Optimal Portfolio Modeling
Founded in 1807, John Wiley & Sons is the oldest independent publishing company in the United States. With offices in North America, Europe, Australia, and Asia, Wiley is globally committed to developing and marketing print and electronic products and services for our customers’ professional and personal knowledge and understanding.

The Wiley Trading series features books by traders who have survived the market’s ever changing temperament and have prospered—some by reinventing systems, others by getting back to basics. Whether a novice trader, professional, or somewhere in-between, these books will provide the advice and strategies needed to prosper today and well into the future.

For a list of available titles, please visit our Web site at www.WileyFinance.com.
Optimal Portfolio Modeling

Models to Maximize Return and Control Risk in Excel and R + CD-ROM

PHILIP J. MCDONNELL

John Wiley & Sons, Inc.
Authoring a book is a labor of love coupled with a dash of inspiration and an abundance of hard effort. Needless to say this takes a toll on family life. With these thoughts in mind, this book is dedicated to my wife Pat and my family. Without their patient tolerance this work would not have been possible.
About the Author

Philip J. McDonnell is an active options trader living in Sammamish, Washington. He and his wife have two grown children.

Mr. McDonnell has been trading options since the first listed options exchange was formed. His trading experience spans more than 30 years. The emphasis and focus of his trading have been on quantitative methods.

Formerly, Mr. McDonnell was president of Dollar/Soft Corporation, a financial software company specializing in options and derivatives-related products. He has done consulting for Charles Schwab & Co. and developed risk-management software to allow the firm to perform what-if analyses of its customer margin accounts.

He has also been the president of Accelerated Data Systems, in which he and his team designed a new scientific high-performance minicomputer. Previously, he was vice president of engineering and manufacturing for a microcomputer company and helped turn critical manufacturing problems around. Prior to that, he served as president of Advanced Information Design, a maker of small business computer systems. His experience also includes stints at Stanford Linear Accelerator and a Northern California investment company.

His academic background is centered around the University of California at Berkeley, where he received his B.A. in 1972. His academic credentials include full satisfaction for all the degree requirements in mathematics, computer science, and statistics. His work also included all of the course work in finance and investments at both the undergrad and graduate levels.

He was employed as a researcher by the U.C. Berkeley Business School, where he worked closely with Prof. Victor Niederhoffer.
## Contents

<table>
<thead>
<tr>
<th>Preface</th>
<th>xi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgments</td>
<td>xiii</td>
</tr>
</tbody>
</table>

### CHAPTER 1: Modeling Market Microstructure—Randomness in Markets

| The Random Walk Model | 3 |
| What You Cannot Predict Is Random to You | 5 |
| Market Microstructure | 7 |
| Efficient Market Hypothesis | 9 |
| Arbitrage Pricing Theory | 10 |

### CHAPTER 2: Distribution of Price Changes

| The Normal Distribution | 13 |
| Reflection Principle | 17 |
| Approximation of the Normal Distribution by Rational Polynomial | 18 |
| Lognormal Distribution | 19 |
| Symmetry of the Normal and Lognormal | 22 |
| Why Pick a Distribution at All? | 23 |
| The Empirical Distribution | 24 |
| The Lognormal as an Approximation | 26 |
# CONTENTS

## CHAPTER 3  Investment Objectives

- Statistician's Fair Game 29
- A Fair Game Is a Loser! 30
- Criteria for a Favorable Game 30
- Gambler's Ruin 31
- Optimal Return Models 32
- Markets Are Rational, Psychologists Are Not 34
- The St. Petersburg Paradox 36
- Compounded Return Is the Real Objective 37
- Defining Risk 38
- Minimum Risk Models 41
- Correlation of Assets 41
- Summary of Correlation Relationships 42
- Beta and Alpha 43
- The Efficient Frontier and the Market Portfolio 46
- The Sharpe Ratio 47
- Limitations of Modern Portfolio Theory 48

## CHAPTER 4  Modeling Risk Management and Stop-loss Myths

- Stop-loss Orders 52
- Stops: Effect on the Mean Return 53
- Stops: Effect on the Probability of Gain 56
- Stops: Probability of Being Stopped Out 56
- Stops: Effect on Variance and Standard Deviation 58
- Effect on Skew 59
- Effect on the Kurtosis 60
- Stop-loss: Summary 61
- Modeling Stops 61
- Identifying When to Use Stops and When Not To 62
- Stop-Profits 64
- Puts and Calls 65

## CHAPTER 5  Maximal Compounded Return Model

- Optimal Compound Return Models 68
- Relative Returns 68
- Average Stock Returns, but Compound Portfolio Returns 70
## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Logarithms and the Optimal Exponential Growth Model</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Position Sizing as the Only Guaranteed Risk Control</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Controlling Risk through Optimal Position Sizing</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Maximize Compounded Portfolio Return</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Maximal Compounded Return Models</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>What the Model Is and Is Not</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Modeling the Empirical Distribution</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Correlations</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>The Enhanced Maximum Investment Formulas</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Expected Drawdowns May Be Large</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td><strong>CHAPTER 6</strong> Utility Models—Preferences Toward Risk and Return</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Basis for a Utility Model</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>History of Logarithms</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Optimal Compounded Utility Model</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>The Sharpe Ratio</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Optimal Model for the Sharpe Ratio</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Optimization with Excel Solver</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td><strong>CHAPTER 7</strong> Money Management Formulas Using the Joint Multiasset Distribution</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>The Continuous Theoretical Distributions</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Maximal Log Log Model in the Presence of Correlation</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Optimal Sharpe Model with Correlation</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>The Empirical Distribution</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Maximal Log Log Model in the Presence of Correlation</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Maximizing the Sharpe Ratio in the Presence of Correlation</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td><strong>CHAPTER 8</strong> Proper Backtesting for Portfolio Models</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>Assuring Good Data</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Synchronize Data</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Use Net Changes Not Levels</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>Only Use Information from the Past</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Predictive Studies versus Nonpredictive Studies</td>
<td>106</td>
</tr>
</tbody>
</table>
Successful investing is more than just picking the right stock. It is not just about timing the market. It requires two essential ingredients. The first ingredient is having a winning edge. The author’s motto is: In order to beat the market, you need an edge. Never let your money leave home without it.

However, that is not enough. Even if one has a winning edge, it is entirely possible to invest too recklessly or too conservatively and not achieve your objectives. Even with a winning system, excessive position sizes can destroy a trade just as surely as lack of an edge. Failure to take adequate risks can also lead to underperformance.

The purpose of this book is to show how to achieve the right balance of position sizing and risk management so as to achieve the investment goals. Many books have covered how to have a winning edge. That is not the purview of this work. Rather, this book focuses on the relatively unexplored realm of money management and portfolio modeling. Managing a portfolio through position sizing is at least as important as finding and maintaining an investment edge.

Optimal Portfolio Modeling provides an introduction to the statistical properties of markets in the early chapters. The book is oriented to intelligent people who may not be full time rocket scientists. The author resisted the very strong temptation to title this work Rocket Science for Average Folks. Instead, the seemingly more dignified title was chosen.

Nevertheless, this work is designed to be very accessible to all, even with a limited math background. Only high school algebra is required to understand this book. Readers with an advanced technical degree may be astonished to find that calculus is not required. They may argue that optimization requires the use of calculus. This is true. However, Excel knows how to do the calculus and so does the statistical language R. Users do not need to know the math in order to understand the result. Only basic high school algebra is required.

These tools know how to do the magic. The user merely needs to know how to invoke the magic spell and how to interpret the results. This book assumes the reader has some beginner level of knowledge of Excel. The text fully explains how to use the built-in Solver, which allows the user to optimize models.
For readers interested in the statistical language R, advanced users will find this work to be an augmentation of their knowledge, with special emphasis on portfolio modeling and optimization problems. Beginning users of the open-source language R will appreciate the fact that the appendices and CD offer both a tutorial and introduction to R targeted to the beginning user. Additionally, the text identifies those functions that are appropriate for optimization in the powerful R library.

An important part of setting up a portfolio model is defining the objective or goal that one wishes to achieve. A significant portion of this book is devoted to the question of what is the right goal for the investor to seek. Often, this book refers to this as the utility function. Not all investors will or should choose the same objective function. The text discusses how to choose an appropriate objective function. A manager who is benchmarked to the Sharpe Ratio should choose that as the objective function. However, for more typical investors, the book strongly argues that maximizing long-term compounded growth of wealth should be an important component of the overall model. An innovative formula is provided that optimizes long-term wealth as well as provides a measure of stability not seen in other optimal money management formulas published to date.
Acknowledgments

Any work such as this builds on the advances and ideas of so many previous giants in the field. In particular, this work is based on the breakthroughs and developments of countless researchers in the fields of finance and statistics. Although they are too numerous to name in this space, the author wishes to acknowledge the foundation that they have built in the field. This book would not have been possible without the giants who preceded us, many of whom are named in the text.

Every person owes a debt of gratitude to those who have taught them throughout their lives. In this regard, I wish to acknowledge and thank my parents, who were my first teachers. They encouraged and inspired my endeavors.

The efforts and dedication of the many teachers and even fellow students who have helped in my education must be gratefully acknowledged. Although too numerous to name individually, the cumulative contribution of their efforts and generosity with their knowledge has been of inestimable value in my personal development. Naturally, it has been a requisite foundation, essential to this work.

There is one particular teacher I met as a professor of finance when we were both associated with the University of California at Berkeley. He is Victor Niederhoffer, who has enjoyed a meteoric speculative career. Prior to meeting him, I had the fuzzy notion that the best approach to the markets was through a rigorous quantitative methodology. At that time, the random walk ruled the thinking in finance. So there was little call for eccentric academics who thought they could beat the market, quantitatively or any other way. Professor Niederhoffer was just such a divergent thinker.

His help and guidance taught me to see things at their simplest. That is the essence of his approach. His enlightenment also helped me to learn how to avoid the numerous pitfalls that can arise in quantitative studies. In fact, one of the things he taught me was what not to do on a quantitative study. Perhaps more than anything, working with him inspired and motivated me to further my efforts to study the markets from a quantitative perspective. His inspiration has lasted a lifetime.

One must also acknowledge the hundreds of friends known as the Spec List who have stimulated my thinking, inspired my work, and helped me in so many ways. Victor
was the founder of the group and remains its chair. We all communicate daily to discuss markets with a quantitative focus. Unquestionably, the help and support of this remarkably intelligent but diverse group has been a source of inspiration to me. Many of the ideas herein are amplifications of some of the ideas I have discussed more briefly with the group. Their collective comments, questions, and debate have helped to refine these ideas. Several of my friends in the group have encouraged me to write this book, and certainly this must be acknowledged.

Perhaps most directly, I must acknowledge the very helpful people at Wiley who were directly responsible for helping with this work. Foremost of those is Pamela van Giessen who edited this work. Her experienced guidance and advice were critical in shaping this book. Her encouragement to become an author was the final impetus that helped to launch this project.

Finally, I must thank Kate Wood and Jennifer MacDonald of Wiley, who were so helpful in supporting and editing this book. There are also many other people at Wiley whose names I do not even know who have contributed to this effort. I am very grateful to all of them.

Even though many have contributed to this work, ultimately, any errors are mine.
Traditionally, portfolio modeling has been the domain of highly quantitative people with advanced degrees in math and science. On Wall Street, such people are commonly called rocket scientists. *Optimal Portfolio Modeling* was written to provide an easily accessible introduction to portfolio modeling for readers who prefer an intuitive approach. This book can be read by the average intelligent person who has only a modest high school math background. It is designed for people who wish to understand rocket science with a minimum of math.

The focus of this book is on money management. It is not a book about market timing, nor is it designed to help you pick stocks. There are numerous other books that address those subjects. Rather, this work will show the reader how to define models to help manage money and control risk. Stock selection is really just the details. The big picture is actually about achieving your overall portfolio goals.

Included with this book is a CD-ROM that includes numerous examples in both Excel and R, the statistical modeling language. The book assumes the user has a beginner’s level knowledge of Excel and focuses mainly on those specific areas that apply to portfolio modeling and optimization. There are many books that offer an introduction to Excel, and the interested reader is encouraged to investigate those.

R is an open-source language that offers powerful graphics and statistics capabilities. Two appendices in this book offer introductory support for users who wish to download R at no cost and learn how to program. Because R is powerful, many functions and graphs can be done with very few command lines. Often, only a single line will create a graph or perform a statistical analysis.
The overriding philosophy of all of the examples is simplicity and ease of understanding. Consequently, each example typically focuses on a single simple problem or calculation. It is the job of the computer to know how to perform the calculations. The user only needs to know how to invoke the right computer function and to understand the results. Understanding and intuition are the primary goals of this book.

This chapter introduces the important background of market microstructure and randomness. This is a foundation for the ideas developed later in this book. The discussion starts with a thorough introduction to the idea of randomness and what a random walk is. The topic of randomness is presented as an essential element in understanding how and why a portfolio works. After all, the primary rationale for a portfolio is intelligent diversification.

From there, the book moves to a discussion of market microstructure and how it affects the operation of markets. Later, the reader is introduced to the efficient market hypothesis, along with its history and development, starting with early pioneers in the field. Augmenting this is the discussion on arbitrage pricing theory and its modern applications. This latter topic shows how the market identifies and eliminates any risk, less arbitrage opportunities.

Trading speculative markets has always been difficult. Over the years, several studies have shown that some 70 to 80 percent of all mutual funds underperform the averages. A study by Professor Terrance Odean of the University of California at Berkeley demonstrated that most individual investors actually lose money. This study analyzed thousands of real-life individual investor brokerage accounts. Thus, it provides a comprehensive look at how real individual traders operate. The inescapable conclusion is that both professional and individual investors find that trading the markets is challenging.

Successful trading is predicated on one thing. Traders must predict the direction of price changes in the future. At a minimum, a successful trader must predict prices so that each trade has an expectation of yielding a profit. This does not mean that each trade must be successful, but, rather, that a succession of trades would usually be expected to result in a profit. This should not be taken to mean that having a positive expectation for each trade is the only thing a successful trader needs. The astute reader will note that the use of words such as usually, average, and expectation naturally implies that the art of forecasting is far from perfect. In fact, it is best studied from a statistical perspective with a view to identifying what is random and what is predictable.

In a recent 500-day period, the stock market as measured by the Standard and Poor’s 500 index was generally a modestly up market. A statistical analysis of the daily compounded returns for the period shows:

- Average daily return: .038 percent
- Standard deviation: .640 percent
- Probability of rise: 56 percent
The standard deviation is simply a measure of the variability of returns around the average. From this simple analysis, we can make some interesting observations:

1. The average daily return is small with respect to the standard deviation.
2. The daily variability is relatively large, at 16 times the return.
3. The market went up 56 percent of the time, or slightly more than half. It also went down the other 44 percent of the days. So even during up markets, the number of up days is only slightly better than 50–50.
4. The variability completely swamps the average return.

Observations such as these have led many early researchers in finance to propose a model for the markets that explicitly embraces randomness at its very core. A cornerstone of this idea is that markets represent all of the knowledge, information, and intelligent analysis that the many participants bring to bear. Thus, the market has already priced itself to correspond with the sum of all human knowledge. In order to outperform the market, a trader must have better information or analysis than the rest of the participants collectively. It would seem the successful trader must be smarter than everyone else in the world put together.

**THE RANDOM WALK MODEL**

To the typical layman, the random walk model is the best-known name for the idea that markets are very good at pricing themselves so as to remove excess profit opportunities. The academic community generally prefers the description the efficient market hypothesis (EMH). Either way, the idea is the same—it is very difficult to outperform the market. If someone does outperform, then it is likely only attributable to mere luck and not skill.

The history of the EMH is a rather long one. The first known work was by Louis Bachelier in 1900, in which he posited a normal distribution of price changes and developed the first-known option model based on the idea of a normal random walk (see Figure 1.1). His seminal paper in the field was quickly forgotten for some 60 years. As an interesting side note, the mathematics that Bachelier developed was essentially the same analysis that Albert Einstein reinvented in 1906 in his study of Brownian motion of microscopic particles. Einstein’s famous paper was published some six years after Bachelier’s work. However Bachelier’s paper languished in relative obscurity until its rediscovery in the 1960s.

Prof. Paul Samuelson of the Massachusetts Institute of Technology offered a Proof that Properly Anticipated Prices Fluctuate Randomly in the 1960s. This provided a theoretical basis for the EMH idea. However, it fell to M. F. M. Osborne to provide the
modern theoretical basis for the efficient market hypothesis. Osborne was the first to posit the idea of a lognormal distribution and provide evidence that the price changes in the market were log normally distributed. Furthermore, he was the first modern researcher to draw the link between the fluctuations of the market and the mathematics of random walks developed by Bachelier and Einstein decades earlier.

Osborne was a physicist by training employed at the U.S. Naval Observatory. As such, he was not an academic, nor did he come from a traditional finance background. Thus, it is not surprising that he is rarely recognized as the father of the efficient market hypothesis in the lognormal form. However, it is very clear that his empirical and theoretical work that described the distribution of stock price changes as log normal and the underlying process of the market as being akin to the process described by Einstein called Brownian motion was the first to elucidate both concepts. Osborne deserves the honor of being the father of the EMH.

As so often happens in academia, others who published later and were fully aware of Osborne’s work have received much of the credit. Statistician and student of mathematical and statistical history, Stephen M. Stigler has whimsically called the phenomenon his law of eponymy. The wrong person is invariably credited with any given discovery.

One aspect of this phenomenon is that when a person is erroneously credited with a discovery for whatever reason, his or her name is attached to that discovery. After much widespread usage, the name tends to stick. So even when it is later discovered by
historians that someone else actually discovered the idea first, it is usually just treated as a footnote and rarely adopted into common usage among practitioners in the field.

Such is the case for Osborne's contribution to the efficient market hypothesis. It was partly because he was a physicist working in the field of astronomy. At the time of his publication, he was not really an accepted name in the field of finance.

One form of the EMH defines the relationship between today's price $X_t$ and tomorrow's price $X_{t+1}$ as follows:

$$X_{t+1} = X_t + e$$  \hspace{1cm} (1.1)

where $e$ is a random error term. We note that this model is inherently an additive model. The usual academic assumption corresponding to this type of model is the normal distribution. The key concept is that the normal distribution is strongly associated with sums of random variables. In fact, there is a weak convergence theorem in probability theory that states that for any sums of independent identically distributed variables with finite variance, their distribution will converge to the normal distribution. This result virtually assures us that the normal distribution will remain ubiquitous in nature.

However, the empirical work of Osborne showed us that the distribution of price changes was log normal. This type of distribution is consistent with a multiplicative model of price changes. In this model, the expression for price changes becomes

$$X_{t+1} = X_t(1 + e)$$  \hspace{1cm} (1.2)

WHAT YOU CANNOT PREDICT IS RANDOM TO YOU

Some would argue that the market is not random. Certainly, almost every single participant in the market believes he or she will achieve superior results. Most of these participants are smarter, richer, and better educated than average. Can they all achieve superior returns? Of course, it would be mathematically impossible for everyone to be above average. Can they all be deluded?

To answer this question it is helpful to look at the long-term history of the market. When we fit a regression line through the monthly Standard & Poor’s closing prices $P_t$ on the first trading day of each month since 1950 until November 2006, we find the following:

$$\ln P_t = .0059 t + 3.06$$

In this case the $t$ values are simply month numbers starting at 1, then 2, and so on for each of the 683 months in the study. The fitted coefficient .0059 can be interpreted as a simple monthly rate of increase in the series. So if we annualize, we get an annual rate of return of about 7.1 percent for the long-term growth rate of the Standard & Poor's 500 average (see Figure 1.2). This is a very respectable long-term upward trend in the
market. The $R^2$ for this regression was 97 percent. Given that 100 percent is a perfect fit, this indicates that the model is a very good one.

The underlying message here is that the market goes up over time. The fact that the natural log model fits well tells us that the growth in the market is compounded and presumably derived from a multiplicative model. But beyond that, it tends to make people think they are financial geniuses who might not be.

*Bull markets make us all geniuses.*

—Wall Street maxim

From the perspective of the long-term time frame, the market has been in a bull phase for at least the entire last century. Human beings have a natural propensity to attribute good luck to their own innate skill. Psychologists call this the *self-attribution fallacy.* The long-term bull market has created a large group of investors who believe they have some superior gift for investing. Few investors ever stop to critically analyze their own results to verify that they are indeed performing better than the market.

Given that the market exhibits long-term compounded returns over time, it is clear the best model is a multiplicative one. This long-term return is often called the *drift*—the tendency of the market to move inexorably upward over time. However, to understand the shorter-term movements of the market, we must look to a different kind of model in which the short-term fluctuations appear to be more random. The reason for that is simply because the marketplace in general will anticipate all known information, and thus, the current market price is the best price available. Thus, by definition, any news that is material to the market and was not anticipated will appear as random shocks in either direction.
The key idea to understand is that the market will not respond to news that it already knows. Or if it does, that response will be contrary to what a rational analyst might have expected. These contrary movements are caused when a large group of investors was expecting a certain piece of news and thus, holding positions that were previously taken. When the news is announced, the entire group may try to unwind their positions, resulting in a market movement in exactly the opposite direction one might expect. Simply put, the market has already discounted the expected news and adjusted the price well in advance. Because this phenomenon is so prevalent, Wall Street has evolved the maxim, “Buy on the rumor, sell on the news.” Although one would never recommend relying on rumors for investment success, certainly buying on the correct anticipation of news is the better strategy.

This leaves us with the realization that, absent informed knowledge of upcoming news, the outcome of such events will be random and unpredictable to us. Some would argue that for most news someone knew the event in advance. Certainly for earnings announcements and government reports, someone did know the information to a certainty. For them, the news was not random but completely predictable. Assuming the information was not widely disclosed, then for the rest of investors, the information remains random and unpredictable.

There is a general principle at work here. *If we cannot predict the news, then it is random to us.* So even if others know the information, then insofar as we do not, and cannot predict it, it remains random for us.

**MARKET MICROSTRUCTURE**

Generally speaking, the market consists of the interactions between four broad classes of orders. These can be grouped into two categories each. There are market orders and there are limit orders. There are orders to buy and sell. Although there are variations and nuances on each, these characterize the main categories of trading orders.

- **Market order**—A market order is an order to buy or sell that is to be executed immediately at the best available price
- **Limit order**—This is an order to buy or sell that is only to be executed at the specified limit price or better. Limit orders may have an expiration, such as the end of the day or 60 days.

The quote at any given time is essentially based on the best limit order to buy, which is known as the bid, and the best limit order to sell, which is the ask. When a market order to sell comes in, it is usually crossed with the bid. Therefore, we can expect the price of a market order to sell to be the bid price. We should note that market orders are
usually smaller in size than limit orders. Usually, this means that the market order will be executed at the bid and that the remaining size of the limit order(s) at the bid will be reduced by the amount of the market order. In effect, market orders nibble away at the larger limit orders. It is only after enough market orders have consumed the bid that the bid–ask quote will drop to a new lower bid.

The other side of this process is when a market order to buy, say, 200 shares comes in. Assume there are 1,000 shares for sale at the ask price of 50.10. In this case, the market order will be crossed with the limit order, resulting in a transaction of 200 shares at a price of 50.10. After the transaction, the ask side limit order will show the remaining 800 shares offered at 50.10. It is only after the 800 shares have been consumed by market orders that the ask price will move higher.

Because limit orders tend to persist longer than market orders and are larger than market orders, there is a tendency for the last sale price to alternate back and forth between the bid and ask until one or the other price barrier is consumed. Only then does the quote move. For example, when the ask price is extinguished, the ask will move to a new higher price—the next limit order up. Quite often, the old bid will be superseded by a slightly higher bid, either from a market maker or an off-floor limit order. Thus, the entire quote has a tendency to move up. To really understand the current market situation, one must really look beyond simply the last sale and consider the current bid and ask and the relative size of the bid and ask.

Another important aspect of this market microstructure can be understood in the sense of news. We can view the arrival of market orders as news of an investor’s decision process. In some cases, orders to sell may simply indicate a need for liquidity. It may be as simple as Aunt Mabel in Peoria sold 100 shares to raise money to buy videogames for all her nieces and nephews this Christmas. Alternatively, the sale may mean that an investor’s views on the prospects for the company have changed. This is certainly a different kind of information, but every trade contains information.

The other side of the coin is that predicting and modeling the market at the microstructure level is very difficult. We do not even know who Aunt Mabel is, much less her plans and how many nieces and nephews she has. Thus, her sale of stock for liquidity needs is unpredictable for us. Therefore, any price change it causes is also random to us. So to model this sort of environment we must explicitly allow for a large degree of randomness in the short-term market movements.

The astute reader may have wondered if all market orders are always crossed with limit orders. The answer, of course, is no. Most market orders are crossed with limit orders for the reasons already mentioned, but certainly it is possible for two market orders to arrive at essentially the same time and be crossed with each other. By the same token, it is possible for an aggressive trader to place a large limit order to buy at the ask price or to sell at the bid price. In this case, a limit order will be crossed with a limit order.
There are also stop orders and other contingencies that can be placed on orders. However, for the most part, such as when the stop price is hit, the order becomes a valid market order. Alternatively, for a stop limit order it becomes a valid limit order. Thus, the four-order model just described adequately covers the vast majority of the cases.

It is also worth noting that the market makers effectively act as though they were placing limit orders. Sometimes the bid and ask will both be from a market maker. At other times, one or the other may be an off floor limit order. Nevertheless, the market makers seek to profit from the tendency for the market to trade back and for the several times between the bid and ask prices before moving either higher and lower. This market-maker strategy does not always work, but it works well enough that market makers tend to make a very good profit. A very telling fact is that New York Stock Exchange seats have routinely sold for millions of dollars for quite awhile now.

**EFFICIENT MARKET HYPOTHESIS**

We are now ready to formalize our efficient market hypothesis as a mathematical model. The general principle is that it is a multiplicative model wherein the near-term price changes are swamped by the short-term variability.

Thus, a good statement of the model is the form expressed earlier:

\[ X_{t+1} = X_t(1 + e) \] (1.3)

Here, we have the price today \( X_t \) related to the price tomorrow by a simple random multiplicative term \((1 + e)\), where \( e \) is the random variable. In Chapter 2, we will discuss the nature of the random variable \( e \) to better understand the structure of the market.

We note that equation 1.3 is the multiplicative form analogous to equation 1.2. This is in contrast to the additive model, which is given by equation 1.1. The multiplicative form is consistent with the log normal distribution put forward by Osborne.

Later in this book, we shall show how the multiplicative and lognormal models are also the most appropriate in order to deal with the compound interest effect known to exist in the equity markets. This forms the foundation for the ideas developed later in this book, which allow an investor to maximize long-term compounded returns on the portfolio.

One of the author’s favorite apocryphal stories is that of the finance professor, the economist, and the nimble trader.

One day, a finance professor, who firmly believed in the Efficient Market Hypothesis, was walking along the street. He spotted a one hundred dollar bill lying on the ground. He paused, realized that in an efficient market no one would leave hundred dollar bills lying around. He continued on his walk, confident that it was only a trick of the light.
Minutes later an economist strolled by and saw the hundred dollar bill. He began to calculate to see if picking up the hundred dollar bill would improve his utility of wealth for the day. While he was still calculating, a quick-stepping trader walked past him, picked up the hundred dollar bill and hastily continued on down the street.

The next section has much to do with quick-footed traders.

**ARBITRAGE PRICING THEORY**

A close cousin of the EMH is a theory called *arbitrage pricing theory* (APT). Essentially, this says that the market will not allow any riskless arbitrage to exist. A simple example of riskless arbitrage is if IBM is selling at $80 per share on the New York Stock Exchange and sells for 79.90 on the Pacific Stock Exchange. A nimble trader can buy shares at 79.90 on the Pacific and sell them for 80 in New York for a quick profit of .10. This trade is essentially riskless if done simultaneously.

Arbitrage pricing theory mandates that such opportunities should not exist, or that if they do, they will be quickly extinguished to the point that they are no longer profitable after expenses. It is easy to see why this should happen. In the case of our arbitrage trader in IBM shares when he buys at 79.90, his buying will tend to increase the price on the Pacific Exchange. When he sells in New York, his selling will tend to drive the price there down. Thus, the two prices will quickly come into line and the arbitrage opportunity will be extinguished.

The ideas of APT have developed largely through the efforts of Stephen Ross and Fisher Black. A broader version of these ideas is the concept that one can arbitrage expectations as well as simple price. So rather than just focusing on price differential, the term can include cross relationships between different assets connected via a common factor.

Suppose the price of oil has risen. Then it might be reasonable to believe that the expectation for the earnings of companies that sell oil would be enhanced as well. They are now able to sell at a higher price. Thus, our expectation for the price of oil stocks is now enhanced and we would buy.

Such buying, if done by many, would tend to force the prices of oil stocks up in response to the rise in the oil commodity itself. It is an example of how one factor can drive many stocks. However, the same factor can have a negative impact on other stocks.

An example of this is obvious as well. Again, assume the price of oil has risen as before. If we consider the impact of this fact on automobile companies, we quickly realize that the impact can only be negative. The effect may vary from company to company, but it is negative. It now costs more to fuel your car and consumers are less likely to purchase new cars or extra cars.
Companies that are heavily into gas-guzzling SUVs will be hurt the most. Consumers have the strongest disincentive with respect to these vehicles. It is much easier for them to defer or cancel any new purchase. However, companies that are strong in the economical car submarkets or in fuel-saving hybrids will likely benefit, relatively speaking.

One might suppose that with the advent and ubiquity of modern computing power, such arbitrage opportunities would vanish. However, it is also the case that there has been an enormous rise in derivative instruments in the last few decades as well. It is now entirely possible to buy a basket of stocks representing some index and to trade a futures contract on the index and to trade an exchange traded fund on the index. The number of arbitrage opportunities increases with the number of combinations of instruments available. So when we add multiple futures contracts to the mix, we have many more combinations. But the real arbitrage opportunities are in the large number of options, both puts and calls, at multiple strike prices and various expiration months. On any given day, the markets will trade over 50,000 equity distinct options. The number of listed options is well into the hundreds of thousands. The number of arbitrage combinations of two, three, or more options on all these stocks is well into the millions.

Thus, even with today’s computing power it is still rather difficult for the market to eliminate all arbitrage opportunities. In fact, the market does a remarkably good job, considering the large number of such arbitrages available. The bottom line is that the arbitrage pricing theory is a pretty good model for market efficiency, but not necessarily a perfect one.

At this point the user is encouraged to begin to explore the CD-ROM that came with this book. Each chapter of this work has corresponding examples on the CD-ROM that relate to the topics developed in the chapter. Although using the CD is not required, it is highly recommended as a way to bring the chapter contents to life. The programs and examples provided are generally intended to be as simple as possible and focus only on a particular topic presented in the text.

With that in mind, the user should find it a very worthwhile exercise at the end of each chapter to take a break and review the examples for that chapter. The exercise should take only a few minutes in most cases, but the hands-on experience should prove very helpful in enabling readers to get a feel for the subjects covered.

The reader should start exploring the CD-ROM by reading the appendix, About the CD-ROM.
CHAPTER 2

Distribution of Price Changes

In order to discuss the idea of randomness in the markets, it is very helpful to understand some basic statistics. This chapter introduces the essential ideas of that discipline from an intuitive conceptual standpoint. The discussion begins with a definition of what a probability distribution is and proceeds to discuss the well-known normal and lognormal distributions as they relate to the markets. A little of the history of these distributions is provided as well.

In keeping with the practical nature of this book, a handy formula is provided that allows us to approximate the normal distribution with a rational polynomial. Do not be dismayed if you are unfamiliar with rational polynomials. It is just a formula, and your computer knows how to do those very well.

One of the essential topics of this chapter is the discussion of the reflection principle. The idea is based on the symmetry and self-similarity of the normal distribution. This section should not be skipped, for it is the foundation of several of the ideas presented later in the book.

THE NORMAL DISTRIBUTION

When Maury F. M. Osborne first studied the market, he examined the changes in price for New York Stock Exchange stocks. His research considered both the actual price levels and changes for the stocks in question. His work was the first to scientifically evaluate the distribution of price levels and of changes. This is a good starting point for any research into the nature of market prices.
First we should consider the definition of the word *distribution*. To a statistician, the probability distribution is the probability that the given random variable will be at a certain value level for a given observation. In other words, a distribution associates a range of values and their respective probabilities of occurrence. It is not just one number. For many distributions there is a known formula to calculate the probability that a given observation will be at level $x$. To calculate the probability, we simply plug $x$ into the formula and calculate. However, for many other distributions there is no known formula or the distribution itself is unknown so the convenience of a closed formula is not available.

Most books on probability and statistics start with a distribution known as the *binomial distribution*. It can be used to model a simple game of coin flipping. For our purposes, the probability of a head will be assumed to be 50 percent as is the probability for tails. If you play this game one time and bet $1 on heads, then the outcome will be $+1$ half the time and $-1$ the other half of the time. After only two coin flips, the outcomes and their probabilities are as follows:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-2$</td>
<td>.25</td>
</tr>
<tr>
<td>0</td>
<td>.50</td>
</tr>
<tr>
<td>$+2$</td>
<td>.25</td>
</tr>
</tbody>
</table>

Note that the outcome of zero or breakeven is the most likely event after only two flips. The more extreme outcomes of plus or minus two are less likely. So even after two flips we see that the outcomes tend to pile up in the middle. The reason for this effect is clear if we examine the four possible paths in getting to these outcomes:

- -  
- +  
+ -  
+ +

There are only these four paths possible, and each is equally likely. An inspection of the paths $-+$ and $+$- shows that these two will both result in the zero outcome. Even though the order is different, the outcome is the same. However, to arrive at plus or minus two, there is only one available path. Both must lose or both must win. Intuitively, we can see that the number of paths to arrive at the center is greater than the number of paths to arrive at extreme values. This principle will generalize to more coin flips. It is easier for a random process to arrive near the center than it is to arrive near an extreme value simply because there are more paths to the center and fewer to the extremes.