Fed ease, on the average day two-year rates fell by about 1/2 of a basis point. During tightening periods, two-year rates rose about 1/3 of a basis point. The total margin of difference came to about 0.86 basis points. Standard statistical tests indicate less than a 1% probability that these results, displayed in Figure 6.3, are random error.

That short-term yields rise during periods of Fed tightening and fall during periods of ease is blindingly obvious to the overwhelming majority of Treasury note traders. In a way that is the point. Even though the

FIGURE 6.3

Two-Sample T-Tests with Equal Variances Comparing Magnitudes of Change Across Easing and Tightening Periods from February 4, 1994 to July 15, 2002

Group	Obs	Mean	Std. Err.	Std. Dev.	[95 %Conf. Interval]	
Ease	796	–.0053894	.0025335	.0714784	–.0103626	–.0004163
Tighten	1109	.0032281	.0016133	.0537242	.0000627	.0063935
Combined	1905	–.0003727	.0014181	.061895	–.0031539	.0024085
Diff		–.0086176	.0028692		–.0142448	–.0029904

Degrees of freedom: 1903

Ho : mean(Ease) – mean(Tighten) = diff = 0

Ha : diff < 0	Ha : diff ! = 0	Ha : diff > 0
t = –3.0034	t = –3.0034	t = –3.0034
P < t = 0.0014	P > |t| = 0.0027	P > t = 0.9986

phenomenon is well known, the data suggest that it still has not been arbitraged away as theory would suggest. Further, this is not simply a phenomenon that is obvious in retrospect. For over 10 years the Fed's policy pronouncements have routinely included forward-looking statements that often strongly suggest what the Fed's future course of action is likely to be. Furthermore, two-year Treasury notes have moved in tandem with Fed policy since they were first issued about 30 years ago. If you know what the Fed is going to do, you can be very confident of what two-year Treasuries are going to do. The close relationship between two-year yields and the funds rate can be easily seen in Figure 6.4.

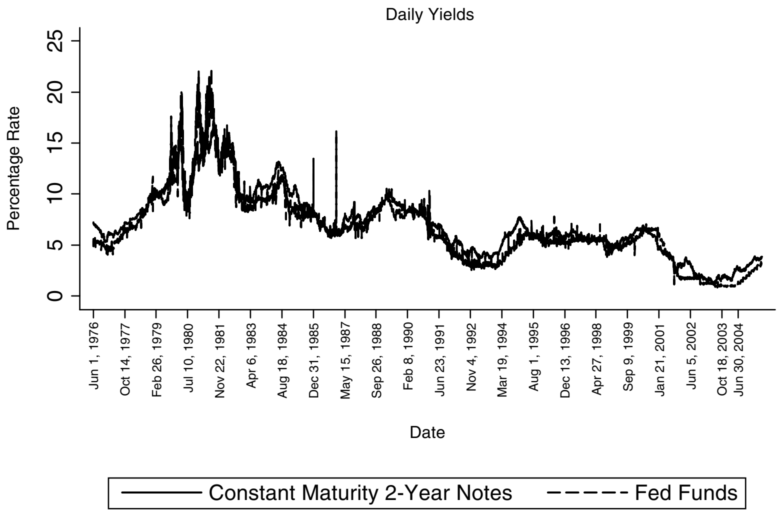
The data on fed funds, Fed policy, and two-year T-note yields are striking in several respects. First, the close relationship between short-term Treasuries and fed funds is evident. Second, yield changes are clearly not randomly distributed. They trend either up or down for considerable periods of time consistent with the Fed's policy stance. Third, the Fed's policy stance is not random, it is not unpredictable, and it does not change day to day. By themselves these results are not the stuff of hard-and-fast trading rules. But they do suggest that traders can tilt the odds in their favor, much as card counters do at the blackjack tables.

COUNTING CARDS TO CALIBRATE RISK

Patrons caught counting the cards at the blackjack tables are invited to leave by the management of the casino—and for good reason. By counting the cards already played, card counters are better positioned to gauge the odds of what they are likely to draw (or not draw) if they take a hit from the dealer. The odds of winning constantly change, but the payoff remains the same. Card counters can tilt the odds in their favor by dynamically

FIGURE 6.4

Federal Funds: Constant Maturity 2-Year Notes



Data source: Yahoo! Finance

adjusting their play as they recalculate the odds of winning, even as the payoff remains static.

Traders in the capital markets can adopt the same types of dynamic strategies by thinking probabilistically and adjusting positions dynamically as circumstances change. The key is to recognize the indicators that signal changing circumstances. Sometimes the indicators are in conventional economic statistics and Federal Reserve policy statements, which are relatively easy to interpret. Sometimes the indicators are obvious only in retrospect and take a while to work their way through the system. But they are the ones that can have big payoffs down the road. That is why it is important to have a broad strategic sense of the state of the world. Consider, for instance, the strange case of Mathias Rust and the story it told.

A CESSNA AIRPLANE, THE COLD WAR, AND THE U.S. MONEY SUPPLY

On May 19, 1987, 19-year-old Mathias Rust took off in his single-engine Cessna 172B Airplane from Hamburg, Germany. He reportedly refueled at Helsinki-Malmi Airport, before telling air traffic control he was going to Stockholm. But, as it turns out, he didn't go to Stockholm.

Instead, he headed for Moscow, eventually landing in the middle of Red Square after ducking underneath Soviet radar. The Soviet establishment that had shot down a Korean civilian flight in September of 1983 was either unable or unwilling to stop a small plane from violating the airspace over its capital city. The Cold War was effectively over. The Soviet regime had lost its nerve.

The Soviet empire began to crumble soon after. In September 1989 Solidarity formed the first noncommunist government in Poland in 40 years. The communist regime of Nicholas Ceausescu in Romania was overthrown in the Christmas Revolution of 1989. In August of 1989 Hungary opened its border with Austria, allowing visiting East German tourists to escape to the West. By October 1989 East Germany's dictator Eric Honecker abandoned ship, and the Berlin Wall came tumbling down on November 9, 1990. It was literally torn apart by thousands of East Berliners. The entire spectacle was carried live and in color on television.

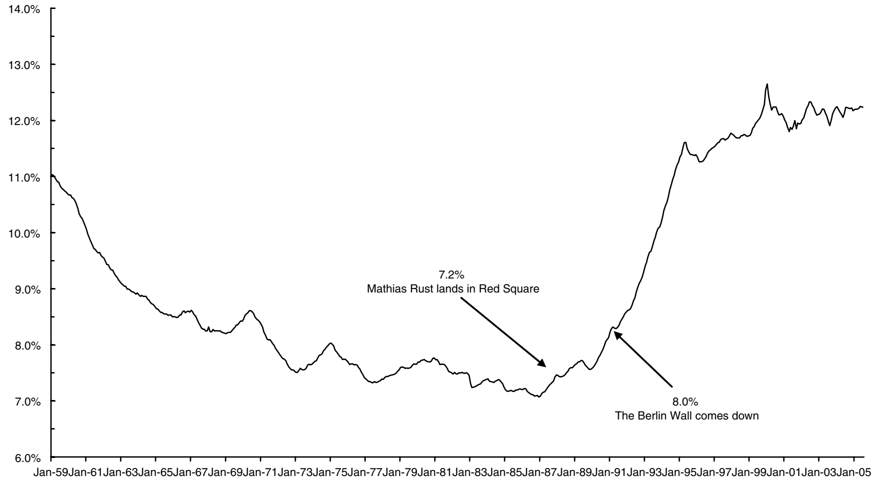
The sudden collapse of the Soviet Union left Eastern Europe without a functioning banking system. There was no real price system; there were no real capital or money markets, and local Soviet-era currency was essentially useless. But there was another currency that mattered: U.S. dollars, and in particular $100 dollar bills. A stunning rise in demand for cash dollar balances coincided with the collapse of the Soviet empire. U.S. dollars became the currency of choice in Eastern Europe, serving as a store of value and a medium of exchange. Still today, about two-thirds of U.S. currency is estimated to be held outside the United States

Cash is expensive to hold. It pays no interest. It is inconvenient for all but the most trivial transactions, and it is easily lost or stolen. In the era of Internet banking, direct deposit accounts, money market funds, ATMs, automatic bill pay, cell phones, and credit cards, it is hard to imagine that demand for cash would rise relative to other forms of money. But it did. After decades of more or less continual decline, all of a sudden demand for U.S. currency began to soar. Measured as a percentage of M2, a measure of the broad money supply, demand for U.S. currency is higher than at any time since 1959. See Figure 6.5.

The collapse of the Soviet Union was a boon to the global economy in general and the U.S. economy in particular. U.S. defense spending fell off rapidly, more capital was available for productive investment, U.S. companies with global brand names profited handsomely, and the dollar cemented its position as the world's reserve currency. The process was neither neat nor linear. It had enormous geopolitical significance that was bound to have a significant impact on market behavior.[3] More to the point, it is preposterous on the face of it to suggest that as the Berlin Wall

FIGURE 6.5

Currency as a Percentage of Money Supply (Seasonally Adjusted)



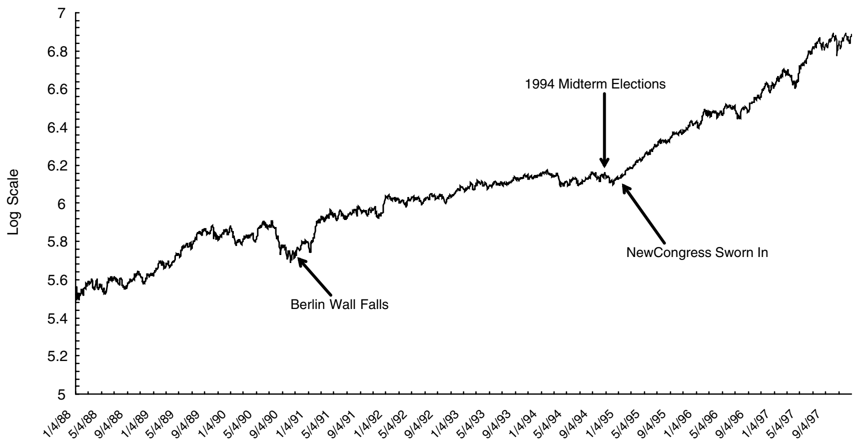
came down, the significance of it was digested by the market virtually instantaneously.

In fact, an examination of recent history suggests that financial markets can take quite a bit of time to digest some types of news. Key events serve as market catalysts, but the full impact of events may take a while to unfold. This suggests that superior insight has a role to play in assessing events and devising strategy. For instance, the evidence indicates that a strategy based on following Fed policy is more likely than not to reap profits. (Techniques for implementing these strategies will be discussed in later chapters.)

Consider Figure 6.6, a graph in log scale of the daily closing prices of the S&P 500 from January 1988 through January 1998. In the period immediately preceding the fall of the Berlin Wall, the stock market is clearly under pressure. Just as clearly, the market stabilized and began to do substantially better after the event. But the gains did not occur all at once. Far from it; the market rallied substantially through the end of the year before a significant downtick. Similarly, in 1994 the stock market had been drifting. However, the midterm elections ushered in a historic change of control in Congress and brought about divided government. The policy debate shifted toward fiscal sobriety and reduced, for a time, government overreach. It was a catalyst for a rally that lasted for years. It wasn't over in a day.

FIGURE 6.6

S&P 500 (January 1988–January 1998)



Data source: Yahoo! Finance

INFORMATIONAL EFFICIENCY RECONSIDERED

The examples cited above point to some real world problems with the idea of informational efficiency. First, events with potential market impact do not all arrive at random. Some are correlated with past information and reflect, or confirm, an existing trend. Second, even if these events did arrive at random, their import may not be fully and immediately digested. The consequences of important events can take time to play out. By implication, it is possible to use high-quality analytic skills to look out over the horizon to generate superior returns. Third, the trade-off between risk and reward is not always perfectly priced. There are times when the risk of loss is relatively low compared to the potential for gain. Certain times are better than others to play the game.

There is another aspect of this to consider. Bachelier's assumption that market prices embed all that is known may be incorrect as a result of strategic behavior. Remember, Bachelier posited that market prices reflected the highest price buyers are willing to pay and the lowest price at which sellers will sell. Otherwise the buyer (seller) would immediately take the offering or hit the bid. But this assumption neglects to take into account the possibility of strategic behavior.

There are several possible reasons why traders sometimes fail to reveal their best prices. There may be price discrimination in which better prices are shown to some but not others. There may be a strategy of holding back, waiting to see if more size shows up on the bid or offered sides of the market. After an initial position is acquired, there may be an intent to buy or sell more at a later time in an aggressive fashion in an attempt to influence market prices. These types of trading tactics are fairly common practice in the markets. So there is ample reason to suspect that Bachelier's operating assumption may be faulty.

The possibility of strategic trading behavior as well as the gradual unfolding of reaction to geopolitical and economic circumstances, business cycles, and market reaction to Federal Reserve behavior all point to the overwhelming importance of superior analysis. What really matters is market insight stemming from data interpretation. In this respect context is all-important. It is important both with respect to economic fundamentals and market psychology. Market fundamentals include both short-term cyclical phenomena as well as larger structural factors that usher in secular bull and bear markets.

It is the contention of this book that secular bull and bear markets do not occur at random. They occur for a reason. Among the most important are economic and financial policy making. Policy that emphasizes price stability and market competition produces conditions that encourage investment, growth, and robust capital markets. That would apply to the great bull market that began in 1982 and lasted for the rest of the century. Policy making that is misguided (1930s) or hostile to the free interplay of markets (1970s) produced the two great bear markets of the 20th century. But most of the time policy just seems to muddle along in the middle. What are the signs that policy is tilting in a way that may have significant market impact?

POLICY AND MARKETS

The answer to the question lies in defining the conditions that markets find friendly. They include protection of property rights, sanctity of contract, the rule of law, transparency in government, low inflation, low tax rates, a simple tax structure, and a regulatory regime designed to promote market competition and a level playing field. Whether, and to what extent, political freedom is conducive to economic growth remains a hotly debated topic.

But there is little doubt that economic freedom produces better economic outcomes. The Heritage Foundation publishes a yearly index of economic freedom based on a nation's trade policies, tax regime, government intervention in the economy, monetary policy, capital flows, banking system, wages, prices, property rights, and informal economic activity. The correlation between freedom (as defined by the index) and per capita income is unmistakable. The countries that are defined by the

Heritage Foundation index as being either "free" or "mostly free" have per capita incomes anywhere from three to seven times higher than the countries defined as "mostly unfree" or "repressed".[4]

The great majority of capital market activity takes place in advanced countries that have liberal market-based economies. In other words, free. From the standpoint of market performance, the presumption is that free institutions and open and competitive markets will produce the best market outcomes. But policy making is not a one-way street. There are cycles of liberalization and backtracking. In this respect policy cycles can be every bit as important as business cycles.

THE THREE PRONGS OF POLICY MAKING

There are three main prongs of economic policy making. The first is fiscal policy. The second is monetary policy. The third is the regulatory regime. All are important, but monetary policy is probably the most important on a day-to-day basis.

Fiscal policy involves taxes and spending. The level of taxation, expressed as a percentage of GDP, is an often-used but crude measure of the size of government and its influence on economic activity. A better measure is government spending, because spending measures resource utilization. Taxes on the other hand only measure a portion of how the spending is financed. The other portion is financed through bond sales. The best (and most relevant) measure from a capital markets perspective is current spending plus the present value of net future liabilities. That would include unfunded liabilities from social insurance programs like Social Security and Medicare. At the moment, most estimates put the present value of those unfunded liabilities north of $50 trillion, a definite source of concern in the marketplace.

Regulatory policy deals with all sorts of rules and regulations designed to promote health and safety, to limit negative externalities, to level the playing field, and–promote vigorous market competition. Unfortunately, regulatory agencies can be captured by the industries they are designed to regulate and then wind up doing more harm than good by limiting competition. While fiscal policy and the regulatory regime are important to the overall economic health of a nation and its capital markets, policy adjustments tend to be on the margin. Consequently, impact analysis of specific tax and regulatory requirements are often industry or company specific. However, it is fair to say that flatter and simpler tax regimes are more likely to attract productive investment, all else being equal.

Monetary policy is the most immediately important policy branch for capital markets because it substantially determines both the risk-free rate of return and the rate of inflation. It heavily influences inflation expectations along with consumption and investment decisions, and therefore the business cycle. All of these are important factors for pricing capital market instruments.

Most of the time policy adjustments are made on the margin in response to wiggles in the business cycle. Under those circumstances, markets react in typical cyclical fashion. However, on some occasions policy adjustments are fundamental and substantial. These periods of regime switching are fraught with peril, but they also offer major market opportunities.

REGIME CHANGE

Does regime switching matter? Compare and contrast two distinct policy eras. The first era begins in the 1960s and extends through the 1970s. It is a time of rapid growth in domestic government programs; Keynesian demand management dominates the economic policy agenda; the Vietnam War and its aftermath impinge on virtually all areas of government decision making. It is an era of an extensive and growing regulatory regime. Marginal tax rates are high; gold ownership (other than for jewelry and dentistry) is prohibited; capital controls are pervasive; interest rate ceilings restrict the competition for funds; and much of American business is cartelized.

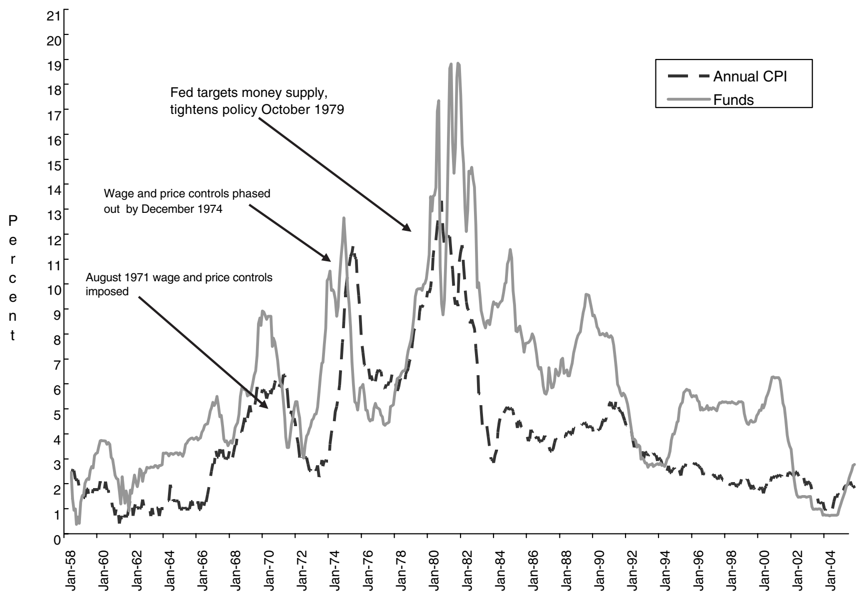
The second era is the one ushered in by the Reagan counterrevolution which began in the 1980s and extended through the Clinton years. This era is characterized by successive waves of deregulation, significantly reduced marginal income tax rates, reductions in capital gains tax rates, freeing up of the capital and money markets, large federal budget deficits, the intensification and end of the Cold War, restructuring of American industry (often by hostile takeovers), and deepening global economic integration.

The most important dividing line between the two eras is the behavior of the Federal Reserve. It began to pursue an independent and nonpoliticized monetary policy in late 1979 and has continued to do so ever since. The results can be seen in the performance of the economy and financial markets. To do this, we can examine two data series. The first is the performance of the economy measured by growth in GDP. The second is a measure of Federal Reserve monetary policy.

Characterizing monetary policy over time can be difficult because the conditions and circumstances under which it is conducted are likely to vary enormously. But ultimately the goal of monetary policy is to maintain price stability, so the best measure of monetary ease or tightness is with respect to inflation. The Fed's policy stance is thus best measured by

FIGURE 6.7

Federal funds and inflation (1958–2005)



the relationship between the interest rate it directly controls (the federal funds rate) and the rate it tries to manage (the inflation rate). A pivotal interruption in the time series occurred on October 6, 1979. On that day the Fed announced a change in its operating procedures; the Fed would no longer target the funds but would target the money supply instead, allowing the funds rate to trade in a wide range in the marketplace. The effect of this policy change can be seen in Figure 6.7, a graph that charts both the federal funds rate and the inflation rate (measured by the CPI) from January 1958 through June 2005.

The most striking feature of the graph is that for the better part of the 1970s the fed funds rate was at or below the inflation rate. In fact the only extended time that the funds rate was nominally higher than the inflation rate was from the beginning of 1972 toward the end of 1974. But this was the period during which the Nixon administration had imposed wage and price controls so the price data reflect only official posted prices, not actual transactions in the marketplace. Nor do they acknowledge nonpriced transaction costs resulting from price controls (like waiting on line for rationed goods). Real prices were actually higher than the posted ones. Nevertheless, as the graph makes clear, the funds rate remained below the inflation rate for the latter half of the 1970s; in fact the funds rate averaged 7% from October 1974 through October 1979, while the inflation rate averaged 7.87%. The effect was to subsidize borrowing to speculate on inventory and commodity price rises, which is precisely what happened.

There is by now little doubt that the Fed's policy stance was extraordinarily harmful, both to the real economy and to financial markets. It brought about the Great Inflation of the 1970s, setting the stage for several recessions including the deep recession of 1981–1982. The Fed fell under considerable pressure to ease monetary policy to secure President Nixon's reelection.[5] Bruce Bartlett of the National Center for Policy Analysis maintains that Arthur Burns willingly subordinated the Fed's policy independence to the goal of securing Nixon's reelection. The only other explanation, he says, is that Burns was an incompetent.[6] Burn's motivation remains an open question. But it is worth noting that Nixon's chief of staff H. R. Haldeman was apparently so pleased with the political pressure on Burns that he was reported to have privately bragged, "We have Arthur Burns by the balls on money supply."[7]

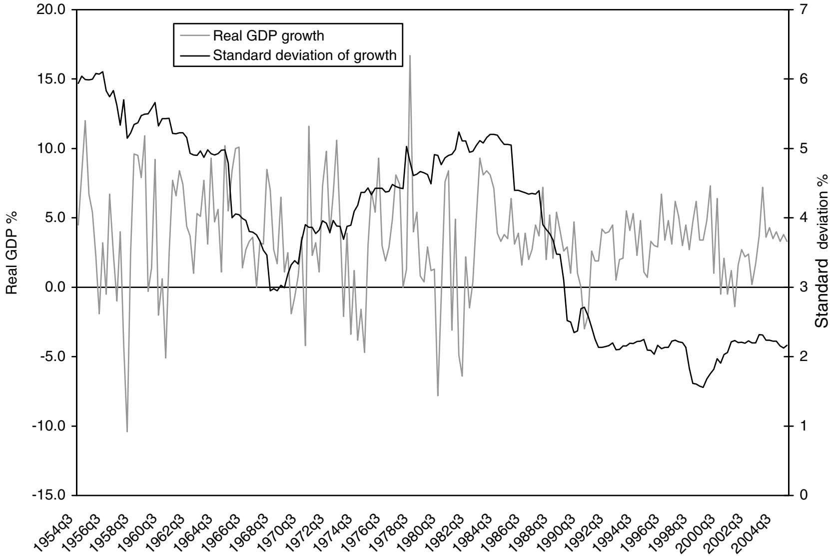
Arthur Burns's motivation notwithstanding, the Fed set off an inflation tidal wave that lasted for years and severely damaged the Fed's reputation for independence. The Fed eventually had to resort to crisis mode to rectify the disaster it created and to restore its credibility. That day came on October 6, 1979. Under the leadership of Chairman Paul Volcker the Fed acknowledged its culpability for the inflation spiral. It then began an aggressive attack on inflation and inflationary expectations by promising to visibly slow the growth of the money supply. Before it was over the daily fed funds rate traded at over 20%. Bond prices plunged, driving yields to over 15%. But the policy gambit worked. The Fed successfully slowed money supply growth and steadily drove down the inflation rate. This was accomplished by keeping the funds rate significantly higher than inflation, making borrowers pay high real rates, and discouraging speculative borrowing. Except for two brief interludes that occurred during the recessions of the early 1990s and 2001–2002, the Fed has maintained this policy stance ever since, setting the stage for 20 plus years of disinflation, and with it, extended bull markets in stocks and bonds.

Another feature differentiates the eras of inflation and disinflation: the performance of the real economy and its volatility. Economic performance can be measured in terms of quarterly sequential growth in inflation-adjusted GDP. Economic volatility can be measured with respect to the variability of that growth. In this case variability is measured by using the standard deviation of the last 30 quarters of real growth. These statistics are displayed as a time series in Figure 6.8.

What is especially striking is how the volatility of economic growth varied across different policy eras. The era from 1966 through 1980 can be

FIGURE 6.8

Growth in the Real GDP and Standard Deviation of Growth (Q2 1953–Q2 2005)



Data source: Bureau of Economic Analysis

characterized as one in which government intervened actively to "manage" the economy. From Q1 1966 through Q4 1980 the average standard deviation of real GDP growth was 4.0%. On the other hand, economic growth was far steadier during the deregulated and low-tax era that began in the early 1980s. The standard deviation of growth from Q1 1981 through Q2 2005 was only 3%, a full 25% less than the earlier era. Moreover, lower volatility beginning in the 1980s did not come at the price of less growth. The average arithmetic rate of growth for both periods was 3.2%

The performance of the capital markets is another matter entirely. Long-term government bond rates fell steadily from peaks that approached 16% in late 1981. By 2005 long bonds yielded about 4.5%. The S&P 500, which had remained essentially flat for 14 years, turned the corner in 1982 to begin one of the longest bull markets in history. Not only did stock prices begin their long rise, market volatility picked up markedly as well. Annualized daily price volatility in the S&P 500, which had averaged about 12% from 1966 through 1980, rose to about 15% from 1981 through the summer of 2005. In other words, reduced economic volatility was accompanied by increased equity market volatility—and rising equity market returns. What was going on?

TWO CHEERS FOR VOLATILITY: FINANCIAL MARKETS AS SHOCK ABSORBERS

One explanation is that liberalization has allowed asset markets to fluctuate freely in response to changing circumstances. This has two important consequences—both of them good. The first is that freely priced financial markets act as buffers that absorb shocks to the system. They can bend without breaking, unlike a strict regulatory regime. The second is that freely priced markets throw off accurate price signals that guide capital allocation toward its highest and best use. In addition, in accordance with the CAPM, we should expect higher returns in the capital markets to accompany increased volatility.

The increase in financial market volatility since the 1980s should be celebrated, not bemoaned. Attempting to bottle up markets distorts price signals, redistributes costs and benefits, misallocates capital, invites stagnation, and puts off the day of reckoning. On the other hand, freely priced asset markets can absorb shocks to the system that otherwise would have been borne by the real economy. Freely traded asset markets allow risk to be spread throughout the system in manageable bits. Not only does this bring enormous benefits to the real economy, but it greatly facilitates the smooth functioning of financial markets. Decentralization of risk significantly reduces the likelihood that any one event will cause catastrophic damage. It allows the system to adjust and adapt quickly to changing circumstances. Conversely, when asset markets are arbitrarily constrained, product markets heavily dependent on external finance careen between artificially induced shortages and surpluses, producing boom-and-bust cycles.

Financial market flexibility, expressed through asset price volatility, simultaneously promotes economic growth through innovation and prudent risk-taking. Robust risk transfer markets make it possible for individuals to pool their money in limited liability companies through which they share risks and rewards as stockholders. And successful companies can raise additional capital to finance innovation and growth through research and development. But financial market volatility gets only two cheers. In the same sense that democracy is a terrible form of government except for all the others, financial market volatility can be thought of as form of systemic flexibility necessary for longer-term innovation and economic growth that rational investors will nevertheless seek to minimize at the firm or household level.

Financial market volatility is the result of the turbulence of the real world and changing expectations about it. The paradox is that by accepting (inevitable) volatility in asset markets, volatility in the real economy is reduced. Moreover as rational investors seek to minimize returns variability by using innovative trading strategies and derivatives, some stand to reap the benefits associated with increased volatility while avoiding the pitfalls. How can this be done?

THE STRUCTURE OF STRATEGY

All trading strategies carry an implicit forecast of what prices are likely to do in the future. Well-constructed trading strategies, like card counting at blackjack, contain several elements designed to improve either the odds of winning (or the payoff) without a commensurate increase in risk. The steps involved are five. The first is data collection. The second is gaining insight through data interpretation. This requires a good deal of circumspection, reflection, placing events in context, and independent thinking. Conventional wisdom will get you nowhere. The third step is forecasting the likely future course of prices based on analytic insight. The fourth step is selecting and executing the strategy that best fits the forecast. The fifth is managing position risk.

The first section of this book discussed some finance theory, the efficient market hypothesis, the CAPM, market models, and the importance of context. The next section will discuss the nuts and bolts of strategy development and execution. It considers the major instruments and trading venues of the capital markets, market institutions, price drivers of various asset classes, strategy types, listed and OTC derivatives, and how to use these instruments to implement trade strategies based on different forecasts.

SUMMARY

The fundamental question is what drives markets. The efficient market hypothesis asserts that markets move as new information comes to light. The CAPM maintains that investors are compensated for risk-bearing. But investors are risk-averting, so they sell off risky assets to buy less risky ones that offer the same return. As the process repeats itself through trading, the market eventually gets to the point where prices properly pair off risk and reward. That point is represented by the efficient frontier.

But several key questions need to be asked. First, how long does it take for markets to price in new information? The conventional wisdom is that the process is relatively fast. And it may be for routine information. But it is likely that the market impact of large-scale events takes a while to unfold. Another question concerns the idea that day-to-day price variation in markets is random. That is problematic in some markets. The evidence strongly suggests that short-term interest rate movements are not random at all. They depend on the Fed, and the Fed exhibits serial-trending behavior. Moreover, this implies that neither the risk-free rate nor the risk premium is random. Nor are economic outcomes random. They are in large measure the result of economic policy decisions, both good and bad.

It would also be a mistake to assume that all that is known is fully embedded in market prices. Strategic behavior may hinder the full expression of information in market prices. Moreover, it is important to distinguish between information, facts, interpretation, and insight. New information may or may not be valuable. It depends on what the new information means and whether it has predictive value. In this regard context is particularly important. How did the market last react to similar circumstances?

The evidence suggests that there are recognizable occasions when market prices are likely to move in a nonrandom manner. There are times when policy changes put a trend in motion or events act as a catalyst. Further, it seems likely that it can take a while for the implications of some occurrences to fully work their way through the system. This suggests that superior insight emanating from rigorous analysis can yield superior returns in the same way that card counting can better the odds at the blackjack table.

NOTES

[1] Perry Mehrling, *Fischer Black and the Revolutionary Idea of Finance*, John Wiley & Sons, 2005.

[2] See http://www.federalreserve.gov/

[3] The acceptance of dollars as a substitute for local currency suggests that a modification of Gresham's Law may be order. In this instance good money is chasing out bad money. Perhaps it is because Gresham's Law only applies when government can use monopoly power to create money and legally mandate its acceptance.

[4] See the Heritage Foundation Index of Economic Freedom at: http://www.heritage.org/research/features/index/index.cfm

[5] Burton A. Abrams, "How Richard Nixon Pressured Arthur Burns: Evidence from the Nixon Tapes," *Journal of Economic Perspectives,* 20/4, Fall 2006, pp. 177–188.

[6] See http://www.nationalreview.com/nrof_bartlett/bartlett200404280812.asp

[7] See Von Mises Institute: http://www.mises.org/fullstory.aspx?control=746&id=74

SECTION II

Instruments, Institutions, and Trading Strategies

This page intentionally left blank

CHAPTER 7

Instruments of the Capital Markets

Before plunging in headfirst to set up trading positions, it is wise to consider the principal characteristics of the instruments to be traded. At the conceptual level, the instruments themselves can be thought of as packets that embed risk and return potentialities. Some instruments relate to capital structure, some to financial technique. Others are mostly characterized by their tax status. For ease of exposition, these instruments are divided into three categories. Financial instruments that constitute the main elements of capital structure are in the first category. The second is options. The third category contains other financial derivatives based on capital market instruments, events, asset classes, and financial market techniques.

Securities, sold as stocks and bonds, are the primary instruments that constitute capital structure for publicly traded firms. Bonds are debt securities; stocks represent equity ownership. Bondholders lend money to the firm. Stockholders own the firm. In general, debt holders have the first claim on a firm's cash flow. Equity holders' claim on cash flow is contingent on debt holders first being paid off.

DEBT INSTRUMENTS

As a rule, debt instruments pay a stated interest rate on a principal amount for a defined period of time. By convention short-term instruments (less than one year to maturity) are referred to as money market instruments. By contrast, longer maturities are capital market instruments. There are significant differences that go beyond terminology. For instance, calculation conventions vary with the type and maturity of the debt instrument.

One of the most important money market instruments is the federal funds rate. Arguably the baseline for all dollar-denominated debt instruments, the funds rate is the interest rate that banks in the United States charge each other for the use of overnight money. It is roughly analogous to LIBOR, the London Interbank Offered Rate. Technically, fed funds transactions are not loans but a matched buy-sell agreement, which allows the market to function without fear of being caught up in legal proceedings over credit precedence.

Federal funds are tradable reserves that commercial banks are required to maintain with the Fed. The size of the reserve requirement varies with the size of a bank's deposit base. Since the Fed does not pay interest on the member banks' reserve deposits, the banks maintain the minimum reserve position possible. Reserves held in excess over requirements are sold overnight (at an interest rate) to other banks short of cash to meet their reserve deposit needs. The interest rate placed on the transaction is the federal funds rate. Typically large city banks are buyers of funds, and smaller regional banks are sellers. Interest on the federal funds rate is calculated on the basis of a 360-day year. A quick rule of thumb is that every 100 basis points racks up $27.77 per day in interest charges per million borrowed. The formula is below.

$$\text{Interest} = \text{Principal} \times \text{Rate} \times \text{Number of Days}/360$$

TREASURY AND FEDERAL AGENCY SECURITIES

Treasury bills are short-term instruments maturing in one year or less and sold at a discount from par. Par is always equal to 100. The Treasury uses the T-bill market to manage its short-term cash needs. Over time the most popular maturities have proved to be three months, six months, and one year. The Treasury also periodically sells cash management bills with very short maturities, generally between a few days and two weeks. By market convention T-bills are quoted at their discount rate. The price calculation for T-bills assumes a 365-day year.

The discount rate is not the same as either the money market yield or the bond equivalent yield. For instance, a one-year T-bill trading at a nominal 10% discount rate would be priced at $0.90. The money market yield would be 90/100 = 11.11%. The various formulas for transforming and annualizing T-bill yields are published in the Federal Register and are embedded as financial functions in Microsoft Excel and other commercial spreadsheet programs.

Treasury notes and bonds are coupon-bearing securities with original maturities longer than one year. Coupons are paid semiannually based on par value. Notes have a maximum maturity of 10 years. Bonds have a minimum maturity of over 10 years. Except for the stated maturity, they are exactly the same. For example a $1 million par value Treasury note with a 4% coupon maturing 8/15/2011 would pay 2% interest or $20,000 every six months (on February 15 and August 15) until the bond matured on August 15, 2011. The last coupon would be paid and the principal amount returned on the final maturity date.

Prices are quoted as a percent of par value. In the Treasury market, fractions of a point have traditionally been denominated in 32nds of a point and fractions of 32nds, with increments as little as 256ths of one point, equal to $39.06 per million of face value. A typical quote displayed on a Treasury bond would look something like this:

$$99.25 - 99.26\ 10 \times 10$$

The quote would signify a bid of 99 and 25 thirty-seconds, an offer of 99 and 26 thirty-seconds, each for $10 million par value. Generally speaking, trading lots are in increments of $1 million. For example, a $1 million par value Treasury bond trading at 95.24 would change hands at $957,500 plus accrued interest. Calculation of accrued interest is on the basis of the actual number of days in the calendar year. (These calculation conventions are built into Excel's financial functions.) American bonds are quoted as "dirty prices," meaning that daily accrued is not built into the price, but added on afterwards. Some other markets have a "clean price" convention in which accrued is built into the net transaction price.

Straight bonds, issued with a defined coupon and maturity, are often called *bullet bonds.* Other bonds have call features that allow the issuer to call them in for early redemption at a predetermined price at certain times before the stated maturity. The Treasury used to sell callable bonds, but it discontinued the practice in 1985.

Treasury inflation protected securities (TIPS) are notes and bonds issued by the Treasury that are linked to the All-Urban Consumer Price Index. Basically the principal amount of the bonds is adjusted for accretion of the nonseasonally adjusted Urban CPI with a three-month lag. Because the principal is adjusted for CPI accretion, interest payments vary as well, since the fixed coupon is multiplied by a changing par value. The CPI is designed to measure consumer inflation, so the idea is that linking the bond's cash flows to the CPI protects investors from inflation. The yield on TIPS therefore approximates the "real" or inflation-adjusted default-free interest rate. Derivatively, the difference between nominal yields and TIPS yields of the same maturity is a measure of expected inflation. Zero coupon bonds are a special case. They are bonds with a stated maturity, but with no coupon. The bond's return comes from the path the bond takes as it rises to par by maturity.

Federal agency or government sponsored enterprise (GSE) bonds are similar to Treasury notes and bonds in many respects. One of the most important differences is that federal agency debt is not backed by the full faith and credit of the United States except for a relatively few instances in which the U.S. government has decided to provide explicit backing. Consequently, agency paper tends to trade at a yield premium to Treasuries all else equal. But agencies also tend to trade at lower yields than comparable corporate bonds, probably because there is an underlying and incorrect belief in the market that the Treasury backs the agencies with an implicit guarantee. Federal government officials have periodically taken great pains to point out that this is simply not the case; there is no federal guarantee, implicit or otherwise. It hasn't seemed to matter very much. Agencies have traded at only very slight yield premiums to Treasuries for years.

Another way in which agencies differ from Treasuries is that for most of these bonds accrued interest is calculated assuming a 30-day month. Treasuries accrue interest on the basis of the actual number of days in the year. Finally, Treasuries are exempt from state and local tax. Some agencies share this feature; others do not.

The biggest issuers of GSE debt are the housing agencies Fannie Mae, Freddy Mac, and the Federal Home Loan Banks. Others issuers include Resolution Funding Corporation (REFCO), the Tennessee Valley Authority (TVA), the Federal Farm Credit System, Financing Corporation (FICO), the Student Loan Marketing Association (Sallie Mae), the Private Export Funding Corp (PEFCO), the U.S. Agency for International Development (US AID), the Financial Assistance Corporation (FAC), the Small Business Corporation (SBC), and the U.S. Postal Service.

MORTGAGE-BACKED BONDS

The structure of mortgage-backed securities is quite different from that of conventional fixed-rate notes and bonds. Mortgage-backed bonds represent pools of mortgages whose cash flows are passed through to The bondholders. Level-pay mortgages have three essential characteristics: The term of the loan is fixed, the interest rate is fixed, and the monthly payment is fixed. The cash flows of mortgage bonds are uncertain because the underlying mortgage can be paid off before the term is due. During a period of sharply falling interest rates mortgagees typically pay off their outstanding mortgages and refinance them at lower interest rates. As a result, mortgage-backed bonds have optionlike characteristics that are similar to callable bonds.

The mortgage finance industry has developed a variety of ways to repackage prepayment and other types of risk normally associated with mortgage-backed securities. For instance, there are collateralized mortgage obligations (CMO) that mix and match different tranches of cash flows. There are mortgage-backed securities whose cash flows derive from only the interest portion, or only the principal portion of the mortgage pool, referred to, respectively, as *interest-only* and *principal-only bonds*. The residuals of some of these repackaging efforts are so difficult to price that the Street refers to them as "toxic waste."

CORPORATE BONDS

Bonds are issued by firms to raise funds either for general corporate purposes or to finance specific investments or other transactions. The famous Modigliani-Miller theorem posits that, in a world of zero transaction costs and no taxes, the economic value of the firm is unaffected by its mix of debt and equity. In other words, its capital structure has no economic relevance. However, taxes and transaction costs do exist, and they have an impact on financial decision making. For example, U.S. corporations have issued tremendous quantities of debt because the tax code (1) reduces the relative cost of debt capital by allowing interest payments to be expensed and (2) raises the cost of equity capital by taxing it both at the corporate and individual levels. Recent shifts in the tax code have mitigated this somewhat.

The credit quality of corporate bonds is evaluated by three officially recognized credit-rating agencies. They are Moody's Investors Services, Standard & Poor's, and Fitch Ratings. The rating agencies score the credit quality of bonds and assign ratings to them based on elaborate econometric models. The ratings are expressed in terms of letters, with the AAA being the best S&P rating and Aaa being the best rating from Moody's. An S&P rating of AA is slightly worse than AAA and so on along a descending scale to D, which signifies that the issuer is in default. Agency ratings are almost always very close to each other; often, if not usually, they are identical.

Many traders argue that agency ratings follow rather than lead the market. In part that may be because the major investment banking firms that act as market makers have their own in-house research departments that analyze credit quality. And it is true that the market does seem to anticipate credit upgrades and downgrades. Partly as a result, Moody's launched a service called market implied ratings (MIR), which analyzes bond prices with respect to credit quality ratings. It is also worth noting that ratings depend on accurate and timely information. The implosions of Orange County California and Enron that caught the markets flat-footed were ultimately the result of criminal behavior that included accounting fraud.

Corporate bonds trade at a yield premium to Treasuries, with the best quality credits having the tightest spreads to Treasuries. As credit quality deteriorates, yield spreads widen. Bonds rated BBB or higher by S&P, or Baa by Moody's, are considered to be investment grade. Lower-rated bonds are thought to be speculative or "junk bonds" in the parlance of the Street.

There are actually two types of junk bonds. The first is made up of "fallen angels" which are bonds that were originally issued as investment grade credits, but whose credit quality (and rating) deteriorated in the interim to below investment grade. The second type is made up of bonds originally issued as junk. It would not be unusual for these to be bonds used to finance highly levered transactions, like a leveraged buyout (LBO) or hostile takeover. Or they may just be issued by a low-quality credit. As an aside, as new issuance of junk bonds picked up momentum, the terminology was upgraded as well, so to speak. Junk bonds are now referred to as *high-yield bonds.* It makes them easier to sell.

MUNICIPAL BONDS

Municipal bonds share certain characteristics of both U.S. government and corporate bonds, but they are different in one very important aspect. By and large municipal bond interest is exempt from federal taxes. Some are triple tax exempt. That generally occurs when they are held in the state of the issuing jurisdiction. Because of the power of triple tax exemption, most municipal bonds are bought by investors who reside in the issuing jurisdiction. Municipals can be bought either directly by purchasing an individual bond or by being purchased indirectly through a mutual fund that invests only in the tax-exempt bonds of a single state and its political subdivisions. Some national municipal funds opt for diversification by purchasing bonds across many jurisdictions.

Tax-exempt securities are generally issued in one of two formats. The first type is that of a general obligation, or GO. In theory general obligations are legally backed by the full taxing power of the issuer. That's the theory. The reality is a bit different though. Orange County, California, spectacularly defaulted in the mid-1990s and refused to raise taxes or cut spending to make good on its defaulted bonds. It succeeded in negotiating a refinancing.

The other type of tax-exempt bond is a revenue bond. These are bonds whose principal and interest are covered by revenues raised by the project they finance. Toll roads, state educational institutions, tunnels, bridges, and economic development projects are typical examples. Default rates for revenue bonds vary widely by project type. The major ratings agencies evaluate municipal bonds with scales similar to the ones they use for corporate bonds.

The power of the triple tax exemption (or alternatively the horror of the tax code if you prefer) is such that municipal bonds are typically grossed up to their taxable equivalent yields in order to showcase their relative attractiveness. That is because nominal yields of municipal bonds are typically lower than comparable maturity Treasuries. But after adjusting for tax, municipals yield more because the credit quality of municipals is not as good. (In fact the great majority of municipal issuers buy credit insurance to backstop their credit.)

For instance, suppose a 10-year Treasury yields 5% and an AAA 10-year municipal bond yields 4.85%. On a comparable after-tax basis the municipals yield 7.46%, assuming a marginal tax rate of 35%. The gross up is accomplished by dividing the tax-exempt yield to maturity by 1 minus the marginal tax rate as shown below.

$$Yield_G = \frac{YTM}{1 - M_{tax}} = \frac{4.85\%}{1 - 0.35} = 7.46\%$$

Not all municipal bonds are tax-exempt. Sometimes otherwise tax-exempt institutions offer taxable securities to the market. In addition, some are subject to the alternative minimum tax.

EQUITY SECURITIES

Common stock represents ownership in a company. It is sold in shares, with each share representing a portion of the company's ownership. Common stockholders taken together as a whole own the company. They receive dividend payouts in proportion to the number of shares they own. Stockholders generally have a right to vote their shares in matters affecting the company, with the election of the members to the board of directors being among the most important. Usually one share represents one vote. Management reports to the company's board of directors, which is charged with representing the interests of the shareholders. Since the shareholders are the owners of the company and posses voting rights, the stock market is sometimes referred to as the "market for corporate control."

Common stockholders are often said to be the residual owners of the firm because their claim to the firm's earnings and assets is contingent on bondholders' and debtors' claims being satisfied first. Common stockholders are last on line if the firm is liquidated. Lenders to the firm are first, with the pecking order among lenders depending on prior legal agreements and bond covenants. Lenders' interests in the firm revolve around the timely repayment of principal and interest. On the other hand, common stockholders share all the business risk, up to 100% of their investment, but no more. In this, a common stockholder has a limited liability investment, unlike a general partner in a firm. A partner's liabilities can extend beyond his capital commitment.

Common stockholders and bondholders have different and sometimes competing interests in the firm. For instance, increasing the leverage of a firm's capital structure by issuance of more debt can increase a firm's earnings per share and return on equity thus boosting its stock price. But it also makes existing bond holdings riskier, without the prospect of greater potential returns. At the extreme, in a leveraged buyout, existing stockholders can receive a substantial premium over market value for their stock. But the value of outstanding bonds can drop precipitously if large quantities of debt are sold to finance the transaction. In this respect it should also be noted that the board of directors' responsibility is to look after the interests of the stockholders, not the bondholders. They are on their own.

Publicly held companies are mostly listed and traded on exchanges. The New York Stock Exchange is the premier stock exchange, followed by NASDAQ. Other trading venues include the American Stock Exchange, the Chicago Stock Exchange, and other smaller regional exchanges. A "round lot" for trading stocks is generally 100 shares. Bid/Ask stock quotes are for lots of 100 shares in one-cent minimum increments. For instance, a quoted market on Procter and Gamble (ticker PG) of 61.05 – 61.06 1,000 × 1,000 would represent a bid to buy 1,000 shares at $61.05 (or $61,050) versus an offer to sell 1,000 shares of PG at $61.06 or $61,06. The quote can be accessed by member firms in increments of 100 shares. Odd lots are easily traded as well.

HYBRID SECURITIES

Some securities have features of both equity and debt. They are known as *hybrids*. Prominent among them are debtlike securities that look like, or can be converted into, common stock. Among the most popular are preferred stock, convertible bonds, and convertible preferred stock.

Preferred stocks typically, though not always, pay an unchanging quarterly dividend for as long as they are outstanding. Sometimes they pay cumulative dividends. Some are listed on the major exchanges; some trade over the counter. They can be quoted either in dollar terms or as a percent of par. Companies often sell preferred stock as an alternative to selling straight debt. Moreover, preferred stock does not count as equity so its sale does not dilute reported earnings per share (EPS).

Since preferred stocks pay a quarterly fixed dividend much the way a corporate bond pays a semiannual fixed coupon, they tend to act like corporate bonds. But yields on preferred stocks are typically lower than they are on corporate bonds issued by the same company. That is because of the U.S. tax code. Corporations are not required to pay federal income taxes on 70% of the preferred stock dividends they receive. Consequently, corporation buying pushes preferred stock yields down lower than they would otherwise be. In addition, preferred stocks are easily accessible by individuals since many are listed on exchanges and trade in small-sized units, with the par value of many issues being $25 to $50 per share.

Preferred stocks are like bonds in the way they react to inflation and changes in interest rates. A disadvantage that afflicts preferred stocks is that preferred stockholders do not have the same privileged position as bondholders in the capital structure. Unlike coupons on bonds, preferred stock dividends can be suspended or passed without driving the firm into default. The result is that preferred stocks are often riskier than corporate bonds, but have lower yields because of their unique tax exemption. Consequently, they should generally be avoided by investors who are not in a position to reap the benefit of the special tax treatment preferred stocks are accorded.

Convertible securities are corporate bonds or preferred stock that can be converted into shares of the common stock of the issuing company. As a result, convertibles are sometimes thought of as deferred equity. By far, most issuance is in the form of convertible bonds rather than preferred stock, but valuation methods are very similar.

Like a conventional bond, convertibles are issued with a coupon and maturity, but they also have a conversion price at which the bond can be tendered for shares of common stock. Normally, a convertible bond has a lower yield than a conventional bond because the convert shares the upside potential of the stock price. For instance, take a hypothetical convertible bond with a 4% coupon, maturing in 10 years, priced at 100 to yield 4%, callable at 110 and convertible into common stock at $50 per share. Assume the common is trading at $40.

The conversion price implies the bond's conversion ratio. A $1,000 face value convertible bond with a conversion price of $50 per share creates 20 potential shares of stock, since 1,000 / 50 = 20. The conversion option can be stated either way—in terms of price or in terms of shares of stock. In the current example, as long as the stock is trading below $50, the bondholder will hang onto his bonds. But if the stock begins to trade above $50, say at $75, the bondholder can exercise the conversion option profitably. The $1,000 bond is worth $1,500, which is the product of 20 shares × $75, the new price of the common.

Convertible bonds usually allow the holder to convert at any time. Also, the bonds are usually callable. If and when the common stock trades at a substantial premium to the conversion price, the issuing company will often move to force conversion. They do this by calling the bonds at the price listed in the bond's indenture, which is typically lower than the conversion price, in effect forcing the convert holders to turn in their bonds in exchange for common stock. This allows the company to raise cheap equity capital.

In the example above, with the common trading at $75, a $1,000 face value bond would be worth $1,500; but it would be callable for only $1,100. Owners of the bonds would therefore covert them into stock worth $1,500 rather than accept $1,100 cash for redeeming them at the call price.

Convertible issues (both preferred stocks and bonds) have the attractive feature that they offer a way to trade off risk and return between straight debt and equity investments. But they do not represent a free lunch. In effect a conversion premium is built into them. Using the original example, $1,000 invested in the convertible bond could have been put to work directly in a straight bond that would have earned a higher yield to maturity; or it could have been invested in 25 shares of common stock instead of the 20 implied by the conversion price.

OPTIONS

Options are contracts that give the owner the right to buy (or sell) a stock, bond, commodity, currency, or any other asset at a fixed price by a set time. Call options give the right to buy; puts the right to sell. The price at which the option holder can exercise his or her option is the strike price. The expiration date is the last day the option is exercisable.

Equity options trade both in the OTC markets and on the major exchanges. OTC options can be tailored to meet specific strike price and expiration demands, but are consequently less liquid. Exchange-traded options typically have standardized expiration dates and strike prices, are transparently priced, and enjoy high levels of liquidity. The Chicago Board Options Exchange (CBOE), the International Securities Exchange (ISE), the American Stock Exchange, the New York Stock Exchange, and the Philadelphia Stock Exchange trade options on stocks. The Chicago Board of Trade and the Chicago Mercantile Exchange trade interest rate, commodity, and equity index options.

The discussion that follows is placed in the framework of stock options, but that is simply for expository convenience. Option trading strategies designed for stocks are, in general, easily applicable to bonds, commodities, indexes, and foreign exchange with relatively minor modifications, depending on the underlying instruments, market practices, and institutional arrangements. The basic drivers of option pricing remain the same regardless of the underlying instrument.

An initial sale is referred to as *writing an option.* Writers who own the underlying stock are *covered writers;* sales without ownership of the underlying stock are *naked (as in uncovered) sales.* Conversely, put writers who are short stock are covered; put writers without a position in the underlying stock are naked put writers. Initial positions are opening positions; unwinding or liquidating an existing position is to close it out. Call options with strike prices below the current stock price are said to be *in-the-money.* If a call strike price is above the stock price, the option is *out-of-the-money.* Options with strikes exactly at the stock price are *at-the-money.* The same is true (in reverse) for puts.

Stock options, which typically expire on the third Friday of the month, are typically quoted in terms of 100 shares of stock. For instance a September 67.5 call option on Apple Computer (AAPL) stock would give the owner the right to buy 100 shares of Apple Computer common at $67.50 per share up until the September 16 expiration. The price of an option is called its *premium,* in this case $2.95. Since options are quoted in terms of 100 shares of stock, the $2.95 premium actually signifies $295 per options contract.

If, at expiration, Apple stock is trading below $67.50 per share, the option would expire worthless. Anyone who wanted to buy Apple stock could do so more cheaply in the marketplace; there would be no reason to exercise an option to buy at $67.50. On the other hand, if the stock is trading over the strike price—say at $70 per share—the option would be worth at least the difference between the market price and the option strike price. The spread between a stock's market price and the option strike price, in this case $2.50, is called the option's *intrinsic value.* The portion of an option premium that is in excess of its intrinsic value represents *time value.*

American-style options can be exercised at any time up until expiration; European options can be exercised only at expiration. American-style options will not trade below intrinsic value because arbitrage will keep a floor at intrinsic value. Suppose that AAPL is trading at $70 and that the AAPL $67.50 call option due to expire in 30 days is offered at $2.00. An arbitrageur would buy the call at $2.00, sell the stock short at $70.00 and then exercise the option. Assuming no transaction costs, the result would be a riskless profit of $0.50. The cost of the stock would be the exercise price plus the premium or $2.00 + $67.50 = $69.50. The sale would be at $70, locking in an arbitrage profit of $0.50.

Since European options are only exercisable at expiration, the arbitrage is a little trickier. The arbitrageurs need to make sure they can finance the transaction through to the option expiration date. In this instance, they would have to make sure that they could borrow the stock to deliver against the short sale until they could exercise their call option.

Puts work essentially the same way, but in reverse. A put gives the holder a right to sell, so the intrinsic value of a put is the difference between the stock price and the exercise price when the stock price is lower than the option exercise price. Any additional premium is time value. For dividend-paying stocks, announced dividends have to be included in the calculation of an option's value. Since owners of stock on the day of record are entitled to any declared dividends, puts on the stock have to discount the value of those and any other expected dividends up until expiration.

Put and call valuations are not independent of each other; they are linked by arbitrage. The price of a call can be said to determine the price of its companion put and vice versa. Suppose AAPL is trading at $70 per share and AAPL September $67.50 calls are trading at $3.00. Of that, $2.50 represents intrinsic value and $0.50 is time value. The corresponding AAPL September $67.50 puts necessarily trade at $0.50. The reason is that any other price would produce a riskless arbitrage profit, as always assuming no transaction costs. To see this, consider an example. Suppose the AAPL September $67.50 puts were trading at $1.25 instead of $0.50. What would happen?

Arbitrageurs would execute a three-part transaction. First, they would sell AAPL stock short at $70. Second, they would buy AAPL $67.50 calls at $3.00. Third, they would write AAPL September $67.50 puts at $1.25. If by expiration the stock goes up, the arbitrageur exercises the call and loses the call's time value of $3.00 – $2.50 = $0.50. But, the arbitrageur makes $1.25 in time value on the put he or she wrote. Net, the arbitrageur makes $1.25 – $0.50 = $0.75 per share.

On the other hand, if the stock goes down, the arbitrageur makes a profit on the short sale, until the stock gets to $67.50, at which point someone will "put" the stock against the $67.5 put he wrote. At a price of $67.50 he makes a profit of $2.50 on the AAPL short sale, loses $3.00 on the calls he bought, and makes a profit of $1.25 on the puts he sold short, for a net profit of $0.75 per share. The only price for the puts that avoids a free-risk arbitrage is $0.50—which is where they will trade. This triangular relationship among the prices of the stock and the companion puts and calls (i.e., the same expirations and strikes) keeps prices in line.

OPTION PRICE DRIVERS

Once we know the price of a call, it is easy to calculate the price of the companion put and vice versa. What determines the initial price level of the put or call to begin with? The answer to that question comes from the famous Black-Scholes option pricing model.

The highly sophisticated mathematics of Black-Scholes is beyond the scope of this book. However, many easy-to-use applications of the model are readily available, including an options toolbox available for free at the CBOE Web site: www.cboe.com. Instead of delving into the mathematics, this chapter will instead present some fundamental concepts behind Black-Scholes to help develop an intuitive grasp of how options are priced. Armed with those ideas and application toolboxes, strategies for using options can be built.

Basically, there are three variables driving the Black-Scholes option pricing model: time to expiration, short-term interest rates, and volatility. Volatility is in some respects both the most important and the most difficult to get a handle on. Volatility can be thought of as an expected range of price variation for a specified period of time. With financial assets, it is generally measured with respect to the standard deviation of price returns. (There will be more about this in later chapters). For now, suffice it to say that the more a stock moves on a day-to-day basis, the more volatile it is considered to be.

As the volatility of a stock increases, so does the premium for options traded against it. All else equal, a stock that on average moves up or down 1% per day will have its options priced higher than a stock that on average only moves ½ percent per day. The reason has to do with the nature of an option's payoff structure. A call option holder can only lose 100% of the premium, but the upside potential is far greater—theoretically infinite. If a high-volatility stock and a low-volatility stock each had options traded against them at the same price, they would each bear the same risk, but the high-volatility stock would offer the possibility of greater reward. We would therefore expect arbitrageurs to buy options on the high-volatility stock and sell options on the low-volatility stock until the expected risk-reward payoff structure for the two became roughly comparable. Options on the high-volatility stock would eventually be expected to rise relative to options on the low-volatility stock.

Time to expiration is similarly an important variable. The longer an option has to go before it expires, the greater the chance an option holder has to profit from an advantageous development in the underlying stock. Conversely, if nothing happens, the time premium in the price of the option will dissipate over the life of the option until it expires. The inevitable erosion of time value makes options "wasting assets" in the sense that they get less, rather than more, valuable with the passage of time.

Short-term interest rates are less important (although related) to time and volatility. All else equal, the higher short-term rates are the more valuable an option is because an option requires less of a cash investment to get the same bang for the buck. Since it requires a reduced cash outlay, the excess cash earns interest in the bank. The higher rates are the more excess cash earns. The more cash can earn, the more willing an investor should be to buy options (with a lower cash outlay) than the underlying stock. Therefore higher short-term rates drive options prices higher, all else equal.

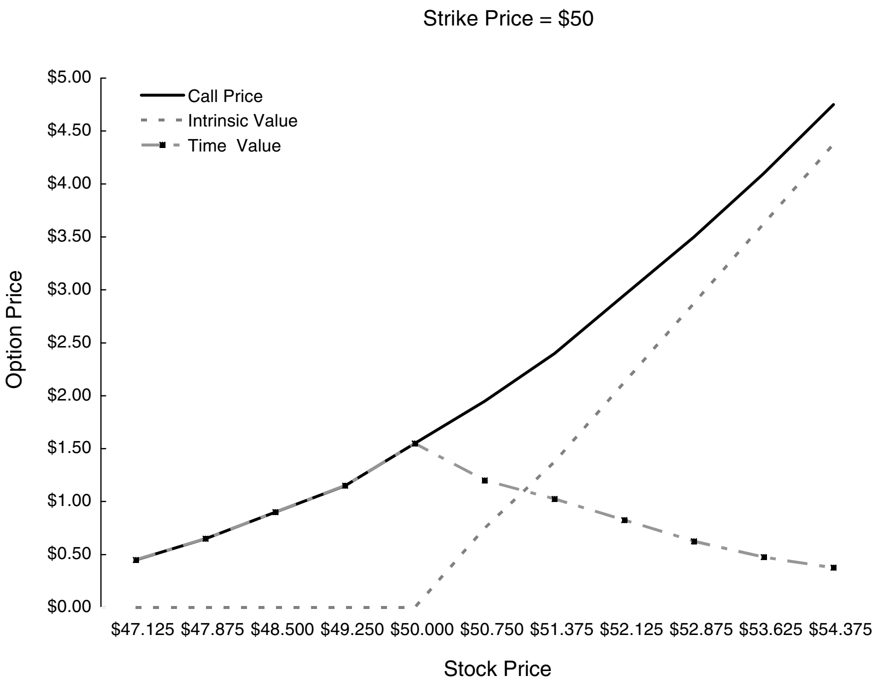
STATIC VERSUS DYNAMIC PRICING

One of the most important characteristics of option pricing is that the price relationship between an option and its underlying stock is dynamic, not static. Consider a 30-day call option on a non-dividend-paying stock trading at $52 a share with a strike price of $50. Assuming annualized volatility of about 25% per year and short-term rates of about 5%, most option pricing models would predict that the call option would trade at around $2.875, which would represent $2.00 in intrinsic value and 0.875 in time value. As the stock moves up and down, the option follows, although not at the same rate. The portion of the option's price that represents intrinsic value will increase, and the portion that represents time value will decrease. The opposite will occur on the downside. The call option price will follow the stock price lower, but the option will pick up time value as it sheds intrinsic value.

This can be shown using a Black-Scholes model to estimate the price of the option as the price of the stock moves up and down, while holding the volatility and interest rate variables constant. The result of doing this type of exercise is on display in Figure 7.1, a graph of the

FIGURE 7.1

Call Option Pricing



sensitivity of the option price and time value with respect to the stock price. It shows first, how the option rises at an increasing rate and falls at a decreasing rate as the underlying stock rises and falls, all else equal. Secondly, it illustrates the change in time value. Time value is represented by the descending dashed gray line, equal to the value of the gap between the dark solid black line (the call price) and the gray dotted line (the call's intrinsic value).

THE GREEKS

Quants have modeled options extensively and assigned Greek letters to represent the different dynamic variables that affect options pricing. Foremost among them is the Greek letter delta (Δ), which represents a change in an option's price with respect to the change in the underlying stock price.

$$\Delta = \text{Option Price Change/Stock Price Change}$$

Rearranging the terms:

$$\Delta^* \text{ Stock Price Change} = \text{Option Price Change}$$

Expressing the relationship this way allows delta to be used to evaluate and manage risk. The baseline is a delta-neutral position, in which the quantity of stock in position (either long or short) is offset by the notional quantity of offsetting options in position. Delta is the quantity that equalizes the two sides of the equation.

For example, take a 30-day call option with a strike price of $50 trading at $1.55. Say the underlying stock is trading at 50, so the option is at-the-money. Assume for the time being, a delta of 0.50. Going by the formula, a delta of 0.50 implies that the price of the option is expected to vary by half as much as the price of the underlying stock. If the stock trades up 2.875 to $52.875, the option would trade up roughly half that amount, or $1.45, bringing it to a price of $1.55 + $1.45 = $3.00. Note that in percentage terms, the stock rises 5.75%, but the call option rises 93.5%.

A hedger could neutralize a 1,000-share position in the stock by purchasing a delta-neutral quantity of options on the opposite side of the market. Since the stock price is expected to move twice as fast as the option price, it takes two times as many shares of stock controlled through options to reach the point of indifference, or neutrality. Since standard exchange-traded options are for 100 shares, 20 calls would equal 2,000 notional shares. A hedger short 1,000 shares of stock and long 20 call options would be delta-neutral since changes in the stock position would be offset by changes in the options. The stock short sale of 1,000 shares would result in a loss of $2,875 on price rise of 2.875 points. The long options position would offset the loss with a gain of 2,000 × $1.45 = $2,900.

The example assumes that the delta is static, but in reality it is not. Delta is actually dynamic. The relationship between the option and its underlying stock changes with the price of the stock. The delta value rises as the stock rises and falls as the stock falls. The relationship is curvilinear rather than linear, which has important implications for trading and risk management. It creates a world of dynamic hedging in which positions are constantly adjusted. To see why, consider why delta has to be 0.5 when an option is at-the-money, and then consider the implications.

Why is the delta 0.50 when the stock is at the money? First we start with the assumption that short-term stock price movements are randomly distributed. If so, the probability of the next trade being an uptick or downtick is the same. Second we assume that the magnitude of the price change is the same for both the upside and the downside. Consequently, we have a fair bet, just like a coin toss. There is a 50:50 chance for each outcome, and the payoff is the same regardless of the outcome. For the option to be priced correctly, the same conditions must hold. There must be an equal likelihood for the direction and magnitude of the next trade. The probability of winning has to equal the probability of losing for the same payoff. It follows that the payoff must be the same if the win/lose odds are the same. The only delta for which those two conditions hold is 0.50 when the stock is at-the-money.

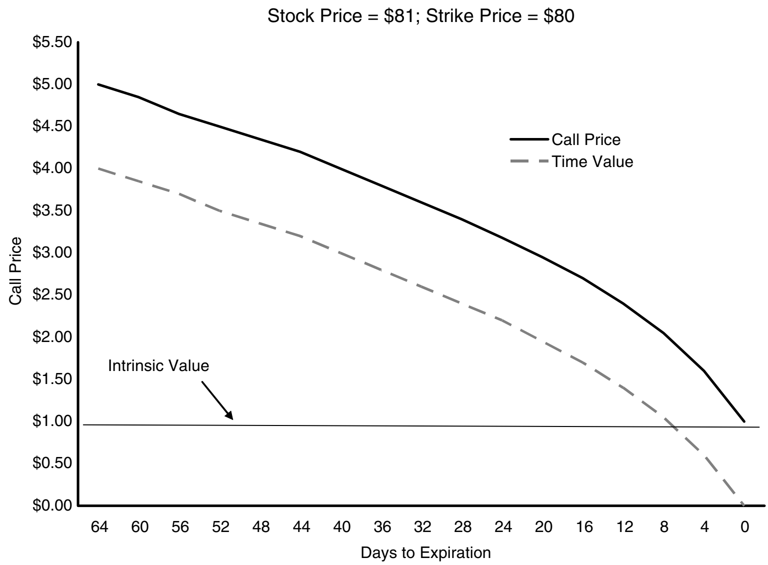
What happens when the stock is not exactly at-the-money? The deltas of the respective puts and calls with the same strikes and expirations always have to sum to 1. Any increase in the delta of a call option is offset by a decrease in the companion put, and vice versa. Since the stock can only go up, down, or remain unchanged, the sum of the put and call deltas exhausts all the possibilities. So the respective deltas of the put and call options vary with the stock price, but the sum of the probabilities does not. It still equals 1.

When a stock is in-the-money, the price of a call option will be higher than when it is out-of-the-money, so there is more money at risk. But the chance of a one-unit uptick in the stock price remains the same as the chance of a one-unit downtick. That implies deterioration in the risk/reward ratio if the payoff for the option position remains the same. But the payoff does not remain the same; it increases. The option will move up with the stock at an increasing rate of speed, increasing the payoff.

This opens the door to dynamic price hedging. As the price of the stock changes, the delta value of the options changes as well, and in the same direction. From an arbitrage standpoint, keeping the risk posture constant requires constant position adjustments in option positions as the

FIGURE 7.2

Option Time-Decay Graph



underlying stock moves up or down. For instance, a trader who is long a delta-neutral quantity of calls and short stock would maintain a market risk-neutral position by selling out his calls on the upside, while buying more on downdrafts. Traders in this position are, in effect, long volatility, hoping to profit by position adjustments; they are selling as the market trades up and buying when it goes down.

But if volatility falls off, so does option premium. Time decay is an issue as well. Options are wasting assets; as options approach expiration, they converge toward intrinsic value, as displayed in Figure 7.2.

In addition to delta, four other option parameters described below often capture the attention of options traders. The first is gamma, Γ, which is defined as the change in delta with respect to price changes in the underlying stock, in other words the second derivative of delta with respect to the underlying. In general, gamma doesn't change much when an option is either far out-of-the-money or deep in-the-money. Gamma is most sensitive—delta is most volatile—when options are close-to-the-money. Some traders like to superimpose gamma hedges on top of delta hedges. But as a practical matter, hedging gamma causes (the generally more important) delta to change, which in turn has to be readjusted in a process of infinite regress.

As it turns out, the next Greek letter, vega (for volatility) is not a Greek letter at all. Sometimes the letters epsilon (ε), eta (η), kappa (κ) or lamda (λ) are used as substitutes, in what most graduate finance students are likely to consider to be a vast conspiracy designed to make options theory as confusing as possible. Nevertheless, Fischer Black, Myron Scholes, and Robert Merton used the term in their original option pricing model to denote the change in an option's value with respect to the change in volatility. It tends to be at its highest when an option is close-to-the-money, and lowest when far out-of-the-money. Volatility, one of the most difficult parameters to predict, is discussed in more detail in later chapters.

Theta (Θ) is a measure of how an option's value changes with its time to expiration. The further an option is out-of-the-money, the less important theta is. Conversely, theta is highest when time remaining is very short and the option is close-to-the-money. Finally, rho (ρ) measures the effect of changes in an option's value with respect to short-term interest rates. Rho is more important for longer-term expirations and for deep in-the-money options.

OPTION STRATEGIES

Options are very flexible instruments that can be used in an extraordinary variety of imaginative ways. Some of the more common uses are briefly reviewed here. At a very basic level options are frequently used as vehicle to speculate on short-term market direction. Bullish traders simply buy calls, and bearish ones may buy puts. Beyond simple long-short directional trades, a wide range of available strike and expiration dates makes possible a dazzling array of strategies.

Covered call and put writing programs can be structured either to be neutral or to tilt toward either side of the market. Vertical option spreads set up long-short trades within the same expiration periods; horizontal spreads set up long-short trades at the same strike price but over different time frames. Diagonal option spreads are trades in which both the strike prices and expiration dates vary. Options can also be used to set caps, collars, and floors on positions. Some examples of these strategies follow.

COVERED WRITES

A covered call writer owns the underlying stock. A 1:1 call writer sells 1 option for 100 shares of stock. He collects premium income, but at the price of limiting the upside. Risk exposure is measured in delta-neutral terms. A 1:1 writer of at-the-money calls can be said to have reduced his initial risk exposure by half. There is still exposure on the downside, but it is partly offset by the premium income received from writing the calls. On the other hand, ratio writers sell options on more notional shares of stock than they own. That further limits exposure on the market's downside, but at the price of additional exposure on the upside, possibly turning the writer into an inadvertent short seller.

For example, consider a pair of hypothetical covered call positions (buy writes) at option expiration date. The first is a straightforward 1:1 covered write on 1,000 shares of a $50 stock with a $50 strike price. The stock is initially purchased at $50, and a total of 10 calls representing 1,000 shares of stock are sold at a premium of $5 per share, producing $5,000 in premium income. The second position is a 2:1 ratio write using the same prices. To simplify, we assume interest rates are 0% and transaction costs are zero as well.

The 1:1 writer has his position protected 5 points down, to $45 a share, but with earnings capped out at $5,000, because the stock will get called way if it is trading above $50. The ratio writer has 10 points protection on the downside because he wrote twice as many calls and therefore collected $10,000 in premium income. However, once the stock goes above $50, the ratio writer is in short-sale territory. The writer owns 1,000 shares of stock, but owes delivery on 2,000; he is short 1,000 shares and is exposed on the upside. The 1:1 writer can be thought of as having a bet with a bullish directional bias. As long as the market stays at $45 or above, he at least breaks even. Profit is maximized when the stock is $50 or higher.

The 2:1 ratio writer has a more complicated bet, with profit potential double that of the 1:1 covered writer. Profit is maximized at $50, but the boundaries are on two sides. At expiration, as long as the market is above $40 and below $60, the writer is at least breakeven on the position. But while the ratio writer has more protection on the downside, he has a lot of potential exposure on the upside. An aggressive takeover bid for instance could leave him with a big problem on his hands. See Figure 7.3.

VERTICAL AND HORIZONTAL SPREADS

Option spreads can similarly be used to establish bullish positions, bearish positions, or volatility bets. As in covered writing, the ratio of longs to shorts is the predominant factor in the profit and loss (P&L). To explore transactions of this type, consider closing market prices for September and December 2006 puts and calls on Google (GOOG), displayed in Table 7.1.

The data in the tables are closing bids and offers for Google puts and calls at the close of business August 18, 2006. On that date, Google's closing price was $383.36. For the sake of convenience, we make the extraordinarily optimistic and wholly unrealistic assumption that all transactions are executed at the bid side of the market. Note though that the bid/offer spread is worth taking a look at because options trading is transaction intensive and commissions and fees pile up quickly. For example in

FIGURE 7.3

Covered Call Writing

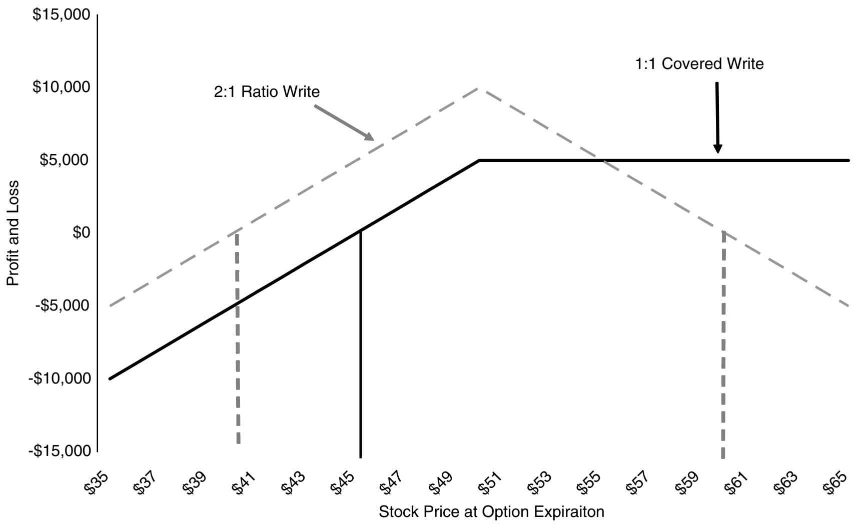


Table 7.1, the average bid/offer spread for call quotes for September is 2.97% and that doesn't count brokerage commissions.

Leaving transaction costs aside, we will use prices from the table to illustrate vertical, horizontal, and diagonal spreads, straddles, and strangles. First, a bullish vertical spread: Long Google September 380 calls at 13.6, short Google 390 September calls at 8.3, with the stock trading at $383.36. The trade has a bullish bias. The market has to rise for the trade to be profitable because more cash is invested on the long side (13.6 points) than on the short side (8.3 points). The maximum exposure is the difference of 5.3 points (multiplied by the size of the trade). Changes in volatility or interest rates will not materially affect the trade, only changes in the stock price. Breakeven is a rise in price of Google stock to $385.30. That is the minimum price Google has to rise to by expiration so that net invested cash of $5.30 can be extracted. Maximum profit potential at expiration is $4,700, which is achieved if the stock closes at $390 or higher, as displayed in Figure 7.4.

As with ratio writes, vertical spreads can be tilted toward either the bullish or bearish side of the market depending on the ratio of longs to shorts. The decision has to do with the extent to which leverage is desired and what the acceptable caps and collars are on the position. Positions with high embedded leverage would be ones in which either at-the-money or deep-in-the-money options are hedged with far-out-of-the-money puts or calls. In effect these are volatility bets. For example, consider a vertical

TABLE 7.1

Vertical and Horizontal Spreads for Google (September and December 2006)

Google closing price August 18, 2006 = $383.36

Google September 16, 2006 Options

	Calls				**Puts**		**Market Spread (Calls)**
Symbol	Bid	Ask	Strike	Bid	Ask	Symbol	
GGDIJ	36.8	37.2	350	1.85	2	GGDUJ	1.09%
GGDIL	28	28.5	360	3	3.3	GGDUL	1.79%
GGDIN	20.2	20.5	370	5.2	5.3	GGDUN	1.49%
GOPIP	**13.6**	**13.7**	**380**	**8.4**	**8.6**	**GOPUP**	0.74%
GOPIR	8.3	8.6	390	13.3	13.5	GOPUR	3.61%
GOPIT	4.7	4.9	400	19.7	20	GOPUT	4.26%
GOPIB	2.55	2.75	410	27.6	28.1	GOPUB	7.84%
							2.97%

Google December 16, 2006 Options

	Calls				**Puts**		**Market Spread (Calls)**
Symbol	Bid	Ask	Strike	Bid	Ask	Symbol	
GGDLJ	51	51.3	350	11.5	11.8	GGDXJ	0.59%
GGDLL	44.1	44.4	360	14.3	14.7	GGDXL	0.68%
GGDLN	37.6	38	370	17.9	18.2	GGDXN	1.06%
GOPLP	**31.9**	**32.1**	**380**	**22**	**22.3**	**GOPXP**	0.63%
GOPLR	26.6	27	390	26.8	27.1	GOPXR	1.50%
GOPLT	22	22.3	400	32.2	32.5	GOPXT	1.36%
GOPLB	18	18.3	410	38.3	38.6	GOPXB	1.67%
							1.07%

Data Source: E*Trade

call spread at a ratio of 4:1. The long side is four Google September 400 calls trading at $4.70; the short side is one Google September 370 call trading at $20.20. As before, the stock is trading at 383.36.

Figure 7.5 plots a theoretical P&L based on the price of Google stock closing between ≥$350 and ≤$450 on expiration day. As the graph illustrates, the transaction is marginally profitable on the downside, loses money in the middle, and reaps substantial gains in the event of a big

FIGURE 7.4

Bullish Vertical Spread

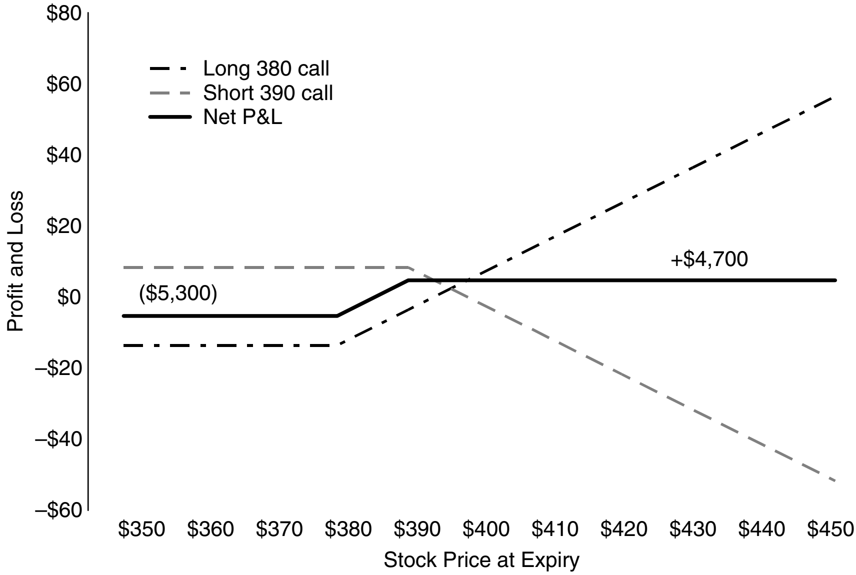
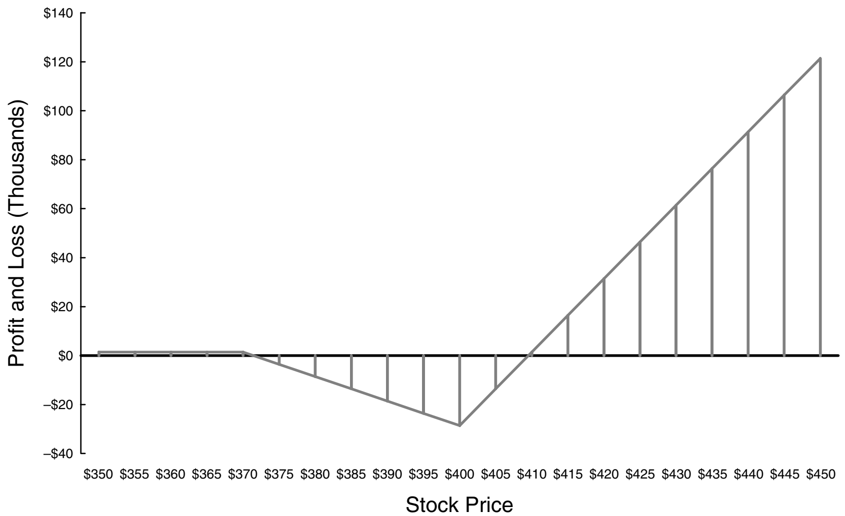


FIGURE 7.5

Google 4:1 Vertical Spread



move to the upside. It is a highly levered bullish position that anticipates increased volatility. Taking the other side of the transaction would, of course, produce a mirror image graph. More importantly, the transaction illustrates the way combinations of options can be used to target price zones. In this particular instance, if Google closes between $375 and $410 at option expiry, the trade loses money (using round numbers). Above $410 the trade becomes profitable and picks up steam to the upside.

In contrast to vertical spreads, horizontal spreads are bets largely driven by time value. A horizontal spread consists of offsetting long and short positions at the same strike, but for different expirations. Going short Google September 380 calls at $13.60 against a long in Google December 380 calls at $31.90 would be an example of such a horizontal spread.

All else equal, interest rates play a role, but in all likelihood it is a small one. The longer the time to expiration, the more important interest rates become. But for the most part the bulk of option trading takes place in the nearby expiration months, so rate considerations (all else equal) are not usually a major issue. Time value is what really matters in this type of trade.

As time stretches farther out, more things can happen. Consequently, longer times to expiration result in greater (absolute) option premiums being paid. The stock has more room to run in either direction; plus unexpected events can lead to volatility spikes. Moreover, time-value or gamma trades can be used to back into synthetic calls or puts. For instance, take the example of going short Google 380 calls for September delivery against going long 380 calls for December delivery. If the September calls are in-the-money at expiration, the short will be called, leaving a position of short Google stock and long December Google calls, which is the mathematical equivalent of being long a synthetic 380 December put on Google. The trade can be neutralized by selling a December 380 put.

On the other hand, the position can be rolled by writing October calls to replace the expiring September calls in the event that Google falls below the strike and the September calls go unexercised. That eventuality presents the choice of which strike to write. If prices remain unchanged, an October 380 call can be sold and the horizontal spread is reestablished, albeit with a time frame shortened by one month. (Failure to roll the position leaves a naked short against the December 380 call). Another possibility is to write October calls, but with a different strike. That would establish a diagonal spread, so called because it combines the directional features of a vertical spread with the time-decay features of a horizontal spread.

Diagonal spreads can be either upward or downward sloping. Going long October 380 calls against short December 400s is an example of an upward-sloping diagonal. Reversing the position would create a downward-sloping diagonal. All else equal, owning the lower-priced strike of an upward-sloping diagonal call spread is a bullish posture. The wider the distance between the long and short strike prices, the more bullish the bias if the trade is long a lower strike on the front month of the diagonal. This assumes a 1:1 ratio of notional longs to shorts. On the other hand, the bias of the position can be neutralized or even completely reversed by changing the ratio of longs to shorts.

STRADDLES AND STRANGLES

Straddles and strangles are plays on volatility rather than direction. To go long a straddle is to purchase both the put and the call on a stock at the same strike price and expiration date. Going long a straddle is making a bet that by option expiration, the stock will trade at a price outside a zone whose boundaries are the sum of the prices of the puts and calls. For instance, assume that with Google trading at a price of $380, the price of a September $380 straddle would be the sum of the 380 puts ($11.00) and the 380 calls ($11.00), or 22 points. A buyer of that straddle is betting that by expiration, Google will be trading (or will have traded) either below $358 or higher than $402, each of which constitutes a boundary 22 points from the strike price.

A market-neutral strangle consists of a combination of puts and calls equidistant from the current market price. For example, going long the same quantity of Google September 400 calls ($4.70) and 360 puts ($3.30) would constitute a long position in a strangle. This is a bet that by expiration day, Google will trade outside the 360–400 boundary points by over 8 points, which is the combined premium of the puts and calls. If Google trades beyond those boundary points, one of the options will have an intrinsic value in excess of 8 points, which is the cost of putting the trade on to begin with.

Straddles and strangles are volatility bets in the sense that the buyer will profit on expiration day if the underlying stock trades through the boundary points by an amount greater than the combined initial cost of the puts and calls. But it is not necessary to wait for expiration; nor is it necessary for the stock to pierce the boundaries to profit. If volatility picks up significantly before expiration, option premium will expand, thus giving the strangle or straddle holder an opportunity to sell at a profit before expiration date. Conversely, a flat market will shrink option premiums, resulting in a trading loss.

Strangles can be given a market tilt by using strikes that are not equidistant from the current market price. An example would be to go long Google September 350 puts and 390 calls with the stock trading at 383. That would give the trade a pronounced upward bias. It is also possible to lever directional trades by selling naked options on one side of the market while going long on the other. An example would be writing naked Google September 370 puts at ($5.20) while going long Google 400 calls

($4.70). The cash from writing the naked puts finances the purchase of the calls. As long as the market for Google at expiration is higher than the put strike price, plus net option premiums paid, the position is profitable.

SUMMARY

Capital market instruments come in two basic flavors: debt and equity. Borrowers issue debt to lenders and pay an interest rate for the use of the money. Debt is issued in many forms. Money market instruments have maturities of one year or less; capital markets extend beyond a year. Coupons can either be fixed or floating; some debt is issued without any coupon at all. Corporations, municipalities, and governments sell debt to finance operations and investments.

Corporations raise equity capital by selling ownership in the business in the form of shares of stock. The shareholders'claim on assets is contingent on the prior claims of the firm's debtors in the event of default. Since the shareholders are the owners of the business, the stock market is the market for corporate control. Other capital market instruments are hybrids that combine features of equity and debt. Still other instruments confer rights but not obligations. Options are one such instrument.

Option contracts give the holder the right, but not the obligation, to buy or sell an asset at a set price within a set time frame. They are extremely flexible instruments for establishing positions and managing risk. The principal drivers of option prices are (1) the price volatility of the underlying stock, bond, index, or commodity; (2) time to expiration; and (3) short-term interest rates. These factors are generally applicable to options markets for many types of instruments. The Black-Scholes option pricing model is the basis of most valuation models.

Perhaps the most important and most difficult variable to model is volatility due to its time-varying nature. Not only is volatility itself highly variable, but volatility tends to cluster. But volatility clusters, like asset price bubbles, are more visible in retrospect than they are in real time.

Options can be used on either (or both) sides of the market. They can be used for speculative purposes. Or they can be used for risk management purposes, including targeting specific price zones and time frames for putting on or taking off positions. Puts and calls are mirror images of each other and can be combined to fine-tune market exposure.

This page intentionally left blank

CHAPTER 8

Swaps and Listed Derivatives

> Derivatives are financial weapons of mass destruction, carrying dangers that, while now latent, are potentially lethal.
>
> *–Warren Buffett, March 2003*

When Warren Buffett referred to financial derivatives as weapons of mass destruction, he had just finished toting up potential losses running in the neighborhood of a few hundred million dollars. So he was having a bad day. But blaming the referee has never been a very good strategy. It's just a distraction from the main event, which is playing the game well.

Derivatives are financial instruments whose value derives (hence the name) from some other reference point. The reference point can be almost anything: an interest rate, stock, bond, commodity, index, currency rate, or economic event. This allows contracting parties to gain or shed exposure to an asset or asset class without the requirement of owning the underlying security or instrument. In effect they are side bets. But, according to Nobel laureate Merton Miller, they are side bets that have made the world safer rather than more dangerous. They provide a mechanism through which risk can be diffused rather than concentrated. Moreover, derivatives can do this transparently, freely, and at market clearing prices.

Derivatives contracts trade both over the counter and on the major exchanges. The market for these instruments is astonishingly large. The Bank for International Settlements, which tracks derivatives, reports that by the end of June 2006 regularly organized exchanges had about $26 trillion in notional derivatives contracts outstanding, the great majority in interest rate derivatives. There were even more outstanding over the counter, about $285 trillion at the end of 2005, although methodological issues may greatly exaggerate the amount through double counting and lack of offsets.

Over-the-counter derivatives range from standard plain-vanilla interest rate swaps to extremely complex exotic options, tailored for specific purposes. The ability to tailor derivatives for specific purposes and clients is one of the great strengths of the OTC markets. Exchange-traded derivatives tend to be standardized contracts on generic instruments. The formidable strengths exchanges bring to the market include price discovery, liquidity, transparency, comparative simplicity, and a ready clearing mechanism for the many different derivatives they trade. Because of their different strengths, the OTC and listed derivatives markets complement each other.

The purpose of this chapter is to provide an overview of the more important exchange-traded derivative instruments without getting too far into the specifics. More discussion of contract specifics will come later in the chapters that explore the use of specific derivative contracts to implement the particular trading strategies for which they are best suited. It is also worth bearing in mind that tailored OTC derivatives possess many of the same structural characteristics as the more generic exchange-traded ones.

CBOT INTEREST RATE CONTRACTS

The Chicago Board of Trade and the Chicago Mercantile Exchange are the two major traders of dollar-denominated interest rate futures in the United States. The CBOT trades futures and options on an entire array of debt instruments. The list includes federal funds, Treasury notes (2-year; 5-year, 10-year) and long-dated Treasury bond contracts. The Board of Trade also has markets in interest rate swap futures. The Chicago Mercantile Exchange, "The Merc," trades futures and options on Eurodollar futures and interest rate swaps.

CBOT fed funds contracts are widely followed as forward-looking indicators of Federal Reserve monetary policy. Federal funds contracts are designed to mimic the average 30-day federal funds rate for each calendar month of the year. The actual rate is the monthly average of the daily trade-weighted rate as reported by the Federal Reserve, based on actual cash market transactions. At any one time 24 contract months are available for trading. Typically, the string of monthly contracts for which there are reasonably liquid markets stretches out about six or seven months, so market expectations extending out about a half year are on display.

Fed funds futures are quoted as an index price. The implied interest rate is determined by subtracting the futures contract price from 100. For instance a federal funds rate of 5.25% would be 100 – 94.75 = 5.25, with the futures price being 94.75. The notional value of the contract is $5 million, and it is quoted in increments of ½ a basis point. Each basis point is worth $41.67; one half a basis point is worth $20.835 per contract.

The largest volume of trade at the CBOT takes place in its complex of Treasury contracts. Treasury contracts are designed to look like a cash Treasury within specific maturity zones. The chosen maturities are the more popular ones along the yield curve where the Treasury does the bulk of its coupon financing. The contracts are structured with notional 6% coupons and par values of 100. No interest accrues on the notional coupon. Treasury futures are listed with quarterly expirations using the months March, June, September, and December.

Treasury contracts positions can be liquidated either by the purchase (or sale) of an offsetting position or by making (or taking) physical delivery of cash Treasury securities. For each Treasury contract there are cash notes or bonds that constitute good delivery against contract short positions. Each deliverable cash Treasury has a factor assigned to it. The factor, which approximates the price of the note at a 6% yield to maturity, is multiplied by the futures price to determine the cash bond's delivery price. The CBOT publishes both the delivery factors for outstanding Treasuries and the list of delivery eligible securities. The details for calculating delivery factors as well as delivery rules are published on the CBOT Web site: www.cbot.com

In addition to trading fed funds and Treasury contracts, the CBOT trades 5-year and 10-year interest rate swap contracts. The contracts settle for cash and are listed quarterly using the same contract expiration months as Treasuries. The final settlement price for each swap contract is based on the benchmark rate (for that particular maturity), as determined by the International Swap Dealers Association (ISDA). The formula for converting the rate to a price, published on the CBOT Web site, essentially treats the swap contract as a synthetic bond. The final trading day is the second London business day preceding the third Wednesday of the delivery month.

CME RATE CONTRACTS

The most actively traded interest rate instrument at the Chicago Mercantile Exchange is the Eurodollar contract which easily trades over 1.5 million on the typical day. The CME Eurodollar contract represents a $1 million 90-day Eurodollar deposit. Most contract listings are for quarterly expirations in the familiar March, June, September, December cycle. For the nearby six months, contracts are listed serially, month by month. The vast bulk of the trade is in the first few years of the quarterly listings, which extend out to 10 years.

The quoting convention for Eurodollar contracts is essentially the same as for fed funds. The implied deposit rate is 100 minus an index number. For instance a 4.5% implied rate would be 100 – 95.5. Each index basis point is worth $25 per $1 million contract. The $25 constant basis point is derived from the calculations used for cash 90-day Eurodollar deposits. Left on deposit in a bank for 90 days, $1 million of Eurodollars would earn $2,500 in interest for each 100 basis points, or one full percentage point. One basis point is one one-hundredth of that amount or $25 per contract per basis point. About 85% of Eurodollar trading is executed electronically via the CME Globex trading platform.

CME Eurodollars are often mixed and matched in various combinations of maturities. The exchange facilitates these trades by offering execution capabilities in packs and bundles. *Packs* are simultaneous transactions of four equally weighted and consecutive Eurodollar futures. They are quoted as an average net change from the previous day's close. *Bundles* allow traders to execute a series of equally weighted and consecutive futures contracts beginning with the first quarterly expiration. This allows traders to execute entire strips of Eurodollar contracts at one shot.

In the very short end of the market the CME trades one-month LIBOR contracts with a $3 million notional value. Its construction and trade is similar to 90-day Eurodollars but for the notional amount being three times larger and the maturity only ⅓ as long. Twelve monthly contracts are listed at any one time. The minimum quoting increment is ¼ of one basis point, worth $6.25 per contract.

Like the CBOT, the CME has listed swap futures contracts, with 2-year, 5-year, and 10-year maturities available for trading. Notional values are $500,000 for the 2-year; $200,000 for the 5-year, and $100,000 for the 10-year. They settle for cash against the ISDA benchmark rate for the same maturity. Unlike CBOT swaps that are quoted like bonds, the Merc's swaps are quoted using an index number like Eurodollars. Owing to differences in notional values, tick values are the same, with one point (or 0.01) equal to $100 per contract. The minimum tick increment in each is ¼ of one basis point, or $25 per contract.

In addition to interest rate contracts, the CME lists futures contracts on foreign exchange (FX) rates. Contract offerings include a full array of currencies like the Euro, British pounds sterling, the Chinese renminbi, russian ruble, japanese yen, mexican peso and canadian dollar. Most settle in the familiar quarterly cycle; Mexican pesos are listed for 12 consecutive months. Details on all CME contracts can be found at the Web site: www.cme.com.

EUREX

Eurex is the world's largest derivatives exchange. In the money markets Eurex trades Euribor, which is analogous to LIBOR, the difference being that it is denominated in Euros rather than dollars. In the capital markets

Eurex trades German government bond futures. The maturities traded are 10-year, 5-year and 2-year notes, better known as Bunds, Bobls and Schatz, respectively. While Bunds, Bobls, and Schatz are German government bond derivatives, they actually represent a far larger portion of the market for European sovereign debt than that would suggest, because the German government bond market is the biggest in Europe. It serves as the benchmark for hedging and pricing virtually all Euro-denominated European sovereign debt. Eurex has traded over 3 million Bunds, Bobls, and Schatz contracts combined in a single day.

The design of Bunds, Bobls, and Schatz derivative contracts closely resembles CBOT Treasury futures contracts. Each of the contracts has a 6% notional coupon and face value is either €100,000 if German Bonds are delivered or CHF100,000 if certain eligible bonds from the Swiss Confederation are delivered to fulfill contract obligations. Schatz notes cover a maturity range of 1.75 to 2.25 years, Bobls 4.5 to 5.5 years, and eligible Bunds span a range of 8.5 to 10.5 years. Quarterly expirations are March, June, September, and December. Details of all Eurex contracts can be found at their Web site: http://www.eurexchange.com/index.html

EURONEXT.LIFFE

Euronext.liffe is a pan-European derivatives exchange that trades a number of debt instruments including a wide range of short-term interest rate contracts (STIRS). Included in this list are Eonia, Euribor, short-term sterling, Euroyen, and EuroSwiss futures and options. Euronext.liffe dominates trading in Euribor with over 200 million futures and options contracts changing hands in the first quarter of 2005. Euribor contracts settle for cash. The reference rate is the benchmark rate for three-month money established by the European Banking Federation and the Financial Markets Association.

In the capital markets Euronext.liffe trades futures and options contracts on British government gilt-edged bonds (gilts). The contract design is based on the CBOT model. The contracts have a notional 6% coupon and face value of £100,000. Contract delivery months are March, June, September, and December. The range of acceptable maturities is 8.75 to 13 years.

Like the Chicago Board of Trade and the CME, Euronext.liffe has launched a suite of swap contracts, called SwapNotes, that includes 2-year, 5-year, and 10-year maturities, denominated both in Euros and U.S. dollars. The structure of Euronext.liffe swapnotes is similar to that of a bond with a notional 6% coupon. It is referenced against ISDA and settles for cash. Specifics of the settlement procedures are available on the Web site: www.euronext.com

EQUITY DERIVATIVES

Beside options there are two major classes of listed equity-based derivatives. The first class is represented by futures contracts designed around equity indexes. The second class is composed of exchange-traded funds (ETFs), which are the equivalent of mutual funds designed in such a way that they can trade real time like a listed stock. ETFs are growing rapidly in acceptance and account for a substantial volume of trading. They are discussed in detail in a later chapter.

By far the most active equity index is the S&P 500, traded at the Chicago Mercantile Exchange. Two other active equity index contracts traded at the CME are the Russell 2000 and the Nasdaq 100. Futures contracts on the Dow Jones Industrial Average are traded at the Chicago Board of Trade. Euronext.liffe trades futures and options contracts on the FTSEurofirst 80 and 100 equity indexes. Eurex, jointly owned by Deutsche-Bourse and SWX Swiss Exchange, trades futures and options in the DAX, Dow Jones Global Titans, and Dow Jones Stoxx indexes.

The design of the various equity index futures contracts doesn't vary much by exchange. A multiplier is applied to the index to give it a notional value. The indexes settle for cash against a special quotation designed for the purpose, and expirations occur along the familiar March, June, September, December cycle. The CME and the CBOT allow margin offsets for opposite sides of the market positions in the Dow Industrials and the S&P 500.

ECONOMIC AND EVENT DERIVATIVES

An entirely new class of derivatives that focuses on events is developing. There are now policy-based, weather, and economic event derivatives being traded at various exchanges. These products have the potential to become powerful capital market tools because they seek to capture actual events and processes that have an impact both on financial markets and on economic decision making on the real side of the economy. In so doing, they may allow businesses and consumers to directly partition off pieces of risk that in the past could only be approached indirectly, if at all.

The CME has launched a whole series of derivatives that include such macroeconomic variables as the core Consumer Price Index, the Institute for Supply Management Purchasing Manager's Index, Non-Farm payrolls, retail sales, U.S. GDP, initial jobless claims, and the U.S. Trade balance. One of the more interesting economic contracts launched recently is a housing futures contract, based on the S&P/Case-Shiller Housing Price Indexes. Weather-related contracts have also been launched that are based on snowfall and heat indexes for different sections of the United States. In theory, these indexes may allow insurers and other commercial hedgers to diversify weather-related risk embedded in their businesses.

On the policy front, products and exchanges for trading pollutants and greenhouse gases are being developed. The Chicago Climate Exchange (CCX) was originally developed after a feasibility study funded by the Joyce Foundation. It offers markets for pricing voluntary but legally binding agreements to reduce greenhouse gas emissions. The European Climate Exchange, a wholly owned subsidiary of the Chicago Climate Exchange, offers futures markets in carbon emission allowances that are listed on the Intercontinental Commodity Exchange (ICE). Details of the various contracts are offered on the CCX Web site: http://www.chicagoclimatex.com/

Small and developing markets are sprouting up everywhere. A U.K. consultancy has developed a secondary market for life insurance policies (http://www.epex-group.com/index.php). The Defense Department took a beating in 2003 when it came to light that the Defense Advanced Research Project Agency was developing the idea of policy futures that, among other things, would allow players to speculate on possible terrorist attacks and other geopolitical scenarios. The project was closed down in the ensuing uproar after its existence was revealed in the press.

There have been many event type markets that aggregate information about possible "noneconomic" events. Robin Hanson, an economics professor at George Mason University, is quoted in *Reason* magazine as saying that the Iowa electronic markets predict election results better than pollsters do just as weather futures predict the weather better than does the National Weather Service.[1] Tradesports.com, a betting exchange, already provides a whole raft of markets in sport, economic, political, and other current events. And *Wired* magazine quotes the *New Yorker* to the effect that traders at the Hollywood Exchange guessed 35 of 40 Oscar nominees in the eight biggest categories.[2] That the market can bring to the surface insight gleaned from lots of disparate individuals should not be surprising to many traders.

SWAPS

The versatility of swaps makes them powerful and important capital market instruments. Not surprisingly dealing in swaps is a big business—and getting bigger by the day. The Bank for International Settlements reports that as of December 2005 the notional amount of outstanding foreign exchange and interest rate swaps was somewhere in the vicinity of $180 trillion, with a market value of about $5.250 trillion. The bulk of the issuance is in the form of interest rate swaps, accounting for about $4.9 trillion. But FX swaps are nothing to sneeze at, especially considering that FX markets are far more volatile than bond prices by an order of magnitude. In addition to interest rate and FX swaps, the major swap dealers also carry equity and commodity swaps on their books, mostly in the form of index swaps.

Swaps are unlike traditional financial instruments. Stocks and bonds are part of the capital structure of a firm. Stocks represent equity ownership; bonds, debt. Futures contracts on commodities (as well as stocks and bonds) are a side bet on market performance. It could also be argued that futures contracts are an implicit claim on an underlying asset when there is a physical delivery requirement. Option contracts similarly represent a contingent claim. Swaps, on the other hand, are a technique for pricing and transferring bits of risk from one party to another. Almost any asset class, index, or commodity is amenable to a swap. The same methods used to construct interest rate, FX, or commodity swaps can just as easily be employed to design a swap of kitchen tables for dining room chairs.

SWAP STRUCTURE

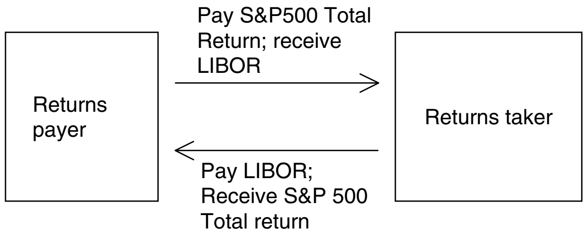
Swaps are an agreement between two contracting parties (the counterparties) to exchange payments based on some agreed-upon criteria. At the outset the swap is priced where the present values of the expected payments are equal. In principle, swaps can be (and are) used to price and transfer exposure to all kinds of market risk and circumstance. It is simply a matter of how the terms are defined and the transactions cleared. They can, for example, be structured in terms of total returns or excess returns. The reference asset or indicator can be interest rates, currency rates, equity indexes, commodity indexes, loans, real estate, art, or dividends, to name a few. They can be used in combination to diversify, shed, or acquire exposure across different types of asset classes. Going beyond asset classes, there are swaps on types of market action, including returns variance swaps that offer exposure to market volatility.

The basic structure of a swap is one in which the first party to the transaction (the returns payer) pays the other a sum of cash based on the performance of an underlying asset. In exchange, the counterparty (the returns receiver or taker) pays the first party a short-term interest rate, generally LIBOR, on the notional amount invested. This structure allows the counterparties to off-load (or acquire) exposure to the asset or asset class without necessarily giving up (or acquiring) ownership of it.

For instance, consider the structure of an equity index total return swap. One party (the payer) agrees to make a cash payment equal to the total return (from price change and dividends) on $1 million invested in the S&P 500 over the next year. In return the payer receives an interest payment on $1 million principal, based on the LIBOR rate over the next year. See Figure 8.1, a diagram of the cash flows.

FIGURE 8.1

Cash Flow Diagram of an Equity Index Total Return Swap



INTEREST RATE SWAPS: FIXED OR FLOATING?

The most widely used swaps are interest rate swaps in which the counterparties exchange fixed for floating interest payments. To see how they work, we will work through an example.

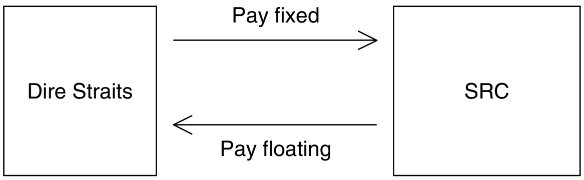
Assume the following. The first party to the transaction is the Dire Straits Savings Bank (DSB). It takes in short-term deposits for which it pays money market interest rates. It uses these deposits to fund 15-year mortgage loans that it has on the books. Dire Straits Savings can borrow short-term money cheaply because the deposits are guaranteed by federal insurance. But borrowing short-term money to fund long-term fixed-rate loans leaves the bank vulnerable. A rise in short rates over the fixed-rate loans on the books will cause carry losses and deplete capital. On the other hand, to borrow money in the long end by issuing bonds, the bank has to pay a 50-basis-point premium over Treasuries to attract investors.

The second party, the Solid Rock Corporation (SRC), has a credit rating of AAA and can borrow long-term money cheaply. But it has no need for long-term money. The balance sheet is not being used to its potential. The resulting duration mismatch between assets and liabilities and sources and uses of funds can be swapped away. Dire Straits Savings can negotiate a swap with the Solid Rock Corporation. The bank agrees to pay the Solid Rock Corporation a fixed interest rate on $100 million principal, semiannually for 10 years. In turn, the Solid Rock Corporation agrees to make an interest payment based on $100 million in principal to Dire Straits every six months at the six month LIBOR rate.

The Solid Rock Corporation issues bonds in the credit markets to fund the transaction. It then invests the proceeds of the bond sale in the money markets. And it will pass through some of the fixed-rate payments it receives from Dire Straights to the bondholders. But only some. SRC can do this profitably because it charges Dire Straights a higher coupon rate than it owes to SRC's bondholders. In effect, SRC is renting out its balance sheet to Dire Straits. On the other side of the transaction, Dire

FIGURE 8.2

A Swap of Fixed for Floating Rate Interest Payments



Straits has borrowed long money from SRC more cheaply than it otherwise could have, and has succeeded in more closely matching the duration of its assets with its liabilities. See Figure 8.2, a schematic of the swap.

The best way to see how this works in practice is to work out the numbers. To evaluate the swap, posit the following series of transactions and rates. SRC sells $100 million 10-year SRC 5.25% notes priced at 100 to yield 5.25%. Upon completion of the sale of the $100 million in notes, SRC reinvests the cash at the LIBOR rate. Dire Straits Savings Bank agrees to pay SRC 5.375% interest on $100 million for 10 years, in semiannual installments. In return, SRC agrees to pay Dire Straits interest on $100 million every six months at the six-month LIBOR rate. The current six-month LIBOR rate is 4.75%. We also note that if Dire Straits had gone to the market directly to issue its own 10-year notes, it would have had to pay a 5.5% interest rate.

Since SRC has sold $100 million of notes with a 5.25% coupon, the firm is obligated to make 20 semiannual payments to the note holders of $2,625,000 apiece. On the other side, SRC will be receiving 20 semiannual payments of $2,687,500 apiece from Dire Straits based on the 5.375% swap rate. That is the fixed-rate side of the transaction. On the floating side, SRC is going to pay Dire Straits the six-month LIBOR rate on $100 million every six months. Effectively, that is a wash transaction. SRC invested the proceeds of its note sale at the LIBOR rate. It then simply rolls over its money market investment every six months and passes the interest through to Dire Straits. In turn, Dire Straits passes these short-term interest payments along to its depositors.

The result is that SRC picks up the difference between coupon payments to its note holders and its fixed payment receipts from Dire Straits. The difference amounts to $62,500 every six months. In present value terms, SRC is better off by $962,916 using a 5.25% discount rate. But Dire Straits is better off as well. If Dire Straits had to go to the capital markets directly, it would have needed to pay a 5.5% rate to borrow 10-year money, rather than the 5.375% rate it pays to SRC.

Dire Straits Savings Bank has a borrowing cost that is lower than it otherwise would have been—to the tune of $62,500 every six months.

For all intents and purposes, SRC has rented out its balance sheet to Dire Straits, for which it receives $962,916 (in present value terms) in compensation. Dire Straits has reduced its borrowing costs by the same amount and reduced duration mismatch exposure between its assets (the mortgage portfolio) and its liabilities (deposits). Both parties are better off than they otherwise would have been.

In principle, the technique of partitioning and trading off bits of risk exposure from many different types of assets is widely applicable. As a result, returns associated with foreign equity markets, market sectors, commodity indexes, currencies, and energy can be (and are) the object of swap transactions. The transactions depend on a reference index or rate, a trade clearing process, and enforcement of legal contracts.

MARKET VENUE

Swaps mostly trade in the OTC markets. There are several reasons for this, not the least of which is that dealers can tailor swaps for the individual needs of their counterparty customers. Another factor is that the OTC swap markets are extremely liquid, in part because plain-vanilla swaps simply represent a plain unadulterated interest rate. In this they are unlike bonds. Each bond issue is distinctive. It has a coupon (which may be zero), a maturity, and a unique distribution of ownership. Selling bonds that are not owned, which is to say selling bonds short, requires finding an owner willing to lend them, which adds to transactions costs. Taking either side of the market via an interest rate swap simply requires writing a trade ticket. No fuss, no muss, no bother.

SUMMARY

The major exchanges have recently made forays into the swaps marketplace. The Chicago Board of Trade, the Chicago Mercantile Exchange; the London International Financial Futures Exchange (LIFFE); and TIFFEX, the Japanese Financial Futures Exchange have each launched some version of a swap futures contract. An important selling point emphasized by the exchanges is that the clearinghouse allows swap transactions to come off the books easily without going back to the original counterparty to clean up the transaction.

Beyond swaps, a wide variety of financial market derivatives trade at established exchanges. The more traditional derivatives reference financial instruments like stocks and bonds and indexes. New innovative products, tools, and trading platforms are being developed that extend beyond traditional financial instruments into derivatives based on macroeconomic variables as well as housing and policy-driven outcomes like pollution emissions. The evidence suggests that derivatives have made the world safer, not riskier. They have done so by creating a mechanism to transfer risk to those most able and willing to shoulder it. And by pricing risk transparently, derivatives traded on exchanges have made valuable price signals available to businesses and consumers.

NOTES

[1] See *Reason Magazine*, July 30, 2003: http://www.reason.com/rb/rb073003.shtml

[2] See *Wired*, July 30, 2003: http://www.wired.com/news/politics/0,1283,59818,00.html

CHAPTER 9

Fed Funds: The Price of Policy

The federal funds rate is quite possibly the most important benchmark interest rate in the world. It is the primary policy tool the Federal Reserve uses to guide money and credit growth; as such it affects the cost of borrowing throughout the entire U.S. economy. Ripple effects from changes in the funds rate spread through the global economy via its effect on consumption and investment decisions and through linkages with foreign exchange markets. As global capital and money markets become more tightly integrated, it is reasonable to expect that the ripple effects will become more pronounced.

Federal funds are non-interest-earning reserves that banks are required to keep on deposit with the Fed. Since the Fed pays no interest for these deposits, the banks try to keep them to a minimum, selling any excess reserves at the prevailing rate of interest. Typically the buyers of fed funds are large money center banks that need to buy money to fund their lending activities. The sellers tend to be smaller banks whose deposits exceed the requirements of funding their loan portfolios. The great majority of federal funds transactions are executed on an overnight basis, although some transactions stretch out longer, sometimes for a week, a month, or even a quarter. The transactions that bear these longer end dates are executed in what is known as the *term market.*

In practice the funds rate is an administered rate, set by the Federal Reserve. There is seldom any mystery about what the target is, because beginning in February of 1994 the Fed adopted a policy of announcing it. The Fed's modus operandi is straightforward: It pegs the funds rate to its announced target. When the funds rate trades higher than the target, the Fed injects money in the system, usually via *repurchase agreements* (otherwise known simply as *repos* or *RPs*), forcing the rate down. Conversely, when the funds rate trades lower than the target, the Fed drains money from the system, pushing the rate back up.

The funds market is gigantic. Dealers in government securities settle their trades and finance their positions in fed funds. According to the stats published on the New York Fed's Web site, it is not unusual for dealers to carry gross financing positions worth well over $1 trillion. These positions are carried on the dealers' "matched books," the equivalent of a banking operation in which money is lent and borrowed via repo and reverse repo transactions. To put the size of the market in perspective, note that every basis point shift on $1 trillion in the fed funds rate is worth about $27,778 in changed borrowing costs.

The Fed executes transactions both for its own account and on behalf of customers, usually foreign central banks and U.S. Treasury subaccounts. When the Fed enters the market to execute transactions for its own account, it uses the system open market account (SOMA). Transactions for the system account are executed regularly. In 2005 the Fed was in the market on all but seven business days, executing overnight RPs on 204 different occasions. The average transaction size was $6.4 billion. It also executed 14-day term repos on every Thursday in 2005. The average size of those transactions was $8.7 billion. In 2004 it executed overnight repurchase agreements on 195 trading days; for 2003 the number was 179 days. Clearly, the funds rate trades where it does because the Fed pegs the rate as part of its open market operations.[1]

While there is little mystery about what the target rate is, there is often considerable uncertainty about what the target is likely to be in the future. The further out in time, the more uncertainty there is. The Fed has taken considerable pains to ameliorate this in its efforts to shape expectations about the future. Not only does it announce its target rate for federal funds, but it often includes forward-looking statements to indicate its thinking about the state of the economy, inflation (or deflation) pressures, and the likely course of future interest rate policy. One result of this policy has probably been some reduction in market volatility. But the reduced volatility comes with a potential monkey wrench. The Fed has become so adept at managing expectations that it may have created an echo effect. Market expectations may begin to reflect the Fed's expectations, depriving it of an independent feedback loop to use in assessing policy efficacy. It may also have built some moral hazard into the system. Traders may be more aggressive taking on large positions because they think future fed policy is more certain than it actually is.

Traders were so confident in Chairmen Greenspan's ability to stave off any large-scale disasters that they coined the phrase the "Greenspan put." No doubt the chairman, famously libertarian and an early devotee of Ayn Rand, must have found that quite amusing, given his marked preference for deferring to the market's judgment.

Gaining and maintaining the market's confidence is vitally important to the Fed. However, even with the market's trust, implementing monetary policy can be a decidedly tricky affair. It revolves around a conundrum that the Fed has been able to finesse in recent years, but it is always lurking in the background. The problem is that the Fed can control either the money supply or the price of money—the federal funds rate—but it can't do both. The Fed can manage the money supply by regulating the quantity of reserves it supplies to the system. But letting the quantity grow at a fixed rate, say 4% a year, means that the price will gyrate up and down, perhaps wildly, as money demand increases or decreases. The Fed can fix the price of money, but then the money supply becomes almost infinitely elastic. To maintain a steady funds rate in the face of increased demand, the Fed needs to expand the money supply by adding reserves to the system. On the other hand, if demand begins to fall off, the Fed will need to sop up excess liquidity to prop the funds rate up.

The problem is that pegging the rate "correctly" is very nearly impossible. But the Fed doesn't have a lot of choice because financial innovation has left it with an uncertain transmission mechanism for policy implementation. In this respect pegging the funds rate may have become like democracy—a very bad form of government except for all the others. The Fed likes to say that its policy inclination is contracyclical; that it tends to "lean against the wind." When the economy is strong and inflation pressures are building, it errs on the side of tight money. And when the economy is weak and inflation pressures are ebbing, it lowers rates. But unless the Fed guesses the rate "just right," contracyclical policy can quickly become procyclical.

Consider what happens if the Fed tightens policy too slowly as inflation pressures are rising. If the Fed pegs the rate below expectations of future inflation, credit demand will expand, and with it the money supply, adding fuel to the fire. On the other hand, if the economy weakens and the Fed props the funds rate up, there is the danger of a deflationary downward spiral, as in the 1930s.

Moreover, managing the money supply is extraordinarily difficult in the age of high technology because it's hard to define what money is, the forms it takes. and how to measure it. There is also a problem with the transmission mechanism for policy. The Fed can add or subtract all the reserves it wants to, but that won't necessarily get the job done. The relationship between the quantity of reserves and the money supply is tenuous, because among other reasons, the velocity (or turnover rate) of money cannot be counted on to be stable.

One possible solution to the policy problem is for the Fed to adopt formal inflation targets. That way the Fed can concentrate on the main target—price stability—while minimizing the impact of operating procedures on achieving policy goals. Fed Chairman Bernanke is a proponent of inflation targeting, but he appears to be in the minority. While formal inflation targeting has not been adopted by the Fed as an operating procedure, the Fed may be moving in that direction. Its current practice of communicating the substance of its thinking, in effect guiding the markets, comes pretty close. Through speeches, policy statements, congressional testimony, and leaks to newspapers, Federal Reserve officials can communicate their policy predilections to the market but leave themselves a bit of maneuvering room. In this regard it is probably useful to think of the current policy regime as one of "soft targeting."

Whether the Fed actually adopts a formal targeting regime remains to be seen. But it is hard to imagine going back to status quo ante where monetary policy decision making was shrouded in ambiguity. The trend in governance is clearly in the direction of transparency and openness. The Fed will very likely continue to become more and more open about how it decides policy and the manner in which it implements policy. And an important guide for policy makers will be expectations. That is one reason why futures markets are so important, particularly futures markets in policy making, which is the essence of fed funds futures traded at the Chicago Board of Trade.

THE FED FUNDS FUTURES MARKET

Futures markets, particularly futures contracts in fed funds, can provide a good way to gauge market expectations with respect to Fed policy. Fed funds futures trade actively at the Chicago Board of Trade, which also launched options on fed funds futures in May of 2003. Before the CBOT launched fed funds futures contracts, access to the market was largely restricted to commercial banks and large dealers in government securities. Now virtually anyone can jump into the game and trade fed funds to lock in short-term borrowing rates, hedge short-term financing risk, or speculate on Federal Reserve monetary policy.

Over the years the funds rate has displayed significant volatility. But there is a twist. The volatility of the funds rate is not the same as that of stocks and bonds. Most financial market theorists argue that stock prices combine a long-term drift (or trend) rate plus day-to-noise (or volatility). Over the long run, the trend dominates; but over the very short run, price variation is just a matter of randomly distributed noise, with prices equally likely to rise or fall.

Leaving aside for the moment whether that is true for stock prices, it is manifestly not the case for the federal funds rate. It can't be as long as the Fed pegs the rate, unless the Fed begins to determine policy by the flip of a coin, an entirely unlikely event, to say the least. In the real world, the

Fed targets a rate and aims to hit it on the average for two-week periods ending every other Wednesday (when the banks are required to settle up their positions). To the extent that the funds rate marginally overshoots or undershoots the target on any given day, the Fed takes corrective action, usually via the RP market. Not only does this greatly dampen day-to-day funds rate volatility, but it also means that it is highly unlikely that the Fed will miss hitting the target.

In this regard the funds rate is very different from securities prices which can rise or fall by a significant margin on any given day as prices adjust to changes in supply and demand. The Fed provides a backstop on either side of the funds market. The fed funds rate will not change by much unless the Fed decides to change it. When the Fed decides that it wants to change the rate, it usually telegraphs its intentions well in advance. Moreover the Fed doesn't easily jump from policy easing to tightening. It relentlessly follows the same policy direction until it gets the job done. The old joke is that the Fed knows only two speeds: fast-forward and reverse.

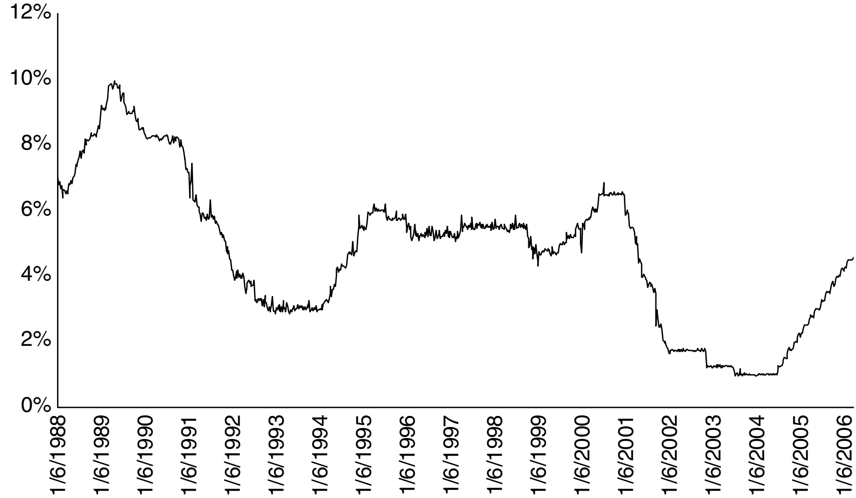
At any point in time, the funds rate is almost never equally likely to go up or down because it is rare that there is an even balance of forces that influence monetary policy. Consider Figure 9.1, a chart of the funds rate that extends from January 1988 through March 2006. There is an obvious stair-step quality to the graph, giving testimony to the fact that the rate is both administered and serially correlated. It is clearly not a random process, but the result of a deliberative one. For most of the time the rate is trending. Moreover the trend does not suddenly reverse. The upward or downward trend plateaus (sometimes for a considerable period) and then finally moves in the other direction, when the Fed changes the direction of its policy stance.

The juncture at which the Fed changes the direction of its policy stance is absolutely critical. From that point on, the Fed can be counted on to execute a series of policy maneuvers in steplike fashion, all moving in the same direction. What's more, it is extremely unlikely the full import of those decisions will be fully and immediately priced into the market.

Observe that there are four distinct tightening cycles on display in Figure 9.1. The first lasts from March 1988 through April 1999; the second, from January 1994 through April 1995; the third, from January 1999 through July of 2000; with the most recent tightening cycle beginning June 2004 and still underway as of February 2006. There are also two distinct easing cycles. One lasts from May 1989 through about December 1992. The other extends from January 2001 through November 2003. Another way to put it is that from January 1988 though February 2006—a period over 17 years—the Fed was either actively easing or tightening policy about 70% of the time.

FIGURE 9.1

Weekly effective fed funds (January 1988–March 2006)



Data source: St. Louis Fed

These large trend moves can be caught and surfed like a wave because that is what they resemble. There are two major prerequisites for doing so. The first is anticipating what the Fed is likely to do. The second is acting on it. Paradoxically, the hard part may be pulling the trigger. Findings from behavioral finance hint that getting up the gumption to act on anticipated events may be more difficult than correctly predicting the outcome, a subject that will be discussed in later chapters. Be that as it may, anticipating what the Fed is going to do isn't all that difficult most of the time. That is because the Fed goes through an elaborate Kabuki dance to inform the market of its intentions. The trick is to understand the dance, or more precisely, *Fedspeak*, the language of the lyrics used by Fed officials and the Federal Open Market Committee (FOMC) when communicating with the market.

READING THE FED

The Fed communicates with the market in numerous ways, including speeches by Fed governors, minutes of FOMC meetings released with a time lag, testimony before congress, and leaks to news organizations. In recent years some of the most important communications have come in the form of FOMC policy statements, which are released immediately after the committee meets. FOMC policy statements sum up the Fed's take on the state of the economy, inflation, and the state of policy with respect to these two critical variables.

FOMC policy statements usually cover four main areas. First, the statement usually leads off with a policy decision (ease, tighten, or stand pat). Second, it is followed by a short statement about the health of the economy. Third, it comments on the inflation picture. Fourth and finally, the FOMC includes a paragraph or so about which way policy is likely to tilt and why. Using a consistent format like this allows the Fed to prepare the market for changes down the road. Fed watchers will compare current and past statements to see if and how the emphasis has changed, if at all. To see how the Fed tees up policy changes, it is a useful exercise to read through and compare the language contained in a pair of FOMC statements, one before and the other immediately after a policy change.

Table 9.1 presents a side-by-side comparison of key statements taken verbatim from the December 2000 and January 2001 FOMC meetings. The statements are available on the Washington Fed's Web site. The table is organized into four key sections: policy action, economic outlook, inflation, and a forward-looking statement.

The first section is straightforward. The Fed announces what it has decided to do. At the December meeting it decided to leave things unchanged. But at the January meeting it decided to lower the funds target rate by 50 basis points. The stage for the policy shift had already been established. The economic outlook section of the December meeting makes it clear that the Fed was already concerned about slowing economic growth. It sees weakening consumer demand, a falloff in profits, and some stress in financial markets. In the January statement it ties its decision to drop rates to mounting concern about issues it raised in December: declining consumer confidence; weakening output, and tightening conditions in financial markets. The use of the word *stress* is important. When the Fed describes financial markets as "strained" or "stressed," the Fed is eying the ease side of the policy ledger.

Against the softening economic outlook it next discusses inflation. In December it discerns some inflation risks, but it sees no sign of rising inflation expectations, which it expects to moderate in the face of slowing economic growth. By January it refers to inflation as contained, and it goes on to emphasize the lack of inflationary pressure in the economy by referring to continued growth in long-term productivity. Productivity growth implies downward price pressure because it results in greater output (more supply) for the same (or less) input (less resource demand). The forward-looking statements for December and January are exactly the same. On balance they see risks mainly weighted toward economic weakness.

TABLE 9.1

Federal Open Market Committee Statements before and after a Policy Change

Release Date: December 19, 2000	Release Date: January 3, 2001
Policy Action	*Policy Action*
The Federal Open Market Committee at its meeting today decided to maintain the existing stance of monetary policy, keeping its target for the federal funds rate at 6½ percent.	The Federal Open Market Committee decided today to lower its target for the federal funds rate by 50 basis points to 6 percent.
Economic Outlook	*Economic Outlook*
The drag on demand and profits from rising energy costs, as well as eroding consumer confidence, reports of substantial shortfalls in sales and earnings, and stress in some segments of the financial markets suggest that economic growth may be slowing further.	These actions were taken in light of further weakening of sales and production, and in the context of lower consumer confidence, tight conditions in some segments of financial markets, and high energy prices sapping household and business purchasing power.
Inflation	*Inflation*
While some inflation risks persist, they are diminished by the more moderate pace of economic activity and by the absence of any indication that longer-term inflation expectations have increased. The Committee will continue to monitor closely the evolving economic situation.	Moreover, inflation pressures remain contained. Nonetheless, to date there is little evidence to suggest that longer-term advances in technology and associated gains in productivity are abating.
Forward-Looking Statement	*Forward-Looking Statement*
Against the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the Committee consequently believes that the risks are weighted mainly toward conditions that may generate economic weakness in the foreseeable future.	The Committee continues to believe that, against the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the risks are weighted mainly toward conditions that may generate economic weakness in the foreseeable future.

Source: The Federal Reserve Board

The December and January statements are straightforward. By citing both economic weakness and moderating inflation, they make the case for policy ease. The evidence they muster points to more of the same in the future, so they communicate a clear predilection for additional easing down the road. Both in format and substance, the FOMC policy statements are typical of the way the Fed makes its case, and they can serve as a template for future Fed watching.

PRICING POLICY: EXPECTATIONS FOR FUTURE RATE CHANGES

Changes in Fed policy tend to ripple through the financial markets quickly, but it takes a while for the impact to work its way to the real side of the economy. Consequently, the policy process is iterative, with the Fed making periodic adjustments as conditions warrant. But the U.S. economy is a very big boat; it takes a while to turn it around. That is one reason why the Fed works very hard to avoid falling behind the curve. Playing catch-up can be an expensive proposition. Accordingly, expectations play an important role in the process. In that vein it is important to consider two factors. The first is that some rate changes are more important than others, and by a long shot. The second is that fed funds futures can be used to gauge expectations as to the future direction of policy.

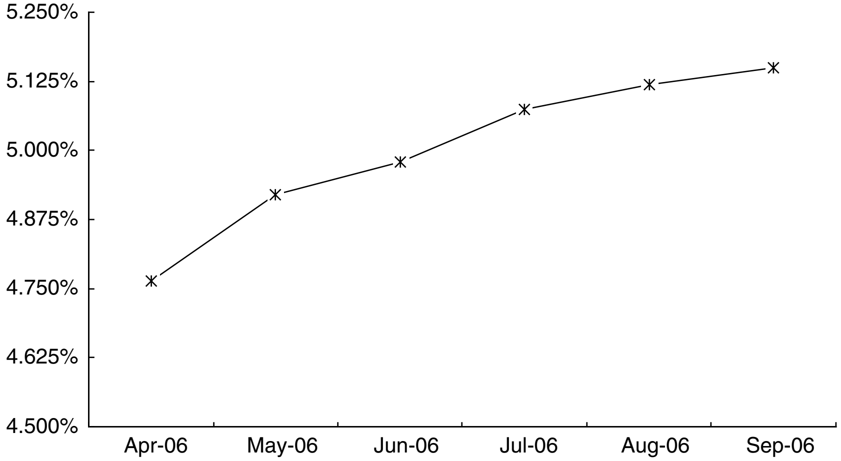
Incremental rate changes in the same direction as previous changes are important. Far more important is a change in the direction of policy. The Fed does not normally turn on a dime. It prepares the market ahead of time using such devices as speeches by Fed governors and adjustments to the tone of FOMC statements. A change in the direction of policy is especially important because it signals a change in the cycle that can be expected to last for a while. Moreover, when the direction of policy changes, expectations begin to change as well, and futures markets begin to price in a series of forward rate adjustments along what can be thought of as a new policy path.

Consider Figure 9.2, a graph that represents market expectations of future fed funds rates. Market expectations of the future fed funds rate can be inferred from the prices of CBOT fed funds futures contracts. The contracts are quoted as index prices, which when subtracted from 100, represent the average effective daily fed funds rate for the calendar month. So for instance, a quote of 95 for the March 2006 expiration would imply a rate of 5%, which equals 100 minus 95. Since each calendar month has a unique expiration, a strip of forward fed funds contracts implies market expectations for the fed funds rate in future months.

By convention, fed funds assume a 360-day year, so 100 basis points is worth \$27.77 per million per day. For example $1\% \times \$1,000,000/360 = \$27.77$. The CBOT fed funds contract is for \$5 million in notional value for 30 days, so by extending the formula, a 100-basis-point change in the rate would be equal to \$4,166.67 per contract. A 1-basis-point rate change is worth \$41.67 per contract, and one-half a basis point rounds out to \$20.835. Specifying the contracts this way allows traders to speculate about or hedge against what they think the fed funds rate is likely to be in each of the contract months for which the exchange lists futures contracts. Most of the trading takes place in contracts out to three months from the spot month, but respectable trading volume takes place out to six months.

FIGURE 9.2

Fed funds futures closing prices (March 31, 2006)



Data source: Chicago Board of Trade

The simple structure of the fed funds contract makes it easy to calculate payoffs from rate change scenarios. For instance, a holder of 20 fed funds contracts priced at 95.375 would be long a notional $100 million worth of federal funds at a rate of 4.625%, If the average rate for the contract month were to finally settle at 4.50%, the holder would receive a payoff of 12.5 basis points, which would be worth 12.5 × 41.67*100 = $10,417.50. On the other hand, if the funds rate for the month were to average 4.75%, the final settlement price would be 95.25, and the holder would lose $10,417.50.

Since the final settlement price is based on the average rate for a given month, it is easy to calculate how much a marginal change in the funds rate on any given day affects the average rate for the month and therefore the final settlement price. Consequently, the futures price embeds an implicit market forecast of the average monthly funds rate. Since the funds rate is pegged by the Fed, the price of the futures contract is really a forecast of Fed policy. And since prospective gains and losses can be easily calculated, the ratio of potential gains to losses is interpreted by some as the market's assessment of the probability of a Fed policy move. To see how this works, consider the following example.

Suppose the Fed's announced overnight target rate for funds is 5% and that the funds rate has traded at the target for the first 14 days of the 30-day month; that the futures contract is trading at a price of 94.90 for a rate of 5.10%; and that the FOMC policy meeting is underway and its decision will be effective the next day, the 15th day of the month. Further assume that Fed policy changes are only decided at regularly scheduled FOMC meetings; that policy is moved in 25-basis-point increments; that policy is established in the direction of tightening; and finally that the Fed unerringly hits the target rate it sets.

Keeping those assumptions in mind, the futures market can be said to be setting the chances of a Fed tightening at 4 out of 5. How is that determined? Consider: The Fed only moves in 25-basis-point increments, and the direction of policy (tightening) is already set. Therefore, the Fed has only two choices. It can decide to tighten policy 25 basis points, or it can leave policy unchanged. If policy is left unchanged, the funds rate will average 5% for the month. If the Fed tightens policy by 25 basis points to 5.25%, the average rate for the month will be 5.125%—15 days at 5% and 15 days at 5.25%.

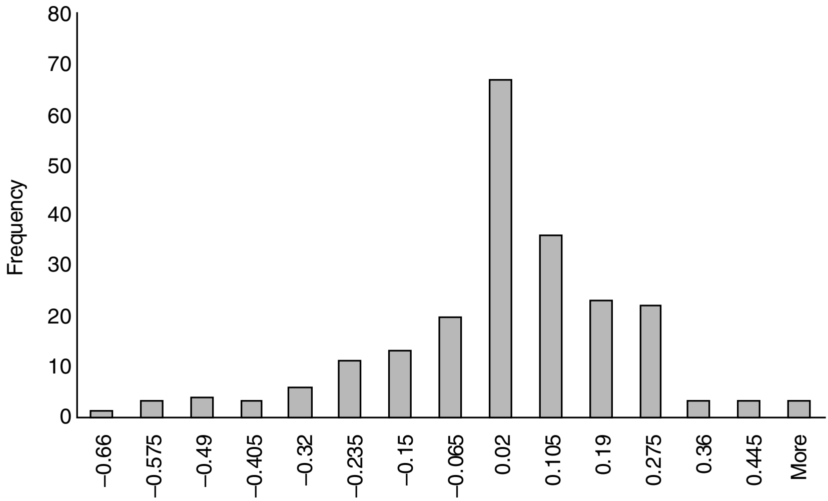
Since there are only two possible average rates for the month—either 5% or 5.125%—the market's assessment of the odds of a policy change can be defined as the ratio of potential gains to losses from the two possible outcomes. The futures contract is priced at a yield of 5.10%. A Fed rate hike will raise the average for the month to 5.125%, or 2.5 basis points over the current futures rate. A Fed decision to leave rates unchanged would result in an average rate of 5%, or 10 basis points lower than the 5.10% rate implied by the futures contract.

In other words, there are two possible payoffs, each associated with a particular policy option: either 2.5 basis points for a rate change or 10 basis points for no change. A fair bet sets the payoff equal to the odds of the outcome. If we assume that the market is efficiently priced—that it is a fair bet—then the ratio of potential gains to losses represents the market's policy expectations. In this case the possible gain is 2.5 basis points versus a possible loss of 10 basis points. The ratio of the two is $10 \div 2.5 = 4:1$, which indicates the market reckons that the likelihood of a rate hike is four times that of an unchanged policy. Odds of 4:1 can be translated as 4 chances in 5, an 80% probability forecast.

Gaining insight on market expectations is an important facet of securities trading, but it is easy to get a bit carried away with this, so some caveats are in order. Mostly they concern the restrictiveness of the initial underlying assumptions about Fed behavior. For instance, in the real world the Fed is not restricted to taking effective policy action only at regularly scheduled meetings. When the occasion calls for it, the chairman will call a meeting or arrange a telephone conference and get the authorization of the board to take action. In response to a genuine crisis, as in the stock market crash of 1987 or the Asian financial crisis of 1998, the Fed

FIGURE 9.3

Monthly change in fed funds (January 1988–February 2006)



Data source: St. Louis Fed

has been known to abruptly change policy from tightening to easing in order to relieve financial stress in the system. Moreover, the Fed is quite capable of moving policy in increments greater than 25 basis points.

An examination of the effective monthly fed funds rate from January 1988 through February 2006 reveals that month-over-month changes in excess of 25 basis points occurred quite frequently. In fact, 48 out of the 218 months during the period experienced changes that exceeded ±25 basis points from the previous month, according to St. Louis Fed data. Not only did changes exceed the assumed boundary limit 22% of the time, but the distribution of changes is skewed, as shown in Figure 9.3. Since 1988 the Fed has shown a predilection to drop rates by a greater margin than it raises them. This is (or should be) entirely unsurprising since it is fully consistent with the Fed's raison d'etre, which includes being lender of last resort to maintain liquidity in the financial system in times of stress.

TRADING STRATEGY

Since fed funds futures settle for cash against the effective monthly rate in the overnight market, all strategy is keyed off the spread between futures and cash. As we have seen, fed fund futures implicitly forecast Fed policy.

In this regard there are two questions that need to be considered in setting up a trading strategy. The first concerns the likely direction of policy; the second, the magnitude of likely policy changes. It may seem patently obvious that the major issue is direction of policy. It is. But in forest-for-the-trees fashion, the market is quite capable of mistakenly anticipating what the Fed is liable to do. That's when big opportunities are created; outright positions can be taken either on the long or the short side of the market when pricing is inconsistent with the likely course of Fed policy.

For some reason or other, there is a widely held belief among bond traders that economic growth is bad for the bond market, notwithstanding mountains of evidence to the contrary. Bond traders tend to associate economic growth with rises in inflation, which causes the Fed to tighten, pushing yields up and prices down. But the Fed is less concerned about economic wiggles and more concerned with inflation and long-term inflation expectations. Globalization, capital market liberalization, along with tax and regulatory reform have greatly expanded the economy's capacity for noninflationary growth. The Fed is well aware of this, taking it into account in its policy making. On the other hand, traders tend to focus on the minutiae of daily economic releases, and sometimes jump the gun trying to anticipate policy changes and market turns. This can lead to some fairly interesting market opportunities. A case in point is the summer of 2002.

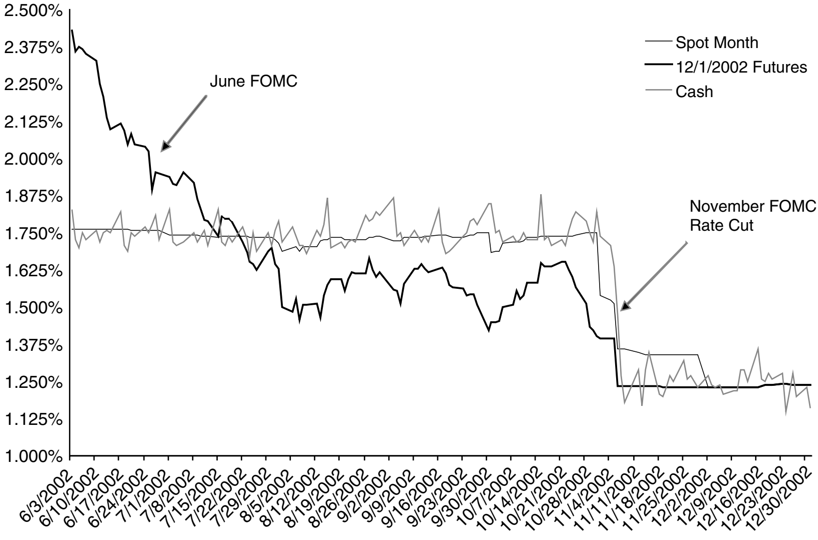
In June of 2002 the fed funds rate was trading comfortably at 1.75%. The Fed had dropped the rate from 6.5% in a series of policy eases beginning early in 2001. The unemployment rate stood at 5.8%, up 1.3 percentage points from the 4.5% rate it registered a year earlier. Year-over-year headline inflation was running in the neighborhood of 1.2%. Real second quarter GDP growth 2.2%, down from 2.7% in the first quarter. Nevertheless the Street was nervous, sensing that the funds rate was too low and that the Fed would have to begin to tighten policy before long. Anticipation of significant policy tightening was in the air. December 2002 fed funds futures contracts were quoted at 2.375%, implicitly forecasting a series of rate hikes (See Figure 9.4).

While back-month futures were forecasting a Fed tightening, the spot month continued to closely track the overnight cash market. In the end, the Fed did not tighten policy; instead it chose to ease policy an additional 50 basis points on November 6, 2002, pushing the funds rate down to 1.25%. A long position in December fed funds futures, taken in early June, would have garnered a full 112 basis points in profit, equal to $4,667 for a single contract. What does it say for trading strategy that spot month contracts closely track cash and therefore Fed policy, while back months can turn out to be wildly off the mark?

For one, it's a lesson in the difference between trades and quotes. December Fed funds futures were quoted at 2.375% in the beginning of

FIGURE 9.4

Daily Fed Funds: Cash and Futures (June 2002–December 2002)



Data sources: Chicago Board of Trade, St. Louis Fed

June, but there were no reported trades until 100 contracts representing $500 million in notional value traded on June 19. Those trades were recorded between 2.09% and 2.05%, still a respectable discount from cash, but not nearly as exaggerated as the initial quotes would suggest. Still, in the time between mid-June and mid-July about 3,500 contracts representing about $17.5 billion in notional funds traded at substantial discounts, suggesting that the market thought the Fed was headed toward tightening.

Research suggests that back-month fed funds futures do not provide very precise estimates of market expectations for future Fed policy. Fed funds futures contracts systematically overestimate the likely future funds rate during periods of policy ease, and they underestimate the likely future rate during periods of rising rates.[2] But they are the most useful money market instrument for predicting the likely course of monetary policy over short-term horizons of a few months. However, for time horizons of five or six months, Eurodollars seem to do a better job, perhaps due to their greater liquidity.[3] This may be changing as liquidity in fed funds futures has picked up markedly in recent years. More to the point, fed funds futures have been shown to react to monetary policy surprises and to reflect changes in expectations for longer time horizons due to FOMC policy announcements.[4]

One conclusion that can be drawn is that near-term probability estimates of Fed policy choices are not easily translated into longer-term probabilities. But that also implies significant trading opportunities based on superior insight about Fed policy. That leaves the question: Was the fall in the funds rate from June through December foreseeable? Was there enough information available at the time to warrant taking a substantial outright long position in fed funds futures?

The simple answer is yes, and it speaks volumes about low-risk, high-reward trades hiding in plain sight. Consider again the circumstances prevailing at the time. Economic growth for the second quarter of 2002 was a modest 2.2%, having decelerated from the previous quarter's 2.7%. Inflation, at 1.3%, was well contained. Japan, the second biggest economy in the world, was in the grip of an ongoing deflation. The United States was at war in Afghanistan and threatening to attack Iraq. Stock markets around the world were under tremendous pressure. Compared to a year ago, the S&P 500 was down 19%; Germany's DAX was off 27%; the UK's FTSE was down 17% and the Nikkei was off 18%. Not only that, but corporate accounting scandals seemed to be mushrooming, further eroding investor confidence.

The Fed had given no hint that it was about to tighten. In fact, by late June the Fed was starting to signal some nervousness about the sustainability of the recovery. But as Figure 9.4 shows, fed funds futures continued to trade at a discount until mid-July. The disconnect between the market perception and the Fed's intentions and the time it took for them to sync up can be seen by once again referring to Figure 9.4. The thick black line is the implied future fed funds rate. The size of the gap between it and the cash funds rate from early June through mid-July illustrates the disconnect between actual policy and anticipated policy changes.

Most importantly, the Fed's thinking during this time can be seen by comparing key elements of FOMC policy statements released in May, June, and November, major portions of which are quoted in Table 9.2.

The Fed acknowledges that monetary policy is accommodative. It makes clear its belief that inflation is well contained. It also references, inflation-dampening productivity gains. It allows it is concerned about slowing growth; it explicitly refers to geopolitical risks. It leaves little doubt that it is determined to keep its easy money policy intact for as long as it takes—for the foreseeable future as the Fed puts it in the press release. Nevertheless it took the futures market at least a month and a half—from early June until mid-July—to take out the discount that implied Fed tightening down the road.

TABLE 9.2

Chronology of FOMC Policy Statements

Release Date	May 7, 2002	June 26, 2002	November 6, 2002
Policy Action	The Federal Open Market Committee decided today to keep its target for the federal funds rate unchanged at 1¾ percent.	The Federal Open Market Committee decided today to keep its target for the federal funds rate unchanged at 1¾ percent.	The Federal Open Market Committee decided today to lower its target for the federal funds rate by 50 basis points to 1¼ percent.
Economic Outlook	Economic activity has been receiving considerable upward impetus from a marked swing in inventory investment. Nonetheless, the degree of the strengthening in final demand over coming quarters, an essential element in sustained economic expansion, is still uncertain.	The upward impetus from the swing in inventory investment and the growth in final demand appear to have moderated. The Committee expects . . . final demand to pick up over coming quarters, supported in part by robust underlying growth in productivity, but the degree of the strengthening remains uncertain.	An accommodative stance of monetary policy, coupled with still-robust underlying growth in productivity, is providing important ongoing support to economic activity. However, incoming economic data have tended to confirm that greater uncertainty, in part attributable to heightened geopolitical risks, is currently inhibiting spending, production, and employment. Inflation and inflation expectations remain well contained.
Forward-Looking Statement	Although the stance of monetary policy is currently accommodative, the Committee believes that, for the foreseeable future, against the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the risks are balanced with respect to the prospects for both goals.	The Committee believes that today's additional monetary easing should prove helpful as the economy works its way through this current soft spot. With this action, the Committee believes that, against the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the risks are balanced with respect to the prospects for both goals in the foreseeable future.	

Source: The Federal Reserve Board

From a trading standpoint, this is an example of a low-risk, high-reward opportunity hiding in plain sight. Fed funds futures were seriously underpriced. Market expectations on Wall Street were way out of kilter with the actual state of the world, a not altogether unusual phenomenon. There was little reason to suppose that the Fed was going to tighten policy and plenty of reasons to think it could ease further in a pinch—which it did. All the Fed had to do was nothing for implied yields in fed funds futures to drop 62 basis points, which was the convergence point with cash. Additional easing was icing on the cake. How could the futures market get this so wrong?

There are a number of possibilities. One is the echo chamber of self-reinforcing expectations. As the quoted market in fed funds futures weakened, some traders and analysts interpreted it to mean that the market thought the Fed would tighten, and in the process convinced themselves that the Fed just might, thereby adding fuel to the fire. It should also be noted that the market was comparatively thin and had a tendency to exaggerate things more than a bit. Another is that the Street just got it wrong because its perspective can be too narrow, kind of like the lady from the upper West Side of New York who could never understand how Reagan got elected. After all, she didn't know anybody who voted for him. Still another thing to consider is that virtually all the fed funds probability models assume some sort of binary decision mode; either tighten/don't tighten or ease/don't ease. This is a case where the decision was to ease, but market pricing embedded a tightening probability—a scenario the models were spectacularly ill-equipped to handle.

Here are some additional considerations. One is that unlike markets that can stay irrational longer than you can stay solvent, fed funds futures contracts have a limited shelf life. They will ultimately converges with the overnight market for bank reserves, which is the province of the Fed. This next step follows from the first. A trader's sense of the market, or his "market feel" is of limited value at best and at worst can be a positive hindrance for trading fed funds futures, at least profitably. The funds market is going to trade where the Fed makes it trade. It will continue to do so unless and until the Fed announces a change in its operating procedures. That could happen. In the meantime the best way to get a handle on the Fed is to learn the Fed's policy language and its operating procedures. Market feel is largely irrelevant and quite capable of reinforcing faulty feedback loops.

There may be times when the Fed gets it wrong and the market prices are a better guide to future policy decisions than statements by the Fed chairman and the FOMC. But don't count on it. The price for being on the wrong side of the Fed can be very high. The better bet is to wait it out, until the situation becomes clearer.

SUMMARY

The fed funds rate, although a money market instrument, is one of the most important benchmark interest rates affecting the global economy and its capital markets. It is an excellent proxy for the risk-free rate of capital asset pricing model fame; it is the dominant tool of Federal Reserve monetary policy; it affects foreign exchange rates; it is the benchmark rate at which dealers finance inventories of government bonds; and it can strongly influence consumption and investment decisions.

Futures markets in federal funds at the Chicago Board of Trade can be used to gauge market expectations of future Fed policy and to hedge financing costs or to speculate on policy. Research suggests that the forward contracts in fed funds are not as efficiently priced as the nearby ones. As a result, trades with low risk but high potential reward surface every now and then. They are worth watching out for. Unlike most other interest rates, the funds rate is an administered rate, put where it is by the Fed. Therefore, the key to profits in the funds market lies in understanding the Fed.

NOTES

[1] See "Domestic Open Market Operations During 2005," Federal Reserve Bank of New York, Markets Group, February 2006, available online at: http://www.newyorkfed.org/markets/omo/omo2005.pdf

[2] Ed Nosal, "How Well Does the Federal Funds Futures Rate Predict the Future Federal Funds Rate?" Federal Reserve Bank of Cleveland, Oct 1, 2001, available online at: http://www.allbusiness.com/personal-finance/investing-trading-futures/1016233-1.html

[3] Refet S. Gürkaynak, Brian Sack, and Eric Swanson, "Market Based Measures of Monetary Policy Expectations," Federal Reserve, August 2002. Available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id = 333640

[4] Refet S. Gürkaynak, "Using Federal Funds Futures Contracts for Monetary Policy Analysis," The Federal Reserve System, July 2005. Available online at: http://www.federalreserve.gov/pubs/fed/2005/200529/200529.pdf

CHAPTER 10

Government Bonds: A Toolbox

> There are two types of people who lose a lot of money in the bond markets: those who know everything and those who know nothing.
>
> *–Henry Kaufman*

Knowing the nuts and bolts of the bond market is critical for developing a broad range of capital market trading strategies. In fact, for several reasons the U.S. government bond market can be thought of as the foundation of the global capital markets. First, yields on very short-term government securities are driven by federal funds, the benchmark policy rate. These short-term rates are a good approximation of the risk-free rate of CAPM fame. Second, yields on long-term government securities represent the benchmark for default-free, but not risk-free, rates of return. Accordingly, riskier securities (like stocks) come with a risk premium attached to them. Third, the Fed uses government securities to manage the money supply, which means that government securities markets are inextricably intertwined with national banking systems as well as national politics. Fourth, the dollar is the world's reserve currency, and the major central banks invest huge portions of their dollar reserves in U.S. Treasuries. U.S. government securities are therefore the linchpin of both the U.S. banking system and international financial markets. For these reasons, U.S. government securities are the baseline for pricing all dollar-denominated capital market instruments and an important (if indirect) variable for pricing many nondollar assets.

The mathematician and philosopher Alfred North Whitehead once remarked that we think in generalities, but live in details; altogether not a bad way to describe trading in the bond markets. Accordingly, this chapter discusses details and intricacies of the government securities markets, the understanding of which is necessary to construct successful bond market trading strategies. The mechanics of how bonds are priced, traded, and financed constitute the tools of the trade.

Rates on short-term Treasury bills are a good proxy for the risk-free rate. There is no chance of default, the T-bill market is extraordinarily liquid, and the holding period is so short that capital risk due to adverse rate changes is negligible. Longer-term government securities are free of default risk. But they are not risk-free. Treasury securities markets are among the most liquid in the world, so bonds can be bought and sold relatively easily. But holding periods are long enough and prices volatile enough, that bondholders face the prospect of significant capital losses if rates move against them.

Government securities therefore contain the principal variables that concern the capital markets: the risk-free rate, default-free rates, price volatility, and liquidity. Consequently, government securities are the benchmark against which all other capital markets are priced, broadly speaking. This chapter takes a detailed look at the working of the government securities markets: the mechanics of how bills, notes, and bonds are priced, how prices and yields are calculated, trading conventions, market institutions, Federal Reserve intervention in the markets, and market structure. In short, this chapter presents a toolbox for implementing strategy.

TYPES OF TREASURY INSTRUMENTS

The Treasury sells three types of securities to finance federal deficits. Short-term Treasury bills are sold at a discount from par. The interest earned is the difference between the discounted price and par, which is always equal to 100. Conventional Treasury notes and bonds bear fixed-rate coupons that pay interest in semiannual installments. At maturity the last coupon is paid and the principal amount is returned at par. As in the case with Treasury bills, par is always equal to 100. Treasury inflation protected securities (TIPS) are structured differently. They bear a fixed-rate coupon, but the par value is linked to the performance of the all-urban CPI. Since coupon payments are calculated by multiplying the coupon by the par amount, semiannual interest payments on TIPS bonds vary, even though the coupon rate is fixed. And the principal amount paid back at maturity will be greater than 100 if the CPI has risen in the interim. All Treasury securities are exempt from state and local, but not federal, income taxes.

T-BILLS

Treasury bill discount yields understate true returns. Accordingly, to simplify returns comparisons, the T-bill rate can be restated as a bond equivalent yield. See Table 10.1. Discount rates on three-month and six-month T-bills are converted to bond equivalent yields. The calculations can be done easily with an

TABLE 10.1

Bond Equivalents of T-Bills

Settlement	Maturity	Discount	Price	T-Bill Yield	Bond Equivalent	Days
10/20/2005	1/19/2006	3.785%	99.043236	3.822%	3.875%	91
10/20/2005	4/20/2006	4.015%	97.970194	4.098%	4.155%	182

Excel spreadsheet, which has all the necessary price and yield calculators built in. (The financial add-in function has to be installed first).

The first column in the table is the settlement date; the second is the maturity or redemption date. The third column gives the discount rate, which is how T-bills are quoted in the marketplace. In turn the discount rate is used to derive the dollar price of the T-bill, expressed as a percent of par (fourth column). The fifth column shows the yield in money market terms. This adjustment from a discount rate to a money market yield is needed because the discount rate understates the true yield. The discount rate is subtracted from par, but the cash return is on the money invested, not the par amount. A one-year T-bill selling at a 10% discount rate, or 90 cents on the dollar, would have a true yield of $10 \div 90 = 11.1\%$.

Another factor that needs to be considered is the fact that conventional money market instruments assume a 360-day year, while Treasury securities (of all kinds) pay interest based on the actual number of days in a year; 365 in a regular year, and 366 in a leap year. Finally, for ease of comparison, Treasury bill rates can be converted into bond equivalent yields using the financial functions embedded in Microsoft's Excel spreadsheet application. The sixth column in Table 10.1 labeled, "Bond Equivalent" displays the results obtained by doing the conversion. The last column lists the number of days to maturity.

In cash terms, an investor who bought $1,000,000 par value T-bills for settlement on October 20, 2005, maturing January 19, 2006, would have to pay $1,000,000 * 0.99043336 = $990,433.36. On the January 19 maturity date, he would receive $1,000,000 back, which would imply that he earned interest of $1,000,000 – $990,433.36 = $9,567.64.

TREASURY NOTES AND BONDS

The structure of Treasury notes and bonds differs from that of Treasury bills. First, the original maturities are different. T-bills all have original redemption dates of one year or less. Treasury notes and bonds have longer maturities. Notes have maturities that extend beyond 1 year out to a maximum of

TABLE 10.2

Treasury Note Maturations

Coupons	Payment	Future value	Rate	Present value
	1	–$19,375.00	1.951%	$19,004.23
	2	–$19,375.00	1.951%	$18,640.55
	3	–$19,375.00	1.951%	$18,283.83
	4	–$19,375.00	1.951%	$17,933.94
	5	–$19,375.00	1.951%	$17,590.75
	6	–$19,375.00	1.951%	$17,254.12
	7	–$19,375.00	1.951%	$16,923.93
	8	–$19,375.00	1.951%	$16,600.07
	9	–$19,375.00	1.951%	$16,282.40
	10	–$19,375.00	1.951%	$15,970.81
Principal	10	–$1,000,000.00	1.951%	$824,299.62
			Σ =	**$998,784.24**

10 years. Bonds have original maturities longer than 10 years. Other than maturity there is no substantive difference between notes and bonds.

The structure of the cash flows differs between T-bills and T-notes. Unlike bills, which are sold at discount, Treasury notes and bonds have fixed-rate coupons, payable semiannually. For instance a $1,000,000 par value five-year note with a 4.25% fixed coupon pays interest of 4.25%/2 * $1,000,000 = $21,250.00 every six months up to and including the redemption date. On redemption date the Treasury pays the last semiannual coupon and makes a principal payment of 100% of par value, in this case $1,000,000.

The price of a Treasury bill, note, or bond is simply the present value of its discounted future cash flows. Table 10.2 shows how the discounted pieces of a $1,000,000 note maturing in five years with a 3.875% coupon sum to the bond's price when the market rate of interest (the discount rate) is 3.902%.

Since the bond has a 3.875% coupon and matures in five years, it will make 10 semiannual interest payments of (0.03875/2) * $1,000,000 = $19,375.00. And it will make a principal repayment of $1,000,000 at maturity. Compounding is semiannual, so the discount rate is converted accordingly: 3.902%/2 = 1.951%. The sum of the discounted coupons and principal is $998,784.24, so the price of the bond, which is quoted as a percent of par, is $99.878424. By convention, U.S. government bonds are quoted in 32nds of a point (and fractions of 32nds), so a price quote would be displayed as 99.28, which would signify 99 and 28/32nds of a point.

The calculation can also be done backwards in Excel to produce a yield once the price, maturity, settlement date, and coupon rates are known. The calculated yield is the discount rate or yield to maturity (YTM). By convention, when bond traders refer to a bond's yield, they mean the bond's yield to maturity. Another measure not to be confused with the YTM is a bond's current yield, which is the quotient of the bond's coupon divided by the price. In the current example, the bond's current yield is 3.875 ÷ 99.875 = 3.88%.

PRICE/YIELD SENSITIVITY

Bond prices change as the discount rate changes, but the rate of change is not constant. The sensitivity of a bond's price with respect to its yield varies, depending on maturity, coupon, and the level of rates. Moreover, the price/yield relationship is not linear; instead it is convex. Hedging and trading strategies therefore depend on dynamic rather than static price/yield sensitivities. Getting a handle on the complex relationships between maturity, rate level, and coupon is critical for managing position risk and setting up successful trading strategies.

The best way to see these relationships is to use concrete examples. Accordingly, to analyze the dynamic nature of bond price/rate sensitivities, we will use a set of examples to explore three different aspects of the relationship. First we compare the sensitivities of several different bonds to rate changes. Second, we examine changes in the price sensitivity of a single bond to changes in rates. Specifically, we are interested in the bond's convexity, or how the price sensitivity of a bond changes as rates change. Third, we will analyze how the convexity varies across different bonds and why it matters.

To begin, we consider five separate government bond issues with maturities ranging from 2 years to 26 years. Unlike T-bills which are quoted in terms of their discount rates, Treasury notes and bonds are quoted as dollar prices that represent a percent of par, or 100. Using Excel, we take market prices and calculate the corresponding yields. See Table 10.3.

Having calculated one set of prices corresponding to one set of yields, we now change the yield on each bond by 1 basis point. By subtracting the second price from the first, we can see the effect of a 1-basis-point change on each bond's dollar price. In Street parlance, this calculation represents a bond's DV01, an abbreviation that stands for the dollar value of 1 basis point, at that particular level of rates. By convention, the Street expresses this relationship in terms of $1 million bonds, so it really refers to the bond's DV01 per million which is obtained by multiplying the price change by $1 million in par value.

TABLE 10.3

Note and Bond Prices and Yields

Issue	Settle	Coupon	Maturity	Price	Yield
2 year	11/15/2005	4.250%	10/31/2007	99.6250	4.451%
3 year	11/15/2005	4.375%	11/15/2008	99.7500	4.465%
5 year	11/15/2005	4.500%	11/15/2010	100.0625	4.486%
10 year	11/15/2005	4.500%	11/15/2015	99.3750	4.579%
Bond	11/15/2005	5.375%	2/15/2031	108.9688	4.760%

To illustrate the calculation process, see Table 10.4 which displays prices calculated at yields 1 basis point higher than Table 10.3. The last column of the table shows the dollar difference per bond issue per $1 million worth of bonds at par value—the DV01.

An examination of the last column in Table 10.4 makes it immediately plain that longer-dated maturities are more sensitive to changes in yield than are short-dated securities. For example, a holder of $1 million two-year notes would have a profit or loss of about $185 per million with a 1-basis-point move in rates compared to $1,550 for a holder of 26-year bonds. Similarly, a 1-basis-point rate change would hit the P&L to the tune of $277 in three-year notes, but a more substantial $792 in 10-year notes. Also note that coupon differences are inconsequential. From these data we infer that, all else equal, for the same change in rates, longer-dated bonds are more price sensitive than short-dated bonds. That, as we shall see, is a critical factor in setting up arbitrage trades.

Changes in bond price/yield relationships do not stop with differences across bonds of varying maturities. The relationship also shifts for the same bond as the level of rates changes. Moreover the rates at which

TABLE 10.4

Bond Yields and Maturities

Issue	Settle	Coupon	Maturity	Price	Yield	DV01
2 year	11/15/2005	4.250%	10/31/2007	99.6065	4.461%	$185
3 year	11/15/2005	4.375%	11/15/2008	99.7223	4.475%	$277
5 year	11/15/2005	4.500%	11/15/2010	100.0181	4.496%	$444
10 year	11/15/2005	4.500%	11/15/2015	99.2958	4.589%	$792
Bond	11/15/2005	5.375%	2/15/2031	108.8137	4.770%	$1,550

DV01 = dollar value of 1 basis point.

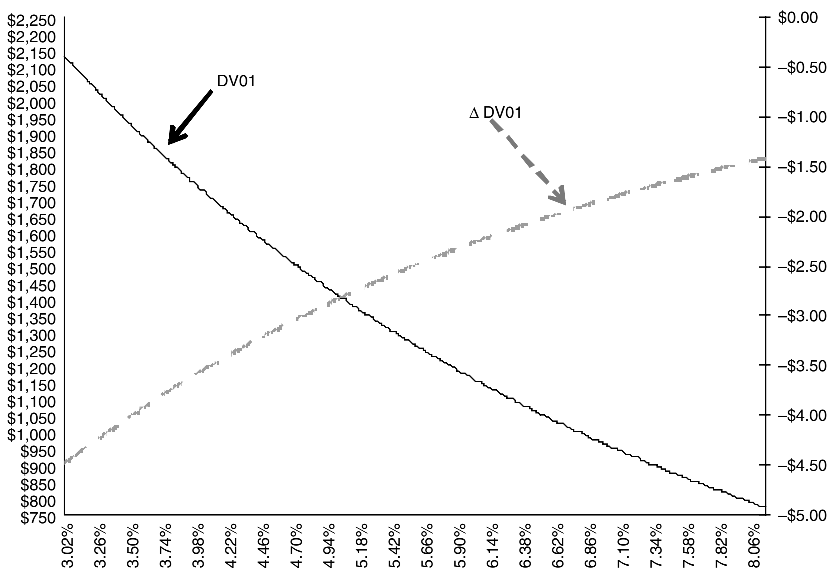
the price/yield sensitivities of bonds change are differentiated by the level of rates, by a bond's coupon, and by its maturity. To examine this, we compare price changes of short-, intermediate-, and long-dated bonds corresponding to yield shifts of several hundred basis points.

A good way to observe how a bond's price/yield relationship shifts is by graphing the changes as they occur. Using Excel, we first calculate prices of a hypothetical 5% coupon 25-year bond, beginning with an initial 3% yield and ending at 8.15%. Between those two boundary points, the price is recalculated at 1-basis-point intervals. After that, the bond's DV01 is recalculated at each yield level and graphed. The results are displayed in Figure 10.1.

The convexity of the bond can be easily observed using the graph.[1] The solid black line is the bond's DV01. The dashed line is the change in the DV01 at each rate level. The DV01 of the bond clearly changes with the level of interest rates, and it does so in a nonlinear fashion. The relationship between rates and a bond's DV01 is inverse. As interest rates go up, the bond's DV01 falls, and vice versa. The curvature of the graphs illustrates the dynamic nature of the process. As rates go higher, it takes larger price movements to change a bond's yield, which is why bond prices tend to be more volatile when rates are high.

FIGURE 10.1

Convexity Graph of a 5% Bond with a 25-Year Majority



THE IMPORTANCE OF DV01S

DV01s are the basic building blocks for implementing bond market trading strategies. The usefulness of the DV01 metric lies in the fact that it can easily be applied to all sorts of bonds, thereby allowing traders to shift risk exposure up and down the yield curve, and across many classes of credit quality, in a yield curve neutral manner.

Consider, for instance, a hedge fund manager with a long position of $100 million 10-year Treasury notes. Suppose she decides that longer-dated 26-year Treasury bonds are more attractively priced than are 10-year notes. However, she is comfortable with her level of market risk. Swapping out of 10-year notes into 26-year bonds would increase her overall risk exposure because, for a given change in interest rates, 26-year bonds are more volatile in price than are 10-year notes. How can she buy the more attractively priced 26-year bonds without increasing her desired risk exposure? The answer lies in the DV01s of the 10-year notes and long-term bonds.

Based on prior calculations, the DV01 of the 10-year notes is $792 versus $1,550 for the 26-year bonds. Since we know the DV01 of each security, we can estimate the relative price sensitivity of one to the other. Dividing the bond's DV01 by the 10-year note's DV01, we see that, all else equal, the bond's price is 1.96 times as sensitive to rate changes as is the 10-year note price: 1550/792 = 1.96. That calculation gives us the rate at which we can substitute one bond for the other while still maintaining very nearly the same level of market exposure. Accordingly, the market-neutral hedge ratio is 1.96:1. It takes 1.96 units of 10-year notes to carry the same P&L impact as one bond unit, for small changes in interest rates. The example that follows will demonstrate how this works in practice.

Assume we buy $50 million long bonds, the 5.375% of 2/15/2031 at 108.9688 to yield 4.76%. Simultaneously we sell $98 million 10-year notes, the 4.5% issue of 11/15/2105 at 99.375 to yield 4.58, picking up 18 basis points in the process. The transaction is very close to market neutral expressed in terms of DV01 equivalents: 50 bonds*$1.96_{\text{DV01 ratio}}$ = 98 ten-year notes. The efficacy of this transaction as a market neutral hedge can be tested by calculating a P&L under the assumption of a parallel 10-basis-point shift in the yield curve, both up and down. The results of these tests can be seen in Table 10.5. (We also make the assumption of no transaction or financing costs.)

As Table 10.5 illustrates, a 10-basis-point drop in rates causes the 10-year notes to go up in price by 0.7960 points, for an opportunity cost of about $78,009 on the $98 million par value notes that were sold. On the other hand, the $50 million par value bonds rose in price by 1.5676 points for a profit of $78,314—essentially a wash. Approximately the same

TABLE 10.5

Buy 50 Bonds; Sell 98 Tens

Coupon	Maturity	Price	Yield	Δ Price	P&L
4.500%	11/15/2015	100.1710	4.479%	0.7960	–$78,009.25
5.375%	2/15/2031	110.5364	4.660%	1.5676	$78,381.41
4.500%	11/15/2015	99.3750	4.579%		
5.375%	2/15/2031	108.9688	4.760%		
4.500%	11/15/2015	98.5865	4.679%	–0.7885	$77,268.46
5.375%	2/15/2031	107.4323	4.860%	–1.5365	–$76,823.11

result obtains on the downside. A rise in rates causes the 10-year notes to fall in price by 0.7885 points, or $77,268 on a $98 million par value position. But that is almost exactly offset by a change of 1.5265 points on the long bonds, which in P&L terms amounts to $76,823 on a par value position of $50 million. Since the P&L of the swap comes out very close to even with a parallel 10-basis-point shift in the yield curve, the hedge ratios based on the calculated DV01s can be considered validated within a 10-basis-point range.

The use of DV01s to establish hedge ratios does not stop at swaps involving only two bonds. The technique is easily extendable for evaluating multiple combinations of bonds. The first step is to construct a matrix of hedge ratios by dividing the DV01s of all the relevant bonds into each other. That way they can each be evaluated against all the others at a glance. Table 10.6 is an example of such a matrix of hedge ratios constructed for the notes and bonds we have been using thus far.

The table can be read across or down. For instance, reading across it can be seen that it takes only 0.67 three-year notes to equal a single two-year note; or 0.42 five-year notes to equal a single two-year note.

TABLE 10.6

Hedge Ratios

	2 Year	3 Year	5 Year	10 Year	Bond
2 year	1.00	0.67	0.42	0.23	0.12
3 year	1.50	1.00	0.63	0.35	0.18
5 year	2.40	1.60	1.00	0.56	0.29
10 year	4.28	2.85	1.79	1.00	0.51
Bond	8.37	5.59	3.50	1.96	1.00

The columns are simply the inverse of the rows. It takes 1.5 two-year notes to equal a single three-year note, and so on.

This example sheds light on some additional issues to consider in setting up bond trades. First, calculating a test P&L beforehand is a good way to validate the hedge ratios under consideration. Second, it pays to run the numbers up and down using incremental changes in yield to see what the effect is at different levels of the market. DV01 levels will change over time and at different rate levels. Position adjustments may become necessary. In fact, depending on how the position is structured, dynamic hedging may be an additional source of profit potential. Third, it is important to note that price changes were asymmetric for both the 10-year notes and the long bonds. A 10-basis-point drop in rates caused the bonds to rise by 1.5676 points, but a 10-basis-point rate increase pushed prices down by the smaller increment of 1.5365 points. That is the result of positive convexity—prices rising faster than they fall for a given change in rates.

Observe too that the effect of convexity is more pronounced in the bonds than in the 10-year notes. In the 10-year notes prices rose 0.796 points and fell 0.7885 points. That compares to 1.5676 points on the upside versus 1.5365 points on the downside for the bonds. The convexity gain is greater in the bonds than in the 10-year notes (0.0312 points versus 0.0076 points).

FEDERAL FUNDS AND THE REPO MARKET

The market for repurchase agreements, colloquially known "repo," or "RP," is another key building block for constructing bond trading strategies. The RP market is extraordinarily important for a host of reasons, both theoretical and practical. The RP rate is the rate at which dealers can borrow money using government securities as collateral. Since the Fed uses the RP and reverse RP markets to set the funds rate, the RP rate is almost always very close to the overnight rates for fed funds and short-term T-bills. The RP rate is a close approximation of the CAPM risk-free rate. To the extent that the risk-free portion of returns can be isolated and managed (via the RP and fed funds futures markets), more energy can be spent managing the risky portion of returns.

At a more practical level, the RP market matters a lot because it is the primary transmission mechanism for the conduct of monetary policy, an enormously important factor influencing market behavior. In addition, since it is the market in which dealers obtain financing to carry their positions, it strongly influences their willingness to carry risk positions and provide liquidity to the market.

In a typical repo transaction a government securities dealer borrows money for very short time periods (usually overnight) using Treasury securities to collateralize the transaction. For all intents and purposes these are short-term loans, but they are structured as "matched buy-sell transactions." The underlying theory is that a matched buy-sell takes precedence in a credit default proceeding. As a result, the RP market can continue to function without seizing up when periodic financial debacles hit the system.

Overnight RP rates move in lockstep with, and are almost always very nearly identical to, the federal funds rate. That is because the Fed controls the funds rate by using RP transactions to peg the price. If the funds rate is higher than the Fed would like, the trading desk of the New York Fed enters the market to inject money into the system, forcing the rate down. Conversely, if the funds rate trades below the target, the Fed enters the market to drain reserves, forcing the rate back up. In each case, government securities are used as the collateral for the underlying transaction.

For instance, suppose the funds rate is trading in the open market at 4.65%—above the Fed's assumed target rate of 4.5%. The Fed's New York trading desk would enter the market to execute RP agreements with the primary dealers by purchasing government securities overnight. After assessing market conditions and their own financing needs, primary dealers would bid to borrow the proffered overnight money, with all borrowings collateralized with government securities. To drive the funds rate down, the Fed would fill all borrowing demand down to (an annualized) rate of 4.5%, provided that the borrowers could produce government securities as collateral to secure the transactions.

The Fed's choice of operating technique—in this case pegging the funds rate—is important for a number of reasons. At a very mundane level, it has the effect of forcing the convergence of general RP rates with the funds rate. More importantly, it means that the Fed has decided to manage monetary policy on the demand side rather than the supply side. This has important implications for the markets. One is that it probably reduces short-term rate volatility. But that comes at a price: Policy makers lose information they would have otherwise received from price signals in the funds market. Another consideration is that, taken in isolation, pegging the funds rate can amount to a procyclical policy stance. This stands in contrast to the Fed's professed predilection for "leaning against the wind." Finally it requires policy makers to possess extraordinary agility in the formulation and execution of policy if they are to avoid the disasters of the past.

In choosing to peg the funds rate, the Fed is implicitly choosing to manage the demand side of the equation. The Fed can control either the price of money or its supply, but it can't control both. In choosing to peg the price of money via the funds rate, the Fed has made the supply of money perfectly elastic. The Fed has implicitly decided to supply all takers at the targeted rate. Conversely, it has also implicitly decided to set a rate floor under which lending and borrowing will not take place.

This can easily turn into a procyclical policy regime: Money demand is most likely to pick up when the economy is strong and vice versa. A procyclical Fed would find itself in the position of exacerbating rather than mitigating cyclical economic activity at what would otherwise be turning points. The Fed is keenly aware of this danger, which it has chosen to address by communicating its forward-looking policy intentions to the market. Arguably, this makes the expected fed funds rate just as important, if not more important, than the actual funds rate.

MARKET STRUCTURE: THE ROLE OF PRIMARY DEALERS

The overwhelming majority of Treasuries trade in the over-the-counter (OTC) markets maintained by large commercial and investment banks. These are firms that have been designated by the Federal Reserve as primary dealers, meaning that the Fed's trading desk accepts them as counterparties for transactions. The Fed has rigorous criteria for selecting firms as counterparties. First, dealers are supposed to make "reasonably good" markets to establish and maintain their trading relationships with the Fed's trading desk. Second, dealers are supposed to "participate meaningfully" in the auction process for new securities, which means roughly that the dealers are supposed to buy bonds in auctions in order to redistribute them to their customers. Third, dealers are supposed to provide the Fed's trading desk with market information and analysis that might be useful to the Fed in the conduct of monetary policy. As a practical matter, however, dealers tend to talk their position when giving the Fed advice, which, of course, the Fed knows. And during these conversations, the dealers try to get a hint of what the Fed is thinking, beyond what is in the press—to no avail.

Dealers provide tremendous liquidity to the market, trading both for their own accounts and on behalf of their customers. On the average day well over $500 billion worth of governments change hands, according to Federal Reserve reports. About 60% of the trades is by dealers trading for their own accounts through interdealer brokers; the remaining 40% is business that dealers execute with customers. The number of dealers has waxed and waned with the vagaries of the business. In 2006 there were 23 dealers reporting to the Fed. At one time in the late 1980s there were over 50.

The top five to seven dealers do the lion's share of the business—between 45% and 50% according to data released by the Fed. See Table 10.7. As a consequence of seeing all that business, the top dealers are

TABLE 10.7

Dealer Market Share Q3, 2005

	Quintile 1 5 Dealers	Quintile 2 5 Dealers	Quintile 3 4 Dealers	Quintile 4 4 Dealers	Quintile 5 4 Dealers
Bills	48%	26%	15%	7%	4%
0–3 Years	48%	27%	15%	8%	1%
3–6 Years	46%	28%	17%	9%	1%
6–11 Years	45%	29%	17%	8%	1%
11 + Years	55%	25%	14%	5%	<1%

New York Fed

likely to possess a significant informational advantage both with respect to money flows into and out of the marketplace and to market psychology. It is an advantage that can be particularly important during auctions of Treasury securities, which is the way the Treasury raises cash to fund deficit spending.

TREASURY AUCTIONS

The Treasury issues bills, notes, and bonds at competitively priced public auctions to raise cash to fund budget deficits. The Federal Reserve acts as the Treasury's agent in the auction process. There are two avenues for participation. Potential buyers can submit bids directly to the Treasury through its Treasury Direct Web site. Or they can submit bids through the commercial book-entry system, which essentially entails going through a dealer or depository institution with an account at the Fed. Bids can be noncompetitive, meaning that the buyer accepts whatever the final auction price is. Or they can be competitive, meaning that the bids have price limits attached to them.

In general, buyers going through the Treasury Direct Web site have to have their bids placed by 12 noon on auction day, although the time may vary depending on when an auction is scheduled. Noncompetitive tenders may be placed for up to $5 million notes, or $1 million worth of T-bills. Placing a noncompetitive bid precludes the buyer from submitting competitive bids for additional securities. In contrast to the Treasury Direct system, depository institutions and Treasury dealers in the commercial book-entry system (essentially an account at the Fed) can submit competitive auction bids electronically right up until the auction deadline at 1 o'clock. In neither case is anyone permitted to purchase over 35% of the bonds offered for sale.

Auction bidding rules produce several important market effects. First, they allow dealers to fine-tune their bids to reflect up-to-the-minute market conditions. Since the auctions are conducted electronically, the results are typically announced within a few minutes after the bids have been submitted. Last-minute bidding coupled with quick announcement of results greatly reduces uncertainty, which leads to more aggressive auction bidding, smaller underwriting spreads, and lower interest costs for the Treasury. Second, large nondealers who wish to fine-tune their bids at the last minute are forced to go to dealers, who are permitted to submit bids on behalf of their customers. This, in the aggregate, provides dealers with a significant information advantage about money flows in the marketplace. For instance, dealers may decide to tag along and bid aggressively when they see large customer interest developing in an auction. On the other hand, they may be inclined to hold back if they see only lackluster interest coming in from the buy side.

GOING DUTCH

The idea of an auction is to clear the market at the efficient price, which is defined as the price where buyers have revealed the highest price they are willing to pay, and sellers the lowest price at which they are willing to sell. The problem is how to get players to reveal their best prices with a minimum of haggling. Haggling over prices is ideally to be minimized because it is a transaction cost. But buyers and sellers each have strong incentives to disguise their best price. Buyers will be inclined to bid low, while sellers naturally tend to offer high, each trying to bluff the other. In the process a lot of false signals are created.

The U.S. Treasury decided to get around this problem by adopting the Dutch auction procedure. In a Dutch auction the bidders (buyers) each specify a bid price (in this case expressed as a yield) and the quantity desired at that yield. Multiple bids at differing prices/yields are acceptable. The seller, which in this case is the U.S. Treasury, ranks all the bids in descending order of price until the quantity bid for is equal to the quantity offered for sale. All successful bidders are then awarded bonds at the lowest winning bid price. Bidders who tie at the low price are awarded bonds on a pro rata basis; successful bidders at higher prices are awarded all the bonds they bid for.

Consider an example of a hypothetical Dutch auction. Assume that the Treasury has $12 billion par value 10-year notes for sale. Suppose that potential buyers place their auction bids at yields ranging from 4.48% to 4.53%. Further suppose that the total quantity of bids is equal to $21.5 billion par value, with $1 billion spoken for by noncompetitive bidders. That leaves $11 billion to be awarded by competitive bidding, although the amount applied

TABLE 10.8

Dutch Auction (All Bonds Awarded at 4.53%)

Bid Yield	Bid Price Equivalent	Bid Quantity	Cumulative Bid	Award
4.48%	100 5/32	500	500	500
4.49%	100 3/32	1,000	1,500	1,000
4.50%	100	2,000	3,500	2,000
4.51%	99 29/32	3,000	6,500	3,000
4.52%	99 27/32	4,000	10,500	4,000
4.53%	**99 24/32**	**5,000**	**15,500**	**500**
4.54%	99 22/32	6,000	21,500	0

for noncompetitively is not known at auction time. Competitive bidders need to estimate the size of the noncompetitive bid as they assess overall demand.

To see how auction bonds from the example would be awarded, examine Table 10.8. The highest yield or "stop-out rate" for the auction is 4.53%. As a result, all bidders who tendered at yields less than 4.53% receive the full quantity they bid for. They are awarded the bonds at a price to yield 4.53%—the stop-out rate. Of the $21.5 billion worth in competitive bids, $10.5 billion were tendered below 4.53%, leaving $500 million to be distributed on a pro rata basis to the 4.53% bidders. At the stop-out rate of 4.53%, the quantity bid for is $15.5 billion worth of bonds. Those bidders are awarded bonds on a pro rata basis. They receive an allocation of 3.22% (500 ÷ 15,500) of the amount they bid for.

In theory Dutch (single-price) auctions create an incentive for buyers to reveal their best prices to the seller. In Treasury auctions conducted this way, the buyers compete with each other to buy a fixed supply of bonds at a predetermined point in time. Since the supply is fixed and dealers are expected to buy at auctions to keep their primary dealer designations in good standing, there is a tendency for dealers to leapfrog over each other, raising their bid prices to ensure buying bonds. In so doing dealers perform an auction underwriting function. They typically buy into position about 50% to 60% of the bonds for sale, hoping to sell them at higher prices in the secondary or resale market.

The underwriting function that dealers perform is critical to the market, a fact that is not lost on the Treasury. Before the modern dealer system was created, the Federal Reserve was obligated to underwrite the Treasury's debt issuance, an arrangement that had several unfortunate effects. One is that it undermined the independence of the Fed in its conduct of monetary policy. Another is that there was little incentive for anyone to create an infrastructure to efficiently distribute Treasury securities.

Making the Fed the Treasury's agent instead of its underwriter necessitated the creation of an alternate system to underwrite the government's financing requirements. What emerged is the modern dealer system in which dealers underwrite auctions in the primary market and provide liquidity as market makers in the secondary market. The system has proved to be so successful that it is emulated around the world. It grew along with the Treasury's financing needs and spawned the huge global distribution network through which dealers market Treasury securities around the world.

It is hard to exaggerate the importance of the auction underwriting process to the orderly functioning of the marketplace, if for no other reason than the Treasury's massive borrowing requirements. Between January 2000 and October 2005, the Treasury sold about $3.5 trillion in notes and bonds in 185 separate auctions. That is about one auction every seven trading days. The dealer community bought just under $1.9 trillion of the bonds auctioned, about 54% of the total.[2]

The dealer community attempts to profit from auctions in two ways: first, by earning an underwriting spread, and second by adjusting the trading positions as new market information is revealed in the auction process. There are three techniques dealers commonly employ to capture auction underwriting spreads. The first is to sell the soon-to-be-issued bonds short in order to cover the short sale in the auction. The second is to sell other bonds short and then lock in an arbitrage profit by buying the new bonds at a discount in the auction. The third is to sell futures contracts short immediately after purchasing new bonds in the auction. (This third technique will be discussed in a later chapter dealing with bond futures).

The underlying idea behind all these strategies is that new issues need to come to market at a concession (however slight) in order to entice investors to buy. When market demand appears weak, the concession is likely to be large. Conversely, when demand is strong, the discount is liable to be small. In some cases, new issues may garner a premium when demand is strong. Dealer underwriting strategies therefore hinge on first assessing the potential for a new-issue discount and then devising ways of capturing it, balancing the probability of a discount against its magnitude and potential profitability. This is part of the process of price discovery that plays out in the preauction market.

Before a note or bond auction even takes place, there is an active forward market in the yet-to-be-issued bond that is scheduled to be auctioned in the near future. The trading takes place on a when-issued (WI) basis. Dealers quote two-sided trading markets (in yield terms) both directly to their customers and to interdealer brokers. Dealers' customers benefit by locking in yields and eliminating the uncertainties inherent in auction participation. By quoting these WI markets, dealers are able to gauge likely demand for the new issue, both outright for cash and for swap against other outstanding issues. The strength of demand in the WI markets aids dealers in assessing whether the new bonds are likely to be auctioned off at a discount. Dealers then try to cover their WI sales by bidding in the auction. The spread they earn (if any) is the underwriting spread. If they read the market incorrectly, they lose.

A second variation on this theme is to sell short existing bonds that are trading at yields that are relatively low compared to where the new bonds are considered likely to be auctioned. The idea is that the discount between the new bonds and the old ones will be eliminated as the new bonds are distributed to new buyers and extra supply no longer overhangs the market place. When the prices of the new and old bonds converge as the discount narrows, dealers sell the new bonds and cover shorts in the old ones.

It is a lot easier to talk about capturing underwriting spreads than it is to actually execute the strategy. This is at least partly due to the Treasury's decision to adopt a single-price auction format to sell new debt. In a series of studies designed to evaluate the impact of switching to the Dutch auction format from the more traditional multiple price format, the Treasury concluded that the policy switch had two effects.[3] First, it induced dealers to bid more aggressively in auctions. Second, it resulted in a wider distribution of participation in Treasury auctions. More aggressive bidding by dealers reduces underwriting spreads. So does wider participation in auctions. Wider participation reduces the funnel effect of few bidders, sets off competition to buy limited supply, and induces potential buyers to reveal the highest price they are willing to pay—to the Treasury—rather than exclusively to market makers. The result is closer convergence (and sometimes negative spreads) between WI markets and new issue auctions, although not necessarily between new issues and futures markets, a subject that will be explored in later chapters.

SUMMARY

It is important to really know the nuts and bolts of market practices and institutions so that insight can be converted into workable strategy. This chapter reviewed the different types of Treasury securities—bills, notes, bonds, and inflation-adjusted securities. It examined the mechanics of how bonds are priced, financed, and traded in the government securities markets. It discussed quoting conventions, repurchase agreements, yield calculation conventions, and the mechanics of how Treasury auctions are conducted. Most importantly, the chapter showed how to construct hedge ratios by calculating DV01s for cash Treasuries. In that sense the chapter can be thought of as a toolbox that can be used when implementing bond market trading strategies.

NOTES

[1] For a detailed discussion of convexity, see Frank J. Fabozzi, *Fixed Income Mathematics,* 3rd ed., McGraw-Hill, 1997, Chapter 15.

[2] See the U.S. Treasury Office of Debt Management at: http://www.ustreas.gov/offices/domestic-finance/debt-management/investor_class_auction.shtml

[3] Paul F. Malvey, Christina M. Archibald, and Sean T. Flynn, *Uniform-Price Auctions: Evaluation of the Treasury Experience*, 1995; and Paul F. Malvey and Christina M. Archibald, *Uniform Price Auctions: Update of the Treasury Experience*, 1998. Both studies are available online at: http://www.ustreas.gov/offices/domestic-finance/debt-management/auctions-study/

CHAPTER 11

The Carry Trade

The essence of the carry trade is borrowing short to lend long. The lending side of the equation is a long position in a fixed-coupon government bond. Borrowing comes into play via the repurchase agreement (RP) market. The fixed-coupon bonds are pledged as collateral in a repo transaction. Since RP rates float with fed funds, the net result is that the bondholder winds up holding a fixed-rate asset financed with a floating-rate loan. Since the slope of the yield curve is normally positive—meaning that long rates are typically higher than short rates—the interest earned on the fixed-rate bond normally exceeds the interest paid on the short-term floating-rate loan. In other words, the expected profit on the transaction comes from capturing the spread between long and short rates.

Most of the risk (and potential reward) in the carry trade is determined by three related factors: the time to maturity of the fixed-rate bonds, the spread between the bond yield and the RP rate, and the Fed's policy stance. Longer-dated bonds generally trade at wider carry spreads. That is to be expected. Longer maturities are more risky. Dollar for dollar, long-dated bonds are more sensitive to changes in interest rates. In addition there is greater uncertainty about what may happen over long holding periods than short ones. Inflation may rise. Fed policy could tighten to the point where the funds rate exceeds the yield of the fixed-rate bond, thereby producing net interest losses.

CARRY AND THE TWO-YEAR TREASURY NOTE

The two-year Treasury note is a commonly used vehicle for the carry trade, for several reasons. The two-year rate is highly correlated with the fed funds rate, as is its market direction. The two-year note is extraordinarily liquid as well. Large positions can be acquired, financed, and disposed of quickly and easily. Typically, over $200 billion of Treasury notes maturing under three years change hands daily. But a word of caution is in order. The liquidity of the two-year market should not give traders a false sense of security. The spread between the two-year and the funds rate is quite volatile. An awful lot of traders have fallen in battle after naively underestimating the risk of the two-year carry trade.

To analyze the carry trade, this chapter makes extensive use of historical interest rate data published by the St. Louis Fed on its Web site. There are several reasons for this, among them being the high quality, consistency, and easy accessibility of the data. Most importantly, government bond market spreads are almost always analyzed in historical terms. That is because there are no fundamental differences among the bonds themselves. As a result, the framework for analysis tends to revolve around the shape and slope of the yield curve, using historical averages as the reference for deciding whether bond prices are rich or cheap.

Normally, the yield curve has a positive slope. Longer-dated bonds have higher yields than do shorter ones. Under these circumstances, the two-year Treasury note typically yields more than the overnight RP financing rate, presenting an opportunity to capture financing profits. For example, consider the sample period beginning February 4, 1994, extending through September 22, 2006. Over the sample period of 3,159 trading days, on 77% of the days the two-year T-note yielded more than the funds rate. Carry was positive.

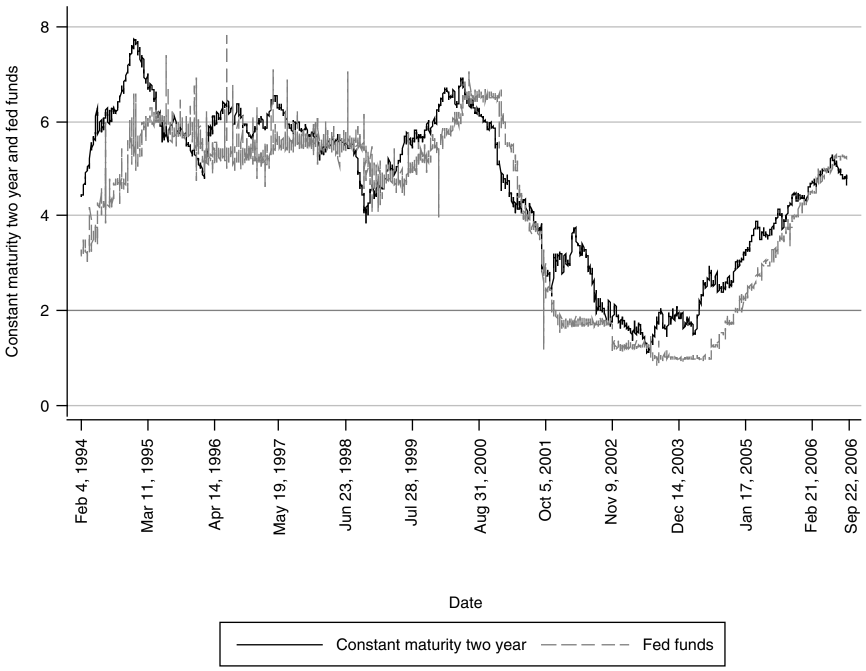
But the spread between two-year notes and fed funds varied quite a bit over the 12-plus-year sample period. The yield spread averaged a positive 49 basis points per trading day between two-year T-notes and the funds rate. But the spread ranged between a negative 184 basis points to a positive 278 basis points. It should be noted that these data reflect the Fed's estimate of a constant maturity two-year note. While that is a convenient formulation, there is no such thing as a constant maturity note. It can mask another aspect of the carry trade, namely the ability to profitably "ride down the curve," picking up incremental returns over time.

The St. Louis Fed data can be used to show how highly correlated, and extraordinarily sensitive, the two-year T-note is to the Fed's monetary policy. Figure 11.1, a graph that includes the funds rate and the constant maturity two-year rate, is a good example of this. When the funds rate rises, so do two-year rates and vice versa. The start date, February 1994, is important because it begins the era in which the Fed implemented policy changes by announcement, removing ambiguity about intentions.

The two-year note is commonly used for the carry trade because its maturity is long enough so that the yield to maturity is significantly greater than are financing rates, but not so long that the holding period

FIGURE 11.1

Two-Year Constant Maturity Rate and Fed Funds Rate



Data source: St. Louis Fed

horizon makes inflation expectations the dominant influence, as with long bonds. Most importantly, the two-year can be used to ride the Fed policy cycle. *The key to the carry trade is to establish a long position at the short end of the yield curve consistent with Fed policy early in the cycle.*

The direction of Fed policy tends to stay intact for considerable periods of time, often years, in concert with the business cycle. There are long and variable leads between shifts in monetary policy and the achievement of policy goals. As a result, the Federal Open Market Committee (FOMC) strategy tends to be executed incrementally, and it stays in place until the objectives, often explicitly stated, have been achieved. Monetary policy takes time to work its way through the system so there is typically plenty of time for traders to jump onboard in response to FOMC policy shifts.

The trending quality of the movement in policy and short-term rates is clearly visible in Figure 11.1. In this respect there are several factors that need to be taken into consideration. The first is that it is not necessary to anticipate turning points in the market to put the carry trade on. Calling market turns is dangerous business—like trying to catch a falling knife with bare hands, as the expression goes. As the graph shows, it's much easier to wait for the market to turn first and then jump onboard once that happens. And it is not necessary to guess when the interest rate policy cycle has turned. As discussed earlier, the Fed announces when it changes the direction of policy. There's not a lot of mystery to it. Based on past history, the trend will likely last for a considerable period of time.

So a reasonable strategy guideline is to wait for the Fed to change the direction of policy from tighten to ease and then go long the carry trade. Buy the two-year and finance the trade in the overnight RP market. But before going into the mechanics of this, there are a couple of other points that need mentioning. The first is that it would be very unusual for the Fed to change the policy stance from tight to easy. Instead the Fed usually reaches a policy plateau. It stops tightening at every meeting or so and then awaits further developments, generally in the form of economic data relating to economic growth and inflation. These can be thought of as neutral periods.

If, after a period of neutrality, the data begin to indicate a fall off in inflation and/or economic growth, the Fed will typically begin to hint through speeches, leaks to newspapers, and official statements that it is preparing to ease policy. Sometime shortly afterwards, the Fed will announce a reduction in the target fed funds rate. That, historically, has been the time to put on the carry trade by buying the two-year note and financing the transaction in the overnight RP market.

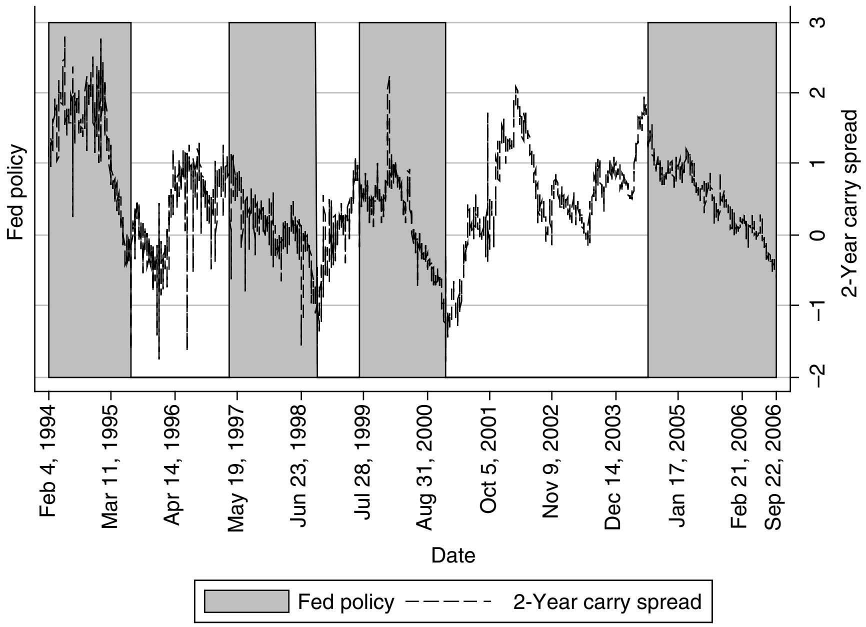
The market will usually anticipate a change in policy before it happens. At the beginning of an easing cycle the spread between the two-year note and the fed funds rate will often be negative. That is because toward the end of the tightening cycle the Fed typically raises the funds rate faster than market rates rise. In this way, market rates begin to anticipate an eventual change in policy direction. Conversely, in the easing cycle the Fed tends to push the funds rate down faster than market rates fall. Spreads widen because the market is fearful of an eventual shift toward higher rates and note buyers demand wider spreads over financing rates to extend along the curve.

The rhythm of the Fed monetary policy cycle is displayed in Figure 11.2, a graph of the spread between the two-year T-note and the funds rate over the February 1994–September 2006 sample period. Fed tightening periods are shaded gray. (For purposes of exposition, tight money times are defined as those in which the most recent Fed policy move was to raise rates; easy money periods are those in which the most recent policy adjustment was a decrease in the funds rate). As the graph illustrates, there is a marked tendency for carry spreads to tighten (and possibly go negative) during the tightening cycle. During the easing cycle, spreads widen out as the Fed drives the funds rate down faster than the market sends two-year rates down.

There are occasional exceptions to this pattern because the policy choices faced by the Fed are asymmetric. In a financial crisis the Fed will

FIGURE 11.2

2-Year T-Note Carry Spread



Shaded areas = fed tightening period

Data source: St. Louis Fed

abruptly shift to an easy posture to relieve financial stress on the system as it did in 1998 during the Asian currency crises. But it is very unlikely that the Fed would ever engage in an abrupt and precipitous turn toward policy tightening. There would be no reason for doing so.

Because of the way the market anticipates Fed policy, buying two-year T-notes at the beginning of an easing cycle often entails assuming negative carry at the outset. However, if the Fed is true to form, it will quickly push the funds rate lower, reducing borrowing costs, turning carry positive. Additionally, capital gains are likely to ensue as two-year yields begin to drop in anticipation of further rate cuts. That pattern of covariation between federal funds and two-year rates is clearly on display in Figure 11.2.

THE MECHANICS OF THE CARRY TRADE

Executing the carry trade has two straightforward components. The first is buying the notes; the second is executing the repurchase agreement. Buying the notes is simple enough. It can be at a Treasury auction, through an interdealer broker, or through a recognized dealer. Nondealers can also secure credit lines from dealers who provide RP services.

Once the bonds are bought into position, the RP part of the transaction is executed as a matched buy-sell. The bond owner borrows money by agreeing to pledge his bonds as collateral. The interest rate for the transaction is the RP rate. The first counterparty has executed an RP transaction; the other side has done precisely the opposite transaction, so that side is called a *reverse RP.* Even though the transaction is executed as a matched buy-sell, ownership of the bonds never changes; the bonds serve only to collateralize the transaction. The entire transaction can be viewed as flows of fixed and floating interest payments. The floating payments are based on the RP rate which reflects demand and supply of short-term money, bounded by Fed policy. The fixed payments are the ones that accrue from the bond's fixed coupon.

CALCULATING CARRY

A good way to see how carry trades work out is with an example in which interest receipts and disbursements are estimated for a defined time period. RP costs can only be estimated because the daily RP rate is likely to vary somewhat over the life of the transaction. In addition, the collateral is marked-to-market, usually daily. As the value of the collateral moves up and down, the principal amount of the loan varies as do its associated interest costs. Compounding is daily as well. On the fixed-coupon side of the transaction, cash flows remain constant.

To avoid needless complications, we will make some simplifying assumptions to create a workable example. We assume that the RP rate stays constant and proceed on the basis of the original marks-to-market, without daily updates or daily compounding. In this slightly simplified example, for settlement date October 2, 2006, we buy at par $100 million 5% coupon notes originally issued August 30, 2006, and maturing September 30, 2008. The notes are put out on RP daily at 4% until October 30, 2006.

Since the notes were issued on August 30, 2006, they settle with accrued interest on October 2, 2006. The exact amount ($0.45558 per $1,000 bond) can easily be calculated using Excel's Add-in financial functions. The same set of calculations for settlement October 30, 2006, will produce the accrued interest for the 29-day holding period ($0.84232 per $1,000 bond). The difference is the accrued due the note holder.

Once the principal plus accrued on settlement date is known for the notes (100.45558), carry costs can be estimated for the 29-day holding period. The daily RP rate is 4%, and by convention RP calculations assume

TABLE 11.1

Carry Calculations

Settle	Coupon	Maturity	Price	Accrued	Money Total
10/2/2006	5%	9/30/2008	100	0.45558	100.45558
10/30/2006	5%	9/30/2008	100	0.84232	100.84232
					0.38674
	RP Rate	No. of Days	Principal	Interest	Net
10/30/2006	4%	29	100.45558	0.32369	0.06305

a 360-day year. Carry costs are therefore estimated: 100.45558 × .04 ÷ 360 × 29 = 0.32369 per $1,000 bond. Net carry for $100 million notes would be the difference between interest accruals and RP expenses, or (0.84232 – 0.45558 – 0.32369) × 10,000 = $63,050. These calculations are displayed in Table 11.1.

Note that despite the fact that the bond's stated coupon rate is 5% versus a 4% RP rate, the bondholder did not receive a carry spread of 100 basis points. The 4% borrowing rate is calculated on a principal amount of $100,455,800, but the 5% coupon accrues on the face amount of $100 million. Also note that as interest continues to accrue on the bonds, borrowing costs rise because the borrowing rate is multiplied by the market value of the bonds plus accrued interest. This is how borrowing costs compound. Similarly, RP collateral is marked-to-market daily, so borrowing requirements either increase or decrease over time as the money value of the transaction changes with market levels.

RIDING DOWN THE CURVE

There are actually two components to the carry trade. The first, which we have just discussed, is really a form of market timing. The attempt to capture the spread between borrowing short and lending long is, at a minimum, a bet that short-term rates will not rise, or at least not rise enough, so that borrowing costs exceed lending costs over the term of the transaction. Ideally, the idea is to put this type of trade on when short-term rates are expected to fall. A policy-induced cut in short rates would reduce borrowing rates and would normally be expected to cause a fall in rates throughout the market, causing bond prices to rise. But even without a drop in borrowing rates a positively shaped yield curve—one in which long rates are higher than short rates—holds out the possibility of capital gains even without a change in the general level of market interest rates.

TABLE 11.2

Riding the Curve

Settle	Coupon	Maturity	Yield	Price	Change
10/2/2006	5.00%	9/30/2008	5.00%	99.9997	
10/30/2006	5.00%	9/30/2008	4.90%	100.1768	0.1771

To see how this works we continue with the example using a hypothetical 5% coupon T-note in the two-year area of the yield curve. Assume that the shape of the yield curve is such that there is a 10-basis-point spread between a 23-month T-note and a 24-month one. In the event that interest rates remain exactly unchanged over the next month, what happens to the 5% coupon two-year note?

In one month the current 2-year note will be a 23-month note. And with no change in market interest rates, it can be expected to trade to yield 4.9%, because that is the market rate for 23-month Treasury paper. A 10-basis-point drop in yield would cause a price rise in the 5% 23-month note of 0.1771 points, as shown in Table 11.2.

Note that the profit on this trade did not require a change in the market level of rates. As long as market rates remained unchanged, the trade generates a profit. The profit comes from two sources. The first is the interest spread between the cost of financing and interest accruals on the Treasury's coupon. The second is the price gain on the notes which derived from riding down the curve, not from a change in the market level of rates. In effect this transaction simply reflected a profitable mismatch between a fixed-rate asset (the notes) and a floating-rate liability (a 30-day loan taken out at the daily RP rate).

Note too that at trade initiation the counterparties agree to swap the bonds back and forth at the same price plus an interest charge over an agreed-upon period of time. Interim marks-to-market are conducted for margining and risk management purposes as well as for calculating interest charges on accurate money balances. Any change in the price of the bonds resulting from a change in interest rates is borne entirely by the holder of the bonds. Another factor to note is the pull to par, which is much more noticeable with short-dated bonds. If two-year notes are used in the RP and they are trading at a substantial premium (or discount) to par at the time of trade initiation, over a period of 30 days or so the premium (or discount) will be cut even if yields remain unchanged.

Net positive carry can be attained on the short side of the market as well as the long side during those times that the yield curve is negatively sloped. From a structural standpoint, however, it is more difficult to set up trades designed to capture carry spreads on the short side of the market. There are two principal reasons for this. First, financing a long simply requires borrowing money using a Treasury security as collateral. The transaction is easy to execute because Treasuries are the best-quality collateral in the market place. From that standpoint, different Treasury issues are, for the most part, interchangeable with each other. But this is not so on the short side of the market. To make delivery, the short seller needs to borrow the specific security he is short.

Search costs for finding the specific security need to be factored into the equation. So too does the reason for choosing the specific bond or note to sell short. Typically, shorts sell an issue that seems to be trading relatively expensively compared to others in the same maturity area. The problem is that when that happens it may very well be because a lot of people are already short that particular issue, can't buy it back, and have to pay a premium to borrow it. That is a danger signal. Not only does the carry spread begin to vanish if the notes begin to carry a borrowing premium, but there is the danger of getting caught in a short squeeze. It is important to differentiate between strategies that are issue specific and those that pertain to the general market.

Another factor that needs to be considered is the possibility of a precipitous, event-driven change in the direction of policy. While rare, such a change is far more likely to result in easing rather than tightening, for the reasons discussed earlier. That is to say that policy change risk tends to be asymmetrical, favoring longs rather than shorts. All in all, trading the short side is a bit more dangerous than trading the long side.

THE CROSS-CURRENCY CARRY TRADE

Another version of the carry trade is one in which the funding currency is not the same as the one in which the fixed-rate asset is denominated. Say for instance overnight interest rates for Japanese yen are 0.5% and three-month Treasury rates are 5%. Some traders will proceed to borrow yen at 0.5%, convert the yen to U.S. dollars in the foreign exchange (FX) markets and then invest those dollars in three-month Treasury paper at 5%. Three months later, the idea is to reverse the foreign exchange transaction after having taken out the 4.5% interest rate differential. In reality, this is really a disguised foreign exchange transaction. The difference between the spot and forward rate in FX is determined by interest rate spreads. In a world free of riskless arbitrage profits, the difference between spot and forward currency rates should erase the interest rate differential. The trader who borrows in one currency to fund positions in another is betting that the funding currency will not appreciate more than the carry spread. This type of transaction has a long history of producing spectacular failures, among the more notable being the implosion of the Mexican peso in the early 1990s and the Asian currency crisis of 1997 in which the Thai baht fell by 50%.

SUMMARY

The carry trade refers to a technique for borrowing short-term floating rate money to buy long-term fixed-rate assets. At its core it is a market-timing trade in which the bet is that the spread between borrowing and lending costs will be at least sufficient to outweigh market risk on the fixed-rate asset. Well-timed and executed carry trades hold out the possibility of profits from the carry spread as well as trading profits from riding down the yield curve. The key is to establish a long position at the front end of the curve consistent with the Fed's policy stance early in the cycle. Then just sit on the position. It needs to be emphasized that successfully riding the curve in the government market is not simply a matter of luck—it is mostly a matter of reading the Fed and its intentions correctly.

CHAPTER 12

Tactical Trading and the Yield Curve

All men can see these tactics whereby I conquer, but what none can see is the strategy out of which victory is evolved.

–Sun Tzu, The Art of War

Tactical trading along the yield curve is mainly driven by microlevel analysis of the pricing of individual bond issues. The pricing analysis mostly depends on relative valuations of various bonds rather than a view on the general level of interest rates. And with a tip of the hat to Sun Tzu, the recommended tactics to exploit issue-level pricing anomalies are embedded in a larger strategy that requires understanding the fundamentals of the bond market, monetary policy, FOMC operating techniques, and the mechanics of position financing.

INTRODUCTION: THE YIELD CURVE

The term structure of interest rates, more popularly known as the *yield curve,* is defined by the relationship between a bond's yield and its time to maturity. The behavior of the curve, its shape and slope, its value as an economic indicator, and what it is liable to do in the future dominate discussions about the bond markets. Usually the curve has a positive slope. Why that is so is a subject of considerable debate. It is of more than academic interest, though. The shape and predictability of the yield curve are key elements of bond market trading strategies.

The yield curve comes in three flavors: positive, negative, and flat. When long rates are higher than short rates, the curve is positive. Because it is usually the case, a positively sloped curve is also commonly referred to as a normal curve. But occasionally the curve takes on a flat or negative slope. The curve is typically depicted graphically as a scatter plot of yield versus time to maturity with yields displayed on the *Y,* or vertical axis, and time to maturity displayed on the *X,* or horizontal axis. Since governments are default-free, they are the baseline for defining the shape and slope of the curve. All other fixed-rate assets trade at a spread to governments of the same maturity, including corporate, federal agency, tax-advantaged, and "junk" credits. For all intents and purposes, the government market *is* the curve.

The shape of the yield curve has enormous significance because it influences behavior. Corporations decide whether to issue commercial paper or longer-term notes based in part on analyzing how the curve affects their cost of funds. Consumers similarly decide whether to use 15-year, 30-year, or floating-rate mortgage loans to finance housing purchases based in part on the relative attractiveness of the rates. The shape of the curve influences the maturities in which bond portfolio managers choose to invest. Moreover, the yield curve is anything but static. It moves around quite a bit, creating trading opportunities in the process. Why does the curve act the way it does?

THE TERM STRUCTURE OF INTEREST RATES: A MODEL

The term structure of interest rates—the yield curve—is derived empirically from observing the relationship between bond yields and their maturities. Douglas A. Ruby developed the term *structure model*, in which the market yield of an individual bond is a function of the required rate of return plus expected inflation, a credit risk premium, and volatility.[1] Formally the model is:

$$i_{market} = \{r + Exp[\pi_t]\} + \rho + \lambda$$

where:
i_{market} = the market rate of interest for a particular bond
r is the required rate of return
$\text{Exp}[\pi_t]$ is the expected inflation rate over the life of the bond at time $_{(t)}$
ρ is the credit risk premium; for government bonds it is assumed to be zero
λ represents a volatility premium that compensates buyers for the risk of holding long-dated securities whose price variability, all else equal, is a function of time to maturity

For government bonds the model reduces to time to maturity as the cause of marginal returns differences. There is no possibility of default with government bonds so the credit risk parameter (ρ) is zero. The required return for all government bonds is the same as is expected inflation for identical periods. The only remaining factor is the volatility premium λ, which is a function of time to maturity.

Numerous theories have been advanced to explain the term structure of interest rates. Three in particular have received a good bit of attention. The first hypothesizes that the fixed-income market is segmented and has heterogeneous buyers. Yield differences across sectors reflect the fact that different segments are dominated by different types of buyers with different sets of preferences. The second maintains that government bonds are perfect substitutes, that the shape of the curve reflects all known information, and that the curve changes only as new (and unexpected) information comes into the market. The third theory (preferred habitat) is a synthesis of the first two. It argues that buyers prefer short maturities but can be induced to invest in longer maturities if a sufficient premium is offered.

MARKET SEGMENTATION

Market segmentation theory may well reflect institutional and regulatory features of the marketplace. As an example, consider the format the Fed uses to report trading volume in the government securities market. The Fed reported volume and position statistics into five maturity categories: bills, 0- to 3-year Treasuries; 3- to 5-year Treasuries; 5- to 10-year Treasuries, and Treasuries maturing in over 10 years. The format reflects the way those maturity areas are regarded by the industry. Dealer firms typically block off market-making responsibilities by maturity area, suggesting differences with respect to market behavior and investor preferences across the various maturity groupings.

Investor preferences can be inferred using securities holdings data and auction participation rates.[2] These data indicate that maturity preferences do indeed vary by investor type. Individuals are more likely to prefer shorter maturities. Certain types of institutional investors like insurance companies and state pension funds are more likely to be disposed to buy bonds at the long end of the market to pair up against long-term liabilities. Auction participation data also tend to support the idea that maturity preferences vary by investor type. Auction data published by the Treasury for the period January 18, 2000, through November 30, 2005, show that individual investors accounted for about 4% of all two-year note auction purchases, but only about 1% of 10-year note orders. Investment funds displayed a preference for longer-dated securities during the same sample period, accounting for 13% of 10-year purchases, but only 7% of two-year year auction orders. However, these data may simply reflect execution preferences, since the vast majority of bond investment fund holders are individuals. By far, the primary dealers are the biggest players in Treasury auctions. Dealers account for 49%, 59%, and 63% of auction bids for 2-year, 5-year, and 10-year note bids, respectively, for the sample period.

While the distribution of auction participation rates by customer type is suggestive, the large participation by dealers is more intriguing. Reporting dealers make markets in government securities across the entire curve and provide trading services to all kinds of customers. Auction participation provides potential for dealers to set up yield curve arbitrage trades. Among other strategies employed, they look to profit by exploiting temporary imbalances of supply and demand of bonds that can cause kinks along the curve. Dealers can sell expensively priced bonds on the curve to be replaced by cheaper bonds purchased at Treasury auctions. It bears noting, however, that the size of the government market, which on average trades well over $500 billion worth of Treasuries a day, makes obvious and persistent mispricings the exception rather than the rule. When mispricings occur, they tend to be issue specific rather than across an entire market segment.

THE EXPECTATIONS HYPOTHESIS

Another explanation for the shape and slope of the yield curve is given by the expectations hypothesis, which posits that short- and long-term Treasuries are perfect substitutes for each other.[3] Consequently, market forces should drive convergence between long and short rates so that the long-term rate is equal to the weighted average of expected short-term rates over the maturity of the long-term bond. For example, if a one-year Treasury bill yields 5%, and a contemporaneous six-month T-bill yields 4.5%, the embedded expectation is that in six months, the next six-month T-bill will yield 5.5%, (with adjustments for compounding). If, on the other hand, the preferred unit of analysis is one calendar quarter, the yield of a two-year Treasury note would, in theory, be equal to the expected yield of eight Treasury notes with three-month maturities arranged sequentially (assuming zero transaction costs).

The expectations hypothesis implicitly assumes that lenders have more flexibility than do borrowers. For instance, a lender can choose to make a five-year loan or a series of shorter loans that sum to five years. He could choose to make a one-year loan and then after one year, reinvest the proceeds for four years (or any time combination). On the other hand, the borrower is presumed to be interested in matching up the maturities of assets and liabilities. If so, the inclination would be to seek a five-year loan to acquire an asset with an expected useful life of five years, much as people (in the United States) generally tend to prefer financing housing purchases with long-term fixed-rate mortgages rather than notes that float with short rates. An interesting sidelight is that mortgage borrowers in

Europe are much more likely to finance home purchases with floating-rate loans, which may indicate differences in expectations or tolerance for risk.

While it may be generally true that lenders have more flexibility than borrowers, the borrowing and debt management practices of the U.S. government are driven by policy goals (and constraints) that are quite different from the ones faced by the private sector. The private sector typically taps the capital markets to finance the acquisition of long-term assets. Part of the calculus includes borrowing costs. On the other hand, borrowing by the U.S. Treasury is determined by the difference between tax receipts and cash disbursements. In turn, the level of receipts and disbursements is a result of the business cycle as well as political decisions made by Congress and the administration. Further, spending by government is increasingly in the form of transfer payments driven by programs like Social Security, Medicare, and Medicaid; demographic time bombs that may yet exact a toll on the bond market.

While the Treasury has little, if any, influence on how much the government borrows, it substantially controls the way it borrows. In this respect, the Treasury's Office of Domestic Finance, which is responsible for managing the government's debt issuance, has one overarching policy concern. It is to "achieve the lowest possible borrowing cost, over time, for the government's financing needs".[4] In furtherance of the policy goal, the Treasury provides regular updates of its cash balances and estimated financing needs. It consults with the dealer community through the Treasury Borrowing Committee. And it regularized its financing schedule into regular cycles of issuance. Weekly auctions of 91-day Treasury bills take place every Monday. Two-year notes are auctioned monthly; five- and ten-year notes eight times a year, and long bonds quarterly beginning February 2007.

As a matter of policy, the Treasury insists that it does not try to time the market; nor does it shift maturities to take advantage of selling new issues along expensive portions of the curve. It simply supplies securities to the market consistent with its borrowing needs and lets the market set the prices. Occasionally the Treasury will reopen existing issues to ensure that sufficient floating supply is available, especially if a supply shortage has been protracted and threatens to disrupt the ordinary functioning of the market.

That is not to say that the Treasury's borrowing plans are set in stone, once and for all. Over time there have been significant shifts in issuance patterns and maturities. Some have been gradual and deliberate; some abrupt. Further, while there is no evidence that the Treasury has tried to time the market or play the curve on a day-to-day basis, it is incontrovertible that the average maturity and form of the debt has changed over time, some of it explicitly the result of policy decisions. For instance, the Clinton administration in 1993 drastically reduced the issuance of 30-year bonds for the express purpose of lowering federal borrowing costs by reducing the maturity of the debt. The Bush administration suspended the issuance of 30-year bonds entirely without any warning in October of 2001 before reinstating bond issuance in 2006. In addition, the Treasury has aggressively marketed inflation-linked bonds to reduce the inflation premium it pays in the marketplace.

The average maturity of Treasury debt turns out to be quite variable. From fiscal years 1967 through 2003, the average maturity of the debt ranged from a low of 31 months to a high of 71 months, averaging 56 months. Of particular note is the 2000–2003 period, during which time the average maturity of the debt dropped sharply, even as publicly held debt rose rapidly, thus setting the stage for increased auction frequency, larger-sized issues, and a resumption of three-year note and long bond issuance. Do Treasury issuance patterns affect the shape of the curve?

Thomas H. Huxley is reported to have quipped that nothing is quite so tragic as a beautiful theory killed by an ugly fact, and in the expectations theory of the yield curve there is an ugly fact that needs to be addressed. It is this: For long periods of time—years, not weeks or months—certain Treasury issues stood out as far too cheap on the yield curve. The specific maturities were 9-month Treasury bills, 7-year Treasury notes and 20-year Treasury bonds. For years, waves of arbitrageurs attempted to profit by buying these issues while shorting the issues around them, but to no avail. They remained cheap until the Treasury took them off the calendar and stopped selling them.

The problem can be illustrated using the example of the Treasury's ill-fated issuance of 20-year bonds, which began in June of 1981. Before the Treasury began selling new issue 20-year bonds on a regular basis, it had intermittently sold 25-year callable bonds. In addition, the Treasury started regular sales of 30-year bonds with 25 years of call protection beginning February 1977. Regular 10-year note sales began in 1976. As a result, there were enough notes and bonds floating around so that economists at Treasury could produce an interest rate time series that included 10-year notes, 20-year bonds, and 30-year bonds even before the 20-year program was formally introduced.

To gauge what effect (if any) the regular issuance of 20-year bonds had on the yield curve, a comparison can be made between the 20-year yield and the average yield of the 10-year note and 30-year bond before and then after June 1981 when regular issuance of 20-year bonds began, using St. Louis Fed data. In theory, the average yield of the 10-year and 30-year bonds should have been about the same as the 20-year yield. And from February 1977 until June 1981, it was. The difference was only about +2 basis points. But from June 1981 until the end of 1986 the average spread more than quintupled to 12 basis points, prompting the Treasury to cease issuing 20-year bonds after February of 1986. Similar stories could be told with respect to four-year notes, seven-year notes, and nine-month bills.

So much for the Treasury's insistence that it doesn't play the curve. More to the point is the clash between prediction and observation. Certain areas of the Treasury curve remained far too cheap for far too long. This led to the modification of the expectations hypothesis to what became known as the *preferred habitat theory.* It sought to reconcile the empirical problem of significant deviations in bond yields with what the expectations theory predicted.

The preferred habitat theory can be seen as an attempt to synthesize the market segmentation and the expectations hypothesis. On the one hand it is consistent with the expectations hypothesis in that it posits a term structure that reflects expectations for the future plus a risk premium. But the risk premium is not necessarily a linear function of time to maturity. Other factors can influence it, including mismatches between investor and borrower maturity preferences. Mismatches are bridged, as always, by price. Issues that are sold in areas of the curve with relatively little demand are priced at a discount in order to entice buyers away from areas of the curve they would otherwise prefer.

Long-term idiosyncratic pricing of identifiable sectors of the curve presents a problem for the expectations hypothesis. If Treasury securities are perfect substitutes for each other as the theory asserts, traders and investors would simply arbitrage away pricing anomalies along the curve by selling the expensive securities and buying the cheap ones. But the evidence is to the contrary. History suggests that perennially cheap areas of the curve remained cheap until the Treasury stopped selling new issues in those areas, reducing the available supply. On the other hand, some investors showed a willingness to alter their portfolios to buy out-of-favor areas of the curve, provided they could get a steep enough discount.

Strictly speaking, market segmentation theory doesn't leave a lot of room for investors to alter portfolio preferences, since the theory tends to stress institutional factors over price. Moreover, by partitioning the market into discrete subunits, market segmentation theory ignores the power of the Treasury to alter its borrowing patterns, thereby shifting market perceptions as well as the balance of supply and demand. Whether this amounts to the Treasury "playing the curve" or securing financing in the most efficient way possible is largely a question of semantics.

The preferred habitat theory seems to be a reasonable framework for analyzing the curve and devising trading strategies. It tips its hat to the power of institutional arrangements and provides a way to account for pricing anomalies along the curve, but it avoids the rigidities of strict segmentation theory. At the same time, the preferred habitat theory maintains the core of the expectations hypothesis. It allows new information to affect pricing across all segments of the curve; it accounts for changes in the behavior of buyers as well as sellers and leaves room for changes in behavioral patterns over time. With that in mind, preferred habitat will be the framework for analysis.

ON AND OFF THE RUN

In the Treasury market there is a distinction between the most current issue in each maturity area and older, more seasoned issues. The most current or on-the-run (OTR) issue tends to trade at a premium to older notes and bonds of roughly the same maturity. They have earned the sobriquet *on-the-run* because they are the most recently issued (and therefore most actively traded) of the Treasury's regularly scheduled sales of bills, notes, and bonds. The schedule currently consists of weekly 3-month bills, monthly 2-year notes, quarterly 3-year notes, 5-year and 10-year notes auctioned eight times a year, and quarterly 30-year bonds. Because these are the issues that trade most actively, they sometimes carry a liquidity premium. In particular, the longest bond almost invariably carries such a premium.

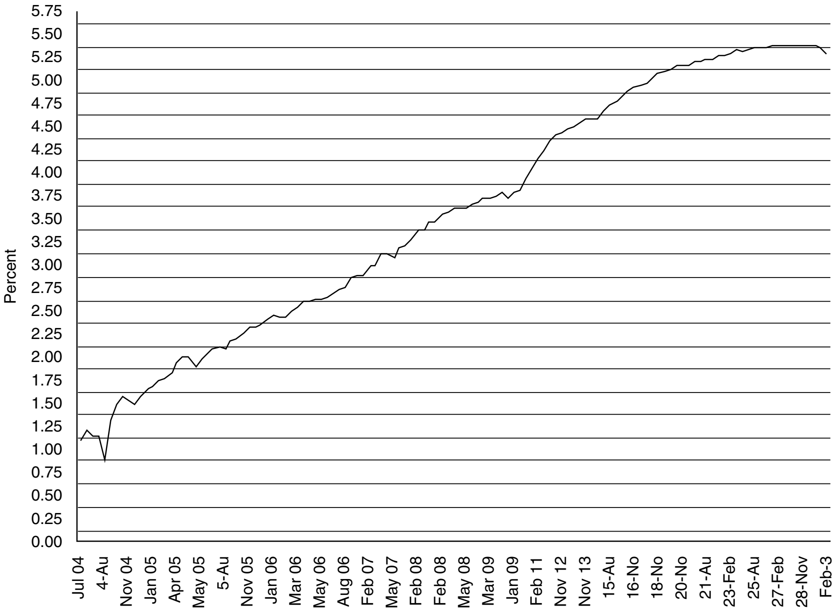
Active OTR Treasury issues serve as benchmarks for their respective market sectors. In so doing, they allow off-the-run issues to be priced relatively easily, consistent with the idea that Treasury issues are near perfect substitutes for each other. Treasury issues whose maturities are nearly identical tend to trade at nearly identical yields with small adjustments made for minor differences in maturity, and larger adjustments made for significant differences in maturity. A sketch of the curve extrapolated from the yields of the current issues may give a false sense of curve dynamics. In the real world when the curve is constructed from the ground up by plotting the yield of each individual Treasury issue along a time-to-maturity axis, the curve takes on a bit more jagged look, as in Figure 12.1.

In theory, a difference in coupon ought to result in some difference in the respective yields to maturity of two otherwise identical bond issues. Bonds with identical maturities but different coupons carry different reinvestment risks. In a rising rate environment high coupons are theoretically preferable because semiannual coupons can be reinvested at higher rates. As a practical matter, however, coupon differences have had a negligible impact on yields to maturity, largely because the Treasury allows bonds to be stripped so that the coupons and final principal payment can be discounted and traded separately.

It should also be emphasized that when Treasuries with roughly the same coupon and maturity are described as near perfect substitutes, the word *near* ought to be emphasized. There have been numerous episodes

FIGURE 12.1

The Yield Curve (June 30, 2004)



Data source: *The Wall Street Journal*

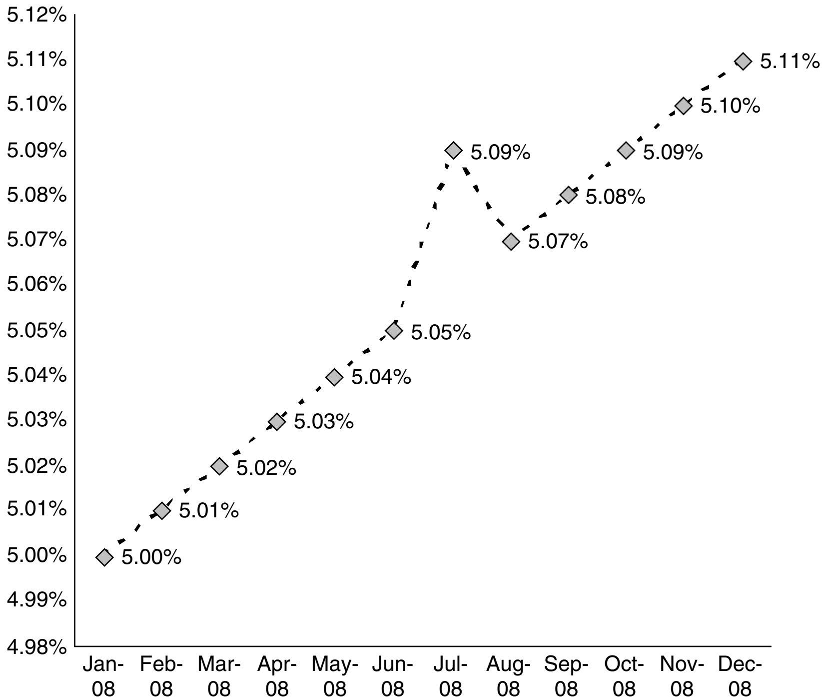
over the years in which near perfect substitutes turned out to be anything but perfect. Bonds that were apparently overpriced were sold short by so many players that they became difficult to borrow, sparking tremendous short-covering rallies with losses running into the hundreds of millions of dollars.

TRADING THE CURVE VERSUS OUTRIGHTS

Unlike the trading of “outrights,” which in essence amounts to bets on the direction of rates, curve trades are market neutral, designed to profit from changes in spreads between issues. Outright trading depends on market timing; yield curve trading can depend on either market timing or issue selection, because the shape of the curve is correlated with both the level and direction of rates. Therefore, this section will begin by examining examples of bottom-up transactions where market timing is of relatively little importance. The sample transactions (exaggerated for illustrative purposes) are designed to show the mechanics of implementing tactical yield curve trades, based on pricing anomalies.

FIGURE 12.2

Sample Yield Curve



Imagine a yield curve that looks like the one displayed in Figure 12.2. It consists of a sequence of 12 Treasury notes, each with a 5% coupon, maturing one month apart. The first in the sequence matures in two years' time; the last in three years. As the diagram illustrates, the slope of the curve is positive. Except for the note maturing July 2008, the series would be smooth and continuous with each note yielding 1 basis point more than the previous one.

The note maturing July 2008 is an obvious bargain. It yields 6 basis points more than the one immediately preceding it, while the rest of the curve seems to indicate a one-month extension is worth only 1 basis point. Presumably, holders of the June maturity ought to be interested in extending for a month to pick up a marginal 6 basis points. Better yet, holders of the August maturity might consider coming in a month to pick up 4 basis points. Let us examine each of these possible transactions.

Suppose that an owner of $10 million par value of June 2008 T-notes is willing to extend out a month to pick up a marginal 6 basis points by swapping out of the June notes and into $10 million of the July notes. Is this the free lunch that it looks like? Not exactly, but with some adjustments, it can come pretty close.

Adjustments are needed because the June and July notes are not perfect par value substitutes for each other. Because of their slightly longer maturity, the July notes are a tad more sensitive than the June notes to interest rate changes. A swap out of the June notes and into the July notes needs to take this into account. The most straightforward way of doing so is to weight a June–July T-note swap by interest rate sensitivities. The best technique for doing so is to calculate a hedge ratio using their respective DV01s, as discussed previously.

The mechanics and impact of weighting (versus not weighting) the swap are displayed in Table 12.1. At the starting point, or initial condition, the June note yields 5.05% and the July note yields 5.11%. Their respective DV01s are $224.54 and $231.52, which implies that June notes are less price sensitive to rate changes. The relative sensitivity of June to July is approximately equal to the ratio of the DV01s: DV01 ratio = 224.54/232.66 = 0.9693. This ratio implies that the proper market neutral weighting for swapping June for July is 0.9693:1.

Table 12.1 shows why. A 25-basis-point parallel rise in interest rates would cause the June and July notes to drop in price by 0.55947 and 0.57682 points, respectively. An unweighted extension swap out of June and into July would cause an additional loss of $173.46 for every $10 million swap units if rates were to immediately rise by a quarter of one point. Weighting the swap by the DV01 ratio mitigates the marginal loss. A weighted swap would entail selling $10 million June notes, buying $9.7 million July notes, and investing the $300,000

TABLE 12.1

Comparison of Weighted and Unweighted Swap Results

	Coupon	Maturity	Yield	Price	DV01	DV01 Ratio
Initial	5.00%	6/30/2008	5.05%	99.8833	224.54	
Condition	5.00%	7/31/2008	5.09%	99.7912	231.66	0.9693
	Coupon	**Maturity**	**Yield**	**Price**	**Price Change**	**P&L**
Unweighted	5.00%	6/30/2008	5.30%	99.3238	–0.55947	$5,594.71
Swap	5.00%	7/31/2008	5.34%	99.2141	–0.57715	–$5,771.49
						–$176.78
Weighted	5.00%	7/31/2008	5.34%	99.2141	–0.57715	–$5,594.17
Swap						$0.54

remainder at the risk-free T-bill rate. As the table demonstrates, weighting the swap controls for a parallel rate change; the P&L for $10 million par value June notes is essentially identical to that for $9.7 million July notes.

It is important to note what the point of the swap is—and what it is not. It is not simply a transaction designed to pick up marginal yield. If that were the point, there would be no reason to stop in July where the yield is only 5.10%; yields for notes maturing in December are even higher at 5.11%. The point of the transaction is arbitrage: to substitute June notes (which are correctly priced) for July notes, solely because the July notes are priced too cheaply along the curve. At the first cut, any of the issues on the list are swappable against the July notes, as long as the trades are weighted correctly. As more people realize that the July notes are too cheap and attempt to buy them versus selling other securities, the buying pressure will drive the July note prices higher (to a lower yield) and drive the other note prices lower (to higher yields) until the spread between July notes and other notes normalizes.

But even as presented, this June–July note swap is subject to some risk. Suppose the slope of the curve were to become more positive, or in trader parlance, to steepen? That could result in a trading loss. The spread between June and July could widen further while the rest of the curve proceeded to assume a smoother, but more pronounced, upward slope. Figure 12.3 displays a before-and-after look at this possibility. It compares the original curve in which month-to-month spreads are worth about a basis point apiece, to a new curve in which month-to-month spreads are worth anywhere from 3 to 6 basis points. Note that as the entire curve assumes a steeper upward the slope, the spread between June and July widens from 4 basis points to 6. In the process, the June–July spread becomes roughly consistent with the overall slope of the curve, so there is little reason to expect a rush of traders to narrow the spread. The result is a trading loss of 2 basis points for arbitrageurs who sold June to buy July on a yield-weighted basis.

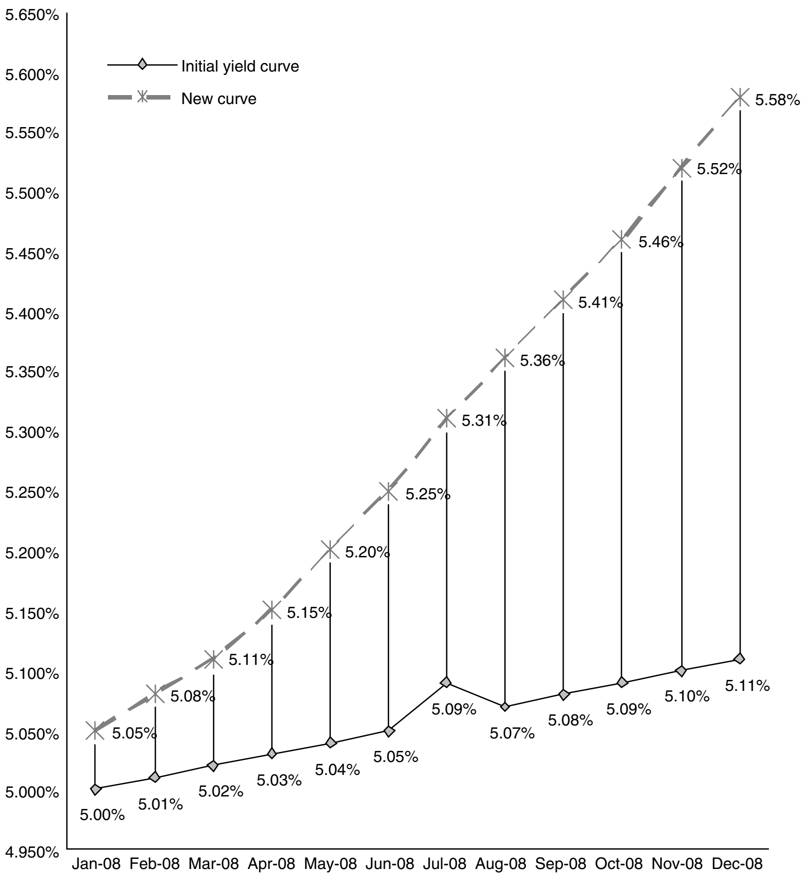
There is a way to exploit yield spread anomalies like this one while minimizing risk from a wholesale shift of the entire curve: butterfly arbitrage trades.

BUTTERFLIES: NOT FREE, BUT REASONABLE

Trading butterfly spreads, as the name implies, involves arbitraging the wings against the body. In this case the trade involves trading July T-notes (the body) against both June and August T-notes, which constitute the wings. This transaction seeks to lock in a position in underpriced July T-notes while neutralizing the impact from shifts in the overall

F I G U R E 12.3

Nonparallel Yield Curve Shift



shape of the curve. As a result, the focus of the trade becomes more centered on the relative price of the July note, and less on the slope of the curve.

Implementing a butterfly trade to exploit pricing inefficiencies leaves room for (but does not guarantee) profit regardless of a change in the slope of the curve. In the present example the idea is that July notes are priced too cheaply against both June and August notes and that the risk of buying July notes can be largely neutralized by selling weighted amounts of the wings against purchasing July notes, the body.

Weighting butterfly swaps can be trickier than it would seem at first blush. In this particular case the most reasonable way to proceed is to

TABLE 12.2

P&L from Butterfly Arbitrage

	Notes	Long (Short)	Yield	Price	DV01	DV01 Ratio
Initial Condition	June	$ (5,160)	5.05%	99.8833	$224.54	1.0317
	July	$ 10,000	5.09%	99.7912	$231.66	
	August	$ (4,855)	5.07%	99.8287	$238.58	0.9710
		Long (Short)	**Yield**	**Price**	**Price Change**	**P&L**
Butterfly Results	June	$ (5,160)	5.25%	99.4354	–0.4479	$23,111
	July	$ 10,000	5.31%	99.2831	–0.5081	-$50,811
	August	$ (4,855)	5.36%	99.1397	–0.6890	$33,453
						$5,753

weight one-half the trade against June and the other half against August. Weighting the wings evenly against the body may seem fairly obvious, but it turns out that in some of the more adventurous butterflies, even weights may not always be the best. When a butterfly's wings stretch out over years rather than months, it may sometimes be advisable to weight one wing more heavily than the other. That aside, we will proceed to set up and analyze an evenly weighted butterfly swap using the market prices and yields of June, July, and August notes, as they are displayed in Table 12.2.

Weighting the swap evenly requires hedging half the July notes against June and the other half against August. In return that implies calculating DV01s for June, July, and August notes. The respective DV01s of June, July, and August notes are $224.54, $231.66, and $238.58. Therefore, we have following weights, using the July DV01 as the numerator in each case:

$$June = \frac{231.66}{224.54} = 1.0317 = \text{June Hedge Ratio}$$

$$August = \frac{231.66}{238.58} = 0.9710 = \text{August Hedge Ratio}$$

Assume for this example that the par value of July notes to be hedged is $10 million, with $5 million on either side. The June short position hedge quantity is calculated as 1.0317*$5,000,000 = 5,158,546 par value notes. The August short position hedge quantity is calculated as 0.9710*$5,000,000 = $4,854,975 par value notes. Rounding off, the final arbitrage position is:

Short $5,160,000 par value June notes, priced at 99.8833 to yield 5.05%

Long $10 million par value July notes, priced at 99.7912 to yield 5.09%

Short, $4,855,000 par value August notes, priced at 99.8287 to yield 5.07%

Note that while the total short position is approximately $10 million par value, the par value of the June short exceeds the par value of the August short because August notes are more price sensitive to changes in yield than are June notes. Note further that the long position yield to maturity (YTM) is 5.09% while the weighted average YTM of the short position is 3 basis points lower at 5.06%.

The efficacy of the butterfly strategy can be tested by calculating a P&L on the position assuming different rate change scenarios. This particular transaction is looking for a change in rates to take the kink out the curve in which the July notes stand out as too cheap (5.09%) relative to the average combined yields of June and August notes (5.06%). In the example, after a nonparallel shift in rates, the June and August notes average 5.31%, the same as the July note. As a result, the transaction produces a net profit of $5,753 for each $10 million unit of the trade.

Referring back to Table 12.2, note the uneven distribution of the P&L. The June wing moved 20 basis points; the July body moved 22 basis points and the August wing moved 29 basis points. A result like this would not be all that unusual. The point of doing a butterfly is to take a position against a kink in the curve and wait for it to smooth out. It is a trade that is essentially indifferent to the overall shape of the curve. It just seeks to take advantage of what appears to be a temporary anomaly in the supply/demand situation.

FINANCING ARBITRAGE POSITIONS

Up to this point we have considered yield curve trades without taking position financing into account. That is unrealistic; yield curve trades typically require some workout time, which necessitates financing positions either overnight or for longer periods of time. One of the reasons for the fantastic liquidity of the U.S. Treasury market is the ability of dealers to easily borrow and lend Treasury bills, notes, and bonds to exploit perceived pricing inefficiencies. This type of arbitrage trading occurs when traders sell bonds that seem to be expensive and buy other bonds that appear to be relatively cheap. Financing considerations can be decisive.

When traders go long (buy bonds into position), they borrow the money to do it using the bonds they bought to collateralize the loan. Technically, these transactions are not collateralized loans; they are repurchase agreements (RP). The long sells his bonds accompanied by an agreement to buy them back by an agreed date (usually the next day) at a specified interest rate. The other side of the transaction executes a mirror image transaction, known colloquially as a *reverse RP.* For some reason or other the Fed for years considered the term *reverses* to be undignified and insisted on referring to reverses as *matched sales.* They finally threw in the towel after hearing RP traders yell across various trading floors that the Fed was "doing matches," which they apparently decided was even less dignified than reverses.

Anyway, the mechanics of RP transactions are important because efficient financing can be the difference between profit and loss. The margins on tactical arbitrage trades tend to be small, so dealers typically put on lots of arbitrage trades in big sizes, hoping to produce substantial profits by capturing lots of little ones. Accordingly, this section discusses the mechanics of financing arbitrage transactions, which is the building block for how all Treasury trading is financed.

In a typical arbitrage transaction a dealer buys one bond and sells a similar one against it. One way to think about all this is to consider the cash flows from all the related transactions, of which there are many. First the long accrues coupon interest on the bonds he owns. Second, he pays interest on the money he borrowed to finance the initial purchase. Third, he owes coupon interest on the bonds he is short. Fourth, he receives interest on the cash he got for selling those bonds short. Fifth, closing out the position by selling out the long and buying back the short generates a P&L on changes (if any) on the prices of the respective bonds.

An example will make this clear. Consider the following hypothetical transaction. On the long side is a $10 million par value position in Treasury 6% notes maturing May 30, 2008. For settlement date February 15, 2006, they are priced at 101.6950 to yield 5.20%. The short side position is in Treasury 5% notes maturing January 31, 2008, priced at 99.9055 to yield 5.05%. Assume that the arbitrage is put on for 28 days, the federal funds rate is 4%, the repo rate is 4.125%, and the reverse repo rate is 3.875%.

This arbitrage is directed at the yield spread between the January and May notes. The seller of the spread thinks that the curve will flatten—in other words the yield differential between January and May will compress from plus 15 basis points to some narrower spread. The bet here is not on the direction of rates so the transaction should be weighted in a market neutral fashion. That can be done by using a DV01 weighting to isolate movement in the general direction of rates from the yield spread between January and May 2008 maturities. Using DV01s, the market neutral hedge ratio is:

$$Hedge = \frac{215.48^{May}}{184.19^{January}} = 1.17$$

In other words it takes 1.17 units of January 2008 notes to hedge one unit of notes maturing May 2008. The market-neutral arbitrage position is therefore long $10 million May 2008 notes and short $11.7 million January 2008 notes. The transaction can be evaluated by calculating the interest accrued by the respective January and May notes over the 28 days the trade is on; the interest (received and paid) on the RP and reverse RP transactions; and the prices of the bonds, assuming no changes in yields at the termination date. See Table 12.3 for a summary.

Notably, the position produces a small loss of $4,331.22. About half is due to the 25-basis-point spread between the RP and reverse RP rate. The position holder borrows money at a 4.125% rate to finance the long, and effectively relends it at 3.875% to borrow bonds on the short side. The other half reflects the pull to par on both the long- and short- side arbitrage positions. The longer the trade is left in position, the more the spread has to move favorably to recoup transaction costs and reap a profit—the opposite of the carry trade.

The breakeven point for the yield spread can be determined by calculating the number of basis points the spread has to change (in the right direction) to achieve a gain equal to $4,331.22. Since the DV01 of the May notes is equal to $215.48, we can estimate the break evenpoint as:

$$B/E = \frac{\$4,331.22}{215.48 \times 10} = 2.16 \text{ basis points}$$

(The DV01 is multiplied by 10 because the position is $10 million notes and the DV01 is expressed in units of $1 million.)

An important factor to consider with respect to workout time for a trade is the difference between RP rates on general collateral and so-called RP specials. RP and reverse rates are based on the fed funds rate and reflect supply and demand conditions for overnight financing in general as well as borrowing demands for particular issues. The spread between RP and reverse RP can be thought of as the bid/ask spread between general borrowing and lending rates.

Sometimes particularly intense borrowing demand surfaces for a particular bond issue. If the demand for an issue is sufficiently powerful, it can push the RP rate significantly below the general collateral, or GC rate, as it is colloquially called. These issues trade in the RP specials market, where borrowing and lending rates reflect scarcity value. Typically, the current long bond trades in the RP specials market because there lots of shorts in them. It is fair to say that large short positions reflect a source of potentially significant market demand, which means that the RP market can produce important signals about the condition of the market. Lots of demand in the RP specials market is suggestive of

TABLE 12.3

Summary of Calculations

Position	Settle	Coupon	Maturity	Price	Yield	Accrued	Cash	
Short	2/15/2006	5%	1/31/2008	99.9055	5.05%	$24,240.33	$11,713,183.85	
Long	2/15/2006	6%	5/30/2008	101.6950	5.20%	$127,624.31	$10,297,126.47	
Position	**Settle**	**Coupon**	**Maturity**	**Price**	**Yield**	**Accrued**	**Cash**	**Principal P&L**
Short	3/15/2006	5%	1/31/2008	99.9057	5.05%	$69,488.95	$11,758,455.63	–$2,316.32
Long	3/15/2006	6%	5/30/2008	101.6406	5.20%	$174,033.15	$10,338,094.57	–$5,440.75
Position	**Financing**	**Rate**	**Days**	**Principal**	**Interest**			
Short	Reverse RP	3.875%	28	$11,713,183.85	$35,302.23			
Long	RP	4.125%	28	$10,297,126.47	–$33,036.61			
Net Interest	**Net Coupon**	**Principal**	**P&L**					
$2,265.62	$1,160.22	–$7,757.06	–$4,331.22					

significant short positions among dealers. Moreover, sustained trading in the specials market can force the price of the underlying bonds to trade at aberrant prices.

Specials RP markets for individual bonds can flash danger signs worth paying attention to. Sometimes the reason that bonds trade very expensively is that there are huge shorts in them and the bonds become especially hard to borrow. There have been classic cases in which a simple, apparently low-risk arbitrage trade turned into a nightmare costing major players hundreds of millions of dollars when they had to scramble to cover because the bonds were unborrowable. The market can be particularly vulnerable to this if a relatively small number of concentrated buyers locks up the available supply. For instance, in 1986 the 9.25% long bond maturing in 2016 was shorted so heavily that before it was all over, it traded at the same price as the 9.875%'s of November 2015—about 7 points higher than would have been normal.

The Federal Reserve, Treasury, SEC, and Justice Department take a dim view of this sort of thing, and they can take action if there is sufficient evidence that groups of traders acted in concert to restrict supply or cause a "short squeeze." But some so-called short squeezes occur naturally. In general it's a lot better to avoid these types of situations by resisting the temptation to sell bonds short that are perennially difficult to borrow.

Cases of short squeezes also point to an often overlooked strategic element integral to managing the short side of any arbitrage position. Arbitrageurs are typically indifferent to whether the position focus is on selling a bond that is expensive or buying one that is cheap. But it actually matters, and it can matter a lot. The underlying idea behind arbitrage trades is that eventually owners of expensive bonds will swap out of them for cheaper bonds, thus filling the shorts and forcing prices back to fair value. But it is far easier to market a cheap bond than it is to buy back an expensive one. Anybody can buy something that's cheap. All you need is cash. On the other hand, there are a limited number of players to sell a bond that is expensive. That's probably why it's expensive to begin with. Finding a seller means finding either an owner or someone else willing to establish a short in the bond. That can be difficult to do. Especially when every trader in the business has heard the famous ditty:

He who sells what isn't hisn't
Must buy it back,
Or go to prison.

TRADING ROLLS

An additional transaction type needs to be considered at this point: forward and reverse rolls. The *forward roll,* as it is called, is essentially an extension swap, but with a twist. An *extension swap* is one in which traders go further out along the curve, for instance selling out of a two-year security to buy a three-year one. The mechanics are the same as in the previous example, which involved swapping notes maturing in June 2008 for notes maturing July 2008. The twist in forward rolls is that they involve swapping out of an existing OTR Treasury issue, into a soon-to-be OTR issue. Reverse rolls are the opposite. They entail selling the soon-to-be issued OTR bond against buying the already existing one. For the most part, pricing is driven by two factors, expected demand for the new issue and repo rates.

For an example, consider a forward roll transaction in two-year T-notes. Assume that the funds rate is 4%, that the current on-the-run two-year issue is trading to yield 5.01%, and that an outstanding seasoned issue maturing the same day as the yet-to-be issued two-year is trading to yield 5.03%. The new two-year note will be auctioned and settled in one week's time, but it is already free to trade on a when-issued, or WI, basis. It is quoted in yield. Where will the WI trade?

All things being equal, the new T-note should trade at the same yield as an outstanding note with an identical maturity, in this case 5.03%. But all things are not equal. The outstanding issue accrues interest daily, while the new issue does not begin to accrue interest until its issue date one week hence. As a result, the WI two-year note has to be discounted so that it offers a competitive rate of return compared to the outstanding issue.

With two-year notes yielding about 5% and a 4% funds rate, the opportunity cost of not being invested in 5% notes for a week is: 7 × (5% – 4%) ÷ 360 = \$194.44. Rounding numbers, the DV01 of a 5% two-year note is about \$188.08. So the discount for the WI two-year note in basis points is: \$194.44 ÷ \$188.08 = 1.03 basis points. If an existing issue with the same maturity (and roughly the same coupon) is trading at 5.03%, the WI should trade at about 5.04%. Otherwise there would be risk-free arbitrage profits available.

SUMMARY

This chapter presents a model for pricing bonds and applies it to governments. The factors that drive bond pricing are expected inflation, credit risk premium, and volatility. Since there is no credit risk for governments and inflation expectations do not vary by bond, the variable that matters for government bonds is volatility, which is largely a function of time to maturity. Position on the yield curve is the dominant factor differentiating government bond yields.

Three theories of the yield curve are considered. The first, segmentation theory, posits heterogeneous investor preferences and institutional arrangements that tend to impede trade gains across various sections of the curve. The second, the expectations hypothesis, argues that government bonds are perfect substitutes for each other and that market forces act to drive convergence between long and short rates such that the long rate is equal to the weighted average of expected short-term rates over the life of the bond. With this theory, the only thing that can be relied on to change the curve is new information. But empirically this is a difficult case to make. There have been numerous examples of entire maturity sectors trading at yields much higher than the expectations hypothesis would allow for.

The preferred habitat theory attempts to synthesize the expectations and segmentation theories. Investor maturity preferences do vary, but given sufficient yield differentials, investors can be persuaded to shift maturities away from what would otherwise be their first choice. Similarly, arbitrageurs strive to exploit pricing anomalies among bond issues by purchasing the cheap bonds versus selling expensive bonds short. Since the goal is to exploit pricing inefficiencies rather than speculate on the level of market rates, the trades are weighted to be market neutral.

The transactions described in this chapter for the most part can be considered to be tactical trades. The focus is often on relatively small pricing discrepancies that crop up between individual bond issues, generally in the same sector of the yield curve. Other trades with a similar structure seek to capture underwriting spreads that can open up when the Treasury auctions off new issues. These are transactions that involve trading new issues in the WI market in preparation for making an auction underwriting bid.

These types of tactical yield curve trades have narrow margins and relatively short holding periods. Financing costs are a significant factor and can easily be the difference between profit and loss. Strategic yield curve trades stretch out across maturity sectors, involve higher degrees of risk, and are more dependent on macroeconomic variables. These types of trades are discussed in the next chapter.

NOTES

[1] Douglas A. Ruby, *The Risk and Term Structure of Interest Rates,* 2003, available online at: www.digitaleconomist.com

[2] These data are available from *The Economic Report of the President* online at: http://www.gpoaccess.gov/eop/tables06.html; and the U.S Treasury's Office of Domestic Finance at: http://www.ustreas.gov/offices/domestic-finance/debt-management/investor_class_auction.shtml

[3] For a recent discussion of the expectations hypothesis see the speech "Understanding the Term Structure of Interest Rates," given by William Poole, president of the St. Louis Fed, to the Money Marketeers in June 2005, and reprinted on the St. Louis Fed Web site: http://www.stlouisfed.org/news/speeches/2005/6_14_05.html

[4] Treasury Office of Domestic Finance Policy Statement.

CHAPTER 13

Strategy and the Yield Curve

> Perception is strong and sight weak. In strategy it is important to see distant things as if they were close and to take a distanced view of close things.
>
> *–Miyamoto Musashi*

While there is no bright demarcation line, the differences between strategic and tactical yield curve trades are very real. Tactical trades focus on the relative pricing of individual bonds with similar coupons and maturities. Typically, these bonds trade at the same yield with only minor adjustments being made for slight variations in duration. If the differences are more than minor, tactical arbitrage traders will buy the cheap bonds and sell the expensive ones, anticipating that the yield spread between them will fall back into line before long. That is because larger than expected yield differences are likely to be attributed to a temporary imbalance of supply and demand of the individual bond issues, rather than anything more fundamental.

Strategic yield curve trades are different. While tactical trades focus on (presumably) short-term pricing anomalies brought about by imbalances between supply and demand, strategic trading focuses on the big picture. From this perspective macrovariables are the ones that matter most, particularly business and electoral cycles. Examples would be real GDP growth, inflation, fiscal policy, the trade picture, and monetary policy. Because the focus is on the big picture, strategic yield curve trades bridge across maturity sectors instead of staying within sector. For example, a typical big picture curve trade would spread the 2-year yield against the 10-year, or the 10-year against the current long bond. Most importantly, strategic yield curve trades forecast a change in the structure of the yield curve. But unlike tactical trades, there is no implicit assumption of mispricing.

Not only do macrostrategy yield curve transactions cross market sectors, but they are liable to have longer holding period horizons. Consequently, when considering weighting schemes for arbitrage transactions of this sort, the idea of market neutrality needs to be explored in more depth. While market neutral weights for a 10-year/long-bond trade are calculated easily enough using DV01s, it turns out that market direction, the level of rates, and the slope of the curve are correlated. Consequently, isolating and then neutralizing directional changes from curve changes can be problematic. Some dynamic hedging techniques may need to be employed to keep strategy on track. First though, it is important to consider the principal factors that influence the shape and slope of the yield curve.

WHAT DRIVES THE CURVE?

The previous chapter discussed three explanations for why the curve is priced the way it is, namely market segmentation, expectations, and preferred habitat. This chapter is concerned with the macrovariables that cause the curve to shift. Three are particularly important: the business cycle; Fed policy, and inflation. Fed policy is a very good predictor of the behavior of the curve. So is the business cycle, but Fed policy is much easier to predict than the business cycle. Moreover, Fed policy, inflation, and the business cycle are closely related.

No two business cycles are exactly alike, but they share important characteristics, so it is worth examining past business cycles as well as the policy and market reactions they provoked. For the purposes of this discussion, business cycles are defined with respect to periods of economic expansion and contraction. Expansion periods are those that extend from the trough of a recession to the peak of the subsequent recovery. Contraction periods are the opposite, extending from the peak of a recovery and ending at the trough of a recession. The peaks and troughs referenced here are those identified by the National Bureau of Economic Research (NBER), which has traditionally served as the national referee.

This expansion-contraction taxonomy has its limits. For one, it leaves a lot of information on the table. As a rule, economies don't suddenly drop from roaring growth into recession. Typically, the rate of growth begins to falter and then heads south and contracts until countervailing forces turn it around. Similarly, economies don't always snap right out of recession. Businesses are typically cautious in the early stages of a recovery, as they wait to see if they are on solid ground. And economists are a pretty glum bunch; they almost invariably underestimate the strength of a recovery when it materializes. They don't call it the “dismal science” for nothing. Nevertheless, characterizing different times as either contraction or expansion periods makes it easy to partition and observe market and policy behavior to see if there are repeating patterns before, during, and after expansions and contractions.

The time frame explored in this chapter extends from April 1953 through December 2005. There are two reasons for this. The first reason is data availability. The St. Louis Fed publishes a good deal of historical interest rate data, much of them based on a constant maturity model. These data turn out to be particularly good for modeling yield curve behavior. There is no noticeable bias in the data; the source is impeccable, and it is freely available. The data points are all expressed as generic yields, which eliminates issue-based idiosyncrasies, making the data particularly reliable as a time series.

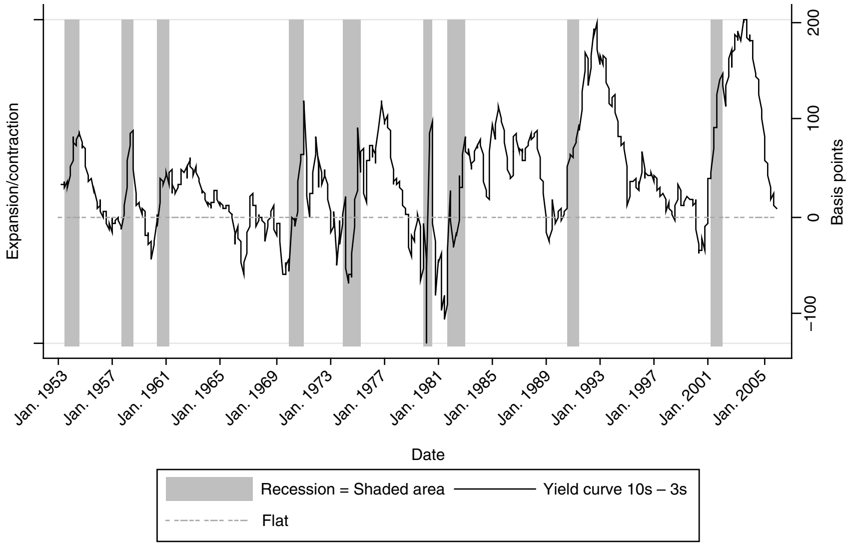
Another reason for beginning in 1953 is that the 1950s represent the beginning of the postwar, postdepression era. For the first time America was the dominant power in world politics and finance. Under the Breton Woods agreement negotiated in 1944, the U.S. Treasury effectively acted as guarantor of the international financial system, or at least the financial system of what had become known as the *free world.* By far, defense spending garnered the largest share of the federal budget. Financial markets were heavily regulated. Tracking market behavior over the huge changes in financial, social, and political systems that transpired from the 1950s to the present allows a broad perspective for evaluating developments in the yield curve and expectations for the future.

As a first cut, a simple measure serves to track the curve over the last 50 plus years. It is the month-by-month yield spread between 10-year and 3-year Treasury notes, based on the St. Louis Fed data. The path of the curve is displayed in Figure 13.1. It tracks the curve against a backdrop of U.S. economic performance, with the time series partitioned into recession or expansion periods. The shaded areas of the graph represent recession periods. The solid black line is the difference between the 10-year Treasury yield and the 3-year Treasury yield. Since the calculations are based on constant maturities, there is no relative change in the benchmarks over time (as there would be if they were real notes).

The behavior of the curve is striking. During every single recession (there are nine) the slope of the curve moved in favor of the short end of the market. It either flipped from negative to positive territory or became more positively sloped. Not only that but the curve steepened throughout each recession, finally peaking fairly early in the postrecession recovery period. Similarly, the curve displayed a marked proclivity to flatten as recovery turned to expansion. These market behaviors held constant through all nine recessions recorded by the NBER from 1953 through 2005. The similarity of behavior across very different eras strongly suggests two things; first, that curve behavior had a common cause across the different eras; second, the pattern is likely to continue in the future.

FIGURE 13.1

The Yield Curve and the Business Cycle (1953–2005)



Data sources: St. Louis Fed, NBER

The reason for the similarity of market reaction—the common factor—is not difficult to pin down. It is largely the Fed. With the notable exception of the 1970s, the Fed has acted to "lean against the wind" to soften the highs and lows of business cycles. Or as Fed Chairman William McChesney Martin once put it, the job of the Fed is to take the punch bowl away just when the party really gets going.

In formulating and implementing monetary policy, the Fed is relentlessly predictable. It has to be if it wants financial markets to function in an orderly fashion. Uncertainty produces unnecessary volatility, which raises risk premiums and increases the cost of capital and reduces investment and economic growth. Market volatility and uncertainty make it difficult for the Fed to read the market, increasing the probability of policy error.

The Fed is institutionally haunted by its twin policy disasters of the 20th century: the deflation of the 1930s and the Great Inflation of the 1970s. In a speech at the University of Chicago honoring Milton Freidman, Fed Governor Ben Bernanke referred to Friedman's work establishing monetary policy error as the cause of the Great Depression. He said, "The brilliance of Friedman and Schwartz's work on the Great Depression is not simply the texture of the discussion or the coherence of the point of view.

Their work was among the first to use history to address seriously the issues of cause and effect in a complex economic system. . . . I would like to say to Milton and Anna: Regarding the Great Depression. You're right, we did it. We're very sorry. But thanks to you, we won't do it again."[1]

In seeking to avoid a recurrence of policy error the Fed necessarily tips its hand to the marketplace. Among other reasons, it needs to gauge market reaction. When the Fed adjusts policy, it strives to induce behavioral change using the price mechanism. If market reaction is not as expected, the policy adjustment may backfire. In order to be effective, the Fed needs to communicate its policy intentions clearly and credibly.

Monetary policy is often most keenly felt on big-ticket items like housing, automobiles, and capital investment. These are not impulse items, and decisions surrounding them are based on long-term considerations. Consequently, for policy to have its desired effect, the Fed needs to be consistent, deliberate, and, above all, credible. Consistency, a form of credibility, must flow through the entire monetary policy signal chain. In this respect the money and bond markets are particularly important as they serve as the transmission belt for policy. Policy execution does not arrive in one fell swoop; rather it tends to consist of a series of consistent and reinforcing steps taken over time. That fact is the key to trading the yield curve successfully. Just as policy is implemented over time, the market adjusts over time. And it does so in the context of a clearly identifiable and definable trend.

Since 1994, the Fed has communicated its policy intentions by issuing statements after FOMC meetings. Its policy intentions and thinking are also conveyed in other ways: for instance, through official testimony, speeches by Fed officials, and newspaper leaks. But FOMC statements are the heart of the matter. Over time a great myth about "Fed-speak" has taken hold in the financial press. The myth is that the Fed chairman speaks in much the same way as the oracle of Delphi. The statements sound wise and knowing, but they are actually masterfully opaque. As a consequence, only the cognoscenti should be entrusted with the sacred task of correctly interpreting the true meaning of Fed pronouncements as they roll off the tongue of the chairman. This is nonsense.

The Fed spends a good deal of time and effort trying to convey its policy thinking to the public. Its statements are in plain English. In general, the format outlines policy options and decision criteria, discusses recent data, and weighs costs, benefits, and risks. It is true that the Fed leaves itself wiggle room. But the discussion is digestible by anyone who is reasonably conversant with the language of undergraduate economics. From a trading standpoint the key element to consider is whether a particular statement implies policy action and whether it is imminent.

That is why a small army of economists dissects every utterance of the Fed. There is good money to be made by paying attention. Trading the curve is a huge source of profit for the bond trading industry, and the Fed has an outsized impact on it; hence all the attention. The Fed's enormous influence on the curve plays itself out principally in three ways. First, when the Fed sets the funds rate, it effectively sets dealer financing costs. Second, monetary policy is a key determinant of inflation expectations. Third, when the Fed signals a change in the direction of policy, it signals the beginning of a trend that will probably last for a considerable period of time.

In the end actions speak louder than words. When the Fed wants to change policy, it will change the funds rate. And it will continue to move the funds rate in the same direction for a considerable period of time. It is therefore not even necessary to anticipate the first policy move. The trigger for establishing a yield curve arbitrage position can simply be Fed action.

Consider Table 13.1 which displays pairwise correlations among fed funds and constant maturity Treasury rates, beginning with the 2-year sector and ending with 20-year bonds. All are positively correlated at statistically significant levels, but the correlations are far greater for short-term securities. For example, according to the table, the level of fed funds explains 96% of the 2-year rate, but only 86% of the 20-year rate.

From a curve-trading standpoint, this raises two issues that warrant consideration. One has to do with the absolute level of the funds rate; the other, the direction of policy. The table only speaks to the level of funds, not policy direction. The effect of policy direction on the yield curve can be observed using a graph that simultaneously tracks the funds rate and a representative measure of the curve. For this purpose a good measure of the curve is the spread between 10-year notes and 3-year notes, which can be tracked against the funds rate on a daily basis, as in Figure 13.2.

TABLE 13.1

Correlation Matrix: Fed Funds and Treasury Notes
Monthly 1953–2005; two-year beginning 1976

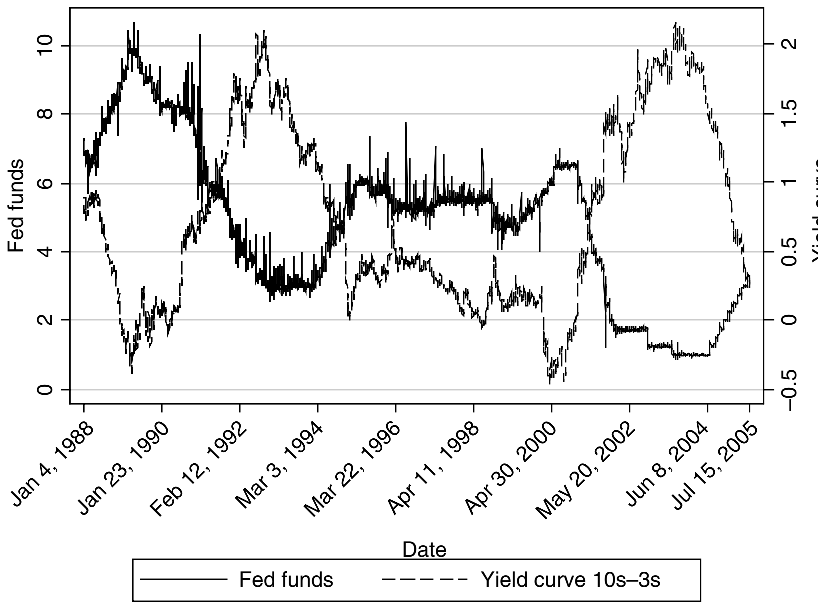
	fedfunds	cm2yr	cm3yr	cm5yr	cm7yr	cm10yr	cm20yr
fedfunds	1.0000						
cm2yr	0.9628*	1.0000					
cm3yr	0.9394*	0.9979*	1.0000				
cm5yr	0.9164*	0.9886*	0.9957*	1.0000			
cm7yr	0.8754*	0.9801*	0.9877*	0.9979*	1.0000		
cm10yr	0.8814*	0.9699*	0.9798*	0.9935*	0.9980*	1.0000	
cm20yr	0.8664*	0.9621*	0.9631*	0.9798*	0.9887*	0.9935*	1.0000

Data source: St. Louis Fed

* p = .05

FIGURE 13.2

Fed Funds and the Yield Curve (1988–2005)



Data source: St. Louis Fed

The Fed's influence on the curve is clearly evident in Figure 13.2. The funds rate and the curve are virtual mirror images of each other. As the funds rate heads lower, the curve becomes more positive and vice versa. As the graph clearly shows, over the 1988–2005 sample period there were five major shifts in the curve. Just as clearly, movements in the curve coincided with changes in the funds rate. These are large and sustained market moves that play out over long periods of time. They aren't over in a day.

ESTABLISHING STRATEGIC CURVE TRADES

Strategic curve trading needs to incorporate dynamic hedging to calculating position weights and financing. Dynamic hedging techniques need to be considered because market weighting requirements can change rapidly in some environments. Market-neutral positions can turn into longs or shorts in the blink of an eye if dynamic hedging is not employed. First, there is some ambiguity as to what constitutes a truly market-neutral weighting because curve shifts are highly correlated with market direction.

Second, the problem of convexity is liable to surface. There are two additional factors that need to be considered. The stakes tend to be larger in strategic curve trades, and workout times are likely to be longer as well.

In the end, strategic yield curve trades, despite being market neutral in form, are essentially a type of market timing, largely because of the correlation of the curve with the level of rates. What is the decision rule, or guide, for entering and existing trades? There is strong historical evidence that the curve moves with the business cycle and that the Fed responds to the cycle. The business cycle, unlike the Fed, is very difficult to read. Therefore the decision rule is: Follow the Fed.

A hypothetical trade will serve as an example of setting up and managing a strategic yield curve trade. Fed action serves as the trigger for trade initiation. But because market timing is critical, this time St. Louis Fed yield data serve as a reference point so that the example reflects actual market conditions and trading opportunities. This is not simply an example of calling the winner after the fact. The pattern in this example recurs over and over again. And to make the example real, the prices are calculated using the St. Louis Fed data, so they replicate the way the market actually behaved.

TRADE INITIATION

On January 3, 2001, the Fed, which had been in tightening mode, changed the direction of policy and reduced the federal funds rate from 6.5% to 6%. The Fed implied there would be more to come in its accompanying statement which said in part: "The Committee continues to believe that, against the background of its long-run goals of price stability and sustainable economic growth and of the information currently available, the risks are weighted mainly toward conditions that may generate economic weakness in the foreseeable future."

Accordingly, following the lead of the Fed the initial trade will be to "buy the curve," by purchasing short-dated Treasuries against establishing a short position in a longer-dated maturity. In this case, the long consists of $100 million 2-year notes; the short is a yield-weighted quantity of 10-year notes. The trade is executed on January 4, the day after the Fed policy announcement, after the market has had a full day and a half to adjust to the policy change. Settlement date is January 5, 2001.

The hypothetical 2-year notes for this arbitrage trade have a 4.75% coupon. They mature on December 31, 2002, and were issued on January 3, 2001. The 10-year notes are 5% coupons maturing November 15, 2010, issued on November 15, 2000. The announced fed funds target rate is 6%, so at the outset the assumed RP rate is 6.125% and the reverse RP rate is 5.875%. Using the DV01 weighting method produces an initial hedge ratio of ≈24.375 million 10-year notes for $100 million 2-year notes.

With these postulates and the yield data published by the St. Louis Fed, a running P&L can be calculated to evaluate the performance of the trade and the working of its component parts day by day. Also note that subsequent to the initial policy ease, the Fed reduced the funds rate two more times by 50 basis points each (on January 31 and March 20). RP and reverse RP rates are adjusted accordingly.

The P&L on this hypothetical transaction (displayed as a time series in Figure 13.3) remained in profitable territory the whole time, net of carry costs. But carry was negative, so for a short while the position was under water. Carry costs started at about $3,400 per day, falling to $1,300 per day as the Fed continued to ease policy. There were also substantial swings in the daily mark-to-market. Before carry, average daily profit was about $14,000 with a standard deviation of about $69,000. The best day produced a profit of $204,000; the worst day showed a loss of $185,000. Another way to view these data is with respect to yield changes. The notes' yields moved in opposite directions. The 2-year yield fell by 41 basis points, while the 10-year yield increased by 14 basis points.

There are three notable features about this transaction that warrant particular discussion. The first is the timing rule. The second is the fact that the yields of the two notes moved in opposite directions. The third concerns financing costs.

With respect to the timing rule—follow the Fed—it is important to note that the position was acquired at the close of business one and one-half days after the Fed's announcement that it was changing policy. The point being that there was more than enough time to get onboard. In fact, Figure 13.3 illustrates that the market continued to follow through with the curve continuing to slope more positively as time went on. This strongly suggests that all known information was not priced into the market. The critical piece of information—Fed easing—was clearly known. But it took a while for the full impact to be felt. Because history has shown that Fed easing leads to curve steepening, this transaction is an example of a low-risk, high-probability trade.

The second issue of interest is that the 2-year and 10-year yields moved in opposite directions. This is unusual, but it is not, as Yogi Berra used to say, "the most unheard of thing he ever heard of." Long-dated notes mostly respond to inflation expectations, while short-dated notes respond more forcefully to Fed policy cycles. Inflation expectations were not lowered by Fed easing, but the short end of the market clearly (and correctly) began to anticipate further policy ease. That is to be expected. What may be surprising to some is the persistence of the lag between Fed policy moves and full market adjustment. But it is there, and it is a rich profit vein to be mined.

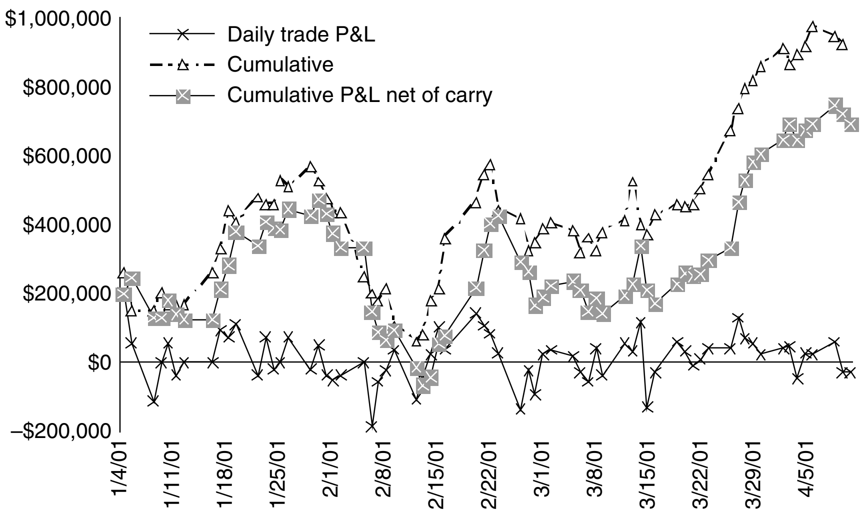
FINANCING

The next issue to be considered is the fact of net negative carry. If carry costs overwhelm arbitrage price gains, then the market has implicitly priced in all that is known, and the apparently easy arbitrage opportunity is illusory. But it turns out that price gains were substantially greater than carry costs, based on the run-of-the-mill assumption of a 25-basis-point spread between RP and reverse RP rates. The effect of the 25-basis-point market spread between RP and reverse RP rates is also graphically displayed in Figure 13.3. It is represented on the graph as the space between the lines that respectively track cumulative P&L with and without carry. The graph makes clear that, accounting for carry, the P&L started off in positive territory, remained there except for a few days, and inexorably trended into more positive territory.

Enduring negative carry is often the price of entry for buying the curve as the Fed begins an easing cycle. In this respect history shows a recurring pattern, displayed in Figure 13.1. As the economy softens and heads into recession, the yield curve is typically flat to negative, including financing spreads. Once the Fed begins to ease, the curve rapidly begins to favor its front, financing spreads turn positive, and the curve continues to move toward a more positive slope as the economy heads into the recovery. The curve then flattens, sometimes rapidly, once the economy is in its expansion phase.

FIGURE 13.3

Profit and Loss "Long the Curve"



Carry costs can be substantial, so managing trade workout time is important. Positioning a curve trade in anticipation of a Fed ease can get expensive fairly quickly when RP rates exceed position yields. Fortunately, it is not necessary to get in early and wait it out. The conservative strategy is to simply wait for the Fed to begin easing and then to act aggressively to get onboard by putting on curve-steepening trades.

It bears repeating that when the Fed begins a cyclical policy change, it will typically go on for a long time. And there is very little mystery about it. History shows that the Fed typically announces its preferred policy stance and sticks to its guns until it achieves the desired result. The main impediment traders face in putting on curve trades is psychological. With eyes fixed firmly on the rearview mirror, there is an overwhelming temptation to fret about having "missed the move." The result is paralysis and missed opportunity. This behavioral issue is a critical factor in successful trading, and it is addressed in a later chapter that discusses market psychology and recent research in behavioral finance.

DYNAMIC HEDGING AND POSITION MANAGEMENT

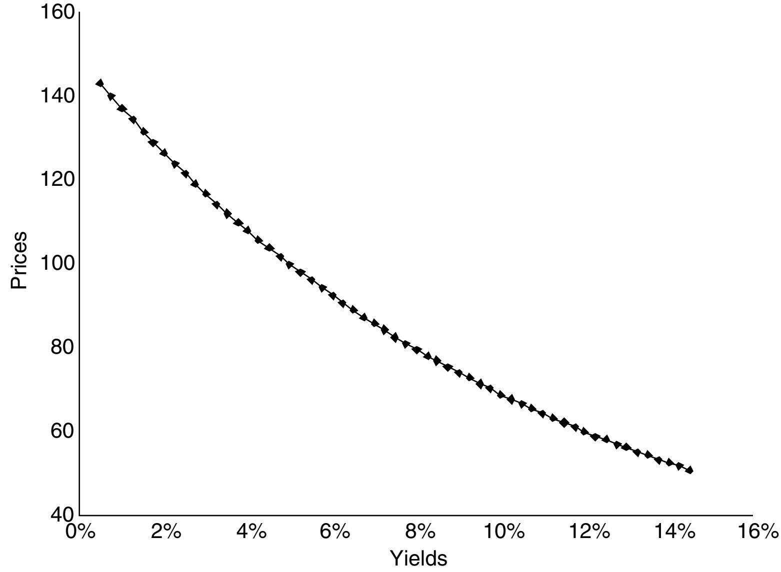
Even after positions have been set up, and the trend established—and there are identifiable trends—it is still easy to get shaken out of a winning position. The reasons are partly psychological. Another reason is not paying sufficient attention to the nuts and bolts of position management. One of the better exercises in position management is contingency planning, which entails running the numbers to see what happens under different interest rate scenarios. That way, market developments can be anticipated and dealt with according to a well-thought-out game plan. It beats panic every time.

Of particular concern for managing bond positions is the issue of convexity, which was briefly touched upon earlier in Chapter 10. The issue arises because the sensitivity of a bond's price to its yield varies with the level of interest rates. The second-order relationship between bond prices and yields is typically a curvilinear function rather than a linear one. This can be seen by plotting instantaneous changes in a bond's price against changes in its yield. The exercise will produce a convex curve, as shown in Figure 13.4, which traces the price/yield relationship of a sample 5% coupon 9-year note in 25-basis-point increments.

The rate scale for Figure 13.4, which ranges from yields as low as one-half of a percentage point to over 14%, is exaggerated to make the effect plainly visible. Rates do not typically move by very large amounts in very short periods of time. But sometimes they do. And that is precisely

FIGURE 13.4

Convexity: 5% T-Notes Due 11/15/2010



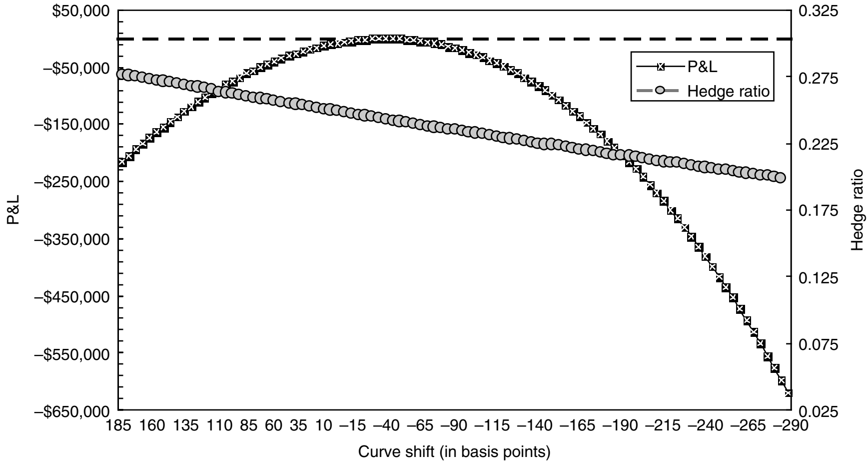
the time when it pays to have completed a scenario analysis by calculating the P&L impact on rapid, large changes in rates for yield curve trades under consideration.

This can be done in the current example by holding the 2-year to 10-year yield spread constant while changing the level of interest rates in 5-basis-point increments. That generates baseline prices for scenario analysis. First, a P&L is calculated at each 5-basis-point increment leaving yield spreads constant. Second, hedge ratios are recalculated at 5-basis-point intervals, again leaving the 2-year to 10-year yield spread unchanged. (Rate changes here are assumed to be instantaneous, so financing is not affected.) The results of these calculations demonstrate a significant (and negative) P&L impact from large rate changes, *regardless of the direction of change in the market*.

Figure 13.5 provides a graphic view of the outcome. With the yield spread between 2-year and 10-year notes remaining constant, a change in the level of rates produces a trading loss for an arbitrage position that is long the curve. The larger the market move, the larger the loss. Conversely, being short the spread produces the mirror image result. The reason is that 10-year notes possess greater convexity than the 2-year notes. Bonds with greater convexity rise more quickly and fall more slowly with respect to corresponding changes in rates than bonds with less convexity.

FIGURE 13.5

The Problem of Convexity: Dynamic Hedge Ratios versus Profit-and-Loss with Static Hedge



A look at the slope of the hedge ratio curve in Figure 13.5 hints at this result. Because the respective convexities of 2- and 10-year notes are unequal and nonlinear, the market-neutral hedge ratio constantly changes as market levels change. As rates fall, the quantity of 10-year notes needed to hedge $100 million 2-year notes falls as well. Referring to the graph at the starting point, it takes about $24.375 million 10s to hedge $100 million 2s. But after a 100-basis-point drop in rates, it takes only $23.9 million 10s to complete the hedge. Conversely, a 100-basis-point rise in rates causes the required short position in 10s to increase from $24.375 million at the outset to about $24.8 million. As a practical matter, being long the curve generally requires being constantly ready to adjust hedge ratios, generally in a disadvantageous direction, buying on the way up and selling on the way down.

Consider the current example. A drop in rates and rise in prices reduces the required size of the 10-year short, and vice versa. Hedge ratios are therefore dynamic, not static. Maintaining market neutrality requires traders who are long the curve to sell on the way down and buy on the way up, at least on the margin, to manage shifting hedge ratios. The opposite is the case for those who are short the curve. They have the advantage of being able to sell into rallies and buy into declines in order to keep their positions market neutral. This implies that optionlike characteristics may be embedded in yield curve positions. One solution is to “hedge the hedge” using delta-neutral bond option trading strategies of the type discussed in Chapter 7.

ARE CURVE TRADES TRULY MARKET NEUTRAL?

Constantly shifting hedge ratios combined with correlation between rate levels and the shape of the curve seem to cast doubt on the idea that yield curve arbitrage positions can be truly market neutral. The empirical evidence is strong that the Fed substantially influences both the shape and slope of the curve. When the Fed begins to ease, rates tend to fall and the curve favors shorter maturities. Seen in this light, going long the curve is biased to the bullish side. It is predicated on the idea that rates will generally fall, and in the process, short rates will fall faster than long rates, thus causing the curve to become more positively sloped.

Substantiation is provided by Table 13.2, which displays monthly means, standard deviations, minimums and maximums for constant maturity Treasuries beginning June 1976 through December 2005. Note from the table that as maturities increase, mean yields increase as well, consistent with the idea that the normal slope of the curve is positive. On the other hand, note that as time to maturity *increases*, the standard deviation *falls,* implying that short rates are more variable than long rates.

Assuming for the moment that rates on government bonds tend to move in the same direction at the same time, the inference would be that changes in the slope of the curve are the result of the typically faster movement of short rates, albeit in the same direction as long rates. Descriptively, that is fine as far as it goes. But it is ultimately unsatisfactory because while it is generally correct to say that government bonds tend to move in the same direction at the same time, the exceptions are

TABLE 13.2

Constant Maturity Treasuries, June 1976-2005

Treasury Yields	Mean	Min	Max	SD	N
fedfunds	6.633521	0.98	19.1	3.676894	355
cm1yr	6.685606	1.01	16.72	3.321385	355
cm2yr	7.042423	1.23	16.46	3.186783	355
cm3yr	7.207606	1.51	16.22	3.066371	355
cm5yr	7.477549	2.27	15.93	2.899414	355
cm7yr	7.679859	2.84	15.65	2.789569	355
cm10yr	7.791099	3.33	15.32	2.701524	355
cm20yr	8.060693	4.34	15.13	2.778717	274

Data source: St Louis Fed

glaring enough to call into question the idea of market neutrality in establishing hedge ratios.

As we have seen, arbitrage positions anticipating a steeper curve that are nominally market-neutral are in fact vulnerable to large, swift parallel shifts in the curve. That is hardly market neutral; it is an implicit forecast based on market timing. In fact, there are several forecasts embedded in a long-the-curve position, most with a decidedly bullish bias. Here are some implicit forecasts of being long the curve:

1. Rates in general will fall, with short rates falling faster and further.
2. Short rates will fall; long rates will remain unchanged or rise.
3. Short rates will remain steady; long rates will rise.
4. The Fed will not suddenly and unexpectedly begin to tighten policy.
5. The Fed may suddenly and unexpectedly decide to ease policy.
6. There will not be a rapid and large parallel shift in the curve.
7. The potential profit from a steepening curve is at least equal to the potential loss from either a flattening curve or a sudden and parallel shift in the curve.

Put another way, curve trades have at least a mild directional bias embedded in them. Consider the seventh implicit forecast on the list, that the probability of profit exceeds the probable risk of loss. It goes without saying that probabilistic calculus of profit and loss goes into any trade. But we know that generating a profit from going long the curve requires forecasting two things correctly. The first is that the slope of the curve will turn more positive. The second is that the magnitude of the curve change will be sufficient to offset convexity risk. These risks stem from the fact that a trading loss will result from either (or both) an unchanged curve accompanied by a sudden, sharp parallel shift in rates and/or a flattening curve from any market level. The risks faced by the arbitrage long position would therefore appear to be asymmetrical, which in turn makes forecast 7 seem problematical.

In fact, it is not. Other factors bring risks and rewards into balance. The first is that large changes in interest rates are almost always accompanied by changes in the slope of the curve. Changes in the curve are almost always sufficient to dominate convexity. The second is that, as discussed earlier, over the last 25 years the curve has repeatedly followed the same steepening or flattening pattern consistent with Fed policy and the business cycle. Third, and perhaps most importantly, the behavior of the Fed in times of crisis is utterly predictable. It is skewed toward policy ease and a more positive yield curve.

When a crisis erupts (either real or imagined), the initial market reaction is a "flight to quality." Money pours into short-dated Treasuries, which are universally regarded as the safest place to be, since they are hyperliquid and free of default risk. The flow of money into the front end typically causes bond prices at the short end to rise, pushing down yields, forcing the curve into steepening mode. If it turns out that a real crisis is under way, the instinctive reaction of central bankers is to flood the system with money, which typically causes the curve to steepen further still.

If an unexpected event threatens the stability of the banking system or the operation of the financial markets, without doubt the Fed will ease policy, perhaps dramatically. But the Fed almost never turns around to move abruptly and unexpectedly to tighten policy. Based on past history, the odds are simply overwhelming that if there is to be a precipitous policy move, it will be in the direction of ease, which in turn will provoke a swift move toward curve steepening. The Fed for instance eased policy in response to the Asian currency crisis of 1998, the September 11 attacks, the October 1987 stock market crash, the collapse of the Penn Square bank in 1982, and the collapse of Drysdale Government Securities in 1981.

The 1987 stock market crash can serve to put into perspective how rapidly the curve can shift when the Fed turns policy on a dime to stave off systemic risk. The weekend before the October 1987 crash, the 2-year to 10-year yield spread stood at plus 101 basis points. On Monday the stock market sold off 23%, and by Black Tuesday the credit system began to freeze up as banks refused to lend money to brokers. In response to a potential financial meltdown, the Fed announced that it would act to ensure that the system had sufficient liquidity to function properly. The government market experienced an explosive rally with short rates dropping about 150 basis points in two days. The Fed flooded the system with money and drove the 2-year to 10-year spread out 39 basis points from 101 basis points to plus 140.[2] Notwithstanding the convexity issue, widening the curve by 39 basis points over two days would have produced significant profits for a position long the curve. The addition of small quantities of call options as a convexity hedge would have added substantial profits.

DURATION WEIGHTS VERSUS DV01s

The trade examples to this point have used DV01s to calculate hedge ratios, but this is not the only way to proceed. Another popular method is to rely on a bond's duration, a measure invented by Frederick Macaulay. A bond's duration is the weighted average time to maturity of its cash flows. The weights are equal to the present values of the various cash flows, each as a percentage of the bond's price. A slight alteration of the duration formula produces a measure known as *modified duration,* which expresses a bond's duration in terms of years. This is often misinterpreted as referring to the length of time a bond investment is outstanding.[3]

The real importance of duration is its link to bond price volatility. A general rule of thumb is that a 100-basis-point change in a bond's yield will change the bond's price by approximately the same percentage as the bond's modified duration. For instance a bond with a modified duration of 7.5 would be expected to drop in price by 7.5% if the yield to maturity increased by 100 basis points. (Duration and modified duration can be calculated with Excel using the program's financial function Add-ins).

While it is true that a bond's price and its duration are very tightly linked for small changes in yield, it is also true that the price/duration link loosens as yield changes become more pronounced. For instance, consider Treasury 5%s of 11/2010 on settlement date January 5, 2001. At a yield of 5.25% the notes are priced at 98.088, with a modified duration of 7.625. A change in yield to 5.3% would reduce the price to 97.71, or 0.3831%, almost exactly what the rule of thumb calls for. But an increase of 100 basis points in the yield to maturity would cause the note's price to fall to 90.89, a drop of only 7.33% rather than the expected drop of 7.625%.

The reason for slippage in the price/duration linkage is convexity, the nonlinear relationship between a bond's price and its yield. There are ways to adjust for convexity, but they are relatively complicated.[4] An easier way to keep the hedge ratios current is to use DV01s and adjust them for changes in market levels. It is simple to do in Excel, and the results are accurate and easy to work with.

SUMMARY

Strategic yield curve trades are based on a macroview of the markets. Unlike tactical yield curve trades which are based on perceived pricing anomalies, strategic yield curve trades reflect a fundamental outlook concerning inflation, monetary policy, and the business cycle. Accordingly, strategic yield curve trades tend to focus on relationships that span market segments, for instance the 2-year versus10-year, rather than within-sector yield spreads.

The volatility of yield spreads tends to increase with the span of the curve: The 2-year/long-bond spread is likely to be more volatile than the 5-year/10-year spread. Consequently, the risks and expected profits for these types of arbitrage positions are typically larger than within-sector trades. While yield curve spreads can be quite volatile on a day-to-day basis, the historical evidence makes it fairly clear that the curve has trending characteristics. During times of Fed ease, the curve favors short maturities; during tightening periods the curve tends to favor longer-dated bonds.

The main drivers of the curve include inflation expectations, position financing costs, and the Fed's policy stance. Periods of policy ease and tightening are highly correlated with the business cycle. Easing is expected during slack times; tight policy would be expected during periods of rising inflation and capacity constraints.

Establishing a winning yield curve position does not require outguessing the Fed. Following the Fed's lead is a much better idea. Typically the Fed tells the market what it intends to do and then proceeds to act. The best strategy is to learn to listen to the Fed when it speaks and then act when the Fed does.

NOTES

[1] Remarks by Ben Bernanke at a University of Chicago meeting honoring Milton Freidman, November 2002. See: http://www.federalreserve.gov/BOARDDOCS/SPEECHES/2002/20021108/default.htm

[2] Based on St. Louis Fed constant-maturity time-series data.

[3] For a thorough discussion of duration see Frank J. Fabozzi, *Fixed Income Mathematics,* 3rd ed., McGraw-Hill, 1997, Chapter 14.

[4] See Fabozzi, Chapter 15.

CHAPTER 14

Trading the Treasury Basis

The future has already arrived. It's just not evenly distributed yet.

–William Gibson

Without question, the heart of the market for trading Treasury futures is the Chicago Board of Trade (CBOT). Consider for instance that on a typical day during the first half of 2006 the board traded just under 2 million Treasury contracts, representing over $210 billion in notional value. In recent years trading volume has picked up markedly. Several factors have contributed to the explosive growth in Treasury futures trading, including increased Treasury issuance due to rising budget deficits, the introduction of electronic trading, and reduced trading costs due to technologically driven efficiency gains.

The first CBOT Treasury bond contract appeared on the scene in the 1970s. Not coincidentally its debut occurred just around the time the bond trading business was beginning to emerge as a major profit center for the major investment banking firms. The original bond futures contract proved to be so successful that by the beginning of the 1990s the CBOT Treasury complex contained a full complement of Treasury futures products that spanned the curve, from fed funds to long bonds.

Each contract represents a particular section of the yield curve, and so there are contracts for 2-year notes, 5-year notes, 10-year notes, and bonds. The basic design of Treasury futures is essentially the same for the entire array of Treasury contracts. Each has a notional 6% coupon. Each has a set of rules limiting delivery to cash Treasuries that are representative of the maturity sector the contract represents. The 5-year, 10-year, and bond contracts each have a $100,000 notional value. The 2-year notional value is $200,000. The similarity of the contract design across the different maturity sectors has contributed to the success of the entire Treasury complex, among other things making the contracts easy to spread against each other.

Chicago Board of Trade Treasury futures are physical delivery contracts, which means that contract positions can be closed out either through delivery of an underlying bond or by an offsetting futures transaction. For example, shorts (longs) can either buy (sell) an offsetting contract (of the same specification), or they can choose to make (or take) timely delivery of a cash Treasury security that meets contract specifications. There is no other way to liquidate a position in CBOT Treasury futures. The result is that delivery requirements drive pricing.

MARKET STRUCTURE: CASH VERSUS FUTURES MARKETS

The market for cash Treasuries is organized as a dealer market. At present there are 22 firms recognized by the Fed as primary dealers in government securities. Becoming a primary dealer requires meeting three criteria. First, the firm is required to make "reasonably good" markets to the Fed when it enters the market to trade for the system account. Second, dealers are expected to participate "meaningfully" in Treasury auctions. Third, they are expected to provide the Fed with market information and analysis that the Fed may find useful in the formulation and execution of monetary policy.[1]

Over time the Treasury market evolved from a loosely affiliated group of bond-trading firms and their customers into a complex international trading and information network that includes recognized dealers, hedge funds, finance ministries, foreign central banks, and institutional investors. At the epicenter sits the Federal Reserve. Although the Fed is an extremely powerful player, its influence depends to a considerable degree on its persuasive powers rather than its raw power to set the funds rate. The Fed's choices are constrained by market expectations embedded in the network. In this sense the networked market structure of the Treasury market can be thought of as a giant system of feedback loops. Information is transmitted through the system by trades and trade reporting.

Decision makers of all sorts—day traders, hedge fund managers, or pension fund managers—draw inferences from market action and adjust their behavior accordingly. And despite the semiobligatory nods to the random walk theory, market participants constantly keep in touch with market action. Anyone who doubts this need only go to a first-rate hotel or any major airport and try to escape the ubiquitous financial market talking heads on CNN, CNBC, or Bloomberg television. Policy makers are not immune to this. The Fed's trading desk checks daily with dealers for market updates and relays the information to policy makers. It is hard to imagine that they would go to all this effort if they truly believed that day-to-day changes in market rates were simply white noise.

Daily trading in the Treasury market can be divided into two closely linked segments. The first is the primary market where new Treasury issues are sold by means of Dutch auctions (discussed previously) in which the Fed acts as the Treasury's agent. The secondary market, or aftermarket, is where T-bill, note, and bond issues trade after they have been auctioned off. In the secondary, primary dealers serve as market makers in all outstanding Treasury issues. In the normal course of business they stand ready to deal on either side of the market in large size more or less on a continually and at very narrow bid/ask spreads. Dealing desks are commonly organized around the various segments of the yield curve. They typically have a separate trader for Treasury bills, short maturity coupons, intermediate maturity coupons, and long-term bonds.

Institutional customers of the dealers include pension funds, hedge funds, bank portfolios, mutual funds, insurance companies, and foreign central banks. (The Fed is in a class by itself.) When a customer wants to buy or sell Treasuries, the usual procedure is to check prices on a dealing screen, or to ask salespeople at a few different dealer firms for a price and then execute at the best price. Acting in their capacity as market makers, the dealers take the other side of customer orders and choose whether to carry the position, hedge it, or liquidate it in whole or in part.

Traders at dealer firms have three potential execution venues. First, they can elect to try to offset their customer trades using their own sales forces to find the other side. Second they can go to interdealer brokers who display bids and offers on dealing screens. These screens are largely restricted to the dealer community and about 300 very large customers. Third, they can turn to the futures market to hedge.

FUTURES EXCHANGES

In many ways futures exchanges are mirror images of the OTC Treasury bond markets, with the Chicago Board of Trade being the dominant Treasury exchange. But there are important structural differences between cash and futures markets. While the cash OTC markets are built from the bottom up, issue by issue, the futures markets are built from the top down. Instead of trading hundreds of different issues, CBOT contracts are generic, designed to capture key characteristics of large swaths of the Treasury note and bond markets. This generic catchall quality of Treasury contract design is a powerful contributor to liquidity. Because of it, each Treasury contract can serve as a hedging vehicle for many different cash securities.

There are two characteristics in particular that differentiate futures exchanges from the OTC bond markets in important ways. First, in contrast to the looser network structure of the OTC markets, futures exchanges are more formally organized as central marketplaces with established rules and operating procedures. Exchange-traded contracts have standardized terms that make them relatively easy to price and trade. By contrast, the OTC markets trade individual securities, any of which may have a wrinkle attached to it. Second, at futures exchanges all trades are guaranteed by the clearinghouse which acts as a principal between the contracting parties. The clearinghouse provides system oversight by requiring traders to post a performance bond and, if necessary, additional margin. The result of this arrangement is that traders can easily liquidate or acquire positions because they are fungible at the clearinghouse. By contrast, credit extension is an individual (or firm) specific affair in the OTC markets.

Order flow routing at exchanges is markedly different from that at OTC markets. In the OTC markets, trades occur in many different venues, essentially wherever the other side can be found. Not so in the futures markets. Organized as central marketplaces, futures exchanges generally require that all trades in exchange contracts go through the exchange where bids and offers are available to all and pricing is transparent—off-board trading is not allowed.

Exchanges argue that centralized markets provide a wide array of benefits to the marketplace. First, they make markets transparent by having all orders openly communicated in the central marketplace. Second, they contribute to market efficiency by distributing price information widely and rapidly. Third, requiring that all trades go through the exchange keeps the playing field level and builds trust.

It is true that listed markets have a transparency advantage. But the significance of differences between listed and the OTC markets can be overstated. When markets are volatile, OTC dealers try to quickly lay off significant risk temporarily acquired as a result of taking the other side of customer trades. In the process the information makes its way into the marketplace quickly—far more quickly than the dealer would like in most cases. First, a dealer faces the danger of tipping his hand trying to liquidate a large position. Second, before a customer executes a large trade, he typically checks prices with a few dealers, so the winning bidder is keenly aware that his competitors are acutely conscious of the fact that a large buyer (or seller) is in the marketplace. As a practical matter it is pretty difficult to keep these things secret for more than a couple of minutes on the Street, where rumors and bad jokes spread like wildfire.

The upshot of it is that there are structural differences between listed and OTC markets, some of which have a significant impact on how business is conducted. But in the end, listed and OTC markets are joined at the hip. Arbitrage trading between cash Treasuries and CBOT futures contracts keeps prices between the two venues closely aligned.

PRICING CBOT TREASURY CONTRACTS

One of the best ways to explore the pricing of CBOT Treasury futures contracts is to first take a short excursion into pricing forwards. A *forward* is an agreement to buy or sell an asset at an agreed-upon price at a specified time. All else equal, the forward price is a function of the risk-free interest rate, storage, and insurance costs. For Treasuries storage and insurance costs are nonexistent to trivial so they will be ignored. As a consequence, the pricing model for forward delivery defaults to the risk-free rate as the driver. It can be easily calculated using the formula below.

$$F = S(t)e^{r(T-t)}$$

where:

F = the forward price;
S = the spot price;
r = the risk-free interest rate;
T = the expiration date;
t = the current date;

Therefore, $T - t$ = the number of days left to expiration. For example, the 181-day forward price of a non-interest-bearing $1,000,000 asset when the risk-free is 6% would be:

$$\$1{,}030{,}626.29 = \$1{,}000{,}000\ e^{0.06/360*181}$$

The main difference between a forward and a future is contract specificity. A forward contract is specific in every detail. The specific asset that is to be delivered, the price, the quantity, and the time of delivery are all established at the outset. On the other hand, a futures contract is a standardized agreement that establishes the parameters of what constitutes good delivery. Standardization makes futures contracts easily tradable, particularly when estimating delivery costs is reasonably straightforward, as it is with CBOT Treasury futures. Pricing the futures contracts simply requires taking into account the risk-free rate plus any other factors in the contract specifications that can affect the cost of delivering the underlying commodity.

Chicago Board of Trade Treasury futures contracts are designed to capture the market behavior of specific segments of the Treasury yield curve. (In this respect the contract design tips its hat to the segmentation theory of the curve.) For instance the Treasury bond contract is designed to emulate the long end of the Treasury curve, the 10-year note contract is designed to capture the long-intermediate curve, and so on. Contracts expire quarterly, every March, June, September, and December. Not coincidentally, when the original contracts were first invented, these were the dates that roughly corresponded to the times of year when the Treasury was traditionally most active issuing securities to finance budget deficits. Contract expiration dates are thus timed to induce hedgers to use Treasury futures to lay off anticipated exposure to the usual quarterly auctions of 3-year notes, 10-year notes, and long-term bonds, which take place in February, May, August, and November.

In order to make sure that Treasury futures accurately mirror the activity taking place in the cash markets, there is a contract delivery requirement. The criteria for delivery eligibility and, to a lesser extent, the rules specifying delivery methods define the essence of the contract.[2] For traders with Treasury contract positions the requirement to liquidate, make, or take delivery is ironclad. Consequently, the contract necessarily mirrors the market behavior of the bond (or bonds) deemed most likely to be delivered at contract expiration.

From a contract-pricing standpoint, the most important variable is a bond's remaining time to maturity. For example, the rules of the long-bond Treasury contract specify that delivery eligible bonds must have 15 or more years remaining to either maturity or call (if the bond is callable). Specifying the bond contract this way ensures that the market reflects activity in bonds with at least 15 or more years to go before they are retired. The requirement is based on the assumption that these bonds collectively mirror long-term investor expectations. It also ensures adequate deliverable supplies. As of this writing, there are outstanding about $280 billion of Treasury bonds with 15 or more years remaining to maturity. At the same time, some key characteristics of delivery eligible bonds can vary considerably. For instance, there is no restriction on the coupon other than that it must be above 0%, paid semiannually. At the moment coupons on delivery eligible bonds range from 4.5% to 8.125%, and eligible maturities range from 15 to 30 years.

Variation in the essential characteristics of delivery eligible bonds prompted the CBOT to establish a system to translate cash bond prices into futures delivery equivalent prices. It was done by assigning each eligible bond a delivery conversion factor. The conversion factor is a close approximation of the price at which the bond would trade if it yielded 6%, because the futures contracts are notional 6% coupon bonds.[3] The delivery price for eligible bonds is a product of the conversion factor and the futures settlement price. Making (or taking) delivery at this price fulfills the contract obligations of a contract short (or long) as the case may be. Actual deliveries, which constitute a tiny minority of contracts traded, take place every March, June, September, and December.

THE BOND BASIS

The *basis* is the spread between the market price of a delivery eligible bond and the price at which it can be delivered into its corresponding futures contract, consistent with the rules. This spread is, by far, the single most important factor affecting how note and bond futures are priced, and it has significant ripple effects throughout the entire marketplace.

An example will serve to show how the basis is calculated and how it can be used to construct and implement various trading strategies. For the first basis trading example Treasury 8.125% bonds maturing May 15, 2021, and the March 2006 Treasury bond futures contract will be used. Market prices are below. Using Street convention, they are quoted in 32nds of a point. A "+" symbol represents an additional one 64th so, for instance 1+ is equal to 3/64. Prices are as of January 19, 2006.

Treasury 8.125% May 15, 2021=139. 01+ or 139.0469 (in decimals)

USH6 (March Bond Futures) =114. 24 or 114.75

RP Rate = 4.35%

Conversion Factor =1.2083

To determine the delivery price for Treasury 8.125% of May 2021, multiply the conversion factor by the futures price:

Delivery Price = 1.2083 * 114. 24/32 = 138.6524.

The basis is the cash price minus delivery price:

Basis =139.0469 – 138.6524=0.3945, or $3,945 per million bonds.

By convention, the basis is quoted in 32nds of a point (and fractions thereof) per million. In this case the basis would be quoted as 12¾/32, because 12.75/32 * $1,000,000 par value = $3984.375, which is reasonably close to the calculated spread.

The basis as calculated here is better described as the *gross basis* because it does not take into account the probable cost of carry. The *net basis* is the gross basis less expected carry profits (or losses). A comparatively simple way to estimate the net basis is to subtract probable financing costs from coupon accruals up through to the contract expiration date. But financing cost calculations are only probabilistic. Daily RP rates are subject to change. So are cash balances and margin deposits which vary with market levels.

With those caveats in mind, a rough carry estimate can be calculated for a hypothetical long position in Treasury 8.125% bonds for a holding period extending from January 19, 2006, to the March 31, 2006, contract expiry. This financing period contains 71 days (with a coupon payment on February 15). The assumed RP rate is 4.35%, and the initial amount borrowed (including accrued interest) is $1,426,851.22 per million at par value. Borrowing costs are therefore estimated as follows.

From the January 19 settlement date through the first coupon:

$1,426,851.22 * 0.0435 * 27/360 = $7,586.09

From the February 15 coupon date through to March 31:

$1,392,187.50 * 0.0435 * 44/360 = $4,542.01.

Total borrowing costs = $7,586.09 + $4,542.01= $12,128.10.

Interest earnings are equal to the coupon earned plus reinvestment of the February 15 coupon payment of $40,625 at the federal funds rate.

Total coupon interest from January 19 through March 31 = $15,836.97.

Reinvestment = $40,625.00 * 44/360= $215.99

Total estimated interest = $16,052.96.

Net expected cash = $16,052.96 – $12,128.10 = $3,924.85, which is almost exactly equal to the dollar value of the bond's basis spread $3,945 per million. There is actually a bit of a hitch in disguise here. In order to assess carry, the price of the bond would ordinarily be recalculated at the termination date using the same yield as that that obtained at trade initiation. That way, amortization of premium would be accounted for. However, in basis trading what matters is the convergence of cash and futures prices, not one price taken in isolation.

Arbitrage trading of the basis, using this cash and carry model, keeps the relative prices of futures contracts and cash bonds in line. If a cash bond begins to trade under its delivery value adjusted for financing, arbitrageurs will go long the basis (long cash/short futures). Arbitrageurs will buy the cash bond and sell the corresponding futures contracts until the cash bond rises enough (or the futures contracts fall enough) to force the basis to where it is at least roughly equal to its estimated carry value through to contract expiration.

But the mirror image transaction—going short the basis—does not necessarily work the same way. That is because there are de facto options embedded in Treasury futures contracts that give basis longs the potential to reap substantial profits. As a result, Treasury note and bond bases often trade at premiums that reflect these embedded option values. To see how this works, it is necessary to review the mechanics of setting up basis a trade.

THE MECHANICS OF THE BASIS

At its most basic level, setting up a long basis trade refers to buying a deliverable Treasury note or bond and selling a weighted amount of futures contracts short. Once established, the position leaves the arbitrageur with two possible options for unwinding it. The position can be liquidated either by delivering bonds or by covering the contract short and selling the cash bonds in the marketplace. Which to do is simply a question of price. But that raises an important question. How many futures contracts should be sold short to hedge any given cash bond? It depends.

The baseline criterion for determining the best contract weighting is whether the bond being hedged is the cheapest-to-deliver (CTD) bond in the basket of delivery eligible securities. The CTD bond is the one whose delivery price is closest to the market price, adjusted for carry. Consider, for instance Table 14.1, which lists some delivery eligible Treasury bonds (for the March 2006 contract) and their prices, yields, conversion factors, the (gross) basis, and the contract's estimated convergence price.

The last column in the table labeled *convergence* is particularly important. It represents an implicit forecast price for where the futures contract will eventually settle on the final expiration date. Assume for a moment that the last trading day and contract expiration day are one and the same. (They are not, but that doesn't matter for now.) At expiration, assuming no ambiguity about which bond issue will be delivered, arbitrage ensures that the factor-adjusted bond price and the futures contract price are the same; otherwise any spread above or below that price would allow for instantaneous risk-free profits. That implies an expiration day basis of zero.

The original basis calculation took the form:

$$\text{Basis} = \text{Cash}^{\text{Price}} - (\text{Futures}^{\text{Price}} * \text{Factor})$$

If the basis is set to zero and the terms are rearranged, the equation reduces to:

$$Futures^{Price} = \frac{Cash^{Price}}{Factor^{CTD}}$$

TABLE 14.1

Delivery Eligible Treasury Bonds

USH6 Settle	114 12/32 Coupon	Maturity	Yield	Price	Factor	Basis	Convergence
1/19/2006	8.125%	5/15/2021	4.590%	138 19/32	1.2083	12/32	114 22/32
1/19/2006	6.250%	8/15/2023	4.612%	119 18/32	1.0265	2 5/32	116 16/32
1/19/2006	6.125%	11/15/2027	4.605%	120 25/32	1.015	4 22/32	119
1/19/2006	5.250%	2/15/2029	4.560%	109 25/32	0.9075	6	120 31/32

Divide the cash market price of the cheapest-to-deliver bond by its conversion factor. All else equal, the result is the expected price of the futures contract on expiration date. Since the basis is specified as zero on expiration date, it is the futures convergence price. In the current example, March bond futures will rise in price by 10/32nds, from 114.12 to 114.22, to converge with the 8.125% cash bond, as shown using the formula:

$$114.22 = \frac{138.19}{1.2083}$$

As long as the 8.125% issue of May 2021 is the CTD bond, the market-neutral hedge ratio of contracts to bonds is the same as the delivery factor. Thus a delivery weighting would call for 1.2083 contracts to be sold short for every $100,000 par value bonds held on the long side of the market. By extension, for every $1 million par value bonds held long, 12 contracts would be sold short, and so on. As long as the 8.125% long bonds remain the cheapest to deliver, the factor adjusted price difference between the cash bonds and the futures contract will gradually converge toward zero as carry value is amortized over the time remaining to contract expiration.

In general, a bond futures hedge neutralizes the risk of adverse price movements brought about by changes in the level of market interest rates. To the extent that the cash and futures positions are locked together until contract expiration by the carry spread, the hedged position is effectively reduced to the equivalent of a money market instrument, albeit with an option attached. The implied option exists because certain factors can cause a shift in the cheapest-to-deliver bond, creating the possibility of substantial arbitrage profits.

EMBEDDED OPTIONS: WILD CARDS AND TAILS

There are mainly three factors that can cause a shift in the cheapest-to-deliver bond. The first comes from the so-called wild card option in a delivery month. The second arises from a change in the general level of interest rates. The third comes from a change in the slope of the yield curve. The probability of any of these things happening depends on the level of rates and market volatility. We will now consider each of these possibilities.

The wild card delivery play rests on the mechanics of the delivery process. The choice of which bond to deliver and when lies with the contract short, within certain constraints. Presumably deliveries will only be made when there is at least a reasonable chance that the process will yield greater profits than simply covering the short in the open market.

When the delivery option is elected, the bond has to be chosen from a basket of eligible securities specified by the contract's rules. Essentially, the basket contains bonds that are coupon bearing and possess certain minimum (and sometimes maximum in the case of notes) maturities. The exchange (and any number of vendors, particularly Bloomberg) regularly posts the list of eligible bonds and their delivery factors for all outstanding futures contracts.

From the trader's perspective, delivery against futures contracts is a multistep process. The first step is the decision to deliver rather than cover the short in the open market. By the rules, the short seller is required to serve notice of intent to deliver. (Declaration of intent to deliver always rests with the short; long contract position holders do not have the right to call for delivery as they do for some commodity contracts.) When contract shorts declare intent to deliver, they do so versus that day's closing futures settlement price. The option is available every trading day of the delivery month up until the last trading day. (The last trading day is the seventh business day preceding the last business day of the delivery month.) After the final settlement price is established on the last trading day, remaining contract shorts have no choice but to deliver against that final settlement price.

Contract shorts not only choose when to serve notice of delivery intent, but they also choose which bonds to deliver. Combining the choices of what and when to deliver produces myriad delivery options for contract shorts. Adroit use of these options can result in substantial arbitrage profits.

TRADING AGAINST THE TAIL

Consider that in a delivery month, but prior to the last trading day, contract shorts can elect to serve notice of intention to deliver against the settlement price established at 2:00 p.m. Chicago time. But the trader can wait until as late as 8:00 p.m. that day to serve notice to his clearing firm that he intends to make delivery. Further, he is not required to go the next phase in the process and declare *which* bonds he will deliver until the following day.[4] The result is that he can deliver against a price established at 2 p.m. Chicago time and frozen in place until 8:00 that night, even though the cash market remains open for business. As a practical matter, during delivery months, arbitrage shorts have several hours a day to shoot at a target that is standing still.

The following example will make clear the potential for arbitrage profits made possible by this arrangement. Assume an arbitrage position that is long $100 million par value 8.125% coupon Treasury bonds maturing May 15, 2021, hedged by a (factor weighted) short position totaling

1208 March 2006 bond futures. As discussed previously, the most accurate hedge ratio for the cheapest-to-deliver bond is determined by the delivery factor. But when it comes time to actually make delivery, the contract short is obliged to supply the same par value in delivery eligible bonds as he is short contracts. In this example the short has to deliver, not a total of $100,000,000 hedged bonds, but a total of $120,800,000 par value bonds against the short of 1,208 contracts with a notional value of $100,000 apiece.

It turns out that the additional 208 contracts, known as the *tail*, is a source of possible arbitrage profit. Continuing with the present example, suppose that on March 6 the bond futures contract settles at 114 14/32 with the 8.125%s of May 21 trading at 138 11/32, which would be a basis of 2/32. Further suppose that sometime after the futures contract settles at 2 p.m. and before 6 p.m. some bad news comes to light and market yields ratchet up 5 basis points. That could touch off an early delivery into the March bond contract, yielding arbitrage profits.

To see how this would come to fruition, consider Table 14.2. It displays prices and yields for the 8.125% bonds, 5 basis points apart. The first row of the table is the market price at 2 p.m. Chicago time; the second a few hours later. At 2 p.m. with the futures contract at 114 14/32, the basis on the cheapest-to-deliver 8.125% bonds is 0.076 (expressed in decimals). To close out the entire position, the contract short could cover 208 contracts at the closing price and then deliver off $100 million cash 8.125% bonds against the remaining short of 1,000 futures contracts. That would result in a loss of $76,121, which is the spread in dollars between the delivery price and the market price—in other words, the basis.

Suppose however that the contract short decided to hold off. Further suppose that market yields increase by 5 basis points after the close but before the notification for delivery period terminates. A 5-basis-point increase in yield would drop the price of the 8.125% bonds by about 21/32nds to 137.6795. But the contract short would still have the option

TABLE 14.2

Prices and Yields for 8.125% Bonds

Coupon	USH6 Maturity	114 14/32 Yield	Price	Delivery	Basis	P&L	Net Delivery
8.125%	5/15/2021	4.59%	138.3510	138.2748	0.076	–$76,121	$100 million
8.125%	5/15/2021	4.64%	137.6795	138.2748	–0.595	$123,825	$20.8 million
						$47,703	Net

to serve notice of intent to deliver the entire position *at the delivery price established by the 2 p.m. settlement*.

If the contract short decides to serve notice of intent to deliver, he has to deliver $120,800,000 par value bonds against the short of 1,208 March futures. Since the short only has $100 million in position, he needs to come up with an additional $20.8 million bonds to make good delivery against his total short position of 1,208 contracts. This he can easily do by entering the cash market to buy an additional $20.8 million bonds which are now trading at $137.6795—substantially lower than the $138.2748 delivery price established at 2 p.m. Chicago time. Just to be sure, he can buy the additional bonds (the tail) in the open market before declaring intent to deliver.

By purchasing $20.8 million cash bonds to deliver against the tail, the contract short locks in a substantial profit on the whole position because the gains from delivering against the tail overwhelm the losses from the bulk of the position. The $20.8 million cash bonds can be bought in the marketplace at a price of $137.6795 and delivered off at a price of $138.2748. That maneuver produces a gain of 20.8*($138.2748 – $137.6795) = $123,825, more than enough to compensate for the loss of $76,121 incurred by delivering the remaining $100 million bonds. The net profit amounts to $123,825 – $76,121= $47,703, as shown in Table 14.2.

In fact, the contract short doesn't even necessarily have to get to the point of declaring intent to deliver to make these arbitrage profits. Suppose, as above, the market drops in price by 21/32nds of a point immediately after the closing settlement price is established. The contract short can enter the market and buy $20.8 cash million bonds to cover the tail of his contract short and then just wait before declaring intent to deliver. If the market were to run back up (for instance to the 2 p.m. closing price) the strategy would be to sell out the $20.8 million bonds, book the profit, and wait for the opportunity to do it again. Leaning against the tail gives the contract short the potential to repeat this over and over again with very little risk.

Another factor to consider is that this delivery maneuver can work in either direction. In the example just discussed the position had "put" value because the conversion factor exceeded 1. As the conversion factor rises above 1, additional put value is created because as the tail increases, more bonds can be delivered or "put" into the contract. The opposite is the case when the factor is less than 1. Then the position would be said to have "call value." In that case, in the event of a rally, the contract short would execute a mirror image transaction of the one cited above. In the event of a market rally after the settlement price, the contract short would sell his excess cash bonds and deliver the remaining bonds into the contract to satisfy his remaining short obligations.

TABLE 14.3

Net Profit for Bond Sales

USH6 Date	103 4/32 Coupon	Maturity	Yield	Price	Delivery	Basis	P&L
3/7/2006	5.250%	2/15/2029	5.750%	93.6689	93.5859	–0.0829	–$75,290
3/7/2006	5.250%	2/15/2029	5.650%	94.8883	0.9075		$112,191
							$36,901

For example, consider a case where the CTD has a coupon under 6% and therefore a conversion factor less than 1. For this example we will take the case of Treasury 5.25% bonds maturing 2/15/2029, the factor for which is 0.9075. For the sake of example we assume that March 2006 bond futures are trading at 103.04/32 and that the 5.25% bonds are trading at 93.6689 to yield 5.75%. The basis is therefore 93.6689 – (0.9075 * 103.04) = 0.0829, or about 3/32nds. Since we posit the 5.25% bonds as the cheapest to deliver, we factor-weight a basis trade, long $100 million cash 5.25% bonds, short 908 March futures contracts.

Now consider a situation in which the bond market rallies after the daily settlement price is set. Let bond rates fall by 10 basis points after futures close but before delivery notification is required. A 10-basis-point drop in yields causes a price increase in the cash Treasury 5.25% bonds of about 1 point and 7/32nds. An arbitrageur who is long $100 million Treasury 5.25% bonds and short a factor-weighted 908 contracts would be able to sell $9.2 million cash bonds into the rally and then deliver $90.8 million bonds into the contract short. Selling $9.2 million cash bonds up 1 point and 7/32nds produces a gain of $112,191. Delivering the remaining balance of $90.8 million bonds against the short of 908 contracts produces a loss of 90.8 * (–0.0829) = $75,290. The result is a net profit of $112,191 – $75,290 = $36,901. See Table 14.3.

Suppose in this example the trader sells $9.2 million of the 5.25% bonds up 1 point and 7/32nds and books a profit of $112,191. But suppose that another two hours remain until the contract short is required to serve notice of intent to deliver. If the rally ends and bonds sink back down to where they were at the outset, the trader can just buy them back, book profits on the sale as a short sale while leaving the position intact for a potential repeat performance. The game is hardly over at this point. There can be repeated rounds of play, leaving the core arbitrage position intact until expiration day.

THE WILD CARD

Far more potentially lucrative trades are available that turn more on *which* bonds to deliver rather than *when* to deliver them. That is where the wild card option comes in. These trades are most likely to work after the final contract trading day, but before the last delivery day. The reason has to do with time and the mechanics of setting the final delivery price.

The final trading day for an expiring bond contract is the seventh business day preceding the last business day of the delivery month. Trading in the expiring contract ceases at noon. Once the final settlement price is established (a few minutes after trading ceases), all remaining deliveries are made against the final settlement price. This element of the delivery process is particularly important. Typically there will be another five or six trading days remaining in the delivery month even though final delivery prices for eligible bonds are frozen in time when the final futures settlement price is established. Since final prices for delivery into the contract are fixed, there is not much point in weighting hedges against potential futures contract price changes because there won't be any.

Consequently, during the final minute of trading, arbitrageurs will typically attempt to adjust futures positions so that the notional value of contract positions equals the par value of cash positions. Keep in mind, though, that only traders with existing positions in the expiring contract are permitted to trade that contract in the last minute leading up to expiration. The market can get pretty thin, and a bad execution is always a possibility.

Leaving aside execution quality, rebalancing the hedge from factor weighting to even par amounts opens up a world of possibilities. Basis positions that formerly had embedded put value can flip to positions with embedded call value, and vice versa. Depending on the shape and slope of the curve, some bond basis positions take on straddle value. Moreover, it isn't even necessary to have a cash position to trade the delivery option. As an alternative it can be profitable to borrow bonds to deliver against a calendar spread. Let's consider these possibilities.

For expository convenience a number of simplifying assumptions are made. First, bond yields and repo rates are approximately the same so there are no carry profits or losses. Second, bond yields are in the area of 5.75%. Third, we posit that three bonds constitute the universe of bonds eligible for delivery into the expiring March 2006 contract. (Typically there would be more than 20, so this assumption underrates the power of the wild card). Fourth, on the last trading day we cover the tail of the contract position at the final settlement price leaving an arbitrage position weighted by par values. For the sake of this example, the par value positions will consist of a short of 1,000 March bond contracts and a long

TABLE 14.4

Data for Three Representative Delivery-Eligible Bonds

Settle	USH6 Coupon	100 Maturity	Yield	Price	Factor	Delivery	Basis (Decimals)
3/22/2006	6.250%	8/15/2023	6.000%	102 21/32	1.0265	102.65	0.0202
3/22/2006	6.125%	11/15/2027	5.990%	101 20/32	1.015	101.5	0.1162
3/22/2006	4.500%	2/15/2036	5.830%	81 9/32	0.793	79.3	1.9734

position of some combination of $100 million deliverable cash bonds. Finally, the settlement price for the expiring March contract is 100.

All this information including the final settlement price for the expiring contract, prices, yields, conversion factors, and final settlement day bases for the three bonds that constitute the hypothetical universe of delivery-eligible bonds can be found in Table 14.4.

Since trading has ended in the expiring March bond contract, the contract short is obliged to deliver $100 million in par value bonds by month end against the outstanding short position of 1,000 contracts. The only question is which bond, or combination of bonds, represents the optimal delivery set, keeping in mind that the cash market in deliverable bonds will trade for several more days even though the final delivery price has already been established for the expiring contract. The answer once again is, it depends.

At first blush the optimal delivery bond would seem to be the 6.25%'s of 8/2023. It has the narrowest basis, and it has some put value to boot. The put value derives from the fact that the conversion factor is greater than 1. But even though the Treasury 6.25% bond has the largest conversion factor, it does not necessarily possess the greatest put value. To see this, consider the impact of an across-the-board 25-basis-point rise in yields, as shown in Table 14.5.

As Table 14.5 shows, a 25-basis-point rise in yields drives prices of the longer-maturity Treasuries down farther and faster than the shorter-maturity Treasury 6.25% issue maturing August 2023. As a result, the 6.125%'s of

TABLE 14.5

Impact of Across-the-Board 25-Basis-Point Yield Rise

Settle	Coupon	Maturity	Yield	Price	PX Change	Basis
3/30/2006	6.250%	8/15/2023	5.750%	105.4414	2.7712	2.7914
3/30/2006	6.125%	11/15/2027	5.740%	104.7266	3.1104	3.2266
3/30/2006	4.500%	2/15/2036	5.580%	84.3770	3.1037	5.0770

November 2027 become more attractive as a delivery candidate. Capitalizing on this turn of events is a three-part process. The first step is to sell the $100 million long position in the 6.25%'s of 2023. The second is to buy $100 million Treasury 6.125%'s of November 2027. The third and last step is to deliver the 6.125%'s into the contract short, thus cleaning up the position. The profit in delivering the 6.125%'s exceeds the loss in selling the 6.25% issue; the transaction nets a profit of $185,000, assuming no transaction costs. Let's work through the numbers.

In the example the 6.25% Treasuries have dropped in price to 99.9913, a loss of 2.6789 points. Selling the $100 million position therefore produces a trading loss of $2,678,918. At the same time the 6.125%'s of November 2027 can be bought in the marketplace at a price of 98.6361 and then delivered into the March contract at a price of 101.5 (referring back to Table 14.4). This transaction nets a profit of 100 * (101.5 – 98.6361) = $2,863,936. Combining the two transactions produces a total net profit of $2,863,936 − $$2,678,918 = $185,018.

The put value that produced this result derives from a duration mismatch that plays to the advantage of the contract short. A 25-basis-point increase in yields was enough to drive the price of the 6.125%'s of 2027 down 2.82%. By comparison, the shorter maturity 6.25% issue due in 2023 fell by only 2.59%. But either issue could be delivered into the same 1,000 contract short in March bonds, thus paving the way for the more volatile issue to be selected for delivery in the event of a substantial market fall.

A similar situation obtains on the upside. Some bonds can be said to have call value. Suppose for instance that instead of rising, yields had fallen 25 basis points across the board. In that case a mirror image, though a bit dicier, transaction would have produced significant profits. This can be shown by taking the same bonds as before, this time with yields reduced by 25 basis points after final contract settlement day, as in Table 14.6. This time we assume a long position of the 4.5% issue maturing February 2036, offset by a short position of 1,000 March contracts.

Note the substantial difference in duration between the 6.25%'s of 2023 and the 4.5%'s of 2036. As discussed previously, longer duration

TABLE 14.6

Impact of 25-Basis-Point Yield Reduction

Settle	Coupon	Maturity	Yield	Price	PX Change	Basis
3/30/2006	6.250%	8/15/2023	5.750%	105.4414	2.7712	2.7914
3/30/2006	6.125%	11/15/2027	5.740%	104.7266	3.1104	3.2266
3/30/2006	4.500%	2/15/2036	5.580%	84.3770	3.1037	5.0770

bonds are apt to be more volatile than shorter duration issues. For instance, a 25-basis-point downward shift in yields sends the price of the 6.25%'s of 2023 up about 2.7712 points or 2.7%; at the same time a 25-basis-point drop in yields sends the longer duration 4.5%'s of 2036 up by 3.1037 points or 3.82%, substantially more than the shorter duration 6.25%'s. So far so good; delivery of $100 million eligible bonds still has to be made to close out the trade.

A holder of $100 million 4.5%'s has some interesting options at this stage. The simplest is to swap out of the $100 million 4.5%'s, replace them with $100 million 6.25%'s of 2023, and then deliver the 6.25%'s. That maneuver would net a profit of $312,259. The sale of $100 million 4.5%'s up 3.1037 points produces a gain of $3,103,660. The 6.25%'s are then purchased in the marketplace at their new price of 105.4414 and delivered against March contracts at 102.65, the price established when the contract went off the board. That produces a loss of $100 * (102.65 – 105.4414) = $2,791,401. Combined, this series of transactions results in a net gain of $3,103,660 – $2,791,401 = $312,259.

Another approach would be to transform the position into a yield curve trade. This is a multistep process. The first step is to borrow $100 million 6.25%'s of 2023 to deliver into the contract to clean up the futures position. The next step is to sell enough 4.5%'s into the market so that the remaining position is evenly weighted against the $100 million short established in the 6.25%'s. In this case the DV01 hedge ratio of 4.5%'s to 6.25%'s is $1,108.47 ÷ $1,241.46 ≈ 0.89. A sale of $11 million of the 4.5%'s would produce a DV01 weighted position of long $89 million 4.5%'s, short $100 million 6.25%'s. The mark-to-market P&L would remain the same, but with the potential for additional profit (or loss) depending on shifts in the yield curve.

POTENTIAL PITFALLS MAKING DELIVERY

Note the lack of symmetry in the above examples. A 25-basis-point rise in yields produced profits of $185,000. On the other hand, a 25-basis-point fall in yields resulted in a far more substantial profit of $312,000. In part it is because the second trade is more risky. The first position, long $100 million 6.25%'s of 2023 and short 1,000 March bonds, has limited risk. The basis risk is minimal since the 6.25%'s are trading at a premium of only 0.0202 over delivery value, or about $20,200 for the whole position. If worse comes to worst, that's about the extent of the risk. Even a substantial jump in financing rates would have minimal impact since the short could declare early delivery.

On the other hand, there is substantially more risk (albeit hidden) in the second trade, long $100 million 4.5%'s of 2036, short 1,000 March bonds. First, there is no guarantee that the yield spread will stay constant between the 4.5% and 6.25% bonds. Note from Table 14.6 that the yield spread between the 4.5% and the 6.25% bonds is negative; the longer bonds yield substantially less than do the shorter ones. A widening of the yield spread (to become more negative) would redound to the benefit of the 4.5%'s, but a narrowing of the spread would work against the trade. And with respect to delivery, there is substantial room for downside risk because the delivery premium on the 4.5% bonds is very high at 1.9734 points (again referring back to Table 14.4). In fact, if the 4.5% bond did not command a liquidity premium for its status as the current long bond, it would be the cheapest-to-deliver security.

But it isn't simply a question of yield spreads. The trade revolves around a combination of yield spread and market direction. Par for par, the 4.5%'s are more volatile than the 6.25%'s. Consequently the 4.5%s can outdistance the 6.25%s on a price basis even though the yield spread is changing. Therefore, the trade can still be profitable on the upside if prices rise faster than yield spreads narrow. But if prices fall, the arbitrage position can be caught with substantial losses, particularly if yield spreads narrow in favor of the shorter maturities.

A second potential pitfall with making a short delivery of a bond is the distinct possibility of having lots of companies looking to buy or borrow that bond precisely because it is the cheapest-to-deliver. The consequences can range from annoying to nightmarish. For instance, if there are too many other people looking to buy or borrow the 6.25%'s to make delivery, an all-too-predictable sequence of events may occur leading to a squeeze. The first warning sign is likely to occur in the so-called specials market, in which bonds in scarce supply command premium borrowing rates. For short periods of time, this is manageable. The real problem arises when bonds are in such short supply that they cannot be borrowed at all. Notwithstanding borrowing difficulties, the contract short absolutely has to make delivery to satisfy the requirements of the contract rules.

There has never been a failure to deliver at the Chicago Board of Trade. The penalties for a failure to deliver are so onerous that traders shudder at the mere thought of them. For instance, failure to deliver against the bond contract calls for a fine of 1 point per day per contract, which would be a cool $1 million per day in the example used here. Faced with the prospect of a fail, traders will opt to deliver the next-cheapest bond in the delivery basket. But that is a generally unsatisfactory solution since the next-cheapest-to-deliver bond may be substantially more expensive than the target issue. Resorting to that strategy as the least bad option may result in substantial losses. To avoid this problem, most professional traders make sure they have extra bonds borrowed ahead of time to be used in a pinch.

STRADDLE VALUE

The strategies examined thus far seek to exploit either call or put values embedded in Treasury futures contracts. Another tactic that may come in handy is playing their straddle value. This can be done by using combinations of cash bonds hedged against a single contract. To explore this possibility, we continue with the price data and market scenarios we used in the previous examples. However, in this instance the structure of the arbitrage position is one in which the long side is a combination of two issues totaling a par value of $100 million bonds versus a short position of 1,000 March contracts.

For this part of the analysis assume a long position that consists of a combination of Treasury 6.25%'s of 2023 and Treasury 6.125%'s of 2027. The position has straddle value because of a confluence of two factors. First, the basis of each bond is relatively narrow, so delivery of either is plausible. Second, the 6.125% bonds are more volatile than are the 6.25%'s, which allows the 6.25% bonds to anchor the position with relatively low risk. If the right circumstances unfold, there is always the possibility of swapping into the 6.125%'s for delivery purposes.

Consider the positions and scenarios displayed in Table 14.7. The final settlement price for the expiring March bond contract is set at 100. Treasury 6.25%'s are trading at 102.6702 to yield 6%. Treasury 6.125%'s are trading at 101.6162 to yield 5.99%. Their respective bases are 0.0202 and 0.1162. As before there is a short position of 1,000 March contracts. However, this time the long side of the position is not split evenly; instead it consists of $65 million par value 6.25%'s of 2023 and $35 million par value 6.125%'s of 2027, for a total of $100 million cash bonds. Now let us examine the P&L impact of three different, but plausible scenarios. In the first, rates are unchanged. In the second, market rates rise by 25 basis points. In the third, rates fall by 25 basis points. The time frame extends from the day the final settlement price is established and the last delivery day. The results of these scenarios are displayed in Table 14.7.

In the event market rates stay unchanged and the strategy is to simply deliver off the combined position of $65 million 6.25% and $35 million 6.125% bonds, a loss of $53,834 will be incurred. That is not the optimal strategy. The loss can be mitigated (although not eliminated) by swapping. If market rates stay unchanged the correct strategy is to sell out the $35 million 6.125%'s at 101.6162, buy a like amount of 6.25%'s at 102.6702 and then deliver off $100 million 6.25%'s at 102.65 into the 1,000 contract short position. That will result in a $20,200 trading loss because the delivery price (102.65) is lower than the market price (102.6702) by 0.0202 per million.

The $20,200 risk can be thought of as option straddle premium. To see this consider what happens in the other two plausible scenarios in

TABLE 14.7

Unlocking Straddle Value

Scenario	**Issue Position**	**USH6 –1,000**	**6.25% 8/2023 $65 Million**	**6.125% 11/2027 $35 Million**					
			USH6 100						
	Settle	*Coupon*	*Maturity*	*Yield*	*Price*	*Factor*	*Delivery*	*Basis(Decimals)*	*Delivery*
Scenario 1:	3/22/2006	6.250%	8/15/2023	6.000%	102.6702	1.0265	102.650	0.0202	–$13,130
No Change	3/22/2006	6.125%	11/15/2027	5.990%	101.6162	1.015	101.500	0.1162	–$40,670
								P&L	**–$53,800**
		Deliver 6.125%							
	Settle	*Coupon*	*Maturity*	*Yield*	*Price*	*PX Change*	*Basis*	*Action*	*Gain /Loss*
Scenario 2:	3/30/2006	6.250%	8/15/2023	6.250%	99.9913	–2.6789	-2.6587	Sell	–$1,741,285
Rates Rise	3/30/2006	6.125%	11/15/2027	6.240%	98.6361	–2.9801	-2.8639	Buy & Deliver	$1,861,535
								P&L	**$79,580**
		Deliver 6.25%							
	Settle	*Coupon*	*Maturity*	*Yield*	*Price*	*PX Change*	*Position*	*Action*	*Gain /Loss*
Scenario 3:	3/30/2006	6.250%	8/15/2023	5.750%	105.4414	2.7712	$65,000,000	Buy & Deliver	–$976,990
Rates Fall	3/30/2006	6.125%	11/15/2027	5.740%	104.7266	3.1104	$35,000,000	Sell	$1,088,640
								P&L	**$98,520**

which rates shift either upward or downward. A 25-basis-point increase in yields would drive the 6.125% bonds down faster than the 6.25%'s. The correct strategy in that case would be to swap out of the $65 million 6.25% bonds into the same amount of 6.125%'s of 2027. The 6.125%'s would then be delivered into the 1,000 short. That series of transactions would result in a profit of $79,580, as shown in Table 14.7.

On the other hand, if yields drop 25 basis points across the board, the price of the 6.125% bonds will rise faster than the 6.25% bonds. At that point the correct strategy would be to swap out the 6.125% bonds and into the 6.25% bonds. After that trade is executed, the arbitrage account delivers $100 million 6.25% bonds into the contract short. The result, also displayed in Table 14.7, is a profit of $98,492. Here it is important to note an upside bias. The potential profit is greater on the upside ($98,492) than the downside ($79,580).

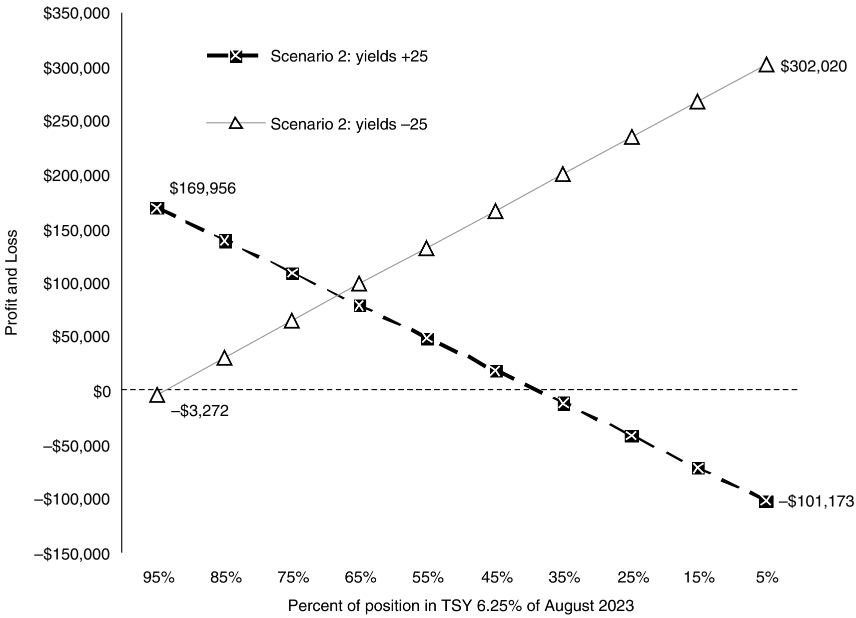
The asymmetry of returns suggests the importance of conducting scenario analyses to evaluate the potential risks and rewards of various position weightings. One way to do this is to calculate and plot a P&L based on a range of plausible position weights and market outcomes. For example, in the case at hand we analyzed the P&L impact of no change, a rate rise, and a rate fall, but with static weights. The analysis can be extended by calculating the P&L for a rate change of ±25 basis points accompanied by different position weights. The results can then be plotted to graphically illustrate how the P&L varies under different market scenarios, as in Figure 14.1.

The graph presents a dynamic look at the P&L impact of different weighting scenarios, assuming a shift in rates of ±25 basis points. The respective slopes of the graphs under the bullish and bearish scenarios make clear the nature of the risk/reward trade-offs. A more bullish posture weights the position more heavily toward the longer duration 6.125% bonds of 2027. A more bearish view would lean toward the 6.25% bonds of 2023. It can also be seen that there is greater potential profit on the bullish side, about $300,000 at the maximum compared to about $170,000 for the bearish weighting. But the extent of potential loss from the most conservative (and bearish) weighting is slight, whereas the risk in the more aggressive bullish weighting is far greater—about $100,000 versus only $3,000.

A weighting around the midpoint, with 65% in short-duration bonds and 35% in long-duration bonds, affords the possibility of making money in either market direction, which is ideally what straddles are set up for. Accordingly, a weighting around the midpoint maintains flexibility and allows for substantial loss mitigation by issue swapping in the event that rates are stable. In addition there is always the possibility that yield spreads could change between the 6.25% and 6.125% issues, the effects

FIGURE 14.1

Scenario Analysis: The Basis and Straddle Value



of which could work in either direction. But it would be unusual for a substantial change in the spread to occur when the issues are so similar.

For simplicity's sake, the examples used here assumed very narrow cash/futures spreads. But in the real world, basis spreads are subject to considerable variation. Two factors drive variation: carry and uncertainty over the bonds most likely to be delivered. When there is little ambiguity about the bonds most likely to be delivered, basis spreads (after carry) are typically narrow. But when there is little certainty about the likely deliveries, basis spreads tend to widen. The more uncertainty, the wider the basis spread is likely to be.

Pricing the delivery option should, in theory, reflect the probability of a shift in the cheapest-to-deliver bonds. In this respect the delivery option value is analogous to deep out-of-the-money options. But deep out-of-the-money options are often underpriced because people tend to systematically underestimate the probability of rapid and unexpected change. As a result, basis-trading opportunities sometimes present themselves when option value can be purchased cheaply. Sometimes option value asserts itself well before the delivery month through basis expansion. It can occur when bond rates move rapidly and decisively enough to cause a marked shift in the cheapest-to-deliver bond or bonds.

SHIFTS IN DELIVERY GRADES

Up to this point the discussion has centered on potential arbitrage profits embedded in the delivery process. But there is also potential for capturing significant arbitrage profits well before the delivery process kicks in. That source of arbitrage opportunity is the ever-present possibility of a substantial change in the cheapest-to-deliver bonds. When a delivery grade shift is slight because the bonds are similar, say from Treasury 8.125% of August 2021 to Treasury 8%'s of November 2021, the impact is slight. But when the change is substantial, for instance from the 8.125%'s of 2021 to the 4.5%'s of 2036, the P&L impact can be significant.

The possibility in a shift in delivery grades arises because there is a mismatch between static CBOT conversion factors and bond duration dynamics. The result is that CBOT conversion factors fail to take convexity into account. As market interest rates change, the sensitivity of bond prices to rates changes as well. Moreover, price/rate sensitivities of bonds vary by coupon and maturity. As a result, hedge ratios between and among bonds vary as market levels change. But since CBOT conversion factors are static, they fail to capture this dynamic. Futures hedge ratios therefore need to be adjusted dynamically to keep them current with market conditions. More to the point, substantial changes in bond rates can lead to a rapid change in the cheapest-to-deliver bond and how the futures contract tracks it.

To see this, consider a rather extreme example. Suppose there are only two bonds in the world: the 8.125%'s of August 2021 and the 4.5%'s of 2036. A contract long of 100 contracts can expect delivery of $10 million of either bond or some combination of both. But a comparison of the respective DV01s of the two bonds implies that they have very different expected volatilities. As Table 14.8 shows, the DV01 of Treasury 8.125%'s of August 2021 is much smaller than the DV01 of the longest Treasury, the 4.5%'s of 2036. The DV01 hedge ratio is 1.18375. It takes $11.8 million of the 8.125%'s to equal $10 million of the 4.5%'s of 2036.

But CBOT conversion factors seem to imply that the hedge ratio runs in the opposite direction—that the 4.5% bonds are *less* volatile than the 8.125%'s. For instance, against June 2006 Treasury futures, the factor

TABLE 14.8

Comparison of DV01s of Two Representative Bonds

Settle	Coupon	Maturity	Yield	Price	DV01	DV01 Hedge	Conversion Factors
3/20/2006	8.125%	8/15/2021	4.90%	134 19/32	1310.50	0.844684	1.2083
3/20/2006	4.500%	2/15/2036	4.73%	9611/32	1551.46	1.1838747	0.7937

weighting for $10 million of the 8.125% issue would require the sale of 121 June bonds. By contrast, *even though they are more volatile, factor weighting of $10 million of the 4.5% issue would require the sale of only 79 June bond futures*. There is a factor/DV01 mismatch. What to do about it?

The general rule of thumb for dealing with this situation is to use factor weights for the cheapest-to-deliver bond, and then back into adjusted factor weights for all other bonds using DV01s. Assume for the moment that the 8.125% bonds are the cheapest deliverable grade. Hedging $10 million of them with futures contracts is straightforward; it requires selling 1.2083 * 100 = 121 contracts. On the other hand, hedging $10 million 4.5%'s, which are very far away from feasible delivery, is a two-step process. The first step is to convert the 4.5%'s into equivalent units of 8.125% bonds due in 2021. The second step is to calculate a hedge ratio based on those equivalents.

Based on the calculated DV01s in Table 14.8, we can see that it takes $11.8 million of the CTD 8.125% bonds of 2021 to equal $10 million of the 4.5%'s of 2036. Therefore, expressed in CTD equivalents, the hedge is 1.2083 * 118 = 143 contracts, which amounts to a hedge ratio that is a whopping 81% greater than direct factor weighting alone would suggest.

The mismatch of DV01 weights and conversion factor values is a source of potential arbitrage profit. Futures contracts tend to most closely track the security that is cheapest-to-deliver. It is hard to argue with the logic of the market. At any given point the security that is currently the cheapest-to-deliver is also the most likely to actually wind up being delivered in the end. But the CTD is not set in stone; it is probabilistic. The bond futures contract can therefore be said to possess certain optionlike qualities which need to be taken into account. Changes in market yields can cause a shift of the cheapest-to-deliver bond, which in turn can cause the contract to change the bond it tracks most closely, which in turn can change the volatility of the contract.

To explore this, we will consider an example of what can happen with a substantial (and plausible) parallel shift in market yields over time. For simplicity's sake we posit the universe of deliverable bonds consists of six issues ranging from Treasury 8.125% of August 2021 to Treasury 4.5%'s of 2036, displayed in Table 14.9. Prevailing market yields from mid-March of 2006 serve as the starting point in the analysis. As of settlement day March 20, the bonds range in yield from 4.89% to 4.72%, and the June bond contract is trading at 111 3/8 as shown. We hypothesize that by mid-June yields have risen by 125 basis points across the board. Yields for the selected issues in June stand between 6.39% and 6.22%. June futures contracts, which were trading at 111 12/32nds in mid-March, have dropped to 95 14/32nds, a fall of almost 16 points.

As Table 14.9 shows, a parallel shift in the curve on the order of 125 basis points would cause the cheapest-to-deliver bond to shift from

TABLE 14.9

A Shift in the Delivery Grade

USM6 Settle	Coupon	Maturity	Yield	Price	Factor	Delivery	Basis
3/20/2006	8.125%	8/15/2021	4.89%	134 23/32	1.2083	134 18/32	5/32
3/20/2006	7.625%	2/15/2025	4.89%	133 16/32	1.1801	131 14/32	2 2/32
3/20/2006	6.125%	11/15/2027	4.88%	116 17/32	1.0148	113 1/32	3 16/32
3/20/2006	5.250%	2/15/2029	4.87%	105 7/32	0.9081	101 4/32	4 2/32
3/20/2006	5.375%	2/15/2031	4.81%	108 5/32	0.9203	102 16/32	5 21/32
3/20/2006	4.500%	2/15/2036	4.72%	96 16/32	0.7937	88 13/32	8 3/32

USM6	95 14/32							Basis
Settle	Coupon	Maturity	Yield	Price	Factor	Delivery	Basis	Change
6/20/2006	8.125%	8/15/2021	6.39%	116 21/32	1.2083	115 10/32	1 12/32	1 7/32
6/20/2006	7.625%	2/15/2025	6.39%	113 11/32	1.1801	112 20/32	23/32	–1 11/32
6/20/2006	6.125%	11/15/2027	6.38%	97 1/32	1.0148	96 27/32	6/32	–3 10/32
6/20/2006	5.250%	2/15/2029	6.37%	86 21/32	0.9081	86 21/32	–0	–4 2/32
6/20/2006	5.375%	2/15/2031	6.31%	88 12/32	0.9203	87 26/32	18/32	–5 3/32
6/20/2006	4.500%	2/15/2036	6.22%	76 27/32	0.7937	75 24/32	1 3/32	–7

the 8.125% of August 2021 to the 5.25% issue maturing February 2029. The 8.125% basis would expand by 1 point and 7/32nds; at the same time the basis in the 5.25%'s of 2029 collapses slightly more than 4 points. To see how this works and how it can be exploited, it is necessary to explore cash/futures convergence in some more depth.

CASH/FUTURES CONVERGENCE

It is easy enough to posit a change in yields and then calculate the effect on bond prices. But to see the impact on basis spreads, we also need to calculate the price of the futures contract, dependent on market rate levels. Estimating the impact on the bond futures contract is complicated. Unlike cash bonds, bond futures prices cannot be determined by simply applying the yield-to-maturity formula. With futures contracts there are no cash flows to discount; therefore, futures contracts do not have yields in the conventional sense. But changes in bond futures prices can be estimated because arbitrage trading will force convergence of cash and futures prices at contract expiration. If cash prices are known, the futures convergence price at expiration is assumed to be the minima of all deliverable bond prices divided by their respective conversion factors, with the basis (set to zero), as we did earlier in the chapter.

$$Convergence = Minima\left(\frac{Prices^{Deliverables}}{Factors}\right)$$

This reverse engineering rests on two assumptions. First, that cash bond prices remain unchanged. Second, that bond futures accrete to meet the cheapest delivery grade. Making all the usual assumptions of continuous pricing, zero transaction costs, and an absence of fails-to-deliver, a zero basis on the last delivery day is the only price where risk-free arbitrage is not possible. Since risk-free arbitrage is not possible in the long run, the futures price has to settle where the basis for at least one bond is zero.

We can see in Table 14.9 that dividing the market prices of the various bonds by their respective conversion factors produces a basis of 0 for the Treasury 5.25%'s of 2/15/2029 with June bond contracts priced at 95 14/32. The 8.125%'s of August 2021 are no longer the cheapest-to-deliver; that honor now belongs to the 5.25%'s of 2029. In other words, there has been a switch in the delivery grade. The rise in market yields has caused the 8.125% basis to widen from 5/32nds to 39/32nds, while the 5.25% basis has collapsed from 4 points and 2/32nds to zero.

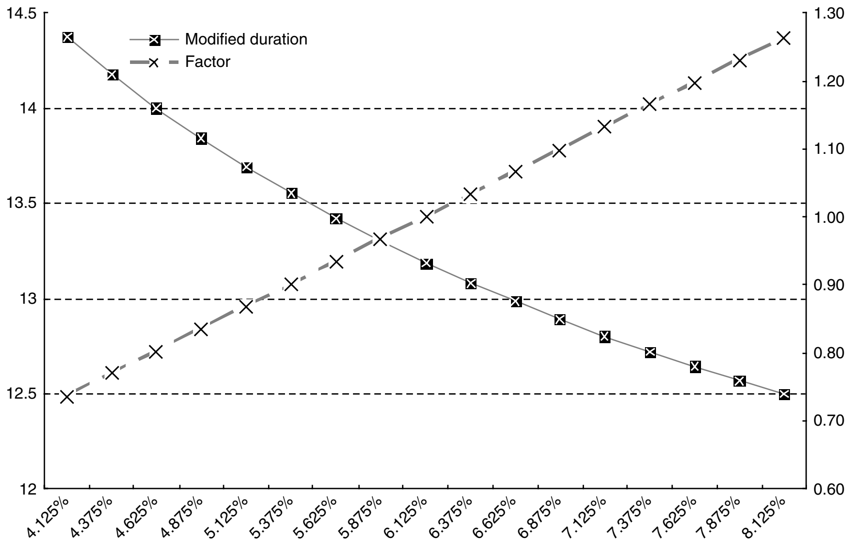
The reason for the CTD shift has to do with differences in the ways by which futures contracts and cash bonds reflect sensitivities to changes in market yields. As market yields change, cash bonds rise and fall at different rates of speed. All else equal, prices of longer-duration bonds move faster than shorter-duration bonds do. On the other hand, conversion factors for bond futures overweight the importance of a bond's coupon. As a result, long-duration low-coupon bonds are assigned conversion factors that imply smaller futures hedge ratios than for short-duration high-coupon bonds, the *opposite* of what cash market hedge ratios imply. The best way to see this is to calculate and plot conversion factors against the modified durations of hypothetical bonds with the same maturity but different coupons, as shown in Figure 14.2.

Note that as the coupon increases, so does the conversion factor. Also note that, holding the yield-to-maturity steady at 6%, the bond's modified duration falls as the coupon rises. Since a bond's price sensitivity to interest rate changes is positively correlated to its duration, Figure 14.2 correctly implies that as the coupon moves up, the bond is less price sensitive (in percentage terms) to rate changes. But the conversion factors run in the opposite direction. As the bond's coupon rises, so does the conversion factor, implying that higher coupon bonds need more rather than fewer contracts for hedging purposes—exactly the opposite conclusion one would reach using duration (or DV01s) as the benchmark for calculating hedge ratios.

Playing this mismatch between duration values (or DV01s) and conversion factors can potentially lead to large arbitrage profits resulting from

FIGURE 14.2

Conversion Factors versus Modified Duration



delivery shifts. The duration/factor mismatch is also the reason why alternative hedge ratios need to be considered for matching up cash bonds with futures contracts. The most straightforward way to analyze potential basis strategies is to observe what happens to bond prices and basis spreads when market yields move substantially in either direction. In this vein, we will continue to use the data in Table 14.9 to consider the impact of a 125-basis-point market move on various cash/futures arbitrage positions. The first strategy to be examined is long the basis.

Going "long" the basis refers to owning cash bonds and selling futures contracts short against them. When the cash bond to be hedged is the cheapest-to-deliver, the correct hedge ratio to use is the one dictated by the delivery factor. For this example we assume a position that is long $100 million Treasury 8.125%'s of August 2021. The short side of the trade is 1.2083 * 1000 = 1,208 June bond futures. The value date is March 20, 2006. In this first scenario, by mid-June the market has moved higher in yield by 125 basis points. As before, we assume continuous markets, the absence of transaction costs, and no carry profits or losses.

As Table 14.9 illustrates, a 125-basis-point rise in market yields causes the June bond contract to fall in price by just under 16 points, from 111 3/8 to 95 14/32. The bonds that constitute the deliverable universe drop in price by varying amounts, with dramatic shifts in their respective bases, implying large P&L consequences for factor-weighted arbitrage trades. For instance, the basis expansion of Treasury 8.125%'s of August 2021 against June bond futures is 1 point and 7/32nds, implying a gain of about $1.2 million for a $100 million basis position on the long side. Even though cash prices have dropped more than those of futures, the trade generates a substantial profit because the size of the futures short is more than enough to offset the relatively smaller price decline of the futures contract.

But the opposite is the case for longer-duration bonds. Their fall in price is far steeper than the factor-weighted number of futures contracts would suggest is likely. For arbitrageurs who decided to purchase long-duration bonds and then hedge with short sales of factor-weighted quantities of futures contracts, the result would be substantial trading losses. As Table 14.9 illustrates, the cheapest-to-deliver would have shifted to the 5.25% bond maturing 2029. Its basis would have fallen from slightly over 4 points to zero, implying a trading loss of over $4 million.

Had yields dropped by 125 basis points, the opposite would have happened. The longer-duration bonds would have risen in price much more quickly than the futures contract. The result would be a substantial basis expansion for the longer-duration bonds with the 8.125%'s being forced into delivery and a loss of the small basis premium. See Table 14.10.

Directionally sensitive hedge ratios capable of producing large P&L swings leave something to be desired. The usual solution to this dilemma in basis trading is to use factor weights to hedge the cheapest-to-deliver bond, and then to use DV01s to adjust the factor weights for other bonds to be hedged. This method for determining hedge ratios has the advantage of allowing contract shorts to play the contract's embedded optionality, while still providing protection against parallel shifts in market yields.

TABLE 14.10

USM6 125 22/32 Basis Impact with Yields Falling 125 Basis Points

Settle	Coupon	Maturity	Yield	Price	Factor	Delivery	Basis	Basis Change*
6/20/2006	8.125%	8/15/2021	3.64%	151 28/32	1.2083	151 28/32	0	–4/32
6/20/2006	7.625%	2/15/2025	3.64%	153 20/32	1.1801	148 10/32	5 9/32	3 7/32
6/20/2006	6.125%	11/15/2027	3.63%	136 29/32	1.0148	127 18/32	9 11/32	5 27/32
6/20/2006	5.250%	2/15/2029	3.62%	125 2/32	0.9081	114 4/32	10 29/32	6 27/32
6/20/2006	5.375%	2/15/2031	3.56%	129 20/32	0.9203	115 21/32	13 30/32	8 10/32
6/20/2006	4.500%	2/15/2036	3.47%	118 31/32	0.7937	99 24/32	19 7/32	11 4/32

* Differences due to rounding

CALCULATING ADJUSTED HEDGE RATIOS

The starting point for adjusting futures hedge ratios is the (often incorrect) assumption of a stable yield-spread relationship between the cheapest-to-deliver cash bond and all the other deliverable bonds. This allows for dynamic adjustment of futures hedge ratios that dampens P&L volatility but retains embedded optionality. However, there is a trade-off. To some degree, reduced exposure to market direction is replaced by exposure to shifts in the yield curve. To see how this works we will continue to work with the same hypothetical universe of six deliverable bonds, except this time the hedge ratios will be adjusted for differences in the DV01s of the individual bonds.

The technique for adjusting cash/futures hedge ratios to reflect cash market DV01s is straightforward. First, each bond's DV01 is calculated as discussed previously. Second, cash-cash hedge ratios are calculated for each bond versus the cheapest-to-deliver bond. Third, using these hedge ratios, each bond is expressed in terms of a CTD equivalent position. Fourth, the futures hedge for each bond is determined by multiplying the conversion factor of the cheapest-to-deliver bond by the CTD equivalent position of each deliverable bond. Let's work through the numbers, extending the current example as it is displayed in Table 14.11.

As shown in Table 14.11, the conversion factor for the CTD Treasury 8.125%'s of August 2021 is 1.2083. With a yield of 4.88%, its DV01 is $1,312. Also yielding 4.88%, the DV01 of Treasury 7.625%'s of February 2025 is $1,497. The conversion factor for Treasury 7.625% bonds is 1.1801. The DV01 hedge ratio of 8.125%'s to 7.625%'s is 1497 ÷ 1312 = 1.1408:1. With the cash hedge ratio in hand, the next step is to translate it into a futures hedge ratio. That hedge ratio will be the product of the DV01 hedge ratio and the CTD conversion factor, which in this case would be 1.1408 * 1.2083 = 1.378.

TABLE 14.11

Adjusting Factor Weights for DV01s

Coupon	DV01	Cash Hedge Ratio	Adjusted Futures DV01	Factor Weights	Difference
8.125%	$1,312.20	1.000	1,208	1,208	0
7.625%	$1,496.93	1.1408	1,378	1,180	198
6.125%	$1,478.19	1.1265	1,361	1,015	346
5.250%	$1,421.92	1.0836	1,309	908	401
5.375%	$1,526.30	1.1632	1,405	920	485
4.500%	$1,555.08	1.1851	1,432	794	638

Another way to think about this hedge ratio is to note that, based on DV01 calculations, it would take $114.1 million par value 8.125%'s of 2021 to hedge $100 million par value 7.625%'s of 2025. Hedging $114.08 million 8.125%'s with June bonds would require 114.08 * 1.2083=1,378 futures contracts on the other side. Since $100 million 7.625% bonds are the functional equivalent of $114.08 million 8.125% bonds, the correct futures hedge is 1,378 contracts. But this 1.3787:1 futures hedge is dramatically different from the 1.1801 ratio suggested by a straightforward application of delivery conversion factors. It represents a change of 14% in the hedge ratio.

As a practical matter, the best way to evaluate the use of adjusted hedge ratios is to do scenario analysis, running the market up and down 125 basis points to see what happens to a hypothetical P&L over time. Results from doing this are displayed in Table 14.12. As before, we assume continuous markets, no transaction costs, no financing profits or losses, and parallel shifts in the yield curve.

TABLE 14.12

Scenario Analysis with DV01 Adjusted Factor Weights

	USM6	95 14/32	–15 30/32			
	Coupon	**Fall in Cash Prices**	**Short Futures**	**Cash P&L**	**Futures P&L**	**Net P&L**
	8.125%	18 2/32	1,208	–$18,052,572	$19,257,281	$1,204,709
	7.625%	20 5/32	1,378	–$20,153,305	$21,968,183	$1,814,879
Rates Rise	6.125%	19 16/32	1,361	–$19,485,679	$21,693,272	$2,207,594
	5.250%	18 18/32	1,309	–$18,550,636	$20,867,407	$2,316,770
	5.375%	19 25/32	1,405	–$19,769,862	$22,399,278	$2,629,417
	4.500%	19 21/32	1,432	–$19,653,482	$22,821,593	$3,168,112
	USM6	**125 22/32**	**14 10/32**			
	Coupon	**Rise in Cash Prices**	**Short Futures**	**Cash P&L**	**Futures P&L**	**Net P&L**
	8.125%	17 5/32	1,208	$17,157,040	–$17,293,794	–$136,754
	7.625%	20 4/32	1,378	$20,124,137	–$19,728,290	$395,847
	6.125%	20 12/32	1,361	$20,378,621	–$19,481,409	$897,212
Rates Fall	5.250%	19 27/32	1,309	$19,840,245	–$18,739,750	$1,100,495
	5.375%	21 15/32	1,405	$21,471,950	–$20,115,430	$1,356,520
	4.500%	22 16/32	1,432	$22,487,495	–$20,494,686	$1,992,809

As before, Table 14.12 shows what happens when yields change by 125 basis points. The P&L on each of the cash bonds is calculated for a $100 million par value position on the long side of the market. However, this time short positions in futures are adjusted to reflect DV01 equivalents (with respect to the CTD 8.125% bonds) rather than strict factor weights. The P&L for each bond's hedge is the product of the short in futures and the change in the futures price. The net P&L is the sum of cash position gain or loss and the offsetting futures hedge.

Note that a long position in the 8.125% basis would produce a trading loss equal to the basis spread premium if the market were to either remain unchanged or to fall in yield by 125 basis points. That potential loss can be thought of as the price of insurance, functionally equivalent to option premium. However, a 125-basis-point parallel yield curve shift *in either direction* produces a net profit for a long basis position in any other bond. The reason for this phenomenon is that the cash bonds tend to be more positively convex than futures contracts. All else equal, a large increase in yields will cause futures prices to fall faster than a large drop in yields will cause futures prices to rise. That is clearly the case in the current example. A 125-basis-point rise in yields causes a drop of 15 points and 30/32nds in futures prices, but a fall in yields of the same magnitude causes a price rise of only 14 points and 10/32nds.

Conversely, the longer-dated bonds in the example are positively convex. From the 6.125%'s of 2027 on out, prices rise faster than they fall for a large change in yields. By implication, long-side basis trades that pair off positively convex cash bonds with futures contracts will tend to yield arbitrage profits for large parallel changes in market yields when the trades are yield weighted.

THERE'S STILL NO FREE LUNCH

Sadly enough, despite appearances there is no free lunch offered by basis trading. Large changes in market yields are typically accompanied by shifts in the slope of the yield curve. These yield curve shifts can easily eat up convexity gains, and then some. In part because of their positive convexities, the longer-dated bonds trade at premiums—which can easily be given up. The 4.5%'s of 2036 in this example trade to yield 4.72% versus 4.89% for the 8.125%'s of 2021—a premium of 17 basis points. If the premium were to fall to 10 basis points from 17, the price drop of the 4.5%'s would be slightly over 1 point, which would produce a trading loss of over $1 million, all else equal.

The magnitude of potential loss in going long the CTD basis is far less than the loss potential of hedging other bonds against futures contracts. When the CTD basis is particularly narrow, the delivery option acts as a kind of P&L backstop. On the other hand, the odds-on bet is that the CTD will remain the CTD; it will probably wind up being delivered into the contract, with a resulting loss of option premium. In the current example, the 8.125% basis would yield a loss option premium if yields were to remain stable or fall.

SUMMARY

At its core, basis trading in Treasury bonds is an implicit forecast of the spread between a Treasury bond and a CBOT futures contract into which it is eligible for delivery. Spreads between cash Treasuries and their companion futures contracts are ultimately determined by short-term financing rates, the level of market rates, the shape of the yield curve, and optionality embedded into the delivery process. Basis premiums in excess of expected financing profits are similar to options premiums.

There are many different types of arbitrage opportunities offered by basis trading. Opportunities can arise during the delivery process; they can stem from mismatches between conversion factors and DV01 weights; they can be the result of large changes in market yields that cause a shift in the cheapest-to-deliver bonds. Finally they can arise from changes in financing rates.

Typically, Treasury bond bases trade at premiums over carry value. When there is greater uncertainty over the most likely delivery grade, futures contracts are liable to trade at a greater discount to the cash markets, raising basis premiums. Market volatility, which increases uncertainty over the cheapest-to-deliver bonds, also serves to widen spreads between cash and futures. As a result, it can be reasonably said that basis spreads possess certain optionlike properties. Deciding whether to approach trading the basis from the long side, the short side, or not at all, depends to some degree on basis premiums (or lack of) and scenario analyses, including projections of market yields, the shape of the curve, position risk, and the payoffs associated with all possible outcomes. Trading the bond basis can be complex, subtle, and highly nuanced. There is no substitute for using scenario analysis to evaluate potential arbitrage positions.

NOTES

[1] See The Federal Reserve of New York, *Administration of Relationships with Primary Dealers* online at: http://www.ny.frb.org/markets/pridealers_policies.html

[2] The specific rules for each Treasury contract are available at the CBOT Web site: www.cbot.com

[3] The precise formula for calculating conversion formulas is available from the CBOT. It isn't really necessary to do all the calculations. The CBOT, Bloomberg, and other vendors publish conversion factor tables on the Web, and they are available free of charge.

[4] Different clearinghouses may require earlier notice times, depending on firm policy.

This page intentionally left blank

CHAPTER 15

Synthetics, Contract Rolls, and Intermarket Spreads

The physical delivery requirement built into the design of CBOT Treasury futures contracts combined with easy market access and low transaction costs ensures minimal tracking error against the cheapest-to-deliver bond. In addition, cash market arbitrage trading keeps spreads among individual cash issues in line with each other. As a result, Treasury futures do a good job of tracking the entire spectrum of notes and bonds along the yield curve. These properties make Treasury futures very effective tools for managing portfolio strategy while minimizing transactions costs. For instance, portfolio duration can be managed using futures as an overlay; financing rates can be locked up, exposure can be shifted along the yield curve, shifts in monetary policy can be anticipated, and leverage can be added to or subtracted from cash positions—all without executing any cash market transactions.

The key to executing these strategies is to use Treasury futures contracts, whether singly or in combination, to replicate the behavior of cash market notes and bonds. That way they can be used to augment, offset, complement, or adjust cash market positions and trading strategies. For instance, note, bond, and fed funds futures can be used in combination to synthetically replicate the yield curve, in whole or in part, and to trade implicit interest rate differentials between calendar months.

THE SYNTHETIC YIELD CURVE

By design, Treasury futures contracts replicate the market behavior of issues at key inflection points along the yield curve, namely 2-year, 5-year, 10-year, and long bonds. The contract specifications are written so that there is a “bucket” of delivery eligible securities for each maturity sector; thus the contract implicitly captures the returns characteristics of the entire sector rather than a single issue. The securities within each of the delivery buckets are unique to that sector. They do not overlap. But arbitrage trading across sectors keeps prices close so that stringing together sequences of Treasury contracts can approximate the behavior of the yield curve in its entirety.

To create a synthetic yield curve using Treasury futures, the first step is to translate them into cash market equivalents. Doing so does not directly create yield spreads between different futures contracts the way yield spreads exist between cash notes and bonds. After all, since futures contracts lack cash flows, they don't have a yield to maturity. Nevertheless, weighting futures contracts in terms of cash market equivalents allows them to closely mimic the market behavior of cash notes and bonds at key inflection points along the curve. Used in combination, they can reproduce the behavior of the Treasury yield curve as a whole, or parts of it. As a result, Treasury futures can be used to execute yield curve arbitrage transactions without the necessity of borrowing and lending individual securities in the RP and reverse RP markets.

The first step is to calculate futures equivalent DV01s for the strip of 2-year, 5-year, 10-year, and bond futures contracts. This involves determining the cheapest-to-deliver (CTD) cash Treasury security for each of the maturity delivery buckets. When that is done, their DV01s are calculated. Those DV01s are then translated into their futures equivalents, as discussed in previous chapters. Dividing the cash note or bond DV01 by its respective delivery conversion factor does the trick; the quotient is a futures equivalent DV01. Assigning equivalent DV01s to futures contracts in this manner provides a way to use them to mimic (and capitalize on) changes in the cash yield curve. This can easily be done without actually entering the cash markets.

To see how this works, we will work through an example using the yields, cash DV01s, and conversion factors of the CTD notes and bonds versus March 2006 futures contracts. The cash DV01s are then converted into futures equivalent DV01s (FE01) as displayed in Table 15.1.

TABLE 15.1

Futures Equivalent DV01s

	Cash DV01		Factor		FE01
2-Year	$184.56	÷	0.9733	=	$189.62
5-Year	$386.11	÷	0.9226	=	$418.50
10-Year	$577.45	÷	0.8937	=	$646.13
Bonds	$1,355.98	÷	1.2083	=	$1,122.22

TABLE 15.2

Hedge Ratios: Cash Versus Futures

Cash DV01 Weights					Futures Equivalent DV01 Weights			
2-Year	**5-Year**	**10-Year**	**Bonds**	**Maturity**	**2-Year**	**5-Year**	**10-Year**	**Bonds**
1.00				2-Year	1.00			
2.09	1.00			5-Year	2.21	1.00		
3.13	1.50	1.00		10-Year	3.41	1.54	1.00	
7.35	3.51	2.35	1.00	Bonds	5.92	2.68	1.74	1.00

Once DV01s have been calculated, it is a simple matter to create a matrix of hedge ratios. The procedure is to divide each bond's DV01 by every other bond's DV01. The result is a matrix that contains the hedge ratios of each bond for every other bond. Similarly, dividing each futures equivalent DV01 by every other futures equivalent DV01 provides the correct yield-weighted hedge ratio of each futures contract for every other futures contract. But these are static hedge ratios. Over time the correct weights will change, so the matrix has to be periodically updated. Table 15.2 is an example of yield-weighted hedge ratios for both cash and futures contracts using this methodology.

For comparison purposes, Table 15.2 places cash and futures weightings side by side. The left side is cash bonds; the right side is futures. The matrices can be read either across the rows or down the columns. For instance, going down the column labeled 2-year on the cash side, it can be seen that the hedge ratio for 5-year notes is 2.09:1. It takes $2.09 million par value 2-year notes to equal the P&L exposure of only $1 million of 5-year notes. The 5-year/10-year hedge ratio (1.50:1) can similarly be found by reading down the rows and across the columns.

Note the variation between cash and futures hedge ratios for the same nominal maturities. Sometimes the difference is relatively slight, as in the 5-year/10-year spread. But sometimes the difference is large, as it is in the 2-year/bond spread. The difference is due to the mismatch between CBOT delivery factors and the duration weights of the underlying bonds, as discussed in previous chapters. Moreover, the correct weights (to the extent that there are correct weights) can and do change with the level of the market, slope of the curve, and the CTD security. To avoid being blindsided, it is important to monitor position weights and keep them up to date.

With futures equivalent DV01s calculated for the full array of Treasury futures, each contract can be arbitraged against every other contract—or combination of contracts in a market neutral fashion. The precise combinations to be used depending on market view, both with respect to the shape of the yield curve and the general level of rates. As always, the best way to evaluate potential trades is by scenario analysis. For example, we can use scenario analysis to see how efficiently a synthetically constructed yield curve trade reproduces actual cash market results. An efficiently constructed futures trade designed to substitute for the cash market ought to produce the same P&L for the same set of circumstances.

Suppose, for instance, that the market outlook is for the yield curve to become more steeply sloped. Regardless of the level of rates, the forecast is that the spread between long and short rates will widen (become more positive than it is at present). Acting on that view implies buying shorter-dated securities and selling longer-dated ones on a yield-weighted basis, or "buying the curve." For the sake of example we test buying cash 5-year notes and shorting cash 10-year notes at a ratio of 1.5:1, based on the DV01s calculated previously. Then to test the accuracy of the weightings, yields of both bonds are instantaneously reduced by 5 basis points so the P&L impact can be evaluated. To the extent that the 1.5:1 weighting is correct, the P&L impact should be minimal.

Table 15.3 shows that this is indeed the case. The trade is executed by going long $15 million par value 5-year notes, the 3.875%'s of May 15, 2010; against that long is a short sale of $10 million 10-year notes, the 4%'s of November 15, 2012. Note that neither the 5-year nor the 10-year exactly lives up to its moniker. They are simply the CTD notes that correspond to 5- and 10-year Treasury futures contracts.

Calculating before and after prices of the notes adjusted for a 5-basis-point parallel change in yields shows that on a DV01-weighted basis, the P&L is only negligible. As shown in Table 15.3, a nickel drop in yields reaps a gain of $28,985 on a long position of $15 million 5-year notes versus a loss of $28,912 on the 10-year short. The difference is only $73, indicating that the position weights are correct.

This result can be replicated with futures contracts weighted by futures equivalent DV01s. Referring to the weights calculated earlier, the futures equivalent weight is 1.54:1 for trading 5-year contracts on the long side against a short sale of 10-year contracts. Therefore, a mirror image futures position requires a long position of 154 Treasury 5-year futures against a sale of 100 Treasury 10-year futures.

To evaluate trade efficacy, we will again move the cash markets by the same 5 basis points. From that starting point we back into where the futures prices would have moved. How to do that? Calculate the basis spread for each note, leave the spread unchanged, and change the cash market price. That will drag the futures market up (or down as the case may be) along with cash, allowing us to calculate new and accurate prices

TABLE 15.3

Going Long the 5- to 10-Year Spread

Position	Issue	Settle	Coupon	Maturity	Yield	Price
Long	Fives	1/20/2006	3.875%	5/15/2010	4.33%	98.221
$15M	Fives	1/20/2006	3.875%	5/15/2010	4.28%	98.414
					P&L	$28,985
Short	Tens	1/20/2006	4.000%	11/15/2012	4.36%	97.891
$10M	Tens	1/20/2006	4.000%	11/15/2012	4.31%	98.181
					P&L	$28,912

TABLE 15.4

Futures Curve Trade: 5s–10s

	Basis	Cash	Factor	Futures
Long 154 FVH6	–1/32	98.2208	0.9226	106 16/32
	–1/32	98.4140	0.9226	106 23/32
				$32,254.50
	Basis	**Cash**	**Factor**	**Futures**
Short 100 TYH6	1/32	97.8914	0.8937	109 16/32
	1/32	98.1806	0.8937	109 26/32
				–$32,350.86

of the futures contracts. As Table 15.4 shows, when this is done, the P&L is once again minimal. The futures transaction is a mirror image of cash. By implication, Treasury futures can be used to synthetically re-create all, or parts, of the yield curve, on either the long or short side of the market.

Note too the basis spreads in Table 15.4. The 5-year basis is negative; the 10-year is positive. The 5-year basis is negative because its yield to maturity is less than the average expected fed funds rate to the end of the March delivery month. The 10-year is not, only partly because 10-year yields are slightly higher. Mostly it is because the 10-year note at this point commands a special rate in the RP market due to high borrowing demand. Differential basis spreads are important because their existence contradicts a widely held but erroneous belief that using futures contracts obviates the need to consider the details of borrowing and lending securities. Not so: It's all embedded in the price.

EMBEDDED FINANCING

The current example, which uses actual market prices, is a case in point. As noted before, the 5-year note basis is negative to the tune of one 32nd of a point, while the 10-year basis is positive by one 32nd. It's easy enough to tell why the 5-year basis is negative; carry is negative. The financing rate is 4.5%, but the notes yield only 4.33%. But that doesn't explain why the 10-year basis is positive. It yields 4.36%, only 3 basis points more than the 5-year, not enough to account for the difference by a long shot.

The reason the 10-year basis is positive is that its RP rate is closer to 4.25% than the 4.5% fed funds rate. Apparently there are lots of short sellers in the 10-year note, stoking borrowing demand in the RP markets. The 10-year note therefore commands a premium in the marketplace. Holders can borrow fed funds at 4.25% in return for lending out 10-year notes as collateral for the loan. But 5-year notes are in plentiful supply; holders have to pay 4.5% to borrow money to finance their positions.

This difference in borrowing rates plays itself out (in part) by differences in the basis. Short sellers of 10-year futures don't avoid the borrowing premium for the 10-year note; they pay indirectly by having to sell the futures contracts at prices that are lower than they would ordinarily be. Similarly, longs in 5-year note contracts do not escape the negative carry of the cash market. They pay indirectly when they buy the futures contracts at a premium to where they would otherwise trade. In fact, in the current example, if the cash market remains unchanged as contract expiration approaches, the 10-year contract will converge upwards while the 5-year will converge downwards. That implies a price decline on the long side of the trade and a price rise on the short side, all else equal.

There are additional factors to be considered when the use of futures is contemplated as a substitute for cash securities. A particularly important one, discussed previously, is that the CTD cash security can change. By definition, when that happens, the futures contract begins to track a different cash bond, quite possibly changing the implicit DV01 of the futures contract in the process. That eventuality would cause changes in volatility weights between cash Treasuries and their underlying contracts. Therefore, it's important to make sure that futures weights are kept up to date by making sure they always track the CTD notes and bonds.

A final caveat has to do with watching calendar spreads. When futures contracts roll into a delivery month, all the delivery options embedded in the contracts come into play. Further, liquidity begins to seep over to the new front-month contract and out of the spot month. Around that point most traders roll their positions out of the spot (or soon to be spot) month and into the new front-month contract. The question is: How are these contract rolls, or calendar spreads, priced and traded?

TREASURY BOND ROLLS

Chicago Board of Trade Treasury futures contracts expire quarterly. Originally designed to roughly coincide with regularly scheduled Treasury bond auctions, contract expirations occur every March, June, September, and December. Beginning in the 1979s as part of a policy initiative to extend the maturity of the debt, the Treasury began to regularly issue long bonds every February, May, August, and November. The essential idea behind the contract expiration dates was that dealers would be natural short sellers who would be able to cover their shorts in the futures market by buying cash bonds in the upcoming auction.

The physical delivery standard for Treasury contracts undoubtedly drives the contract pricing. That said, actual deliveries as a percentage of the open interest tend to be quite small, typically under 5%. Most players roll their positions forward as they approach quarterly expiration. Owing in part to the very large open interest in Treasury contracts, the market in forward rolls is very active and becomes especially active as a delivery month approaches.

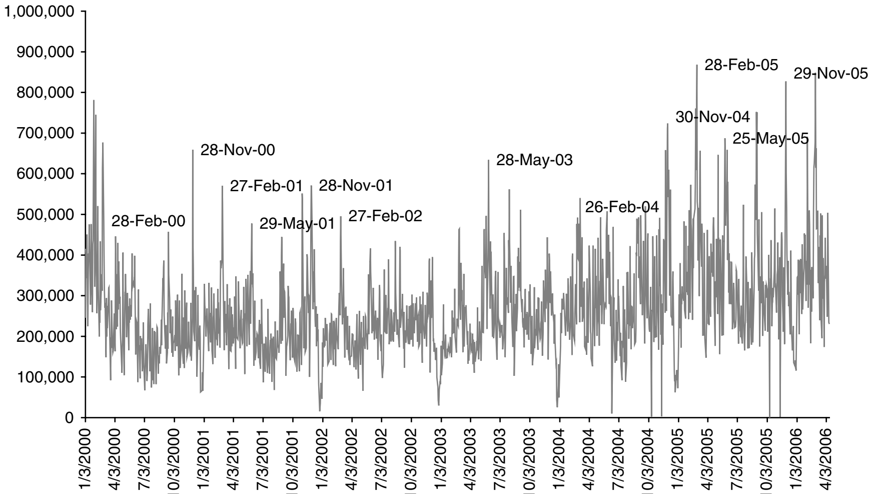
Trading the calendar spreads is a very big business. The best way to see this is to observe seasonal patterns of bond trading volume at the CBOT, depicted in Figure 15.1. As the calendar approaches the spot months of March, June, September, and December, trading volume builds until it reaches a crescendo going into the final days of February, May, August, and November as traders roll their positions forward. To the extent that positions are routinely rolled forward rather than liquidated, pricing the roll is a critical element of hedging and risk management strategy.

Two factors tend to dominate the pricing of forward rolls. The first is expected carry; the second is change in delivery conversion factors. Other factors include potential changes in the cheapest-to-deliver and opportunities to deliver against the spread. Each of these will now be discussed in turn.

First, though, a definition of terms: a *contract roll* refers to buying one contract expiration month and selling another. *Buying the roll* means extending the position forward, for instance, by buying June 2005 bond contracts and selling a like amount of September 2005 bonds. Selling the roll is the opposite. By convention, rolls are quoted with respect to the calendar month that is nearby. If, for instance, June is trading at 100.16 and September at 100.06, the roll would be 10. The market quote for the roll might be 9 bid/10 asked, meaning that buyers of the spread are willing to pay a 9/32nd premium over September to buy June, and sellers are willing to sell at a 10/32nd premium over where they can buy September.

FIGURE 15.1

Chicago Board of Trade Daily Bond Volume (January 2000–April 2006)



Data source: Chicago Board of Trade

The essential difference between two Treasury contracts for the same maturity bucket is the difference in expiration dates. That difference translates into more interest days for the new front-month contract over the expiring one. Longer-dated contracts also have more time for embedded option value to kick in, which may be offset by the value of actually making delivery against the spot month. In addition, delivery factors frequently change with the passage of time, which has to be taken into account. Table 15.5 shows how this works.

On August 17, 2005, Treasury 7.875% bonds due February 15, 2021, were the cheapest-to-deliver into the expiring September 2005 futures

TABLE 15.5

Closing Prices, Yields, and Spreads

Expiry	Futures Price	Settle	Coupon	Maturity	Price	Yield	Factor	Basis
Sept	115 19/32	8/17/2005	7.875%	2/15/2021	137 11/32	4.501%	1.1855	10/32
Dec	115 7/32	8/17/2005	7.875%	2/15/2021	137 11/32	4.501%	1.1838	30/32
Roll	12/32						Differential	20/32

contract. At the close, they were trading at 137 11/32nds. September bond contracts were trading at 115 19/32, making the September basis 10/32nds, which was equal to carry value until the last delivery day, September 30. December bond futures closed at 115 7/32nds, making the December basis 30/32nds, approximately equal to carry until December 31, 2005, the last delivery day for that contract. The 12/32nd roll between September and December bond futures is the spread at which the CTD bond trades at its carry value to each contract. Any other spread between September and December bond futures would create a free arbitrage opportunity.

Consider what would happen at any other price. An arbitrageur, for instance, could take a position of long $100 million Treasury 7.875%'s while hedging half against September and half against December. That would require selling 1.1855 * $50M = 593 September contracts and 1.1838 * $50M = 592 December contracts. Suppose the calendar spread were to widen out to 14/32nds (from 12) by the September contract moving up two ticks to 115 21/32nds. The September basis against the 7.875%'s would then be priced at 8/32nds even though carry is worth 10/32nds. At that point arbitrageurs would simply sell the September–December spread and deliver off their positions in the Treasury 7.875%'s into the September contract, locking in a risk-free 2/32nds profit in the process.

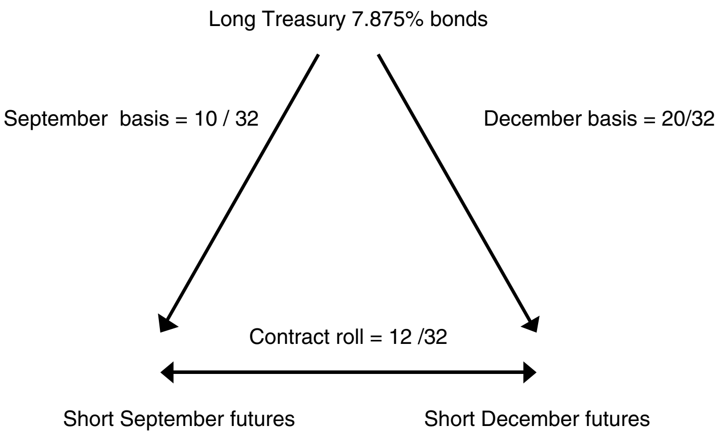
Similarly, if the spread were to narrow to 10/32nds because of a fall in the price of September bonds, arbitrageurs could still lock in guaranteed profits. They would simply buy the spread, lock up financing until December, and deliver off Treasury 7.875%'s in December if they remained the cheapest-to-deliver. The point is that once the two contracts are each trading at their respective cash-and-carry values, the spread between the two contracts is locked in. Any deviation will be ruthlessly exploited by the arbitrage community.

Note in this example that the arbitrage is structured as a triangle: long 7.875%; short a combination of June and December bonds. See Figure 15.2. This position structure allows the spread to be traded back and forth while still retaining option value. Absent this type of structure, trading the spread from the long side in a delivery month is a very different proposition. Owning the spread—long September, short December—would expose the spread holder to delivery options embedded in the spot month.

Another facet of the roll that requires discussion is the apparent disconnect between carry values (20/32) and the roll price (12/32). Carry to September 30 is about 10/32nds while carry to December is about 30/32nds, a difference of about 20/32nds. The forward roll from September to December is only 12/32nds, yet each contract is priced at its cash-and-carry value. What accounts for the difference?

FIGURE 15.2

Triangle Structure of Arbitrage



The reason for the seeming 8/32nd disparity (between 20/32nds of carry and the 12/32nd roll) is largely due to factor shrinkage. The conversion factor for September is 1.1855, but for December it is slightly less, at 1.1838. That factor difference produces a December delivery price that is about 6 and ¼ thirty-seconds lower than the September delivery price. This can be seen by multiplying the price of the September contract by the difference: 115 19/32 * (1.1855 – 1.1838) = 0.1965. Once rounding errors are included, the roll price for the example can be explained in its entirety by factor shrinkage, carry, and transaction costs.

Assuming that Treasury futures contracts retain their notional 6% coupons, cash bonds with coupons above 6% will see their delivery conversion factors shrink as their maturities shrink with the passage of time. Conversely, bonds with coupons below 6% will see their conversion factors expand, because, as discussed earlier, each bond's conversion factor is essentially what the price of the bond would be if it yielded 6%. Premium bonds lose their premium as the maturity shortens; discount bonds rise toward par as they get closer to maturity, all else equal. Bond conversion factors with coupons other than 6% will tend to see them drift, causing delivery prices to change slowly over time even without a change in market levels.

BASIS SPREADS AS IMPLIED RP

Up to this point the examples of basis, calendar, and other spread trades have relied on actual prices so that a P&L could be generated to see how the trades actually would have worked out in dollars and cents. But using dollar prices gives an after-the-fact look. Moreover, using dollar prices for basis spreads may not always be the best way to look at basis spreads. For on the spot comparison purposes most traders use a measure called *implied repo,* or *implied RP.*

Implied RP is defined as the rate that would be needed for breakeven financing of a basis trade. So, for instance, if a bond basis traded at 4/32nds and it would take a repo rate of 5% to earn 4/32nds in net interest until the end of the delivery period, implied RP would be 5%. Many vendors, Bloomberg being the most prominent, calculate and display implied RP rates based on cash/futures quotes. But it is reasonably easy to calculate with Microsoft Excel using the following formula:

$$IRR = \frac{(F * C) + AD - MV}{MV \times t}$$

where:
IRR = implied repo rate
F = futures price
C = conversion factor
AD = accrued interest at delivery, including coupons received and reinvested
MV = current market value of the bonds, meaning the price plus accrued interest
t = time in years from settlement to delivery date

Once the implied RP rates of the various deliverable bonds are known, it is much easier to evaluate them with respect to futures contracts and each other. The bond with the lowest implied RP is generally considered the cheapest-to-deliver. And knowledge of the breakeven financing rate provides an anchor for devising basis and calendar spread strategies that can employ fed funds futures contracts to lock in financing rates.

LOCKING IN FINANCING WITH FED FUNDS FUTURES

Changes in financing rates can have a large impact on basis and calendar spreads. As the spread becomes more positive between bond yields and financing, futures prices tend to be discounted more deeply. The discount also increases between nearby and distant expirations by virtue of the fact that there is more time to accrue net positive carry. The situation is reversed when carry is negative. Distant contracts are priced at a premium to nearby months.

While financing spreads are an important component of bond basis trades, they are not usually the reason that the trades are put on to begin with. Other considerations go into the mix, like speculating on a change in the cheapest-to-deliver, changes in the yield curve, or just plain hedging. Accordingly, some traders may wish to avoid the vagaries of overnight RP rates, opting instead to lock in a financing rate through to the contract expiration date using fixed-rate long-term RPs. But the cost is a sacrifice of flexibility. Locking in the financing rate with term RP requires locking up the collateral, which brings up the problem of RP "specials" mentioned briefly in earlier chapters.

Essentially, there are two types of RP rates quoted in the marketplace. The first is for general collateral, which trades pretty much right on top of the funds rate. The other is the *specials rate,* which is called this because the collateral is in high demand. Holders of a bond in high demand can typically borrow against it at rates that are lower than for general collateral. The greater the demand, the more "special" the bond is and the lower the borrowing rate. In extreme cases, lenders have been known to pay negative interest rates to borrow bonds in unusually high demand.

This aspect of the RP/reverse RP game is not intuitive, so a short recap of the process is in order. When a bondholder repos his position, he borrows money at a specified interest rate using the bonds as collateral for the loan. The other side of the transaction—the money lender—has a "reverse repo" on the books. He is lending cash and taking in the bonds as collateral. When a particular bond is in very high demand, its collateral value is high. Borrowers of bonds in high demand have to accept a low interest rate in order to acquire that bond as loan collateral.

That is why it is problematic for a bondholder to lock up term financing, particularly if the bond becomes special. In order to liquidate a long position, the seller has to make delivery. If the holder/seller has his bonds locked up until a certain date, he has to borrow replacement bonds to that date in order to make delivery against his sale. The difference between the original RP rate and the new RP rate needs to be taken into account when evaluating the transaction.

For instance, suppose that a particular bond is put out on RP for 90 days at 5%. Then suppose that 30 days later the holder wishes to sell the bonds, but they have become special in the RP market, commanding a borrowing rate of 3%. Selling the bonds requires borrowing replacement bonds for 60 days at a rate of 3% to facilitate delivery. The result is a loss of 200 basis points on the loan (5% – 3%) for the 60-day lockup period.

One way to get around this quandary is to hedge term financing rates using fed funds futures. RP rates for general collateral are generally very close to the daily fed funds rate. Since fed funds futures contracts settle for cash at the average effective funds rate for the calendar month, the contract is a reasonably good approximation of financing rates. Accordingly, it is possible to go long cash bonds, finance the position daily in the overnight fed funds market, sell bond futures contracts short to hedge the cash bond position, and sell a strip of fed funds futures short to lock in a hedge against the daily financing rate. The result is a basis trade protected against either an unexpected tightening, or larger than expected tightening, in the Fed's monetary policy. But because it locks in the rate, it comes at the price of not benefiting from a surprise ease in Fed policy. However, if the bond begins to trade as a "special" in the overnight RP market, the hedger will pick up some of the rate spread differential between general and special collateral.

To see the import of this, consider an example that spans three delivery months and includes a switch in the CTD bond. The data for this example include prices for June, September, and December bond futures, yields and implied RPs for three different cash bonds, and yields on a strip of fed funds futures extending from April 2006 through December 2006 inclusive. These data are displayed in Table 15.6.

Note from Table 15.6 that the overnight fed funds rate, tantamount to the RP rate, is 4.75%. Also note that for each of the three delivery months, a different bond is cheapest-to-deliver and that the implied RP (IRR) of each bond exceeds the overnight fed funds rate. At first blush this would appear to suggest the possibility of riskless arbitrage. But on closer inspection it is apparent that the free money is merely a mirage. The fed funds strip has a relentless positive slope, indicating that the market expects the funds rate (and therefore financing) to move higher though December. The implied RPs of the CTD bonds in each of the delivery months is approximately equal to the weighted average expected fed funds rate, based on the strip of fed funds futures. There is no risk-free arbitrage here.

On the other hand, fed funds futures can be used to lock financing in place. Financing overnight after selling a strip of fed funds contracts provides protection against additional hikes in the fed funds rate over and above what is already priced in. But it precludes profits that would come from an unexpected policy ease and leaves open the possibility of RP specials coming into play.

Another factor to consider is that a different cash bond is the cheapest delivery grade for each of the contract expirations. Why would this be the case? The contract specifications allow bonds with a maturity of 15 years or longer to be deliverable. As a result of the passage of time over the different contract expirations, the shorter-dated bonds will drop from eligibility as their maturities become too short. Consequently, the futures market will track other bonds with longer maturities that remain eligible for delivery. By implication, changes in yield spreads among cash bonds can change the pricing of the calendar spreads.

TABLE 15.6

Three-Delivery Month Example Including a Switch in the CTD Bond*

Bond Futures	June 2006	Sept 2006	Dec 2006
Price	107.03	107.00	107.06

Cash Treasuries	Price	Yield	IRR June	IRR Sept	IRR Dec
8.125% Aug 2021	129 23/32	5.27%	5.05%	—	—
8% Nov 2021	128 19/32	5.28%	4.96%	5.03%	—
7.25% Aug 2022	121 9/32	5.29%	3.84%	4.53%	5.10%

Fed Funds Futures Strip

	O/N	April	May	June	July	August	Sept	Oct	Nov	Dec	Avg.
Fed funds	4.75%	4.78%	4.94%	5.02%	5.15%	5.21%	5.24%	5.25%	5.24%	5.22%	5.12%
Weighted avg.		4.78%	4.88%	4.94%	5.00%	5.04%	5.08%	5.10%	5.12%	5.13%	5.01%

*Data as of April 2006

BOND YIELDS AND CALENDAR SPREADS

Our working hypothesis is that calendar spread pricing is driven by marginal carry profits due to extension of the hedge period, adjusting for factor shrinkage or expansion. The pricing calculation implicitly assumes that the bond delivered into one contract expiration is most likely to be redelivered into the next one. For instance, if the Treasury 8% issue maturing in 2021 is delivered into the March contract, the assumption is that the 8%'s will be redelivered into the June contract three months later. And as a matter of fact, that is what tends to happen. It can be confirmed at the Web site of the Commodity Futures Trading Commission (CFTC) (http://www.cftc.gov/). Under the Reports and Publications heading the CFTC lists the full history of all deliveries for all contract months. When the delivery data are observed in sequence, it quickly becomes apparent that the usual cycle is that of delivery and redelivery of the same bonds into the succeeding contract expiration.

While that is the usual course of events, it is by no means the only possible outcome. Consider the example we have been using (from Table 15.6) in which the CTD for the first two contract expirations are almost certainly going to be different by virtue of the fact that two of the bonds are in the process of falling out of delivery eligibility. With the passage of time from June to September, and then from September to December, the maturities of the 8.125% and 8% bonds will have shrunk sufficiently so that they no longer meet the requirement of having at least 15 years left to maturity. Consequently, respective pricing of the June, September, and December contracts is driven by the respective prices of three different cash bonds, each of which is cheapest-to-deliver into one contract but not the others. Implicitly this means that the yield spreads of cash bonds in fact drive the price spreads between June, September, and December bond contracts.

This phenomenon is worth considering for two reasons. First, since the Treasury started selling 30-year noncallable bonds in 1985, there are lots of outstanding older issues that are currently deliverable but will not be in the near future. This is a situation that is liable to recur for years to come. Second, large swings in long-term interest rates have left deliverable coupons outstanding that range from the 4%'s to the 7%'s. Moreover, because of the way rates trended over the years, the highest coupons have the shortest remaining maturities. As a result, potential delivery grade shifts could have a pronounced impact on contract pricing.

To see how yield spreads between cash bonds can affect calendar spreads, we will continue working with the same example as before using the June, September, and December Treasury contracts. This time, however, we make a small adjustment in the cash bond yields. Suppose that instead of yielding about 1 basis point more than the 8.125% issue,

TABLE 15.7

Change in Futures Attributable to Yield Spread Change

	Contract Prices			Calendar Spreads		
	June	Sept	Dec	June-Sept	June-Dec	Sept-Dec
Before	107 3/32	107	107 6/32	3/32	–3/32	–6/32
After 1 BP Δ	107 3/32	107	107 9/32	3/32	–6/32	–9/32
Net Δ	0	0	3/32	0	–3/32	–3/32

BP = basis point

the 7.25% bonds were to trade even yield. To trade at the same 5.28% yield as the 8.125%'s, the price of the 7.25%'s would have to rise from 121 9/32 to 121 12/32. By leaving the December 2006 basis unchanged and consistent with carry, the December futures contract would similarly need to rise by about 3/32nds, from 107 06/32nds to 107 09/32nds.

The result is that both the June–December and September–December calendar spreads have to adjust. The December contract would go up three ticks, but both June and September would remain stationary. The June contract remains tethered to the 8.125% bonds, while the September contract continues to track the 8% bonds, neither of which has budged. So both calendar spreads will have to adjust to a shift in yield spreads among the underlying cash bonds, as shown in Table 15.7.

The lesson here is that the driving force behind the very big business of trading calendar spreads is, without a doubt, priced in the cash Treasury bond and RP markets. During the time that the forward roll markets are busiest, large securities dealers who are active market makers in cash bonds and who participate actively in cash repo markets have a very significant trading advantage. They are constantly in touch with the cash markets that effectively price the forward rolls, and they have access to financing in the cash RP markets. It's a tough combination to beat, and it's usually best for nondealers to avoid the temptation.

TRADING THE TED SPREAD

The TED spread is, without doubt, one of the most widely followed and actively traded intermarket arbitrage spreads on Wall Street. It amounts to trading the yield spread between Treasuries (T) and strips of three-month Eurodollar (ED) deposits, hence the name. Originally the trades were of fairly short duration, pairing up Eurodollars against three- and six-month Treasury bills. By the 1990s traders began stringing together longer strips of Eurodollars to match up against Treasury two-year and five-year notes; some embark on forays out ten years but these are relatively rare.

Eurodollar futures represent the three-month London Interbank Offered Rate (LIBOR) on U.S. dollar deposits outside the United States Contract expiration months follow the familiar March, June, September, December quarterly cycle plus serial expirations which are the next four months not in the quarterly cycle. Expirations extend out 10 years for a total of 40 quarterly cycle contracts. The last trading day is the second London bank business day prior to the third Wednesday of the contract month. Contracts settle for cash to the British Banker's Association (BBA) daily survey of three-month LIBOR.

The BBA survey group consists of 16 banks. At 11:00 a.m. each bank submits its rate estimate for maturities ranging from one day to one year. The four top and bottom rates are eliminated from each of these samples. The BBA survey group sets the fix for each maturity using the arithmetic average of the remaining eight samples. The rates are published daily by Dow Jones, Bloomberg, and Reuters.

For convenience, CME Eurodollars are quoted as an index number rather than a rate, with the implied rate equal to 100 minus the price. For instance a price of 95 would signify a yield of 100 – 95 = 5%. The dollar value of 1 basis point is always $25. The $25 per 1-basis-point valuation stems from the fact that the contract represents a 90-day deposit in a 360-day year, or $1,000,000 * 0.01%/360/90 = $25. By the rules, the front-month contract may be traded in increments as small as ¼ of a basis point, or $6.25 per contract. Trading is in ½-basis-point increments in all the quarterly expirations going out 10 years. A sample representation of daily prices for Eurodollar contracts is displayed in Table 15.8.

It is possible to draw a LIBOR-based yield curve that extends out 10 years by stringing together all the outstanding quarterly Eurodollar contracts in time sequence. But the construction of this curve is different from that of the Treasury curve. The Eurodollar curve represents a series of forward-starting 90-day LIBOR rates; the Treasury curve represents yields available along a continuum of maturities. Differences between Treasury and Eurodollar rates and what they represent are the basis for trading in the TED spread.

THE EURODOLLAR CURVE VERSUS THE BENCHMARK CURVE

The benchmark curve is the one derived from government bond rates. Government bond rates (for paper issued in the national currency) are necessarily the baseline by virtue of the fact that they are free of default risk. The government can simply print up the cash needed to pay off bondholders at par. Cash is legal tender, the government maintains a legal monopoly over its

TABLE 15.8

Sample Representation of Daily Prices for Eurodollar Contracts

Month	Open	High	Low	Last	Settle	Change
May-06	94.81	94.82	94.8025	94.8 1B	94.81	1.0
Jun-06	94.725	94.76	94.715	94.735	94.735	2.0
Jul-06	94.705	94.705	94.705	94.695B	94.7	1.5
Aug-06	—	—	—	—	94.7	3.5
Sep-06	94.66	94.695	94.66	94.685	94.685	3.5
Oct-06	—	—	—	—	94.695	4.0
Dec-06	94.67	94.715	94.66	94.7	94.7	4.5
Mar-07	94.71	94.76	94.695	94.755	94.75	5.0
Jun-07	94.745	94.79	94.74	94.78	94.78	4.5
Sep-07	94.74	94.8	94.74	94.79	94.79	4.0
Dec-07	94.75	94.795	94.75	94.78	94.78	3.5

creation, and lenders are generally required to accept it as good payment. The risk is inflation, not default. All same-currency debt instruments are by definition lesser-quality instruments than government bonds. To attract buyers, they have to pay higher rates for comparable maturities as compensation for greater risk. By implication, spreads between the Eurodollar and Treasury curves are at least partly attributable to these quality differences.

Comparing the Treasury and Eurodollar curves necessitates making some adjustments in order to put them on the same footing. Eurodollar yields are expressed in money market terms. A money market year has 360 days. On the other hand, Treasury interest calculations are based on the actual number of days in a calendar year. For a true comparison, Eurodollar rates need to be converted to bond equivalent yields.

For very short maturities, less than six months, converting money market yields to bond equivalents is a simple matter of multiplying the money market yield by 365/360. For example, a money market yield of 5% would be a bond equivalent yield (BEY) of $365/360 * 5\% = 5.069\%$.

But bond equivalent yields incorporate semiannual compounding; money market yields do not. Bonds pay coupons every six months and yield calculations assume that the coupons are regularly reinvested at the bond's yield to maturity. Eurodollar deposits are add-on instruments that pay interest at maturity. A 270-day deposit therefore does not allow for semiannual compounding; rather compounding takes place with the reinvestment of the principal and interest paid at maturity. For strict comparability, the calculations need to be adjusted. There are several ways to do this; in fact there is an extensive literature devoted to the subject.[1]

THE TERM TED

The quality spread between Eurodollars and Treasuries—the TED spread—can be traded by pairing up Treasury cash flows with a matching sequence of Eurodollar contracts. Alternatively, CBOT Treasury two-year note futures can be traded against CME Eurodollar futures. But before taking that step, the relationship between Eurodollar futures and cash Treasuries has to be sorted out. For example, consider the hypothetical hedge of a cash Treasury note with slightly less than two years to go until maturity. The notes are Treasury 3.625%'s of June 30, 2007, bought for settlement date August 23, 2005.

The first step is to calculate the DV01 of the Treasury notes. The hedge ratio for Treasuries against Eurodollars can then be expressed as the ratio of the respective DV01s. Eurodollar DV01s are always $1,000,000 * 0.01% * 90/360 = $25. In this example, with Treasury 3.625% notes priced at 99.10 to yield 4.00%, and the DV01 is $176.16. The hedge ratio is therefore $176.16 ÷ 25 ≈ 7.05. In other words, it takes 7.05 Eurodollar contracts to hedge a $1 million Treasury 3.625% coupon note maturing June 30, 2007, as of settlement August 23, 2005.

But not just any Eurodollar contract will do. As a starting point the distribution of contracts used for hedging should resemble the distribution of the Treasury's cash flows. That can be done rather easily by spreading out short positions in ED contracts over the life of the long position in Treasuries. The Treasury cash flows are spread out over two years. The first coupon payment occurs on December 31, 2005. After that, coupon payments alternate between June and December until there is a final coupon payment and a return of principal on June 30, 2007.

This is not a static process, at least with respect to risk. The DV01 of the Treasury note will change as its remaining time to maturity shrinks. To the extent that the Treasury DV01 changes while the ED DV01 remains constant, the proper hedge ratio changes as well. However, by spreading out an offsetting ED hedge over the life of the Treasuries, the hedge can be made to self-liquidate fairly closely in line with changes in the DV01 of the Treasury as it rides down the yield curve. Rough estimates of future Treasury DV01s can be made by using implied forward rates. More to the point, the DV01s of the notes will shrink over time as their maturity inexorably drops. With the strip of estimated DV01s in hand, Treasury/ED hedge ratios can be estimated over the life of the notes since the Eurodollar DV01 remains constant at $25. An example is displayed in Table 15.9.

As Table 15.9 makes clear, the number of ED contracts required to hedge the Treasury position is expected to shrink over time. As a result, many traders are inclined to set up the ED hedge so that it is automatically synchronized with the progressive decline in the maturity of the

TABLE 15.9

Example of Treasury/ED Hedge Ratios

Expiration	DV01	Implied Forward Rate	Hedge Ratio
Aug-05	$176	4.00%	7.0
Dec-05	$144	3.94%	5.8
Mar-06	$121	3.90%	4.8
Jun-06	$97	3.87%	3.9
Sep-06	$73	3.85%	2.9
Dec-06	$49	3.81%	2.0
Mar-07	$25	3.53%	1.0

Treasuries on the other side. Normally, the way to do that is to spread out the ED positions over the contract months that coincide with the life of the Treasury, as shown in Table 15.10.

The table displays the way the ED hedge position is spread across the contract months that encompass the life of the Treasury 3.625% note. When the initial Treasury position is taken in August 2005, the Eurodollar hedge ratio is 7:1. Assuming a $10 million par value position, 70 ED contracts are needed to fully hedge the T-notes. To spread out the hedge over the Treasury's term to maturity, short sales (of 10 contracts apiece) are executed in each of the first seven consecutive contract months beginning with the December 2005 expiration. The makes the total hedge $7 \times 10 = 70$ contracts over the Treasury's term to maturity.

TABLE 15.10

ED Hedge Position over a Treasury Lifetime

Eurodollar Contract	Position	Position	Position	Position	Position	Position	Position
Dec-05	10						
Mar-06	10	7					
Jun-06	10	10	8				
Sep-06	10	10	10	9			
Dec-06	10	10	10	10	9		
Mar-07	10	10	10	10	10	10	
Jun-07	10	10	10	10	10	10	10
Position Totals	**70**	**57**	**48**	**39**	**29**	**20**	**10**
Date	Aug-05	Dec-05	Mar-06	Jun-06	Sep-06	Dec-06	Mar-07

Eurodollar Position Self-Liquidates

As the trade rolls on over time, the front-month contracts automatically liquidate when they settle for cash. As a result, very little needs to be done to keep the hedge ratio close to where it's supposed to be. For all intents and purposes, it's almost on autopilot. For instance, at the end of December, the position will automatically adjust from 70 to 60 contracts as the December contracts expire. By that time, the DV01 of the Treasuries will have shrunk sufficiently so that only 57 contracts are needed to hedge the position, so it's simply a matter of covering three contracts for the March 2006 expiry. The procedure repeats itself until finally, the position settles down to a 1:1 ratio in March. By that time the Treasury position is a three-month piece of paper maturing in June 2007, and the last remaining part of the hedge is a three-month Eurodollar contract expiring in June 2007.

The efficacy of the hedge can be tested using scenario analysis. To do this, market yields are subjected to parallel shifts on two different dates. First, we move the markets higher in yield by 20 basis points across the board from the original starting 4% level on the Treasuries. This hypothetical rate shift takes place on the original August 25, 2005, date. A second hypothetical rate shift is examined as well, this one taking place on December 30, 2005. The base rate for this second test is 3.94%, which in August was the implied forward rate for Treasuries.

Note that the weighting scheme has been altered for the second test in that the short position totals 57 contracts, less than the 70 needed in August. By December, 10 contracts would have expired, thus reducing the size of the outstanding short position. In order to keep the proper weights, the March 2006 contract position is adjusted by the purchase of three contracts bringing the March short to seven ED contracts. The other ED contract short positions are left untouched, so the total ED short is 57 contracts.

The acid test of the hedge is how the position fares after a change in the market. Accordingly, we test the impact of an instantaneous 20-basis-point increase in rates across the board. These tests are carried out twice. The first is immediately after acquisition of the position in August of 2005. The second is at the end of December 2005, after the December Eurodollar contract has expired, and along with it, that part of the original hedge. The results are displayed in Table 15.11.

Note the difference in P&L impact with the passage of time. In August an upward 20-basis-point spike in rates induced a principal loss of $35,072 in the Treasury 3.625% notes. By December, four months later, a 20-basis-point upward shift in rates was worth only $28,761, or 18% less. That is because the passage of time reduced the time to maturity and therefore the price volatility of the notes. Similarly, applying an across-the-board increase of 20 basis points to the Eurodollar market produced

TABLE 15.11

Impact of a Sudden 20-Basis-Point Increase in Rates

	Settle	Coupon	Maturity	Price	Yield	P&L
Scenario 1	8/25/2005	3.625%	6/30/2007	99.3346	4.000%	
	8/25/2005	3.625%	6/30/2007	98.9839	4.200%	–$35,072.68
	Settle	**Coupon**	**Maturity**	**Price**	**Yield**	**P&L**
Scenario 2	12/30/2005	3.625%	6/30/2007	99.5446	3.940%	
	12/30/2005	3.625%	6/30/2007	99.2570	4.140%	–$28,761.93

Eurodollars

					Scenario 1	**Scenario 2**
	Contract Expiry	**8/25/05**	**12/30/05**	**Yields**	**August P&L**	**Dec P&L**
	Dec-05	10	–	20	$5,000	
	Mar-06	10	7	20	$5,000	3,500
	Jun-06	10	10	20	$5,000	5,000
P&L	Sep-06	10	10	20	$5,000	5,000
	Dec-06	10	10	20	$5,000	5,000
	Mar-07	10	10	20	$5,000	5,000
	Jun-07	10	10	20	$5,000	5,000
	Totals	70	57		$35,000	$28,500
	Net				–$72.68	–$261.93

an almost identical outcome. In August, the upward rate move generated a $35,000 gain on the short position of 70 Eurodollars spread across seven contract expirations. In December, the 20-basis-point hike was worth $28,500 on 57 contracts spread over six contract expirations. The reduction in the size of the short (and the P&L) corresponds to the reduced duration of the Treasuries.

In each case the loss on the long position in Treasuries was almost exactly offset by a gain on the corresponding short position in Eurodollars. As the Treasuries become less sensitive to changes in rates (due to reduced time to maturity) *by design*, the Eurodollar hedge-adjusted automatically by virtue of the fact that pieces of the hedge rolled off along with the expiring contracts.

Actually it isn't necessary to go into the cash markets to trade the TED spread. The transaction can be replicated by substituting CBOT two-year futures contracts for cash two-year T-notes. In the example used here, for example, Treasury 3.625%'s of June 2007 happened to be the cheapest-to-deliver cash note into the CBOT Treasury two-year futures contract. Figuring out the proper hedge ratio is therefore a matter of estimating the DV01 of the futures contract and then dividing it by the $25 DV01s of CME Eurodollars.

There is an important difference between a TED spread that uses cash securities versus Eurodollar futures and one that pairs up CME Eurodollar futures against CBOT Treasury futures. The cash version of the transaction rolls down the curve and effectively self-liquidates as time moves on. The futures-only version of the trade does not. CBOT two-year futures contracts are designed to replicate the market behavior of cash notes with remaining maturities between 21 and 24 months. A program of rolling CBOT two-year futures positions forward to maintain the position will continually extend the effective duration of Treasury side, requiring adding to the Eurodollar hedge—the exact opposite dynamic of the Treasury cash/Eurodollar futures hedge.

USING TREASURY FUTURES FOR BENCHMARKING

Because Treasury futures contracts track the cash markets so well, they can easily be used to benchmark the performance of traders and portfolio managers. In addition, they can be used as risk benchmarks. It turns out that benchmarking bond market performance can be a surprisingly difficult endeavor. But it can be made easier by using a Treasury futures-based index.

The difficulty in benchmarking bonds stems from the fact that, unlike stocks, the essential risk/reward components of bonds change daily. All else equal, one year from today, 100 shares of IBM will represent the ownership stake they represent today. Not so with bonds. One year from today, a two-year Treasury note will be a one-year Treasury note. The risk/reward characteristics of a one-year note are markedly different from those of a two-year note. And government bonds in particular are priced by the maturities they represent. That presents a problem for bond market indexing.

Bond index providers often try to get around this problem by creating indexes differentiated by maturity buckets. For instance, Lehman Brothers, possibly the biggest name in bond indexing, offers Treasury indexes that track short, intermediate, and long maturities. It also publishes an index that tracks the whole Treasury market, issue by issue. Another, the Lehman Aggregate Index, tracks the taxable U.S. fixed-income market in its entirety. They are available from Lehman on a subscription basis.

Along with Dow Jones, the CBOT publishes a real-time index that in many respects is simpler, more transparent, and easier to use. The Dow

Jones CBOT Treasury Index (DJCBTI) is constructed using the prices of front-month CBOT 5-year notes, 10-year notes, and bond futures contracts. Unlike other bond indexes, the DJCBTI is weighted by the modified durations of its components, so the index is yield curve neutral. It updates every 15 seconds for 22 hours a day and is rebalanced quarterly with a divisor used to maintain returns continuity.[2]

The DJCBTI has several advantages for benchmarking market performance. One is that it is real time, so positions can be instantly updated and measured against it virtually anytime the market is open. Another is that prices are transparent, based as they are on trades in the underlying Treasury contracts. A third is that the simplicity of its design allows the index to be replicated easily. A fourth and more subtle feature of the index is that, unlike other bond indexes, the DJCBTI captures and isolates the beta component of bond market returns.

Recall from previous chapters discussing the CAPM that there are two components to returns: risky and risk free. The risk-free rate can be approximated by the fed funds rate. The risky component is the market rate of return over and above the risk-free rate. That portion of the return (the risky portion) derives from interest, dividends, and price changes. Since Treasury futures contracts replicate the price behavior of the cheapest-to-deliver Treasury less the fed funds rate, it is fair to say that Treasury futures contracts capture the risky component of Treasury returns (keeping in mind that Treasuries are default-free, not risk-free.)

Since the Dow Jones CBOT Treasury Index is an agglomeration of the 5-year, 10-year, and bond contracts, it represents the risky portion of the Treasury market in its entirety—at least from four or five years out. Add fed funds back in, and the DJCBTI becomes a total returns index. That allows the index to be used as a transparent benchmarking device because not only does it give a reliable real-time replication of the larger market, but it also separates out the risky and risk-free portions of Treasury market returns. Moreover, because it is in real time and neutral with respect to the curve, it can be used to measure and manage execution quality as well as risk-adjusted performance.

SUMMARY

Spread trading in debt futures is a big and potentially very lucrative business. But it is important to remember that the most realistic pricing models are the ones in which the cash market is the driver. With that in mind, futures markets can be used to replicate the yield curve, create synthetic bonds, and lock in financing spreads. These types of transactions are possible because futures contracts combined with T-bills are the mirror image of the cash markets.

Moreover, quarterly contract rolls in Treasury futures reflect expectations both for Federal Reserve monetary policy and for future bond deliveries. Just as individual Treasury contracts can be strung together to replicate the Treasury yield curve, they can be used to create a yield-curve-neutral benchmark index. The Dow Jones CBOT Treasury Index, a modified duration-weighted index based on Treasury futures, is an example. Its comparative simplicity and transparency make it an ideal instrument for benchmarking trading and portfolio management skills. Using the DJCBTI as the benchmark, risky and risk-free components of return in the bond market can be isolated, measured, and managed 22 hours a day.

In addition to Treasury futures, Eurodollar futures can be used to hedge cash Treasuries, particularly at the short end of the yield curve. Derivatively, they can be also be used to set up spread trades with Treasury futures, once the implied DV01s are calculated.

This chapter has sketched out a framework for considering how various interest rate futures markets can be used for arbitrage trading both by maturity sectors and quality rankings; how combinations of contracts can be used to lock up financing rates; and how Treasury contracts can be strung together to replicate all or parts of the Treasury curve. These techniques can be extended for risk management and performance benchmarking purposes. There are so many ways to use these very flexible instruments that half the game is to always be on the lookout for new opportunities to use them creatively. The other half is acting when the opportunities show up. Nobody ever got rich by trading on paper.

NOTES

[1] See, for instance, Galen Burghardt, *The Eurodollar Futures and Options Handbook,* McGraw-Hill, 2003, as well as the CME Web site: www.cme.com

[2] More detail, including the precise index formula, can be found at the CBOT Treasury Index Web site: www.cbot.com/treasuryindex

This page intentionally left blank

CHAPTER 16

The Credit Cycle

Everything comes if a man will only wait.

—Benjamin Disraeli

Significant fluctuation in business activity is a fact of life in modern industrial economies. These periods of expansion and contraction, commonly referred to as *business cycles,* were first discovered and analyzed by Arthur Burns and Wesley Mitchell in their 1946 book *Measuring Business Cycles*. In their research Mitchell and Burns observed the marked tendency of many economic indicators to rise and fall together. During expansion periods, output and employment rise, and the rate of unemployment falls. During contractions, the unemployment rate tends to rise while output growth falters. That this is now taken for granted is testament to the power of their work.

Employment and output are among the most important cyclical indicators. But they are not the only ones. A plethora of economic reports can provide telltale signs of the state of the economy. Housing starts, inflation indexes, the index of leading economic indicators, capital spending plans, indexes of business plans published by the Institute for Supply Management, budget and trade deficits (or surpluses), and the University of Michigan Consumer Confidence surveys are among the perennials that are closely followed by many analysts. Plenty of others once thought important have fallen by the wayside. M2, a broad measure of the money supply, comes to mind.

In theory, there is no reason why boom and bust cycles need to occur. Economies should be able to expand with population growth and new technologies. Christina D. Romer, an economist from the University of California at Berkley and an authority on business cycles, notes that the term *business cycle* can be misleading because it seems to connote regularity in the timing of economic expansions and contractions.[1] There is little evidence of this. Regular or not, business cycles occur and they have causes. Among them are disturbances to the economy stemming from policy changes such as large shifts in taxes and spending. The issue is one of timing. There can be long lags between policy changes and the full realization of their effects.

A case in point is monetary policy, often the cause of boom and bust cycles. When the money supply grows too quickly, it can foster the money illusion. In the short run people feel wealthier even though inflation is eroding real incomes. Feeling wealthier, they spend more and give the economy a short-term boost until they realize that their increased wealth was only an illusion, at which point they retrench. On the other hand, if the money supply grows too slowly, real interest rates will rise, stifling investment in new productive capacity, slowing growth potential, and possibly causing a recession.

New Keynesian economists tend to view cycles as stemming from rigidities or kinks in the economy that do not allow prices to perfectly adjust. The inability of prices to fully adjust opens the door to changes in output driven by changes in fiscal and monetary policy. Neoclassical economists citing rational expectations theory tend to dispute that, arguing that the economy is flexible enough to anticipate and adjust to policy changes, obviating their effects.

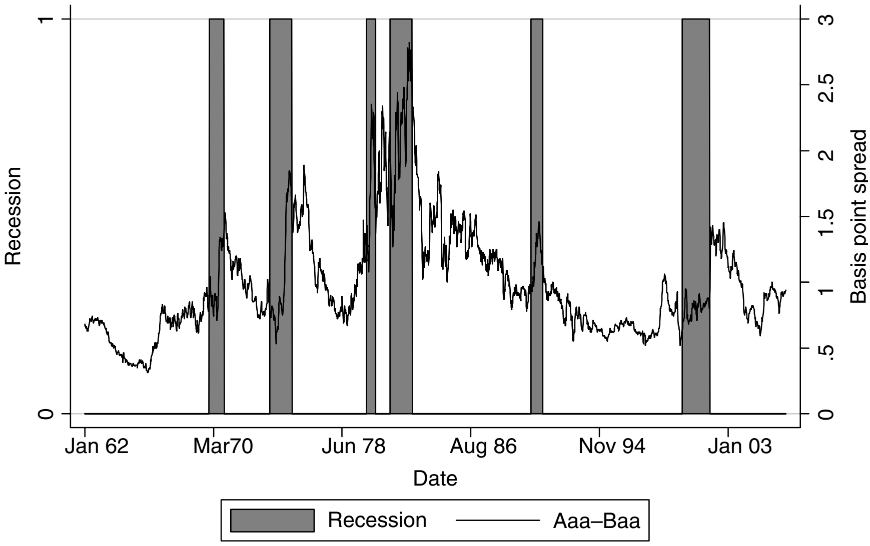
The bulk of the evidence suggests that markets instantaneously and fully adjust to changing circumstances only in the faculty lounge. Everywhere else it takes time. New information comes with a good deal of noise. It has to be filtered and evaluated by businesses, consumers, and financial markets. And the reaction of financial markets feeds back into business and consumer expectations. Nor is there much to be gained by assuming that the market's reaction to events is prescient. The stock market, it is commonly said, discounted eight of the last three recessions.

Uncertainties abound regarding the depth of recessions as they unfold. Similarly, the exuberance of expansions, their timing and likely longevity, are the focus of much debate. Inherent ambiguity in the data makes pinpointing the beginning and end of a recession difficult, even after the fact. So difficult that there is a designated nonpartisan national referee that keeps the score. In the United States, the National Bureau of Economic Research (NBER), an independent research organization, is charged with determining the official dates marking the peaks and troughs of the economy.

Notwithstanding the considerable difficulty of calling turns in the economy, certain patterns emerge when historical data are used to analyze market behavior with respect to the business cycle. During recessions, short-term interest rates tend to fall, the yield curve gets progressively

FIGURE 16.1

Aaa–Baa yield spreads



Data sources: St. Louis Fed and NBER

steeper, and credit spreads widen. During the latter part of an expansion, rates tend to rise, the curve flattens, and credit spreads tighten.

The recurring nature of this market behavior can be observed in Figure 16.1, a graph of the weekly yield spread between seasoned Aaa and Baa corporate bonds from Q1 1962 through Q3 2006. During this time there were six recessions, represented by the shaded areas of the graph. In each of those recessions, the better-quality Aaa-rated bonds outperformed the lesser-quality Baa-rated bonds, evidenced by widening yield spreads. Beyond the simple correlation of spread direction and the business cycle, several other factors need to be taken into consideration.

First, note from the graph the manner in which the yield spreads have historically widened. It is not a gradual, linear process. Yield spreads have tended to widen with a jolt, well into, if not toward the end of the recession. This suggests that the widening of credit spreads tends to be episodic within the larger picture of the business cycle. A credit incident (e.g., a surprise default) occurs, and credit spreads widen generally in sympathy. Second, from 1970 onward, yield spreads displayed a tendency to tighten shortly after coming out of a recession, and for the most part, they continued to narrow on an irregular path for most of the expansion. Third, the times during which spreads are likely to widen (recessions) are the times when the Fed is liable to be easing policy. This may be related to a correlation between periods of financial market stress and the Fed's policy response.

Variation in credit spreads is the market's response to changes in perception of credit quality. As credit quality improves and investors become more confident, spreads tend to tighten. Conversely, when credit quality deteriorates, quality spreads widen as investors demand higher rates on the margin to compensate them for assuming greater default risk. As credit spreads widen and narrow with changing circumstances, quality spread trading opportunities present themselves.

ESTIMATING CREDIT RISK

As discussed in previous chapters, the major ratings agencies assign credit quality ratings to publicly traded debt issues. In addition there are sophisticated credit models available that provide statistical scores of creditworthiness, some of which are made available by the ratings agencies themselves. For instance, Standard & Poor's has developed CreditModel, which produces a quantitatively derived estimate of a Standard & Poor's credit rating.[2]

S&P's CreditModel generates credit scores within industry groupings by applying statistical analyses of relationships between financial data and within-sector S&P credit ratings. But model scores are not actual ratings. A rating from S&P represents the agency's opinion of the creditworthiness of the rated entity. The rating process normally includes analysis of qualitative variables such as management prowess, projected capital spending, and market share. These variables are not included directly in the strictly quantitative inputs for deriving CreditModel scores. They do, however, feed back into the model indirectly; the models are designed to emulate S&P ratings. In fact, CreditModel scores are very highly correlated with official S&P ratings.[3]

In addition to S&P's CreditModel there are many other models of credit risk in the market place. JP Morgan's CreditMetrics model is based on a probability analysis of a credit risk moving from one credit state to another within a defined time horizon, for instance a downgrade from AAA to AA within one year.[4] This methodology allows for the creation of a transition matrix, as in Table 16.1, and the application of value-at-risk (VaR) modeling.

The transition matrix displays probability estimates for credit upgrades, downgrades, and no change. Those probabilities can be combined with a credit curve, which illustrates differences in yields to maturity for bonds based on their credit quality. Unanticipated changes in credit quality would presumably result in higher or lower yields, the dollar magnitude of which can be calculated based on the yields in the matrix.

TABLE 16.1

Hypothetical Credit Transition Matrix (Probability Estimate of Credit Rating in One Year in Percent)

Credit Rating	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	90	9	1	1	0	0	0	0
AA	7.5	85	8	5	0	0	0	0
A	2.5	4.5	80	10	4	0	0	1
BBB	0	0.50	8	75	5	1	1	2
BB	0	0	2	7	70	8	5	5
B	0	0	0	2	20	60	15	8
CCC	0	0	0	0	0	30	50	20
Credit Curve	**5%**	**5.25%**	**5.5%**	**6%**	**6.75%**	**8%**	**20%**	

As a practical matter, markets often anticipate ratings changes. To some degree that is because the major market-making firms have their own (very sophisticated) in-house credit modeling capabilities. Another is that there is a trade-off between ratings accuracy and stability. Ratings, by design, take a long view. Their utility does not hinge on small day-to-day changes in circumstances.[5] But sometimes credit disasters hit like a bolt out of the blue. Orange County California was rated investment grade only days before it defaulted. And the corporate scandals involving Enron, WorldCom, and Global Crossing indicate that the markets (and many analysts) were caught off base. All of which suggests that rigorous research may be able to ferret out issues (or sectors) that are over- or undervalued due to the market's faulty perception of credit quality. In such cases, transition matrices of the sort displayed in Table 16.1 may be valuable tools for analyzing the potential price impact of credit quality changes.

One way to use transition matrices is for credit quality scenario analyses. Transition matrices and credit scoring can be used to examine quality distances between individual bond issues in the same industry, between individual bonds across industries, or between portfolios of bonds representing different sectors of the economy. The way that the transition matrix is set up illustrates that credit spread trading is based on more than probability-of-default (PD) and recovery estimates in the event of default. It includes trading off incremental changes in quality without necessarily contemplating default. For instance the matrix can be used as a tool to assess whether the risk of a downgrade from AA to A is worth the yield spread differential between the two bonds.

A number of caveats need to be mentioned with respect to trading credit spreads and the use of transition matrices. First, credit modeling is an incredibly complex business, and there are lots of wrinkles that have to be considered. The probability estimates contained in transition matrices are only as good as the data and the modeling techniques used. Second, good price data on corporate bonds are notoriously hard to obtain, so estimating bond price changes resulting from changes in credit quality is inherently dicey. Moreover credit modeling techniques use quite sophisticated mathematics, and it is important to understand how the models work instead of relying blindly on them. JP Morgan explicitly makes this point in its reference paper on CreditMetrics when it says, "We remind our readers that no amount of sophisticated analytics will replace experience and professional judgment in managing risks."[6]

A third factor to consider is that statistical models are based on averages (and sometimes averages of averages) that may not translate easily to individual bonds. It is worth remembering that AAA-rated bonds can only stay the same or be downgraded; their ratings can't improve. Furthermore, a rating of AAA or AA may turn out to be cause for hollow laughter if having good quality credit sets the stage for an LBO that results in a soaring stock price at the expense of a large credit downgrade and crashing bond price. The classic of the genre is the 1988 RJR-Nabisco LBO, memorialized by Bryan Burrough and John Heylar in *Barbarians at the Gate*.[7] RJR-Nabisco's AA/Aa rating plunged to Ba/BB virtually overnight in the wake of the LBO bidding war, crushing the bondholders. In order to avoid this predicament, some AAA and AA bondholders also maintain positions in the bond-issuing company's common stock.

A MACROFRAMEWORK FOR TRADING CREDIT SPREADS

The discussion of credit spreads thus far has implicitly focused on first isolating the credit portion of spreads by matching up maturities of different quality instruments before trading the quality spread. There is, however, another piece that can be layered on top of matching maturity credit spread differentials. The historical evidence strongly suggests that both credit spreads and the yield curve are correlated with the business cycle. Weaker *credits* perform relatively better in expansions; shorter *maturities* perform better during recessions. This hints at a strategy that simultaneously incorporates both yield curve and credit spread considerations.

The basic idea is illustrated Table 16.2, which lays out a strategy framework along the two dimensions: credit quality and bond duration. Depending on the approach and weighting scheme, layering the yield curve over credit spreads can make a bond portfolio (or arbitrage position) behave in a procyclical, contracyclical, or neutral fashion.

TABLE 16.2

Business Cycle Quality/Duration Matrix

Preferences	Recession	Expansion
Quality	High	Lesser
Duration	Short	Longer

The credit quality/yield curve strategy matrix can be viewed in either of two ways. The first is profit maximization; the second is risk minimization. An optimal strategy would be one that, consistent with the CAPM, finds the best trade-off between curve and credit risk. As a practical matter that may be beyond current data and modeling capabilities, at least at the level of precision that would be needed to make the effort worthwhile.

From a profit maximization standpoint, the idea is to marry up strategies whose outcomes have been reliably correlated in past business cycles. From a hedging or risk management standpoint, the strategy is to diversify by offsetting credit spread risk with an implicit yield curve hedge. The profit-maximizing position can be considered procyclical, whereas the hedge position can be thought of as contracyclical or neutral, depending on the weighting used. The approach chosen depends on the immediate goals, market outlook, and risk propensity.

PROCYCLICAL PROFIT-MAXIMIZING STRATEGY

For example, consider the following scenario. The economy is (or appears to be) in recession with the Federal Reserve beginning to cut interest rates. If history is any guide, two things are likely to happen. First, the yield curve will become more positively sloped. Second, credit quality, will begin to deteriorate along with business conditions. As credit quality deteriorates, yield spreads between higher- and lower- quality bonds will widen.

Under these conditions the profit-maximizing strategy would be to go long high-quality short-dated bonds, preferably Treasuries, against establishing a short position in low-quality, longer-dated corporate bonds. The positions would be DV01 weighted, using the methodology discussed in previous chapters. The result is a box trade that incorporates both the curve and credit quality. The expectation is that profits will be made two ways. First, a steepening of the curve would be expected to bring in profits. Second, even absent a favorable curve shift, the spread between low-quality corporate bonds and Treasuries of the same maturity would normally be expected to widen, which would result in a profitable trade.

In an expansion the opposite is expected to happen, so the preferred strategy would be to go long lower-quality corporate bonds against a yield-weighted short position in shorter-dated Treasuries. This anticipates profits from a curve flattening, combined with tightening quality spreads as investor confidence builds in the ability of lower-quality credits to pay off.

CONTRACYCLICAL RISK MINIMIZATION

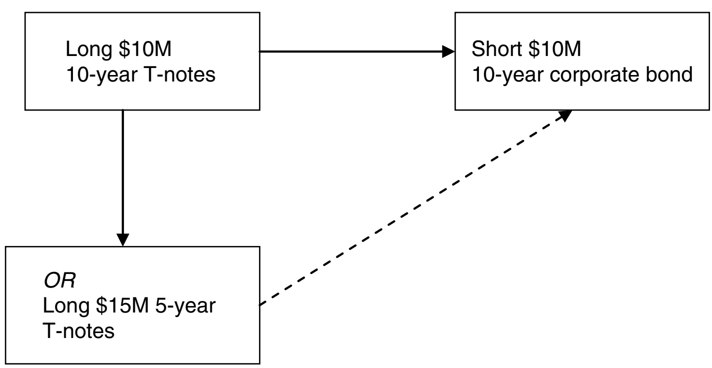
The contracyclical or risk management approach would seek to offset a curve trade with a credit quality trade. In this scenario, a long position in a lower-quality corporate bond would be married to a yield-weighted short position in a shorter-duration Treasury. Conversely a short position in a lower-quality corporate bond would require a yield-weighted long position in a longer-duration Treasury.

Layering a yield curve trade over a credit spread can provide both additional flexibility and implicit liquidity by expanding the array of hedging instruments available. For example, hedging a short position in a 10-year maturity corporate bond with an offsetting long position in 5-year T-notes (on a yield-weighted basis) provides the added flexibility of trading the 5-year T-notes against 10-year T-notes while keeping the corporate short in place. This creates a triangle structure as shown in Figure 16.2.

In principle, combining yield curve trades with quality spreads can be executed up and down the curve with a wide variety of credit instruments at various quality gradations. In general, the wider the space between the chosen credit instruments, the more volatile the spread is liable to be, just as volatility widens along the curve as the distance between maturities grows. A wide variety of credit instruments trade in the public markets, including bank loans, securitized credit card liabilities, OTC interest rate

FIGURE 16.2

Credit/Yield Curve Triangle



swaps, and more conventional corporate and federal agency paper. Which instrument to use when depends on market view. Note however that some instruments, in particular OTC swaps, add an additional layer of counterparty risk to the transaction. One way to avoid this is to use exchange-traded swaps. They are reasonable proxies for corporate credit risk and can be used to layer credit spreads over yield curve trades.

MUNICIPAL NOTES AND BONDS

Municipal notes and bonds are credit instruments that, for the most part, are exempt from federal tax as well as taxes levied by the issuing jurisdictions. Some are exempt from the alternative minimum tax; some are not. In order to make a fair comparison between tax-exempt and taxable securities, by convention the tax-exempts are grossed up to their taxable equivalent yields (*TEqY*). That is done by dividing the yield to maturity by 1 minus the marginal tax rate as shown in the following equation.

$$TEqY = \frac{YTM}{1-r}$$

As a general rule of thumb, municipal bonds do not lend themselves to the same types of arbitrage/spread trading that corporate and government bonds do. There are several reasons for this. The first is that it is usually impermissible to sell tax-exempt bonds short, although the ban seems to be honored in the breech. The reason for the prohibition is that the short seller creates de facto tax-exempt income, which he is not authorized to do. The second reason is that individual municipal bond issues are small as is the volume of trade in the tax-exempt secondary market. Average daily volume is only about 3% of governments. The average par value of a municipal bond is in the neighborhood of $10 million. Selling short (and trying to borrow) an issue that small is suicidal. After they are placed by the underwriters, the vast majority of tax-exempts disappear until maturity when they show up for redemption. There is not a lot of point in secondary market trading. People buy tax-exempts to avoid taxes, not to rack up taxable transactions.

A final reason to avoid using municipal bonds as a trading vehicle is the paucity of information available about the finances of the issuer. By and large, neither states nor municipalities are required to meet anything close to the accounting and disclosure standards of *pre* Sarbanes-Oxley corporate America. A close look at the books of most big cities in the United States (if you could get them) would probably make Enron look pristine by comparison. Weak credit quality and accounting opacity are among reasons why most tax-exempt issuers have to get their bonds insured before they can be brought to market.

SUMMARY

A bond's credit spread is the yield premium over Treasuries needed to attract investors to assume default risk. Treasuries serve as the benchmark because they are the default-free rate for a given maturity. Other bond yields for different quality credits, controlling for maturity, can be thought of as representing a credit curve. But the credit curve is not static; it is integrally linked to the business cycle and therefore indirectly to fluctuations in the yield curve.

The major rating agencies (Standard & Poor's, Moody's, and Fitch) publish opinions of bond credit quality in the form of ratings. Beyond official ratings, credit risk, which includes probability of defaults, downgrades, upgrades, and recovery potential, can be estimated using sophisticated mathematical scoring models. Combining probabilistic models of credit scores, transition matrices, and the credit curve provides a benchmark for conducting value-at-risk and trade scenario analysis. Yield curve analysis can be layered over credit curve analysis.

Microlevel credit-trading strategies tend to be either issue or industry specific. Strategy is focused on exploiting pricing anomalies in which it is believed that credit spreads do not accurately reflect credit fundamentals. Yields may be too high (low) relative to the chances of an upgrade (downgrade) or default. On the other hand, macrolevel strategies place more emphasis on the business cycle and the yield curve. Arbitrage positions may have yield curve components embedded in them. Generic instruments like exchange-traded swaps may be used in addition to, or as a substitute for, individual corporate bonds to represent the credit side of a transaction.

NOTES

[1] See "Business Cycles" by Christina D. Romer in *The Concise Encyclopedia of Economics* online at: http://www.econlib.org/library/Enc/BusinessCycles.html

[2] For a white paper explaining the details of the model see "CreditModel Technical White Paper" by Craig Friedman at: https://www.creditmodel.com/

[3] See, for instance, CreditModel performance statistics available at www.creditmodel.com

[4] JP Morgan has published an extensive technical document discussing methodology, "CreditMetrics," April 1997, which can be found at: www.creditriskresource.com/papers/paper_125.pdf

[5] See for instance, "Analyzing the Tradeoff between Ratings Accuracy and Stability," Richard Cantor and Chris Mann, Moody's Investor Services, Global Credit Research Special Comment, September 2006.

[6] JP Morgan, "CreditMetrics," April 1997, p. 1.

[7] Bryan Burrough and John Heylar, *Barbarians at the Gate,* McGraw-Hill, New York, 1988.

CHAPTER 17

Trading Equities: An Overview

> October: This is one of the peculiarly dangerous months to speculate in stocks in. The others are July, January, September, April, November, May, March, June, December, August, and February.
>
> —*Mark Twain,* Pudd'nhead Wilson's Calendar

It is hard to exaggerate the differences in the way stocks and bonds are traded. The instruments are constructed differently, they represent different tiers of a firm's capital structure, and different legal rights are attached to them. Not only are the underlying instruments different, but the respective regulatory regimes and agencies responsible for rule making and enforcement are different as well. In addition, market structures, cultures, and institutions vary significantly. For instance, the overwhelming majority of bond trading takes place over the counter with the major investment banking firms acting as principals and market makers. On the other hand, stock trading is dominated by exchanges, in which major firms are more likely to act as brokers than as dealers. Even an OTC stock market like Nasdaq has the essential characteristics of an exchange, minus the trading floor.

To lay a foundation for considering stock-trading strategies, this chapter first examines some key characteristics of equity securities. Second, it reviews pricing theories and equity market valuation models. Third, it considers market institutions, including differences between OTC and exchange markets. It also summarizes exchange listing requirements and corporate governance standards.

WHAT DOES A SHARE OF STOCK REPRESENT?

A share of stock represents partial ownership of a company, hence the word *shares*. The key difference between shares and bonds is that stockholders own the company; bondholders merely lend the issuing entity money. Shareholders have an equity interest in the company, but with liability limited to the original investment. They own the firm's assets, and the management of the firm is theoretically answerable to the stockholders. In general, the board of directors is supposed to oversee the firm's management and represent the interests of the stockholders.

Corporate governance is a very complicated issue and one that can have considerable impact on stock market returns. One of the more vexing problems is finding ways to align the interests of management with the company's shareholders. Compensation plans aimed at aligning interests are difficult to design and may have unintended consequences. Recent scandals involving stock options are a case in point. Ostensibly designed to reward the common interests of managers and shareholders, stock option awards induced some managements to implement tactics designed to hype short-term stock market performance at the expense of long-term profitability. They have also led to backdating scandals in which options exercise prices were picked at low prices after the fact thereby eliminating the performance incentive that the options were supposed to create. These types of problems can arise fairly easily, particularly when the board is less than vigilant about looking out for the firm's shareholders.

Knotty problems designing governance structures notwithstanding, the ultimate weapon shareholders have to protect their interests is the public marketplace. If a firm's management is sufficiently inept, it is likely that the company's stock will trade at a discount. A deep discount from true value can put the firm in play as a tempting takeover target. That is why the stock market is often referred to as the market for corporate control. By implication, true value, management quality, and corporate governance are critical variables to consider in formulating stock selection strategies.

The true value of a stock is widely thought to be the present value of its future cash flows. But no one knows what future cash flows will be; nor does anyone know with certainty what the appropriate discount rate ought to be. Therefore analysis of stock values deals in terms of probabilities and expectations. The price of a stock can be said to represent the market's estimate of the present value of expected future cash flows. Cash flows basically come from dividend payments, although they can come from future sales of stock as well. Dividend payments come from earnings; the percentage of earnings returned to stockholders by way of dividends is the *payout ratio*. Because earnings are so important, an army of Wall Street analysts follows the major companies (and lots of smaller ones) to gauge their current earnings prospects, as well as those far into the future.

DO STOCK ANALYSTS MATTER?

Analysts publish their research, generally using ratings systems that benchmark the relative attractiveness of a particular company compared to its industry or the market as a whole. Typically the analysts either recommend an action (buy, sell, hold) or assign a rating (market neutral, overweight, underweight) that implies taking action. More often than not, analysts pick a target price for the stock to reach over some time interval, usually a year.

Whether analysts' opinions count for much, if anything, is a matter of debate. An intriguing paper by Hans G. Heidle of Notre Dame and Xi Li of the University of Miami has presented evidence that market makers' quotes in Nasdaq stocks anticipate the recommendations of in-house analysts, leading them to conclude that nonpublic information is being put into these stocks by opportunistic dealers and customers who trade through them.[1] Chan, Chang, and Wang found confirming evidence that financial firms tend to trade in the same direction of their analysts recommendations, including in the quarter after the recommendations were issued. This would seem to indicate that the firms believe their analysts' recommendations are worthwhile and that it takes time for the information to be fully priced into the market.[2]

On the other hand, a spectacular series of miscalls by technology analysts in the vanguard of the Internet stock boom cast considerable doubt on the value of much of what goes by the name of fundamental analysis. *BusinessWeek* noted that one analyst dubbed "Queen of the Net" was an early proponent of the new economy theory (this time it's different) and an early bull on Yahoo! She liked the stock at $10, then $25, then $100, and all the way up to $237. And she liked it all the way back down as it plunged by 94% to $15.[3] She was not alone in letting her enthusiasm get the best of her; lots of analysts did. Others cynically manipulated the research process to serve their own ends.

Geniuses, it is said, are born in rising markets. In their respective sectors of analytical "expertise" (telecoms and Internet services), Jack Grubman and Henry Blodget were rated number 1 by The Street.com in June 2000.[4] Blodget was particularly enthusiastic about the Time-Warner AOL merger, one of the most spectacular deal failures in history. He was exceptionally bullish on prospects for the new combined entity, calling for the stock to rise smartly. Most would have found it difficult to value the new firm since the merger had only recently been announced and the newly combined company had not disclosed its business plans. That did not appear to bother Blodget one iota. "They are starting with a blank sheet, you have to take it on faith it's a positive," he reportedly said.[5] His analyst-as-rock-star status was confirmed when *Time* magazine reported that a Blodget sighting prompted the announcement that "Elvis has entered the building."[6]

During this era Jack Grubman was busily advising telecom companies like Worldcom and touting the stock at the same time. Selling the benefits of these stratospherically priced stocks had the dual advantage of bringing in investment banking business and getting his daughter a sufficient edge to get admitted into the notoriously competitive 92nd Street Y kindergarten class. It's nice to see fatherhood taken seriously. But the cheerleading of Grubman and Blodget pales in significance compared to Enron, which became one of the largest and most spectacular bankruptcies in history. Here is a company riddled with fraud whose stock soared with each new press release. No major investment bank or rating agency (publicly) displayed any doubt about the enterprise. And then in the space of a few short months it all crashed back to earth, the stock worthless, the gullible relieved of their cash.

In the end, the authorities took a very dim view of all this. The SEC, the New York attorney general's office, NASD, and the New York Stock Exchange permanently barred Grubman from the business, fining him $15 million in the process.[7] Grubman buddy Bernie Ebbers, CEO of Worldcom, is now in prison. Blodget escaped with a relatively light $4 million fine, although he too is permanently barred from the securities business.[8] Merrill Lynch was fined $100 million, and it agreed to separate research from investment banking, an agreement that became the model for the industry.

The Internet bubble is a cautionary tale that illustrates the importance of looking with a jaundiced eye when the next big story comes around. And it will. This type of thing happens periodically, typically toward the end of a bull market when there is a frenzy to get onboard. In the early 1970s the rage was about the "nifty 50 one-decision stocks," so called because they needed only to be bought, never sold. They would go up forever. That too ended when the bear's paw slapped the market hard. Some of the glamour stocks lost 90% of their value. Others like Polaroid, or just "Roid" as the traders use to call it, eventually went out of business.

Another thing to note is that stocks are not bought—they are sold. People talk about "the market" all the time, but what really gets them revved up is how individual stocks fare. For bragging rights there is nothing quite like picking a big winner. But most of the research strongly suggests that the search for a big winner is hopelessly quixotic. Over time, few people have shown themselves to be able to match, much less beat, the returns generated by the popular indexes, after adjustments are made for risk. The paradox may be that it is difficult, although not impossible, to beat the market precisely because it tends to be priced reasonably well—because of the work of good analysts.

HOW ARE STOCKS PRICED?

Investors buy stocks to earn returns, which they can receive in two ways. The first is by price appreciation; the second is through dividends. What matters more than any one component of returns is the total return, which is the combination of price change and dividend payments, as shown below. Stock prices can therefore be said to reflect the total returns investors expect to receive over time.

$$Total\ Return = \frac{Dividends + \Delta Price}{Initial\ Price}$$

VALUATION MODELS

There have been many models developed to assess stock market pricing. A concise summary of standard valuation models can be found in *Finance* by Groppelli and Nickbakht.[9] Perhaps the most famous is the dividend discount model (DDM), originally developed by John Burr. Published by Harvard in 1937, the DDM was an outgrowth of Burr's "Theory of Investment Value" dissertation. A few years later it was elaborated more fully by Graham and Dodd in their classic work *Security Analysis*.[10] The underlying idea is that the intrinsic value of a share of stock is equal to the present value of its expected future dividend payments.

One of the factors that attracted Burr to using dividend payments to value a company was transparency in corporate accounting, although that's not how they phrased it in those days. It was common then to refer to stocks as "watered" when their assets were inflated on the books for the purpose of hyping the stock, particularly during the roaring twenties. Which, when you think about it, it doesn't sound all that different from the behavior of Enron, Worldcom, Tyco, Global Crossing, and a host of others.

Dividends are tangible (but not foolproof) evidence that the company actually is making at least enough money to pay out some of the profits directly to shareholders. Since dividends are presumably paid out of earnings, estimating the value of a stock with the DDM implies a forecast of future earnings and the dividend payout ratio, which is the portion of earnings that the company returns directly to the shareholders in the form of dividends. The balance remains in the company as retained earnings. Determining the value of expected future dividends requires choosing a discount rate, also referred to as the *expected rate of return* or the *required rate of return*. The inverse of the required rate of return is the *multiple,* also known as the *price/earnings ratio* (P/E) or the *capitalization rate*. The P/E ratio is one of the more popular measures for valuing stocks.

In addition to the P/E ratio, other key financial measures are used to value stocks; some of the more important ones that we will employ in this chapter are defined by the formulas that follow.

$$\text{Earnings per share: } EPS = \frac{Earnings}{No. of\ Shares}$$

$$Dividend\ Payout = \frac{Dividends\ Declared}{Earnings}$$

$$Current\ Yield = \frac{Dividend}{Stock\ Price}$$

$$P/E = \frac{Stock\ Price}{EPS}$$

$$Earnings\ Yield = \frac{1}{P/E}$$

$$\text{Return on Equity: } ROE = \frac{Net\ Income}{Equity}$$

For example, assume a company called Ron's Flawless Jeans (ticker RFJ) earns net income of $200 million, the company pays 45% of its earnings as dividends, there are 100 million shares outstanding, and the stock trades at $21 a share. Further assume equity capital of $2 billion and that earnings grow at 7% a year. Earnings per share (EPS) would be $200 million ÷ $100 million = $2. The dividend payout would be 45% of $200 million = $90 million or $0.90 per share. The current yield of the stock would be $0.90 ÷ $21 = 4.29%. The stock would be trading at 10.5 times earnings, another way of saying the P/E ratio is $21 ÷ $2 = 10.5. The earnings yield would be the inverse, or 1 ÷ 10.5 = 9.52%. Return on equity would be $200 million ÷ $2 billion = 10%.

The value of a share of stock in Ron's Flawless Jeans can be estimated with a model developed by Myron J. Gordon, emeritus finance professor at the University of Toronto's Rotman School of Management. The model values a share of stock as a function of dividends, the dividend growth rate, and the required rate of return. For stocks with a constant growth rate the formula that follows can be applied:

$$P_0 = \frac{D_0(1+g)}{K-g}$$

where:
P_0 = the current stock valuation
D_0 = the current annual dividend

K = the required rate of return
g = the dividend growth rate

In the case of RFJ, the calculation would be:

$$P_0 = \frac{0.90(1+0.07)}{0.116-0.07} = 20.93$$

Note that the growth rate is the key driver of the model. The faster dividends are expected to grow, the more the stock is worth. But with this model, the expected growth in dividends reaches its limit at the required rate. If the expected growth in dividends were to exceed the required rate, the denominator (which represents the risk premium) would be negative, in turn producing a negative stock valuation. It would imply that someone would be willing to pay you to take stock off their hands, clearly a nonsense solution, or at least one that is unlikely to take place anytime soon.

Calculating the likely growth rate of a company is a daunting task. There are lots of variables that can affect it, including position in the business cycle. One relatively simple way to get a handle on the dividend growth rate is to treat it as dependent on the firm's return on equity. This method implicitly incorporates the firm's investment decisions into the stock price. To the extent that a firm's capital investment decisions affect earnings growth, they affect dividend payouts as well as expected future payouts and therefore the stock price.

$$G = ROE \times (1 - Payout\ Ratio)$$

For Ron's Flawless Jeans, the estimated growth rate is easily calculated by substituting into the formula:

$$5.5\% = 10\% \times (1 - 0.45)$$

Valuing stocks on the basis of dividends is fine, but some fast-growing companies do not pay dividends. These companies argue that the rate of return they can earn by reinvesting in the business likely exceeds the rate of return a stockholder could earn in the public markets. The sensible thing for such a company is to minimize the dividend payout and plow earnings back into the business. Stockholders will benefit from increased earnings resulting from investment of internally generated funds.

This approach to valuation, championed by Nobel Prize winners Franco Modigliani (1985) and Merton Miller (1990), asserts that the market value of a firm is unrelated to its dividend payments. The value of the firm is the present value of its earnings before interest and taxes (EBIT). Whether the firm pays out the earnings in the form of dividends is irrelevant as far as the firm's actual economic value is concerned.

A variant of this theory includes the impact of taxes and depreciation. Firms pay taxes on their earnings, but corporate taxes (and tax rates) are not the same as individual taxes (and tax rates). To the extent that corporate tax rates are lower than individual tax rates, individuals are arguably better off having corporations reinvest their earnings back into the business because the earnings can compound at a lower after-tax rate. To boot, dividend receipts are a taxable event, so individual stockholders are taxed twice upon receipt of dividends.

That leaves the question: How are stockholders supposed to profit from earnings if they don't receive dividends? The answer is capital gains. Stockholders can choose to realize their earnings by selling all, or some, of their stock. In theory, the stock of a company that minimized dividend payouts in favor of reinvesting in the business ought to have a higher multiple and stock price. The marginal (positive) difference in the stock price works to the benefit of the stockholder. Moreover, when long-term capital gains are taxed at lower rates than dividends, stockholders are better off receiving their income in the form of capital gains.

The basic model for this value of the firm (V_f) is the present value of its earnings before interest and taxes at time $_t$, using discount rate K_0.

$$V_f = \frac{EBIT_t}{K_0}$$

This conception of the firm has been criticized by Warren Buffett, among others, for failing to recognize consumption of the firm's capital. It also fails to recognize the importance of tax, hardly a matter of insignificance evidenced by the enormous amount of time and money spent on tax matters by corporate America. A slightly more complicated, but much more realistic way of modeling the firm is to use free cash flows, which takes these factors into consideration. *Free cash flows* can be thought of as a combination of after-tax EBIT, depreciation (a measure of capital consumption), taxes, and new investment. It is represented by the formula that follows:

$$FCF = EBIT\,(1 - T) + Dep(T) + Ts - I$$

where:

FCF = free cash flow
T = tax rate
$EBIT$ = earnings before interest and taxes
Dep = depreciation
Ts = tax subsidy
I = (new) investment

Note the importance of the tax regime in the free cash flow model. EBIT multiplied by the effective "take home rate" $(1 - T)$, represents the cash portion of operating earnings; depreciation adds to cash by reducing reported profits; the tax subsidy (a plus) comes about from borrowing because corporate interest payments are tax deductible, and new investments (I) are a subtraction from cash (before tax).

The advantage of valuing a firm in terms of its free cash flow is that it gives a reasonably good view of the firm's operations. And in the end, dividends are paid with cash, which has to be earned. To illustrate how free cash flow can be used to value the firm, we will go back to Ron's Flawless Jeans and assume a tax rate of 35%, depreciation of $10 million, and new investments in plant and equipment of $35 million. In addition, RFJ has issued $125 million par value bonds due in 10 years with an 8% coupon (the market rate of interest) payable semiannually.

Free cash flow is therefore equal to $\$200\text{M} \times (1 - 0.35) + \$10\text{M} \times 0.35 + \$10\text{M} - \$35\text{ M}$, or $2,372,608,696. From a firm valuation standpoint that is not the end of it. RFJ still has outstanding debt obligations maturing in 10 years in the par amount of $125 million. The present value of the outstanding debt discounted at 8% is equal to $84,445,521. When the present value of the debt is subtracted from free cash flow, the remainder is $2,288,163,175, which represents the value of the firm. Since there are 100 million shares outstanding, the value of each share is $22.88.

Clearly, expected earnings growth is a critical driver of stock prices. The more quickly earnings rise, the quicker dividends rise, assuming the same payout ratio. All else equal, it is therefore preferable to own stock in a company whose earnings are growing quickly. But all things are not equal. People are willing to pay up for earnings growth, which implies that companies with faster earnings growth will tend to have their earnings valued more highly than companies with slower earnings growth. Similarly, markets abhor uncertainty, so companies whose earnings grow steadily will tend to have their earnings valued more highly in the marketplace than those whose earnings fluctuate with the business cycle. The heart of the question is: What is the right price to pay for expected earnings growth?

A metric often used as a common denominator for stock valuations is the price/earnings ratio. The multiple, or P/E, is simply the stock price divided by earnings per share. Companies with faster, steadier growth are assumed to be less risky and worth more. As a result their earnings are valued more highly as evidenced by higher price/earnings ratios. Since stocks are valued on the basis of expected future earnings, the P/E ratio can be thought of as the price for purchasing expected future earnings growth.

A stock's multiple can be very misleading, so it's important to understand just what it is and what it represents. That requires making a distinction between the current and forward-looking multiple. The current multiple is based on the most recent trailing earnings. The forward-looking multiple is based on a combination of expected earnings and the certainty of the earnings estimate. The more reliable the estimate, the higher the multiple is likely to be. And the greater the shock when the estimate turns out to be faulty.

There are many ways to estimate a capitalization rate for a company's expected future earnings. One way to do it is to make the P/E conditional on the dividend payout ratio, the required rate, and the growth rate, as in the formula below. All things being equal, this method produces higher valuations for large dividend payout ratios. But that can be misleading. Growth companies tend to have high multiples and low dividend payouts, whereas mature companies tend to have higher dividend payouts and lower multiples. Growth companies are likely to reinvest free cash flow in their fast-growing businesses, while mature companies pay out a greater portion of their earnings to stockholders who can reinvest (or consume) the cash as they wish.

$$P/E = \frac{Payout\ Ratio}{K - g}$$

As before, the key driver is the growth rate. An increase in the growth rate compresses the risk premium and drives the multiple higher. But as the risk premium compresses, good news is built into the stock price, leaving little margin for error. In a way it's similar to the racetrack: As more and more money is bet on a particular horse, the payoff for placing a winning bet drops. By contrast, the payoff for a long shot is commensurately greater. By analogy, the horse that is the heavy favorite is the high-multiple horse.

For instance, assume Rich's Quilt Factory (RQF) has a dividend payout policy of 40% of earnings, a required return (*K*) of 15%, and a growth rate of 8%. Based on the formula, the expected P/E would be 5.7, which translates to an earnings yield (the inverse of the multiple) of 17.5%. Leaving aside the particulars of the company for the moment, that sounds like an awfully cheap valuation. But whether the valuation is cheap or rich is not absolute; it stands in relation to all the other available alternatives. Therefore, it pays to assess the relative value of a stock in relation to the market as a whole, to its peer group, and with respect to the business cycle. One way to do that is by comparing the stock to the overall market from the top down to see how it is likely to respond to changing circumstances.

TOP-DOWN VALUATIONS

Top-down valuation models seek to predict stock market behavior by estimating the impact that macrolevel variables have on economic activity. From there the question is how it cascades down over various market sectors and the companies that reside within those sectors. By far, the two most important variables are economic growth and inflation. Growth can be further subdivided into cyclical growth, related to the business cycle, and secular growth that is less sensitive to the business cycle and often tied to innovation.

Macromodels focus on two layers of analysis. One concerns the level of the market as a whole; the other is on the relative value of sectors and companies in the market. Assessing whether the overall market is cheap or rich is a market-timing issue. On the other hand, the relative attractiveness of particular companies or sectors has more to do with evaluating a company's competitive position, management, and long-term earning power. Even so, virtually all companies are subject to the vagaries of the business cycle. The real question is to what extent. Consequently, it is important to consider models that take business cycles into account as well as investors' appetite for risk.

In a working paper titled "Stock Market Returns in the Long Run," Roger Ibbotson and Peng Cheng reviewed four classes of models designed to estimate the marginal return to stocks over bonds—the equity risk premium.[11] The first used historical returns data for stocks and bonds to estimate the marginal return accruing to equities. The second examined the type of fundamental data security analysts typically use, like earnings, dividends, and productivity. The third focused on the discount investors demanded for bearing the risk of buying stocks. The fourth relied on surveys of market professionals.

For the sample time period (1926–2000), Ibbotson and Cheng calculated the long-term total return on stocks to be 10.7%. They found the long-term risk premium, defined as the marginal return of stocks over government bonds, to be about 6% when calculated arithmetically, or 4% calculated geometrically. To estimate forward-looking equity risk, Ibbotson and Cheng subdivided stock market returns into various supply side factors such as earnings, the dividend payout ratio, inflation, productivity, GDP (per capita), return on equity, and book value. Their results generally fall in line with historical measures of earnings and productivity.

The study also produced several other important findings relative to stock market valuations. First, growth in corporate productivity was consistent with the economy as a whole. Second, historical expansion of P/E ratios over time accounts for only 1.25% percentage points of the total 10.7% returns for the period. Third, they found a secular decline in dividend yields and the payout ratio, making dividend growth and the payout ratio suspect as measures of corporate profitability and likely future growth. In this regard it is worth noting that tax policy changed to give more favorable tax treatment to dividends after the study was completed. It is possible that the policy change may arrest the downtrend in dividend payout ratios. And going forward, it may provide empirical evidence of whether and to what extent tax policy directly affects stock pricing.

Ibbotson and Cheng's study used six different methods to analyze stock market returns; two were historical, and four were mathematical models. Three of the models are particularly interesting from a trading standpoint because they are simple, convincing, and lend themselves to rules of thumb for evaluating market levels. The three are (1) the building blocks model, (2) the earnings model, and (3) the return on book equity model.

The building blocks model uses three factors to explain long-term returns: inflation, the risk-free rate, and the equity risk premium. The earnings model includes inflation, real growth in earnings per share, growth in the P/E multiple, and income received and reinvested. The return on book equity model incorporates the growth of company book value and return on equity. It is important to note that inflation and growth play critical roles in each of the models. The impact of inflation is direct in the building blocks model. In the others, inflation is felt indirectly. Its effect is refracted through nominal book values and earnings.

The building blocks model is particularly easy to work with. It sets up a framework for evaluating the market's level by backing into an implied P/E forecast, using inflation, the real risk-free rate, and the risk premium as predictor variables. Based on Ibbotson and Cheng's study, a model can be specified as follows:

$$\text{Expected Return} = (1 + \text{CPI}) \times (1 + \text{TIPS}_{10\text{-Yr}}) \times (1 + \text{ERP}) - 1$$

where:

CPI = all-urban CPI, a measure of inflation
$\text{TIPS}_{10\text{-Yr}}$ = the yield to maturity on constant maturity 10-year TIPS
ERP = the equity risk premium

Plugging in reported CPI, the market rate for Treasury TIPS and a range of values for the equity risk premium provide a range of possible discount rates for the market as a whole. Once the expected return is calculated, the sensitivity of the P/E ratio can be evaluated with respect to inflation, real rates, the equity risk premium, and the growth rate. From before, the formula for estimating a market P/E based on the dividend payout, the required return, and the growth rate was:

$$P/E = \frac{Payout\ Ratio}{K - g}$$

By substitution, using Ibbotson and Cheng for the required return *K* the model can be expressed as:

$$P/E = \frac{Payout\ Ratio}{[(1+\text{CPI})\times(1+\text{TIPS}_{\text{10-Yr}})\times(1+\text{ERP})-1]-g}$$

In this form the model can be used to gauge market sensitivity to changes in key variables like inflation, interest rates, and, especially, expected earnings growth. CPI is reported monthly, and TIPS rates are available daily. The long-term average dividend payout ratio, about 48% for the S&P 500, can be used in the denominator. Setting boundary constraints on the various predictor variables allows the model to be used to test sensitivity to changes in the other variables. For instance, holding the real interest rate, the equity risk premium, and the inflation rate constant while varying the growth rate predicts how the P/E ratio (and therefore the market) can be expected to respond to changes in economic growth.

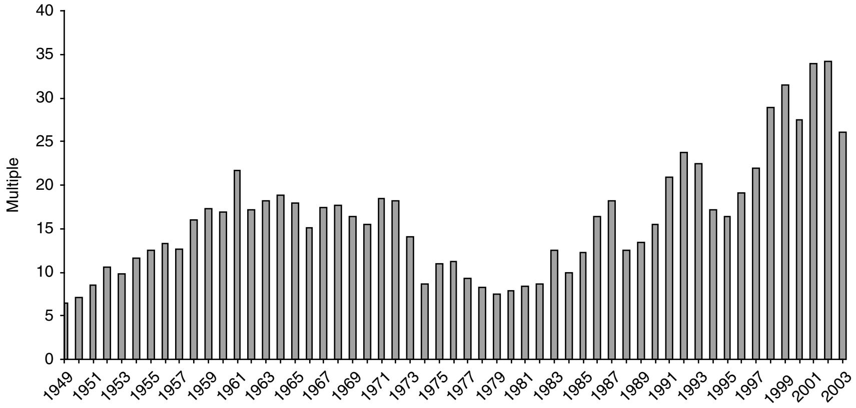
The format of the model lends itself to experimentation by use of different predictor variables, such as GDP growth, Fed policy, the slope of the yield curve, and expected inflation. Regardless, experimenting with the model makes clear the importance of the growth rate estimate. Small percentage changes in expected growth have a dramatic effect on price/earnings ratios. Unfortunately out-of-sample estimations tend not to be as accurate as in-sample simulations would suggest—often by very wide margins.

All forecasts are based on the past; gauging the attractiveness of the market's valuation level is an implicit comparison of present valuations to past ones. Historical analysis tries to ascertain how markets are likely to perform going forward based on how they responded in the past to similar circumstances. Consequently, it provides a much-needed check against the four most dangerous words in the English language: "This time it's different." Especially when valuations get particularly frothy or during bear markets when guys appear seemingly out of nowhere wearing sandwich signs declaring that the sky is falling.

Simple valuation models like the one described here provide a reasonable starting point for gaining perspective, but they can easily leave a false sense of security. Financial markets tend to be far wilder than most models admit. For instance, consider Figure 17.1, a display of the price/earnings ratio of the S&P 500 annually from 1949 through 2003. It averaged about 16, but careened between a low of about 6.5 and a high of 34. Similarly, between December 1990 and April 2006, the average 12-month return of the S&P 500 was about 12.46%, but it ranged between a low of minus 26.62% in November 2001 and a high of 52.14% in July of 1997.

FIGURE 17.1

S&P Trailing P/E (1949–2003)



Data source: Economic Report of the President

DOES THE CAPM REALLY WORK?

A number of empirical studies have been published that cast doubt on whether the CAPM accurately describes real-world stock market pricing. James Davis of Dimensional Fund Advisors has reviewed and summarized several of them.[12] For example, as reported by Davis, Sanjoy Basu documented the "P/E effect" in a 1977 paper. He found that stocks with low price-earnings multiples produced significantly higher returns than did stocks with high multiples. The time period covered in the study was 1957 through 1971. Rolf Banz in 1981 found that small cap firms produce above-average returns, but Basu showed that the small-firm effect was separable from the P/E effect. In another study published in 1983 he found that the P/E effect could not be attributed to firm size, a finding later confirmed in 1989 by Jaffe, Keim, and Westerfield.

DeBondt and Thaler in 1985 found a pattern of return reversals. Stocks that produced inferior returns over a three- to five-year time period had a tendency to produce significantly higher returns in subsequent periods.[13] The opposite result was found for stocks showing superior returns over three to five years. Rosenberg, Red, and Lanstein found that stocks with low prices relative to book value have higher returns than stocks with high price-to-book value.[14] Similar results were found in the Japanese stock market by Chan, Hamao, and Lakonishok.[15] In a related vein of research, Bhandari in 1988 found that highly leveraged firms produced above-average returns, but he also found that increased risk from greater leverage was not accounted for by a higher beta.[16] Jegadeesh and Titman found evidence of short-term momentum in stock prices.[17] Stocks that have done well recently have a tendency to repeat superior performance in the short run.

The critical factor in all these studies is that, on average, differences in returns were not fully explained by beta. According to the CAPM, beta is the great equalizer. To the extent that a stock is riskier than average and is therefore capable of producing superior returns, it ought to have a higher beta that explains away the difference. But in the studies cited above, beta was not sufficient. Other variables seemed to account for the returns differentials, thereby casting doubt on the CAPM framework.

In 1992 Eugene Fama and Kenneth French published a paper titled "The Cross-Section of Expected Stock Returns," which considered firm size, leverage, price-earnings multiples, price-to-book, and beta.[18] Their paper showed that the relation between risk (measured by beta) and return, the foundation for the CAPM, was largely a statistical quirk. Higher returns were actually attributable to firm size, or market cap, a paint that was masked by the fact that small firms tend to have higher betas. When Fama and French employed statistical controls for market cap, beta faded into insignificance as a predictor of returns. However, they found that book-to-market and firm size were strongly correlated to returns.

Fama and French were subsequently able to show that returns differentials could be explained by three factors: the market rate less the risk-free rate, market capitalization, and book-to-market. Note that market capitalization and book-to-market are sources of risk independent of general market risk (beta). According to one interpretation, the coefficients for the small-cap and book-to-market variables can therefore be viewed as risk premiums for those factors.

Lakonishok, Shleifer, and Vishny contend that the Fama/French results stem from market inefficiency. Investors assume (incorrectly) that recent performance will continue into the future indefinitely. Consequently, they overestimate the sustainability of good performance and are overly pessimistic about recent poor performance. They are unpleasantly surprised when a well-performing firm stumbles and surprised again when the management of a poorly performing one successfully takes steps to turn it around. As a result, investors pay too much for growth stocks and discount value stocks much too deeply. Fama and French disputed that interpretation, showing in another paper that factors such as size affect earnings and therefore flow through to stock returns. However, their three-factor model has not been able to explain all the variation in returns. Most notably, it has not been able to explain short-term momentum effects.

SOME CAVEATS

There are a few threads running through the academic research studies that need to be considered. One is that study results have been known to vary with the size of the data set. Another is that the results can vary from era to era. But this may not reflect data problems so much as institutional change. The mechanics of trading, clearing costs, transparency, and liquidity have changed dramatically over the years. Institutional differences may account for how returns have varied over time. For instance, there has been a secular expansion of P/E multiples since the 1920s. Although the impact on returns has been comparatively modest, the time power of money is such that a little bit can go a long way.

Another thread running through the story is the importance of growth. There is little doubt that growth matters, and it matters a lot. The prospect of growth is what attracts fresh capital, entrepreneurs, risk taking, and innovation. The real question is how much to pay for it. Typically the framework for discussion is the tension between valuations for rapidly growing companies and more mature, seasoned ones. If the market is efficiently priced, it should make no difference. But it does.

In 1997 Fama and French published a paper "Value versus Growth: The International Evidence," in which they compared returns of growth and value stocks.[19] They found that value stocks outperformed growth stocks, which is (or was) precisely opposite the conventional wisdom. In discussing this phenomenon, Fama noted that people had a tendency to think that because stocks were priced low that the market thought the companies were relatively distressed, and vice versa. People think that because companies are good, well-run firms that the stock return will be high. But in their study, Fama and French found that growth stock prices were pegged so high that the returns turned out to be low.

In addition to fundamental factors such as earnings and dividends, stock prices may be affected by other factors. Institutional arrangements, including market structure, transaction costs, and the choice of trading venue, need to be examined. Among other considerations, listing standards vary across the different stock exchanges within the United States. Exchanges outside theUnited States are subject to different regulatory regimes.

NYSE LISTING STANDARDS

The New York Stock Exchange (NYSE) is the world's premier stock exchange. As of the end of 2005, the Big Board had 2,672 issuers with securities listed for trading, with a market value of $21 trillion. These include operating companies, closed-end funds, and exchange-traded funds. Of those listing, 435 were foreign companies with a market value of $7.1 trillion.[20]

Domestic companies that wish to have their company's stock listed on the NYSE face certain minimum listing requirements, which are published on the Exchange's Web site. In general, the company's shares must be distributed widely; trading volume must meet a minimum threshold; a company must be financially healthy according to earnings, revenue, or cash flow criteria; and the firm must have been in business long enough to be seasoned. Distribution thresholds can be met by having at least 2,000 U.S. shareholders with 100 shares each, or 2,200 total shareholders and average monthly trading volume of 100,000 for the last six months. Companies with only 500 shareholders need to generate average monthly turnover of 1 million shares.

Listed companies' financial health must pass muster by earning at least $10 million in aggregate pretax income over the last three years and $2 million in each of the most recent two years. An alternative measure allows a firm seeking listed status to show revenues of $75 million for the most recent fiscal year and global market capitalization of $750 million. Real estate investment trusts (REITs) and mutual funds may meet the standard with $60 million in stockholder equity or net assets, respectively.

Corporate governance standards have become increasingly important, and they play a key role in securing a listing on the NYSE. They are laid out in Section 303A of the *NYSE Listed Company Manual*. With certain exceptions the rules require listed companies to comply with Section 303A and with the Sarbanes-Oxley Act passed in 2002. The framework for compliance is a disclosure regime codified in a series of 12 rules designed to protect stockholders, accompanied by enforcement provisions.

The Big Board requires that a majority of a listed company's board of directors be independent. The reason is that the board is supposed to provide oversight of the management and look after the interests of the stockholders. The theory is that a board with a majority of independent directors will be less prone to conflicts of interest and more likely to protect shareholder interests. There is a definitional issue of what constitutes an independent director. The NYSE has therefore adopted an elaborate set of qualifying tests to establish independence. In addition, it requires companies to identify which directors are independent and the basis for the determination. For instance, a director who was a recent employee or who received more than $100,000 in direct compensation from the company over the last three years or who is related to such a person would not be considered independent.

In order to make sure that nonmanagement directors serve as a check on management, the NYSE requires that nonmanagement directors meet regularly without the company's management. The idea is to make communications among nonmanagement directors routine and also to avoid rumor mongering and negative inferences about what is going on in the company. Listed companies are also required to have a nominating committee that is composed entirely of independent directors. Further, the nominating committee must be given the mandate to identify, select, or recommend nominees for the board. The committee is also required to have a mandate to develop a set of guidelines for oversight of the board, the management, and the nominating committee itself.

The NYSE requires that the compensation committee be composed exclusively of independent directors. The committee is responsible for review and approval of corporate goals relevant to CEO compensation, as well as CEO performance with respect to those goals. It is also charged with recommending compensation for executive officers and is responsible for evaluating the performance of the committee.

Listed companies must have audit committees consistent with applicable SEC regulations. The audit committee must have at least three members who possess the requisite financial and business acumen. In addition, the members of the audit committee must be independent. The committee provides oversight of the integrity of the firm's financial statements and compliance with legal and regulatory requirements. It is responsible for overseeing the qualifications and independence of the firm's auditors and the performance of the company's internal audit function and independent auditors. The committee is also charged with preparing the audit committee report for the annual proxy statement as required by the SEC and an evaluation of the committee.

The audit committee is also responsible for annually reviewing the auditor's report with respect to internal quality control procedures, material questions raised by internal audits, and investigations by government authorities. If needed, it is to take steps to deal with these issues and assess the independence of the independent auditor. In addition, the audit committee must meet to discuss the company's quarterly and annual reports, both with management and the independent auditors, and must specifically review disclosures under "Management's Discussion and Analysis of Financial Condition and Results of Operations." With respect to public communications, the committee is responsible for discussing the firm's earnings press releases, financial information, and earnings guidance provided to analysts and the rating agencies.

Finally, the audit committee provides risk management oversight. It meets periodically with internal auditors, management, and external auditors. It reviews management's response to audit problems, sets hiring policies with respect to former auditors, and reports to the board of directors.

Listed companies are all required to adopt and disclose corporate governance guidelines. These include matters such as qualification standards for and responsibilities of directors; their access to management; compensation, and continuing education. The management succession plan should include criteria both for selecting a CEO and evaluating his performance. The board should also conduct a self-evaluation at least once a year to determine its effectiveness and the effectives of its committees.

Listed companies are required to adopt a code of business conduct and ethics. The code applies to directors, officers, and employees; any waivers granted to directors or executive officers must be disclosed. The code of conduct should define what constitutes a conflict of interest, company confidentiality policies, fair dealing, and proper use of company assets; it should promote compliance with law and regulations and encourage reporting of any illegal or unethical behavior. Foreign companies need to disclose how their governance practices differ in material ways from NYSE standards. Each listed company's CEO is required to certify each year that he is not aware of any violations of NYSE corporate governance listing standards.

Finally, the NYSE may issue a public letter of reprimand to any NYSE company that violates a NYSE listing standard.

NASDAQ LISTING STANDARDS

Nasdaq's listing requirements cover much the same ground as those of the NYSE, but there are some important differences. Most notably, the financial requirements for a Nasdaq listing are far less stringent than those of the NYSE. For instance, an initial listing on Nasdaq can require minimum stockholder equity of only $15 million, and the market value of a company's publicly held shares can be as little as $8 million. By way of contrast, the minimum market value of publicly held shares for a Big Board listing is $60 million, and that is for IPOs, carve outs, and spin-offs. The market value of a Big Board company's shares needs to be $100 million. Nasdaq requires a minimum of 400 shareholders; the NYSE requirement is five times as high. Nasdaq requires that a firm's stock have at least three market makers; the NYSE assigns a specialist to the stock, which goes to the heart of the difference between the two exchanges.

Nasdaq corporate governance rules carry much the same spirit as those of the NYSE, although some of the thresholds are different. The rules center on disclosure. Like the NYSE, major Nasdaq rules cover the distribution of annual and interim reports; the independence of directors; the makeup and functions of the audit committee; proxy solicitations; conflicts of interest; stockholder voting rights; and business codes of conduct. The major differences in the two exchanges are not governance rules. The differences are primarily in company size and exchange market structure.

Nasdaq is a dealer market; the NYSE is an auction market. In a dealer market, market-making firms quote bid/offer spreads in the price of a stock, denoting the prices at which they are willing to buy (on the bid) or sell (on the offer). Dealers try to profit by picking up the spread between the bid and the offer. An auction market is different. Prices are set by buyers and sellers when they reveal the best prices at which they are willing to act. An auction market represents the highest price buyers are willing to pay and the lowest price sellers are willing to take. When those prices are the same—when the buyers and sellers meet—a transaction takes place.

The function of the specialist in the New York Stock Exchange system is to maintain an orderly market. When there is a dearth of order flow on either side of the market, the specialist is responsible for stepping in to provide liquidity to the market to keep it running smoothly. The responsibility for assigning a specialist for each listed stock rests with the management of the exchange. This arrangement creates de facto local monopolies on the trading floor and has been the source of considerable controversy over the years.

DOES EXCHANGE VENUE MATTER?

The question arises as to whether the choice of exchange matters, either with respect to liquidity or stock price. Naturally enough, the exchanges say that it does. There has also been considerable academic research on the subject that also suggests that the listing venue matters. In addition to its own research, the New York Stock Exchange underwrites academic research on these matters and publishes the results on its Web site.

One such study published by the NYSE surveyed 67 companies that transferred their primary listing from the Nasdaq market to the NYSE between 2002 and 2005. Of these, 63 were common stock listings; the other 4 were American depository receipts (ADRs). The Exchange study found that intraday volatility for these stocks decreased by 34% following the transfer. They also found that execution costs were lower for the stocks after being listed on the NYSE, with the effective bid/ask spread falling 38%. In addition, the NYSE found that profits accruing to intermediaries fell from an average of 4 cents a share on the Nasdaq to 0.2 cents afterward on the NYSE, a whopping decline of 95%.

These data are suggestive, but the findings need to be placed in perspective. The exchange study does not address whether stocks in general were less volatile in the 2002–2005 period than the years immediately prior. If so, the significance of the reduced volatility in the stocks that changed listing may be overstated or attributed to the wrong cause.

Another factor to consider is the company rationale for changing the listing in the first place. It may be that the companies switching to the NYSE had grown more profitable and transparent with the passage of time, in which case the cause of reduced volatility would be attributable to the company rather then its trading venue.

There is evidence that market structure is an important factor in market quality. Researchers have found that market fragmentation can result in reduced liquidity and increased volatility, although some studies have found that competition can mitigate these effects. In a study of 39 stocks that transferred from Nasdaq to the NYSE during 2002 and 2003, Paul Bennet and Li Wei found an improvement in market quality after the switch due to reduced market fragmentation. They found order flow consolidation to be particularly valuable for less liquid securities.[21] In another study comparing the market quality of NYSE and Nasdaq stocks, Daniel G. Weaver found that Big Board stocks were less volatile than Nasdaq stocks on an intraday basis. Weaver's findings were robust, with the result holding during periods of both market stability and stress, as well as across almost all industry groups studied.[22]

There is also some evidence that trading volume has a tendency to migrate to the New York Stock Exchange in times of stress. Amihud and Mendelson used the fact that NYSE-listed stocks can be traded in different venues to investigate the relative performance of alternative trading platforms, including Nasdaq and electronic communications networks (ECNs). They found that other exchanges tended to compete when conditions were beneficial, but not adverse. NYSE market share tended to rise for higher-priced stocks and during periods of increased volatility. The effect becomes stronger during broad market sell-offs and generally declining liquidity. The authors attribute this to "cream skimming." Other exchanges grab volume during good times, but when push comes to shove during tough market environments, the trade migrates back to the Big Board.[23] Along this line, it should be noted that Nasdaq trading volume is probably overstated, notwithstanding some changes in order handling that were mandated by the SEC and implemented in 1997.[24]

The reason for the volume discrepancy is that Nasdaq is a dealer market, so there is some double counting going on when trades are between dealers and customers. The Street's rule of thumb is that Nasdaq's reported volume is twice its real volume. Trading volume is important for a number of reasons. It provides a sense of market depth and liquidity; it is used to compare brokerage firms; and it is a measure used by the SEC and in financial market research.

Finally, there is the question of trading costs. In general, the research suggests that trading costs are higher for Nasdaq stocks than they are for NYSE stocks. Kee H. Chung, Bonnie F. Van Ness, and Robert A. Van Ness found significantly larger average, realized, and quoted spreads on Nasdaq stocks.[25] The wider spreads could not be fully attributed to differences in the stocks themselves. The authors suggest internalization and payment for order flow as explanatory factors. In another paper, Kee H. Chung and Youngsoo Kim attribute increased Nasdaq trading costs to the structural difference between Nasdaq dealer markets and NYSE auction markets. Specifically, they cite the NYSE specialist's direct responsibility to maintain a liquid market as the causal factor.[26]

SUMMARY

The stock and bond trading businesses are very different. Stocks tend to be priced from the bottom up, while bonds tend to be priced from the top down. As a result, returns correlations among individual bonds are likely to be far higher than those among stocks. Diversification to minimize risk is far more important in stocks than in bonds.

Fundamental stock valuation models are based on the idea that stocks are worth the present value of expected future cash flows. The CAPM valuation model asserts that there are two types of risk embedded in a stock: systematic and idiosyncratic. Idiosyncratic risk can be diversified away; systematic risk cannot. However, recent studies have cast doubt on the efficacy of the CAPM model in practice. New evidence has come to light that suggests that stock returns are determined by factors not adequately explained by beta: the risk premium, book-to-market, and market capitalization. The extent to which these factors influence stock prices, and under what circumstances, would seem to be useful territory to explore in the development of stock-trading strategies.

Market structure is different for stocks and bonds. Bonds tend to trade in the OTC markets; stocks trade on exchanges. Bond firms make their money acting as principals in customer transactions. In the stock-trading business, firms seek to profit mostly by charging commissions. OTC bond trading rules are relatively informal compared to the rules enforced by stock exchanges. Stock exchange trading rules cover rights, privileges, and responsibilities of specialists and member firms. Listing rules define the financial and governance standards companies must meet before they can be listed on either the NYSE or the Nasdaq, the two major U.S. stock exchanges.

The market structures of the NYSE and Nasdaq are different. The NYSE is organized as an auction market. Specialist firms are charged with maintaining orderly and continuous markets in return for which they are designated as exclusive market makers. Nasdaq is organized as a dealer market with competing dealers. Research indicates that NYSE markets are broader, deeper, more liquid, and better able to withstand stress than are Nasdaq markets. Companies whose shares are listed on the Big Board tend to be the bigger, older, and more mature firms. Transaction costs are significantly higher on Nasdaq than they are on the NYSE.

NOTES

[1] Hans G. Heidle and Xi Li, "Information Leakage and Opportunistic Behavior before Analyst Recommendations: An Analysis of the Quoting Behavior of Nasdaq Market-Makers," American Finance Association, 2004 San Diego Meetings.

[2] Kin Wai Chan, Charles Chang, and Albert Wang, "Put Your Money Where Your Mouth Is," draft paper, April 2004, available online at: http://web.mit.edu/finlunch/Spring04/why%20do%20firms%20issue%20analyst%20recommendations%202004.04.05.pdf

[3] See *BusinessWeek:* http://www.businessweek.com/magazine/content/01_18/b3730088.htm

[4] See The Street.com: http://www.thestreet.com/markets/analystrankings/962392.html

[5] See Tech Web: http://www.techweb.com/wire/story/TWB20000223S0009

[6] See Time Digital: http://www.time.com/time/digital/digital50/33.html

[7] See SEC document: http://www.sec.gov/news/press/2003-55.htm

[8] See SEC document: http://www.sec.gov/news/press/2003-56.htm

[9] A. A. Groppelli and Eshan Nikbakht, *Finance*, Barron's Business Review Series, 2000

[10] Benjamin Graham and David Dodd, *Security Analysis: The Classic 1940 Edition,* McGraw-Hill, 2002.

[11] Roger Ibbotson and Peng Cheng, "Stock Market Returns in the Long Run," Yale ICF Working Paper No. 00-44, available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=274150

[12] James L. Davis, *Explaining Stock Market Returns: A Literature Survey,* Dimensional Fund Advisors, 2001, available online at: http://ifa.com/Library/Support/Articles/Scholarly/ExplainingStockReturns.pdf

[13] Werner F. M. De Bondt and Richard Thaler, "Does the Stock Market Overreact?" *Journal of Finance,* 40/3, July 1985, pp. 793–805.

[14] Barr Rosenberg, Kenneth Reid, and Ronald Lanstein, "Persuasive Evidence of Market Inefficiency," *Journal of Portfolio Management,* 11, 1985, pp. 9–17.

[15] Louis K. C. Chan, Yasushi Hamao, and Josef Lakonishok, "Fundamentals and Stock Returns in Japan," *Journal of Finance,* 46, 1991, pp. 1739–1789.

[16] Laxmi Chand Bhandari, "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence," *Journal of Finance,* 43, 1988, pp. 507–528.

[17] Narasimhan Jegadeesh and Sheridan Titman, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance,* 48, 1993, pp. 65–91.

[18] Eugene F. Fama and Kenneth R. French, "The Cross-Section of Expected Stock Returns," *Journal of Finance,* 47, 1992, pp. 427–465.

[19] Eugene F. Fama and Kenneth R. French, "Value versus Growth: The International Evidence," *Journal of Finance,* 53, 1998, pp. 1975–1999.

[20] See the NYSE Web site: http://www.nyse.com/about/listed/1089312755443.html

[21] Paul Bennet and Li Wei, "Market Structure, Fragmentation, and Market Quality," *Journal of Financial Markets,* 2005.

[22] Daniel G. Weaver, "Intraday Volatility on the NYSE and NASDAQ Essays in Market Microstructure," *NYSE*, August 2002.

[23] See Yakov Amihud and Haim Mendelson, "Intermarket Stock Returns: Evidence from NYSE Listed Stocks," November 2004 NYSE;, and *The Exchange* 12/2, February 2005, available at: http://www.nyse.com/pdfs/02-05_newsletter_full.pdf

[24] See Anne-Marie Anderson and Edward A. Dyl, "Market Structure and Trading Volume," *NYSE*, April 2004 (Working paper from July 2003 available online at: http://207.36.165.114/Denver/Papers/MarketStructureandtradingvolume.pdf

[25] Kee H. Chung, Bonnie F. Van Ness, and Robert A. Van Ness, "Are NASDAQ Stocks More Costly to Trade Than NYSE Stocks? Evidence after Decimalization." Working paper, July 24, 2001. Available online from SSRN at: http://ssrn.com/abstract =277817

[26] Kee H. Chung and Youngsoo Kim, "Volatility, Market Structure and Liquidity": http://www.efmaefm.org/efma2006/papers/783236_full.pdf

CHAPTER 18

Equity Instruments: Indexes and Exchange-Traded Funds

If you don't find it in the index, look very carefully through the entire catalogue.

–Sears, Roebuck, and Co. Consumer's Guide, 1897

Market professionals have a well-developed arsenal of strategies for trading in equities markets. Some strategies are likely to work better in down markets; some in up markets. Some trading strategies are flat-out directional; some are market neutral and revolve around relative valuations. Some strategies are focused on individual stocks; some on market sectors. Then there are approaches like risk arbitrage, in which traders speculate on takeover and LBO candidates, some announced, some not. Some spreading strategies involve arbitraging cash stocks against equity index futures and options; others involve statistical arbitraging trading of equity indexes against each other. Making sense of these different approaches requires some sort of benchmark as a frame of reference. Equity indexes, of which there are thousands, are a good place to start.

Four stock indexes stand out in particular: the Dow Jones Industrial Average, the S&P 500, the Russell 2000, and the Nasdaq 100. Each index is widely followed and has actively traded futures contracts listed on it by an exchange. But they use different methodologies and represent different market segments. Accordingly, this chapter begins by examining these indexes in detail. It discusses what they represent, how they are constructed, and what they may tell us about the equity markets.

Benchmark indexes are more than just passive measuring instruments. To the extent that they can be differentiated and replicated, they attract investable funds. As a result, the makeup and properties of a benchmark index can have an important effect on stock prices. To put it in perspective, consider the fact that over $1 trillion in investment portfolios are benchmarked to Standard & Poor's U.S. family of equity indexes.[1] Because so many managers are benchmarked against the index, inclusion in the S&P 500 can have a large short-term impact on a stock's price. In fact, the impact can be sufficiently large that regulators have implored S&P to find a way to find a way to mute the market impact of index inclusion.

PROPERTIES OF GOOD INDEXES

Good indexes are difficult to construct. Inevitably they involve trade-offs, so it is worth reviewing what constitutes a good index and why. A summary of the factors required was published in the spring of 1992 by Jeffrey V. Baily in the *Journal of Portfolio Management*.[2] According to Baily's widely accepted criteria, good benchmark indexes accurately represent a well-defined returns universe; their historical data are readily available; the securities in the index are investable; the daily prices of the components are available; the turnover of index components is rare; and the selection rules for index inclusion are specified in advance.

These criteria may seem self-evident, but there is a lot under the surface that needs to be explored. For instance, defining the targeted returns universe is crucial because it plays a central role in how an index is used. Is it a sector index or a broad-based measure designed to capture the entire equity universe? Historical data need to be available for back-testing. The securities in the index need to be investable—investors have to be able to buy and sell them in the real world—or index valuations are just an illusion. Daily data are necessary for portfolio tracking. Turnover of components needs to be kept to a minimum so that the index reliably measures the same returns universe over time. Rules for index inclusion need to be specified in advance to guard against arbitrary changes in composition.

Equity indexes exert a powerful influence on how and where investment funds are allocated in the marketplace and therefore how the stock market trades. For instance, recent New York Stock Exchange data suggest that as much as 60% of daily volume is attributable to program trading, a form of arbitrage that involves trading baskets of stocks against futures contracts and other derivatives.[3] By implication, both equity index construction methodology and futures contract pricing techniques are crucial variables for developing equity market trading strategies.

EQUITY INDEX CONSTRUCTION

Constructing a stock index involves three major decisions. The first is defining the target returns universe. The second, which flows from the first, is selecting components for index inclusion. The third is choosing a weighting scheme for index components. The process needs to be integrated so that the final product achieves the goal of accurately replicating the targeted returns universe.

Major providers of equity indexes have adopted somewhat different approaches to the craft of index construction. In order to develop trading strategies involving indexes, their companion futures contracts, and the funds based on them, this chapter examines how four indexes are made. Specifically they are the Dow Jones Industrial Average, the S&P 500, the Russell 2000, and the Nasdaq 100. They are examined with respect to the returns universes they are designed to represent; the selection criteria for index inclusion, and the weighting scheme each employs.

THE DOW JONES INDUSTRIAL AVERAGE

The Dow Jones Industrial Average of 30 stocks is the oldest and most widely recognized stock market index in the United States, and probably the world. It has the stated objective of representing large and well-known U.S. companies that are leaders in their industries. Even market professionals benchmarked against the S&P are not immune to the pull of the Dow brand. And in the popular mind, the Dow and "the market" are virtually synonymous.

The simple index methodology used to construct the Dow is markedly different from that used by most of the other major index providers, particularly Standard & Poor's and Russell. The Dow's simplicity is probably a legacy that stems from the era of its invention. First published in May of 1896, the Dow Jones Industrial Average consisted of only 12 stocks. The index was calculated simply by adding up the prices and dividing by 12. Today the index has 30 components, and the divisor is adjusted to maintain returns continuity when there is a component change or a stock split.

Since its initial launch, only one of the original components remains. Although General Electric (GE) was in the original index and is in the index today, there was a hiatus of nine years during which GE was absent from the Dow. It was removed in 1888 and then put back nine years later as a replacement for Tennessee Coal, which had been taken over by J. P. Morgan's U.S. Steel. To this day, the takeover of a Dow stock is one of the few reasons that a stock is replaced in the index.

There are essentially two ways that components can be selected for an index. They can be chosen objectively by applying a set of rules, or the components can be determined subjectively, relying on human judgment. Selection for inclusion in the Dow is subjective, at the discretion of the editors of *The Wall Street Journal*. There are no formal criteria for index inclusion, other than a requirement for the firm to be an established U.S. company and a leader in its industry.

Changes in the Dow's components are made only rarely, in order to maintain index continuity over time. When a change is made, it is usually in response to a major shift in one or more of the components. For instance, a takeover of a Dow component would necessitate a search for a replacement. Or a dramatic shift in a Dow company's core business might prompt the editors to remove it from the index. To minimize disruption to returns continuity, when the editors consider a shift in the Dow, they consider the entire index. This sometimes results in multiple changes to the Dow being implemented at one shot.

By design, the Dow represents the bluest of the blue chips. The total capitalization of the stocks in the index stood at about $3.5 trillion in the fall of 2006, with the average Dow stock having a market cap of about $119 billion. Unlike most stock indexes, which are weighted by some form of market capitalization, the Dow Jones Industrial Average is price-weighted. It is calculated by summing the prices of the 30 Dow stocks and dividing them, not by 30, but by a special divisor. When the Dow was first published, the divisor was simply the number of stocks included in the average. But over time, stock splits, additions, substitutions, and deletions made it necessary to implement a divisor to maintain index continuity.

To see this, consider how a change in index composition affects the price of the index. If a $50 stock were to be removed from the Dow and replaced by a $100 stock, the index would jump without any move in the market as a whole simply because a higher-priced stock replaced a lower-priced one. Conversely, a 2-for-1 stock split would have the effect of reducing the calculated value of index, but would only be a statistical artifact with no economic rationale. Consequently, Dow Jones adopted the policy of adjusting the divisor to compensate for changes in index composition and stock splits. The divisor is set so that the value of the Dow Industrials remains the same at the transition point when index components change. The result is returns continuity. Sort of.

The methodology that Dow Jones employs has the benefit of maintaining returns continuity in the sense that its price changes in response to changes in the prices of the underlying components, made seamless by the use of a divisor. There are no price gaps due to either index substitutions or stock splits. But while returns of the Dow technically maintain continuity in the sense that there are no gaps in the time series, the meaning of what the Dow represents may have changed, thus obviating apples-to-apples comparisons, which is the entire point of returns continuity. There is also a more subtle issue that needs to be considered, unrelated to changes in index components. Stock splits can systematically change the meaning of the index over time.

Consider how a stock split or a component substitution can affect index performance. Assume that the highest-priced stock in the Dow is trading at $100 a share and that its weight in the index is 6%. Because it is the highest-priced stock in the index, its price movements have the greatest impact on index valuation. Now suppose the stock splits 4 for 1. Nothing fundamental will have changed. The shareholders now own four pieces of paper trading at $25 instead of one certificate trading at $100. But the index will have changed in meaning, although not in price. An adjusted divisor keeps the index price the same, but the influence of what had been its biggest component is drastically reduced. If its weighting had been 6% (when the stock was $100), it would drop to about 1.5% (6% * 0.25) after the split, all else equal.

This hypothetical example illustrates an important aspect of index performance: namely, that an event devoid of economic significance (like a stock split) can have a material effect on the index. The effects can be subtle. If for instance the stock that split was a computer manufacturer and the next-highest-priced stock is a drug company, the emphasis of the index will have changed. That can be a problem for measurement continuity.

It actually goes beyond that. A change in calculating the Dow as a result of stock splits can have the effect of altering the volatility of the index as a whole. Consider the impact of a 5% change in the price of IBM before and after a hypothetical 2-for-1 stock split. Before the split a 5% rise in IBM would cause the Dow to jump by 33 points. This can be calculated by taking 5% of IBM's price over the index divisor: $(5\% * 81.75)/0.12493117 \approx 33$. However, the same 5% rise in the split shares would have far less impact because the Dow is price-weighted. Using the same methodology, a 5% rise in split IBM shares would only cause an 18-point jump in the Dow Industrials, because the base number is smaller. In this example the calculation with a new hypothetical divisor adjusted for the split shows the impact to be $(5\% * 42.91875)/0.113962 \approx 18$. And since the same percentage price move in an index component results in a smaller price move for the index, the effect of stock splits is to lessen volatility. A closely related consideration is that stock splits tend to happen to companies with high (in absolute terms) prices. So the Dow likely produces a systematic bias toward lower volatility than the underlying stocks possess.

For the same reason, the price-weighting methodology of the Dow may tend to slow its upside and to cause leadership to rotate within the index, quite apart from external economic forces in the real economy. Companies are more likely to announce a stock split when the stock price is relatively high, typically when the stock price approaches triple digits. In percentage terms the effect is to act like a brake on the Dow's performance. The reason is that stock prices tend to go up when things are going well, and high priced stocks—the ones going up—are more likely to be split. In effect, the Dow's price-weighting scheme biases the index against stocks doing well.

A final consideration is the discriminatory effect that arithmetic price weighting has on potential candidates for index inclusion. It would be difficult for the editors of Dow Jones to include very highly priced stocks in the Industrial Average because they would tend to dominate performance. As of this writing, Google (GOOG) is trading at $455 per share; over five times greater than the highest-priced stock in the Dow, IBM at $91.

THE S&P 500

The S&P 500 is the index of choice for institutional investors. It is the most widely used benchmark of the U.S. equities markets. Its focus is on the large cap segment of the market, accounting for about 80% of the capitalization of U.S. stocks. The market cap of the S&P 500 in its entirety is about $11.5 trillion. The average stock in the index has a market cap of about $23 billion. Note that the 500 stocks of the S&P have a total cap three times the size of the 30 stocks of the Dow, but the average S&P stock has a market cap only a fifth as big.

The aim of the S&P 500 is to reflect the essential risk-return characteristics of the large cap market, consistent with the liquidity normally associated with large cap stocks. At the same time the S&P seeks to minimize substitutions and replacements of index constituents. An oversight committee consisting of economists and index analysts is charged with ensuring that the S&P 500 remains the leading equity index for the United States.

The construction and maintenance of the S&P 500 involves a fairly elaborate series of published guidelines, both for index additions and removals. To maintain transparency in the process, S&P publishes on its Web site an extensive collection of policy statements and research papers relating to the S&P index family and its maintenance.[4] Decisions to change the index, including component additions and deletions, are cloaked in secrecy and announced to the public via press release after the close of trading.

The criteria for adding a component to the S&P include seven areas of consideration. The firm must be a U.S. company; market cap must exceed $4 billion; the company must be financially viable; the stock must be reasonably liquid; the public float must be at least 50% of the company's stock; addition to the index must add (or maintain) sector balance; and the issuer must be an operating company. Companies that substantially violate the criteria are eligible for index removal, as are companies that have merged, been acquired, or changed substantially enough so that they no longer meet the original criteria for addition to the index.

The rationale for first criterion, that the firm must be a U.S. company, is that the S&P 500 is supposed to be representative of the U.S. market. To determine if a company is in fact a U.S. company, S&P considers such factors as the geography of a company's operations, its corporate structure, accounting standards, and where its securities are listed for trading. Determination of financial viability is based on reported earnings over the last four consecutive quarters. Mutual funds generally fall outside the criteria for index inclusion, but REITs are acceptable. Initial public offerings (IPO) need to be seasoned for 6 to 12 months before they are considered for index inclusion.[5]

The S&P strives to minimize turnover of index components. Consequently, temporary violation of a criterion for index addition does not necessarily constitute automatic cause for removal; ongoing circumstances must warrant it. Changes made to the index are as needed, not by annual or semiannual reconstitution. Nevertheless, changes to the index are not infrequent, and they can have substantial market impact. For instance, in between January 31 and June 5, 2006, there were 13 additions/deletions from the S&P 500, involving 26 firms. One of the firms added to the S&P 500 was Google.

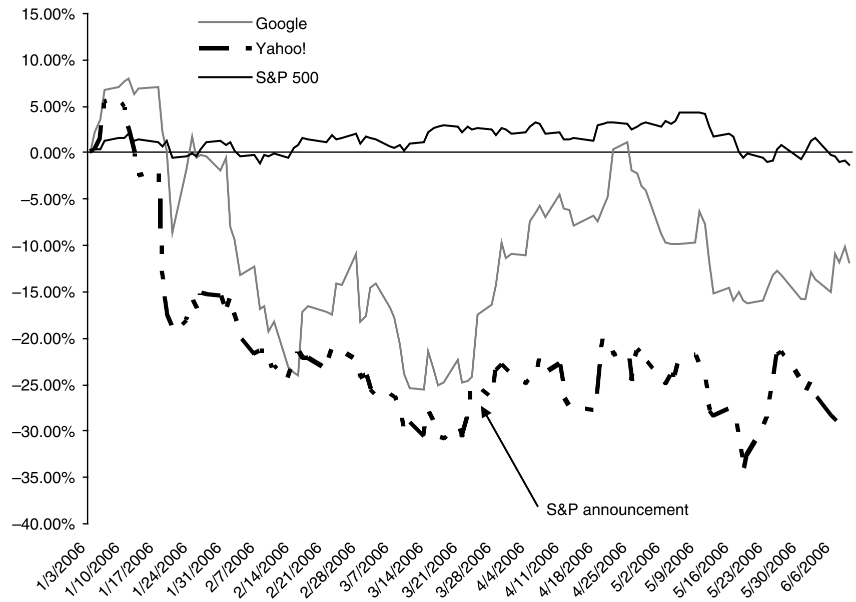
A glance at Figure 18.1 gives a good snapshot summary of the story. From the beginning of 2006, Internet search engine firms like Google and Yahoo! were under pressure compared to the S&P 500. Then after the close of trading on March 23, Standard & Poor's announced that Google would be added to the S&P 500 on March 31. Google, which closed at $342 per share just before the announcement, opened 26 points higher at $368 the following morning. It quickly began to outperform search engine competitor Yahoo!, and over the next month made up most of its performance deficit versus the S&P 500, as can be seen in Figure 18.1.

The story of Google's entry into the S&P 500 illustrates the importance of benchmark indexes to modern stock markets and trading strategy. Once the announcement was made that Google would be included in the benchmark index, it was bound to outperform. Estimates are that somewhere in the neighborhood of $1 trillion worth of investments are benchmarked to the S&P equity index family. Moreover, a good deal of the money is specifically dedicated to index replication. As a consequence, many funds designed to replicate the S&P 500 had no choice but to buy Google once it was added to the index. The irony is that many were not permitted by their rules to buy the stock before the March 31 implementation date; they had to wait until the stock was actually included in the index calculation. The result was that the arbitrage community had a field day. They simply bought stock after the announcement knowing full well that the funds had no choice but to buy, essentially at the market, when the March 31 implementation date rolled around.

As important as index additions and deletions are to the S&P 500, the weighting scheme adopted by Standard & Poor's is probably more important. Unlike the Dow Jones Industrials, which is weighted by price, the S&P 500 is weighted by market capitalization. This feature has played a prominent role in its success as the institutional benchmark of choice.

FIGURE 18.1

Cumulative Price Returns: Google, Yahoo!, S&P 500



Data source: Yahoo! Finance

WEIGHTING THE S&P

In addition to being measuring instruments, the entire family of Standard & Poor's U.S. equity indexes is meant to support a variety of investment products. The list includes index mutual funds, portfolios, exchange-traded funds, index futures contracts, and options. Accordingly, the indexes are designed to be liquid and tradable. The weighting scheme for the S&P 500 reflects that design philosophy.

One of the original criteria for index construction involved defining the available universe of returns. The word *available* is the operative one. For an index to be accurate (and tradable), the universe of returns has to be available in the sense that it can be replicated. Portfolios are not made up of theoretical issues purchased at hypothetical prices. They are made up of real stocks and bonds, bought and sold in a marketplace in which there is a buyer for every seller. That is why it makes sense for an index to be weighted by the number of shares outstanding that are actually available for trading.

The weighting scheme that Standard & Poor's adopted for the S&P 500 is a float weight, which it implemented with a phase-in that was completed in September 2005. Prior to the adoption of a float weight, the index was calculated using a straight cap weight, in which the number of shares issued was multiplied by the stock price to give it an index weight.

The S&P index value is the quotient of the total float-adjusted capitalization of the 500 stocks and an index divisor. The index divisor is used to maintain index returns continuity, in similar fashion to the Dow Jones Industrials. It is adjusted when there are changes in the constituents' share capital, including changes resulting from index additions, deletions, new share issuance, share buybacks, rights issues, and spin-offs. The divisor is adjusted so that the newly weighted index would produce the same index value as the old index at the instant in time the index is transitioned over to its new weights. This cap-weighting scheme causes the largest 50 or so stocks in the index to dominate, as a result of which the S&P 500 correlates rather closely with the Dow 30.

But the float weight does not erase differences between the two indexes, and S&P goes to considerable effort to get the float right. For the purpose of index calculation, S&P counts shares outstanding as defined by the rules of the Financial Accounting Standard Board (FASB) consistent with generally accepted accounting principles (GAAP). Using that count as the base, S&P implements an adjustment to float-weight the count so that it reflects only shares that are available for trading.

For the purpose of determining stock available for trading, the basic distinction Standard & Poor's makes is between investment buyers and strategic buyers. Investors in a company's stock do so presumably because they wish to generate returns through price appreciation, receipt of dividends, or both. Strategic buyers own the stock because they are interested in control. Shareholders deemed interested in control would include board members, founders, and owners of large blocks of stock. Holdings by other corporations and governments are similarly deemed to be for the purpose of control as opposed to investment purposes. Holders of large blocks are not ipso facto defined as holding for control. Mutual funds and trust departments typically hold large positions in companies, but typically for investment purposes.

To implement the float count, Standard & Poor's lists three holding categories of presumed strategic owners of stock. The first includes holdings by other corporations, venture capital, private equity, leveraged buyout groups, and strategic partners. The second includes all levels of holdings by government entities. That does not include public pension funds for which the beneficial owners are government employees, as opposed to the government. The third category includes current and former officers and directors, founders, and family trusts run by the same. Holdings of trusts, foundations, pension funds, employee stock ownership plans (ESOP), and other investment vehicles controlled by the company are treated as strategic rather than investment holdings.

Once the available float has been determined, each company is assigned an *investable weight factor* (*IWF*), which is defined as the available float *A* divided by the total number of shares outstanding *T*.

$$IWF = \frac{A}{T}$$

Another factor that affects company index weights has to do with the not unusual practice of a company's issuance of multiple classes of common stock. For instance, some types of shares may have different rights attached to them. Some may have voting rights; others may not. However, in Standard & Poor's indexes a company is only represented by one class of stock, generally the one that is most liquid. For these companies Standard & Poor's calculates a weighted average IWF for the shares. Details of the calculation methodology are posted on the S&P Web site.

Once the index weight factors of the constituent companies have been determined, the value of the S&P 500 is determined as the sum of the product of the prices of the stocks, the number of shares of each outstanding, and its IWF over the divisor. All are summed over the index divisor.

$$\text{S\&P 500} = \frac{\sum_{n=1}^{500} (P_j \cdot S_j \cdot IWF_j)}{\text{Divisor}}$$

where:
P = the price of the stock j
S = the number of shares outstanding j
IWF = investment weight factor j

THE RUSSELL EQUITY INDEXES

The methodology of the Russell family of broad-based indexes is similar to that of Standard & Poor's. They are designed to be comprehensive representations of investable U.S. equity markets. The indexes are float-weighted using an adjusted market cap; they include only common stocks with the constituent firms incorporated only in the United States and its territories. The Russell index design methodology seeks to achieve three goals. First, the indexes serve as a performance benchmark for money managers. Second, they are a reference point for asset allocation decision making. Third, they are designed to be investable and replicable so as to accommodate passive index investment strategies.

The Russell family of broad-based indexes has three main components: the Russell 3000, 2000, and 1000. The Russell 1000 represents the top 1,000 companies ranked by market capitalization in descending order. The Russell 2000 is the bottom 2,000 companies ranked by market capitalization, or the Russell 3000 minus the Russell 1000.

Of primary interest for this chapter is the Russell 2000, which is an index designed to be a comprehensive representation of the small cap segment of the U.S. equity markets. The difference in capitalization between the Russell 2000 and both the Dow and the S&P 500 is stark. To put it in perspective, as of the Q3 2006, the market cap of the average Russell 2000 stock was in the neighborhood of $1 billion. That is less than 5% of the size of the average S&P stock and less than 1% of the size of a Dow Industrials component. The largest stock in the Russell 2000 was a shade over $5 billion, a small fraction of the average cap of a Dow stock. (Stats describing the stocks in the Russell indexes as well as methodology are available at the Russell Web site).

The Russell 3000 represents 98% of the U.S. equity market. The Russell 2000 is a subset, so the first order of business is to take a look at how the Russell 3000 is made. Russell starts by ranking U.S. companies by market capitalization with the largest on top. The largest 3,000 stocks that are not excluded by some other criteria are included in the Russell 3000. The index is then reconstituted annually.

The main, but not exclusive, criterion for index inclusion is a company's total market capitalization, which is calculated by summing all classes of common stock as of May 31 of that year. The primary trading vehicle is the most liquid stock, generally determined by trading volume adjusted for price. Classes of stocks, like tracking stocks, that are independent of the primary trading vehicle are considered for inclusion separately. Beyond market cap, there are several other initial criteria that have to be met for inclusion in the 3000. One is the requirement that the stock is trading at or above $1 per share. Once in, it is not automatically deleted from the index if it falls below $1 per share. However, if it is still trading below $1 at the next index reconstitution, it will be deleted.

Foreign companies (like Schlumberger) listed in the United States are excluded from consideration, the idea being that the index is supposed to represent U.S. investment opportunities. Other types of securities that do not truly represent a company's primary equity financing instrument are excluded from consideration. These include preferred stock, convertible stock, redeemable shares, warrants, rights, trust receipts, royalty receipts, closed-end investment companies, and limited partnerships. Berkshire Hathaway is excluded as an exceptional case.

The Russell 3000 is float-weighted, which is to say that the stock price is multiplied by the outstanding shares available for trading. Large blocks of stock (10% or more of capitalization) held by other listed companies or individuals are not considered part of the floating supply. Russell uses SEC corporate filings as its source of information for these decisions. If these data are missing or appear questionable, Russell goes to other sources.

In addition, Russell defines ESOP shares, shares not listed on U.S. exchanges, and IPO lock-ups as not available for trading.

Russell reconstitutes its index family annually. Actual reconstitution is implemented effective the last Friday of June. Changes in index constituents are announced ahead of time, subject to change if there is an interceding corporate event before the change is implemented. The most important difference between the S&P and the Russell 2000 is not so much methodology as the vast difference in their market caps.

THE NASDAQ 100

The Nasdaq 100 is made up of the 100 largest market capitalized, nonfinancial domestic and international companies listed on the Nasdaq stock market. The index is dominated by technology companies, although many industry groups are represented, including computer software and hardware, telecommunications, biotechnology, and the retail and wholesale trade.

In many respects the Nasdaq 100 is quite different from the more conventional indexes published by Dow Jones, Russell, and Standard & Poor's. For instance, the Nasdaq 100 includes foreign and domestic securities. Nasdaq uses a modified cap-weighting scheme, which Nasdaq reviews quarterly. After the quarterly review Nasdaq adjusts component weights using a proprietary algorithm with predetermined criteria.

To be included as a member of the Nasdaq 100, a company must be listed exclusively on the Nasdaq stock market, with exceptions for companies that had a dual listing prior to 2004. The securities must be issued by a nonfinancial company, must not be in bankruptcy, and must have average daily trading volume of at least 200,000 shares. Foreign companies must have listed options on a recognized U.S. exchange or meet the eligibility requirements for listing options. Only one class of security per issuer is eligible for inclusion, and the security must be "seasoned," which generally means that it has to have been listed for a minimum of two years.

To be eligible for continued inclusion in the index, the company must generally maintain compliance with the original eligibility criteria. Further, the security needs to have an adjusted market capitalization equal to no less than 0.10% of the adjusted aggregate capitalization of the Nasdaq 100 at the end of each month. Companies failing to meet this criterion for two consecutive months are removed from the index. Finally, to be included in the index, a company may not have a financial statement with an audit statement that is currently withdrawn.

Nasdaq annually reviews rankings of securities by market capitalization. Index eligible securities already in the index and ranked in the top 100 by market cap stay in the index. Securities previously in the top 100 and currently ranked between 101 and 125 are retained in the index. Securities that fail to meet these criteria are replaced by the one with the largest market cap that meets all the other eligibility criteria.

Nasdaq indexes are price return indexes so they are not adjusted to take into account dividend payments. When adjustments are made due to events such as special cash dividends or rights issuances, they are implemented after the market has closed. The basic formula used to calculate Nasdaq indexes is:

$$Index = \left(\frac{AIMV}{ABPMV} \right) \times Base$$

AIMV is the aggregate index market value, which is the aggregate value of the securities in the index; ABPMV is the adjusted base period market value, which Nasdaq uses to maintain index returns continuity, in similar fashion to the divisor for the Dow Industrials. The current base value for the Nasdaq 100 of 125 was set on January 1, 1994.

EXPLORING AND EXPLOITING INDEX DIFFERENCES

Differences in the index methodologies and targeted returns universes chosen by Dow Jones, Nasdaq, Russell, and Standard & Poor's have resulted over time in nontrivial differences in returns and volatilities. Those differences provide a source of arbitrage trading profit that can be exploited using futures, options, and exchange-traded funds (ETFs). Exploiting them depends on a two-step process. The first is identifying and isolating differences in the indexes and analyzing when and how they are likely to assert themselves in pricing. The second is to acquire familiarity with index-based derivatives—how they trade and how they are priced. Equity index derivatives listed on exchanges come in two flavors. One is in the form of ETFs. The other is in futures contracts that trade against the cash index.

EXCHANGE-TRADED FUNDS

Exchange-traded funds represent shares of stock in a fund or unit investment trust. ETFs are relatively new, combining the liquidity of a listed stock with the diversification advantages normally associated with a traditional mutual fund. The structural features of ETFs that facilitate this are briefly touched on here. But for a detailed analysis readers are advised to consult David Lerman's book: *Exchange Traded Funds and E-Mini Stock Index Futures*.[6]

Typically an ETF seeks to replicate the returns of a benchmark stock or bond index, although recently some ETFs have been launched against baskets of commodities. While ETFs are similar in many respects to traditional mutual funds, there are some important structural differences that allow ETFs to trade like individual listed stocks on an exchange during market hours (hence the name). The major innovation that made this possible is a creation/redemption process that makes an ETF easy to arbitrage against its underlying benchmark. As a result, ETF prices tend to be kept closely in line with the underlying benchmark, thereby enhancing liquidity.

There are substantial differences between ETFs and the more traditional fund structures. Unlike traditional mutual funds, ETF issuers do not sell shares directly to the public. Instead, they issue shares in large blocks called *creation units* usually in increments of 50,000 shares and up. Creation units, which replicate a target index, are typically bought by large institutions. The creation units are then broken up into shares in a fund that can be sold in the secondary market. The process can be reversed. Creation units can be returned to the issuer. They can also be swapped for the basket of securities that constitutes the underlying index in the same proportion as their index weights. But they cannot be redeemed for cash like conventional mutual funds. Consequently, ETFs may not market themselves as mutual funds.

One major advantage of the ETF structure is that the component parts can be easily arbitraged against the benchmark index the ETF is designed to replicate. By design, the creation/reconstitution process facilitates the arbitrage trade, to keep the market liquid and efficiently priced. For instance, if and when the sum of the parts trades at less than the value of the whole, arbitrageurs buy up the pieces and issue new ETF shares. Conversely, if the component parts trade at a premium to the underlying index value, arbitrageurs will typically step in and sell the individual stocks short, buy the underlying ETF, and redeem it for the underlying shares. The relative ease of this creation/redemption process is not accidental. It is an integral feature of the ETF structure, designed to remove barriers to arbitrage.

THE ETF CREATION/REDEMPTION PROCESS

Creating an ETF is a multistep process. First, authorized participants assemble a collection of stocks that replicates an underlying benchmark index. (Authorized parties would typically be market makers, specialists on the exchange floor, large institutional investors, and professional arbitrageurs.) Second, upon collecting the underlying stocks, the authorized participant notifies a distributor, usually either APLS Mutual Funds Services or SEI Investment Services, of its intention to assemble a creation unit.

The creation units are composed of large blocks of stock that replicate, or very closely mimic, the behavior of the underlying benchmark index. Usually creation units come in blocks of 50,000 shares of stock, although they can range from 20,000 shares on up. The distributors handle the back-office work and certify to the custodian that the creation unit is in good form. After certification, the basket of stocks is delivered to the trustee by the authorized participant, along with the requisite fees, accrued dividends, and transfer charges. Upon receipt of the portfolio, the custodian issues shares in the ETF.

The redemption process is a mirror image of the creation process. A request for redemption notice is sent to the trustee. Using close of business prices, the trustee delivers a portfolio of stocks that is identical to the tracking index in composition and weighting. In addition to stock, the redeeming owner receives a cash payment equal to accrued dividends, accrued interest, and capital gains (or losses) accrued since the last distribution, net of redemption fees.

The ease of the ETF creation/redemption process serves to increase liquidity and keep pricing efficient. It is relatively simple to arbitrage an ETF basket against its component stocks, or against futures contracts on the same underlying index. To the extent that arbitrageurs see prices get out of line, either between ETFs and their component parts, or between ETFs and their companion futures contracts, they will buy one side and sell the other. The result is a steady supply of bids and offers for the underlying securities, futures contracts, ETFs, and options. This combination of instruments and trading venues minimizes ETF tracking error, draws in liquidity and helps to explain the growing popularity of ETFs.

ETF STRUCTURES

ETFs generally fall into one of three types: unit investment trusts (UITs), open-ended mutual funds, or grantor trusts. Unit investment trusts fall under the Investment Company Act of 1940 and are thus subject to its regulatory requirements. One consequence is that statistical replications of the index are not permitted. A unit trust ETF must exactly replicate its underlying reference index. Moreover, the trustee is not permitted to lend out the securities held in the replication portfolio. Dividends cannot be reinvested. Typically they are paid to the trustee and redistributed to ETF shareholders. Some of the more popular ETFs use this structure, among them SPDRs, DIAMONDS, and QQQQs, based, respectively, on the S&P 500, the Dow Industrials, and the Nasdaq 100.

Other ETFs are structured as open-ended mutual funds. As such they are not encumbered by the same restrictions as unit investment trusts. For instance, an ETF structured as an open-ended mutual fund is not required to exactly replicate its underlying index. It may instead use statistical techniques to construct a portfolio that mimics the index without exactly replicating it. In addition, an ETF using this structure may lend securities held in the portfolio, reinvest dividends, and include derivatives in the portfolio. Unlike closed-end mutual funds that typically trade at a discount, the ETF creation/redemption process keeps the shares trading close to intrinsic value. Barclay's iShares is an example of an ETF structured as an open-ended mutual fund.

A third type of ETF structure, the grantor trust, was created by Merrill Lynch. This structure involves using holding company depository receipts (HOLDRS). Typically they represent ownership in a basket of 20 companies' securities which remain in the trust. Over time attrition may occur though mergers and corporate restructurings. The HOLDRS structure does not permit portfolio optimization; the underlying set of securities is exactly replicated. Consequently, the use of derivatives is prohibited. Nor can securities in the portfolio be lent out. Investors in HOLDRS retain voting rights in the underlying shares, and dividend payments are simply passed through to the investors. Creation and redemption units can be produced with as few as 100 shares.

Specific structure aside, ETFs enjoy a number of advantages not available to investors in traditional mutual funds. The most important is that ETFs can be traded during normal market hours. Moreover, because they are listed just like individual stocks, the full array of order types can be used to effect transactions in them, including stop, limit, and market orders. Typically, traditional mutual funds can only be bought or sold at calculated net asset value (NAV) based on the market's close, normally 4 p.m. Eastern Standard Time. In addition, short selling of ETFs is permitted and the uptick rule does not apply. Traditional mutual funds cannot be sold short. ETFs can be traded on margin; in general, mutual funds cannot be.

Certain tax advantages are associated with ETFs. In-kind redemption in which shares in the fund are exchanged for the underlying stocks is not usually treated as a taxable event. (However, sales of the underlying shares afterwards are treated as taxable events.) By contrast, traditional mutual funds distribute capital gains and losses reflecting fund market activity. But this has more to do with fund benchmarking than fund structure. ETFs generate low turnover to the extent that benchmark composition is stable, a situation that applies equally to conventional but passive mutual fund index funds. It is turnover by actively managed mutual funds that generates taxable events over which fund owners have no control. In addition, it should be noted that active trading of ETFs will generate tax and transaction costs as well.

Annual expense fees for ETFs vary by the fund, but tend to be low for the bigger, more widely followed indexes. For instance, SPDRS have an annual expense ratio of only 0.1%. The expense ratio for DIAMONDS is 0.18%; for the Russell 2000 it is 0.20%; for the QQQQs the annual expense ratio is 0.20%. On specialty and foreign indexes the expense ratios can be far higher, often ranging from as low as 0.3% to as high as 1% for the iShares Asia/Pacific ex Japan Fund.[7] When calculating trading costs for ETFs, commissions have to be added to the mix because brokers charge fees for buying and selling them, just the way they do for individual stocks.

Traditional mutual funds do not charge commissions. Some charge up-front sales loads; others charge fees at liquidation; while others charge penalty fees to liquidate positions before certain minimum time periods have passed. Expense ratios and fees for these funds vary widely, with up-front sales loads sometimes reaching as high as 6%.

THE IMPACT OF FEES ON RETURNS

Fees matter, and they matter a lot. To put it in perspective, consider the fact that over a 30-year investment horizon the difference between $100,000 invested at 8% and 9% (compounded semiannually) is a whopping $350,000. One percentage point is about the difference between the fees charged by the big ETFs cited above and the fees charged the average mutual fund investor. According to the *2005 Investment Company Fact Book* the average investor paid mutual fund fees of 1.25% in the year 2003.[8] That is a solid 1 percentage point higher than the fees charged on SPDRS, DIAMONDS, and QQQQs, and 95 basis points higher than fees on the Russell 2000. To be sure, buyers of ETFs have to pay commissions to purchase them, but the commissions are trivial for buy-and-hold investors.

An important source of hidden fees is transaction costs. John Bogle, founder of the Vanguard funds (and scourge of much of the industry) points out that portfolio turnover of mutual funds, which averaged around a 17% annual rate in the 1950s, soared to 108% by 2000, with no appreciable gain in fund performance.[9] Which is not to say that brokers were not appreciative. Adding insult to injury is the tax code. Fund managers do not manage for tax impact, with the result being that active managers make both the IRS and the brokers very happy. Capital gains realized by portfolio managers flow through to investors as taxable events, not to mention the transaction costs associated with high levels of portfolio turnover.

Transaction costs can exact a punishing toll on returns. But there are ways to minimize or at least reduce transaction costs while still maintaining an active trading profile. Execution costs for trading in equity index futures are comparatively low. Virtually all the major equity indexes have futures contracts listed on them; the markets are transparent and easy to access electronically. For active equity index trading, futures contracts warrant a close look. Accordingly, the next chapter takes a look at trading strategies for equity index futures.

SUMMARY

Equity indexes are excellent measuring instruments for evaluating portfolio managers and trading strategies. They also present low-cost alternatives for targeting specific returns universes, diversifying holdings, and implementing absolute return trading strategies as well as relative value and arbitrage strategies. The key to strategy implementation, particularly with respect to arbitrage, is to understand differences in equity index design.

Equity indexes can be used to execute trading strategies directly through the use of ETFs, and indirectly through the use of equity index futures contracts. Virtually all the major equity indexes have companion listed futures contracts on them as well as ETFs and options. Just as it is important to understand the mechanics of index construction, it is important to know the nuts and bolts of the ETF creation and reconstitution processes. In addition, futures contracts can serve as efficient substitutes for ETFs and can be used to acquire or lay off various segments of equity market risk exposure.

NOTES

[1] See the S&P Web site: http://www2.standardandpoors.com

[2] Jeffrey V. Baily, "Are Manager Universes Acceptable Performance Benchmarks?," *Journal of Portfolio Management,* 18/3, Spring 1992, pp. 9–13.

[3] These data are available online at: http://www.nyse.com/marketinfo/datalib/1152267398806.html

[4] See http://www2.standardandpoors.com/

[5] Details can be found in the Policies and Methodology section of the S&P Web site: http://www2.standardandpoors.com

[6] David Lerman, *Exchange Traded Funds and E-Mini Stock Index Futures,* John Wiley & Sons, 2001.

[7] These and other expense and turnover data are available at Yahoo! Finance.

[8] See *The Investment Company Fact Book* at: http://www.ici.org/factbook/05_fb_sec3.html#trends

[9] See the Bogle Financial Markets Research Center at: http://www.vanguard.com/bogle_site/sp20010128.html

CHAPTER 19

Equity Index Futures, Basis Trading, and Market Timing

> If stock market experts were so expert, they would be buying stock, not selling advice.
>
> *–Norman Augustine*

The S&P 500 is by far the most actively traded stock index futures contract. It is listed on the Chicago Mercantile Exchange (CME or the Merc). Initially launched as a pit-traded contract, the Merc launched an electronic version on its GLOBEX trading platform in 1997. When it was first launched, the CME E-mini® S&P 500 traded about 7,500 contracts a day. But easy electronic access combined with a bull market in stocks drove explosive growth in trading volume. By the end of 2004, the E-mini S&P average daily trading volume had risen to about 700,000 contracts. Two years later trading volume grew enough so that on some days trading volume exceeded 1 million contracts. At the Chicago Board of Trade, which lists a futures contract on the Dow Jones Industrials, volume also rose. The Dow has recently traded over 200,000 contracts in a single day.

Equity index futures contracts are intended to replicate the performance of an underlying stock index. In so doing, the futures contract turns a measuring instrument (the index) into a tradable asset. By allowing an index to be bought or sold at one shot, the futures contract enables market participants to synthetically gain or shed exposure to the risks and returns offered by an entire basket of underlying stocks, weighted in proportion to their weights in the benchmark index. Moreover, the transaction can be either highly leveraged or fully collateralized; the extent of the leverage depends on the preferences of the trader or risk manager, within a wide set of parameters.

Just as the underlying indexes have their differences, the companion futures contracts at the exchanges have some slightly different design features. There are also some important differences in the rules and regulations that govern trading in stocks and stock futures contracts. Stock index futures are regulated by the Commodity Futures Trading Commission (CFTC); stock trading is regulated by the Securities and Exchange Commission (SEC). Stock index futures, which settle for cash, mimic the price action of an underlying stock index, but futures contracts do not give owners an equity stake in any of the underlying index companies. A position in a stock index future is a side bet. A share of stock is not a side bet, it represents an ownership stake in a company.

While positions in futures contracts can be fully collateralized with cash deposits or Treasury bills, it is infrequently done. Typical margin deposits on stock index futures run in the neighborhood of 5% to 20% of the notional value of the contract position. That implies leverage of anywhere from 20:1 to 5:1. On the other hand, the typical initial margin deposit on a levered stock position is 50%, implying a far lower leverage factor of 2:1. The clearinghouse serves as a safeguard for stock index futures trades; the Securities Investors Protection Corporation (SIPC) acts as a safeguard for investors' segregated funds in the event of a broker's bankruptcy. Rules for short selling vary as well.

Stock index futures can be sold short whenever the market is open; shares of stock can be sold short either on an uptick or on a zero-plus tick. (*An uptick* occurs when the current trade is at a higher price than the one immediately preceding it. *A zero-plus tick* occurs when the most recent price is higher than the most recent different price.) Under certain circumstances shares of stock can be sold short when there is an offsetting derivatives position. The uptick rule, a legacy of "bear raids" of the 1930s may not be around forever. The SEC has been conducting pilot studies by relaxing the rule on some of the bigger, more liquid stocks, to assess its usefulness.

NOTIONAL VALUES AND STOCK INDEX FUTURES

The design features of stock index futures at the major derivatives exchanges are essentially the same. The notional value of the equity index contract is the product of the underlying cash index and a multiplier. The methodology is flexible enough to accommodate various sized traders by allowing several different sized versions of the same index to trade simultaneously. In general, the larger and smaller versions of the same indexes at a single exchange are fungible; positions in the smaller version can be aggregated and delivered into the larger version. That keeps prices in line across the various sized contracts.

The notional values and multipliers of equity futures contracts vary by index. The CME's E-mini S&P futures contract has a $50 multiplier;

but the E-mini Russell indexes have \$100 multipliers, and the E-mini Nasdaq 100 has a \$30 multiplier. The mini-sized Dow, by far the most popular version at the Chicago Board of Trade, has a \$5 multiplier. The many different indexes and multipliers result in a wide variety of notional values for electronically traded stock index futures.

For instance, the notional value of the E-mini S&P 500 toward the end of 2006 is about 1400 × \$50 = \$70,000. The mini-sized Dow is about 12,350 × \$5 = \$61,750, and the big Dow is five times larger at \$308,750. The tick values of the various contracts vary as well. (*The tick value* refers to the dollar value of a price change at the minimum increment permitted by contract rules.) For the S&P E-mini the minimum tick increment is 0.25 index points, so the tick value of the contract is equal to \$12.50, or 0.25 × \$50. The multipliers, minimum price ticks, tick values, and approximate notional values of all the indexes traded at the Merc and the CBOT are listed on their Web sites.

Trading in stock index futures is increasingly electronic. The Chicago Merc has trademarked its electronic stock index futures as "E-mini" contracts. They trade the E-mini S&P 500, the E-mini Russell 2000, and others. At the Chicago Board of Trade, the most actively traded equity index is the Mini-sized Dow Jones Industrial Average, so-called because of its \$5 multiplier. The electronic versions of stock index futures often have smaller notional values than the ones traded physically in the pits. Prices of the larger and smaller versions are kept in line by traders who are constantly on the lookout for the opportunity to sell in one venue and buy in the other at a cheaper price. It should also be noted that while the S&P 500 trades exclusively at the Merc, Russell indexes are dual listed. The New York Board of Trade lists futures and options on the Russell 1000, 2000, and 3000 Indexes as well as the NYSE composite, the Russell 1000 Value, and the Russell 2000 Growth and Value Indexes.

PRICING EQUITY INDEX FUTURES

Trading the *basis*—the spread between the cash value of a stock index and its companion futures contract—drives equity index futures prices. Although there are lots of equity index futures listed at the various futures exchanges, there is one pricing model that is broadly applicable to all. It is basically the cash-and-carry model, which we applied earlier to T-bond futures. But some important modifications have to be applied to the bond futures version of the model to take into account institutional and contract design differences.

For instance, stock index futures settle for cash. Consequently, the delivery options that apply to bond futures are not attached to stock index futures. The financing mechanisms for stock and bond positions are different as well. Bonds are financed through the RP markets. Normally, there is positive carry associated with positions on the long side of the bond market and vice versa. For stocks the opposite is true; positive carry normally accrues to positions on the short side of the market. A short sale generates cash proceeds, but the seller need only put up 50% margin on the value of the transaction, which can be pledged in the form of interest-earning Treasury bills.

Stock index futures contracts typically have quarterly expirations (in March, June, September, and December) and settle for cash. The final settlement price is established using a "special opening quotation" on the third Friday of the contract month. The special opening quotation is generally based on the opening price of each index stock on the final trading day, no matter when during the day it opens. If the stock does not open at all on the last trading day, the closing price the night before is used. For the Nasdaq 100 index, the Nasdaq official opening price (NOOP) is used.[1]

Because stock index futures contracts settle for cash at expiration, the futures price necessarily converges with the cash index value at expiration. In the time leading up to contract expiration, the spread between the cash index value and the futures contract—the basis—is determined by two variables: namely, a short-term interest rate and expected dividends. The short-term interest rate determines carry costs, while dividends represent cash inflows. *Net carry* is the difference between dividends received and interest paid.

To see this, assume two mutually exclusive choices. One possibility is to buy a basket of stocks (for cash, not on margin) that exactly replicates the target index. The other is to buy a stock index futures contract on the index. Now consider the opportunity costs for each option. Buying the basket of stocks means that cash will have to be spent, which means that the cash can no longer earn interest. On the other hand, buying the stocks that constitute the index entitles the holder to the dividends paid to the stockholders of record. Conversely, buying a futures contract enables the holder to earn interest, but not dividends. Only stockholders are entitled to dividends.

A futures contract position is like a side bet; it entails no equity ownership in the companies that comprise the index. However, a futures trader can buy Treasury bills and pledge them as margin (technically a performance bond) against an index futures position. Owning the Treasury bills entitles him to the interest they pay. *The fair value of an equity index futures contract* can therefore be defined as the value of the cash index less forgone expected dividends plus interest earned on Treasury bills, or:

$$\text{Fair Value} = \text{Spot Index Price} + \text{Interest on Cash Balances} - \text{Dividends}$$

The idea is that fair value represents the price at which riskless arbitrage is not possible. There are no profits to be had by buying the index and selling a mirror image basket of stocks short or vice versa. However, the spread between a basket of stocks that exactly replicates the index and its companion futures contract can sometimes trade at either a premium or discount from fair value for short periods of time. When that happens, arbitrageurs (also known as *program traders*) will buy the cheap side of the market and simultaneously sell the expensive side. Then they wait for the spread to close back to fair value or converge to zero at contract expiration.

As noted previously, program traders who execute these types of transactions account for a large portion of the volume on the NYSE, sometimes over 60%.[2] To do these trades they need to be able to quickly and accurately calculate fair value, replicate the underlying basket of stocks, and manage transaction costs. This entails minimizing the rate risk of financing positions, correctly anticipating dividend payments, as well as minimizing commissions, fees, and execution slippage.

CALCULATING FAIR VALUE

Getting fair value right depends on three variables. The first, the spot price, is known. The second, the short-term financing rate is uncertain but can be hedged. The third, dividends, is estimable. In operational terms, a reasonable formula for calculating fair value once the variables are collected is:

$$FV = \left[Spot \times \left(\frac{1+r}{360} \right)^{Days} \right] - Dividends \pm TC$$

where:

FV = fair value
Spot = spot price or cash index value
r = short-term interest rate
Days = number of days to contract expiration
Dividends = dividends expected to be paid through contract expiration
TC = transaction costs

Consider an example. Suppose the Dow is trading at 11,000; there are 45 days to go until contract expiration, and the total of expected dividend payments through contract expiration equals $50. Using the fair value calculation formula, assuming for the moment that there are no transaction costs, the Dow Jones futures contract is worth 11,018.

$$11{,}018 = \left[11{,}000 \times \left(\frac{1+.05}{360} \right)^{45} \right] - 50$$

Continuing with the same example assume that the Dow Jones futures contract is trading at 11,050, while the underlying stocks are trading at a price that implies a cash index value of 11,000. Program traders would likely buy the basket of Dow stocks and sell the futures contracts short, figuring that they would make a profit of 11,050 – 11,018 = 32 Dow points. In dollar terms, 32 points is worth \$5 × 32 = \$160 per mini-sized Dow Jones Industrials futures contract at the Chicago Board of Trade.

Let's change some numbers around and examine this type of transaction more closely to see how to manage it and what the potential pitfalls are. Suppose the following: The cash index value of the Dow is 11,079; the cash index divisor is 0.12493117; there are 84 days left to go before the Dow Jones futures contract expires and expected dividend payments are \$50. Going by the formula, the fair value of the futures contract is 11,159.

Suppose that the futures contract is trading at 11,200, even though its fair value is 11,159. At that point, we would expect canny arbitrageurs to buy the stock of the 30 firms that make up the index in exact proportion to their index weightings and simultaneously sell the index futures contract short. For every unit of the transaction (one futures contract and one basket of 30 Dow Jones stocks) arbitrageurs would expect to make the 41-point difference between fair value (11,159) and the actual price (11,200) at which they sold the futures contract. In the mini-sized Dow, the 41 point difference is worth 41 × \$5 = \$205.

Note that while the expected profit comes from the spread between fair and actual value, the realized P&L will actually come from the combination of dividends received, interest paid, and convergence of cash and futures. To examine this further, we examine the cash flows that result from a 100-unit transaction in mini-sized Dow futures. For simplicity's sake, we assume 0% margining of positions (the stock is paid for with cash) and no transaction costs. A 100-unit transaction is approximately equal to the purchase of 4,000 shares in the stock of each of the companies in the Dow Industrials. The same number of shares for each company is required because the Dow is weighted by price rather than capitalization. An easy way to calculate the number of shares needed to offset a futures position is to use the following formula:

$$S = m \times \frac{1}{d} \times f$$

where:

s = number of shares of each company in the Dow Industrials
m = the index multiplier
d = the index divisor
f = the quantity of futures contracts

Using the formula, it is easy to see that it takes 4,000 shares of each Dow stock to offset a futures position of 100 mini-sized ($5 multiplier) Dow contracts.

$$4{,}000 \approx 5 \times \frac{1}{0.12493117} \times 100$$

We can use market prices to show that, in the example, 100 mini-sized Dow Jones futures contracts offset by 4,000 shares of each Dow stock is market neutral. Using the closing prices of the 30 stocks of the Dow as of June 21, 2006, it would cost a total of about $5,536,680 to buy 4,000 shares of each stock. A sale of 100 mini-sized Dow contracts at the cash index value of $11,079 would represent a notional value of about $5 × 11,079 × 100 = $5,539,500, which is almost exactly equal to the dollar value of the cash stocks purchased. The actual contract sale price of 11,200 has a slightly higher notional value of $5 × 11,200 = $5,600,000, because the higher sale price has embedded in it expected interest and dividends that will be amortized over the life of the contract.

Over the remaining 84 days until contract expiration, the trader who is short futures contracts and long the individual stocks will forgo interest earnings on the cash spent to buy stock. But he will receive dividend distributions from the companies whose stocks he owns. Depending on the timing of the cash flows and compounding, forgone interest is estimated to be about $64,973 using an interest rate of 5%. Expected dividends over the 84-day period are $50 per unit. Using the $5 multiplier for the mini-sized Dow, 100 units would be expected to throw off $5 × 100 × $50 = $25,000 in dividends.

By contract expiration, the futures price will converge to the cash index price. To illustrate the P&L, we hold the Dow stationary and allow futures to converge to cash. In this case the futures contract that is trading at 11,200 is expected to converge downward to the cash index value of 11,079. Using the $5 multiplier, convergence produces gross revenue of $5 × 100 units × (11,200 – 11,079) = $60,500. Using all the cash flows, it is easy to see that the transaction can be expected to yield a net profit of $25,000 $_{\text{(Dividends)}}$ – $64,973 $_{\text{(Interest)}}$ + $60,500 $_{\text{(Price Convergence)}}$ = $20,527. Leaving aside rounding, the profit is equal to the spread between theoretical fair value (11,159) and the actual contract price (11,200) times the $5 multiplier.

COMPLICATING FACTORS

There are several complicating factors that need to be brought into the example at this point. One is that the effect of financing on program trades is asymmetric. A second is the impact of a change in dividend payouts. A closely related consideration is the timing of the cash flows. Another is that program trading involves numerous simultaneous transactions that have to be monitored and managed. Still another is that some indexes contain so many different securities that exact replication is impractical. Consequently, some program traders have taken to using statistical samples to create smaller baskets that they deem to be close enough to the underlying index they are trying to replicate. The Dow does not pose much of a problem in this regard. But the Russell 2000 does, not only because of the volume of securities but also because the securities are so much smaller that they can pose liquidity problems.

The effect of financing on basket trading is asymmetric in that the impact on futures contract prices favors short sellers. Short selling typically generates positive cash flow before dividends, unlike trading from the long side which paradoxically imposes greater carry constraints. Short sellers are required to make delivery of the stock they sell. In order to complete the transaction, short sellers borrow stock to make delivery. When the sellers borrow stock, they are typically required to put up margin equal to 50% of the stock's market value. Upon delivery the seller receives 100% of the sale price in cash, at least a portion of which can be invested in short-term interest-bearing money market instruments. However, short sellers are required to pay dividends on stock sold short. Those dividends are passed through to the clearing system to the ultimate owner of the stock. Finally, fees charged by stock loan brokers for borrowing stock are transaction costs that need to be taken into account.

DIVIDEND PAYOUTS

Another issue to be considered is the impact of a dividend change by an index component. The price of an index futures contract is discounted by expected dividend distributions over the life of the contract. Because futures longs forgo receiving cash dividends from owning actual stock, they reduce the price they are willing to pay for stock futures by an amount equal to expected forgone dividends. But companies periodically change dividend policies. Dividends can be increased, reduced, or eliminated altogether. Sometimes companies also announce the payment of one-time special dividends.

When companies pay already announced (or correctly anticipated) dividends, there is not really any material impact on futures prices. Expected dividend distributions are already priced into the index futures contract. But when a surprise dividend is announced, the impact can be substantial. The basis spread between cash and futures has to adjust to reflect the change in expected cash flows.

From an absolute valuation standpoint it is important to note that index methodology can magnify the effects of a change in dividend payouts. That is because the Dow Jones Industrial Average trades ex-dividend along with the stocks in the index. When a stock goes ex-dividend, the stockholder of record on that day is entitled to the dividend when it is paid. The stock price adjusts to reflect that fact. For instance, assume a closing price of Boeing at $85.85 and that Boeing pays a quarterly dividend of $1 a share. On the ex-date Boeing would be quoted at $84.85, unchanged on the day. It is considered unchanged because the dollar difference in the price quotation will be paid out to the stockholder of record in the form of a cash dividend.

But the price impact of going ex-dividend spills over into the calculation of the Dow Jones Industrials because its value is based on the prices of the component stocks; the index calculation is not normally adjusted for dividend payouts. Consequently, the price impact on the Dow on the ex-dividend date is the quotient of the dividend and the index divisor. For instance, with a 0.12493117 Dow divisor, a stock paying out a $1 quarterly dividend would reduce the Dow by $1/0.12493117, or 8 points.

Certain circumstances, including the payment of unusually large cash dividends, can lead to a change in the Dow's divisor. Microsoft's decision to pay a one-time special $3 dividend to holders of record on August 25, 2004, was such a case. Similarly, special cash distributions from events outside a company's normal course of business, such as lawsuits, asset sales, and restructurings, can bring about a change in the divisor.

RATE CHANGES AND THE BASIS

With respect to stock index futures, short-term interest rates work in the opposite direction from dividends. All else equal, as stock dividends increase, stock index futures fall in price relative to the underlying cash index value. As discussed above, the reason is that the opportunity cost of owning index futures rises and falls with interest rates all else equal. The opportunity cost of owning stocks can be thought of as forgone earnings on other investments. The best measure of this opportunity cost is the risk-free rate. As the risk-free rate rises, the opportunity cost of owning stocks rises as well. Not so for holders of stock index futures contracts who retain the opportunity to invest cash at the risk-free rate. As the risk-free rate rises, they should be willing to pay higher and higher premiums to own stock futures instead of actual stocks.

In setting up cash/futures index arbitrage positions, it would seem that interest rate risk is far easier to manage than dividend exposure. Companies can change dividends for any number of reasons. On the other hand, short-term interest rates are largely controlled by the Fed, and the Fed responds in rather predictable ways to changes in inflation and economic growth. Even were that not the case, short-term rate exposure can be hedged relatively easily by using either CBOT fed funds futures or CME Eurodollar futures, as discussed in earlier chapters.

The basic technique for hedging short-term rate exposure is to buy (sell) futures contracts in fed funds or Eurodollars in an amount equal to the notional value of the cash index trade. The notional amount of the trade essentially represents the quantity of dollars exposed to changes in short-term rates. Traders who wish to lock in financing rates for stock bought on margin can sell fed funds futures in the amount of the notional dollar exposure over the life of the index futures contract. Alternatively, some traders may choose to use Eurodollar contracts the same way. Choice of contract to manage rate exposure depends on how well each of the instruments correlates with the financing used to establish the position.

TRANSACTION COSTS

Transaction cost is a critical element that needs to be considered when setting up arbitrage trades between stocks and equity index futures. Program trades have lots of moving parts, each of which has transaction costs associated with it. Program trades using the Dow are among the simplest. The index only has 30 stocks in it, it is price weighted, and all the Dow stocks are very liquid. Nevertheless, there are still a multitude of steps required for executing a successful—meaning profitable—transaction.

The steps include estimating dividend payouts over the life of the contract for the 30 index stocks and borrowing stock for the short side or arranging financing for the stock on the long side. On the long side the decision needs to be made on whether to lock up financing rates. On the short side, if there is considerable demand for borrowing a particular stock issue, premium fees will likely be charged. In addition, program trades can be commission and fee intensive. When the time for order execution arrives, it involves a transaction in each of the 30 Dow stocks as well as an offsetting futures contract trade. Each of those transactions involves fees, commissions, and a bid/offer spread. There is also risk of execution slippage, but that risk has been shrinking. Lightning-fast computer systems allow traders to program algorithms that touch off order entries when prespecified trade parameters are met. And electronic trade matching engines at exchanges now provide virtually instantaneous execution reports.

Managing transaction costs is a matter of cataloguing the incremental costs associated with each part of the transaction chain. It is a matter of adding marginal transaction costs when calculating upper and lower breakeven boundary points for fair value. These boundaries can then serve as trade entry and exit points. Moreover, setting up the boundaries this way may also serve as a market guide. Other traders will undoubtedly be doing the same calculations. Furthermore, such a process keeps a check on whether transaction costs are competitive. A list of transaction costs is shown in Table 19.1.

TABLE 19.1

Transaction Cost Checklist

Transaction Costs by Item	Per Transaction		Per 100 Shares	
	Low	High	Low	High
Stock Commissions and Fees Per Share	$0.01	$0.03	$1.00	$3.00
Average Stock Bid/Ask Spread	$0.01	$0.05	$1.00	$5.00
Futures Contract Commisisons	$2.00	$4.00	$2.00	$4.00
Futures Contract Bid/Ask Spread	$5.00	$10.00	$5.00	$10.00
Futures Exchange Fees	$0.15	$0.25	$0.15	$0.25
Total Transaction Costs/1 Contract Unit			**$9.15**	**$22.25**

Estimating transaction costs this way provides a feasibility guide for executing trades based on the fair value model, which is dependent on transaction costs as well as expected dividends and net carry. For example, consider the data in Table 19.1. The breakeven discount (or premium) for buying (or selling) the basis ranges between $9.15 and $22.25. In order to be reasonably sure of at least breaking even on the arbitrage, it would be necessary to either sell futures contracts at a 22.25-point premium or buy futures at a 9.15-point discount from fair value. These can be thought of as upper and lower boundary points for triggering program trades. They are also indicative of how competitively trading costs are being managed. If other traders or firms have lower trading costs, they will be first in to execute and will have a leg up.

STATISTICAL REPLICATION OF INDEXES USING BASKETS OF STOCKS

Up to this point the discussion has revolved around pricing futures contracts based on fair value calculations. The implicit assumption is that the underlying index can be exactly replicated, that dividend distributions can be estimated reasonably well, that short-term rate risk can be managed, and that trades can be executed cleanly with little price slippage. Those prerequisites seem reasonable enough when the underlying index is the Dow Jones Industrials. There are only 30 stocks in the index, weighting is by price, and all the stocks are big cap issues with plenty of liquidity.

But when the underlying index has 500, 2,000, 3,000 or more stocks in it, replicating an index is a horse of a different color. Leaving aside the merits of cap-weighted versus price-weighted indexes for the time being, cap-weighted and float-weighted indexes can be harder to track than price-weighted indexes. And the transaction costs incurred for exactly replicating larger indexes are far greater than for smaller ones.

Consider for instance that a change in float requires a weighting adjustment. Portfolio adjustments imply additional transaction costs from commissions, fees, and bid/ask spreads. Also consider how transaction costs are liable to multiply when the cash basket involves 500 (or more) separate stocks, all with commissions, fees, and bid/ask spreads. Consider too how transaction costs are liable to differ between the Dow, a big cap index with only 30 stocks, and the Russell 2000, a small cap index with 2,000 stocks. It is unlikely that the Russell 2000 stocks will have the liquidity of the Dow's components. Large orders are far more likely to cause prices to be jumpier in the smaller stocks of the Russell than the big stocks of the Dow. One consequence is that bid/ask spreads in the small stocks tend to be larger (in percentage terms).

High transaction costs associated with exact replication of indexes with many components has led to the adoption of index tracking strategies that accept some tracking error in return for reduced transaction costs. Index funds designed to match the returns of an index are the most likely users of this strategy; the goal is not to try to beat the index, but to match it. On the other hand, index arbitrageurs and hedge funds *are* looking to beat the index—that is how they get compensated. So the arbs and hedge funds are constantly on the lookout for ways to create portfolios that outperform benchmark indexes either by market-timing or stock selection strategies. These are multistep processes.

The first step is to determine the minimum number of issues needed to replicate the target index. The next is to determine the maximum number of stocks consistent with minimizing execution costs. The third step is to identify returns drivers unique to the benchmark. The fourth step is to identify and classify stocks by whether they are highly correlated, lightly correlated, noncorrelated, or negatively correlated with the stocks that drive the returns of the benchmark. Then a position or portfolio can be constructed and traded against a futures contract on the benchmark index.

This strategy is premised on the theory that the portfolio's returns will deviate significantly from the returns generated by the benchmark index. If the portfolio is expected to outperform the benchmark, the strategy is to go long the stocks in the portfolio and short the futures contract. Vice versa if the portfolio is expected to underperform the benchmark. The idea is to identify the index drivers and isolate individual stocks or stock groups whose returns are expected to deviate significantly from the rest.

While passive portfolio managers simply seek to replicate a benchmark index, active portfolio managers seek to beat the index by superior stock selection, market timing, or both. Trading benchmark futures contracts against an actively managed portfolio can be thought of as strategic basis trading. The portfolio managers seek to generate *alpha*—risk adjusted returns in excess of the benchmark index. Generating alpha, either through successful stock picking or market timing, requires identifying significant gaps between market valuations and true value. Stock picking concentrates on relative returns; market-timing strategies emphasize the importance of absolute returns. Stock-picking strategies place stocks in different categories (e.g., value, growth, small cap) and tend toward bottom-up analysis. Market-timing strategies put the context in terms of the business cycle and tend toward top-down analysis. These differences are easily overstated. Each contains elements of timing and selection. Timing strategies often revolve around the types of stocks to be long (or short), depending on the business cycle.

CATEGORIZING STOCK TYPES

The ways to categorize stocks are almost limitless. Among the most common categorizations are by industry groups, capitalization, growth versus value, foreign versus domestic, style (e.g., defensive versus aggressive, high dividend), and cyclical versus growth. These categories contain overlaps and subcategories. Leaving market direction aside for a moment, price leadership of the market changes as buyers and sellers rotate from group to group. Sector rotation is heavily influenced by circumstances, particularly the business cycle.

The importance of stock categories is of more than just academic interest. They are a way to differentiate stocks for diversification purposes. Moreover, mutual funds, index funds, hedge funds, and others attract investors with rules specifying the types of securities and market sectors in which they trade. To put it in perspective, it's worth considering the amount of money under management by the mutual fund industry. According to the Investment Company Institute, the trade organization of the mutual fund industry, about $4.9 trillion was invested in long-term equity funds in the United States, distributed across 188 million shareholder accounts.[3]

A classification model widely accepted by the industry, officially known as the Global Industry Classification Standard (GICS) was developed jointly by Standard & Poor's and Morgan Stanley Capital International (MSCI) in 1999. The GICS is a four-layered system. It is divided into 10 sectors, 24 industry groups, 67 industries, and 147 subindustries as of April 2006. It includes about 33,500 publicly traded companies worldwide and covers about 95% of all global equity market capitalization.[4]

The 10 major sectors in the GICS classification scheme are: energy, materials, industrials, consumer discretionary, consumer staples, health care, financials, information technology, telecommunications, and utilities.

TABLE 19.2

Market Sectors

	No. Firms	% Firms	Mkt. Value ($)	% Mkt.
S&P 500	500	100%	11,620,664	100%
Energy	29	6%	1,143,596	10%
Materials	31	6%	358,792	3%
Industrials	53	11%	1,360,389	12%
Consumer Discretionary	87	17%	1,200,242	10%
Consumer Staples	40	8%	1,113,687	10%
Health Care	56	11%	1,426,008	12%
Financials	86	17%	2,504,201	22%
Information Technology	78	16%	1,766,328	15%
Telecommunications Services	9	2%	368,422	3%
Utilities	31	6%	378,997	3%

Data Source: Standard & Poor's

Standard & Poor's and MSCI jointly assign GICS codes at the company level. Each firm has only one code; there is no dual coding. All firms included in the various S&P equity indexes receive a GICS classification code.

In fact, S&P publishes a daily a breakdown of the number of companies from each sector included in the S&P 500 along with the market caps of the companies aggregated by sector. The result is that it is relatively easy to benchmark the performance of various sectors against the market as a whole. Similarly, the performance of specialty portfolio mangers can be tracked against recognized sector and subsector indexes. Table 19.2 for instance, displays the 10 major sectors of the GICS classification scheme, the weight of each sector in the S&P 500, the number of companies in each sector, and the market cap of each sector as of May 26, 2006.

The GICS codes are meant to reflect the principal business activity of a company as determined by S&P and MSCI. Among the criteria used are revenues, earnings, and market perception. These criteria are a bit more ambiguous than they appear at first glance. GE, for instance, is classified as an industrial company, which in some sense it undoubtedly is. But GE Capital, one of its most profitable subdivisions, is a major player in financial markets, including such businesses as structured finance and leasing.

But industry classifications are important nonetheless because a key element of many equity trading strategies has to do with sector rotation. Some sectors (e.g., consumer staples) are widely believed to perform relatively better when the economy is soft. Others (e.g., information technology) are thought to perform better during periods of robust economic growth. Accordingly, active money managers over- or underweight sectors based on economic forecasts. One way to do so is by using indexes, subindexes, and index futures to gain or shed exposure to the various sectors.

In this respect industry classification systems do provide a good deal of interesting and important information, especially for arbitrageurs who trade equity indexes in the futures markets. For instance, it is worth noting that three sectors (health care, information technology, and finance) account for 49% of the market cap of the S&P 500—with 22% coming from the financial sector alone. Contrast that with the consumer discretionary category. It comprises 17% of the firms (the same as financials) but only 10% of the market cap.

Sector rotation strategies are focused on the characteristics of targeted companies or industries. A different vantage point is to think of trading strategies not in terms of particular companies or industries but in terms of style. Among the more important investment styles, two in particular stand out: investing in growth and investing in value.

STYLE INDEXES: GROWTH OR VALUE?

Stocks are often categorized as "growth" or "value" shares. Although the definitions of growth and value are murky at best, the terms are generally meant to distinguish between young, fast-growing firms and older, mature companies. In an attempt to systematize the difference, the major index providers have developed rules to categorize firms as either growth, value, or hybrid "blend" companies that have characteristics of each.

To distinguish between growth and value companies Dow Jones has developed a six-factor model. The factors are: (1) the projected price/earnings ratio, (2) projected earnings growth, (3) the price-to-book ratio, (4) the dividend yield, (5) trailing revenue growth, and (6) trailing earnings growth. In order to implement the model, Dow first separates the 5,000 stocks of the Dow Jones Wilshire Composite Index into four categories: large cap, mid-cap, small cap, and microcap. Then each market cap category is divided evenly into growth and value components based on the six factors cited above.[5]

Similarly, Standard & Poor's launched a series of style and growth indexes to address two distinct needs it identified in the marketplace. The first is for broadly based conventional indexes to serve as references for index funds and financial engineers seeking exposure to particular style segments. The second is for narrower "pure" indexes with very high factor loads (for growth or value). They serve as vehicles for concentrating investments by style.[6]

Like Dow Jones, Standard & Poor's uses a multifactor model to evaluate stocks of companies for growth and value characteristics. Three factors are used for growth; four are used for value. The growth factors are the five-year growth rate in earnings per share; the five-year growth rate for sales; and the company's five-year internal growth rate, which Standard & Poor's defines as return on equity (ROE) times the earnings retention rate. The four value factors are the book value to price ratio, cash flow to price ratio, sales to price ratio, and the dividend yield.

Once the stocks have been scored along the dimensions of growth and value, they are sorted into style baskets and ranked. Broad style indexes are created by dividing each market capitalization evenly into growth and value indexes. A more narrow set of pure style indexes for each market cap is created by restricting the constituents to the top-ranked 33% (value or growth) stocks to populate the respective pure style index baskets.

Categorizing stocks as either growth or value securities has enormous significance for stock trading. Index funds and derivatives get created around these indexes and money flows into them. Moreover the debate over value versus growth stock investing has fascinated market professionals for a very long time. In general, the rule of thumb has been to view value stocks as defensive plays and growth stocks as aggressive ones. Market timers think in terms of owning growth stocks on the way up and value stocks when the bear shows up.

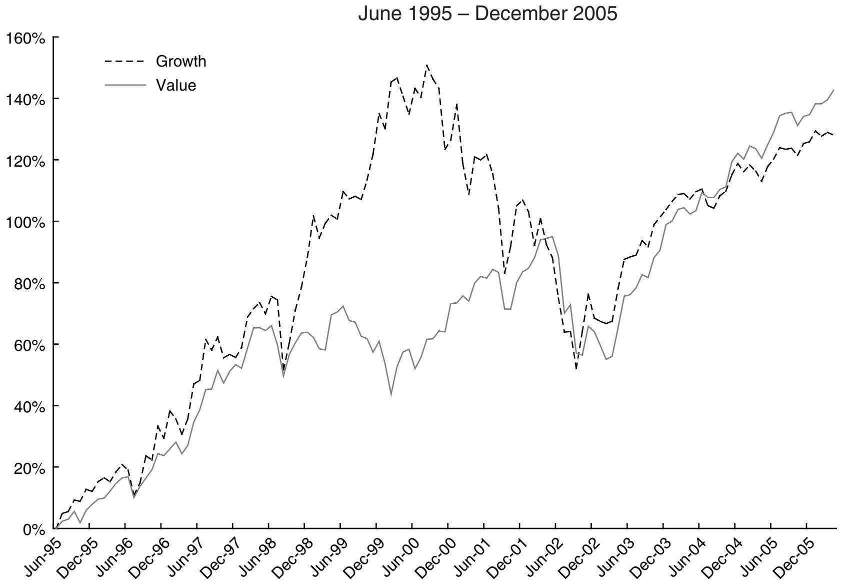
The difference between value and growth stocks can be clearly seen in Figure 19.1, which is a graph of the cumulative total returns of the S&P 500 pure growth and pure value indexes from July 1995 through April 2006. This encompasses periods of Fed ease and tightening as well as recession and strong growth, so the graph provides a reasonably good visual of stock market behavior under different circumstances. Over time, the value index provided a higher total return, with considerably less volatility. On the other hand, the growth index is manic-depressive, at times greatly outpacing the value index before crashing back down to earth.

There are a number of approaches to trading growth and value stocks against the benchmark indexes. One is based on the Fama-French research that suggests that value stocks have systematically outperformed after adjusting for risk.[7] It seeks to avoid the trap of being overly rigid about portfolio weightings, so as not to get caught in the flurry when stocks are added to (or subtracted from) a benchmark, as in the case of Google discussed earlier. This is a core-satellite strategy that marries positions in small cap and mega cap stocks. The small cap stocks increase expected returns; the mega cap stocks (like Exxon and General Electric) reduce portfolio volatility. A second variant is to mix bonds with small cap stocks.

Another approach to trading value versus growth stocks is market timing. While value stocks seem to outperform in the long run, during

FIGURE 19.1

Growth versus Value



Data source: Standard & Poor's

the short run and in bull markets growth stocks often soar ahead of conventional indexes and value stocks. To what extent, if at all, is this predictable? Is market timing a viable strategy for catching market rotations, allocating among asset classes, or alternating between the long and short sides of the market?

MARKET TIMING

Market-timing strategies revolve around the idea that traders can forecast when markets are more likely to rise or fall and position themselves accordingly. Academic researchers have tended to believe that this is a fool's errand, although some recent studies have shown evidence of market-timing abilities among certain investors, including hedge funds.[8] Regardless, the idea of buying stocks when they're going to go up and selling them when they're going to go down has nearly irresistible intuitive appeal. And it's not likely to disappear any time soon.

Market timers like to advertise their dexterity at getting into and out of the market at the right times. Risks associated with stocks are usually considered in terms of the losses that might be incurred in the event of a stock market sell-off. But that only tells part of the story. Another part is the risk of being out of the market at the wrong time thereby missing a big move to the upside. A study published by Merrill Lynch makes the point by noting that in the 10-year period ended December 31, 2003, an investor who missed the market's best 10 days would have forgone $6,469 in earnings out of an original $10,000 investment made in the S&P 500. An investor who remained fully in the market would have earned a compound return of 11.06% compared to the 2% return earned by anyone who missed the top 20 best days of the same 10-year period. Similar results have been published by Ibbotson and AXA.[9]

In the end, market timing only makes sense insofar as it demonstrably improves returns or lowers portfolio risk. That would seem to imply a set of decision rules to determine exposure to shifts in market direction. A good deal of academic research has been devoted to devising and testing decision rules such as these based on past market experience. For the most part academicians have come up empty handed, unable to formulate decision rules capable of beating passive index investments after adjustments for risk. But that has not stopped the quest. It would therefore seem to be useful to review some highlights of this research.

To do this, a more precise description of market timing is in order. To begin with, market timing is not an all-or-none proposition. Market-timing strategies fall along a continuum that ranges from "pure timing" in which the decision is 100% cash or 100% stocks, to a more nuanced approach in which the mix of portfolio assets is adjusted based on market levels. Then there is the decision-making process to consider. A buy-and-hold approach represents an implicit (but passive) market-timing strategy.

Consider for example a portfolio strategy that sets out an initial asset mix comprising 50% stocks and 50% government bonds. Then suppose a year later stocks have risen 15% and bonds have declined 5%. The change in prices over time will have changed the portfolio asset mix to 55% stocks and 45% bonds. The result is a change in portfolio strategy. In contrast, a more active approach to market timing would be one in which positions or asset mixes are adjusted by portfolio managers in anticipation of changes in market levels. This type of market timing is often referred to as *dynamic asset allocation*.

These three basic types of market timing suggest a frame of reference for considering strategy. For the sake of convenience, label the three types as (1) active, (2) passive, and (3) dynamic tactical asset allocation. *Active market timing* refers to strategies in which the decision is a wholesale adoption of one asset class for another (including cash). *Passive* refers to acceptance of a changing portfolio mix as a result of changes in market prices. *Dynamic tactical asset allocation* refers to marginal changes in portfolio asset mix as the result of active decision making.

In a paper entitled "Market-Timing Strategies That Worked," an economist from the Kansas City Federal Reserve Bank, Pu Shen, back-tested some market-timing strategies against a passive buy-and-hold approach using a data set of stock prices from 1970 to 2000.[10] To test the efficacy of market timing, Shen compared returns from two portfolios. The baseline portfolio was based on a buy-and-hold strategy; the other changed composition based on anticipated changes in market values. To conduct his test, Shen used spreads between the earnings/price ratio (E/P) of the S&P 500 and interest rates as predictors of future price action. Specifically, Shen set up a series of "horse races" in which the performance of the S&P 500 (the baseline) was compared to a portfolio usually invested in the S&P 500, but sometimes invested entirely in cash.

Underlying Shen's study is the idea that the E/P ratio provides important information about market valuations, especially at extremes. Shen used the E/P ratio, also known as the *earnings yield,* as a market sentiment indicator. Because P/E ratios (the inverse of the E/P) tend to be high when interest rates are low and vice versa, Shen used the earnings yield in conjunction with interest rates to construct two spread indicators. The first is the spread between the S&P 500 earnings yield and the yield on 3-month Treasury bills; the second is the spread between the S&P 500 earnings yield and the 10-year Treasury note yield.

The T-bill yield essentially represents the risk-free interest rate. It is largely determined by Federal Reserve policy, which is sensitive to both inflation and the business cycle. It is widely believed that FOMC policy is an important predictor of stock prices, and there is considerable evidence to suggest that stock prices are more apt to rise during periods of expansive monetary policy than they are during tight money periods. Longer rates are less sensitive to FOMC policy, but they have embedded in them longer-term expectations of monetary policy.

To test the signaling power of these spreads, and hence their usefulness as predictors of future stock prices, Shen examined the distribution of returns of the S&P 500 with respect to the E/P interest rate spread. He found that when spreads between earnings and interest rates were narrow (implying that stocks were priced expensively), stock market returns were generally lower than average in subsequent months. In addition, he found that volatility tended to be higher when E/P ratios were low.

With those findings in hand, Shen devised some simple trading rules for identifying market entry and exit points. The basic strategy is defensive. It is to remain fully invested in the market except when the E/P interest rate spread is extremely low, signaling that valuations are very expensive. Since the strategy revolves around exiting the market only at extremely high valuations, trading is minimized as are transaction costs.

When the trading rules were implemented (through back-testing with historical data), Shen found that the portfolio switching strategy produced slightly higher returns than did the buy-and-and-hold strategy. Moreover, returns variances were lower, implying lower risk levels, another measure of superior performance. Finally, Shen found that the switching strategy was robust, not being sensitive to the choice of starting time in the sample. Based on these results, Shen concluded that market-timing strategies based on these rules would have been successful from 1970 through 2000.

The Shen paper generated considerable interest. While it has long been known that changes in interest rates have an impact on stock prices, the market-timing question has to do with the predictive power of rate changes. In theory rate changes should be of little value because stock prices are thought to adjust rapidly to the new information. But there is some evidence that rates do contain some predictive ability. A 1995 study by Jenson and Johnson found that stock prices rose in periods after a Federal Reserve discount rate cut and fell in periods after a rate hike.[11] In a similar vein Prather and Bertin found that a simple trading rule of buying stocks during periods of discount rate cuts and switching to Treasury bills after rate rises produced superior risk adjusted returns.[12] Other studies have suggested that an inversion of the yield curve leads to lower stock prices in the future.

That rate changes may very well have predicative power over future stock prices should not be surprising. As discussed previously, changes in FOMC policy have tended to have an enormous impact on both the level of short-term rates and the shape of the yield curve. Further, it appears to be reasonably clear that it takes a while for the impact of a change in Fed policy to be fully priced into the market for Treasury bonds. If so, it is easy to see why policy changes would take even longer to work their way through the equities markets.

The E/P ratio has been used extensively in market timing studies, based on the idea that if stocks can be bought cheaply (measured by high earnings-to-price ratios), they ought to provide higher returns over time. And in fact, Sanjoy Basu showed that stocks with high earnings yields performed better over time than randomly constructed portfolios.[13] Similarly, Fama and French found a positive correlation between high earnings yields and stock market returns.[14] Basu explored this further and found that superior returns remained even when adjusting for small-firm effects.[15]

Jerome Hennet, a skeptic when it comes to market timing, decided to test how well various market-timing indicators performed when applied across five different markets: the United States, the United Kingdom, Japan, Germany, and Switzerland. Hennet found that short-term rate indicators, E/P short-rate spreads, and fed funds indicators did not perform well as predictors of future stock prices. Further, he found that other indicators that seemed to work well, including the long-rate E/P spread, were sensitive to the observation period, suggesting a lack of robustness in the original results. Hennet concludes by suggesting that market timing has the awful reputation it does among academics because it is well-deserved. Similarly, Keith V. Smith found that investors would have been better off using a buy-and-hold strategy instead of implementing quarterly asset allocation adjustments based on their brokers' advice during the first half of the 1990s.[16] Before dismissing market timing out of hand, a recent paper by Ken Fisher and Meir Statman is worthy of examination.[17] They make the point that stock prices are driven by two factors: value and sentiment. If stock prices were only influenced by value, then stock returns would be determined exclusively by information. But in the real world there are noise traders as well as information traders. The sentiment of noise traders drives (or can drive) prices away from value.

If that is true, then predicting future stock prices involves forecasting not only the fundamental factors likely to influence prices, but also the likely sentiments of noise traders. If noise diffused at a constant or at least a predictable rate, it would not constitute a large problem. But of course, it doesn't. And then there is the question of how to separate the sentiment of noise traders from the calculus of information traders. How much of the high P/E ratio is the result of expectations of fast earnings growth, and how much of it is sentiment?

To investigate the predictive power of pure market sentiment, as well as P/E ratios, Fisher and Statman chose to use the Investors Intelligence Sentiment Index, which is based on the writings of over 100 independent financial market newsletters. Investors Intelligence examines the newsletters and then puts them in one of three categories: bullish, bearish, or correction. The bulls recommend buying, bears recommend selling, and the correction camp is waiting to buy at lower prices. The Bullish Sentiment Index is the number of bullish newsletters divided by the sum of the bullish and bearish letters.

Fisher and Statman found the Bullish Sentiment Index to be a mildly successful contrary indicator. Investors who switched from stocks to T-bills when the Bullish Sentiment Index was above its median, and bought back in only when the index fell below the median, exceeded the returns gained by adopting a buy-and-hold strategy. It also produced better predictive results than did switching strategies based on price/earnings multiples and dividend yields. Investors who substituted the market P/E ratio for the Bullish Sentiment Index earned substantially less than the buy-and-hold strategy. In another variation on the theme, using an interest rate adjusted P/E ratio produced a similarly disappointing result.

The authors concluded that while the sentiment indicator did better as a market-timing guide than did P/E ratios, neither appeared to be very reliable. Certainly one of the arguments in favor of the efficient market hypothesis is the very absence of a viable market-timing trading rule that produces superior risk-adjusted returns. On the other hand, market volatility makes it painfully obvious that stock prices can stray far from intrinsic values. Moreover, there may be an asymmetry at work here.

Before the 1980s when the takeover and leverage buyout boom began, entrenched corporate managements could easily run their firms with an astonishing disregard for the owners, the stockholders. Takeover financing provided a way for outsiders to hold managements accountable. If a company's miserable management caused the stock to trade too cheaply, a raider with sufficient financing could come along, buy the company, and fire its management. That eventually kept prices from falling too low.

On the other hand, there is no ceiling on prices when a speculative flurry bursts out, as the Internet craze showed. Short sellers can sell all they like, but they can't affect control or deployment of corporate assets. They can only hope the stock goes down, a fact that speaks to the power of sentiment in the marketplace, especially when it is detached from, or at best only loosely tethered to, fundamental values. Still, when sentiment shifts from bullish to bearish and everybody starts heading for the exits, they are liable to find that the doors are kind of small. One result is that, on balance, sell-offs tend to be more violent than rallies.

Couple this with the fact that the stock market has a long-term upward drift, and the problems faced by market timers begin to come into sharper focus. The market's upward drift should create a bias toward the long side. But in the short run, daily volatility dominates the trend. Furthermore, the down days are likely to be of greater magnitude (on average) than the up days, a fact which works in the opposite direction of the long-term trend. In addition, long-term returns are heavily influenced by a relatively limited number of big days, a fact alluded to by the Merrill Lynch study cited earlier. Missing a small number of the big up days can severely depress long-term returns. So a danger with market timing is not just the possibility of getting caught on a big down day; it also includes missing the big up days.

All this suggests the importance of being able to read crowd psychology when establishing or managing trading positions. Sentiment can change on a dime, just when you least expect it. Changes in sentiment, especially when there is an event trigger, can bring about a wholesale reevaluation of market fundamentals leading to pronounced market moves. There may be key signs or psychological trip wires to watch out for, as discussed in the next section, but they are not likely to lend themselves to hard-and-fast quantitative rules. Trading is not only science; there is art to it as well.

SUMMARY

Virtually all the major equity indexes have futures contracts listed on them. The market leader is the CME E-mini S&P 500. The Russell 2000 and the Nasdaq 100 are popular indexes that trade at the CME as well. The Dow Jones Industrials trade futures at the Chicago Board of Trade.

Equity futures contracts share a common design and a common pricing model. They are essentially the cash index value times a multiplier and settle for cash. The CME and CBOT offer margin offsets for Dow and S&P positions on opposite sides of the market. Equity index futures contracts are priced on a cash-and-carry model in which expected dividend payments and financing charges drive the basis spread between the futures contract and the underlying cash index.

Strategies for trading equity index futures include both cross-sectional stock picking and market-timing subsets. Cross-sectional or relative value strategies seek to profit by owning stocks or index subsets that outperform the major benchmark indexes. In this respect there is evidence that value stocks and smaller stocks produce superior risk-adjusted returns compared to the large benchmark indexes, which suggests arbitrage opportunity. Basis trading, another type of relative value trading, can be successful, depending on financing and dividend prediction acumen.

Market-timing strategies come in several varieties. Some are passive, allowing changes in market prices to alter asset allocations. Others are more direct, seeking to profit by calling market tops, bottoms, turns, and trends. Somewhere in the middle are dynamic asset allocation strategies in which marginal changes in the portfolio asset mix depend on active decision making.

There has been a good deal of academic research on market-timing strategies. Some recent studies have detected momentum effects; others have shown a predictable inverse relationship between market performance and published market sentiment reports. However, on balance the research has failed to produce convincing evidence that any particular set of rules is likely to yield superior results. Moreover, a critical but overlooked danger of market timing is the possibility of missing big up days, thus trimming returns.

NOTES

1 For a more detailed discussion of the final settlement price, please go to the CME Web site at www.cme.com

2 See the NYSE Web site for program trading statistics: http://www.nyse.com/

3 *Investment Company Fact Book 2006*.

4 Detailed information on the GICS classification system is available at the S&P Web site: http://www2.standardandpoors.com/

[5] See *Dow Jones Style Indexes Methodology Review*, at: http://www.djindexes.com/mdsidx/index.cfm?event=showStyleMethod.

[6] *See S&P U.S. Style Indices Methodology*, Standard & Poor's April 2006.

[7]Eugene Fama and Kenneth French, "The Cross-Section of Expected Stock Returns," *The Journal of Finance,* 1992.

[8] See, for instance, Yong Chen, "Timing Ability in the Focus Market of Hedge Funds," 2005: http://www.fma.org/Chicago/Papers/hftiming-FMA.pdf; and Cheng-few Lee and Shafiqur Rahman, "Market Timing, Selectivity and Mutual Fund Performance: An Empirical Investigation," *The Journal of Business,* 63/2, April 1990, pp. 261–278.

[9]See, for instance, *Money Digest* at: http://www.findarticles.com/p/articles/mi_m0JQR/is_2_13/ai_30430016; or Johnathan Burton, "Taking the Long View: Time in the Stock Market, Not Timing the Market, Brings Greater Rewards," *Market Watch,* December 3, 2006.

[10] Pu Shen, "Market-Timing Strategies That Worked" FRB of Kansas City Research Working Paper No. 02-01, May 2002.

[11] Gerald R. Jensen and Robert R. Johnson, "Discount Rate Changes and Security Returns in the U.S., 1962–1991," *Journal of Banking and Finance*, 19/1, April 1995, pp. 79–95.

[12]Laurie Prather and William J. Bertin, "A Simple and Effective Trading Rule for Individual Investors," *Financial Services Review,* 6/4, 1997.

[13] Sanjoy Basu, "Investment Performance of Common Stocks in Relation to Their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis," *Journal of Finance,* 32/3, June 1997, pp. 663–82.

[14] Eugene F. Fama and Kenneth R. French, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47/2, June 1992, pp. 427–65.

[15] Sanjoy Basu, "The Relationship between Earnings' Yield, Market Value and Return for NYSE Common Stocks: Further Evidence," *Journal of Financial Economics,* 12/1, June 1983, pp. 129–156.

[16]Keith V. Smith, "Asset Allocation and Investment Horizon," *Financial Services Review,* 6(3), 1998, pp. 201–209; and Jerome Hennet, "Could Market-Timing Strategies Really Be Lucrative?" Draft Master's Thesis, HEC International Business School, December 2004, available at: http://www.hec.unil.ch/cms_mbf/master_thesis/0411.pdf

[17] Kenneth Fisher and Meier Statman, "Investor Sentiment and Stock Returns" *Financial Analysts Journal,* March–April 2000, pp. 16–23.

SECTION III

Market Psychology, Trading, and Risk Management across Asset Classes

This page intentionally left blank

CHAPTER 20

Pairs Trading

> The generation of random numbers is too important to be left to chance.
>
> –*Robert R. Coveyou, Oak Ridge National Laboratory*

In some respects all trading is based on the intuition that market prices of similar types of assets behave in a similar fashion. For instance, real estate agents typically use recent sales of comparable homes to benchmark house prices. And it would be a rare trader who, upon observing a rapidly declining stock market, decided to go long by lifting sellers' offerings instead of brandishing a bid below the last sale. The type of correlation trading that will be discussed in this chapter is a bit more formal than that.

Pairs trading is based on the theory that securities prices tend to be correlated to the degree that they share common factors. Some share more common factors than others, so their prices are liable to be more highly correlated. Pairs-trading strategies revolve around the idea that commonalities can be identified and that differences can be isolated and statistically cordoned off. That allows the price spread between similar types of securities to be modeled as a single unit, so that pricing anomalies can be exploited.

The framework for pairs trading can be found in the CAPM. In the CAPM stocks possess two types of risk. The first, measured by beta, is generalized market risk. The second is idiosyncratic risk, peculiar to that particular stock. An important corollary is that stock prices oscillate around their true fundamental values. From that corollary two conclusions follow. The first is that stocks with similar characteristics will tend to behave in a similar, but not identical, fashion. The second is that there is a true equilibrium value for the price spread between the companies' stocks.

Companies' stock prices may oscillate around true value, but there is a random white noise element as well. As a result, even though two companies share similar characteristics, their stock prices can diverge in the short run even without any fundamental changes having taken place. This innocent-sounding statement is actually a lot dicier than it sounds. Divergence can be a random event. But it can also represent a significant shift in the fundamentals. In this respect stock arbitrage trading is fundamentally different from most types of bond market arbitrage, particularly, government bond arbitrage.

There are several ways to think about this. In theory, the same arbitrage pricing mechanisms that keep bond prices in line serve to keep stock prices in line. Arbitrageurs will identify and buy the cheap securities and sell the expensive ones, betting that the spread will move back into line. With government bonds, identifying relative value is a comparatively simple affair. The only variables that really matter are coupon and maturity, because a bond's price simply represents the present value of expected future cash flows. And with government bonds, it isn't really *expected* future cash flows—because the cash flows are already known. So the discount rate (or yield to maturity) of a government bond maturing in nine years and six months is going to be pretty much the same as that of a government bond maturing in nine years and four months.

With stocks it's a much more complicated story. In theory the price of a stock represents the present value of expected future cash flows. But the determinants of those cash flows are many and varied. And those determinants may very well affect different companies in very different ways—even if they are in the same industry. Consider for instance an increase in the price of oil. It's fine for oil companies, but not so fine for airlines. Moreover, a rise in the price of oil doesn't even affect all oil companies or all airlines the same way. Southwest Airlines for instance, managed to do very well despite the run-up in oil prices in 2005 and 2006 because, unlike most of its competion, it used the futures markets to hedge its expected fuel demand. On the other hand, before oil companies commit billions of dollars to new energy exploration, they need to consider the long-term outlook for prices. A misstep may put them at a competitive disadvantage.

Another way to consider the vagaries of equity arbitrage trading compared to bond arbitrage is that bonds are (mostly) priced from the top down. Stocks, however, contain a good deal of idiosyncratic risk. For instance, an unexpected rise (or fall) in inflation will affect all bonds pretty much the same way. But not stocks. Gold stocks will be apt to rise, while financials will likely come under pressure. Moreover, the fundamental nature of a company's business can change in a heartbeat with a strategic acquisition or a change in its business model. The marriage of Time-Warner (formerly Time Inc.) and AOL is a case in point.

Keeping these caveats in mind, there are ways that arbitrage traders seek to take advantage of short-term statistical discrepancies in the price relationship between pairs of financial instruments whether they be individual stocks, bonds, or indexes. The basic idea is to identify common factors that cause price movements and treat the remaining variation as white noise. To the extent that stocks actually do oscillate around their true values, the spread between a pair of highly correlated stocks or stock indexes can be similarly expected to oscillate around its true value. Arbitrageurs seek to identify points at which the spread deviates significantly from true value and generate profits by trading the spread on the theory that it will return to true value.

Consistent with the efficient market hypothesis, the true value of the spread—its equilibrium—is its long-term mean. Over time, variations from the mean are expected to sum to zero. If that is the case, significant variation from the mean represents arbitrage opportunity because the spread is expected to exhibit mean-reverting behavior. Consequently arbitrage traders expect to profit by either selling the spread higher than its mean value or buying the spread below its mean value, provided there is a sufficient margin of safety. Statistical methods are used to estimate appropriate safety margins.

SPREAD VALUES AND SAFETY MARGINS

Pairs trading relies on various types of statistical analyses: first, to determine whether there is a tradable spread and, second, to estimate optimal exit and entry points. In part, the analysis depends on the types of instruments being traded. For instance, it may involve stock pairs (Coke and Pepsi), stock indexes (Dow Jones and the S&P 500), baskets of stocks in the form of ETFs, or sector and style indexes representing various returns universes. Other examples of suitable pairs would be baskets of stocks against equity index futures contracts, and baskets of bonds against bond ETFs or bond futures contracts. The focus of this chapter is on pairs trading using stock indexes, ETFs, and futures contracts rather than individual stocks.

The first step in pairs trading is to determine a value anchor, which is done by using a historical mean value for the spread. Historical dispersion around the mean is used for establishing confidence levels. For expository purposes this chapter uses the spread between the Dow Jones Industrials and the S&P 500 to investigate the potential for pairs trading. The Dow and the S&P are particularly well suited for studying pairs trading. Not only do the two indexes possess important similarities, but they also possess significant differences that may help explain returns variations. For instance, the Dow's components are exclusively mega cap firms whose stocks are highly liquid. The average Dow stock has a market cap of about $119 billion. The average market cap of an S&P 500 stock is $23 billion. But the S&P is dominated by large cap firms by virtue of the fact that it is float-weighted. That also points to a major difference: The Dow is price-weighted rather than float-weighted like the S&P 500.

A straightforward way to model and evaluate the spread between the Dow and the S&P 500 is to run a correlation and regression analysis, easy enough to do with a conventional statistical software package.[1] In this example we take the logs of the daily closing prices of both the Dow Jones Industrials and the S&P 500. The sample period extends from January 2, 2002, through July 14, 2006. The calculated correlation coefficient between the Dow and the S&P for the period of four and a half years is 0.98, which means that 98% of the log price of the S&P is explained by the log price of the Dow. Going further, we can estimate a regression equation using the Dow as the independent variable and the S&P as the dependent variable. Because we use natural logs (ln), index comparisons are in percentage terms. The regression equation is:

$$\ln(\text{S\&P}) = \alpha + \gamma * \ln(\text{Dow Jones}) + \varepsilon$$

Estimating the regression equation using a statistical software package (in this case STATA), produces the result below. The *t* statistics are in parentheses.

$$\ln(\text{S\&P}) = \underset{(-71.77)}{-4.573449} + \underset{(181.54)}{1.257573} * (\ln \text{Dow Jones}) + \varepsilon$$

$$R^2 = 0.966$$

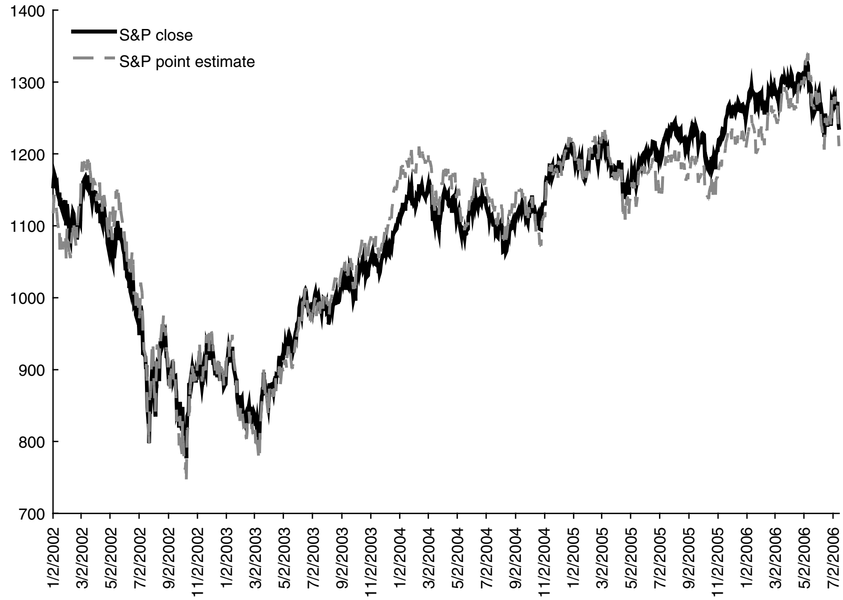
The results of estimating the model can be interpreted as meaning that, on average, the natural log of the price of the S&P 500 is expected to be roughly equal to 1.257573 times the natural log of the price of the Dow Jones Industrials minus 4.573449. The model's residual—the difference between the actual price of the S&P and the model's prediction—is represented by the error term ε. The γ term (in this case 1.257573) is known as the *cointegration estimate*.

Variation between actual and predicted values of the S&P based on the Dow's price is slight. This can be seen in Figure 20.1, a graph of both the actual and predicted prices of the S&P 500. The spread between them is the error term, and it is small. One conclusion is that the price of the Dow is a good predictor of the price of the S&P. That conclusion is unsurprising: The model R^2 is 0.966. The model explains almost 97% of the price of the S&P 500 index using only the Dow as the predictor variable.

The model's error term ε is white noise and the focal point of the correlation trading strategy. It represents deviation from the underlying true value of the spread between the two indexes. Over time, the spread

FIGURE 20.1

S&P Actual versus Estimate



Data source: Yahoo! Finance

between the actual value of the S&P and its predicted value is expected to average out to zero. Therefore, the arbitrage strategy is to sell the S&P short and buy the Dow when the S&P sells at a significant premium to its predicted value. Conversely, when the S&P sells at a significant discount to its expected value, the trade is to buy the S&P and sell the Dow Jones Industrials short. That is because the Dow has been shown to be a good predictor of the S&P, and when the S&P trades significantly higher or lower than the model predicts, it is either too expensive or too cheap relative to the Dow. (The model has nothing to say about the absolute value of either, only the relative value).

TRADE IMPLEMENTATION

As discussed in previous chapters, there are numerous ways that the S&P and the Dow can be bought and sold in toto. Each has a liquid futures market, each has an ETF that trades against it, and baskets of stocks that constitute the indexes can be assembled. For the sake of convenience we assume that each of the methods is easily accessible, efficiently priced, and without transaction costs.

An important question concerns identification of trade entry and exit points. They can be thought of as boundaries that separate opportunity zones from neutral territory as well as points that delineate margins of safety. How much of a premium or discount does it take to make it worthwhile to put on a trade? How far can the trade veer away from neutral territory? How long is the workout time likely to be? One way to answer these questions is to examine the history of the spread. Then the historical depth and frequency of variation from true value can be analyzed and risk/reward parameters estimated.

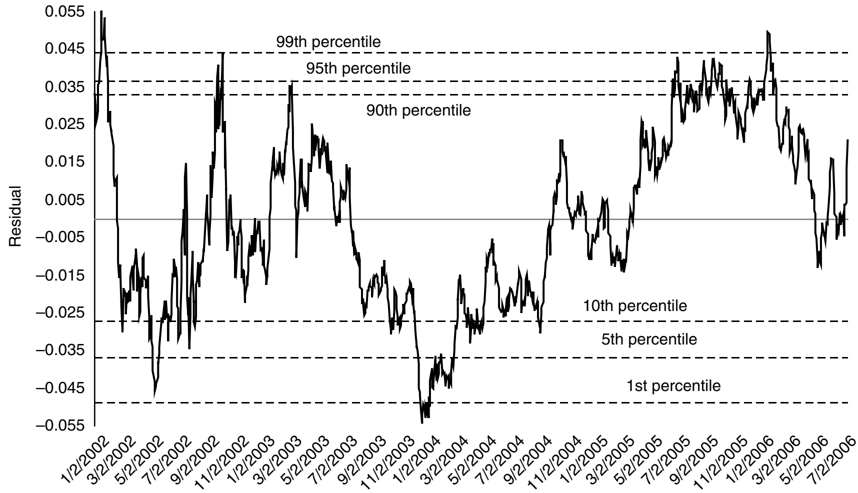
Both frequency and depth are important. Frequency is important because the more often the spread varies from true value, the more numerous are the opportunities for putting on and taking off the trade. Depth of variation is extremely important. The ideal outcome is to have larger quantities of a trade toward the extremes of the distribution. But getting there can be uncomfortable; it requires enormous psychological stamina to maintain staying power, and there is a fine line between stubbornness and conviction. Stubbornness has a long history on Wall Street of separating traders from their cash.

To analyze trade feasibility, we develop a history of the Dow Jones/S&P spread trade, using the parameter estimates from the model. The γ cointegration estimate of 1.257573 controls for general market volatility. It is the functional equivalent of beta in the CAPM. Since the model is in log form, it is expressed in percentage terms. Therefore, the spread trade can be easily converted to dollar units. For instance, using the γ cointegration estimate from the model, it would take $1,257,573 worth of the Dow Jones Index to equal $1,000,000 worth of the S&P 500 Index. (To simplify matters, we do not consider either interest on the difference in cash balances or dividends.) To the extent that the Dow and the S&P measure the same underlying returns universe, day-to-day differences between predicted and actual index values (i.e., the error term) are assumed to be white noise, after controlling for market volatility.

A history of the spread (the model residual) can be constructed and plotted with the data used to estimate the model. The results of doing this are shown in Figure 20.2. The horizontal axis marks off the dates of the time series. The vertical axis presents the residual, which is the difference between the predicted value of the S&P and the actual value. The solid black line traces the daily history of the residual from January 2, 2002, through July 14, 2006. The dashed black lines represent key percentiles of the distribution. As the graph shows, the spread oscillates around zero, which can be thought of as neutral territory. The further the spread veers away from zero, the closer it is to an opportunity zone. Statistically speaking, the areas below the 10th and above the 90th percentiles can be thought of as those opportunity or arbitrage zones.

FIGURE 20.2

Time series: Dow/S&P model residual



As Figure 20-2 illustrates, the S&P/Dow Jones spread can move from over- to underpriced with breathtaking rapidity. In part that is because the two indexes are so highly correlated. In absolute terms in doesn't take all that much price movement for the spread to veer into arbitrage territory. Moreover, rapid movements into and out of arbitrage territory allow the spread to be put on and taken off relatively quickly. All things equal, trades with fast turnover are preferable to ones that have longer workout times.

The graph illustrates another aspect of this type of trade that has to be managed: momentum. The underlying theory behind the trade is that variation between actual and predicted values is simply white noise, a random event. But noise traders, trend followers, or momentum players may very well jump onboard and keep pushing prices away from true value. Paradoxically, they may do so in the belief that the market "knows" something, as a result of which they decide to "follow" a nonexistent trend or a faux fundamental change in true value. For the arbitrage trader this poses a dilemma about how to manage entry and exit points.

One way to see this is to examine the prices of the Dow and the S&P at the points where they cross boundary lines—for instance the 90th through 99th and 1st through 10th percentiles. Percentage risk-adjusted returns for the trade can then be calculated from the point of position acquisition and at various boundary points until market exit. For example, Table 20.1 displays net percentage returns on a γ-weighted hypothetical

TABLE 20.1

Dow/S&P Spread Trade Progress (January 14 to February 20, 2002)

Date	Percentile	Dow Jones (Long)	Δ Dow	β Adjusted Dow Δ	S&P (Short)	S&P (Change)	Net
1/14/2002	95	9891			1138		
1/23/2002	99	9730	–1.63%	–2.05%	1128	–0.88%	–1.17%
2/5/2002	90	9685	–2.08%	–2.62%	1090	–4.22%	1.60%
2/20/2002	0	9941	0.51%	0.64%	1098	–3.51%	4.15%

Note: β = 1.2576

Dow/S&P spread trade first acquired when the spread hit the 95th percentile of the frequency distribution. Subsequent marks-to-market are at the 99th, 90th, and 0 percentiles.

Table 20.1 also shows the progress of the trade in percentage terms over the passage of time. Trade initiation (long the Dow/short the S&P) takes place on January 14, 2002, when the spread first hits the 95th percentile of the distribution. About a week later the Dow has dropped 1.63% while the S&P has dropped 0.88%. But the Dow position is weighted by the cointegration parameter of 1.2576, so on a weighted basis the Dow position is off 1.2576 * 1.63% = 2.05%. That puts the trade underwater by 1.17% at the 99th percentile of the distribution.

A week or so later the S&P has caught up and overtaken the Dow on the downside. The Dow is down an adjusted 2.62%, while the S&P is down a solid 4.22%. That puts the trade back into positive territory to the tune of 1.6%. And a few weeks after that, the spread has reverted to its mean of 0, so the trade is ahead by 4.15%. It is important to note that the pattern of being underwater and then bouncing back is not all that surprising for this type of trade. By definition, the occasional visit to the 99th percentile first requires piercing the 95th percentile. The real issue concerns position management—primarily selecting trade entry, exit, and scaling points.

POSITION MANAGEMENT

There are implicit trade-offs embedded in the selection of entry and exit points. Choosing aggressive entry points, at either at the 10th or 90th percentiles, ensures participation in a relatively large number of potential arbitrage opportunities. But aggressive entry comes at a price. Statistically speaking, there is increased potential of having to sit through losses for those times when the trade goes through the first barrier (the 10th or 90th percentile) and heads toward the 95th and 99th or 5th and 1st percentiles. One compromise solution is scaling in and acquiring larger positions if and when the trade pierces the first boundary and hits the second—for instance putting some of the arbitrage on at the 90th percentile and more on at the 95th.

It is a fair bet that most correlation traders would opt for scaling into a trade. Some would be put on at the 95th percentile, possibly more at the 97th, and more again at the 99th percentile. Generally, the allotment decision is made ahead of time, with bigger units being put on the deeper the spread goes. And as a general rule it's almost always wise to keep some reserve ammo, just in case.

However, scaling into trades can be difficult. For one, there is the inevitable heartburn that accompanies buying more of something that has already gone against you, meaning you are already solidly in the loss column. Second, scaling in is a tactic, not a strategy. The two should not be confused. The best strategy is picking the right spread and modeling it correctly. Tactics involve position acquisition techniques. In this respect it might be worth ruminating a bit about the St. Petersburg paradox.

The St. Petersburg paradox refers to the classical decision theory problem presented by Bernoulli in 1738.[2] Basically it revolves around this question: How much would you be willing to pay to bet on the flip of a coin if you could double down after each losing flip, up to the amount you paid to play? In theory you should be willing to pay a lot to play the game. Statistically speaking it is a virtual certainty that you would eventually win more than enough to cover the cost of entry. (That is one reason why casinos enforce betting limits at the tables.) But no sensible person would ever make the bet. It wouldn't take very many consecutive losing flips to put you in the poorhouse. By the 25th coin flip an initial $1 bet doubled down would be for $16,777,216.00, a bet most would be more than a bit hesitant to make.

Strictly speaking, scaling in is not the model of the St. Petersburg paradox, because (among other reasons) scaling is not an all-or-nothing proposition. But applying probability theory is useful, and most would say crucial, for sensible position management. Still, when a trade goes against you before the ink is dry on the ticket (which in the electronic era means pretty fast), it is reason to give pause for thought, sophisticated statistical models notwithstanding.

Nick Leeson of Bearings Bank apparently decided not to take this lesson to heart and doubled down on a series of wrong way bets on the Japanese stock market. For his efforts, with the notable help of a little back-office fraud, he managed to lose just under a billion dollars in a remarkably short period of time. In the process he put the centuries old Barrings Bank out of business. Now out of prison, Leeson is busy on the lecture circuit. His Web site helpfully notes that he spends much of his time presenting to companies on the subject of risk management, based on his "life experiences."[3]

Charms of the lecture circuit aside, there are better ways than the Leeson method to approach position management. The first lesson of trading comes to mind: Always keep an eye on the exit door. Have a stop-loss point in mind. Scaling in is fine, but there is a reason why it's called the 99th percentile and not the 100th. This type of trade (like all trades) is an exercise in applied probability, not certainty. Things can go wrong and sometimes do.

HIGH-FREQUENCY PAIRS TRADING

Thus far we have examined pairs trading using daily cash closing prices of the Dow Jones Industrials and the S&P 500. Although common practice, the use of closing prices may bias the results. For one, closing prices reflect the last trade, which may or may not have occurred at the close. So the prices of each member of the pair may not have been generated at the same moment. Another factor to consider is that the bulk of the trading takes place during the day, not at the close. Exclusive use of closing prices may result in measurement error and model estimates that are less accurate than they could be.

We can take advantage of the fact that modern technology makes high-frequency intraday trade data available. Not only does availability of intraday prices present more data points for analysis, but the data reflect nearly simultaneous transactions in the pair components. The result is data that do a better job of representing real-world behavior. Moreover, the same techniques used to model time series of pairs over days, weeks, and months can be employed with minute-by-minute data.

High-frequency price data can therefore be used to create rapid-fire trading models designed to take advantage of the many small pricing anomalies that can easily crop up during the trading day. To explore this, we take a sample of intraday prices of futures contracts on the S&P 500 and the Russell 2000, traded on the Chicago Mercantile Exchange. The CME makes these data available for download on its Web site.

To get a good sample pairing of actual transactions, prices on the sample day (July 21, 2006) are partitioned into segments of 15 seconds each, starting at 8:30 a.m. lasting until 3:15 p.m. If both the S&P and the Russell traded within those 15 seconds, they were treated as simultaneous transactions. A cleaned sample taken from those transactions is used to estimate a regression model as before. The results are shown below.

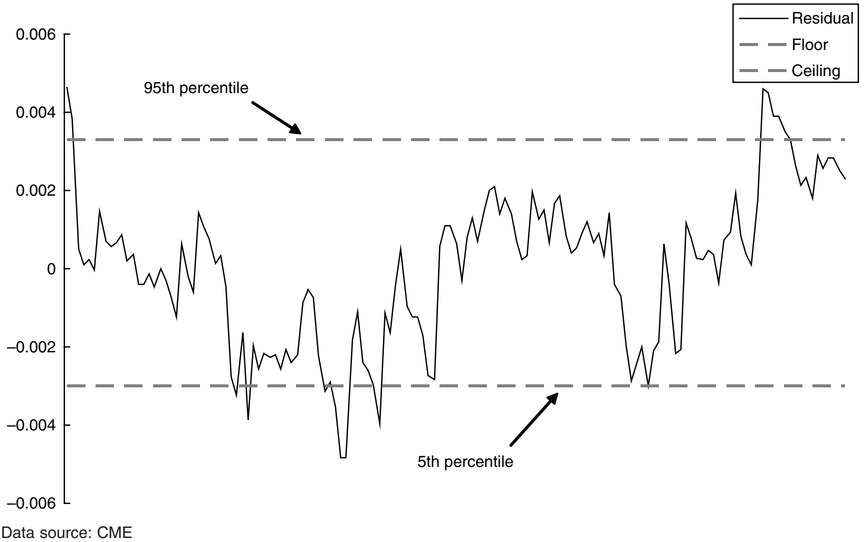
$$\ln(\text{Russell } 2000) = \underset{(-4.47)}{-2.4742} + \underset{(16.59)}{1.26083} * \ln(\text{S\&P } 500) + \varepsilon$$

With the model estimated, the error term can be calculated and graphed as in Figure 20.3. As in the Dow/S&P model, the error term oscillates around zero, with the 5th and 95th percentiles marked off as trade entry points. Even though the time period is only a few hours, the pattern displayed is very similar to the one shown previously, even though that one tracked the S&P/Dow Jones closing cash spread over a period of several years. But there are some important differences. Not surprisingly, variation in the intraday Russell/S&P spread is much smaller than variation in the closing daily price model of the Dow and the S&P 500. And due to the frequency of the observations, in absolute terms the spread can be expected to touch the arbitrage boundary points more often per unit of time. For instance, the Russell/S&P spread touches either the 5th or 95th percentile 14 times in the course of the sample trading day, a situation made to order for day traders.

There is a caveat here, and it has to do with the stability of the model. There is no guarantee that the arbitrage boundary points for any one day (calculated ex post) will be accurate going forward, particularly for intraday models. The best way to deal with this issue is to constantly update and adjust the model as new data become available. It is also worth noting that changes in index composition may cause the boundary points to move.

FIGURE 20.3

High-Frequency Intraday Pairs (Russell–S&P)



Data source: CME

IMPLIED BASIS: INDEX PAIRS USING FUTURES AND ETFS

Another way to trade index pairs is to use futures contracts against an exchange-traded fund (ETF) based on the same underlying benchmark index. As discussed previously, futures contracts on equity indexes settle for cash against the target index. Similarly, the value of an ETF with respect to its underlying cash index is kept in line by arbitrage that is driven by the ETF creation/redemption process. As a result, correlation between an ETF and an index futures contract that targets the same underlying index will tend to be very high.

There are a few hitches though. ETFs pass through dividends to the shareholder. On the other hand, futures contracts discount expected dividend payments over the life of the contract. Buying an ETF requires cash payment. Equity index futures can be bought using interest-earning Treasury bills pledged as the performance bond. Therefore the spread, or basis, between an ETF and the companion futures contract will reflect the impact of dividends and interest payments. As the futures contract approaches expiration, it will converge with its underlying index, which by implication means that it will also tend to converge with the companion ETF.

As a result, the spread between an ETF and a futures contract that shares the same benchmark index oscillates around an implicit convergence path that the futures contract follows against the underlying index. Running a simple ETF/futures regression model will produce a biased result because the time series is trending, not stationary, insofar as it follows the convergence path. Nevertheless the ETF/futures spread can be modeled and traded by treating the convergence path as a trend variable and controlling for the trend.

The best way to see this is to compare the closing prices of the Dow Jones Industrials ETF and Dow Jones futures contracts. The American Stock Exchange trades the Dow ETF known as Diamonds (ticker symbol DIA). The Chicago Board of Trade lists several different sized futures contracts on the Dow, with the electronically traded mini-sized Dow being the most popular. For analytical purposes there are actually two comparisons to be made. The first is of the daily closing prices of the Dow futures contract and the companion ETF. The second is a comparison of the daily closing prices of the Diamonds and an estimated fair value of the futures contract. Since the fair value calculation takes into account interest, dividends, and time value, it serves as a control variable for the convergence path. Subtracting the fair value premium (or discount) from the futures contract price ought to make it roughly equal to the price of the ETF.

After these calculations are carried out, the futures contract/ETF relationship can be explored in two ways. The first tracks the spread between the futures contract and the companion ETF. That spread is the convergence path and can be expected to approach zero at contract expiration. The second spread is defined as the price of the futures contract less the fair value premium (or discount) less the closing price of the ETF.

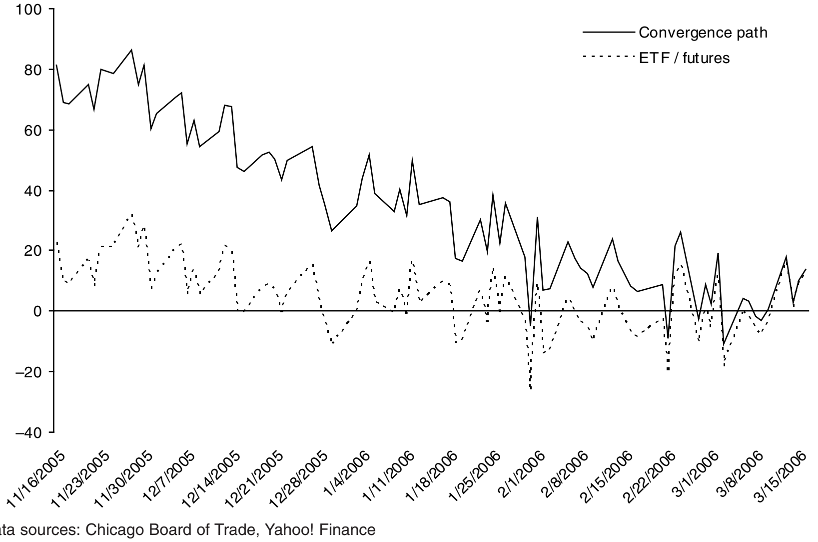
Spread = Futures ± Fair Value Premium (Discount) – ETF

That spread should oscillate around zero since the only reason for any difference in the value of the futures contract and the ETF is the premium (or discount) on futures for net interest and dividends. By this reasoning the first spread should follow a trending path; the second a flat path. To test the proposition, daily closing prices of the CBOT mini-sized Dow futures and American Stock Exchange Diamonds are compared in two ways. The first comparison is the price difference for equal par amounts. The second comparison is between the Diamonds and the calculated fair value of the mini-sized futures contracts. The sample period is November 15, 2005, through contract expiration March 16, 2006. The results are displayed graphically in Figure 20.4.

As expected, the spread between the Diamonds and the Dow futures contracts travels a relentless convergence path toward zero through to contract expiration. Also as expected when the futures contract is adjusted for fair value, the spread between it and the Diamonds oscillates around zero in what appears to be random fashion. This suggests that variation in the spread

FIGURE 20.4

Dow Futures versus Diamonds



Data sources: Chicago Board of Trade, Yahoo! Finance is white noise and amenable to pairs trading modeling techniques. But it also must be noted that trading a futures contract against an ETF is an implied basis trade, carrying with it all the stock loan, dividend, and interest rate risk associated with more conventional equity index basis trades.

SUMMARY

Pairs trading is a form of statistical arbitrage. The underlying assumption is that markets are priced reasonably efficiently and contain elements of white noise. To the extent that prices of individual stocks (or indexes) are priced efficiently and the companies' fortunes are strongly influenced by common factors, it is reasonable to expect their stock price returns to be positively correlated as well. Statistical techniques can be employed to test if and how well stocks and stocks indexes actually are correlated and by implication how well suited they are for statistical arbitrage trading.

To the extent that pairs of financial instruments are sufficiently well correlated, statistical models can be built that predict the price of one based on the other. The difference between model prediction and observed value (i.e., the error term) can be interpreted as white noise. White noise is the focus of pairs trading because paradoxically it captures spread pricing anomalies at the extremes of its distribution by virtue of the fact that the pair components are assumed to be priced reasonably efficiently.

When white noise reaches an extreme, pairs traders will look to sell one stock or stock index and buy the other, waiting for the spread to revert to the mean. There are some caveats that attend this strategy. One is that sometimes a statistically unusual spread change is not white noise at all, but a fundamental repricing. Another is that statistical models have to be reestimated frequently so that they are up-to-date with new developments. Finally, when establishing thresholds for trade entry and exit, it is important to consider the frequency of trading opportunities, workout time, and the merits of scaling into a position.

NOTES

[1] This relatively simple model is used for illustrative purposes. More sophisticated models might include other variables to generate a more precise estimate, or might estimate the γ coefficient using sample time frames of differing lengths. For a very good discussion of theory and estimation techniques, see Ganapathy Vidyamurthy, *Pairs Trading*, John Wiley & Sons, 2004.

[2] *The Stanford Encyclopedia of Philosophy* has a brief discussion of the St. Petersburg Paradox which can be accessed at this Web address: http://plato.stanford.edu/entries/paradox-stpetersburg/.

[3] http://www.nickleeson.com/index.html

CHAPTER 21

Technical Analysis

Every ship at the bottom of the sea has a chart room.

–Anonymous

A time-honored ritual on the Street is the morning traders' meeting. Coffee in hand, the firm's traders and economists amble into a conference room to discuss market goings-on. Often they are linked to overseas offices to widen participation and hear more input. One by one the traders give a brief outline of market developments, economic forecasts, market chatter, various rumors floating around the Street, and, most importantly, how they see the market trading over the near future. After they have all finished, the firm's head trader, who has been silent up to this point, looks around the room. His gaze eventually settles on a guy sitting at the other end of the long conference table. Laptop booted up, the guy is staring intently at the screen which is resplendent with multicolored graphs of price and volume indicators. The head trader focuses his gaze on the guy, and in a raspy voice he asks a single question: "What do the charts say?"

Technical analysts study market price, volume, and sentiment indicators in order to determine and implement trading strategies. For the most part technical analysts argue that these and other indicators contain all the information they need to make successful predictions about future market behavior. And so they claim to rely on these indicators more or less exclusively, concerned only about what the market is doing, not why. Among the most famous is John Magee.[1]

They don't care about why because the "why part" has to do with market fundamentals. Thinking about fundamentals will only muddy the waters when it comes to interpreting technical market data. And since everything is in the price anyway, why worry about why? This line of reasoning eventually leads to an obvious problem. If everything is in the price and certain chart patterns predict future prices, isn't that in the price as well?

In the end there is only one question about technical analysis that matters: Does it work? There has been quite a bit of academic research on this question. Thus far no one has been able to offer convincing evidence that trading outcomes based on technical analysis are any better than those based on the flip of a coin. In part, this may be because there is little agreement among technical analysts about exactly how to read and interpret the patterns they study. That turns out to be remarkably convenient. It means that there are no falsifiable hypotheses.

When a chartist is caught on the wrong side of the market, it is never because the charts were wrong. Almost inevitably the claim is made that the charts were misinterpreted. But there is not much agreement about what it is, exactly, that the charts are supposed to be able to predict that goes much beyond an oracle-of-Delphi generality. The typical market prognostications offered by technical analysts have a way of making a political stump speech sound like a model of clarity.

Here for instance is some sample tech talk that, while utterly fabricated, is typical of the genre and would fit nicely into more than a few of the hundreds of technical analysis newsletters that flood the industry: "The market traded up yesterday with ABC stock closing at 41¾ after hitting an intraday high of 42, the old short-term resistance point. ABC has been trending sideways in a narrow range between 38 and 42 for the last three months, building a base for a possible breakout to the upside. The stock is poised to challenge its old long-term high of 50, but it first has to convincingly break through the $42 short-term barrier and then pierce through heavy resistance at $45. On the other hand, failure to achieve a decisive breakout to the upside leaves the stock vulnerable. If it can't hold support at $38, it could slip to $34. Major support at $25 won't be tested unless intermediate support at $30 cracks."

Let's see now. A stock trading at 41¾ has to go above $45 before it can get to $50; or it has to trade below $30 to get to $25. To an innocent bystander who is reasonably proficient in fifth grade arithmetic, this might seem to be fairly obvious.

There is nevertheless a huge audience for technical analysis. In the first place it's easy to pick up. In the second place there is no requirement to be right. The psychological lure of looking at a chart to discover recurring patterns that will make you rich is almost overwhelming. A trip to the bookstore provides the evidence. The shelves of the markets and investments section brim with title after title proclaiming the wonders of one system or another as the road to riches. And, not to put too fine a point on it, virtually every trading floor on the planet has a multitude of computer screens with charts flashing closing prices, moving averages, advance-decline lines, trading volumes, and other standard measures used by financial alchemists. Charting is so pervasive that even those who dismiss technical analysis out of hand still latch onto its indicators. Who, for instance, ignored Dow 10,000 on the way up the first time or on the way down after the tech crash or on the way back up in the rally that began in 2002?

Some technicians (and lots of traders and portfolio managers) argue that it's important to watch the charts because they have a tendency toward self-fulfilling prophecy. As long as everyone thinks Dow 10,000 is important, it is. There may be some truth to this, although it nimbly dances around the fact that there are two sides to every trade. That said, a paper published by a pair of finance professors, Kenneth Kavajecz and Elizabeth Odders-White, found evidence that technical analysis revealed information about liquidity on the order book that supports and resistance correlated with the depth of the order book and that moving averages reveal information about the depth of the limit order book.[2] Because of the pervasiveness of technical analysis, this chapter provides a brief discussion of some key measures and techniques that are closely followed by the industry.

PATTERNS AND INDICATORS

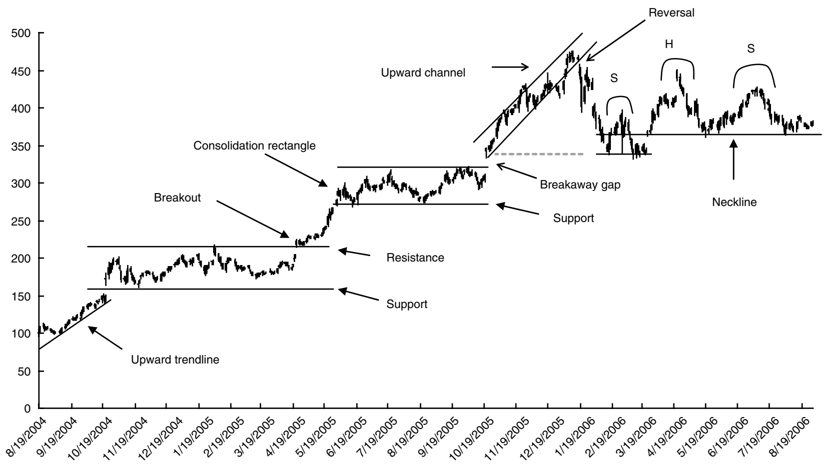
Technical analysis revolves around pattern recognition. The idea is that certain patterns of price and/or volume reflect an internal dynamic that predicts what the market is likely to do in the future. The patterns include representations of support and resistance, trends, strength and duration of trend, trend divergence, reversals, as well as market tops and bottoms. But all this comes with considerable hedging on the part of the cognoscenti over what constitutes a distinctively important pattern, how to recognize it, and how to interpret it.

Among the more important measures are accumulation patterns that supposedly detect money flows into a stock, which presumably presages future upward price pressure. Conversely, distribution patterns are said to represent money leaving a stock, predicting lower prices down the line. Trend lines and channels chart a directional path. As long as the price activity stays on the trend line (or in the channel), the technician assumes that the trend remains intact and will continue.

Support areas are price floors where buyers can be expected to appear, giving the market a platform from which to stage a rally. But beware the dreaded "dead cat bounce" in which the market rallies marginally from support levels only to plunge later on. On the other side, resistance areas are presumed market ceilings where sellers are likely to emerge. Don't be too cavalier about shorting at resistance though. If the market breaks through resistance to the upside, it may be a buy signal and the sellers will turn to buyers on the breakout.

FIGURE 21.1

Google Common Stock: Daily High and Low Opening and Closing Prices (Aug. 2004–Aug. 25, 2006)



Data source: Yahoo! Finance

So what do these patterns look like and how are they to be recognized? For some examples, observe Figure 21.1, a chart of the daily high and low opening and closing prices of Google common stock. It contains many of the more popular patterns and indicators used by technical analysts. The graph is based on data downloaded from Yahoo! Finance and covers the history of Google from the time it went public in August 2004 up to and including August 25, 2006.

In the jargon of technicians, the graph displays a primary uptrend, punctuated by periods of consolidation, during which time the stock lays the groundwork for the next upward thrust. Within the primary trend there can be short and intermediate countertrends that do not necessarily disturb the primary long-term trend. In addition, there are numerous patterns that are used to indicate whether the trend remains intact, where support and resistance lie, and where entry and exit points can be established.

SUPPORT, RESISTANCE, AND MARKET GAPS

Probably the two most important weapons in the technician's arsenal are support and resistance. They represent the respective floor and ceiling prices where buyers and sellers are likely to come into the market counter to the trend. They are determined by historical patterns of behavior. If buyers come in whenever the stock hits $35 and sellers come in when it hits $40, those prices would become recognized as support and resistance levels. The more times it happens, the stronger the support or resistance.

Technical analysts maintain that when a stock pierces through what was previously a support (or resistance) level, the price turns into a mirror image of its former self. Support becomes resistance; resistance becomes support. For instance, if a stock that has been trading in a range of $35 to $40 finally breaks convincingly above the $40 barrier, at that point $40 would become the new support level. The firmness of the support depends primarily on how well it stands up under fire. The more times it holds while under attack, the stronger it is thought to be. There is also a pecking order for support and resistance. Most technicians track intensity levels of support and resistance, generally considering them to be either short-term, intermediate-term, or long-term targets, depending on circumstances.

Support and resistance levels can be stationary, as when the market lacks a clear trend. Or they can be dynamic, as when the market is trending in a channel. Long trendless periods when the market meanders in a narrow trading range are often thought to produce consolidation patterns during which the market lays the groundwork for further gains (or losses) as the case may be. In this regard, one pattern that particularly interests chartists is the head-and-shoulders formation, so called because it resembles the silhouette of a person's head and shoulders. The neckline represents support; a break below it suggests that the stock could trade down by an amount equal to the width of the distance between the extremities of the shoulders. A reverse head and shoulders is supposed to work the opposite way.

Some support and resistance levels are thought to be dynamic rather than static in that they move with the market. The idea that there are dynamic support and resistance levels is based on the more general belief that trends in motion must either stay in motion or begin to deteriorate. An upward-trending stock needs to make higher highs and lower lows to keep the trend intact. A slowdown, made manifest by failure to make higher highs, might be an early warning that a trend reversal may be in store.

Chart pattern breakouts can occur in a number of ways, but two stand out. One is a slow grind; the other is the result of a dramatic price spike due to some surprising news or corporate development. Slow grinds can be deceptive. At first blush tiny incremental steps might not seem like a big deal. But it is in retrospect, after the incremental steps have been aggregated, that their collective impact becomes clear.

Price gaps are a different beast altogether. For one, they are attention grabbers. While some are relatively small, others can be quite large, and technicians consider them to be important. Price gaps come in many varieties. There are, for instance, breakaway gaps, so called because the price leaps out from the range in which it previously traded. There are runaway gaps, so labeled because the existing trend accelerates as the price jumps ahead of resistance (or craters below support). Then there are exhaustion gaps. Exhaustion gaps represent an emotional blow-off signifying the end, and possible reversal, of an existing trend.

Large gaps can be very disconcerting. Among other things, they make hash of a lot of models. Modelers tend to assume that markets are free-flowing and continuous. Large gaps put the lie to that theory. Newer jump diffusion models are being developed to cope with discontinuities, but they tend to deal with smaller, rather than larger, price gaps.

Like modelers, traders usually prefer markets to be continuous and free-flowing. But traders know that markets are frequently subject to all types of discontinuities. Some occasions, like important economic release and earnings dates, are more prone than others to produce large price discontinuities. Gaps like these represent danger zones where mistakes can be especially costly. Liquidity is likely to be shallow (that's why the gap happened to begin with), so it may be difficult to execute transactions. What's more, if the market begins to reverse course, it's easy to get whipsawed. Consequently, many traders hold their fire and wait for things to settle down before diving back in.

MOVING AVERAGES AND WAVES

While basic support and resistance remain key underlying measures, other tools in the technician's kit include various types of moving averages and oscillators designed to gauge "overbought" or "oversold" conditions, momentum, and market sentiment.[3]

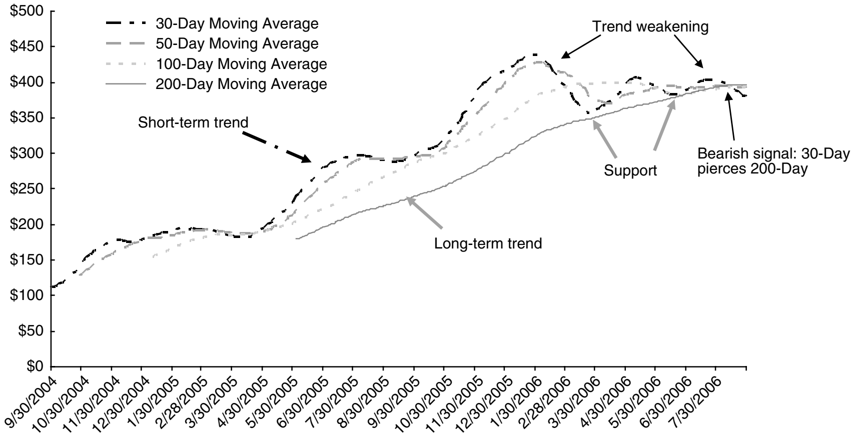
Moving averages come in various forms: simple, smoothed, weighted, and with differing time lengths, to name a few. At root, moving averages are designed to separate day-to-day noise from the larger trend, if any. Practitioners argue that traders can use moving averages to concentrate on positioning with the trend, where the big money lies, and to avoid getting sliced up in the day-to-day noise. These types of trend-following strategies are frequently used in commodities markets in the managed futures industry. Published results though, have tended, on average, to be less than stellar.

Moving average (MA) measures (and combinations of them) are typically used to establish sets of rules to identify market entry and exit points, to monitor trading positions, and to gauge risk exposure. Analyzing differences of moving averages of varying time lengths lies at the core of how they are used to construct trading strategies. The shorter the moving average time length, the more sensitive it is to day-to-day wrinkles in the market. Moving averages with a longer time frame are more sluggish, but they convey a larger sense of the long-term trend.

Provided that the shorter- and longer-term trends are in sync, trend followers will establish and hang onto positions. But when they diverge, it's

FIGURE 21.2

Google Moving Averages



Data source: Yahoo! Finance

time to bail out. In this respect one of the more critical indicators for technicians is the spread between the 30- and 200-day moving averages. As long as the 30-day is higher than the 200-day, the trend is supposed to be bullish, and vice versa. But if the 30-day decisively crosses the 200-day (in either direction), many technicians would be inclined to read that development as the sign of a trend reversal taking place and time to head for the exit door.

To get a sense of how moving averages are used in practice, observe Figure 21.2, a chart of the 30-, 50-, 100-, and 200-day moving averages of the price of Google. It is an example of how these data can be used to set and monitor trading rules. For instance, a typical bullish posture would be to go long and stay long Google, provided the 30-day MA exceeded the 200-day MA. It should also be noted that with the widespread use of electronic trading platforms, it is a comparatively simple matter to build trading rules into software programs so that transactions can be executed instantly when decision rules are triggered.

Other market measures similar to moving averages rely on drawing envelopes around current market prices. The envelopes are based on volatility, past price trends, or simple percentages. For instance, Bollinger Bands (named after John Bollinger, in "preguru days" a business news reporter for MSBNC) bracket a moving average (usually 20 days) by 2 standard deviations. Since 95% of all occurrences are expected to fall within 2 standard deviations in the normal course of events, strategies tied to Bollinger Bands implicitly rely on the idea that market valuations are mean-reverting. When the security trades outside the 2-standard-deviation boundary, according to the theory, it should be either sold or bought, depending on whether the market puts it 2 standard deviations higher or lower than the mean.

Another branch of technical analysis, closely associated with R. N. Elliot and later popularized by Robert Prechter, uses waves. The idea is that markets move in recognizable cycles or waves that can be identified, analyzed, and used to predict how trends are likely to develop. The essence of Elliot Wave theory is that markets unfold in five-wave packets. In bull markets, the first, third, and fifth waves are up waves; the second and fourth are countertrend waves or corrections within the larger trend. Bear markets are the opposite, with waves one, three, and five leading the way down, while the second and fourth act as countertrend waves in the larger cycle. After the fifth wave, the trend reverses.

If it were that simple and direct, the Elliot Wave theory could be subjected to empirical testing. Needless to say, it isn't. It turns out that the five-wave cycle has subcycles and sub-subcycles, and sub-sub-subcycles that potentially go on ad infinitum. Rules of interpretation do not exist a priori, only ex post, so the theory is impossible to falsify. As a result, wave counting and analysis remain the exclusive province of the Elliot Wave high priests, who on occasion seem to bear an eerie similarity to the interpreters of Nostradamus.

MONEY FLOWS

The age-old practice of watching the tape, also known as "ticker tape analysis," has been updated and rebranded as "money flow analysis." The basic idea is that money flows, which can be systematically tracked, provide clues about future market performance. The tracking methodology is straightforward. Money flows are calculated by multiplying the quantity of stock traded by the price of the transaction. Price upticks are treated as buys, or inflows. Downticks are treated as sells, or outflows. Netting the two out provides a picture of money flowing into or out of the stock. The longer a pattern of inflows or outflows persists, the more reliable it is supposed to be as a forecaster.

Laszlo Birinyi, perhaps the technique's best-known practitioner, likens the technique to analyzing sports matches.[4] Consider an analysis of a soccer team. Instead of just looking at the recent won/lost record, it might make sense to also consider the average number of shots on goal to get a more balanced picture. In baseball it might make sense to see how many hits a team had, how many walks, and how many men got on base in the last five games, rather than just the won/lost record.

The analogy in the stock market is divergence. If money is flowing into a stock but the stock is trending sideways, champions of money flow analysis would argue that accumulation is going on that is not fully reflected in the market's valuation of the stock. It is an indication that the stock may be ripe for buying, recent losses, or apparent lack of trend notwithstanding.

OSCILLATORS, MOMENTUM, AND CONTRARY OPINION INDICATORS

When something oscillates, it swings back and forth. With respect to technical analysis, *oscillate* implicitly refers to some sort of fluctuation around a mean value. With that in mind technical analysts have devised oscillators that purport to measure overbought and oversold conditions. Typically they rely on price momentum, rates of change, relative strength, and convergence/divergence indicators. Based upon their reading of the oscillators, technical analysts seek to predict future market behavior.[5]

Price oscillators measure differences between moving averages of different time frames. When the oscillator turns up, meaning that the short-term moving average is higher than the long-term moving average, it is considered a bullish sign. A popular variation on this theme is the moving average convergence/divergence (MACD) indicator. This MACD indicator is generally calculated as the difference between a security's 12- and 26-time period exponentially weighted moving averages. A signal, or crossover line, is created by calculating a 9-period moving average of the difference. A buy (sell) signal is generated when the MACD rises above, or falls below, the crossover line.

Rate of change (ROC) and momentum oscillators are calculated to show how a stock is faring with respect to a prior period. The idea is to measure the pace of a trend to see whether it is picking up or losing steam. Momentum indicators are generally calculated as rolling price differences over fixed time periods. For instance, a 30-day momentum indicator would be the price today less the price 30 days ago: $M_t = P_t - P_{t-30}$. This measure allows momentum to be tracked as a time series that can be used as a comparison against other time periods.

ROC oscillators are very similar, the main difference being that they are expressed as ratios. The current market price is divided by the price from a prior starting date. For instance, the 30-day would be:

$$ROC_t = \frac{P_t}{P_{t-30}}$$

By virtue of the way they are calculated, both momentum and ROC measures oscillate at around zero, which technicians consider to be neutral ground. Scores above zero are thought to be bullish; below zero, bearish. A steeply sloped line indicates that momentum is picking up, but too steep a slope may indicate a trend reversal is in store. Some chartists argue that a slowdown in momentum augurs a trend reversal. Others consider it to be the pause that refreshes. Go figure.

Neither of these oscillators is scaled, so technicians put them in historical context for interpretive purposes. The scaling problem was addressed by the development of relative strength indexes (RSI) and "stochastics" oscillators which use mathematical formulas to normalize the results. Essentially, the RSI takes a relative strength measure and converts it to index form. The relative strength measure (RS) is the average of up-day price moves divided by down-day price moves over a fixed time period. The RSI forces the score to fall between 1 and 100 by using the following formula:

$$RSI = 100 - \frac{100}{1 + RS}$$

Stochastic oscillators are quite similar to RSI oscillators. They use a formula that includes price ranges over short-term and longer-term time periods, with the score normalized to range between 0 and 100. Stochastic oscillators are mostly used to identify overbought and oversold market conditions. Scores below 20 are thought to indicate an oversold condition and are therefore a buy signal. Scores above 80 are said to occur when the market is overbought and ripe for sale.

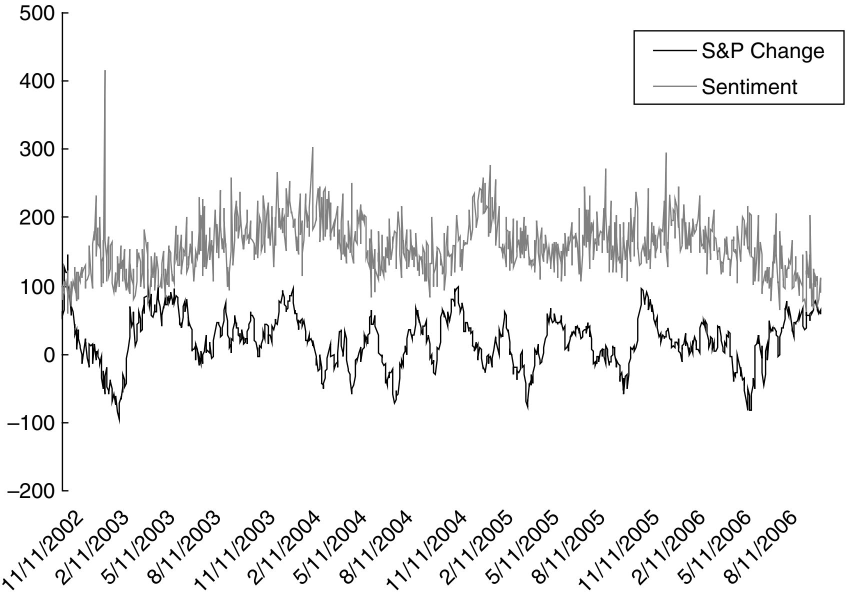
The McClellan indicator measures market breadth rather than price. Its methodology is similar to many oscillator type indexes. It is the smoothed difference of advancing versus declining stocks. The underlying belief is that market conditions reflect breadth or participation as well as price. Markets that have larger numbers of rising stocks over falling stocks are more bullish and vice versa. Another indicator, similar in construction as well as underlying belief, is an index based of new highs versus new lows.

Another popular type of measure is the sentiment indicator. Sentiment indicators, as the name suggests, attempt to measure the general psychological disposition toward the market. They are generally used as contrary indicators, the theory being that if "everyone" is bullish, they have already bought and are looking to sell (and vice versa). Two polls of bullish/bearish sentiment that are widely followed are the Investors Intelligence Survey and the American Association of Individual Investors. Following the contrarian logic of the sentiment surveys, some technical analysts follow the put/call ratio for common stocks. If it tilts too much in one direction or the other, the general belief is that the market will head in the other direction.

One such index is published by the International Securities Exchange (ISE) on its Web site.[6] The ISE put/call index attempts to get to the heart of investor sentiment by dividing the customer opening purchases of calls by customer opening purchases of puts. The theory is

FIGURE 21.3

S&P 500 30-Day Change versus Lagged Sentiment Indicator



Data sources: Yahoo Finance and the International Securities Exchange

that customer transactions are a better indicator than those of market makers. Whether or not market sentiment is a reliable indicator of future price action is another matter entirely. By and large technicians look for extreme readings to use as pivot points. Figure 21.3 illustrates this with a graph of the ISE sentiment index versus the 30-day change in the S&P 500. For the sample period, it does seem that peaks in sentiment have been accompanied by market fall-offs 30 days later. Although not rigorous, it is consistent with findings of Fisher and Statman with respect to market timing, cited previously.[7]

SUMMARY

At the heart of technical analysis is the notion that both prices and trading volumes contain information that can be used to infer future prices. Technical analysts argue that all that is known about the stock or bond (including fundamental research) is already embedded in the price. Despite asserting that all relevant information is embedded in the price, technical analysts part company with mainstream academicians who assert that markets are informationally efficient. Efficient market theorists argue that prices follow a random walk, precisely because all that is known is already in the price. Technical analysts argue that they have mastered the art of interpreting price and volume information in such a way that their predictions are better than the mere flip of a coin. There is, however, little empirical evidence to support the technicians' case.

Perhaps that is because the technicians' argument is fundamentally contradictory. If all that is known is in the price, how can a chart pattern have predictive value? Doesn't it (or its interpretation) have to add something to the mix? And if it adds to the mix, then all that is known is not already in the price. There is no way to escape the logic trap. Apparently, that has been very little deterrent. A great deal of time and effort continues to be invested in technical research. Books on the subject sell briskly, and even skeptics allow that short-term technical indicators can be self-fulfilling. And in this regard, some research suggests that technical analysis may be able to ferret out the depth of book under certain circumstances, while other research finds some negative correlation between market sentiment and market direction.

An important arrow chartists claim to have in their collective quiver is the ability to determine support and resistance points for stock, bond, index, and commodity prices. Closely related is the professed ability to interpret other market indicators like momentum, moving averages, and money flows as predictors of future market action. It is entirely possible, perhaps even probable, that some people are systematically better than others at gauging crowd psychology and are thus able to reap short-term trading profits by coolly exploiting the trading mistakes of others. But there is little empirical evidence that a trading system (as opposed to a gifted individual) has the ability to do this. In the end, the efficacy of technical analysis remains in the eye of the beholder.

NOTES

[1] See, for instance, Robert D. Edwards, John Magee, and W. H. C. Bassetti, *Technical Analysis of Stock Trends*, 9th ed., CRC Press, 2007.

[2] Kenneth Kavajecz and Elizabeth Odders-White, "Technical Analysis and Liquidity Provision," Rodney L. White Center for Financial Research Working Paper No 11-02. Available online at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=315660

[3] For a detailed discussion of these measures, see Leigh Stevens, *Essential Technical Analysis: Tools and Techniques to Spot Market Trends*, John Wiley & Sons, 2002.

[4] Birinyi Associates can be found here: http://www.birinyi.com/

[5] Leigh Stevens, 2002. It should also be noted that various formulas and methods for calculating and then displaying these types of measures can be found on Bloomberg.

[6] http://www.iseoptions.com/

[7] See the discussion of market timing in Chapter 19.

CHAPTER 22

Psychology and Markets: Behavioral Finance

> If you don't know who you are, the stock market is an expensive place to find out.
>
> –*Adam Smith,* The Money Game

Until comparatively recently, the economist's view of financial markets was almost antiseptic. Economists traditionally depicted traders as hyper-rational agents who carefully, coolly, and impartially weighed probable risks and rewards before jumping into a transaction. Other social scientists are considerably more skeptical as to the accuracy of this description. So is everyone else who has ever spent time on a trading floor.

Behavioral finance grew out of nagging doubts about the way economists traditionally viewed financial decision making. It rests upon two building blocks. The first is that there are limits to arbitrage, which implies that market inefficiencies can remain uncorrected indefinitely or at least for very long periods of time. The second is psychology; specifically the psychology of decision making and its relationship to rationality as defined by economists.

The traditional view of traders as rational utility-maximizing agents has come under assault by behavioral economists. In theory, individuals operating in their own self-interest are expected to buy the stocks they think are likely to produce the best risk-adjusted returns. By the same token, they sell the others. In the process all relevant information is priced into the market. Prices change as new information is introduced. Traders respond to new information by buying and selling stocks affected by changing circumstances. In the process stocks are efficiently priced as the new information is discounted. But what if traders can't be counted on to act rationally in the first place?

In theory, stock prices reflect the present value of expected future cash flows. To the extent that circumstances change in unexpected ways (e.g., there is a sudden change in interest rates or earnings fall shy of expectations), stock prices ought to rise or fall. But research has shown that stock prices are far more volatile than can be reasonably explained by changes in earnings and dividends. Trading volume is also much higher than would be expected if it were driven purely by market reaction to news. Moreover the return differential between stocks and bonds—the equity premium—is similarly much greater than risk differentials would suggest is appropriate. Small stocks and stocks with low price-to-book ratios have been shown to consistently outperform the market on a risk-adjusted basis—something that shouldn't happen if the market is efficiently priced by rational traders. These anomalies have been well publicized, but they persist. Why?

Perhaps the explanation is a failure to distinguish between information and insight. New information is subject to differing interpretations and frames of reference. Do some people have superior insight, and are they able to act on it? What if the market has both rational and partly rational players? Can the rational traders be counted on to overwhelm the quasi-rational traders so that efficient pricing prevails? Or will the herd mentality trample the rational players? Can rational traders systematically identify mistakes being made by quasi-rational traders and exploit them? How long can they hold out against the momentum of the crowd, bearing in mind Keynes's famous dictum that the market can stay irrational longer than you can stay solvent? Less ambitiously, are there recognizable psychological traps that traders can avoid?

PSYCHOLOGY, TRAPS, AND TRADING

Numerous studies have identified ways in which people (including financial experts) systematically err in their approach to decision making. Nicholas Barberis and Richard Thaler have summarized much of the research in a working paper published by the National Bureau of Economic Research.[1] In it they explain why they believe there are limits to arbitrage. They discuss a series of puzzling stock market pricing anomalies, and they identify repeated sources of error in financial market decision making. Key elements of their discussion are reviewed here because understanding systematic (and often hidden) biases in financial market decision making can be especially useful for traders, both to gauge market psychology and to avoid common traps.

Most market models assume that markets oscillate around true value. Arbitrage can be counted on to keep prices in line. To the extent that valuations are faulty, arbitrageurs will buy the cheap stocks and sell the expensive ones, forcing prices back toward true value. The problem is that arbitrage is not costless; nor is it without risk. Arbitrageurs, for instance, may value a security incorrectly. As a result of having been burned before, arbitrageurs may become gun-shy, thus allowing flawed valuations to persist. Similarly, noise traders may keep prices out of line for long periods of time. Finally, high transaction costs may present a barrier to arbitrage, allowing prices to stay uncorrected for long stretches.

More to the point, there is an extensive catalog of psychological and social factors that can impede rational decision making. In a trading context, a rational decision is one in which the expected gain is at least equal to the potential loss. Impediments to this type of thinking can be the source of costly trading error, so it is especially useful to be able to identify and counteract them.

The mind plays psychological tricks on us, even when we're on the lookout for them. One to which traders are particularly prone is finding patterns where none exist, a subject treated extensively (and well) by Nassim Nicholas Taleb in his book *Fooled by Randomness*.[2] Another is finding causes for every little wiggle in the markets. The financial press routinely falls into this trap. Rallies coming after declines are inevitably described as "short-covering"; sell-offs are "profit-taking." Psychologically, it's almost impossible for active market players to resist the temptation to read meaning into short-term market behavior. Plus talking to reporters looking for quotes gets your name in the newspaper.

While there is a lively debate about the impact of psychology on market efficiency, there is little doubt that it plays an important role in financial decision making. Individuals' beliefs, tastes, and preferences affect both the substance and style of their decision making. Social factors also come into play. Moreover, these factors can skew decision making away from what financial market economists would normally consider to be rational. For example, in her book *Out of the Pits* Caitlin Zaloom describes the thrill-seeking rather than risk-averting behavior of local floor traders who sometimes executed especially risky trades to boost their social status in the trading pits.[3]

In the economist's world people act in what they subjectively consider to be their own best interest. But this definition may be problematic. Lei Gao points out that economic agents change the definition of what they value to adjust to circumstances.[4] When people can't get what they want, they downplay its significance to rationalize why they didn't want it in the first place. This may not be irrational, but it is surely not the quasi-fixed subjective rationality economists refer to when considering decision making.

It is therefore important to examine research on factors that influence financial decision making. One emerging train of thought is that human evolution leaves us ill-equipped to judge financial markets. The survival of the fittest leaves our brains conditioned to pattern recognition with respect to what worked in the past. But markets are forward-looking and treat backward-looking strategies unkindly.[5] In addition, behavioral finance researchers, taking aim at economists' traditional assumptions about rationality, have identified several traits associated with "irrational" decisions. These include overconfidence, wishful thinking, representativeness, conservatism, belief perseverance, anchoring, narrow-framing heuristics, and availability biases.

For example, repeated testing has shown that people routinely overestimate their ability to assess probabilities. They discount too deeply events they think are unlikely to occur, and they overestimate the probability of events they consider likely. Researchers Fischoff, Slovic, and Lichtenstein found confirming evidence of this propensity. Their testing showed that sure things (in the subjects' views) happened only about 80% of the time.[6] "Nearly impossible" events (again in the subjects' view) actually happened about 20% of the time. Similarly, in financial market experiments Allen and Evans found that approximately 40% of their subjects had more confidence in their decisions than was warranted by the information they had.[7]

Professional investors display overconfidence in their abilities as well. In two studies of stock market forecasts, Torgren and Montgomery found that professionals and laymen produced the same subpar results, but each group incorrectly expected the professionals to fare better.[8] Overconfidence may not be distributed evenly. Gokul Bhandari and Richard Deaves found in a study that highly educated males are more subject to overconfidence.[9] A closely related problem is excessive optimism, or wishful thinking. Thaler and Barberis cite (in the paper mentioned earlier) studies showing that very large majorities (sometimes over 90%) consider themselves to be above average with respect to driving ability, physical appearance, sense of humor, and so on. Most people are overly optimistic planners, typically underestimating the time it takes to get a job done.

Not surprisingly, people feel more confident about things that they think they know something about. This can easily result in underutilizing the benefits of diversification. It is not at all unusual for people to concentrate their investments in the companies that employ them on the theory that they know something about the company. But this concentrates risk in a number of ways, not the least being that if the company goes down, employee-investors lose both their money and their jobs, as tragically happened at Enron. A similar phenomenon is home country bias: People tend to overconcentrate holdings in their own countries (although globalization may be mitigating some of the excess risk).

As it turns out, overconfidence leads both to excessive trading and to systematic trading errors. Barber and Odean found evidence that overconfidence led to traders systematically hanging onto losing trades while selling winners.[10] Moreover, the more people traded, the worse they did. Among other things, lots of trading racks up lots of transaction costs. They also found that men were more overconfident than women, traded more than women, and did worse than women.

Decision making can be biased by overgeneralizing from small information samples that lead people to make associations that appear to be (but are not) reasonable. Nobel laureate Daniel Kahneman and Amos Tvsersky illustrate the problem, referred to as *representativeness,* with the following exeriment. They describe a woman to a group of subjects as being a bright college graduate who majored in philosophy. She cares deeply about social justice, and she participates in antinuclear demonstrations. Then they ask the subjects if it is more likely that the woman is either (A) a bank teller or (B) a bank teller *and* active in the feminist movement.

Most participants in the experiment chose answer B. They thought it more likely that the woman was both a bank teller and active in the feminist movement. But that is a mathematical impossibility. To be both active in the feminist movement *and* a bank teller you have to be a bank teller first. There can't be more feminist bank tellers than there are bank tellers. But the people who were surveyed thought the woman's description sounded like someone who would be a feminist, incorrectly framed it that way and gave what sounded like the right answer even though it could not possibly be correct.

The problem of poor representativeness due to small sample bias can manifest itself in many ways. In a time series it can lead to an overemphasis on the most recent developments rather than on longer-term averages. It comes up in sports all the time, in the form of a winning streak or a player's "hot hand." In reality the hot hand is a statistical illusion that comes from paying more attention to recent performances than past ones. A related phenomenon is the gambler's fallacy, in which the bettor assumes his number is "due" to come in. The psychological grip of this fallacy is so powerful that when confronted with the evidence of it, otherwise rational people ignore the evidence and continue to rely on it. Go to a casino and watch how many people keep track of which numbers have already been spun on the roulette wheel, figuring that red or black, or odds or even are "due" even though each spin of the wheel is independent of the others.

Then there is the issue of mental accounting, in which a single decision is treated as two separate ones. Traders routinely do this. Consider a spread trader who agonizes over which way to leg into or out of a spread trade. The issue is the spread, not the long or short side of the transaction. Then there is narrow framing, which as the name suggests, involves a lack of circumspection. Too narrowly framing a situation can easily bias decision outcomes. For instance, the question, "Which stock should I buy" biases the decision in the first place toward investing in stocks rather than some other asset class, such as bonds.

Three loosely related traits likely to bias decisions are anchoring, conservatism, and belief perseverance, more popularly known as plain old stubbornness. *Anchoring* occurs when judgments or decisions revolve around a starting point—which may or may not be arbitrary. Arbitrary or not, psychological experiments have shown that people assign too much weight to the starting point, which becomes the anchor for subsequent discussion. Think about salary negotiations or putting a house on the market: Bargainers try to set the anchor either high or low depending on which side of the bargaining table they're on. In valuing securities, almost everyone is heavily influenced by the current price, assuming that it is at least in the ballpark of true value.

Conservatism refers to slowness to react to changing circumstances. The importance of past experience is overweighed in the decision calculus. In this, conservatism closely resembles anchoring in that acceptance of change is incremental, taking the form of marginal adjustments to the status quo. But conservatism may be at odds with representativeness. Conservatism places too much emphasis on long historical patterns; representativeness assigns too much weight to small samples and very recent events.

Belief perseverance, the innate tendency to cling to past beliefs, is closely related to conservation, manifesting itself in several ways. One is refusal to consider evidence that challenges existing beliefs (stubbornness). Another is to discount or ignore contradictory evidence. The extreme version is to argue that contradictory evidence actually supports the existing belief system. This tendency runs rampant among traders who ride losing positions into the ground. An example: A trader decides he is bullish on the bond markets because the price of oil is coming down, thereby reducing inflation pressures. Sure enough, the price of oil soars—and the trader is still bullish. He now contends that rising oil prices will slow the economy forcing the Fed to ease policy and bring down rates. Rising oil prices are bullish; falling oil prices are bullish.

Availability, or omitted variable bias, can skew decision making. When asked to judge the likelihood of some event happening, people tend to draw on their own memories and experiences to reach a conclusion. But memories may be faulty or incomplete. Witnesses called on to testify about the same event are notorious for coming up with very different versions of what they saw happen. Moreover, more recent events tend to be weighted more heavily than are older histories.

There can also be a good deal of difference between the way people actually behave and the way economics textbooks predict they are supposed to behave. Among other things, people tend to treat profit and loss differently, even when the amounts are the same. The evidence that loss stings more than profit satisfies has been demonstrated many times in psychological experiments. For instance, consider the following experiment. In phase one, subjects are given a choice. They can accept $100 in cash or they can bet on the flip a coin. If heads come up, they win $200; if tails come up, they win nothing. Given the choice, subjects overwhelmingly choose to take the $100 even though the coin flip has precisely the same expected value: $(1 - p) * 200 = 100$.

For phase two of the experiment, the terms of the bet are altered. The subjects can choose to pay a fine of $100, or they can bet on the flip a coin. If the coin comes up heads, the $100 fine is forgiven. If tails show, the fine is doubled to $200. In this version of the experiment the subjects mostly go for the coin flip, trying to gamble their way out of the fine. This result is the exact opposite of the previous experiment even though the terms are mathematically identical.

Does this happen in trading? Probably so; anecdotal evidence suggests that traders are more likely to try to leg out of losing spread trades than winning ones. Other research, cited previously, has found that they hang onto losing trades longer than winning trades.

Barberis and Thaler (in the paper cited earlier) note that economists tend to be rather skeptical about the importance of this body of research as it pertains to financial markets. Most economists believe that people learn by experience to correct their mistakes, investment professionals are less prone to error to begin with, and that error can be further reduced if investment banking firms adopt the right incentive structures. It turns out, however, that investment professionals are more likely than others to be overconfident of their financial acumen. Anyway, they argue, bias can never be wiped out entirely. But that goes without saying. The real question is how important the biases are on the margin, if they are important at all.

THE NEOCLASSICAL RESPONSE

Steven Ross of MIT, inventor of arbitrage pricing theory, is dismissive of behavioral finance. He argues that the neoclassical model of efficient markets is the superior one. Behavioral finance, he says, is ad hoc, devoid of theory, with little distance between hypotheses and conclusions. Moreover, it has little to offer by way of an alternative theory of financial markets. He argues further that the examples of irrationality cited by behavioral economists are insignificant in the grand scheme of things. Stock pricing anomalies tend to be small and occur in illiquid securities, calling into question their applicability in larger markets. Psychology may be useful for marketing purposes or analyzing flows of funds, but it has little to say about valuation. In the long run, Ross says, firms with the same cash flows will trade the same.[11]

Ross goes on to attack behavioral finance as amounting to little more than a collection of interesting anomalies lacking a theory that holds them together. Moreover, he says, the observations are sometimes contradictory. Behavioralists assert that people are both too pessimistic and too optimistic and that their conservatism prevents them from reacting to rapid change—but that they overweight the importance of recent events. He goes on to argue that the limits to arbitrage posited by behavioralists are simply an artifact of models that limit markets and institutional structures.

For the purpose of this discussion, which is to apply psychology to trading decisions, it is important to avoid conflating arguments over efficient markets with findings of psychologists about the way humans make decisions. No one seriously doubts that humans routinely make biased decisions. Whether those biases are large enough or systematic enough to cause prices to diverge from true value for significant periods of time is another matter entirely. But traders can certainly benefit from understanding what behavioral finance has to say if only to avoid falling into the traps that so many people appear to make to the detriment of their bottom lines. Reducing error can be just as important as occasional brilliant market calls. More important in fact: One too many errors and you are out of business.

STRATEGIC BEHAVIOR, PSYCHOLOGY, AND THE RATIONAL TRADER

Discussion of the psychology of financial market decision making often seems to focus on academic theories of economic rationality or its apparent absence. On the one hand behavioral economics point to numerous anomalies that suggest "irrational" behavior; taken together, they argue, enough of a pattern emerges to challenge the idea that financial markets are rationally priced. Neoclassicists retort that (1) a few anomalies do not constitute a serious attack on economic rationality, (2) arbitrage will put prices right in the long run, and (3) arbitrage failures are not due to irrationality but to institutional constraints. Seemingly left out is any discussion of strategic behavior.

Strategic market behavior can be thought of as trading with intent to influence market prices. For example, a nonstrategic trader who thought the market was likely to go up would simply buy. On the other hand, a strategic trader is one who would buy in a manner designed to make it go up. As it turns out, strategic trading appears to be common practice.

At quarter end, when portfolio managers are marked-to-market, the financial press routinely reports that trading reflected "window dressing." *Window dressing* refers to a practice in which portfolio managers buy more stock in companies they already own around the close in order to push the price up, thus hyping their reported quarterly returns.[12] A similar situation would be one in which a market maker initially buys (or sells) only a portion of the position size he really wishes to acquire. There are many possible reasons for doing this, some of which involve strategic calculation.

An example would be buying 80% of a position in the morning, waiting for the close to aggressively buy the remaining 20%, hoping in the process to push the closing price up and therefore the official mark-to-market. Another would be to acquire only a portion of the desired final position in order to save ammunition to be used in the event that it is needed to defend the position by buying more if under attack. The term *attack* is used advisedly in this context. Traders (as opposed to portfolio managers) tend to view trading as a zero sum game. My win is your loss. In that sense trading is analogous to a war game. To the extent you can weaken the opposition's morale by having the closing marks-to-market go against him, he may be more likely to bail out leaving you more likely to prevail.

It should be noted that the regulatory authorities almost certainly take a dim view of this type of jockeying. With perfectly competitive markets the preferred state of affairs, neither producers nor consumers can influence the market price. Accordingly, there are rules against market manipulation. But in financial markets the line between manipulation and influence can be blurry, hard to detect, and even more difficult to enforce. So it happens.

The fact of strategic behavior in financial markets has important implications for market psychology and efficiency. First, it means that markets are not as transparent as they might appear to be. Remember the underlying postulate of Bachlier that is embedded in modern finance theory: In auction markets traders put their best foot forward. The market is efficient insofar as it reflects the highest price buyers are willing to pay and the lowest price sellers will accept. But if some traders systematically hold back for the purpose of influencing prices later, current prices do not truly reflect all that is known. Markets are neither as transparent nor as informationally efficient as they would appear.

Second, if some traders systematically hold back and some put their best foot forward, that implies that there are different types of traders in the market, equipped with different styles and psychological makeups. Academic research has indeed detected fascinating systematic differences in trading with respect to attitudes about money, conceptions of morality, and sociability to name a few.

Carmen Keller and Michael Siegrist performed a cluster analysis of investors in Switzerland and found that they could be classified as one of four types: safe players, open books, money-dummies, and risk-seekers.[13] Harrison Hong, Joseph Chen, and Jeremy Stein found that sociable households were far more likely to invest in stocks, other things being equal.[14] Garvey and Murphy found that individual traders adopt riskier strategies than institutional traders to try to cover the costs of higher trading commissions.[15] In examining the behavior of day traders, Jordan and Diltz discovered that 65% held losing trades longer than winning trades, an example of asymmetric treatment of profits and losses.[16] In a study of Japanese investors, Konari Uchida was able to determine that online traders were more likely to be young men and that employed people trade more, but successful past trades did not produce increased trading volume. Further he found that Japanese online traders prefer returns in the form of capital gains; they do not prefer low-volatility stocks; they choose stocks on their own and rely on stock charts when making trading decisions.[17] By way of contrast, Louis Lowenstein studied a group of value fund managers who were distinctly inner-directed, confident, willing to swim against the crowd, and had low portfolio turnover. These managers outperformed their benchmarks by a whopping average of 11% per year from 1999 through 2003.[18]

While behavioral research provides clear evidence that some traders (or even most) make decisions that are based on flawed reasoning, it does not provide convincing evidence that market prices are systematically biased in a predictable way. Nor does it provide much empirical evidence that particular decision-making techniques yield superior results. But it does suggest that there are traps that should (and can) be avoided and approaches that are worthy of further exploration.

SUMMARY

Behavioral finance examines how the psychology of decision making affects the substance of the decision. Studies have found ample evidence that people are not automatons; they are liable to make impulsive trading decisions, run up transaction costs, miscalculate risk and reward probabilities, have difficulty admitting error, focus their attention too narrowly, and become overconfident of their abilities. Behavioral finance has discovered numerous pricing anomalies and shown that arbitrage can be difficult. The neoclassicists respond that anomalies do not constitute theory and that arbitrage difficulties are the result of institutional constraints, not market inefficiency.

It is worth noting that studies of trader behavior seem to have retail traders as subjects. That is probably because there is little reason for professional traders to want anyone to discover trade secrets, so to speak. Since studies have clearly shown stylistic differences among traders, it is entirely possible that much of the behavioral research suffers from omitted variable bias due to the relative paucity of professional trader studies, although some work suggests superior returns accrue to market makers. But that may be due more to privileged position than trading ability.

Nevertheless, behavioral research provides valuable guidelines for traders. The first is to avoid overtrading. The evidence on this score is fairly clear: Too much trading racks up transaction costs rather than gains. The second is that stubbornness can be detrimental to P&L. The third, holding onto losses while taking quick gains, is closely related to stubbornness. Problem trades need to be recognized and dealt with. Putting off the inevitable just makes it worse in the end. There is a lot of truth in the old Wall Street adage that the best loss is the first loss. Finally, studies that show superior performance by value-based inner-directed portfolio managers hint at important, but often ignored, lessons about trading. A thorough understanding of the fundamentals, combined with a certain amount of stoicism and hard work, is more likely than not to pay off over time. Kind of like golf.

NOTES

[1] Richard Thaler and Richard Barberis, "A Survey of Behavioral Finance," NBER Working Paper No. 9222, September 2002. See also Richard Thaler, *The Winner's Curse*, The Free Press, 1992.

[2] Nassim Nicholas Taleb, *Fooled by Randomness*, Texere LLC, 2001.

[3] Caitlin Zaloom, *Out of the Pits*, University of Chicago Press, 2006, especially pages 98–109.

[4] Lei Gao, "Self Is Never Neutral: Why Economic Agents Behave Irrationally," *Journal of Behavioral Finance,* 6/1, 2005, pp. 27–37.

[5] Terry Burnham, *Mean Markets and Lizard Brains*, John Wiley & Sons, 2005, p. 6.

[6] B. Fischoff, P. Slovic, and S. Lichtenstein, "Knowing with Certainty: The Appropriateness of Extreme Confidence," *Journal of Experimental Psychology: Human Perception and Performance* 3, 1977, pp. 552–564.

[7] W. David Allen and Dorla A. Evans, "Bidding and Overconfidence in Experimental Financial Markets," *Journal of Behavioral Finance,* 6/3, 2005, pp. 108–120.

[8] Gustaf Torngren and Henry Montgomery, "Worse Than Chance? Performance and Confidence among Professionals and Lay People in the Stock Market," *Journal of Behavioral Finance,* 5/3, 2004, pp. 148–153.

[9] Gokul Bhandari and Richard Deaves, "The Demographics of Overconfidence," *Journal of Behavioral Finance,* 7, 2006, pp. 5–11.

[10] Brad M. Barber and Terrance Odean, "The Courage of Misguided Convictions," *Association for Investment Management and Research,* Nov.–Dec. 1999, pp 41–55.

[11] Ross has delivered his broadsides against behavioral finance in a series of lectures as well as in his book *Neoclassical Finance*, Princeton University Press, 2005. For a link to the lecture series see: http://www.ima.umn.edu/public-lecture/2003-04/ross/index.html

[12] See, for instance, the story "Dow Tops 12500 in Year-End Run" by Scott Patterson, *The Wall Street Journal,* December 28, 2006.

[13] Carmen Keller and Michael Siegrist, "Money Attitude Typology and Stock Investment,"*Journal of Behavioral Fianance,* 7/2, 2006, pp. 88–96.

[14] Harrison Hong, Joseph Chen, and Jeremy Stein, "Social Interaction and Stock-Market Participation," Graduate School of Business, University of Chicago, Working paper, June 2001.

[15] Ryan Garvey and Antony Murphy, "Commissions Matter: The Trading Behavior of Institutional and Individual Active Traders," *Journal of Behavioral Finance,* 5/4, 2004, pp. 214–222.

[16] Douglas Jordan and J. David Diltz, "Day Traders and the Disposition Effect," *Journal of Behavioral Finance,* 5/4, 2004, pp. 192–200.

[17] Konari Uchida, "The Characteristics of Online Investors," *Journal of Behavioral Finance,* 7/3, 2006, pp. 168–177.

[18] Louis Lowenstein, "Searching for Rational Investors in a Perfect Storm: A Behavioral Perspective," *Journal of Behavioral Finance,* 7/2, 2006, pp. 66–74.

CHAPTER 23

Breaking Up Is Easy to Do: Portable Alpha

Up to this point the strategies discussed in the book have been market specific, focusing on the instruments, institutions, and price drivers of particular asset classes. Bond market strategies considered yield-to-maturity, the business cycle, the shape of the yield curve, convexity, and duration. Price drivers included Fed policy, basis trading, inflation expectations, and quality differentials. Equity market strategies focused on valuation techniques, market levels, business cycles, market sector differentiation, and growth versus value. Strategy implementation focused on the use of equity index futures and exchange-traded funds.

The focus of this chapter is rather different. It considers mixing and matching strategies and methods from different asset classes and blending them into a "metastrategy." The technique, used by many hedge funds, is called *portable alpha*.

The idea of portable alpha comes from the CAPM framework. In the CAPM framework there are two sources of return. The first is the risk-free rate; the second is return that compensates for risk taking. Within a given asset class risk can be minimized by diversification, but there still remains an irreducible level of risk. The return associated with that irreducible risk is that of the market portfolio, or *beta*. A portfolio manager exactly replicating the market portfolio (with no transaction costs) would realize the market rate of return. Beta would measure 1.

In the real world, transaction costs along with portfolio strategy and its implementation matter a great deal. As the saying goes, you can't manage what you can't measure, so in order to gauge performance, portfolio managers are benchmarked against a target index. To the extent that portfolio returns differ from those of the market portfolio, variation can be attributed to portfolio management skills. This means that total portfolio returns (P_r)

can be disaggregated into two segments: the market return, or beta (β), and the contribution of the portfolio manager, denoted as *alpha* (α), or:

$$P_r = \alpha + \beta^* (\text{Market}_R)$$

Separating the elements of return in this manner allows for portfolio attribution analysis and provides the basis for developing and executing more sophisticated trading strategies. To the extent that portfolio total return is the same as the market return ($P_r = \text{Market}_R$) after controlling for risk, the portfolio manager's net contribution is nil ($\alpha = 0$) However, if the portfolio total return exceeds the market return ($P_r > \text{Market}_R$), the portfolio manager has made a positive contribution ($\alpha > 0$). If the portfolio's total return is less than the market return ($P_r < \text{Market}_R$), the portfolio manager's efforts have been a drag on performance ($\alpha < 0$).

This simple model of performance can be used to analyze risk posture, portfolio management styles, and strategy. For example, risky portfolios ($\beta > 1$) would be expected to have greater than market returns.[1] Returns variation among portfolios with the same beta might be attributed to differences in transaction costs or market timing. In fact, variation in return is necessarily limited to differences in security selection, market timing, and transaction costs. But it is important not to frame the returns universe too narrowly. This requires thinking about the combined issues of performance benchmarking and strategy selection.

In targeting a particular returns universe, the benchmark exerts a powerful influence on trading strategy for both active and passive portfolio management styles. For instance, passive strategies simply seek to mimic the benchmark index and efficiently replicate the benchmark's returns. The obvious strategy for passive managers is to minimize transaction costs. On the other hand, active managers need to create strategies that beat the benchmark index. Paradoxically, it is not unusual for active managers to construct portfolios that differ from the benchmark only on the margin, not in a fundamental way. This is particularly true as portfolio size increases. If the benchmark is the S&P 500, the portfolio will tend to look a lot like the S&P 500, even for an active portfolio. Consequently, active managers are sometimes referred to as *closet indexers*.

A potential problem with this approach is that it can easily conceive the returns universe in too narrow a fashion. The S&P is a good measure of big cap stocks in the United States, but the larger returns universe includes bonds, small cap stocks, overseas equities, and real estate. It might also include collectibles like art, precious metals, and soft commodities. In this respect it is well to remember that the CAPM, upon which so much modern finance is based, is a very broadly conceived theory of equilibrium. The market portfolio encompasses the whole returns universe, including different asset classes. And to the extent that expected returns can be improved or risk reduced by asset class diversification, competitive market forces will induce money managers to broaden both the range of asset classes they trade and the types of trading strategies they offer.

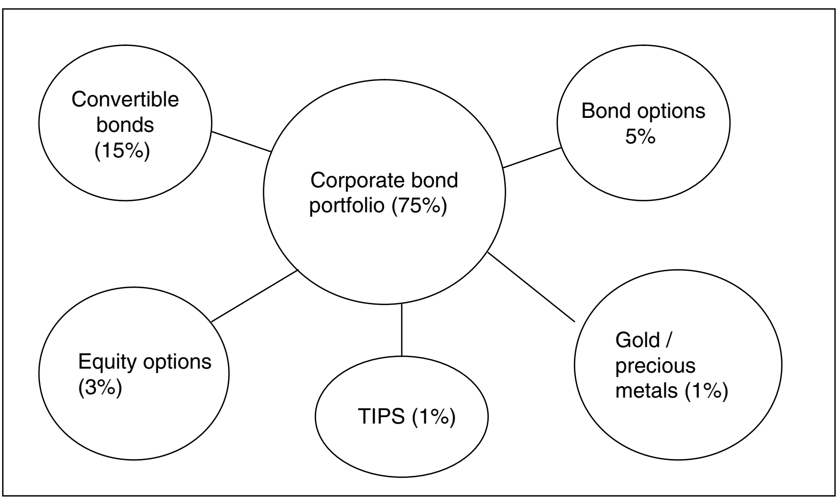
But certain trading strategies and portfolio management skills may work better with different asset classes. The challenge is to bring those skills to bear across a broad array of asset classes and trading styles without diluting money management expertise. One way is to adopt a core-satellite portfolio structure that, with a little financial engineering, allows trading skills and strategies to be ported across asset classes.

CORE-SATELLITE PORTFOLIO STRUCTURE

A core-satellite portfolio structure is what it sounds like. The core of the portfolio is invested in one asset class. Other subsidiary holdings are at the periphery and are designed to accomplish specific tasks for the purpose of enhancing overall portfolio returns. Some strategies may be designed to dampen overall portfolio volatility; some may be designed to diversity holdings across different asset classes; and some may be designed to offset particular types of unwanted exposure in the portfolio core. For example, long equity positions in LBO candidates might serve as an offset to the risk inherent in owning bonds of potential target companies. Satellite positions in options or volatility indexes might offset portfolio volatility. A schematic of this type of structure is shown in Figure 23.1.

FIGURE 23.1

Core Satellite Portfolio Structure



PORTING ALPHA

The core-satellite graph shows capital being deployed across different asset classes, with the percentage deployments listed in parentheses, summing to 100%. Portable alpha strategies borrow from the core-satellite structure, but with two critical differences. First, porting alpha requires leverage. Second, rather than simply mix together returns from different asset classes, portable alpha uses leverage to graft excess returns from superior trading skills (alpha) onto core portfolio holdings (beta). Doing this requires some financial engineering.

For example, consider an actively managed portfolio benchmarked against the Lehman Aggregate Bond Index (Lehman Ag). Suppose that the portfolio is permitted and able to use leverage and is open to market-neutral strategies in other asset classes, for instance, equity pairs trading. Porting alpha extracted from other asset classes and grafting it onto the returns of the portfolio core can enhance overall performance. A hidden advantage of using other asset classes to execute alternative strategies is that they may be less highly correlated with the portfolio core, thus providing a measure of diversification.

Implementing a portable alpha strategy entails three preliminary steps. The first can be thought of as *levered beta acquisition*. It entails gaining exposure to the target index (beta) on a levered basis, which frees up cash for alpha generation. The second step is to seek and find alpha wherever it resides. Then transactions designed to extract alpha can be executed in the appropriate markets using the freed up cash. All this is predicated on a third prerequisite, which is a risk budget.

Because strategy implementation requires leverage, the risk profile is greater than it would otherwise be. The trade-off between increased risk and potentially greater returns can be evaluated by using historical risk and returns data. That analysis can be used to determine acceptable levels of leverage. Once risk parameters are established, the core position is acquired.

There are a number of ways the portfolio core (or beta) can be acquired. For instance, in this example, corporate bonds constitute the core asset class holding, with the Lehman Ag serving as the benchmark index. One possibility is to purchase (on a levered basis) a representative sample of the bonds that comprise the index. The bonds could be purchased and financed (all or in part) using the RP markets. The excess cash gained from the RP transactions can then be used to finance purchases of additional securities in other asset markets. A second possibility is to execute a total returns swap against the Lehman Ag with a dealer. The dealer guarantees the total returns of the Lehman Ag over a specified period of time in return for interest payments every quarter at the three-month LIBOR rate. Since exposure to the Lehman Ag is gained via the swap route, no cash changes hands. The cash that ordinarily would have been used to purchase bonds is available for other transactions designed to produce alpha from other asset markets. If no alpha strategies are immediately available, the funds can be invested in the three-month LIBOR, which effectively makes interest payments to the swap dealer a wash, except for transaction costs.

A third way to acquire levered beta is with futures contracts. In fact, futures contracts in financial instruments can be thought of as beta generators for the particular asset class they represent. In this example the Lehman Ag, the target index, could be replicated by using combinations of CBOT Treasury futures contracts, or Treasury futures contracts combined with OTC swaps and purchases of lesser-quality credits. The Chicago Board of Trade also publishes the Dow Jones CBOT Treasury Index (Bloomberg ticker DJCBTI) that can be used as an index model or as a swap reference point. Futures contracts require a relatively small performance bond. Normally, the performance bond requirement can be met by putting up between 2% and 5% of notional principal value with T-bills. That leaves between 98% and 95% of the original cash position to be deployed in alternative strategies.

Once levered beta has been acquired, the next step is to find and execute potential alpha-generating transactions with the freed up excess cash. For the purpose of example we assume that alpha returns (negative or positive) come from executing a market-neutral equity pairs trading strategy. Since the pairs-trading is market neutral, excess returns are being added to the portfolio as a result of trading skills (alpha), not as compensation for increased risk taking in the stock market. Total portfolio returns are the sum of alpha returns produced by equity pairs transactions plus beta returns acquired by using derivatives (swaps and futures) or cash freed up via the RP markets or both.

Using this example as a conceptual baseline, there are a number of factors that need to be considered. Alpha returns can come from any number of strategy types employed in various asset markets. By definition, beta returns are the market returns associated with a specific asset class. Total returns are a combination of the two. By implication, for the strategy to be considered successful, it should generate total returns that are greater than those that would ordinarily accrue to a passive strategy in the targeted asset class, provided one of two conditions is met. First, the increase in marginal returns should not be accompanied by greater portfolio risk. Alternatively, increased portfolio risk must generate additional expected returns within the constraints and trade-offs established by the risk budget. Second, the strategy can generate the same returns but with reduced portfolio risk.

It is also important to note that the freed up cash used to finance alternatives strategies is not free. Its implicit cost is likely to be somewhere in the vicinity of the risk-free rate. To see this, consider again the cash-and-carry model of futures contracts. When combined with the three-month T-bill rate (effectively the risk-free rate), a financial futures contract approximates the total return of the underlying bond, index, or commodity. When the cash that would have been invested in T-bills goes into an alternative investment, that T-bill rate is the opportunity cost; that is, the implicit cost of investing those funds. A successful transaction therefore needs to generate returns that are greater than the opportunity cost of the available funds.

YOU CAN'T HAVE ONE . . .

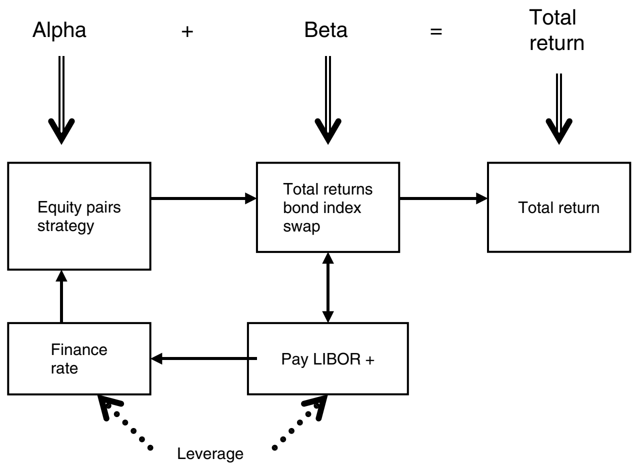
Portable alpha combines an array of diverse trading tactics in different markets with market specific beta returns, usually measured by an underlying benchmark index, with the aim of generating higher risk-adjusted returns. Active trading strategies like these can generate significant transaction costs that need to be managed. For instance, there is a cost associated with beta acquisition, so alpha returns need to be sufficiently robust to overwhelm any potential added trading costs. Futures contracts are often the weapon of choice here because they are beta equivalents and trading costs tend to be low. But in the end, trading strategies designed to produce portable alpha need to generate returns higher than the marginal cost of funds plus beta acquisition costs. Otherwise total returns will fall below those of the benchmark, marking failure.

There are quite a few moving parts in a portable alpha strategy, so sometimes it can be helpful to use a flow chart, as in Figure 23.2, to visualize the transaction structure.

The flow chart tracks both the core asset holding and the diversification inherent in the portable alpha strategy. In addition, it hints at the relationship between financing and overall portfolio strategy. The top of the diagram tracks the transaction path; the bottom illustrates position-financing mechanisms. In the flow chart example, the preferred core asset class is bonds. In a conventional portfolio, the bonds would be purchased and the coupons periodically reinvested. In this example, the core bond holding is achieved with a total returns swap versus paying LIBOR. All or part of the cash saved from executing the swap is either reinvested at LIBOR or used to finance alterative strategies whose returns are added to the swap returns. The combination of all those returns produces total portfolio returns. What is clear from the flow chart is that the floating rate on the swap is the effective financing rate for the leverage needed to implement alternative portfolio strategies to graft onto the portfolio core. In this case the alternative strategy added to the mix is equity pairs trading.

FIGURE 23.2

Structure of a Portable Alpha Transaction Strategy



HIDDEN RISKS: UNINTENDED CORRELATION

In principal the portable alpha model can be easily extended to include many different types of trading strategies executed in different asset markets. But it is important to make sure that the anticipated returns from peripheral strategies are not highly correlated with returns from the core asset class. That could result is concentrated risk, the opposite of the intended outcome. Correlation avoidance across financial assets can be more difficult than it sounds. One way to proceed is with historical data, testing for significant returns correlations across different asset classes and strategy types.

Table 23.1 is an example of a correlation matrix calculated for this purpose. It displays pairwise daily returns correlations of several different asset classes and hedge fund trading strategies from 2000 through 2005. The asset classes are fixed income, represented by the Dow Jones CBOT Treasury index; stocks, represented by the S&P 500; along with oil markets and agricultural commodities, captured by the Dow Jones AIG Commodity Index. Historical returns from popular hedge fund trading strategies include: convertible bond arbitrage, distressed securities (emerging from bankruptcy proceedings), event driven (such as mergers and acquisitions), equity neutral (pairs trading), merger (risk arbitrage), and equity long/short, which is a pairs type trading strategy with a long bias.

The matrix displays both the degree and significance of correlation. The coefficient for each pair indicates the degree to which the returns of

TABLE 23.1

Correlation Matrix across Different Asset Classes and Strategy Types

Pairwise Correlations	DJCBTI	S&P 500	CBOE Oil	DJ-AIG	Convertibles	Distressed	Event	Equity neutral	Merger	Equity long/short
DJCBTI	1.00									
	1417									
S&P 500	–0.2670*	1.00								
	1417	1489								
CBOE Oil	–0.0829*	0.4500*	1.00							
	1417	1489	1489							
DJ-AIG	0.008	0.0198	0.3437*	1.00						
	1417	1489	1489	1489						
Convertibles	0.4016*	–0.2823*	–0.1107*	0.1100*	1.00					
	465	493	493	493	493					
Distressed	–0.0528	0.3778*	0.2794*	0.1046*	–0.0754	1.00				
	465	493	493	493	493	493				
Event	–0.0636	0.7367*	0.5318*	0.1358*	–0.1897*	0.5434*	1.00			
	465	493	493	493	493	493	493			
Equity Neutral	0.0846	0.1736*	0.3241*	0.1707*	–0.1176*	0.1907*	0.2504*	1.00		
	465	493	493	493	493	493	493	493		
Merger	–0.1015*	0.5712*	0.3478*	0.0279	–0.2102*	0.3450*	0.6840*	0.2013*	1.00	
	465	493	493	493	493	493	493	493	493	
Equity Long/Short	–0.0005	0.8007*	0.6980*	0.2379*	–0.2072*	0.4508*	0.4508*	0.3468*	0.585*	1.00
	272	288	288	288	288	288	288	288	288	288

* Significant at =0.05

Data source: CBOT; CBOE, Dow Jones. Time period = 2000 through 2005.

one asset class or strategy are explained by the other. Pairs labeled with an asterisk are statistically significant at $p = 0.05$, which indicates that there is only a 5% chance that the correlation is the result of chance. The sign before the coefficient shows the direction of correlation. For instance going by the matrix, correlation between fixed income and equities is negative; about 27% of negative Treasury returns are explained by positive equity returns.

These correlations can provide valuable information when deciding which strategies to pair up with which core asset classes. Since stocks and bonds have had negative correlations and equity long/short strategies are uncorrelated with bonds, it would seem to be worth examining market neutral equity pairs trading strategies as an overlay for core fixed-income holdings. Not surprisingly, convertible bond arbitrage has had a highly positive correlation with Treasury returns, but a negative correlation with stocks. Consequently, it would seem to be a better overlay candidate for equities than for fixed income. However, it should be noted that correlations across asset classes and strategies can easily drift over time and may act differently as circumstances change.

SUMMARY

The idea behind portable alpha is to separate returns attributable to investment management or trading skills from returns that adhere to a particular asset class. Then returns to management skills can be ported over to core portfolios in other asset classes. Properly executed, this strategic approach affords the possibility of reducing risk (by strategy diversification) and increasing returns. But implementation requires the use of leverage, so risk budgets need to be established. In addition, satellite strategies need to generate excess returns over the combined costs of financing and beta acquisition.

Portfolio managers typically advertise alpha-generating skills, but that is only part of the picture. In a portfolio management context, total return, which is what really matters, is a combination of market return (beta) and portfolio management skills. Producing superior total return, properly adjusted for risk, requires efficient acquisition of beta, risk budgeting, and market-beating trading strategies. Because index futures contracts are natural low-cost beta generators, they often serve as the beta core of a portable alpha strategy.

NOTES

[1] Refer to Chapter 6 for calculating the required rate of return when beta is known.

CHAPTER 24

Managing Trading and Position Risk

If you don't risk anything, you risk even more.

–Erica Jong

In the early 1970s the bond business began its emergence from near obscurity, eventually to become the epicenter of the Street's trading and risk-positioning activities. Risk management as we know it today simply didn't exist. It mainly consisted of getting marks-to-market right so that the P&L was a useful guide to how the firm was doing day to day. Consider for example the real live case of Patrick O'Shea, a big blustery Irishman who for a time ran the government desk of one of the larger dealers. Patrick O'Shea is not his real name, course, but the character is real. For the sake of innocent bystanders we'll use that pseudonym for him and refer to his firm as EF Walston & Co.

Known around the Street as "the Major," befitting his rank as a former Air Force pilot, O'Shea was nothing if not colorful. Anyway, the Major decided to go long the bond market. Nothing is ever halfway with the Major. When he likes the market, he *really* likes the market. As he put it that morning in the weekly department managers' meeting, he planned to "buy everything in sight." And so on this particular Monday, right at the market's open, his entire trading desk began to buy. That turned out to be relatively easy since the market was in the throes of a substantial sell-off. The Major was, however, undeterred.

As was the custom, each trader would mark his positions to market, estimate a P&L, and turn it in to the Major after the close. At this point the Major would tote up the individual P&L estimates to arrive at what became known as the daily WAG, an acronym standing for wild ass guess, a description the traders considered hilarious. The partners, whose money was on the line, were not as easily amused. Accordingly, every day the

Major would turn in the WAG to George Dunn, the partner to whom he reported. (The firm, like many back then, was a partnership).

The slaughter in the bond market that started at the open that particular Monday was mercifully ended only by the market's close. Since Patrick and his traders had spent the better part of the day buying, the daily WAG was pretty grim. The Major nevertheless duly wrote it down on behalf of the department, marched it over to George Dunn's desk on the trading floor, and wordlessly handed it to him. Whereupon Dunn handed it back without comment. Pretty much the same thing happened Tuesday and Wednesday with the market continuing to sell off while the traders continued to buy.

By Thursday morning the Major had accumulated several hundred million bonds, a position most would consider to be fairly substantial, at least by the standards of the day. More to the point, George Dunn thought it was a pretty hefty position. So that morning about ten o'clock or so Dunn decided to stroll over to the government trading desk to check out how the market was doing. The conversation, as recounted to the author, went something like this.

GD: Good morning, Patrick.

PO: The name is Major.

GD: OK, Major. How is the market doing?

PO: They are beating the living hell out of it.

GD: I see. *Long silence. Then*, What are you doing about it?

PO: I'm buying. *Longer silence.*

GD: Patrick, do you think the sell-off is approaching an end?

PO: The name is Major.

GD: *Irritation now beginning to show.* All right then, Major, is it your opinion that the selling is almost over?

PO: There doesn't appear to be any let up in sight.

GD: *Now through gritted teeth,* Why then, may I ask, do you continue to buy when the market is going down?

PO: Because that's the only time you can buy any size.

Risk management techniques have changed since then.

DEFINING RISK

Risk can be defined as the possibility of suffering harm or loss. Sources and types of risk are numerous and varied. Some types, like operational risk, are internally generated and are more or less controllable. Others are external, like weather and markets, and are more difficult to manage. The General

Association of Risk Professionals (GARP), a nonprofit professional organization for risk managers, has an extensive library of papers on sources of risk and ways to manage it, many of which are published on its Web site. GARP also runs education and training programs in risk management, and it is a valuable source of information for developments in the field.

In financial markets, risk is treated probabilistically. The question is not whether things can go badly; of course they can. The risk assessment process considers the likelihood that things will go wrong and the extent of the damage if and when they do. Trading risk models typically express risk exposure with respect to the potential for loss on any given trading day. Commonly used metrics are daily earnings at risk (DeaR) and daily value at risk (VaR). These metrics are designed to estimate the outer bounds of what might reasonably be expected to happen based on past market performance. A more recent version is crash metrics, which, as the name suggests, is designed to provide an estimate of what can happen if things get really bad. Once calculated, these metrics can be used to compare risk with expected gains from trading positions. Then a risk/reward profile can be drawn.

SOURCES OF RISK

Within the general category of market-based risk there are several subcategories that need to be considered. Among the more important are valuation risk, volatility risk, liquidity risk, event risk, and credit risk. Valuation risk is at the top of the list for the obvious reason that the market value of a security has to be discoverable in order for the position to be managed. Although some securities are easy to value, others can be difficult. Consequently, there is a hierarchy of valuation risk that depends on where the security trades, its transparency, and its level of complexity.

Plain-vanilla listed securities are the easiest to value. Usually the last sale, the bid or offer on the listing exchange, serves as an end-of-day mark-to-market. For example, there is little dispute about how to approximate the dollar value of 10,000 shares of Coca-Cola at the end of the trading day: It's the last sale price times the number of shares.

Relying on exchange prices has enormous advantages. Among them are transparency and third-party pricing, which makes it difficult to incorrectly mark positions to market. Mindful of their stewardship responsibilities, exchanges keep copious records of transactions and publish open, high, low, and closing prices daily. Prices of listed derivatives can be found in the financial press and are generally available on exchange Web sites and through data vendors like Bloomberg. The exchanges, which are mostly self-regulating organizations, enforce trading rules to keep the markets free of price manipulation. Regulatory agencies, including the National Association of Security Dealers (NASD), the Securities and

Exchange Commission (SEC), and the Commodities Futures Trading Commission (CFTC), monitor trading as well. Violators are subject to severe sanctions, including criminal prosecution.

Over-the-counter securities and derivatives can be more difficult to value, sometimes due to the type of security, sometimes due a lack of trading volume, sometimes due to the complexity of the instrument itself. The fact that a security is listed on an exchange should not be taken to mean that the listed market is the main trading venue. Although lots of corporate bonds (and all U.S. government bonds) are technically listed on the NYSE, as a practical matter virtually all trading in them takes place in the OTC markets.

Government bond prices are published in the financial press, but only a limited number of representative corporate and municipal bonds are. Because the government market is so big and liquid, it's generally fairly easy to get good quotes on individual bonds. Corporate and municipal bonds are another matter entirely. In order to facilitate price transparency in the corporate bond market, NASD issued a regulation that requires all broker/dealers to report transactions in corporate bonds to TRACE (an acronym that refers to the NASD Trade Reporting and Compliance Engine) within 15 minutes of the transaction. NASD publishes these data daily. Finally, in the taxable bond area, mortgage-backed securities are mathematically complex; noncurrent coupons can be very difficult to price.

The market for municipal bonds is even more opaque than in the market for corporate bonds. There are several reasons for this, not the least of which is that there are millions of small issues outstanding. Many get sold at the underwriting, are put away, and rarely trade afterwards. Another reason is that bid/offer spreads in municipals tend to be extremely wide by the standards of the rest of the bond markets, a fact that the market-making community is not especially eager to advertise.

Information about individual municipal bond issues is on file with each of the nationally recognized municipal securities information repositories (NRMSIRs). These data include information about issuer financial condition, operating results, and defaults. Some data services, in particular Standard and Poor's and FT Interactive, sell end-of-day prices to market users, including mutual funds who use the prices to mark positions to market for determination of daily net asset values (NAV).

Marking some OTC derivatives to market can be especially problematic. Plain-vanilla derivatives like rate and currency swaps can be benchmarked rather easily off exchange-listed products like Eurodollars, published rates by the International Swap Dealers Association (ISDA), and active interdealer broker markets. Other OTC derivatives like options on cash bonds can be a bit more difficult, especially for longer-dated strikes. Exotic options and structured products are more opaque and present more of a challenge.

MARKING POSITIONS

Exotic derivatives, collateralized mortgage obligations (CMOs), and mortgage securities often wind up being "marked-to-model" instead of to market, because active trading markets are not always readily available for the particular security that needs to be valued. In fact, some of the residuals (the leftover bits from repackaging mortgage pools into different slices or "tranches") used to be referred to on the Street as "toxic waste," which ought to give pause for thought when the subject of valuations comes up.

Marking-to-model is unavoidable when active, liquid markets are not readily available. But successfully marking-to-model requires having the right valuation model to begin with, and checking it against other sources. Countless disasters have resulted from positions being marked incorrectly, either through malfeasance or model error. Incorrectly marked positions do not merely hide losses or profits; they give a false picture of the firm's true risk posture and distort decision making. It is also worthy of note that mismarked positions are rarely brimming with undiscovered profits. More likely than not, they disguise losses. In the end, proper risk management depends on having the right people as well as good systems. Systems can always be subverted by people with a mind to do so. Character matters.

LIQUIDITY RISK

Liquidity generally refers to ease of transacting, a crucial factor to consider for managing position risk. Operationally it is often defined in terms of bid/offer spreads, the depth of standing bids and offers, and funding availability. These are related. For analytic purposes it is convenient to think of two types of liquidity. The first is security specific. The second is at the system level and can be thought of as market liquidity. These too are related.

Liquidity varies by security type, market sector, time zone, and funding availability, among other factors. Securities that are less risky are typically more liquid. U.S. government securities are a case in point. Hundreds of billions trade on a daily basis. By far, they are the safest securities in the world, and there are typically many willing buyers and sellers at every price increment. Stocks are far riskier and trade in far smaller quantities. To put it in perspective, for the first half of 2006 the NYSE traded about $71 billion worth of stocks on the average day. During the same time period, according to the New York Fed, average daily volume in government securities was about $548 billion, almost eight times the volume of stocks.

In a recently published paper entitled "Market Liquidity and Funding Liquidity," economists Markus K. Brunnermeiser (Princeton) and Lasse Heje Pederson (NYU) illustrated the importance of the mutually reinforcing relationship between the ease of trading and funding availability.[1] On a day-to-day basis, traders provide market liquidity. They do so with borrowed money, which is readily available when things are going well. But when markets turn sour, lenders may ask liquidity-providing dealers to pony up more in margin against their positions, thus reducing market liquidity just when it is most needed. Exchanges also typically raise their margin requirements in times of crisis, further reducing liquidity.

The link between funding availability and market liquidity points to an inherent structural fragility. In the event of financial turmoil, liquidity can dry up at a moment's notice if lenders pull their credit lines to dealers. Partly because the major dealers are active in most (if not all) of the major (and not so major) markets, a liquidity crisis in one market can easily spill over to others—the so-called contagion effect commonly seen in FX trading. Further, the often-observed flight to quality during times of financial stress can exacerbate the problem, as does the fact that volatility generally moves opposite the market. That is, volatility tends to rise when markets fall and vice versa.

THE FLIGHT TO QUALITY

During times of financial stress the first market reaction is a flight to quality. As a practical matter, that means the scramble to buy U.S. Treasuries commences. Calling it a flight to quality doesn't quite capture the emerging bedlam though. As traders race to buy Treasuries, prices rise. And as Treasury bond prices rise, Treasury futures rise as well. At the same time that traders are buying Treasuries, they are unceremoniously bailing out of other, riskier securities, thereby putting downward pressure on those prices. Quality spreads zoom out. This causes horrific trading losses in the arbitrage community because the arbs typically go long lesser-quality credits against going short Treasuries waiting for spreads to narrow in convergence trades.

The process begins to feed on itself, and, before long, credit lines are under review or are being pulled, forcing dealers to redouble efforts to liquefy, thus exacerbating the situation. The search for bids in a falling market adds to volatility, increasing nervousness in the market. At around this point firms begin to be especially careful about the creditworthiness of their counterparties. To boot, clearinghouses may increase margin requirements on positions. Dealers start to hang onto their collateral, the delivery system can back up in a series of round robins, and before long the system can freeze up. That's when the Fed steps in.

THE 1987 STOCK MARKET CRASH

In October of 1987 as the stock market began to crash, it seemed like everyone was trying to bail out at once. In fact specialists on the floor of the NYSE were selling faster than they were buying. (So much for specialists stepping in to maintain orderly markets.) When the Fed saw that the system was beginning to freeze up, the FOMC announced that the Fed would make credit available to the system as needed—in effect acting as lender of last resort. Government bond prices began to soar.

But that presented another problem. As government bond prices skyrocketed in the flight to quality, it became increasingly difficult to make delivery of the current 10-year T-note. Arbitrageurs had amassed huge short positions in the notes against futures contracts, making them virtually impossible to borrow for delivery purposes. The dealer community began to call on the Fed to lend the 10-year notes it had in its portfolio to break up the logjam. Customarily, the Fed permitted dealers to borrow Treasuries from its own portfolio (albeit at penalty rates) to plug short-term timing gaps while waiting for a delivery from another source.

As the 1987 panic began to spread, credit lines were being cut; at the same time short positions and fails-to-deliver piled up all over Wall Street. The Fed decided to restore liquidity by figuratively taking the banks to lunch. Senior officials at the Fed called on the senior managements of the New York City banks and "asked" them to make additional credit available to the Wall Street trading houses. When the Fed "asks" banks to do something, the request is made in the same spirit as when a fifth grade teacher "asks" a student to go to the principal's office. There is not an abundance of give and take.

The Fed then turned its attention to breaking up the logjam in the Treasury market by taking aggressive steps to lend out the much sought after 10-year T-notes it held in its portfolio. Senior officials at the Fed began to call around the Street to the government bond trading houses, seeking borrowers for the 10-year T-notes. Strictly speaking, this wasn't the way it was supposed to happen. Normally, a senior trader from one of the trading houses would make a request of the Fed to borrow securities for a short time, whereupon the Fed would ask if he (the trader) was expecting the notes to arrive from another source before too long. At this point the trader would say yes and the Fed would approve the borrowing request.

On this particular day though, with the financial system tottering on the brink, the Street's senior traders were pretty much all on the phone, furiously trying to manage their positions. That's about the time the Fed began calling about the 10-year notes. When the Fed's trading desk wants to talk to a dealer firm to get a bead on the market, it calls on a direct line. And when the Fed calls, it gets picked up right away. You just don't leave the nation's central bank on hold. So when the direct line from the Fed rang at one particularly large dealer firm, it was picked up by a very junior trader—a kid. In fact it was only his second week on the trading desk. But he knew that the Fed line was supposed to get picked up pronto, bedlam or no. And he was assigned to the RP desk, specifically tasked with borrowing notes and bonds. So by the time the Fed called, he was kind of morose, fretting over his inability to borrow 10-year notes to cover the firm's $250 million short position and wondering what his exit interview was going to be like.

Needless to say, the Fed knew that the firm was short the 10-year notes because the Fed was monitoring the situation closely and had previously asked the dealers to report their positions. And to boot, when the Fed called this particular firm on the direct line, it wasn't just the Fed. It was a very senior fed official (VSFO); a name that would have been recognized instantly on the Street. So when this VSFO called to inquire of the firm if it was short and needed to borrow 10-year notes, he undoubtedly expected to go through the formalities quickly, agree to lend the T-notes, get a market update, and carry on calling other big firms.

Unfortunately, "the kid" picked up phone. The conversation, as recounted to the author, went something like this.

The Kid: *A little tentatively*. Hello.

VSFO: Good morning, this is so and so; are you looking to borrow 10-year Treasury notes?

The Kid: *A bit more firmly*. Yes, we are.

VSFO: How many would you like to borrow?

The Kid: *Brightly now*, $250 million.

VSFO: *Ritualistically*, Are you expecting delivery of $250 million notes from somewhere else?

The Kid: No. That's why I want to borrow them from you.

VSFO: *After a brief silence, with an edge to his voice*. The correct answer to this question is yes. Now we are going to try it one more time. Are you expecting delivery of $250 million notes from somewhere else?

The Kid: *Clearing his throat*, Yes.

VSFO: Very good answer. Now please put your head trader on the phone.

The notes were delivered in short order. The Fed flooded the system with liquidity, and the system pulled through.

LEVERAGE AND LIQUIDITY

There have been numerous books, articles, and studies of the 1987 crash that enumerate its allegedly distinctive features. The role of portfolio insurance, the huge risk arbitrage positions on the Street, the breakdown of several LBO deals all come to mind. More to the point, there are the similarities between the 1987 crash and other, smaller ones. Two in particular stand out. The first is that they inevitably involve the use of too much leverage. The second is the assumption that markets will always be free-flowing and continuous. They won't be.

The list of disasters from overuse of leverage and its ripple effects is impressive. The collapse of Drysdale Securities in the early 1980s cost hundreds of millions of dollars and required Fed intervention. The Orange County default in 1994 was the direct result of leveraging up what should have been run-of-the-mill money market investments. The Asian currency crisis, followed by the 1998 Russian default, signaled the doom of the highly levered hedge fund Long Term Capital Management, complete with a Fed orchestrated quasi-bailout. The current record holder is Amaranth Advisors, a hedge fund that reportedly lost more than $6 billion in a week and proceeded to go out of business after losing on highly leveraged natural gas positions. Investors, according to *The New York Times,* "were baffled by the apparent lack of risk management at the fund."[2]

The particulars of any given incident may vary a bit (or a lot) from previous ones. But liquidity dry-ups and market crashes are inevitably accompanied by leverage, too much of it, when the music stops. They are typically the result of the deleveraging that accompanies tightening of monetary policy. Tightening policy does not make money unavailable directly; it makes money more costly, and fewer deals get done. Liquidity dry-ups are more likely to emerge when lenders step back over credit concerns. When policy is easy and things are going well, credit never seems to be a problem. But when policy tightens, problems hidden by easy money suddenly begin to emerge. And if there is a sudden shock to a system that has a lot of leverage embedded in it, the shock is hard to absorb and liquidity flies out the window as people run for cover. It is an unedifying spectacle, kind of like watching people flee a burning building.

The strategy to avoid getting caught in a liquidity crunch is pretty obvious. Don't get overleveraged in the first place. That is easier said than done. Markets feel pretty good at the top, and it usually seems like nothing can go wrong. But go wrong they can and will. One way to keep things in perspective is to stress-test positions using various scenarios of what might go wrong and to what extent. The fundamental method most risk managers use for doing this is the statistical technique mentioned earlier, value at risk (VaR).

VALUE AT RISK

The idea behind risk modeling is that, unlike uncertainty, risk is quantifiable. A commonly used risk metric in the trading business, *value at risk,* seeks to quantify risk in terms of maximum loss from adverse market changes, based on past history. Loss potential is measured in terms of price volatility. The measure most frequently used for volatility is the standard deviation of daily returns, calculated over a period of several days. The model can be adapted for estimating 5-day, 10-day, 22-day risk, or for any time frame of average annualized volatility, the conventional measure of risk.

The initial assumption made by VaR models is that returns are log-normally distributed. That convenient assumption allows the statistical tools of the economist to be deployed to estimate a plausible range of how much the price of a particular stock, bond, index, or commodity is likely to move within a given time frame. That range of possibilities is typically estimated using the standard deviation of daily logged returns for the time period of interest. The standard deviation is then used to calculate confidence intervals for the extent of risk exposure.

In a normally shaped distribution, approximately 68% of all occurrences fall within 1 standard deviation of the mean, 95% within 2 standard deviations, and 99.7% within 3 standard deviations. Consequently, once the mean and standard deviation of a distribution are known, it is a relatively simple matter to estimate the probability that an event will happen— if the past is a predictor of the future and if returns are normally distributed.

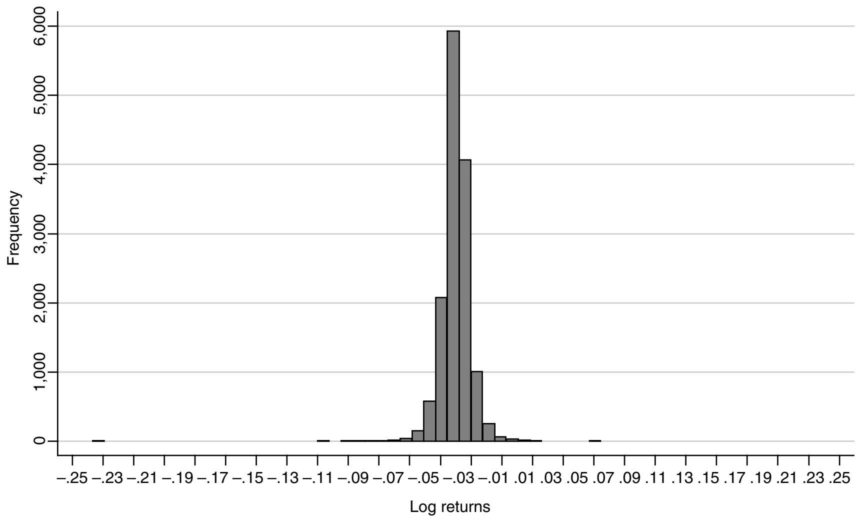
For instance, consider a $100,000 position in the S&P 500. Assume the annual mean return for the S&P 500 is 10% with a standard deviation of 15%. The expected return (omitting dividends) is 10% or $10,000 over a one-year period. However there is a 68% chance that the return for a given year will fall within 1 standard deviation of the mean, or –5% on the downside and +25% on the upside, which in dollar terms falls between a loss of $5,000 and a gain of $25,000. Similarly, there is a 95% chance that the return will fall between –20% and +30% (2 standard deviations), and there is about a quarter of a percent chance that the return will fall between –35% and +55%.

Unfortunately, this oversimplifies things by more than a bit. For one, the world is not quite as tidy as all that. Returns in financial markets are not normally distributed; they are actually fat-tailed. The extremes are far wider than the normal curve would suggest, meaning that high-impact but low-probability events are actually far more likely to occur than conventional models suggest. In addition, the distribution is typically skewed to the left, implying a tilt to the downside. Big surprises are likely to be bad ones.

For example, consider some simple descriptive statistics and a graph of the daily logged price returns of the S&P 500. See Figure 24.1. The data, which encompass the time period from 1950 through mid 2006, illustrate the tendency for changes in stock prices to skew to the left and reach far beyond what the normal distribution would in fact suggest is normal. The skew of the distribution is –1.32 rather than zero. The kurtosis of the distribution, which measures how peaked it is, comes in at 38. The kurtosis

FIGURE 24.1

Frequency Distribution S&P 500 Daily Log Returns (January 1950–June 2006)



Data source: Yahoo! Finance

of a normal distribution is 3. A kurtosis measuring more than 12 times what a normal Gaussian distribution would produce indicates a fat-tailed and highly peaked curve, which is precisely what the graph displays.

Most conventional VaR models have implicitly embedded in them the assumption that securities returns are distributed normally, an assumption that Figure 24.1 demonstrates is manifestly incorrect. Historical experience that is markedly at odds with fundamental assumptions built into a risk model ought to be disquieting. As a practical matter, the fact that securities returns have historically been fat-tailed (as in Figure 24.1) indicates that the probability of a high-impact event is far greater than a conventional risk model would typically suggest. And since the distribution is skewed to the left, high-impact events are more likely to be negative than positive. All of which implies that risk positions are likely to be more risky than they are acknowledged to be. As it happens, this is consistent with findings from behavioral finance.

There is a human tendency to discount too deeply the probability of (negative) high-impact or catastrophic events. (It isn't going to happen, and if it does, it isn't going to happen to me.) Modelers (who should know better) are not immune to the phenomenon. It is very likely related to the psychological bias toward overoptimism that pervades the world of finance. You would be hard pressed to find a group more self-confident than traders. Shortly after Tom Wolf skewered bond traders (in particular), sardonically referring to them in *Bonfire of the Vanities* as “Masters of the Universe,” bond traders commonly began to refer to themselves as masters of the universe without a hint of irony.

Although most traders acknowledge the possibility of a high-impact event in the abstract, it is not clear that many take it seriously enough to adopt serious countermeasures. Many act as day traders on the theory that going home flat limits the possibility of getting caught off base. Which it does, but catastrophes don’t always conveniently hit after the close. The result is that when disaster strikes, many traders are needlessly caught flat-footed. They and their fellow traders will all start heading for the exit doors at once, but the doors will be too small to accommodate more than a few. To boot, there is also a bit of moral hazard built into the system. The too-big-to-fail syndrome leads to risk-shifting and overleveraging. Too much leverage, overconfidence, and overdiscounting of improbable events is a recipe for trading disaster. It is a pity because there are some relatively low-cost and interesting ways to manage it. These involve understanding volatility, the judicious use of deep out-of-the-money options, and the CBOE VIX volatility contract (discussed below).

VOLATILITY AND VOLATILITY CLUSTERING

Financial market volatility can be thought of as a measure of price variability over relatively short periods of time. How short depends on the subjective perspective of the analyst. For some purposes the spread between daily high and low prices is a good measure. Other popular measures include average daily, weekly, monthly, and quarterly logged returns. Recently, as real-time data have become increasingly available, realized volatility over very short intervals (five minutes and shorter) has become a focus of inquiry. For detailed discussions of volatility theory and modeling techniques, readers are advised to consult Ser-Huang Poon’s *Practical Guide to Forecasting Financial Market Volatility* and Paul Wilmott’s *Introduction to Quantitative Finance*.[3]

It is important to differentiate between long-term returns and short-term price swings (volatility). Over the very long run the stock market has generated total nominal returns in the neighborhood of 11%, with an annual standard deviation of around 20%. This trend dominates over the very long run. But in the short term, volatility does. So in order to compare risk with returns, volatility (the measure of risk) is usually annualized.

There are two ways to approach volatility calculations. One is implied from options prices. The idea is that when two of the three variables of the Black-Scholes model are known (the risk-free rate and time to expiration), the third (volatility) can be inferred. The problem is that when this method is used, all options on the same security ought to have the same implied volatility. Typically they don't.

Another way to approach volatility is to build a forecast model based on historical data. A common method for doing this is to define volatility as the annualized standard deviation of returns for a time step (t) over a sample period T. Calculating this is a two-step process. First, calculate the standard deviation of returns for the chosen time step using Excel or any standard statistical package. Second, multiply the result by the square root of the number of time steps in a year. The result is annualized volatility for the desired time step, be it days, weeks, or months. The method is expressed formulaically as:

$$\text{Annual Volatility} = \sigma \text{ Returns}_{(t)} \sqrt{T}$$

For instance, suppose the daily volatility $_{(t)}$ of a stock is being investigated. After calculating the standard deviation of daily returns for a sample period (T), the result would be multiplied by $\sqrt{252}$, since 252 is usually used for the number of trading days in a typical calendar year. For weekly volatility the standard deviation of weekly logged returns would be multiplied by $\sqrt{52}$; for monthly volatility, the standard deviation of monthly returns would be multiplied by $\sqrt{12}$, and so on. For example, if the standard deviation of daily logged returns during 2003 to 2005 equaled 0.0089, annualized daily volatility for that sample period would be $0.0089 * \sqrt{252}=14.1\%$.

There are a number of problems associated with this approach, however. For one, since the calculation of the standard deviation of returns is based on historical data, the implicit assumption of the forecast model is that future volatility will look like past volatility. There is little empirical evidence to suggest that this assumption is likely to hold except in the most general way. Secondly, the choice of time frame has an important influence on the result. Daily, weekly, monthly, and quarterly volatility are liable to have different structures. Longer periods allow for greater diffusion of returns and hence volatility. But using closing prices of longer time periods leaves a lot of potentially valuable information unused. For instance, using weekly closing prices as data points necessarily misses large day-to-day and intraday price swings. The best time step to use (days, weeks, or months) really depends on the time horizon of interest and what the forecast model is to be used for.

To get a sense of volatility structure, some analysts use historical data to construct volatility cones (as shown in Figure 24.2) that display both the range and stability of annualized volatility over different time horizons, treating the time horizons as if they were option expirations.

Another technique is to calculate and plot rolling averages of volatility for different time horizons, as shown in Figure 24.3. This technique provides a way of observing how volatility unfolds over time in the marketplace.

FIGURE 24.2

SPY Volatility Cone (January 3, 1996–August 18, 2006)

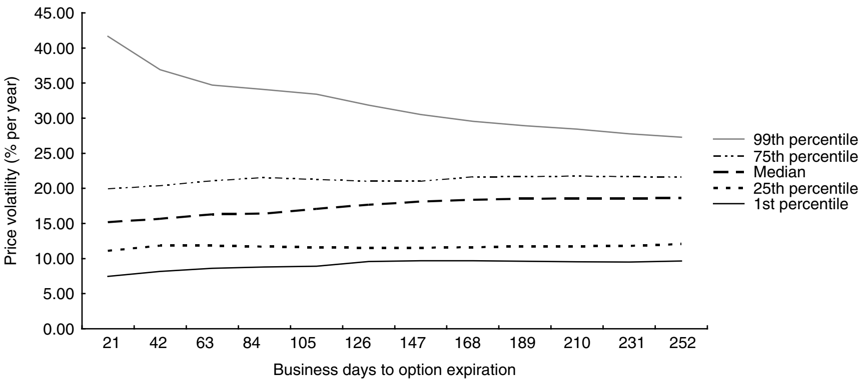
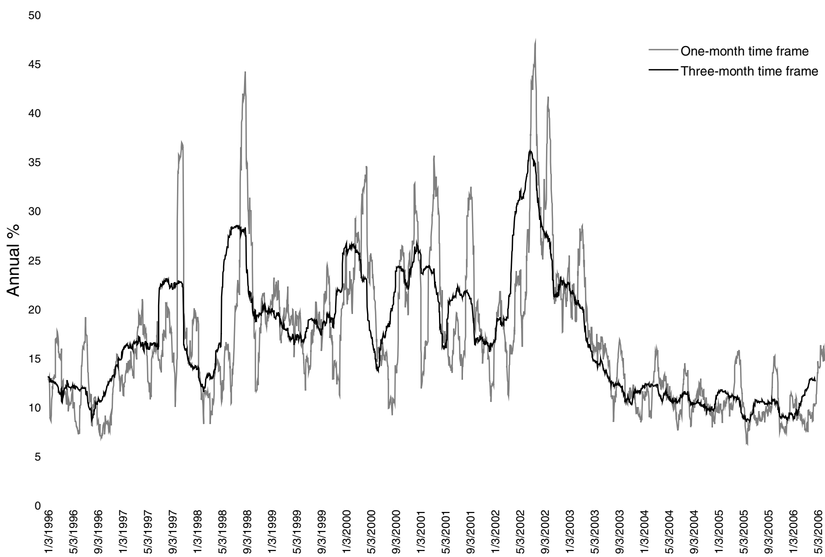


FIGURE 24.3

Average Daily SPY Volatility by Time Horizon



As Figure 24.3 suggests, one reason why volatility modeling is so vexing is that it is not stable. Volatility begets volatility, which is to say that it tends to cluster. And since risk (or volatility) is the flip side of reward, high volatility tends to suppress stock prices. Investors demand a greater discount (potential reward) as compensation for the greater risk produced by high volatility. That seems to have been empirically borne out over time, with big down moves in stock prices accompanied by increasing volatility. Both variability and clustering can be seen in Figure 24.4, a graph of S&P 500 daily logged returns from 1950 through June 2006. Note that the graph is not centered around zero. The upper boundary is +0.15%; the lower boundary at –0.25%. This accommodates the asymmetry of daily returns, heightened by the inclusion of the October 1987 crash.

TRADING AND HEDGING VOLATILITY

Thanks to the Chicago Board Options Exchange VIX contract, volatility can be traded easily. CBOE introduced the CBOE Volatility Index in 1993, and it soon became an industry benchmark. Originally based on the S&P 100, it was later updated to reflect the volatility of the S&P 500. A VIX futures contract was launched during the fourth quarter of 2004.

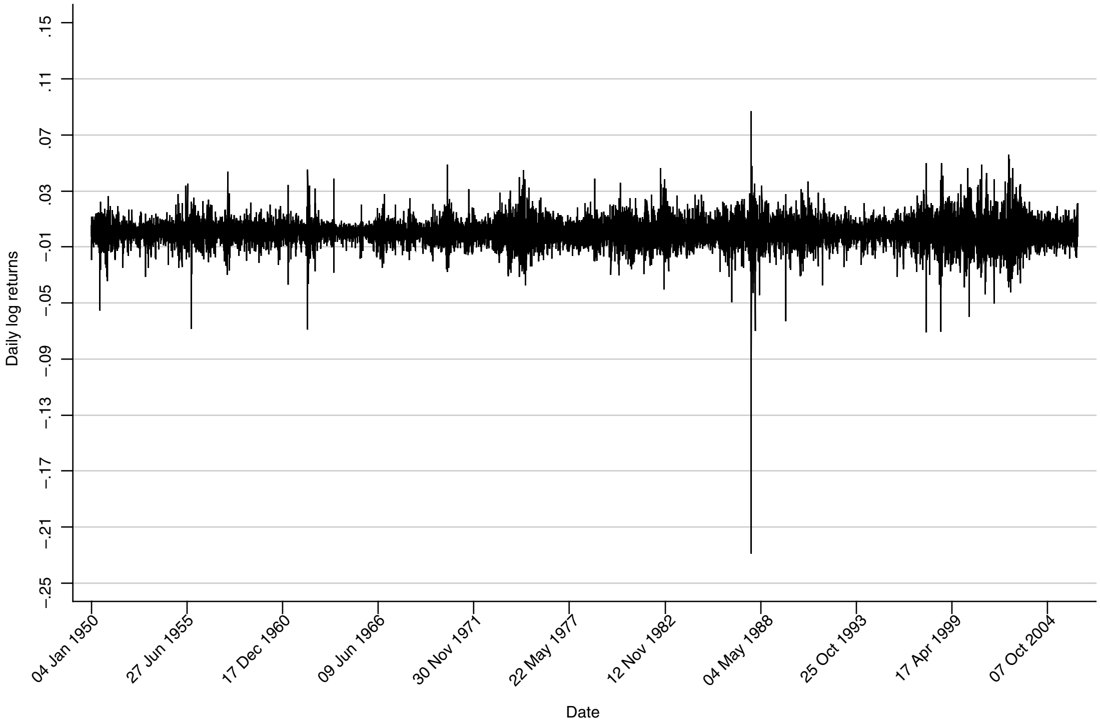
The VIX measures near-term (30-day) expected stock market volatility. Since options prices have expected volatility embedded in them, the VIX uses a formula that extracts expected volatility of the S&P 500 using prices of a weighted strip of S&P index options calculated in real time. The result is a volatility measure expressed in index points: VIX $=100 \times \sigma$. The derivation of the formula and a step-by-step guide to its calculation can be found online at www.cboe.com/micro/vix/vixwhite.pdf.

The VIX has been referred to as the *investor fear gauge* because it has tended to hit its peaks during times of financial stress. Then after panic subsides, the index tends to fall back. This tendency is on display in Figure 24.5, which graphs the interplay between volatility and market direction from 1990 through mid-2006. The stock market is represented by the log of the Dow Jones Industrials. The graph shows that during times of sharp market sell-offs, volatility (as represented by the VIX Index) tends to rise, with particularly dramatic upward spikes during times of market turmoil. For instance, note the timing of large volatility spikes: first in 1990 at the beginning of the first Iraq war; a brief spike in early 1994 when the Fed began to tighten monetary policy; the 1997–1998 Asian financial crisis, Russian default, and the Long Term Capital Management collapse; the September 11 terrorist attacks; and the lows of the stock market in the 2002 sell-off.

The VIX is now used as a market-timing device, as a tool for macrohedging, and as a speculative vehicle. It is used as a market-timing tool because stock market returns have been shown to be significantly higher than is usual 30 days after volatility spikes. That may well be because volatility spikes often arrive during sharp sell-offs. When the dust settles 30 days later, prices have bounced back in mean-reverting fashion producing short-term superior returns.

FIGURE 24.4

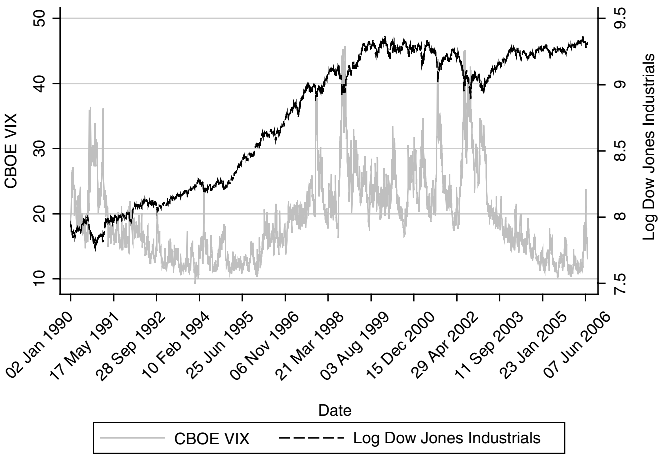
Volatility and Clustering in the S&P 500 (January 1950–June 2006)



Data source: Yahoo! Finance

FIGURE 24.5

Volatility and Market Direction (January 1990–June 2006)



Data sources: CBOE and Yahoo! Finance

That may partly explain the usefulness of the VIX as a macrohedge. Volatility has a pronounced tendency to rise during times of market turmoil. A long position in volatility can sometimes help offset damage to stock positions held on the long side of the market during periods of market stress. One thing that makes this a particularly attractive strategy is that the VIX is skewed to the right, unlike the stock market, where the skew is to the left. Finally, the VIX can serve as a speculative vehicle for a pure play on volatility, absent a view on market direction.

EVENT AND CREDIT RISK

Two other sources of risk that bear mentioning are event and credit risk. Event risk constitutes such things as political risk, takeover risk, and weather risk. Credit risk has taken on increasingly complicated forms with the emergence of leveraged buyouts, prepackaged bankruptcies, and class action lawsuits that can bring a company, and sometimes an entire industry, to its knees.

As the name implies, event risk is difficult to foresee. However, new and emerging markets are developing that allow certain types of event risk to be priced. In the wake of hurricane Katrina there has been increased interest in the issuance of catastrophe bonds. Environmental and regulatory risk with respect to greenhouse gases and other pollutants can be hedged at the Chicago Climate Futures Exchange (CCFX). The CCFX now trades standardized futures contracts on emission allowances and other environment products. In addition, the European Climate Exchange (ECX), a wholly owned subsidiary of the CCFX, manages sales and marketing for European environmental products.

There have been fledgling attempts to create markets for terrorist attacks. And there is the Iowa futures market in presidential elections. The political risk consulting business is picking up as globalization requires informed decisions about the political environment before commitments are made to what may be very illiquid markets. One of the more important measures of political risk in this regard may have to do with country corruption, or lack of it, as the case may be. It is worth noting that Russia's default in 1998 set off the chain of events that brought down Long Term Capital Management.

An extremely valuable source of information about country corruption and (by implication) political risk, is Transparency International. A nonprofit, nongovernmental organization (NGO) dedicated to fighting corruption, Transparency International publishes a series of governance and corruption indicators, including a Corruption Perceptions Index (inception 1995), a global corruption barometer, a bribe payers index, and other regional and national surveys and indexes. Beyond the more exotic types of event risk, one of the more difficult types to manage is, paradoxically, the danger of having good credit.

Merton Miller of the University of Chicago and Franco Modigliani of MIT showed that the economic value of a firm is independent of the way it is capitalized, and they won the Nobel Prize in economic science for their efforts.[4] According to Miller-Modigliani, a firm can finance growth by issuing equity or debt; intrinsically it makes no difference. But firms in the United States have had a tendency to borrow by selling bonds (rather than raise equity capital by selling stock) because of the tax code. Corporations can charge bond interest payments (but not dividends paid) against earnings. To add insult to injury, dividends get taxed twice: first at the corporate level as earnings and second as income to stockholders.

This was not lost on LBO and private equity firms that reasoned that they could easily take over conservatively capitalized firms that had little debt and lots of shareholder equity. First, the target firm's conservative capital structure is unlocked to free up cash that can be used to buy the firm's assets. The takeover firm pledges the target firm's own assets as collateral against the issuance of low-grade high-yield bonds. Next the acquirer uses proceeds of the bond sale to buy out the shareholders. Interest on the bonds is tax deductible, reducing the after-tax cost of capital for the acquisition.

The upshot of the transaction is an upward spike in the stock of the takeover target, but a plunge in the company's outstanding bonds. These asset price changes reflect a shift in capital structure, not a fundamental shift in the business. As a possible risk management solution, some high-grade bond portfolio managers took to buying small positions in the stocks of the companies whose high-grade bonds they owned. In the event the company became the target of an LBO, the gain on the stock could more than offset the loss on the bonds.

LBOs are a special, but important, case. Other forms of credit risk can come from changes in the business cycle, poor management, a change in the competitive landscape, or a host of other reasons. The rating agencies, primarily Moody's and Standard & Poor's and to a lesser degree Fitch, assign ratings to bonds based on these and other considerations. But ratings, which take the long view, are sometimes lagging indicators. The market tends to anticipate upgrades and downgrades by changes in quality spreads. Yield spreads between corporate bonds and full-faith-and-credit Treasuries, which generally reflect quality differentials, adjust as credit conditions change.

There are other mechanisms for pricing and trading credit quality differentials. In addition to yield spreads, new mechanisms have developed for managing credit risk. A particularly important one that has developed over the last several years is the market for credit default swaps.

MANAGING CREDIT RISK WITH CREDIT DEFAULT SWAPS

Credit default swaps are derivative instruments used to price and transfer default risk from one party to another.[5] Normally bonds of inferior credit quality trade at a discount to better-quality bonds. As quality differentials increase, so does the discount. Corporate bonds, for instance, trade at a discount to governments because a corporation may default. Governments are presumed to be free of default risk when borrowing in their own currency.

The yield spread between a corporate bond and a government bond of the same coupon and maturity is mostly determined by credit quality. Other factors like liquidity can come into play, but by far the main driver of the yield spread is credit quality. However, institutional features of the market may make it difficult to price corporate bonds efficiently through arbitrage. Corporate bonds may be difficult to sell short, so it can be difficult to trade the spread from the short side. The bonds may be hard to borrow. Issue sizes are typically small compared to governments, and the relative lack of liquidity and transparency in the corporate market makes transaction costs high. Partly for these reasons, the market for credit default swaps developed.

The basic structure of a credit default swap is straightforward. For a price, two parties agree to transfer the default risk of a third party from one to the other. One side of the transaction, the investor, assumes default risk in return for a fee. That party is known as the protection seller. The other side, often a bank or a dealer, pays the fee to be rid of it. In that case the dealer is known as the protection buyer. If nothing happens, the investor collects his fee and has no further obligation. However, in the event of a default, the dealer delivers the defaulted asset to the investor. The investor then pays the dealer the remaining interest and the par value of the defaulted asset. In effect, the default risk seller owns a contingent put, exercisable upon default.

Suppose, for instance, a holder of a five-year corporate bond wanted to buy two years' worth of credit insurance on the bonds by use of a credit default swap. In this case the corporate bond would serve as the reference asset. The bondholder would pay the other side a fee to assume two years' worth of default risk. In the event that no default occurred, the holder of the bonds would receive interest payments from the bond's issuer. If, however, a default occurred within the two-year period of the contract, the insurance purchaser would "put" the bonds to the swap seller (investor). The swap seller would be obliged to pay par for the bonds and make good on the missed interest payment.

This basic structure can be modified in any number of ways, depending on precisely what the counterparties are willing to agree to and transact on. For instance, negotiable variables could include the length of time the swap is outstanding, what constitutes a default or otherwise triggers a payment requirement, what the payment price will be for the bond in the event of default, what constitutes good delivery, whether the payoff will be determined by cash settlement instead of delivery, what the fee is, and over what time span it will be paid.

Other considerations include the creditworthiness of the contracting counterparties themselves, the degree of credit diversification, as well as clearance and settlement systems. Obviously, it is important that the credit guarantor be creditworthy. It is less obvious, but also true, that credit protection should be bought from a firm whose credit is not highly correlated with the credit being insured. If the creditworthiness of two firms is highly correlated, then the quality of the insurance protection may be significantly degraded.

OPERATIONAL RISK

It is unfortunately the case that money made trading (and then some) can be lost in the back office. Credit derivatives are a case in point. The explosive initial growth in credit derivatives trading was not matched by a buildup of a back-office infrastructure needed to document trades and clearly monitor positions. On this score, Thomas Huertas of the British Financial Services

Authority sounded the alarm in a speech he gave at the annual Rhombus Research Conference in April 2006. He pointed out that estimates for the notional value of credit derivatives outstanding approached $17 trillion and that as of mid-1995, large banks were taking 44 days on average to confirm plain-vanilla credit derivative trades and more than two times as long for complicated transactions. He also noted that credit derivatives were being reassigned without proper notification, rendering problematic counterparty risk assumptions made at the time of the initial transaction.[6]

The Counterparty Risk Management Group, an industry consortium, released a report on the situation and has successfully prodded the industry both to clean up backlogs in trade confirmations and to move forward to bring trade documentation and confirmation processes up to speed.[7] Still, there remains a good deal of work to do, and it is extremely important for firms to get their trades cleared and confirmed properly, especially when the point of the trade is credit risk transfer.

SUMMARY

In practice, financial market risk has come to be defined in terms of the width of the distribution of possible outcomes. The standard deviation of returns, a measure of returns dispersion, is used to compute volatility, the preferred operational measure of risk in the marketplace. For the sake of consistency and comparability, volatility for different time horizons is annualized. Risk management has come a long way in a short period of time. Seat of the pants is out; probability and statistics are in.

There are many sources of risk lurking in the markets, and they need to be recognized and dealt with. That is because only half the game is making trading profits; the other half is holding onto them. Lack of contingency plans for unexpected events is asking for trouble. Among the risks that need to be taken into account are (inaccurate) marks-to-market and overdependence on marks-to-model. Models have to be maintained, updated, and specified properly. And their limitations need to be respected.

Contingency plans should acknowledge worst case possibilities, a sudden drying up of market liquidity being one of them. One problem with many risk management models is that they assume continuous and liquid markets. But liquidity has a bad habit of disappearing just when it's needed most. One of the worst consequences of evaporating liquidity is that it forces people to dispose of positions at bargain basement prices because of their inability to ride out the storm. The way to avoid this—and better yet be able to take advantage of the situation—is to avoid getting overleveraged to begin with. That is easier said than done.

Value-at-risk models can provide a sense of how much risk exposure there is in a position under most, but not all, circumstances. Similarly, measuring volatility can put risk assumptions into perspective and provide a valuable tool for trading off risk against expected returns. The VIX, listed at CBOE, allows volatility to be traded easily for both speculative and risk management purposes. Credit and counterparty exposure can be managed by exercising due diligence and by using OTC credit derivatives, but it is important to make sure that credit risks are diversified and that operational controls are sufficiently developed so that trading profits are not held hostage to the mercies of lax monitoring.

NOTES

[1] Markus K. Brunnermeier and Lasse Heje Pedersen, "Market Liquidity and Funding Liquidity," March 2005. AFA 2006 Boston meetings paper available at SSRN: http://ssrn.com/abstract=688065.

[2] Gretchen Morgenstern and Jenny Anderson, "Hedge Fund Shifts to Salvage Mode," *New York Times,* September 20, 2006, C1.

[3] Ser-Huang Poon, *A Practical Guide to Forecasting Financial Market Volatility*, John Wiley & Sons, 2001; and Paul Wilmott, *Paul Wilmott Introduces Quantitative Finance*, John Wiley & Sons, 2001, especially Chapter 6.

[4] Merton Miller and Franco Modigliani, "Dividend Policy, Growth and the Valuation of Shares," *Journal of Business,* 34, Oct. 1961, pp. 235–264.

[5] A good reference work written by Janet Tavokoli is, *Credit Derivatives and Synthetic Structures: A Guide to Instruments and Applications,* 2nd Edition, Wiley Finance, 2001.

[6] Thomas Huerta, "Credit Derivatives: A Boon to Mankind or an Accident Waiting to Happen?" Speech by Thomas Huertas, Rhombus Research Annual Conference, London, April 26, 2006.

[7] The report is available at http://www.crmpolicygroup.org/

CHAPTER 25

Summary and Conclusion

> Never pay the slightest attention to what a company president ever says about his stock.
>
> *–Bernard Baruch*

What predicts success in trading? This book implicitly argues that success results from the interplay of three principal factors and one that hides in the background. The first factor is analytic. It is the process of collecting, analyzing, and interpreting market information. The second factor is action. It is the ability to formulate and execute strategy based on analytic insight. The third factor is position management. This requires recognizing and minimizing potential risk from various sources. These include external sources of risk (unexpected market developments) as well as internal ones (e.g., psychological overconfidence). Progress on all these fronts can (and should be) subjected to rigorous assessment and periodic review. But in the end it is not all science and the application of quantitative methods. There is an additional factor not so easily quantified. There is an art to trading.

INSIGHT AND INTERPRETATION

The first factor, analysis, seems reasonably straightforward. At least it does until you think about the immense task of filtering out background noise in order to focus on the information flow that really matters. The age of information overload means that data filtering may be just as important as data collection. And how do you determine what information is important? It is information that is likely to cause a market reaction. And what type of information does that? At the microlevel it is information that pertains to a particular security or company. Examples would be earnings reports or credit downgrades. At the macrolevel it is data that could affect the economy or a sector. Examples would be employment and inflation data.

It is easy to draw too fine a distinction between micro- and macrolevel data. Microlevel data may be a precursor of future events and can bleed into the macropicture. A drop in a homebuilder's reported sales may be due to losing business to a competitor or it may be due to a general economic slowdown. As more homebuilders report earnings, the microlevel picture can aggregate into a bigger picture, with larger implications for the state of the economy. Similarly, macrodata (for instance inflation reports) are important not only because they can change perceptions of traders, but because the data may also cause policy makers to act, thereby altering the fundamentals on the ground.

A key skill for traders is the ability to sift through and analyze data for clues to the market's likely reaction. This book emphasizes the importance of data interpretation over fact collection because raw facts are meaningless in themselves. Context is crucial. Is a reported annual inflation rate of 4% high or low? It depends. By the standards of 2006, it would be cause for alarm and aggressive Fed tightening. By the standards of 1980 it would be a godsend. Is a stock market multiple of 14 high or low? It depends among other things on the level of interest rates and the business cycle. Do budget deficits matter, or matter much? It depends to some degree on whether the deficits are cyclical or structural.

The difference between cyclical and structural change is a major determinant of market behavior. Unfortunately, it can be difficult enough to gauge the length, breadth, and depth of a normal business cycle. Recognizing structural change is more difficult still. But both of these things can be done. It requires a wide-ranging intellectual curiosity, constant scanning of economic, political, and social landscapes, as well as a sense of history so that events can be placed in context.

Markets are not priced deterministically. Instead they reflect the collective (and subjective) valuations of all the participants, including those who choose to put their money elsewhere. In this markets are adaptive; they reprice as events unfold and are reinterpreted. And markets teach. Markets produce price signals that affect behavior. Traders, professional portfolio managers, policy makers, hedge funds, and retail investors watch to see how financial markets react to events; then they take their cues and adapt if need be.

Market prices are feedback loops. They are the impetus for new ideas, technologies, methods, and organizational forms conjured up by entrepreneurs who seek to exploit new opportunities that seem to beckon. It is the process of Schumpeterian creative destruction that dispenses with the old and brings in the new. In this respect it is worth remembering that the architecture of modern financial markets is borne of the wreckage wrought by the two major monetary policy calamities of the 20th century: the Great Depression of the 1930s and the Great Inflation of the 1970s. Free and competitive international capital markets emerged in the 1980s as a necessary corrective to the financial straitjacket imposed by a regulatory regime conceived in the 1930s. And global liberalism came to the fore on the back of the telecommunications revolution of the 1990s and the collapse of the Soviet empire.

The Berlin Wall came down; the Cold War ended; China began the process of economic (though apparently not political) liberalization; a competitive two-party system emerged in Mexico; the Euro was born. The Internet went mainstream, pushing aside traditional intermediaries in the process. Companies that barely existed 10 years ago became global economic players. Think Google. Securities and futures exchanges went public and introduced electronic trading. The Chicago Mercantile went public in December 2002 at $35 per share. By November of 2006 the stock was trading at over $525, and the Merc had agreed to acquire its crosstown rival, the Chicago Board of Trade, in a stock swap.

Apple Computer revolutionized the music business when it negotiated licensing deals with the major record companies that allowed the computer company to sell songs for download on Apple computers and iPods. The day before Apple announced the deal (April 25, 2003) its stock closed at $13.35. By mid-November of 2006, it was trading at $85. In similar fashion, when the Fed finally began to concentrate on establishing and maintaining price stability and gave up on economic fine-tuning, inflation collapsed and economic growth soared.

In retrospect none of this seems very remarkable. But does the mere recitation of facts—the Merc being the first to go public, Apple negotiating licensing deals, the Fed changing its operating methods—predict the results that actually transpired? No. Because as important as facts are, they need to be interpreted. Facts are relatively easy to obtain. Insight is not so easy to come by. It requires understanding the dynamics driving the process: economic and managerial, in the case of the CME; social and cultural in the case of Apple; and political, in the case of the Fed. Hence the quotation from Charlie Munger in the introduction to this book explaining why he is a voracious reader.

STRATEGY FORMATION AND EXECUTION

Casey Stengel once said of the Mets that they all wore the uniform, but none of them knew how to play the game. Trading can be like that. Then it's called "shoulda, woulda, coulda." Brilliant and insightful analysis does not produce a dime of revenue until somebody swings the bat and a trade ticket is written.

This book treats strategy formation and trade execution as implementation processes. Analysis leads to forecasting the likely course of events. Strategy formation is a process of adopting the best means to achieve the desired end of profit generation based on the forecast. Execution is putting the strategy into play the most efficient way possible. The forecast-strategy-execution process requires understanding what the instruments are, what they represent, how they are priced, as well as market structures and institutional constraints.

The instruments of the capital markets can be classified as debt, equity, hybrid, and their respective derivatives. Debt holders have a contingent claim on a firm's assets; equity holders own the firm. As a result the stock market is commonly referred to as the market for corporate control. Hybrid products take on features of both debt and equity; which one predominates depends on the particulars of the security. Derivatives are, in effect, side bets that derive their value from some other underlying asset. Some derivatives trade on organized exchanges; others trade over the counter.

What can be said about the pricing of securities? A security is worth the present value of its expected future cash flows. But future cash flows are not known for sure and the choice of discount rate is problematic because some cash flows are riskier than others. How to deal with this? The model that lies at the heart of securities pricing theory is the capital asset pricing model. The CAPM assumes that investors are rational players who seek to maximize gains but minimize risk. Securities are thus priced in terms of expected returns, adjusted for risk.

In the CAPM world, there are two types of return: risky and risk free. Risk comes in two varieties as well. It can be idiosyncratic, which refers to security specific risk. Or it can be the market risk—the risk associated with a particular asset class, or more precisely, virtually all risks associated with ownership. Idiosyncratic risk can be diversified away, but market risk cannot. Marginal or excess returns over the risk-free rate are the reward for bearing the risk of ownership. The riskiness of a security can be measured with respect to its volatility compared to other securities in its asset class. It is commonly estimated using the beta of a regression equation in which percentage changes in the price of an individual stock are compared to percentage changes in an underlying benchmark stock index.

One implication of the CAPM is that investors and arbitrageurs will trade off risk for potential reward until they reach the efficient frontier, the point where risk/return profiles are equalized across all securities and asset classes. But that is theory. Transactions costs, market opacity, and institutional constraints may leave prices short of the frontier. Efficient frontier notwithstanding, there is persistent evidence that small cap stocks systematically generate excess returns. Moreover, noise traders and market bubbles may impede arbitrage. As Lord Keynes once said, markets can stay irrational longer than you can stay solvent.

MARKET INSTITUTIONS AND ASSET CLASSES

There are many different institutions in the marketplace serving many different functions. The stock and commodity exchanges, the regulatory agencies, the trade associations, all have important roles to play. But when it comes to market impact there is one institution that dwarfs them all. It is the Fed.

By its nature the Fed is the dominant player in the financial markets. The Fed sets the federal funds rate (roughly equivalent to the risk-free rate of CAPM fame) and effectively determines the inflation rate. In so doing, it also strongly influences inflation expectations, the shape of the yield curve, and the working of the business cycle. Inflation, inflation expectations, and the business cycle all exert an important influence on debt, equity, commodity, and precious metals markets.

The Fed is almost always on a tightrope. If policy is too easy, inflation picks up; too tight, and there is danger of a downward deflationary spiral. It should be noted that the Fed's job have gotten more difficult lately. Globalization and extensive use of derivatives have allowed savvy market participants to offload interest rate and currency risk to third parties. This has had several consequences. One is that the transmission mechanism for monetary policy has been altered in ways that are not fully understood. Another is to attenuate the impact of monetary policy decisions.

One possible solution is for the Fed to change its operating procedures to embrace inflation targeting in place of targeting fed funds. Chairman Bernanke has spoken favorably of this as a policy option. Regardless of its specific operating procedures, the Fed is likely to remain the dominant influence on market behavior for the foreseeable future. And so with respect to trading strategy formation, Fed policy ought to be a central consideration.

U.S. government notes and bonds are the capital market instruments most directly affected by Fed policy. They represent the default-free rate for a given maturity, and they serve as the benchmark for all other dollar-based debt securities. Arguably they are the benchmark for all debt pricing worldwide. They trade almost exclusively over the counter.

In the money markets, Fed policy is the effective determinant of yields; in longer-dated notes and bonds, inflation expectations dominate pricing. With governments, time to maturity is the primary determinant of yield differentials between individual issues. In other credit sectors, credit quality is the most important variable after controlling for maturity. Lower-quality credits trade at a discount (higher yield) to Treasuries, all else equal. Tactical bond market arbitrageurs keep prices in line, buying cheap issues and selling rich ones. Strategic arbitrageurs make bets on the shape of the yield curve. They trade long-dated securities against short-dated ones. Or they may trade quality spreads by going long corporate bonds against selling comparable maturity Treasuries short.

The impact of monetary policy extends beyond fixed-income securities. Fed policy is a critical factor in equity markets as well. The level of rates affects valuations. But not all sectors are affected the same way. Value and growth stocks, cyclicals, and consumer staples are liable to react differently to changes in the interest rate/business cycle. Similarly, credit spreads are likely to be affected by the rate cycle.

Relationships between cash securities and futures contracts are powerfully affected by the interest rate cycle as well. The cash-and-carry model for pricing futures contracts, applicable to stocks, bonds, and commodities, implies that rate changes will ripple through the derivatives markets by altering carry costs. In the end, the CAPM can be applied across asset classes, and since the Fed largely determines the risk-free rate, its influence must be incorporated into virtually all trading decisions.

TRADE IMPLEMENTATION

There are basically two types of trades: directional and arbitrage. Directional trades are driven by market-timing considerations. Traders try to buy when they see the market going up. They sell, or sell short, when the market is expected to go down, whatever that market might be. But with one large (one might say glaring) exception, there is little evidence that any formal trade-timing rules can generate trading results any better than would be produced by chance. The one glaring exception to this is the movement of short-term interest rates. Unlike stock prices which exhibit short-term random behavior, short-term interest rates often move in a trending fashion.

There is a good reason for this. Short-term rates respond to the Fed, and the Fed's behavior is not random. When the Fed chooses a policy stance, it tends to stay on course until it reaches (or if history is a guide, overshoots) its objective. But that takes time. U.S. GDP is about $13 trillion as of this writing. It takes a while to turn a boat that big around. When the Fed embarks on a change in the direction of policy, it typically follows up with a long series of eases (or tightenings). In the beginning of the process, the odds favor a directional strategy in sync with the Fed at the short end of the yield curve. When the Fed begins to ease policy, it is reasonably safe bet that a long position in short-term Treasuries will produce profits, and vice versa. That is not necessarily true for either long-dated bonds or equities. Directional trades are best done at the front end of the Treasury market.

Arbitrage trading is a horse of a different color. Arbitrageurs go long what they perceive to be undervalued securities and sell other issues short on the theory that the spread will eventually work its way back to fair value. There are a number of considerations that go into arbitrage trading. First is the extent of the perceived mispricing, which implies the profit potential when things get back to normal. Second is the ease or difficulty of financing the transaction; cash securities that are shorted have to be borrowed. Third is the workout time. The longer it takes a trade to work, the longer capital is tied up.

Some arbitrage trading is based on fundamental conceptions of value/disvalue. Credit quality spreads would fall into this category. Other arbitrage trades are based on a market view; strategic yield curve trades would be an example. Still others are statistical arbitrage trades, based on historical spreads. That is mostly what pairs trading is about. Basis traders seek to profit by faulty pricing of the spread between cash securities and a futures contract. It is similar in technique to pairs trading. The difference is that convergence is necessarily driven by the contract delivery process.

Beside getting the spread right, the most important factor to consider for implementing arbitrage trades is how to weight longs against shorts. Although the underlying idea is to weight by volatility, techniques for doing so vary by asset class. Fixed-income arbs are generally weighted by a measure of price/yield sensitivity. The most common is a bond's DV01, an acronym that stands for the dollar value of 1-basis-point change in the bond's yield to maturity. Other methods rely on a bond's duration. But it needs to be kept in mind that DV01s change with the level of rates, and do so at different rates of speed—the convexity issue—so traders need to constantly update their hedge ratios to make sure that they are in sync with the market.

Equity arbitrage weightings are more likely to be based on historical behavior than deterministic formulas. In pairs trading, the spread between two highly correlated stocks is traded as if it were a single security. The buy/sell decision is based on the history of the spread, and the weighting factor is a regression coefficient that serves the same function as beta does in the CAPM.

In basis trading the weighting depends on the particulars of the futures contract. In bonds, a delivery factor is used as the base weight for the cheapest-to-deliver bond. Other bonds can be weighted by their published delivery factors, or the weights can be adjusted for DV01s. In equity index trading, roughly equal dollar weights are used when the object is to precisely replicate a benchmark index. When a sample of stocks is used to mimic the benchmark, it is advisable to stress-test the sample against the benchmark. Then volatility weights for the sample portfolio can be based on regression analysis of past history.

POSITION MANAGEMENT

The first thing most good traders do when they walk into a room is look for the location of the exit door, just in case. People put on trades because they expect to make money, but it doesn't always work out that way. Position management requires knowing when to fold, understanding the risks, and having an exit plan. Then when things go wrong, you can get up and dust yourself off to fight another day.

But that is easier said than done. Behavioral research strongly suggests that market professionals suffer from overconfidence just as much as the average man—if not more so. Quirks of human psychology—among them bias, overconfidence, narrow framing, overgeneralizing, wishful thinking, belief perseverance, and anchoring—can lead to putting on bad trades and staying in them rather than admitting error. These types of trading errors can get expensive rather quickly, so one of the first rules of position management is to learn to take losses quickly when things go wrong.

There are other types of dangers lurking in the background that many, if not most, risk management models fail to capture adequately. It is worth reviewing some of the implications of this. The first one is that markets can be a lot more volatile than you counted on. Most risk models assume that financial market returns are normally distributed. In fact they are not. The distribution of financial market returns is fat-tailed, meaning that "unlikely" but high-impact events (the long tails of the distribution) are in fact a lot more likely to happen than they are given credit for. A related second consideration is that markets are not as free-flowing and continuous as they are portrayed to be in the textbooks. Liquidity is most likely to dry up when a crisis erupts, making it difficult to get out of positions.

The main enemy of good position management is overleveraging. When traders are strapped for cash, they are forced to liquidate good positions to finance bad ones. It is an unenviable position to be in. The best way to avoid it (and be able to take advantage when the next crisis inevitably hits) is to leave a margin of error—one that is larger than conventional models suggest is necessary.

THE ART OF TRADING

In the 1960s lots of traders came to the game without a lot of formal training, equipped mainly with street smarts. By the late 1990s lots of traders came to the game equipped with Ph.D.'s in physics, mathematics, economics, and engineering. Quantitative analysis became increasingly important as the markets became bigger, more complex, and global. And yet some of the bigger disasters have their roots in sophisticated quantitative strategies that went wrong and imploded, taking down firms along with them. That is worth thinking about.

In quantitative analysis it is common to refer to models of the market. And traders often refer to "the market" as if it were an independent entity, acting on its own. But it isn't. Organized markets aggregate human behavior. Prices reflect the range of (and maybe fight between) human emotion and reason, usually summed up as the struggle between fear and greed. The reason why markets careen between mild and wild, between periods of complacency and frenetic activity, is that they reflect the way humans behave. Returns distributions have fat tails because crowd psychology has fat tails. Go to a hockey game and watch the emotions of the crowd range from jubilation to despair and back again in a little over an hour or so. In this respect trading floors bear no small resemblance to sports fields.

Some people are good at watching a play unfold; some are adept at reading crowds. In a similar vein, some traders are good at reading the flow of the crowds' emotions, embedded as they are in the market's prices. They do it by juxtaposing actual and expected market performance in the face of market-moving information. This model of using market performance as a feedback loop implies both an intuitive sense of crowd psychology and a fundamental understanding of the context in which the game is played. Trading is more than economics, statistics, and market models. There is an art to it as well.

SUMMARY

The trading business can be highly lucrative, and like all highly lucrative endeavors it is intensely competitive. Four elements are crucial to successful trading. The first is analytic ability. Information needs to be collected, filtered, analyzed, and interpreted with an eye toward taking action. The second is the formation and execution of strategy. Post analysis, a strategy that fits the analysis needs to be selected and implemented. The third is risk management. Profit potential should be at least equal to the risk assumed. The exit doors should be clearly marked, and worst-case scenarios should be considered. Finally, trading is not science. There is an art to it. And good traders, like good artists, know how to get the most from their tools without confusing their tools with their craft.

This page intentionally left blank

APPENDIX

Suggestions for Further Reading

Benning, Joseph F. "Differentiating Fixed-Income Performance with Portable Alpha." *Alternative Investment Management Journal,* Summer 2006.

Bernstein, Peter. *Capital Ideas: The Improbable Origins of Modern Wall Street*. John Wiley and Sons, 2005.

Biggs, Barton. *HedgeHogging*. John Wiley and Sons, 2006.

Buchanan, James. *Constitutional Economics*. Basil Blackwell, 1991.

Burghardt, Galen. *The Eurodollar Futures and Options Handbook*. McGraw-Hill, 2003.

Burghardt, Galen, and Belton, Terry. *The Treasury Bond Basis*. McGraw-Hill, 2005.

Burnham, Terry. *Mean Markets and Lizard Brains*. John Wiley and Sons, 2005.

Chorfas, Dmitris N. *Credit Derivatives and the Management of Risk*. Prentice Hall, 2000.

DeRosa, David F. *In Defense of Free Capital Markets*. Bloomberg Press, 2001.

Edwards, Robert D.; Magee, John; and Bassetti, W. H. C. *Technical Analysis of Stock Trends,* 9th ed. St. Lucie Press/CRC, 2007.

Friedman, Milton. *Money Mischief: Episodes in Monetary History.* Harcourt, Brace, Jovanovich, 1992.

Friedman, Milton, and Schwartz, Anna Jacobson. *A Monetary History of the United States,* 1867-1960. Princeton University Press, 1963.

Gitmann, Lawrence J., and Joehnk, Michael D. *Fundamentals of Investing*. Addison Wesley, 2004.

Global Association of Risk Professionals: http://www.garp.com/index.asp

Grieder, William. *Secrets of the Temple: How the Federal Reserve Runs the Country,* Simon and Schuster, 1987.

Groppelli, A. A., and Nikbaht, Ehsan. *Finance*. Barron's Educational Series Inc., 2000.

Jones, Cen. *Back to Basics: Portable Alpha*. Pensions Management, August 1, 2004. Available online at: http://www.pensions-management.co.uk/news/fullstory.php /aid/1072/Back_to_basics:_portable_alpha.html

Kissinger, Henry. *Diplomacy*. Simon and Schuster, 1994.

Koehler, Adele. *Implementation Guide for Portable Alpha and Efficient Beta Exposure*. State Street Global Advisors Essays and Presentations, December 13, 2004. Available online at: http://www.ssga.com/library/esps/Implementation_Guide_for_Portable_Alpha_Adele_Kohler_12.13.04rev.pdf

Lerman, David. *Exchange-Traded Funds and E-Mini Stock Futures*. John Wiley and Sons, 2001.

Mackin, John H., and Ornstein, Norman J. *Debt and Taxes*. Times Books, 1994.

Malkiel, Burton G. *A Random Walk Down Wall Street*. W. W. Norton and Co., 2003.

Mandelbrot, Benoit B., and Hudson, Richard L. *The Misbehavior of Markets*. Basic Books, 2004.

Mehrling, Perry Fischer. *Black and the Revolutionary Idea of Finance*. John Wiley and Sons, 2005.

Miles, Marc A.; Holmes, Kim R.; O'Grady, Mary Anastasia. *2006 Index of Economic Freedom, Executive Summary,* Heritage Foundation. Available online at: http://www.heritage.org/research/features/index/about.cfm

Moen, J., and. Ellis, T. "Lessons from the Panic of 1907." *Federal Reserve Bank of Atlanta Economic Review,* May–June 1990: 2–13.

Neiderhoffer, Victor. *The Education of a Speculator*. John Wiley and Sons, 1997.

New York Stock Exchange. "NYSE, New York Stock Exchange." Available online at: http://www.nyse.com/

North, D. C. *Structure and Change in Economic History*. W. W. Norton and Co., 1981.

Peters, Edgar E. *Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility*. John Wiley and Sons, 1996.

Poon, Ser-Huang. *A Practical Guide to Modeling Financial Market Volatility*. John Wiley and Sons, 2005.

Postrel, Virginia. *The Future and Its Enemies: The Growing Conflict over Creativity, Enterprise, and Progress*. The Free Press, 1999.

Schumpeter, Joseph. *Capitalism, Socialism, and Democracy*. Peter Smith Publisher, 1984.

Securities and Exchange Commission. "Exchange Traded Funds." Available online at: http://www.sec.gov/answers/etf.htm

Shafritz, J. M., and Hyde, A. C. *Classics of Public Administration*. Harcourt Brace College Publishers, 1997.

"Special Reports, Portable Alpha." *Financial Times*, n.d. Available online at: http://www.ftmandate.com/news/categoryfront.php/id/77/PORTABLE_ALPHA.html

Steiner, Robert. *Mastering Financial Calculations*. Prentice Hall, 1998.

Stevens, Leigh. *Essential Technical Analysis*. John Wiley and Sons, 2001.

Taleb, Nassim Nicholas. *Fooled by Randomness: The Hidden Role of Chance in the Markets and Life,* 2nd ed. TEXERE, 2004; Random House, 2005.

Tavakoli, Janet M. *Credit Derivatives and Synthetic Structures*. John Wiley and Sons, 2001.

Teebagy, Nicholas, and Aczel, Amir D. *The Math behind Wall Street: How the Market Works and How to Make It Work for You*. Four Walls Eight Windows, 2000.

Vidyamurthy, Ganapathy. *Pairs Trading*. John Wiley and Sons, 2004.

Wilmott, Paul. *Paul Wilmott on Quantitative Finance*, 2 vols. John Wiley and Sons, 2001.

Wilson, Woodrow. "President Wilson's Fourteen Points: The Basis of the New World Order" [International League for Permanent Peace, 1918]. Available online at: http://www.woodrowwilson.org/learn_sub/learn_sub_show.htm?doc_id=377217

Zaloom, Caitlin. *Out of the Pits*. University of Chicago Press, 2006.

Index