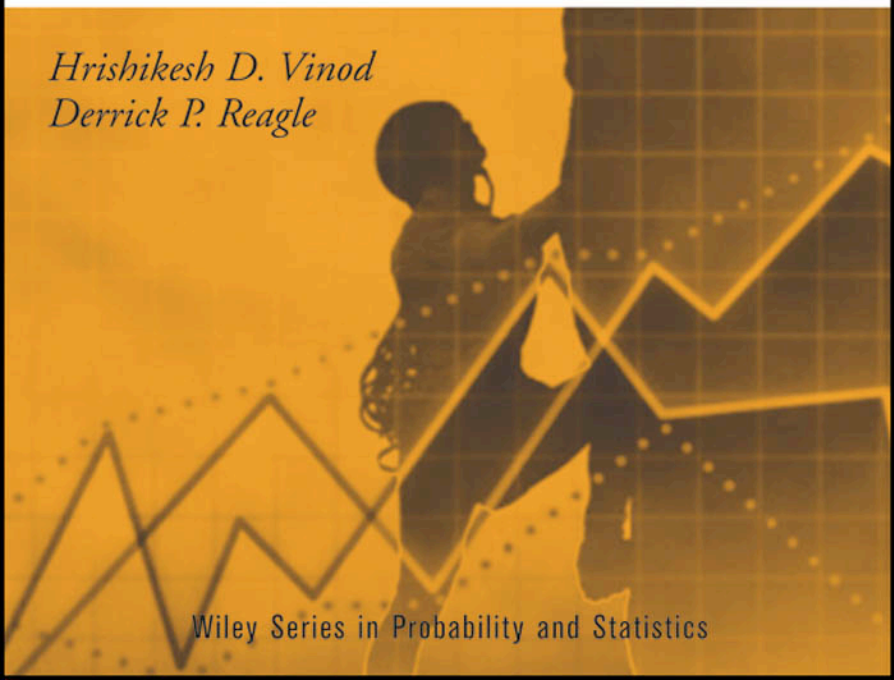
WILEY

# Preparing for the Worst

## *Incorporating Downside Risk in Stock Market Investments*

*Hrishikesh D. Vinod*
*Derrick P. Reagle*



Wiley Series in Probability and Statistics

# Preparing for the Worst

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Noel A. C. Cressie, Nicholas I. Fisher, Iain M. Johnstone, J. B. Kadane, Geert Molenberghs, Louise M. Ryan, David W. Scott, Adrian F. M. Smith, Jozef L. Teugels*
Editors Emeriti: *Vic Barnett, J. Stuart Hunter, David G. Kendall*

A complete list of the titles in this series appears at the end of this volume.

# Preparing for the Worst

## Incorporating Downside Risk in Stock Market Investments

HRISHIKESH D. VINOD

Fordham University
Department of Economics
Bronx, NY

DERRICK P. REAGLE

Fordham University
Department of Economics
Bronx, NY

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.



***Library of Congress Cataloging-in-Publication Data:***
Vinod, Hrishikesh D., 1939–
Preparing for the worst: incorporating downside risk in stock market investments / Hrishikesh D. Vinod, Derrick P. Reagle.
p. cm.
"Wiley-Interscience, a John Wiley & Sons Inc. publication."
Includes bibliographical references.
ISBN 0-471-23442-7 (cloth: alk. paper)
1. Stocks. 2. Investments. 3. Risk management. I. Reagle, Derrick P. (Derrick Peter), 1972– II. Title.

HG4661.V56 2004
332.63′22—dc22

2004047418

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To

ARUNDHATI AND RITA VINOD

ELIZABETH REAGLE

# Contents

# List of Figures

# List of Tables

# Preface

This book provides a detailed accounting of how downside risk can enter a portfolio, and what can be done to identify and prepare for the downside. We take the view that downside risk can be incorporated into current methods of stock valuation and portfolio management. Therefore we introduce commonly used theories in order to show how the status quo often misses the downside, and where to include it.

The importance of downside risk is evident to any investor in the market. No one complains about unexpected gains, while unexpected losses are painful. Traditional theories of risk measurements treat volatility on either side equally. To include downside risk, we divide the discussion into three parts. Part 1 covers the current theories of risk measurement and management, and includes Chapters 1, 2, and 3. Part 2 presents the violations of this theory and the need to include the downside, covered in Chapters 4, 5, 6, and 7. Part 3 covers the quantitative and programming techniques to make risk measurement more precise in Chapters 8 and 9. The discussions in these two chapters are not for the casual reader but for those who perform the calculations and are curious about the pitfalls to avoid. Chapter 10 concludes with a summarized treatment of downside risk.

Our aim in this book will have succeeded if the reader takes a second look when investing and asks, "Have I considered the downside?"

CHAPTER 1

# Quantitative Measures of the Stock Market

## 1.1 PRICING FUTURE CASH FLOWS

Our first project in order to understand stock market risk, particularly downside risk, is to identify exactly what the stock market is and determine the motivation of its participants. Stock markets at their best provide a mechanism through which investors can be matched with firms that have a productive outlet for the investors' funds. It is a mechanism for allocating available financial funds into appropriate physical outlets. At the individual level the stock market can bring together buyers and sellers of investment instruments. At their worst, stock markets provide a platform for gamblers to bet for or against companies, or worse yet, manipulate company information for a profit. Each investor in the stock market has different aims, risk tolerance, and financial resources. Each firm has differing time horizons, scale of operations, along with many more unique characteristics including its location and employees.

So when it comes down to it, there need not be a physical entity that is *the* stock market. Of course, there are physical stock exchanges for a set of listed stocks such as the New York Stock Exchange. But any stock market is the combination of individuals. A trading floor is not a stock market without the individual investors, firms, brokers, specialists, and traders who all come together with their individual aims in mind to find another with complementary goals. For any routine stock trade, there is one individual whose goal it is to invest in the particular company's stock on the buy side. On the sell side, there is an individual who already owns the stock and wishes to liquidate all or part of the investment. With so much heterogeneity in the amalgam that is the stock market, our task of finding a common framework for all players seems

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

intractable. However, we can find a number of features and commonalities which can be studied in a systematic manner.

1. The first among these commanalities is the time horizon. For any investor, whether saver or gambler, money is being invested in stock for some time horizon. For a young worker just beginning to save for his retirement through a mutual fund, this time horizon could be 30 years. For a day trader getting in and out of a stock position quickly, this time horizon could be hours, or even minutes. Whatever the time horizon, each investor parts with liquid assets for stock, intending to hold that stock for sale at a future date $T$. When we refer to prices, we will use the notation, $P_T$, where the subscript represents the time period for which the price applies. For example, if the time $T$ is measured in years, $P_0$ denotes the current price (price today) and $P_5$ denotes the price five years from now.

2. The next commonality is that all investors expect a return on their investment. Since investors are parting with their money for a time, and giving up liquidity, they must be compensated. We will use $r_T$ to represent the return earned on an investment of $T$ years. Therefore $r_1$ would be the return earned on an investment after one year, $r_5$ on an investment after five years, and so on. Using our first two rules, we can derive a preliminary formula to price an asset with a future payment of $P_T$ which returns exactly $r_T$ percent per year for $T$ years. We use capital $T$ for the maturity date in this chapter. Lowercase $t$ will be used as a variable denoting the current time period.

We start with the initial price, $P_0$, paid at the purchase date. After the first year, the investor would have the initial investment plus the return:

$$P_0(1+r_T). \tag{1.1.1}$$

For the second year, the return is compounded on the value at the end of the first year:

$$P_0(1+r_T)(1+r_T) \quad \text{or} \quad P_0(1+r_T)^2. \tag{1.1.2}$$

Thus the price that the investor must be paid in year $T$ to give the required return is

$$P_0(1+r_T)(1+r_T)\ldots(1+r_T) = P_0(1+r_T)^T = P_T. \tag{1.1.3}$$

This is the formula to calculate a *future value* with compound interest each period. For example, interest compounded quarterly for two years would use the quarterly interest rate (annual rate divided by 4) and $T = 8$ periods.

Now let us find the fair price for this asset today, $P_0$, that will yield a return of exactly $r_T$ every year for $T$ years. Clearly, we would just need to divide through by $(1 + r_T)^T$ to obtain

$$\text{Present value} = P_v = \frac{P_T}{(1+r_T)^T}. \tag{1.1.4}$$

This tells us that given a return $r_T$ of periodic future payoffs, we can find the present price in order to yield the correct future price $P_T$ at the end of the time horizon of length $T$. This formula is called the discounting, or present value, formula. The discounting formula is the basis of any pricing formula of a financial asset. Stocks and bonds, as well as financial derivatives such as options and futures, and hybrids between various financial instruments all start with the discounting formula to derive a price, since they all involve a time interval before the final payment is made.

Lest we get too comfortable with our solution of the price so quickly and easily, this misconception will be shattered with our last common feature of all investments.

3. All investments carry risk. In our discounting formula, there are only two parameters to plug in to find a price, namely the price at the end of time horizon or $P_T$ and the return $r_T$. Unfortunately, for any stock the future payment $P_T$ is not known with certainty. The price at which a stock is sold at time $T$ depends on many events that happen in the holding duration of the stock. Company earnings, managerial actions, taxes, government regulations, or any of a large number of other random variables will affect the price $P_T$ at which someone will be able to sell the stock.

With this step we have introduced uncertainty. What price $P_0$ should you pay for the stock today under such uncertainty? We know it is the discounted value of $P_T$, but without a crystal ball that can see into the future, $P_0$ is uncertain. There are a good many investors who feel this is where we should stop, and that stock prices have no fundamental value based on $P_T$. Many investors believe that the past trends and patterns in price data completely characterize most of the uncertainty of prices and try to predict $P_T$ from data on past prices alone. These investors are called *technical analysts*, because they believe investor behavior is revealed by a time series of past prices. Some of these patterns will be incorporated in time series models in Section 4.1.

Another large group seeks to go deeper into the finances and prospects of the corporations to determine the fair value for the stock price represented by $P_T$. This group is called fundamental value investors, because they attempt to study intrinsic value of the firm. To deal with the fact that future prices are not known, fundamental value investors must base their value on risk, not uncertainty. By characterizing "what is not known" as risk, we are assuming that while we do not know exactly what will happen in the future, we do know what is possible, and the relative likelihoods of those possibilities. Instead of being lost in a random world, a study of risk lets us categorize occurrences and allows the randomness to be measured.

Using risk, we can derive a fundamental value of a firm's stock. As a stockholder, one has a claim of a firm's dividends, the paid out portion of net earnings. These dividends are random, and denoted as $D^s$ for dividends in state $s$. This state is a member of a long list of possible occurrences. Each state represents a possibly distinct level of dividends, including extraordinarily high, average, zero, and bankruptcy. The probability of each state is denoted by $\pi^s$ for $s = 1, 2, \ldots, S$, where $S$ denotes the number of states considered. The more likely a state is, the higher is its probability. An investor can calculate the expected value of dividends that will be paid by summing each possible level of dividends multiplied by the corresponding probability:

$$E(D) = \sum_{s=1}^{S} \pi^s D^s, \quad \text{where} \sum_{s=1}^{S} \pi^s = 1, \tag{1.1.5}$$

where $E$ is the expectations operator and $\Sigma$ is the summation operator. Outcomes that are more likely are weighted by a higher probability and affect the expected value more. The expected value can also be thought of as the average value of dividends over several periods of investing, since those values with higher probabilities will occur more frequently than lower probability events. The sum of the probabilities must equal one to ensure that there is an exhaustive accounting of all possibilities.

Using the framework of risk and expected value, we can define the price of a stock as the discounted value of expected dividends at future dates, namely the cash flow received from the investment:

$$P_0 = \sum_{t=1}^{T} \frac{E(D)_T}{(1 + r_T)^T}, \tag{1.1.6}$$

where each period's expected dividends are discounted the appropriate number of time periods $T$ by the compound interest formula stated in (1.1.6).

Formula (1.1.6) for the stock price is more useful, since it is based on the financials of a company instead of less predictable stock prices. One must forecast dividends, and thus have a prediction of earnings of a company. This approach is more practical since the other formula (1.1.4) was based on an unknown future price. One may ask the question: How can we have two formulas for the same price?

However, both formulas (1.1.4) and (1.1.6) are identical if we assume that investors are investing for the future cash flow from holding the stock. Our price based on discounted present value of future dividends looks odd, since it appears that we would have to hold the stock indefinitely to receive the entire value. What if we sell the stock after two years for a stock paying quarterly dividends (8 quarters)?

The value of our cash flow after including the end point price would be

$$P_0 = \frac{E(D)_1}{(1+r_1)^1} + \frac{E(D)_2}{(1+r_2)^2} + \ldots + \frac{E(D)_8}{(1+r_8)^8} + \frac{P_8}{(1+r_8)^8}, \tag{1.1.7}$$

where $r_t$ is the quarterly return for the quarter $t$ with $t = 1, \ldots, 8$. But realizing that the buyer in quarter 8 is purchasing the subsequent cash flows until they sell the stock one year later we have

$$\frac{P_8}{(1+r_8)^8} = \frac{E(D)_9}{(1+r_9)^9} + \frac{E(D)_{10}}{(1+r_{10})^{10}} + \frac{E(D)_{11}}{(1+r_{11})^{11}} + \frac{E(D)_{12}}{(1+r_{12})^{12}} + \frac{P_{12}}{(1+r_{12})^{12}}. \tag{1.1.8}$$

And so on it goes. So that recursively substituting the future prices yields $P_0$ equal to all future discounted dividends. This means that even for a stock not currently paying any dividend, we can use the same discounting formula. The stock must eventually pay some return to warrant a positive price.

Using the value of dividends to price a security may be unreliable, however. The motivation for a company issuing dividends is more complex than simply paying out the profits to the owners (see Allen and Michaely, 1995, for a survey of dividend policy). First, growth companies with little excess cash flows may not pay any dividend in early years. The more distant these dividends are, the harder they are to forecast. Dividends also create a tax burden for the investor because they are taxed as current income, whereas capital gains from holding the stock are not taxed until the stock is sold. This double taxation of dividends at the corporate and individual levels leads many to question the use of dividends at all, and has led many firms to buy back shares with excess cash rather than issue dividends. Also dividends are a choice made by the firm's management. Bhattacharya (1979) shows how dividends can signal financial health of a company, so firms are seen paying out cash through dividends and then almost immediately issuing more shares of stock to raise capital.

Alternatively, since dividend amounts are chosen by the management of the firm and may be difficult to forecast, price can be modeled as the present value of future earnings, ignoring the timing of exactly when they are paid out in the form of dividends. This model assumes that earnings not paid out as dividends are reinvested in the company for $T$ years. So that if they are not paid in the current period, they will earn a return so that each dollar of "retained earnings" pays $1 + r_T$ next period. This makes the present value of expected earnings identical to the present value of dividends. Hence a lesson for the management is that they better focus on net earnings rather than window dressing of quarterly earnings by changing the dividend payouts and the timing of cash flows. The only relevant figure for determining the stock price is the bottom line of net earnings, not how it is distributed.

## 1.2 THE EXPECTED RETURN

Once the expected cash flows have been identified, one needs to discount the cash flows by the appropriate return, $r_T$. This is another value in the formula that looks deceptively simple. In this section we discuss several areas of concern when deciding the appropriate discount rate, namely its term, taxes, inflation, and risk, as well as some historical trends in each area.

The first building block for a complete model of returns is the risk-free rate, $r_T^{\mathrm{f}}$. This is the return that would be required on an investment maturing in time $T$ with no risk whatsoever. This is the rate that is required solely to compensate the investor for the lapse of time between the investment and the payoff. The value of the risk-free rate can be seen as the equilibrium interest rate in the market for loanable funds or government (FDIC) insured return:

*The Borrower*

A borrower will borrow funds only if the interest rate paid is less than or equal to the return on the project being financed. The higher the interest rate, the fewer the projects that will yield a high enough return to pay the necessary return.

*The Lender*

A lender will invest funds only if the interest rate paid is enough to compensate the lender for the time duration. Therefore, as the interest rate increases, more investors will be willing to forgo current consumption for the higher consumption in the future.

*The Market*

The equilibrium interest rate is the rate at which the demand for funds by borrowers in equal to the supply of funds from lenders; it is the market clearing interest rate in the market for funds. As can be seen from the source of the demand and supply of funds, this will be the return of the marginal project being funded (the project just able to cover the return), and at the same time this will be the time discount rate of the marginal investor.

A common observation about the interest rate is that the equilibrium return tends to rise as the length of maturity increases. Plotting return against length of maturity is known as the yield curve. Because an investor will need more enticement to lend for longer maturities due to the reduced liquidity, the yield curve normally has a positive slope. A negative slope of the yield curve is seen as a sign that investors are expecting a recession (reducing projected future returns) or that they are expecting high short-term inflation.

To see how inflation affects the required return for an investor, we can augment our return to get the nominal interest rate:

$$r_T^{\mathrm{n}} = r_T^{\mathrm{f}} + \pi_T^{\mathrm{e}}, \tag{1.2.1}$$

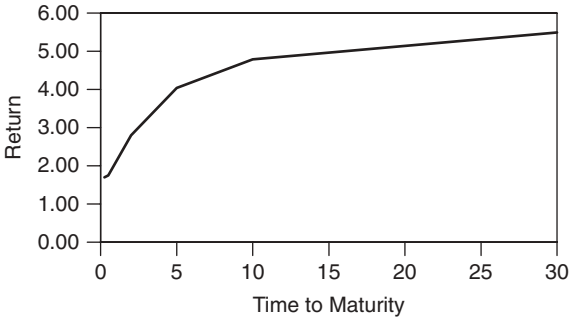



**Figure 1.2.1** Yield curve for US treasuries, June 1, 2002

where $\pi_T^e$ is the expected rate of inflation between time 0 and time $T$. For an investor to be willing to supply funds, the nominal return must not only compensate for the time the money is invested, it must also compensate for the lower value of money in the future.

For example, if \$100 is invested at 5% interest with an expected inflation rate of 3% in January 2002, payable in January 2003, the payoff of the investment after one year is \$105. But this amount cannot buy what \$105 will buy in 2002. An item that was worth \$105 in 2002 will cost \$105(1.03) = \$108.15 in 2003. To adjust for this increase in prices, to the nominal interest rate is added the cost of inflation to the return.

One may wonder about the extra 15 cents that the formula above does not include. According to the formula for nominal rate, an investor would get 5% + 3% = 8%, or \$108 at the payoff date. That is because the usual formula for nominal rate is an approximation: it only adjusts for inflation of the principal but not the interest of the loan. The precise formula will be

$$r_T^n = r_T^f + \pi_T^e + r_T^n \pi_T^e \quad \text{or} \quad r_T^n = \frac{r_T^f + \pi_T^e}{1 - \pi_T^e}. \tag{1.2.2}$$

As the additional term is the interest rate times the expected inflation rate, two numbers are usually less than one. Unless either the inflation rate or the interest rate is unusually high, the product of the two is small and the approximate formula is sufficient.

Our next adjustment comes from taxes. Not all of the nominal return is kept by the investor. When discounting expected cash flows then, the investor must ensure that the after-tax return is sufficient to cover the time discount:

$$r_T^{at} = r_T^n (1 - \tau), \tag{1.2.3}$$

where $r_T^{at}$ is the after-tax return and $\tau$ is the tax rate for an additional dollar of investment income.

It is important to note that taxes are applied to the nominal return, not the real return (return with constant earning power). This makes an investor's forecast of inflation crucial to financial security.

Consider the following two scenarios of an investment of $100 with a nominal return of 12.31% at a tax rate of 35% of investment income. The investor requires a risk-free real rate of interest of 5% and expects inflation to be 3%. The investment is to be repaid in one year.

*Scenario 1—Correct Inflation Prediction*

If inflation over the course of the investment is, indeed, 3%, then everything works correctly. The investor is paid $112.31 after one year, $12.31(0.35) = $4.31 is due in taxes, so the after-tax amount is $108.00. This covers the 5% return plus 3% to cover inflation.

*Scenario 2—Underestimation of Inflation*

If actual inflation over the course of the investment turns out to be 10%, the government does not consider this an expense when it comes to figuring taxable income. The investor receives $112.31, which nominally seems to cover inflation, but then the investor must pay the same $4.31 in taxes. The $108.00 remaining is actually worth less than the original $100 investment since the investor would have had to receive at least $110 to keep the same purchasing power as the original $100.

Scenario 2 shows how unexpectedly high inflation is a transfer from the investor, who is receiving a lower return than desired to the borrower, who pays back the investment in dollars with lower true value.

The final element in the investor's return is the risk premium, $\theta$, so that the total return is

$$r_T^{\mathrm{n}} = \frac{r_T^{\mathrm{f}} + \pi_T^{\mathrm{e}} + \theta}{1 - \tau}. \tag{1.2.4}$$

The risk premium is compensation for investing in a stock where returns are not known with certainty. The value of the risk premium is the most nebulous of the parameters in our return formula, and the task of calculating the correct risk premium and methods to lower the risk of an investment will be the subject of much of the balance of this book. At this point we will list some of the important questions in defining risk, leaving the detail for the indicated chapter.

1. How do investors feel about risk? Are they fearful of risk such that they would take a lower return to avoid risk? Or do they appreciate a bit of risk to liven up their life? Perceptions of investors to risk will be examined in Chapter 6.

2. Is risk unavoidable, or are there investment strategies that will lower risk? Certainly investors should not be compensated for taking on risk that could have been avoided. The market rarely rewards the unsophisticated investor (Chapters 2 and 3).
3. Is the unexpected return positive or negative? Most common measurements of risk (e.g., standard deviation) consider unexpected gains and losses as equally risky. An investor does not have to be enticed with a higher return to accept the "risk" of an unexpected gain. This is evidenced by the fact that individuals pay for lottery tickets, pay high prices for IPOs of unproven companies, and listen intently to rumors of the next new fad that will take the market by storm. We explain how to separate upside and downside risk in Chapter 5, and evidence of the importance of the distinction in Chapters 7, 8, and 9.

## 1.3 VOLATILITY

In order to develop a measure of the risk premium, we must first measure the volatility of stock returns. The term "volatility" suggests movement and change; therefore any measurement of volatility should be quantifying the extent to which stock returns deviate from the expected return, as discussed in Section 1.2. Quantifying change, however, is not a simple task. One must condense all the movements of a stock throughout the day, month, year, or even decade, into one measure. The search for a number that measures the volatility of an investment has taken numerous forms, and will be the subject of several subsequent chapters since this volatility, or movement, of stock prices, is behind our notion of risk. Without volatility, all investments are safe. With volatility, stocks yield gains and losses that deviate from the expected return.

As an example, we will use the annual return for the S&P 500 index from 1990 through 2000 shown in Table 1.3.1. The average annual return for this time period is 13.74%. Compare this return to the return of U.S. three-month Treasury bills for the same time period (Table 1.3.2) that is on average 4.94%.

**Table 1.3.1 S&P 500 Index Annual Return**

| 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|
| –6.56% | 26.31% | 4.46% | 7.06% | –1.54% | 34.11% | 20.26% | 31.01% | 26.67% | 19.53% | –10.14% |

**Table 1.3.2 US Three-Month Treasury Bills Return**

| 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|
| 7.50% | 5.38% | 3.43% | 3.00% | 4.25% | 5.49% | 5.01% | 5.06% | 4.78% | 4.64% | 5.82% |

The average return of the stock index is almost three times the average return on T-bills. There must be a reason, or there should be no investor buying T-bills. Both are in dollars, so inflation is the same. Both are under the same tax system, although T-bill interest is taxed as income, and stock returns as capital gains, but that should give a higher return to T-bills. Capital gains taxes are usually lower than income taxes, and they can be delayed so they are even lower in present discounted value.

Common sense tells us the reason for the difference in returns is the volatility. While the T-bill return is consistently around 4% or 5%, the stock return has wide swings in the positive and negative range. In a free market economy, if investment in risky assets creates economic growth, new jobs, and new conveniences, these risky activities have to be rewarded. Otherwise, there will be no one taking the risks. This means the market forces must reward a higher return for investors in certain wisely chosen risky activities. Such higher return is called risk premium. Volatility is therefore very important in determining the amount of risk premium applied to a financial instrument.

To measure volatility, the simplest measure would be the range of returns, when the range is defined as the highest return less the lowest return. The S&P 500 has a range of [34.11 – (–10.14)] = 44.25. For T-bills, the range is [7.5 – 3] = 4.5. The range of returns is much larger for the S&P 500, showing the higher volatility.

The range has the benefit of ease of calculation, but the simplest measure is not always the best. The problem with the range is that it only uses two data points, and these are the two most extreme data points. This is problematic because the entire measure might be sensitive to outliers, namely to those extreme years that are atypical. For instance, a security will have the same average return and range as the S&P 500 if returns for nine years were 13.74%, the next year has a money-losing return of –8.39% and the next year has a spectacular return of 35.86%. But the volatility of this security is clearly not identical to the S&P 500, though the range is the same. This security is very consistent because only two years have extreme returns.

In order to take all years into account, one simply takes the deviations from the mean of each year's returns

$$r_t - \mu, \tag{1.3.1}$$

where $\mu$ is the average return for the respective security for each time period $t$. To condense these deviations into one measure, there are two common approaches. Both approaches try to put a single value on changes of the returns. Since values above or below the mean are both changes, the measure needs to treat both positive and negative values of deviations as an increase in volatility.

The mean absolute deviation (MAD) does this by taking the absolute value of the deviations, and then a simple average of the absolute values,

**Table 1.3.3 S&P 500 Index Deviations from Mean**

| 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|
| –20.30% | 12.57% | –9.28% | –6.68% | –15.28% | 20.37% | 6.52% | 17.27% | 12.93% | 5.79% | –23.88% |

$$\text{MAD} = \frac{\sum_t |r_t - \mu|}{T}, \tag{1.3.2}$$

where $\Sigma_t$ denotes sum from $t = 1$ to $t = T$ and where $T$ is again the total number of years.

For the S&P 500, the mean absolute deviation is 13.72%. For the T-bill series, the mean absolute deviation is 0.84%. This shows the dramatic difference in volatility between the two securities.

The other way to transform the deviations to positive numbers is to square them. This is done with the variance, $\sigma^2$:

$$\sigma^2 = \frac{\sum_t (r_t - \mu)^2}{T}. \tag{1.3.3}$$

*Note*: This variance formula is often adjusted for small samples by replacing the denominator by $(T - 1)$. A discussion of sampling is in Chapter 9.

The variance formula implicitly gives larger deviations a larger impact on volatility. Therefore 10 years of a 2% deviation ($0.02^2 \times 10 = 0.004$) does not increase variance as much as one year of a 20% deviation ($0.2^2 = 0.04$).

The variance for the S&P 500 is 0.0224, for T-bills, 0.0001. It is common to present the standard deviation, which is the square root of the variance so that the measure of volatility has the same units as the average. For the S&P 500, the standard deviation is 0.1496, for T-bills, 0.0115.

The advantage in using the standard deviation is that all available data can be utilized. Also some works have shown that alternate definitions of a deviation can be used. Rather than strictly as deviations from the mean, risk can be defined as deviations from the risk-free rate (CAPM, ch. 2). Tracking error (Vardharaj, Jones, and Fabozzi, 2002) can be calculated as differences from the target return for the portfolio. When an outside benchmark is used as the target, the tracking error is more robust to prolonged downturns, which otherwise would cause the mean to be low in standard deviation units. Although consistent loss will show a low standard deviation, which is the worst form of risk for a portfolio, it will show up correctly if we use tracking error to measure volatility.

Other methods have evolved for refining the risk calculation. The intraday volatility method involves calculating several standard deviations throughout the day, and averaging them. Some researchers are developing methods of

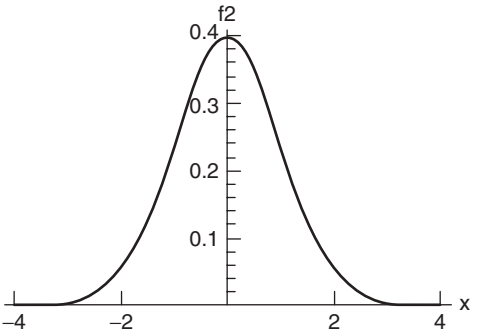



**Figure 1.3.1** Probability density function for the standard normal N(0, 1) distribution

using the intraday range of prices in the calculation of standard deviations over several days. By taking the high and low price instead of the opening and closing prices, one does not run the risk of artificially smoothing the data and ignoring the rest of the day. The high and low can come at any time during the day.

Once the expected return and volatility of returns are calculated, our next step is to understand the distribution of returns. A probability distribution assigns a likelihood, or probability, to small adjacent ranges of returns. Probability distributions on continuous numbers are represented by a probability density function (PDF), which is a function of the random variable $f(x)$. The area under the PDF is the probability of the respective small adjacent range of the variable $x$. One commonly used distribution is the normal distribution having mean $\mu$ and variance $\sigma^2$, $N(\mu, \sigma^2)$, written

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{((x-\mu)/\sigma)^2/2}, \tag{1.3.4}$$

where $\mu$ is the mean of the random variable $x$ and $\sigma$ is the standard deviation. Once we know these two parameters, we know the entire probability distribution function (pdf) of $N(\mu, \sigma^2)$.

It can be seen from the normal distribution formula (1.3.4) why the standard deviation $\sigma$ is such a common measure of dispersion. If one assumes that returns follow the normal distribution, with the knowledge of only the average of returns ($\mu$) and the standard deviation ($\sigma$), all possible probabilities can be determined from widely available tables and software sources. Therefore an infinite number of possibilities can be calculated from only two statistics. This is a powerful concept. (We will discuss the validity of using the normal distribution for stock returns in Chapter 4.)

The normal distribution is a common distribution because it seems to possess several characteristics that occur in nature. The normal distribution has most of the probability around the average. It is symmetrical, meaning the

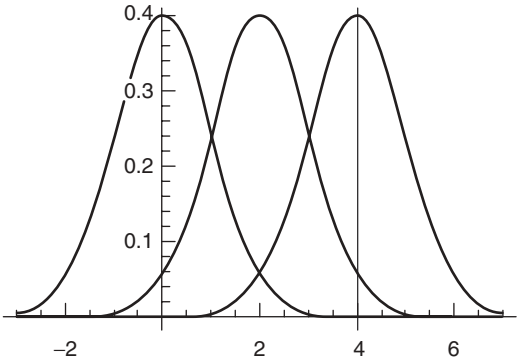



**Figure 1.3.2** Normal distribution with change in the average $\mu(= 0, 2, 4)$ with $\sigma = 1$

probability density function above and below the average are mirror images. The probability of getting outcomes an extreme distance above or below the average are progressively unlikely, although the density function never goes to zero, so all outcomes are possible. Children's growth charts, IQ tests, and bell curves are examples of scales that follow the normal distribution.

Since one need only know the average and standard deviation to draw a specific normal distribution, it is a useful tool for understanding the intuition of expected value and volatility. Probability statements can be made in terms of a certain number of standard deviations from the mean. There is a 68.3% probability of $x$ falling within one standard deviation of the mean, 95.5% probability two standard deviations of the mean, 99.74% probability three standard deviation from the mean, and so on.

The normal probability can change dramatically with changes of the parameters. Increases in the average will shift the location of the normal distribution. Increases in the standard deviation will widen the normal distribution. Decreases in the standard deviation will narrow the distribution.

Because the normal distribution changes with a change in the average or standard deviation, a useful tool is standardization. This way the random variable can be measured in units of the number of standard deviations measured from the mean:

$$z = \frac{x - \mu}{\sigma}. \tag{1.3.5}$$

If $x$ is normally distributed with mean $\mu$ and standard deviation $\sigma$, then the standardized value $z$ will be standard normally distributed with mean of zero, and standard deviation equal to one. In statistical literature this relation is often stated by using the compact notation: $x \sim N(\mu, \sigma^2)$ and $z \sim N(0, 1)$. It can be verified by some simple rules on the expectations (averages) of random numbers stated below. Given $a$ and $b$ as some constant real numbers, we have:

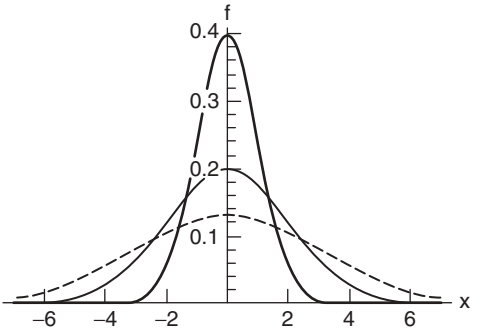



**Figure 1.3.3** Normal distribution with change in σ, the standard deviation

1. If the average of $x = \mu$, then the average of $a(x) = a(\mu)$.
2. If the average of $x = \mu$, then the average of $(x + b) = \mu + b$. Therefore the average of $x - \mu = \mu - \mu = 0$.
3. If the standard deviation of $x = \sigma$, then the standard deviation of $a(x) = a(\sigma)$. (Note: The variance of $a(x) = a^2\sigma^2$.)
4. If the standard deviation of $x = \sigma$, then the standard deviation of $(x + b) = \sigma$. Therefore the standard deviation of $(x - \mu)/\sigma = (1/\sigma)\sigma = 1$.

Through standardization, tables of the area under the standard normal distribution can be used for normal distributions with any average and standard deviation. To use the tables, one converts the $x$ value under the normal distribution to the standardized $z$ statistic under the standard normal and looks up the $z$ value in the table. The probability relates back to the original $x$ value, which is then the number of standard deviations from the mean. With the wide availability of Excel software workbooks, nowadays it is possible to avoid the normal distribution tables and get the results directly for $x \sim N(\mu, \sigma^2)$ or for $z \sim N(0, 1)$.

Therefore, for the normal distribution, mean and standard deviation are the end of the story. Since symmetry is assumed, the distinction of downside risk is moot. However, for this reason the normal distribution is not always a practical assumption, but it provides a valuable baseline against which to measure downside adjustments. The next section provides a dynamic framework of modeling stock returns following the normal distribution.

## 1.4 MODELING OF STOCK PRICE DIFFUSION

A probability distribution gives the likelihood of ranges of returns. If one assumes the normal distribution, then the distribution is completely defined by its average and standard deviation. Knowing this, one can model the dis-

crete movement of a stock price over time through the diffusion equation, which combines the average return μ and the volatility measured by the standard deviation σ.

$$\frac{\Delta S}{S} = \mu \Delta t + \sigma z \sqrt{\Delta t}, \tag{1.4.1}$$

where Δ is the difference operator ($\Delta S = \Delta S_t = S_t - S_{t-\Delta t}$), $S$ is the stock price, $\Delta t$ is the time duration, and $z$ is N(0, 1) variable. Note that $\Delta S/S$ is the relative change in the stock price, and the relative changes times 100 is the percentage change. Equation (1.4.1) seeks to explain how relative changes are diffused as the time passes around their average, subject to random variation.

The diffusion equation (1.4.1) has two parts: the first part of the percentage change is the average return μ per time period (or drift), multiplied by the number of time periods that have elapsed; the second part is the random component that measures the extent to which the return can deviate from the average. Also we see that the standard deviation is increased by the square root of the time change. The root term arises because it can be shown that (1.4.1) follows what is known as a random walk (also known as Brownian motion, or Weiner process). If $S_t$ follows a random walk, it can be written as $S_t = S_{t-1} + \delta + \varepsilon$, where the value of the stock price at any point in time is the previous price, plus the drift (= δ), plus some random shock (= ε). The diffusion process is obviously more general than a simple random walk with drift. The more time periods out you go, the more random shocks are incorporated into the price. Since each one of these shocks has its own variance, the total variance for a length of time of $\Delta t$ will be $\sigma^2 \Delta t$. Thus the standard deviation will be the square root of the variance.

The cumulative effect of these shocks from (1.4.1) can be seen by performing a simple simulation starting at a stock price of 100 for a stock with an average return of 12% per year and a standard deviation of 5% and the following random values for $z$. For a complete discussion of simulations, see Chapter 9.

At any time $t$, the price the next day ($\Delta t = 1/365 = 0.0027$) will be

$$S_{t+1} = S_t + \Delta S_t = S_t + \mu S_t \Delta t + \sigma S_t z_t \sqrt{\Delta t}. \tag{1.4.2}$$

Simulating random numbers for 30 days yields the stock prices in Table 1.4.1.

Looking at the stock prices in a graph, we can see that simulation using a random walk with drift gives a plausible series of stock prices. A few other insights can be gained from the graph. We can see that a random walk, as the name implies, is a movement from each successive stock price, not reverting back to an average stock price (this is the basis for another term associated with random walk: nonstationary). Also, as the stock price gets higher, the movements get larger since the price is the percentage of a larger base.

**Table 1.4.1 Simulated Stock Price Path**

| $t$ | $z_t$ | $\Delta S_t$ | $S_t$ |
|---|---|---|---|
| 0 | | | 100.00 |
| 1 | 1.24 | 0.35 | 100.35 |
| 2 | 0.23 | 0.09 | 100.45 |
| 3 | –1.08 | –0.25 | 100.20 |
| 4 | 0.36 | 0.13 | 100.33 |
| 5 | –0.21 | –0.02 | 100.30 |
| 6 | 0.07 | 0.05 | 100.35 |
| 7 | –0.37 | –0.06 | 100.29 |
| 8 | –0.30 | –0.05 | 100.25 |
| 9 | 0.37 | 0.13 | 100.38 |
| 10 | –0.03 | 0.02 | 100.40 |
| 11 | –0.16 | –0.01 | 100.39 |
| 12 | 0.61 | 0.19 | 100.58 |
| 13 | 0.03 | 0.04 | 100.62 |
| 14 | –0.44 | –0.08 | 100.54 |
| 15 | 0.02 | 0.04 | 100.58 |
| 16 | –0.67 | –0.14 | 100.44 |
| 17 | 0.19 | 0.08 | 100.52 |
| 18 | 1.04 | 0.30 | 100.82 |
| 19 | 1.42 | 0.40 | 101.23 |
| 20 | 1.49 | 0.42 | 101.65 |
| 21 | 0.48 | 0.16 | 101.81 |
| 22 | –0.52 | –0.10 | 101.71 |
| 23 | 0.53 | 0.17 | 101.88 |
| 24 | 1.41 | 0.41 | 102.29 |
| 25 | 2.05 | 0.58 | 102.86 |
| 26 | 0.45 | 0.15 | 103.02 |
| 27 | –0.79 | –0.18 | 102.84 |
| 28 | 0.78 | 0.24 | 103.08 |
| 29 | 2.05 | 0.58 | 103.67 |
| 30 | –0.58 | –0.12 | 103.54 |

This general diffusion model of stock prices has gone through many alterations for specific stock pricing situations. The descriptions that follow cover only a few of these adaptations.

### 1.4.1 Continuous Time

For empirical use, or for producing simulations, we can only work with discrete time changes, but theoretically a continuous time approach (as $\Delta t \to 0$) can model the path of stock prices at each moment in time. This often simplifies calculations and yields more elegant results. The continuous time diffusion equation is

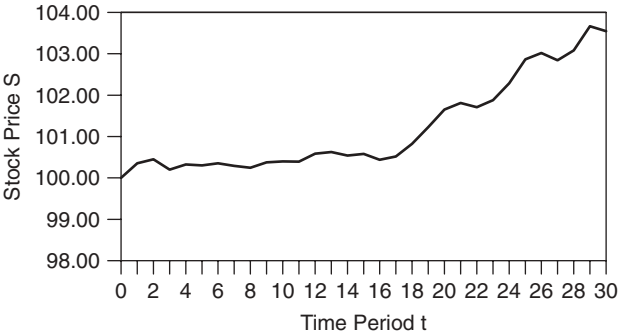



**Figure 1.4.1** Simulated stock prices over 30 time periods

$$\frac{dS}{S} = \mu dt + \sigma dz, \tag{1.4.3}$$

where $d$ denotes an instantaneous change. The $dz$ in (1.4.3) represents a standard Wiener process or Brownian motion (Campbell et al., 1997, p. 344) that is a continuous time analogue of the random walk mentioned above.

### 1.4.2 Jump Diffusion

The jump diffusion process recognizes the fact that not all stock movements follow a continuous smooth process. Natural disasters, revelation of new information, and other shock can cause a massive, instantaneous revaluation of stock prices. To account for these large shocks, the normal diffusion is augmented with a third term representing these jumps:

$$\frac{dS}{S} = (\mu - \lambda k)dt + \sigma dz + dq, \tag{1.4.4}$$

where $\lambda$ is the average number of jumps per unit of time, $k$ is the average proportionate change of the jump (the variance of the jump is $\delta^2$ to be used later), and $dq$ is a Poisson process. The adjustment to the drift term ensures that the total average return is still $\mu$: $(\mu - \lambda k)$ from the usual random walk drift, plus $\lambda k$ from the jump process leading to a cancellation of $\lambda k$.

In the Poisson process, the probability of $j$ number of jumps in $T$ time periods is determined by the Poisson discrete probability function

$$P(j) = \frac{e^{-\lambda T}(\lambda T)^j}{j!}. \tag{1.4.5}$$

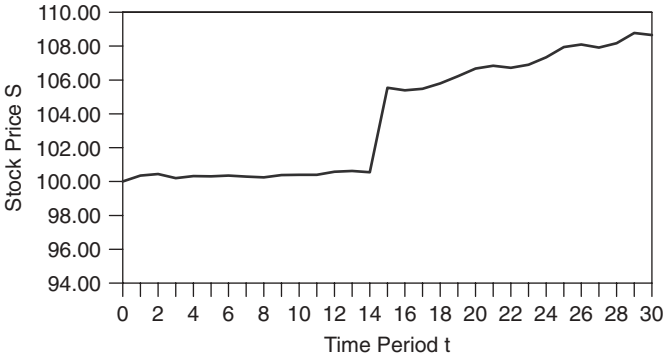



**Figure 1.4.2** Simulated stock prices for jump diffusion process

A graph of the same stock diffusion in Table 1.4.1 with a jump of \$5 occurring on the 15th day is given in Figure 1.4.2. As can be seen from the graph, a jump will increase the volatility of the stock returns dramatically, depending on the volatility of the jump and the average number of jumps that occur. The total variance of the process is then $\sigma^2 + \lambda\delta^2$ per unit of time. This can also be used as a method to model unexpected downside shocks through a negative jump.

### 1.4.3 Mean Reversion in the Diffusion Context

For stock prices that should be gaining a return each period, a random walk with drift seems a reasonable model for stock prices. For some investments, however, it does not seem reasonable that their price should constantly be wandering upward. Interest rates or the real price for commodities, such as oil or gold, are two such examples in finance of values that are not based on future returns, and thus have an intrinsic value which should not vary over time. Prices that revert back to a long-term average are known as mean reverting (or stationary).

Mean reversion can be modeled directly in the diffusion model

$$dS = \eta(\overline{S} - S)dt + \sigma dz, \tag{1.4.6}$$

where $\overline{S}$ is the average value of the financial asset and $0 < \eta < 1$ is the speed at which the asset reverts to its mean value. Since a mean reverting process is centered around $\overline{S}$ and always has the same order of magnitude, the diffusion need not be specified in terms of percentage changes. Graphing the diffusion using the random numbers as above and an average price of \$100 gives the path for two different speeds $\eta = 0.8$ and $\eta = 0.2$ of mean reversion.

From Figure 1.4.3 both processes stay near 100. The solid line path with the higher reversion speed ($\eta = 0.8$) snaps back to 100 quicker, even after large shocks to the average price level. For the stock with the lower reversion speed,

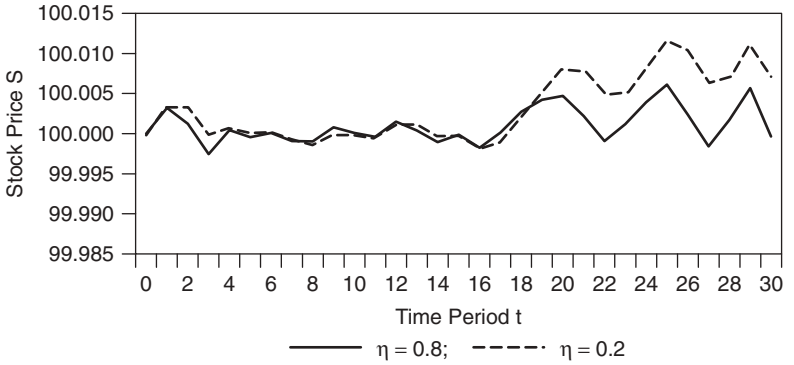



**Figure 1.4.3** Simulated stock prices for two mean-reverting processes (——— $\eta = 0.8$; ------ $\eta = 0.2$)

large increases or decreases linger because the stock takes smaller steps back to its average price similar to the dashed line. We discuss mean reversion in a general context in Section 4.1.2.

### 1.4.4 Higher Order Lag Correlations

In the mean reverting process, the change in a stock price $S_t$ is affected by how far the previous period is from the mean $(S_{t-1} - \mu)$. But with the high frequency at which stock data is available (e.g., hourly), it is realistic that correlations could last longer than one period. Would one expect that a boost in sales at the end of February would immediately be gone in the beginning of March? Would a news report at 10:00 am on Tuesday morning be completely reflected in the stock price by 10:01 am? No.

A method to account for these holdovers from past periods $(S_{t-2}, S_{t-3}, \ldots)$ is the ARIMA model. The AR stands for autoregressive, or the previous period returns that are directly affecting the current price. The MA stands for moving average, or the previous period shocks that are directly affecting the current price. The "I" in the middle of ARIMA stands for integrated, which is the number of times the data must be transformed by taking first differences $(S_t - S_{t-1})$ over time. If $L$ denotes the lag operator $LS_t = S_{t-1}$, $(S_t - S_{t-1})$ becomes $(1 - L)S_t = \Delta S_t$. If $(1 - L) = 0$ is a polynomial in the lag operator, its root is obviously $L = 1$, which is called the unit root. Most stocks have unit root and are said to be *integrated of order 1*, I(1), meaning that one time difference is necessary to have a stationary process. Since taking returns accomplishes this, returns would be *stationary* or integrated of order zero, I(0). So we can work with returns directly.

When the stock price increases, $\Delta S_t$ is positive. Let us ignore the dividends temporarily, and let $r_t = \Delta S_t / S_t$ denote the stock return for time $t$. The

autoregressive model of order $p$, AR($p$), with $p$ representing the maximum lag length of correlation, would be

$$\begin{aligned} \text{AR}(p): \quad r_t &= \mu(1-\rho_1-\rho_2-\ldots-\rho_p)+\rho_1 r_{t-1}+\rho_2 r_{t-2} \\ &\quad +\ldots+\rho_p r_{t-p}+z_t\sigma, \end{aligned} \tag{1.4.7}$$

whereas the MA process of order $q$ includes lags of only the random component

$$r_t = \mu+\phi_1 z_1\sigma+\phi_2 z_2\sigma+\ldots+\phi_q z_{t-q}\sigma+z_t\sigma, \tag{1.4.8}$$

where $\mu$ is the average return.

The AR process never completely dies since it is an iterative process. Consider an AR($p$) process with $p = 1$. Now let the first period be defined at $t = 0$, and substitute in (1.4.7) to give

$$r_0 = \mu+z_0\sigma. \tag{1.4.9}$$

During the next period we have the term ($\rho_1$ $r_0$), as this value gets factored into the return by a proportion $\rho_1$. As a result

$$r_1 = \mu(1-\rho_1)+\rho_1 r_0+z_1\sigma = \mu+\rho_1 z_0\sigma+z_1\sigma. \tag{1.4.10}$$

Because $z_0$ influences $r_1$, it also gets passed through to the next period as

$$r_2 = \mu(1-\rho_1)+\rho_1 r_1+z_2\sigma = \mu+\rho_1(\rho_1 z_0\sigma+z_1\sigma)+z_2\sigma \tag{1.4.11}$$

so that a shock $t$ periods ago will be reflected by a factor of $(\rho_1)^t$. In order for the process to be stationary, and eventually return to the average return, we need $|\Sigma_s\rho_s| < 1$, meaning only a fraction of the past returns are reflected in the current return. The flexible nature of this specification has made the ARIMA model important for forecasting. The estimation of the parameters and use of the ARIMA model for simulations will be discussed further in Section 4.1.2.

### 1.4.5 Time-Varying Variance

All of the diffusion methods used to define the change of returns can also be applied to the variance of stock prices. Stocks often go through phases of bull markets where there are rapid mostly upward price changes and high volatility, and bear markets where prices are moving mostly downward or relatively stagnant. As seen in Figure 1.4.4, the standard deviation of returns for S&P 500 has gone through several peaks and troughs over time. In the basic diffusion model (1.4.1), however, the standard deviation $\sigma$ is assumed to be constant over time.

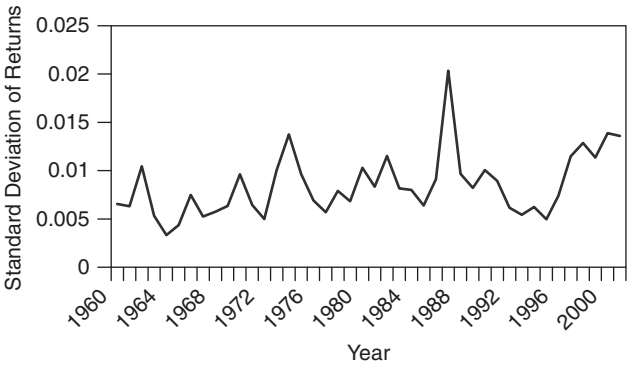



**Figure 1.4.4** Standard deviation of the S&P 500 over time

A more general notation for the diffusion model would be to reflect that both the drift and the volatility are both potentially a function of the stock price and time:

$$dS = \mu(t,S)dt + \sigma(t,S)dz. \tag{1.4.12}$$

A comprehensive parametric specification allowing changes in both drift and volatility is (see Chan et al., 1992):

$$dS = (\alpha + \beta S)dt + \sigma S^{\gamma} dz. \tag{1.4.13}$$

This model is flexible and encompasses several common diffusion models. For example, in the drift term, if we have $\alpha = \eta \overline{S}$ and $\beta = -\eta$, there is mean reversion of (1.4.6). If $\alpha = 0, \beta = 1$, and $\gamma = 1$, it is a continuous version of the random walk with drift given in (1.4.1). When $\gamma > 1$, volatility is highly sensitive to the level of $S$.

Depending on the restrictions imposed on the parameters $\alpha$, $\beta$, and $\gamma$ in equation (1.4.13) one obtains several nested models. Table 1.4.2 shows the eight models (including the unrestricted) considered here and explicitly indicates parameter restrictions for each model.

The first three models impose no restrictions on either $\alpha$ or $\beta$. Models 4 and 5 set both $\alpha$ and $\beta$ equal to zero, while models 6, 7, and 8 set either $\alpha$ or $\beta$ equal to zero. Model 1, used by Brennan and Schwartz (1980), implies that the conditional volatility of changes in $S$ is proportional to its level. Model 2 is the well-known square root model of Cox, Ingersoll, and Ross (CIR) (1985), which implies that the conditional volatility of changes in $S$ is proportional to the square root of the level. Model 3 is the Ornstein-Uhlenbeck diffusion process first used by Vasicek (1977). The implication of this specification is that the conditional volatility of changes in $S$ is constant. Model 4 was used by CIR (1980) and by Constantinides and Ingersoll (1984) indicates that the

**Table 1.4.2 Parameter Restrictions on the Diffusion Model $dS = (\alpha + \beta S)dt + \sigma S^{\gamma}dz$**

| Model Name | $\alpha$ | $\beta$ | $\gamma$ | $\sigma$ | Diffusion Model |
|---|---|---|---|---|---|
| 1. Brennan-Schwartz | | | 1 | | $dS = (\alpha + \beta S)dt + \sigma S dz$ |
| 2. CIR SR | | | 0.5 | | $dS = (\alpha + \beta S)dt + \sigma S^{0.5}dz$ |
| 3. Vasicek | | | 0 | | $dS = (\alpha + \beta S)dt + \sigma dz$ |
| 4. CIR V | 0 | 0 | 1.5 | | $dS = \sigma S^{1.5}dz$ |
| 5. Dothan | 0 | 0 | 1 | | $dS = \sigma S dz$ |
| 6. CEV | 0 | | | | $dS = \beta S dt + \sigma S^{\gamma}dz$ |
| 7. GBM | 0 | | 1 | | $dS = \beta S dt + \sigma S dz$ |
| 8. Merton | | 0 | 0 | | $dS = \alpha dt + \sigma dz$ |

conditional volatility of changes in $S$ is highly sensitive to the level of $S$. Model 5 was used by Dothan (1978), and model 6 is the constant elasticity of variance (CEV) process proposed by Cox (1975) and Cox and Ross (1976). Model 7 is the famous geometric Brownian motion (GBM) process first used by Black and Scholes (1973). Finally, model 8 is used by Merton (1973) to represent Brownian motion with drift.

This flexible parametric specification is useful since the parameters may be estimated (see Chapter 9) to determine the model that best fits a particular security. Vinod and Samanta (1997) estimate all these models to study the nature of exchange rate dynamics using daily, weekly, and monthly rates for the British pound sterling and the Deutsche mark for the 1975 to 1991 period and compare the out-of-sample forecasting performance of the models in Table 1.4.2. The models Brnnan-Schwartz, CIR-SR, and Vasicek performed poorly, whereas CIR-VR and GBM were generally the best.

Now that we have examined some of the possible ways that stock prices can move, the next section explores the way that a stock price should move in an efficient market.

## 1.5 EFFICIENT MARKET HYPOTHESIS

We have concentrated thus far on the similarities of investors and investments. There are many commonalities that market participants share. The mantra of Wall Street is "buy low and sell high," and all investors would prefer to do so. Looking at the big picture, this makes stock investing a tricky proposition. Any stock trade involves a buyer and a seller. They cannot both be right.

For speculators, the mantra may be valid, since they try to profit on short-term price movements. But for the average investors who are interested in receiving a fair return on their investment, a stock trade need not be a zero-sum game. While stock prices are certainly risky, it is not the same as a card game where one player's gain is another player's loss. Stock trades can be mutually beneficial, since each investor has different needs and endow-

ments at different times in the life span. Consider a trade involving an investor who is just entering retirement and needs to liquidate some of his portfolio for living expenses selling to a new father who needs an investment for his new child's college fund. Note that both are better off if they were not able to complete the trade. Without the trade, the retiree may have to do without basic necessities if he does not have ready cash; without the college fund, the child would be worse off.

In order for investors to be confident that the stock market is better than a gamble, stocks must be priced fairly. Any major stock market in the world insists that investors have all available information about a company before purchasing shares. This ensures that the purchase is an informed decision. If a stock is trading at a price lower than its fair value, then sellers will be missing out on value due to them. Stocks that are overvalued go against the buyers. The idea that stocks are continually priced at their fair value is known as the efficient market hypothesis (EMH).

The EMH is an attractive idea for stock investors, since even investors who are not stock analysts, and do not have time to perform in depth research for every company in their portfolios, are assured of trading at a fair price by simply buying or selling at the current market price. Of course, this argument taken to its extreme has all investors relying on the market to price fairly, and no one doing the homework to figure out what this fair price should be. We can see, then, that the EMH creates a market of "smart money" investors who are well informed, as opposed to investors trading at whatever the prevailing price happens to be. It is a waste of resources to have all investors well informed if prices are fair already, but there must be some smart money to ensure the prices get to this fair value.

The EMH comes in three strengths depending on how informed we assume the market price is: weak, semi-strong, and strong.

### 1.5.1 Weak Form Efficiency

This form of the efficient market hypothesis assumes that all historical information is factored into the market price of a stock. Evidence of weak form efficiency can be seen by the demand for high-speed information, and the high price charged for real time financial news. Leinweber (2001) documents the speed at which earnings announcements are reflected in stock prices. He shows that in the 1980s earnings surprises could take up to two weeks to be incorporated into the stock price. In the 1990s, only a decade later, stock prices jumped within minutes of earnings announcements.

What happened to cause this change? In the 1980s, investment news was still very much a print industry. Earnings announcements were on record in the *Wall Street Journal*, which has an inevitable lag because of printing time and delivery. In the 1990s technology took over, and electronic news services such as Reuters and Bloomberg, and Internet news services were the source for late-breaking news, with the newspapers providing analysis and often

ex post credibility. The speed at which transactions could take place also increased with computerized trading and discount brokers. The increased technology makes it hard to argue against weak form efficiency in a market where old news has little value.

### 1.5.2 Semi-strong Form Efficiency

This second level of the efficient market hypothesis encompasses weak form efficiency, and adds the additional requirement that all expectations about a firm are incorporated into the stock price. There is no reason why an investor who thought an interest rate hike was an inevitability would wait until the formal announcement to trade on the information. In previous sections we have implicitly used semi-strong form efficiency in our pricing formulas by pricing *expected* earnings or *expected* dividends.

Semi-strong form efficiency can also explain some of the counterintuitive movements in the stock market, such as the market going up after bad news, or declining on seemingly good news. If the bad news was not as bad as expected, the price can actually rebound when the uncertainty is resolved. Consider the following example: A firm has a fundamental value of \$100 per share at the current level of interest rates of 6%. Inflation starts to pick up, and the Federal Reserve Bank (the Fed) considers an interest rate hike. An interest rate hike increases the cost of funds for the company and therefore reduces its value. The firm's analysts have come up with the possible scenarios listed in Table 1.5.1.

Operating under weak form efficiency, the stock price of this firm would not necessarily move, since no announcement has been made. However, we see that all of the scenarios in Table 1.5.1 involve a rate hike and a share price decrease. Hence it stands to reason that in the last column of the table investors pay less than the original price of \$100 per share.

Under semi-strong form efficiency, the investor prices the stock based on expected price:

$$E(P) = 0.2(95) + 0.5(89) + 0.3(80) = \$87.50 \quad \text{per share.} \tag{1.5.1}$$

Since the price already reflects the consensus rate hike of a little over $\frac{1}{2}$%, the only price movement on the day of the Fed announcement will be the

**Table 1.5.1 Hypothetical Probabilities of Interest Rate Hike from the Current 6%**

| Probability | Interest Rate | Share Price |
|---|---|---|
| 0.2 | 6.25 | 95 |
| 0.5 | 6.5 | 89 |
| 0.3 | 6.75 | 80 |

unexpected component. So, if the Fed increases the rate by $\frac{1}{4}\%$, the stock price will increase from \$87.50 to \$95.00 even though it was still a rate hike.

### 1.5.3 Strong Form Efficiency

This is the extreme version of the EMH. Strong form efficiency states that all information, whether historical, expected, or insider information, is already reflected in the stock price. This means that an investor will not be able to make additional profits, even with insider information. In markets with strict insider-trading regulation, it is probably an overstatement to assume that investors are always correct. In essence, there would have to be a powerful contingent of "smart money" constantly driving the price to its true level. In emerging markets, however, and in markets without restrictions on insider trading, it is not so far-fetched that this "smart money" exists. However, it may be just as far-fetched that this "smart money" is able to drive such volatile markets.

In any of the formulations of the EMH, the underlying result is that profit cannot be made from old news. At any point in time, news is reflected in the stock price, and the price change to the next period will be a function of two things: the required return and the new information or expectations that hit the market. Therefore it should be no surprise that advocates of the EMH find it convenient to model stock price movements as a random walk, as in Section 1.4, with the volatility portion representing unexpected changes in information.

Empirical evidence for or against the EMH is a tricky concept, since it is a hypothesis that states that old information has no effect on stock prices. Usually in empirical tests, researchers look for significant effects, not for the lack thereof. Therefore, disproving the EMH is much easier than proving it. In Chapter 4 we examine a number of anomalies that attempt to disprove the EMH by finding predictable patterns based on old news. These tests usually take the form of

$$r = \alpha + \delta(\text{Anomaly}) + \varepsilon, \tag{1.5.2}$$

where $r$ is the actual stock return, $\alpha$ is the average return, $\delta$ represents the excess return for the anomaly in question, and $\varepsilon$ is the random error term. If $\delta$ is statistically significant, then there is evidence of market inefficiency, since on those anomaly days the market gets a predictable excess return. For a summary of regression analysis, see the Appendix.

Another factor to be taken into account is transactions costs. If a stock price is 2 cents off of its fundamental value, but the brokerage fee is 5 cents per share to take advantage of the discrepancy, it is futile to undertake the transaction and suffer a loss of three cents. The EMH still holds in this case where no transaction takes place, since there are no net profits to be made after the transactions cost is considered.

One method that has been employed to test for evidence of efficiency rather than inefficiency has been an application of the two-sided hypothesis test (Lehmann, 1986) to the EMH by Reagle and Vinod (2003). As we stated above, an excess return less than transactions costs is not necessarily inefficient, since investors have no incentive to trade on the information. Reagle and Vinod (2003) use this feature of the EMH by setting a region around zero where $|\delta|$ is less than transactions costs, and then test that $\delta$ falls in this region at a reasonable confidence level. This is akin to a test that the anomaly is not present, and therefore it would be evidence in favor of the EMH. As with the time it takes for information to be incorporated into a stock price, many other anomalies that were common in stock prices a decade ago have gone away as transactions costs have decreased and the volume of information has increased.

If one accepts the EMH, risk becomes the central focus of the investor. All movements other than the average return are unknown. Research and data collection do not aid in predicting these movements along a random walk. So if these risks cannot be avoided, the next step is to measure and quantify the risk involved for an investment. This will be the topic of the next chapter.

## APPENDIX: SIMPLE REGRESSION ANALYSIS

When there is a relationship of the form

$$y = \alpha + \beta x + \varepsilon, \tag{1.A.1}$$

where $y$ is a dependent variable that is influenced by $x$, the independent variable, and the random error is $\varepsilon$, then regression analysis can be employed to estimate the parameters $\alpha$ and $\beta$.

The most common method of estimation for regression analysis is ordinary least squares (OLS), which can be used given that the following assumptions hold:

1. The regression model is specified correctly; that is, in the above case $y$ is a linear function of $x$.
2. $\varepsilon$ has a zero mean, a constant variance for all observations, and is uncorrelated between observations.
3. $x$ is given, not correlated with the error term, and in the case where there is more than one independent variable, the independent variables are not highly correlated with each other.
4. There are more observations than the number of parameters being estimated.

These assumptions basically state that the parameters can be estimated from data, and that all the available information is used. These assumptions

may be tested, and in most cases the regression model may be modified if one or more assumptions fail. In this appendix we cover OLS estimation and interpretation. For a complete reference on extensions to OLS see Greene (2000) or Mittlehammer, Judge, and Miller (2000).

OLS involves finding the estimates for the parameters that give the smallest squared prediction errors. Since ε is zero mean, our prediction of the dependent variable is

$$\hat{y} = a + bx, \tag{1.A.2}$$

where $a$ and $b$ are the estimates of $\alpha$ and $\beta$, respectively.

Prediction error is then the difference between the actual value of the dependent variable $y$, and the predicted value $\hat{y}$:

$$e = y - \hat{y}. \tag{1.A.3}$$

The OLS solution, then, is the value of $a$ and $b$ that solve the optimization problem:

$$\min_{a,b} \sum_n e_i^2, \tag{1.A.4}$$

where the minimization is with respect to the parameters $a$ and $b$ and where $n$ is the number of observed data points and $\Sigma_n$ denotes summation from $i = 1$ to $i = n$.

The values of $a$ and $b$ that solve the OLS minimization are

$$b = \frac{\sum_n (y_i - \bar{y})(x_i - \bar{x})}{\sum_n (x_i - \bar{x})^2} \quad \text{and} \quad a = \bar{y} - b\bar{x}, \tag{1.A.5}$$

where $\bar{y}$ and $\bar{x}$ denote the average of $y$ and $x$, respectively. The numerator of $b$ is also known as the covariance between $x$ and $y$ (multiplied by $n$), and the denominator is identical to the variance of $x$ (multiplied by $n$). The presence of the covariance is intuitive since b estimates the amount of change in $y$ for a change in $x$. Dividing by the variance of $x$ discounts this movement of $x$ by its volatility, since high volatility independent variables may move a large distance for small changes in $y$.

These are the estimated values of the OLS parameters that give the lowest sum of squared prediction errors. They do not give a precise value, however. One can think of the estimates as being in a range around the true value. Given a large enough sample size (over 30 observations) this range can be determined by the normal distribution.

Given that the OLS assumptions hold, the estimates $a$ and $b$ in (1.A.5) are unbiased, meaning that on average they fall around the true parameter value, and they are the "best" estimates in that they have the lowest variance around

the true parameter of any other unbiased estimator of a linear relationship. These properties are referred to as BLUE, or best linear unbiased estimator.

Since the estimators will have some error compared to the true parameters, this error can be quantified by the standard deviation of the normal distribution around the true parameter, also known as the estimate's standard error:

$$\sigma_b = \sqrt{\frac{\sum_n e_i^2}{(n-z)\sum_n (x_i - \bar{x})^2}} \tag{1.A.6}$$

From the properties of the normal distribution, 95% of the probability falls within 1.96 standard deviations of the mean. This allows us to construct a 95% confidence interval for the true parameter based on the estimated value

$$b \pm 1.96\sigma_b. \tag{1.A.7}$$

Roughly speaking, the true unknown parameter β will only fall outside of this range 5% of the time. The 5% error is known as the significance level.

A further method of statistical inference using the OLS estimate is hypothesis testing. Hypothesis testing sets up two competing hypothesis about the parameter, and then uses the estimates from the data to choose between them. Usually the accepted theory is used as the null hypothesis $H_0$, and the null is assumed to be valid unless rejected by the data, in which case the default hypothesis is the alternative hypothesis $H_A$.

To test the significance of $b$, the usual null hypothesis is $H_0: \beta = 0$. If we assume the null is true, the observed $b$ should fall within 1.96 standard deviations 95% of the time. Therefore $[b - 1.96\sigma_b, b + 1.96\sigma_b]$ is our acceptance region where the null hypothesis is reasonable (alternatively, $b$ could be divided by the standard error to obtain a $z$ value—or $t$ value in small samples—and compared to $-1.96 < z < 1.96$). If the estimated statistic falls outside this region, the null is not reasonable since this would be a rare event if the true parameter were, zero. Then the null hypothesis would be rejected, and the alternative $H_A: \beta \neq 0$ would be accepted (strictly speaking, not rejected). In the case of rejection of the null, we say that $b$ is statistically significant, or statistically significantly different from zero.

The regression model is often extended to allow for several independent variables in a multiple regression. The interpretation of regression coefficients, their estimators, standard errors, confidence intervals, are analogous to the simple regression above, although a compact solution requires matrix algebra, as we will see in later chapters.

CHAPTER 2

# A Short Review of the Theory of Risk Measurement

## 2.1 QUANTILES AND VALUE AT RISK

We know intuitively that high-risk ventures should yield a risk premium and thus higher returns. In Chapter 1 we even quantified risk in terms of a standard deviation $\sigma$ of a probability distribution of returns. What we want to do in this chapter is to go one step further to answer the question "How much more should I get for taking additional risk?" We will answer this question according to traditional methods, and point to places that these methods can be improved or, in some cases, where these methods have already been improved but the improvement has not been widely implemented.

Standard deviation is a valuable concept for modeling volatility of stock returns, but the estimated standard deviations often seem foreign to an investor and not easily transferred to the bottom line: the price. Since business managers and bankers think of risk in terms of dollars of loss, standard deviation does not fit very well into the language of business. It is not necessarily true that the higher the risk as measured by $\sigma$, the higher are the losses. The senior management needs a number that gives the worst-case scenario of dollar loss, in a prescribed holding period.

Enter value at risk (VaR). VaR puts a dollar amount on the highest expected loss on an investment. VaR has become popular since the US financial giant J.P. Morgan published RiskMetrics™ methods in 1995. We will provide an overview of VaR in this chapter and many extensions that have been developed. For further details see surveys of VaR by Duffe and Pan (1997) and Jorion (1997). Some issues involving the time horizon for VaR need

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

methods developed for forecasting volatility in Chapter 4 and their discussion is postponed until Chapter 5.

The first section discusses various approaches to VaR computation. Recent VaR computations can involve variances, covariances, Monte Carlo simulations, and other types of historical simulations. In a typical VaR the aim is use the data to construct a "loss distribution" for a given time horizon and amount of investment. Then we ask: What upper bound can we put on losses such that higher losses would be highly unlikely, say, a 1 in 100 (or 1 in 20) chance? The first step in constructing this loss distribution is to identify quantiles.

A quantile of a return distribution is the return that is above a certain proportion of all possible returns. Special cases are percentiles, which are above a certain percentage of returns, and deciles, which split returns into tenths (i.e., the first decile is about 10% of returns). Similarly quartiles split returns into quarters. Quantiles are easiest to calculate if we specify normal distribution, since then all quantiles can be calculated based on only the average and standard deviation. This is the original approach taken by RiskMetrics™.

To find the percentiles of a normal distribution one starts with the PDF for a standard normal distribution (mean of zero and standard deviation of one), seen in Figure 1.3.1. The cumulative distribution function (CDF) is the area of the PDF below a specific $z$ value. Given a proportion $\alpha' \in [0, 1]$, the associated quantile $Z_{\alpha'}$ is defined by the following probability statement $\Pr(z \le Z_{\alpha'}) = \alpha'$. For our examples if $\alpha' = 0.05$ (the 5th percentile), we have $Z_{\alpha'} = -1.645$. If $\alpha' = 0.025$, we have $Z_{\alpha'} = -1.96$, and if $\alpha' = 0.01$, we have $Z_{\alpha'} = -2.326$. These numbers can be obtained from the traditional print tables for the standard normal distribution, or from computer software as detailed in Chapter 9. Computer software is often preferred as it allows more decimal places and can be conveniently added to existing spreadsheets.

Another way to think of normal distribution quantiles is as the inverse function of its CDF. The CDF of the unit normal N(0, 1) is written as an integral of the PDF:

$$\Pr(z \le A) = \Phi(A) = \int_{-\infty}^{A} \left( \frac{1}{\sqrt{2\pi}} \right) \exp\left[ \frac{-z^2}{2} \right] dz. \tag{2.1.1}$$

Thus $\alpha' = \Phi(Z_{\alpha'})$, and then applying the inverse operator $\Phi^{-1}$ to both sides yields $\Phi^{-1}(\alpha') = Z_{\alpha'}$. This notation is very convenient when obtaining percentiles from spreadsheet programs such as Microsoft Excel.

We have discussed percentiles of the standard normal, but the normal distribution can easily be generalized since any linear transformation of a normal variable is still normally distributed. So, if we assume that interest rates $R$ are normally distributed with mean $\mu$ and standard deviation $\sigma$ (variance $\sigma^2$), it can be written as

$$R \sim \mathrm{N}(\mu, \sigma^2),$$

which is related to the standard normal $z$ by the linear relation

$$R = \frac{z - \mu}{\sigma},$$

where we are using capital $R$ for the random variable of returns as opposed to lowercase $r$ for the realizations. Such a distinction is not always convenient to sustain.

Hence the quantiles for $R$ are obtained by a linear transformation of the quantiles of $z$, the standard normal.

$$R_{\alpha'} = \mu + Z_{\alpha'}\sigma, \tag{2.1.2}$$

where $Z_{\alpha'}$ itself is negative when it is in the left tail of N(0, 1) for lower quantiles. See Figure 1.3.1. We will directly use the lower quantile from (2.1.2) in our formula for VaR.

Downside risk is represented by the lower quantiles of $R$ (upper quantiles when dealing with the distribution of losses instead of returns). Thus our formula for value at risk includes dollar units by multiplying the lowest probable return by the capital invested:

$$\text{VaR}(\alpha') = -R_{\alpha'} * K, \tag{2.1.3}$$

where $K$ denotes the capital invested, and the quantile of $R$ is from (2.1.2) in decimals. So the dollar value of VaR($\alpha'$) represents an extreme scenario where one would only see higher losses $\alpha'$ proportion of the time. Common proportions to choose for $\alpha'$ are 0.01 (1%), 0.05 (5%), and 0.1 (10%).

We expect $R_{\alpha'}$ itself to be negative in most practical cases, so the VaR is positive. If, however, in rather unusual cases, $R_{\alpha'}$ is positive, we will have a negative estimate of VaR. This simply means that the worst-case scenario loss is actually a (small) profit. We emphasize that the quantile $R_{\alpha'}$ has been calculated here using the normal distribution. In general, this quantile can be estimated by any number of parametric or nonparametric methods, and can come from any probability distribution function (PDF) of returns $f(R)$. Alternate distributions to the normal will be discussed in Chapter 4.

First, let us illustrate an application of the formula (2.1.3) to the artificial data from a diffusion model we have already encountered in Section 1.4. Let us assume that the *time horizon* $\tau$ equals one day. Let $r_t$ in the first column of Table 1.4.1 be the percent daily returns, originally obtained from the standard normal distribution. A sorting of 30-day data reveals that the smallest value is –1.08 and the largest value is 2.08. If we assume that this range is valid, and if one invests $K = \$100$ on a given day in that stock, one can lose \$1.08 at worst or gain \$2.08 at best. Various percentiles of the data given in Table 2.1.1 go beyond mere sorting. The first row gives the percentage, the second row the quantile associated with it from raw data, and the third row gives the quantile

**Table 2.1.1 Percentiles of 30-Day Returns from Table 1.4.1 for $100 Investment**

| Percent | 1 | 5 | 25 | 50 | 75 | 95 | 99 |
|---|---|---|---|---|---|---|---|
| Raw data | –0.9959 | –0.736 | –0.2775 | 0.21 | 0.7375 | 1.798 | 2.05 |
| Normality | –1.5865 | –1.0273 | –0.2311 | 0.3223 | 0.8758 | 1.6720 | 2.2312 |

under the normality assumption. For example, the five percentile of –0.736 for raw data means the following probability statement for the random variable denoted by upper case $R$: $\Pr[R < -0.736] = 0.05$.

Since this is simulated data, this is one case where the appropriateness of the normal distribution need not be debated. From Table 2.1.1, with the one percentile of –1.5865 suggests that a $100 investment in this stock can be $1.59 or lower with a 1 in 100 chance. The value at risk (VaR) then is $1.59 on a given day if the probability distribution of returns follows the available 30-day data and time horizon is one day. We see from the last two rows of Table 2.1.1 that the empirical percentiles from the raw data do not match the normal distribution quantiles exactly. If these were actual returns, there could be some debate that the normal distribution makes the losses look much larger than they should be. The theoretical distribution has the advantage of being continuous. With empirical data, no observation is below the lowest value even though much lower returns than the lowest in the sample of 30 cannot be ruled out.

If we recognize that the normal distribution is not appropriate for the data at hand there are literally hundreds of possible distributions developed by statisticians over the last century to consider. However, such consideration will lead us far astray into theoretical statistics. Instead, we take a "systems approach" and consider one important family of distributions, the Pearson family, which represent a generalization of the normal. Many of the probability distributions discussed in elementary statistics courses including normal, binomial, Poisson, negative binomial, gamma, chi-square, beta, hypergeometric, Pareto, Student's $t$, and others are all members of this family.

### 2.1.1 Pearson Family as a Generalization of the Normal Distribution

The Pearson family of distributions has been known since the 1890s and has been used in engineering and natural sciences, but has remained a theoretical curiosity in social sciences (see the Appendix at the end of this chapter for details). However, thanks to modern computers and software tools, Pearson family has now become practical for applications in Finance. For example, Rose and Wood (2002, ch. 5) discuss computer tools for estimating any member of Pearson type I to VII from data on excess returns in a fairly mechanical fashion. Therefore it is now possible to apply a "systems approach" and estimate any suitable member of the Pearson family by following one general method for all members. In particular, there is a straightforward way

to make the VaR more general by using a member from the Pearson system of probability distributions.

Since the Pearson system allows the distribution to be more general than just normal, it adds more parameters than just the mean and standard deviation to describe the probability density. Before we explain the Pearson system, we need to define Pearson's measures of skewness and kurtosis. Skewness represents the magnitude to which a PDF has higher probability in the positive or negative direction. Positive skewness means that extreme outcomes above the mean are more likely than extreme outcomes below the mean. A negatively skewed distribution will have relatively higher probability for extreme outcomes below the mean. For the normal distribution, skewness is zero. Kurtosis measures the degree to which extreme outcomes in the "tails" of a distribution are likely. The normal distribution has a kurtosis of 3 (mesokurtic). Distributions with fatter tails are leptokurtic, and distributions with smaller tails are platykurtic.

Definitions of Pearson skewness and kurtosis measures are

$$\beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3} \quad \text{and} \quad \beta_2 = \frac{\mu_4}{(\mu_2)^2}, \tag{2.1.4}$$

where population central moments of order $j$ are denoted by $\mu_j$ ($\mu_2 = \sigma^2$). Pearson considered the limiting distribution of the hypergeometric distribution, which encompasses several related distributions, and found that it can be viewed in terms of a differential equation:

$$\frac{df}{dx} = (x-a)\frac{f}{b_0 + b_1 x + b_2 x^2}, \tag{2.1.5}$$

where $a$ is the mode (the value of $x$ when the frequency is the largest) of the distribution and the coefficients of the quadratic in the denominator of (2.1.5) can be expressed in terms of the moments of the distribution. If the origin is shifted to the mode, the differential equation can be written as

$$\frac{d\log f}{dy} = y\left[B_0 + B_1 y + B_2 y^2\right]^{-1}, \tag{2.1.6}$$

where $y = x - a$. The explicit density is obtained by integrating the right-hand side of this equation. As with any quadratic equation, there are a few possibilities for the solution. The quadratic in the denominator of (2.1.6) can have real roots of the same sign, real roots of the opposite sign or imaginary roots. Define $K' = (B_1)^2/[4B_0B_2]$, and note that if $K' \geq 1$, we have real roots of the same sign. If $K' \leq 0$, we have real roots with opposite sign, and if $0 < K' < 1$, we have imaginary roots of the quadratic. These three possibilities give rise to the three main types of the Pearson family of distributions. Instead of check-

ing whether $K' < 0$ on $K' \in (0, 1)$, or $K' > 1$, it is convenient to estimate $\kappa$, which can be expressed in terms of Pearson's skewness and kurtosis coefficients in (2.1.4) above as

$$\kappa = \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)}. \tag{2.1.7}$$

The normal distribution has $B_1 = 0$, $B_2 = 0$, and $B_0$ is such that $\beta_1 = 0$, $\beta_2 = 3$. By changing the coefficients in the quadratic and skewness and kurtosis parameters, it is possible to consider various members of the Pearson family from type I to type IX.

When the quadratic has real roots of the opposite sign, it leads to the "beta distribution of the first kind," dubbed as Pearson type I, which can be U-shaped or J-shaped. Its symmetrical form is called Pearson type II, with zero skewness ($\beta_1 = 0$). Type III is the gamma (and chi-square) distribution. If the roots of the quadratic are imaginary numbers (complex conjugate), it is Pearson type IV. Common statistical distributions do not belong to type IV, and its evaluations often need numerical quadratures. If the quadratic is a perfect square $[B_0 + B_1 y + B_2 y^2] = (y - \lambda)^2$, it leads to types V, VIII, and IX. If the two real roots are of the same sign, it leads to the "beta distribution of the second kind," dubbed as Pearson type VI and an important example of this type is Student's $t$ distribution. Student's $t$ distribution is of interest in finance, since it has fat tails (higher probability of extreme outcomes), which are often present in financial data on returns. The Weibull distribution is close to type VI. It is also known as Pareto distribution of the first kind. Type VII arises from a mixture of the normal and gamma and is also of interest in finance, since mixtures can yield a wide range of realistic parametric flexible forms for the pdf of excess returns, $f(x)$.

Which distribution from the Pearson family is the right one for a given data? To answer this question, it is useful to consider some known guidelines developed over 100 years ago by Karl Pearson based on his measures defined in (2.1.4). The guidelines involve a graph with $\beta_1$ on the horizontal axis and $\beta_2$ on the vertical. One simply finds the point where the estimates of skewness and kurtosis fall in this graph and the plot tells which member of the Pearson system, type I to type VI, should be used. For example, in Figure 2.1.1 the normal distribution is at a single point $N$ in the graph, the type I has three regions shown by I, I($U$), and I($J$) to indicate where the curve will be similar to a usual density, when it is U-shaped and when it is J-shaped.

Figure 2.1.1 depicts our estimates $\hat{\beta}_1 = 0.282958$ and $\hat{\beta}_2 = 2.49385$ from the simulated data as coordinate values of the bold black dot appearing in the region marked by type I curve of the Pearson system of curves. Accordingly we find that the following fitted member of the Pearson system defined below in (2.1.8) best represents the underlying PDF:

$$f(R) = 0.0821792\,(2.89435 - R)^{1.60402}\,(0.961295 + R)^{0.299601} \tag{2.1.8}$$

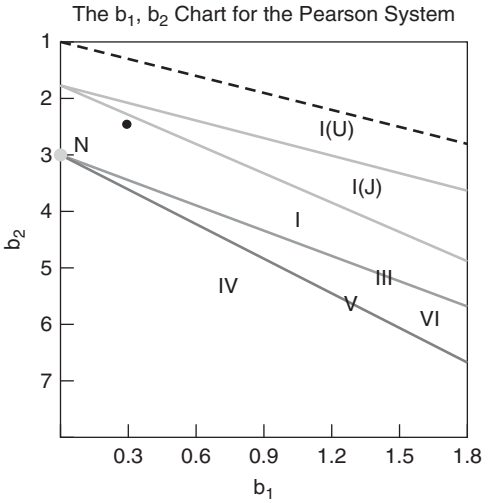



**Figure 2.1.1** Pearson types from skewness-kurtosis measures

**Table 2.1.2 Descriptive Statistics for AAAYX Mutual Fund Data**

| Max | Min | Median | Mean | Variance | Standard Deviation | Kurtosis $\hat{\beta}_2$ | Skewness $\sqrt{\hat{\beta}_1}$ |
|---|---|---|---|---|---|---|---|
| 10.724 | –21.147 | 1.088 | 0.7387 | 14.174 | 3.765 | 10.81264 | –1.60407 |

defined for $R \in [-0.961295, 2.89435]$. These coefficients are related to the coefficients of the underlying differential equation used by Pearson. The Appendix to this chapter gives the formulas for getting such coefficients from estimates of central moments $m_2$, $m_3$, and $m_4$ obtained from the data.

Now we consider a real world example. Let Tb3 denote the risk-free return from the three-month US Treasury bills. Now our $r$ denotes the *excess* return (defined as return minus Tb3) from a mutual fund in Morningstar (2000). Our excess return data are for a mutual fund named Alliance All-Asia Investment Advisors Fund with the ticker symbol AAAYX. We consider a period of 132 months from January 1987 to December 1997. The estimated central moments $m_2$, $m_3$, and $m_4$ are 14.1743, –85.6006, and 2172.38, respectively. Karl Pearson's betas defined (2.1.4) and other descriptive statistics for our data are given in Table 2.1.2. Skewness (column 8) bears the sign of $m_3$.

First, we compute the VaR for the mutual fund by ignoring the skewness and simply assuming that the underlying PDF is normal with mean 0.7387 and variance 14.174 (standard deviation = $\sigma$ = 3.76484). Let $\alpha' = 0.01$, and recall that the theoretical 1 percentile of the standard normal distribution is –2.326348. For the mean and variance of the data the relevant quantile for

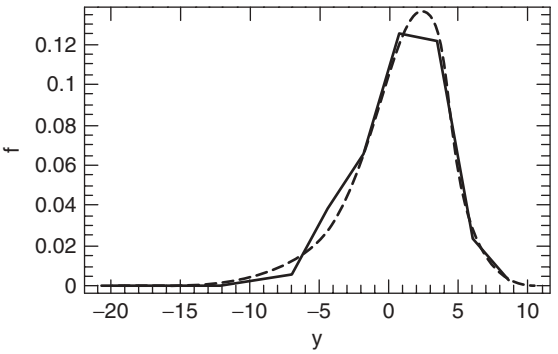



**Figure 2.1.2** Empirical PDF (–) and fitted Pearson PDF (––) based on (2.1.9)

VaR is $R_{\alpha'} = -8.019621$. The negative value of this quantile times $K$ times 0.01 (to be in decimals) then becomes the VaR. If the investment is $K = 100$ million, the value at risk over a fixed time period $\tau$ is \$8,019,621. Thus VaR exceeds eight million.

### 2.1.2 Pearson Type IV Distribution for Our Mutual Fund Data

We can see from the estimates in Table 2.1.2 that the normal distribution is not appropriate. Kurtosis is high, and the returns are negatively skewed. If we plot the observed point in a diagram with Pearson's skewness $\beta_1$ on the horizontal and $\beta_2$ on the vertical axis, as in Figure 2.1.1, we note that AAAYX mutual fund falls in the region of Pearson's type IV curve. This is not good news. Type IV is the hardest to work with since it involves imaginary roots of the quadratic in (2.1.6), so we expect difficulties in computing the quantiles for our VaR calculation. However, we can obtain an analytical expression (2.1.9) (the calculation is given in the Appendix at the end of this chapter and in Chapter 9) for its observed density using four moments (which are parameters of the underlying distribution):

$$f(y) = \frac{\exp[4.83928 \tan^{-1}(1.00414 - 0.166965y)]}{(9.61299 - 1.60502y + 0.133437y^2)^{3.74707}}. \tag{2.1.9}$$

The long left tails in Figure 2.1.1 for empirical probability distribution function (PDF) and for the fitted Pearson curve (2.1.9) together confirm negative skewness. Figure 2.1.2 also reveals the good fit from the closeness of the two curves to each other.

Integrating the Pearson PDF (2.1.9) to get the VaR at the 1% level can be done only by numerical approximation. We set the lower limit of the range to –50 and evaluate various upper limits, where each integral takes about three minutes. Hence the inverse CDF is difficult to compute. However, sequential

computations with different upper limits [–11.55, –11.54] yields the integral of [0.00998124, 0.00100078]. Thus the VaR = $11,540,000 is a good approximation. This is higher than $8,019,621 based on the normal distribution. Using the more flexible Pearson system reveals extra risk in the left tail ignored by the normal distribution. Whereas, if the normal distribution were appropriate, it would be the one selected by Pearson's system.

### 2.1.3 Nonparametric Value at Risk (VaR) Calculation from Low Percentiles

Normal distribution $N(\mu, \sigma^2)$ is parametric because it depends on estimating $\mu$ and $\sigma^2$. The data driven distributions represented by the solid lines in Figure 2.1.2 are called nonparametric (or empirical) because they do not use moments or any other shape parameters. Suppose that a portfolio manager invests $K$ = $100 million in AAAYX mutual fund for $\tau = 1$ month based on the data for 132 months. The one percentile of –8.708737 from the raw data means that she may lose $\text{VaR}(\alpha' = 0.01)$ = $8,708,737 with a chance of 1 in 100. All we have done is computed the 1 percentile (1%LE), changed its sign and multiplied by one million (= 0.01*$K$). The 1%LE is based on the nonparametric or empirical PDF from the observed data. The comparable VaR from Pearson type IV is more conservative, suggesting a larger potential loss of $11,540,000.

### 2.1.4 Value at Risk for Portfolios with Several Assets

In the example above, we considered only one portfolio (AAAYX) in isolation. If a set of funds and stock market index funds is considered, it becomes necessary to incorporate the correlations among the various portfolios. Table 2.1.3 reports annualized returns on three components of a portfolio. Assume that the investor invests 20%, 30%, and 50% of available funds in Fidelity Magellan Fund, Vanguard 500 index fund, and three-month Treasury bills, respectively.

Table 2.1.4 gives the 3 × 3 matrix of variances (along the diagonal) and covariances (off-diagonal) among the three investments. (For an explanation of matrix and variance-covariance notation, see Chapter 8.) The variance-covariance matrix $V$, and the variance $s^2$ of the portfolio, respectively, are

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix},$$

$$s^2 = (p_1 \quad p_2 \quad p_3) \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}. \tag{2.1.10}$$

**Table 2.1.3 Returns on Three Correlated Investments**

| | Fidelity Magellan | Vanguard 500 | Three-Month T-Bill |
|---|---|---|---|
| Jan–97 | 4.389 | 6.232 | 0.429 |
| Feb–97 | –1.342 | 0.789 | 0.426 |
| Mar–97 | –3.443 | –4.136 | 0.439 |
| Apr–97 | 4.464 | 5.956 | 0.442 |
| May–97 | 7.134 | 6.088 | 0.436 |
| Jun–97 | 4.14 | 4.454 | 0.423 |
| Jul–97 | 8.391 | 7.966 | 0.431 |
| Aug–97 | –4.448 | –5.609 | 0.437 |
| Sep–97 | 5.886 | 5.469 | 0.423 |
| Oct–97 | –3.405 | –3.35 | 0.423 |
| Nov–97 | 1.949 | 4.598 | 0.439 |
| Dec–97 | 1.138 | 1.725 | 0.442 |
| Jan–98 | 1.081 | 1.11 | 0.432 |
| Feb–98 | 7.58 | 7.192 | 0.435 |
| Mar–98 | 5.039 | 5.104 | 0.429 |
| Apr–98 | 1.149 | 1.008 | 0.424 |
| May–98 | –1.976 | –1.743 | 0.428 |
| Jun–98 | 4.261 | 4.072 | 0.425 |
| Jul–98 | –0.748 | –1.054 | 0.422 |
| Aug–98 | –15.486 | –14.474 | 0.418 |
| Sep–98 | 6.046 | 6.411 | 0.396 |
| Oct–98 | 7.701 | 8.164 | 0.34 |
| Nov–98 | 7.76 | 6.072 | 0.378 |
| Dec–98 | 9.619 | 5.805 | 0.376 |
| Jan–99 | 5.248 | 4.204 | 0.369 |
| Feb–99 | –3.193 | –3.124 | 0.379 |
| Mar–99 | 5.402 | 3.997 | 0.38 |
| Apr–99 | 2.389 | 3.852 | 0.362 |
| May–99 | –2.875 | –2.389 | 0.384 |
| Jun–99 | 6.517 | 5.556 | 0.394 |
| Jul–99 | –3.368 | –3.13 | 0.388 |
| Aug–99 | –1.236 | –0.496 | 0.407 |
| Sep–99 | –1.478 | –2.739 | 0.403 |
| Average | 1.94803 | 1.926667 | 0.410879 |
| Standard deviation | 5.220064 | 4.960944 | 0.027211 |

We report corresponding correlations in parentheses in Table 2.1.4. It is interesting to note that stock market funds and Treasury bill returns are inversely correlated during the time period. The portfolio consists of $p_1 = 0.20$, $p_2 = 0.30$, and $p_3 = 0.50$ proportions of the total investment. Let the column vector of these proportions be denoted by $p_v$. The average returns for each investment are reported at the base of Table 2.1.3.

**Table 2.1.4 Covariances and Correlations among Three Investments**

| | Fidelity Magellan | Vanguard 500 | Three-Month T-Bill |
|---|---|---|---|
| Fidelity Magellan | 26.42334 | | |
| Vanguard 500 | 24.40821 (0.971986) | 23.86518 | |
| Three-month T-bill | –0.02557 (–0.18563) | –0.01697 (–0.12964) | 0.000718 |

Note: Correlations are in parentheses.

Let the column vector of averages be denoted as $\bar{z}_v$. Now the overall expected return for the portfolio is $\bar{x} = (p_v' \bar{z}_v) = 1.173046$, where the prime denotes the transpose, which is simply the weighted sum of column averages with weights 0.2, 0.3, and 0.5, respectively. The variance of the combined portfolio with weights $p_v$ is the matrix multiplication $s^2 = p_v' V p_v$, where $V$ is the $3 \times 3$ variance covariance matrix from Table 2.1.4. This is analogous to our result in Chapter 1 where $\text{var}(ax) = a^2\text{var}(x)$. Note that this multiplication gives a number (scalar) equal to 6.121181 with the square root $s = 2.474102$ (the standard deviation of the portfolio). The 1 percentile for this portfolio consisting of correlated securities is obtained by substituting the $\mu$ and $\sigma$ (s) in $R_{\alpha'} = \mu + Z_{\alpha'}\sigma$ to yield –4.582576. Hence VaR = \$4,582,576 for $K$ = \$100 million.

In conclusion we have discussed the idea of quantiles from first principles and given formulas for computing the value at risk from the lower quantiles of the distribution of returns. We have considered both parametric and nonparametric methods and considered multiple investments, some with negative correlations among themselves. We will extend these methods in later chapters as we account for downside risk (Chapter 5), and expand our repertoire of simulation techniques (Chapter 9).

## 2.2 CAPM BETA, SHARPE, AND TREYNOR PERFORMANCE MEASURES

Value at risk is an important concept in risk management, since it can put a dollar amount on the downside risk. One drawback of VaR, however, is that it often is used to analyze an investment portfolio independent of alternative investments, and independent of the market it is in. The performance measures discussed in this section again are used to put a number on risk but do so in the context of the broader market. In these measures also the decision process of the investor can be incorporated, giving valuable insight for adjusting the models for downside risk.

Once a collection of securities is put into a market for sale, the ball is in the court of the investor. An investor can reduce risk exposure through diversification among the collection of securities. Consider for simplicity only two dif-

ferent securities $x_1$ and $x_2$ with average returns $\mu_1$, $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$, respectively. Now we show that combining these securities into a portfolio can yield an investment with lower risk than either security. For example, if we put proportion $w$ of the portfolio into security 1, and thus the proportion $(1 - w)$ is invested in security 2, the portfolio has $wx_1 + (1 - w)x_2$. The mean return for this portfolio is $w\mu_1 + (1 - w)\mu_2$, and the variance of the portfolio is

$$\text{var(portfolio)} = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\text{cov}_{1,2} \tag{2.2.1}$$

where $\text{cov}_{1,2} = E(x_1 - \mu_1)(x_2 - \mu_2)$ is the covariance between the two securities defined in terms of an expectation operator (i.e., averaging over time). The covariance equals

$$\text{cov}_{1,2} = \left(\frac{1}{T}\right)\sum_{t=1}^{T}(x_{1t} - \mu_1)(x_{2t} - \mu_2), \tag{2.2.2}$$

which measures co-movement of the excess return on two securities, $x_{1t}$ and $x_{2t}$, over time. As seen in the formula, if security 1 and 2 are both above their respective means, the term in the summation at time $t$ is positive and hence time $t$ makes a positive contribution to the covariance. If the two securities are both below their means, $(x_{1t} - \mu_1)$ is negative and $(x_{2t} - \mu_2)$ term is also negative. Since their products remain positive, the $\text{cov}_{1,2}$ term makes a positive contribution at time $t$ and the covariance is increased. Only if one security moves above its mean, and one security moves below its mean at the same time, the product is negative and covariance $\text{cov}_{1,2}$ is decreased.

This covariance (2.2.2) is the key to diversification. If the securities move in opposite directions, then volatility of the portfolio is lowered as the lows of one security are offset by highs of the other. However, we emphasize that the two assets don't have to be negatively correlated with each other for diversification to take place from a portfolio combining the two. Consider the following hypothetical portfolio equally weighted between two assets:

$$\text{Portfolio} = wx_1 + (1-w)x_2 = 0.5x_1 + 0.5x_2 \tag{2.2.3}$$

with $w = 0.5$, $\sigma_1^2 = \sigma_2^2 = 1$. Portfolio variance using (2.2.1) is $0.5 + 0.5*\text{cov}_{1,2}$, which is generally fractional. This means that unless $x_1$ and $x_2$ are perfectly correlated (i.e., unless $\text{cov}_{1,2} = 1$ in this case), the variance of the combined portfolio (2.2.3) will be less than $\text{var}(x_1) = \sigma_1^2 = 1$ and $\text{var}(x_2) = \sigma_2^2 = 1$, the variance of each of the individual securities. Here $\text{cov}_{1,2}$ only needs to be fractional (not negative) to achieve the diversification benefits.

So far we have seen that combining two securities into a portfolio can reduce risk by diversification. But how many securities do we need in a portfolio to adequately diversify? A good rule of thumb is that after 30 stocks, there are limited benefits to any further diversification. It is an interesting coin-

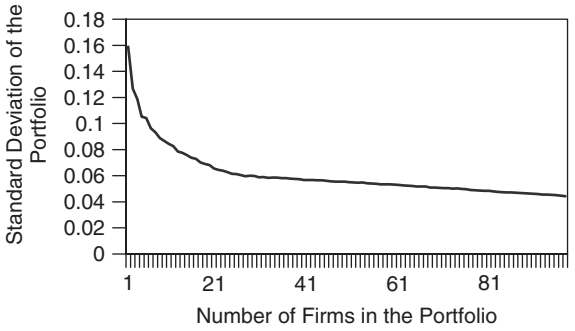



**Figure 2.2.1** Increasing the number of firms in a portfolio lowers the standard deviation of the aggregate portfolio

cidence that the Dow Jones Industrial average chosen over hundred years ago has 30 industrial corporations.

From Figure 2.2.1, the first stocks that are added to a portfolio lower the standard deviation of returns dramatically. The next few stock achieve some diversification and lower the standard deviation somewhat. But after 30 or so stocks, the standard deviation is still positive and does not fall below 0.04. This is because risk in stocks can never be completely eliminated by increasing the size of one's portfolio. There are going to be some events that affect all stocks to some extent, simply because they are all on the same planet, running some business and trading on some market. This risk, which cannot be eliminated, is called market risk, or systematic risk. Some examples of risk that cannot be diversified are currency crises, worldwide recessions, wars, and extreme weather conditions, and these are often exactly what is feared in downside risk.

Consider the following analysis of variance:

$$\text{Total variance} = \text{Asset-specific variance} + \text{Undiversifiable market variance.} \quad (2.2.4)$$

Assuming that the risk is represented by variance, (2.2.4) implies that

$$\text{Total risk} = \text{Diversifiable risk} + \text{Systematic undiversifiable risk.} \quad (2.2.5)$$

Recall that the difference between diversifiable and nondiversifiable risk is an important step in asset pricing and risk management. We noted in Chapter 1 that investors require a higher return for a riskier investment, namely a risk premium. But there is no reason to award a risk premium for diversifiable risk, since it can be avoided simply by holding a portfolio of several stocks. The stock market rewards risk, not stupidity. If an investor can easily get rid of the risk, the market forces will not compensate for it. The capital asset pricing

model (CAPM) uses this insight to price stocks by measuring only the systematic risk in order to put a value on what the risk premium should be. We show the derivation of the CAPM result in the Appendix and directly proceed to the pricing results of CAPM.

To find out how much market risk a particular security or portfolio has, CAPM compares the returns of the security to the overall returns of a market portfolio. These returns are generally calculated in excess of a risk-free rate, $r_f$, where the subscript $f$ refers to its risk-free character. The risk-free rate, as in Chapter 1, is the rate that can be earned on a safe investment, so a risky investment would earn an "excess" return to compensate for risk.

The CAPM formula sets up excess returns $r_i$ of the $i$th stock or security as a linear function of excess return $r_m$ of the market. The CAPM regression is

$$\underset{\text{Systematic}}{(r_i - r_f) = \alpha + \beta(r_m - r_f)} \quad + \quad \underset{\text{Nonsystematic}}{\varepsilon_i} \tag{2.2.6}$$

where $\varepsilon_i$ is a random shock uniquely affecting the $i$th security and representing the nonsystematic or diversifiable risk. The equation above lends itself to regression analysis (explained in the Appendix to Chapter 1). The coefficient $\beta$ (beta) is most often referred to in the context of CAPM, since it measures how much movements in the market return are transmitted to movements in the security return, meaning how much market risk is present in this security.

There is a separate regression for each security or each asset portfolio and hence a separate beta for each. Since this is a time series regression of $(r_i - r_f)$ at time $t$ on $(r_m - r_f)$ at time $t$, the estimate of $\beta_i$ may also be calculated from (1.A.5). It is the ratio of the covariance of the returns of the $i$th asset portfolio with the market portfolio, to the variance of the market portfolio returns over the available time series:

$$\beta_i = \frac{\text{cov}[r_i, r_m]}{\text{var}[r_m]} \tag{2.2.7}$$

If portfolio $i$ coincides with the efficient portfolio for the entire market, the corresponding $\beta_m$ is unity, since $\text{cov}[r_m, r_m] = \text{var}[r_m]$ in (2.2.7). The beta for $i$th asset $\beta_i$ is interpreted as the intrinsic nondiversifiable risk for the $i$th asset. Securities riskier than the market portfolio will have a beta greater than one, and thus have a higher return than the market portfolio. Securities less risky than the market will have a beta less than one. Since the market portfolio is made up individual securities in the same market, it is rare to have a beta less than zero. It should be cautioned, however, that poorly collected or atypical data can give a spurious estimate of beta, so the beta should not be taken at face value.

Since the random, nonsystematic risk is assumed to have a zero mean, the CAPM derivation suggests that the constant $\alpha$ in (2.2.6) should be zero

(testing its significance is one test of market efficiency and portfolio selection). If $\alpha = 0$, for any security for which we know the beta, we can calculate the expected return, which will compensate the investor for market risk by substituting it in the following formula:

$$E(r_i) = r_f + \beta_i [E(r_m) - r_f], \tag{2.2.8}$$

### 2.2.1 Using CAPM for Pricing of Securities

It is possible to use CAPM to determine the price of the security, Copeland and Weston (1992). Recall from Chapter 1 that that spot price of a security can be found by discounting the expected future value as $P_t = E[P_T]/(1 + r_i)^{T-t}$. The difficulty in Chapter 1 was pinning down $r_i = r_f + \theta$, the risk-adjusted return. CAPM lets us put a number on the risk premium by using beta, so our pricing formula becomes

$$P_t = \frac{E[P_T]}{(1 + r_f + \beta_i [E(r_m) - r_f])^{T-t}}. \tag{2.2.9}$$

The price of the $i$th security at time $t$ depends on the expected future value at the end of time horizon, $E[P_T]$, the length of the time horizon $(T - t)$, as well as the variables in the CAPM at (2.2.6): the estimated beta, the risk-free rate (e.g., the return on three-month T-bills) and the excess return of the market (historically around 8%).

### 2.2.2 Using CAPM for Capital Investment Decisions

Since CAPM beta is widely reported and commonly known by investors in the stock market, the management of a corporation knows the value of their firm's own beta. It also determines the (equity) cost of capital to the firm. Hence it is recommended that the management decisions regarding the choice among investment projects should check whether the project yields a rate of return that exceeds the cost of capital. If the cost of debt financing is assumed to be equal to that of equity financing, a simple way to think about this is to compute the beta for each project and reject those projects with betas exceeding the firm's own beta based on its time series data.

As seen by the formula, the higher the market risk, the higher is the required return in excess of the risk-free rate. Risk is no longer an ethereal number subject to the whim of each investor. In CAPM all rational investors require the same risk premium for identical securities regardless of the level of risk aversion. In fact a version of CAPM calls for each investor to hold the exact same portfolio of stocks regardless of the level of risk aversion, and for that portfolio to be the market portfolio. But this appears to go a little too far.

We know that most mutual funds and financial advisors tailor portfolios to the investor, and change allocation depending on his or her risk aversion

profile in terms of years to retirement, age, social values, and so on. So this leaves two possibilities:

1. Financial advisors made up these profile calculations to make themselves look busy.
2. There is something fundamentally wrong with CAPM.

We will now look at the assumptions of CAPM to test possibility 2 before we deem to accuse the financial advisors of malfeasance.

### 2.2.3 Assumptions of CAPM

1. All investors are price takers. Their expectations are homogeneous based on identical information about future returns, where the returns satisfy a multivariate normal distribution.
2. All investors are risk averse. They maximize the expected utility associated with aggregate return at the end of the time horizon $\tau$, which is common for all investors.
3. Investors can borrow or lend any amount. Within their budget, but otherwise without any penalty they do this at the risk-free rate of $r_f$.
4. Total supply of financial assets is fixed and all assets are marketable and perfectly divisible.
5. Markets are "perfect". There are no taxes nor brokerage commissions (transaction costs), and investors can do unlimited short-selling.
6. All investors know the components of the market portfolio, and that returns are normally distributed with a known means, variances, and covariances.

## 2.3 WHEN YOU ASSUME . . .

The CAPM model has led to numerous publications and there is a controversy regarding whether it can ever be properly tested by data. For example, Roll (1977) discusses in detail why various tests of CAPM themselves are flawed.

Let us consider some of the issues. Most issues arise from a possible lack of empirical validity of the assumptions of CAPM listed above. However, merely providing evidence that the assumptions do not hold true is not a complete rejection of the model, unless its insights can be shown to also be invalid.

### 2.3.1 CAPM Testing Issues

***Expected Returns.*** The CAPM theory refers to "expected" returns, whereas observed data are invariably only actual historical returns. There is no way to

know the mental expectations in the minds of market agents scattered all over the world. Since averaging over a set of stocks can be used to solve the first problem of discrepancy between observed and expected returns, Fama and MacBeth (1973) divided the New York Stock Exchange stocks into 20 portfolios and found the "market line" for each. The CAPM is said to be valid if the actual returns fall close to one of these 20 market lines.

The variability of unknown true expected returns is put in the proper context by thinking about beta as a measure of risk or volatility in relation to other investments in the *same asset class*. For example, suppose that the beta for Yahoo is 3.8. In relation to the beta of 1.2 for General Electric, Yahoo appears to be a lot riskier than GE. However, GE and Yahoo do not belong to the same peer group. So CAPM is valid to the extent that investors can use the beta to compare their investments, only if care is taken to consider the proper peer group for comparison.

***Market Portfolio.*** The data on the entire market (minimum variance efficient) portfolio are generally unavailable. In CAPM testing the measurement of market return is one of the issues. Unless true data on market portfolio return is available, instead of some proxy similar to a market index such as S&P 500 index, one does not get a true test of the CAPM per se. For example, the average firm in the S&P 500 index has a larger capitalization than the average firm in the market. The S&P 500 is based on US firms, so it is certainly not a global market portfolio. If our choice of a market portfolio is not correct, then we are not finding a true beta. Gibbons and Ferson (1985) and Campbell (1987) suggest dealing with this problem by explicitly creating another proxy for the unobservable market return in terms of various observable variables and an error term.

***Normality of the Distribution of Returns.*** The asymptotic validity of the tests of CAPM type models depends crucially on the iid-normal assumption (identically and independently distributed). The empirically observed distribution of returns, however, is generally not iid-normal in practice. In particular, one observes skewness, excess kurtosis, conditional heteroskedasticity (changing variance), as well as, serial dependence over time (this will be discussed in Chapter 4). The nonnormality by itself does not reject the CAPM. The issue is whether the results of statistical tests of CAPM-type models are reversed when the iid-normality assumption is abandoned and corrections are made. We now turn to the available evidence suggesting that conclusions of CAPM tests may well be reversed if the iid-normality assumption is invalid.

Monte Carlo simulation evidence presented by Affleck-Graves and McDonald (1989) shows absence of normality. For the US data, sophisticated econometric tools including the generalized method of moments (GMM) estimator, the so-called *J* test, and the bootstrap are used by MacKinley and Richardson (1991) and Chou and Zhou (1997) for robust tests of CAPM in

the absence of normality. They show standard CAPM test statistics have only a small bias. Tests for other country data also yield similar results, illustrated by Faff and Lau (1997). Gronewold and Fraser (2001) conclude that the iid-normal assumption does matter for the data of some countries. However, the overall literature suggests that only some CAPM implications are sensitive to the normality assumption, while many are not.

***Risk-Free Rate.*** In order to estimate beta, returns need to be calculated in excess of the risk-free rate of interest. Usually the return on the three-month US T-bill is used as the risk-free rate since there is no default risk associated with it. But default risk is not the only type of risk. As seen in Chapter 1, taxes, inflation, and market factors can change the value of the T-bill. All of these factors will affect the return. A true risk-free security should have a real return of $r_f$, period. The riskiness associated with our empirical "risk-free" security violates the theoretical existence of a market portfolio. Without a true $r_f$ and $r_m$ our estimate of beta may be meaningless.

***Other Risks.*** Beta only estimates the risk premium due to market risk, and only stock market risk at that. Any other type of risk is assumed to be either nonsystematic, or encompassed by market risk. These other types of risk could come from interest rates changing, natural disasters, fraud by employees, or political change (see Section 7.4). These sources of risk are so large that they cannot be diversified, and short-term historical measures will not forecast these risks well.

For example, betas are usually calculated with historical data from the previous one to three years (sometimes up to five years). This range will contain many ups and downs in the market with which to identify market risk using regression analysis. But how many political changes will it contain? For the United States, at most one and rarely two. For some economies there will be no changes. For volatile economies that have many coups or regime changes in a short time period, political risk and uncertainty will be much higher.

## 2.4 EXTENSIONS OF THE CAPM

There is considerable and growing literature devoted to extending and improving the CAPM model by relaxing some of the assumptions. For brevity, we will not attempt to discuss all of these extensions in detail.

Black (1972) shows that if a risk-free asset does not exist, all investors are not confined to the market portfolio and instead diversify within risky assets as follows. He shows that some risky assets, which have a zero correlation with the theoretically determined most efficient "tangency" portfolio will have zero beta. Hence one can use the return on the zero beta choice as the risk-free return. This point is mostly of theoretical interest (see Section 8.2), since most practitioners have no difficulty accepting the presence of a risk-free asset.

Some authors have extended CAPM regression (2.2.6) by relaxing the implicit normality assumption about errors. Fama used the lognormal or stable Paretian distribution in 1965, and several extensions have since been made. Myers (1972) extended CAPM by allowing for nonmarketable assets. Lintner (1969) showed that even if investors are allowed to have heterogeneous expectations, CAPM could survive. Merton (1973) and others have allowed for continuous time. In macroeconomics Consumption-CAPM is an extension of CAPM model where the portfolio that mimics the aggregate consumption (per capita) is shown to be mean-variance efficient. Intertemporal-CAPM model studies whether the changes induced by hedging over time are mean-variance efficient. Bollerslev, Engle, and Wooldridge (1988) developed conditional CAPM where the covariances are time varying. They used generalized autoregressive conditional heteroscedasticity (GARCH) to model these changes (Chapter 4).

There is only one beta in the CAPM model, and the idea of extensions by Fama and French (1996) is to introduce several sources of risk associated with directly measurable variables called factors such as firm size, price–earnings ratio (P/E), book-to-market equity ratio (BE/ME), and cash flow–price ratio (C/P). Since directly measurable factors are easier to understand and do not need the multivariate statistical technique called "factor analysis," we discuss them first, although they were developed much later than Ross's (1976) arbitrage pricing theory (APT) based on factor analysis.

### 2.4.1 Observable Factors Model

A version of the arbitrage pricing model with k *observable* factors for *i*th asset when $i \in [1, n]$ is given by

$$r_i - r_f = \alpha + \beta_{i1}(F_1) + \beta_{i2}(F_2) + \beta_{i3}(F_3) + \ldots + \beta_{ik}(F_k) + \varepsilon_i. \tag{2.4.1}$$

In (2.4.1), $\alpha$ and $\beta_{ij}$ represent the regression coefficients similar to the CAPM regression (2.2.6), except that here there are more sources of systematic risk than just the market return. This equation can be written in a more compact matrix notation by combining the $k$ time series of factors into one $T \times k$ matrix denoted by $F_{1:k}$ and combining the beta coefficients into one $k \times 1$ vector $\beta_i$, where the subscript suggests that coefficients are distinct for each asset. The CAPM extension with observable factors in matrix notation is

$$x_i = \alpha + F_{1:k}\beta_i + \varepsilon_i. \tag{2.4.2}$$

How do we find the observable factors? Basu (1983) found that low price–earnings ratio (P/E) stocks meant higher returns, suggesting that P/E can be a factor. Banz (1981) found that small-cap stocks had higher returns. Lakonishok, Shleifer, and Vishny (1994) found higher returns associated with average return and book-to-market equity ratio, and cash flow–price ratio.

Fama and French (1996) gave a more comprehensive study of the factors mentioned by the authors cited above. They found that by observing the returns of corporations that stocks with higher betas did, indeed, yield a higher return on average but that beta only explained a small fraction of the differences in returns between stocks. There were, however, other factors noted above that aided the CAPM in explaining excess returns. Fama and French thus theorized that these items must be additional risk factors based on their empirical evidence that (2.4.2) provided a good fit to the data.

***Size Factor.*** Fama-French factors try to pick up sources of risk not represented by the market portfolio. The size factor, for instance, could proxy for the ability of a firm to rebound in a down market. It has long been known that small firms give a higher stock return on average than large firms with the same beta. The Fama-French model uses this to imply that perhaps access to credit, reserves of resources, and the like, make being a small firm inherently more risky. While large firms may be able to weather a storm, small firms are more easily blown away.

The same intuition applies to the book-to-market equity value of a firm or BE/ME ratio. As we saw in Chapter 1, the stock value of a healthy firm is the present value of all future earnings. This number should be in excess of the price paid for the assets of a company if the company is productive. Seeing a firm with a book value close to the market value is a sure sign that investors are seeing the company more as a garage sale than an investment. Thus book-to-market value signals likelihood of default.

It stands to reason that the probability of default cannot be reflected in the CAPM beta, since as soon as a firm files for bankruptcy, it is no longer in the market and no longer used in calculating market returns (known as survivorship bias). Therefore the Fama-French factors try to fill in the gaps that the single factor beta leave out in CAPM. The construction of factors $F_j$ in (2.4.1) can be done in sophisticated ways. One can first classify stocks into deciles or similar categories (high, medium, and low) values for the factors and then focus on one factor, while trying to hold others fixed (see D'Souza, 2002, for international Fama-French estimations). This method gives a multifactor CAPM that starts with market risk and adds additional factors as needed within each category.

### 2.4.2 Arbitrage Pricing Theory (APT) and Construction of Factors

One of the strongest statements made by CAPM is that the only risk that should be compensated is systematic risk as shown by a representative stock market portfolio. Arbitrage pricing theory (APT) broadens CAPM by not putting all of its eggs in the market portfolio basket. Fama-French type APT is stated above in (2.4.1) and includes additional *directly observable* risk factors that contain nondiversifiable risk that may not be completely reflected in

the market portfolio. The APT proposed by Ross (1976) considers *indirectly observable* factors constructed by using "factor analysis," a multivariate statistical techniques.

Factor analysis has long been popular in psychometrics to assess personality from psychological test scores, and this achieves (1) dimension reduction without losing common variance of original data and (2) closeness to the normality assumption by reducing error variance. Factor analysis uses the results of a multivariate technique called principal component analysis (PCA) and "rotates" them to obtain fundamental representations. Section 8.3 uses a constructed finance example to explain PCA and the related matrix algebra. The aim is to construct a low-dimensional summary of the data on $n$ assets in the market, and construct the minimum number of factors needed for the APT regression.

Recall that in (2.4.1) we have one regression for each of the $n$ assets (or portfolios) indexed by $i \in [1, n]$ over T time periods indexed by $t \in [1, T]$. When we do not assume observable time series for the factors, we have to construct them by somehow summarizing the information in the original data. The "market return" in CAPM usually refers to hundreds of stocks in the market (e.g., S&P 500), but these stocks may be redundant, washing out important information through aggregation. Summarizing the information in their comovements over time into a handful of factors as index numbers requires some algebraic tools discussed in Chapter 8, Section 8.3. The tools include the singular value decomposition (SVD), principal components analysis (PCA) and factor analysis.

Let $X$ denote an $n \times$ T matrix of "excess returns" with elements $\{x_{it}\}$ for the $i$th asset at time $t$ (see Chapter 8 for more on this notation). The technique of decomposing the return matrix into eigenvectors distills the characteristic vectors that can replicate the entire matrix. Therefore no information is lost, but there is also no redundancy. Using a matrix $V$ of eigenvectors of $X'X$ associated with the largest $k$ eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ the corresponding columns of $V$ are $(v_1, v_2, \ldots, v_k)$. Using this subset ($k < n$), we construct principal components that are then used to construct the $k$ factors $F_{1:k}$ appearing in (2.4.2). Note that the principal components $p_j$ for $j = 1, \ldots, k$ involve "dimension reduction" and summaries of the entire data on market returns for $n$ assets. After dimension reduction we retain only the coordinates of the data along the most important $k$ "principal axes." The first principal axis captures the most variable (highest variance, highest eigenvalues) dimension in the $n$-dimensional space for $n$ assets over $T$ time periods. The second principal axis is orthogonal (perpendicular) to the first and maximizes the variance. This is done sequentially until all dimensions are exhausted.

Now write the singular value decomposition $X = USV$, where the matrix $U$ contains the n columns with coordinates along the principal axes and $S$ is a diagonal matrix of square roots of eigenvalues of $X'X$ known as singular values in decreasing order of magnitude, and $V$ contains the corresponding eigen-

vectors. The *SVD* makes it is intuitively clear how replacing the trailing singular values by zeros reduces the dimension from $n$ to $k$. The principal components are equivalently obtained simply by the matrix multiplication (derived in Chapter 8):

$$X(v_1, v_2, \ldots, v_k) = (p_1, p_2, \ldots, p_k) = F_{1:k}, \tag{2.4.3}$$

where both $X$ and the eigenvectors $v_j$ of $X'X$ are observable, although the latter are sensitive to units of measurement. Thus we have constructed the $T \times k$ matrix of principal components as factors ready for substitution in (2.4.2). It is possible to use a mixture of these and some macroeconomic variables including interest rates, exchange rates, industrial production, in an extended definition of $F_{1:k}$. Since the columns of $X$ in our example are asset returns in comparable units for all columns and mean something in dollars and cents, it is possible to work with $X'X$ defined in original units. If we include interest *rate*s, *indexes* of industrial production, or other macro variables, their units are not necessarily comparable to return on assets. If this leads to strange results, one can always include these data as additional columns of $X$ and then standardize the $X$ data by making $X'X$ a correlation matrix, as in Vinod and Ullah (1981). Since standardization is a linear operation involving means and standard deviations, it is a simple matter to un-standardize the data. The main point is that APT extends the CAPM by incorporating additional regressors from selected few principal components of returns on most (if not all) market assets, as well as, macroeconomic variables.

If the length of time series $T$ is short compared to the number of assets $n$, then one can reduce the number of principal components by taking the smaller of $n$ and $T$. It may then be better to write the data matrix $X$ after transposing it as $n \times T$ whenever $n > T$, and then compute the principal components to limit the number calculations. In short, the idea behind factor analysis is to focus on a matrix of covariances or correlations among possible variables describing the expected returns instead of a single beta. It is not surprising that the statistical fit of the CAPM equation can be improved by adding more right-hand variables and in the process some unrealistic assumptions of CAPM can be avoided (Chen et al., 1986).

### 2.4.3 Jensen, Sharpe, and Treynor Performance Measures

Among the practical implications of asset pricing models we recognize the tools offered by Jensen, Sharpe, and Treynor to compare the performances of different portfolios. We consider these tools practical, because ordinary investors can use them in choosing among mutual funds and portfolio managers can use them in comparing a wide variety of possible financial instruments. We start with Jensen's alpha defined as

$$\text{Jensen's alpha} = E(r_i) - r_f - \beta_i[E(r_m) - r_f]. \tag{2.4.5}$$

Jensen's alpha measures abnormal performance by the intercept of the CAPM equation (2.2.8). It is easy to compute Jensen's measure from the time series regression of excess returns on excess returns of market benchmark assets. Also the sampling properties and confidence intervals for Jensen's alpha are well known from regression theory.

A generalized Jensen's measure is defined in terms of a similar regression except that CAPM regression is replaced by one where, instead of risk-free rate, one uses portfolio of benchmark assets with the property that they have zero beta (zero slope) of the regression equation. The generalized Jensen measure relaxes the assumption that market portfolio is mean-variance efficient. It answers the question whether the investor can improve the efficiency of the portfolio by incorporating a new asset in the portfolio.

Sharpe (1966) and Treynor (1965) portfolio performance measures are widely cited and used in the literature and pedagogy of finance. Consider the following scenario in which the relative performance of $n$ portfolios is to be evaluated. In this scenario, $r_{i,t}$ represents the excess return from the i-th portfolio in period $t$, where $i = 1, 2, \ldots, n$. A random sample of $T$ excess returns on the n portfolios is then illustrated by $r_t' = [r_{1t}, r_{2t}, \ldots, r_{nt}]$, where $t = 1, 2, \ldots, T$ and where $r_t$ is assumed to be a multivariate normal random variable. The unbiased estimators of the mean vector and the covariance matrix are

$$\bar{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_{i,t} \quad \text{and} \quad S = \{s_{ij}\} = \frac{1}{T-1}\sum_{t=1}^{T}(r_{i,t} - \bar{r}_i)(r_{j,t} - \bar{r}_j), \tag{2.4.6}$$

where the notation is close to Jobson and Korkie (1981). These two estimators are then used to form the estimators of the traditional Sharpe and Treynor performance measures.

The population value of the Sharpe performance measure, usually called Sharpe ratio for portfolio $i$, is defined as

$$\mathrm{Sh}_i = \frac{\mu_i}{\sigma_i}. \tag{2.4.7}$$

The corresponding sample estimate uses the mean excess return over the standard deviation of the excess returns for the portfolio. Consider a basic graph of the risk-return trade-off, where we plot the return $\mu_i$ on the vertical axis and risk measured by the standard deviation $\sigma_i$ on the horizontal axis. This simple graph represents the notion that higher returns are generally associated with higher risk. The Sharpe ratio in such a graph is the slope of the straight line from the bottom left to the top right. The Sharpe ratio $\mathrm{Sh}_i$ is the tangent of the angle of that straight line with the vertical axis. The higher line has a higher slope and represents a higher average return for the same risk. Hence it is clear that a "more efficient" or more desirable portfolio should have, in general, a higher Sharpe ratio.

If one wishes to find out what will happen to the maximum attainable Sharpe ratio described below, Jensen's alpha is related to Sharpe's measure in interesting new ways. DeRosen and Nijman (2001) use Jensen's alphas together with the regression error covariance matrix to estimate the potential improvements to maximum Sharpe ratio.

The population value of the Treynor (1965) performance measure gives return per unit of market risk:

$$\mathrm{Tr}_i = \frac{\mu_i \sigma_m^2}{\sigma_{im}} = \frac{\mu_i}{\beta_i} \tag{2.4.8}$$

for $i = 1, 2, \ldots, n$, where m is the market proxy portfolio often denoted by the S&P 500 index and $\beta_i$ is the beta from the CAPM. For new tools including bootstrap for statistical inference on Sharpe and Treynor performance measures, see Vinod and Morey (2000).

This chapter has concentrated on measuring the risk that cannot be eliminated from a stock portfolio. This risk factor can be characterized by market risk, but more commonly it is seen as coming from several inherent factors in the economy. In the next chapter we will detail how even this risk can be completely eliminated (for a price, of course) through buying and selling derivatives contracts such as call and put options.

## APPENDIX: ESTIMATING THE DISTRIBUTION FROM THE PEARSON FAMILY OF DISTRIBUTIONS

The Pearson system was referenced in equations (2.1.4) to (2.1.6). It is the family of solutions $f(y)$ to the differential equation

$$\frac{d \log f(y)}{dy} = -(a+y)[B_0 + B_1 y + B_2 y^2]^{-1} f(y). \tag{2.A.1}$$

A solution of this equation $f(y)$ is always a well-defined density function. The shape of $f(y)$ depends on the Pearson shape parameters ($a$, $c_0$, $c_1$, $c_2$) given below. These parameters can be expressed in terms of the first four raw moments of the distribution. However, these expressions are tedious. In terms of the central moments we have the following simpler expressions:

$$c_0 = \frac{\mu_2\left(3(\mu_3)^2 - 4\mu_2\mu_4\right)}{2\left(9(\mu_2)^3 + 6(\mu_3)^2 - 5\mu_2\mu_4\right)}, \tag{2.A.2}$$

$$a = c_1 = \frac{-\mu_3\left(3(\mu_2)^2 + \mu_4\right)}{2\left(9(\mu_2)^3 + 6(\mu_3)^2 - 5\mu_2\mu_4\right)}, \tag{2.A.3}$$

$$c_2 = \frac{6(\mu_2)^3 + 3(\mu_3)^2 - 2\mu_2\mu_4}{2\left(9(\mu_2)^3 + 6(\mu_3)^2 - 5\mu_2\mu_4\right)}. \tag{2.A.4}$$

Hence we use empirical estimates of central moments to plug on the right sides of (2.A.2) to (2.A.4) and then solve the differential equation to get the underlying density. This was done in (2.1.9) for the artificial data of Chapter 1 in Section 1.4. Computer-aided calculations are discussed in Chapter 9.

CHAPTER 3

# Hedging to Avoid Market Risk

## 3.1 DERIVATIVE SECURITIES: FUTURES, OPTIONS

We learn from Chapter 2 that all risk cannot be diversified away from a portfolio, no matter how many stocks are included. Idiosyncratic risk of individual companies can be diversified away, but systematic risk is common to all companies to some degree. This systematic risk, or market risk, arises because all companies are on the same planet, employing human beings, and trying to make a profit. There are events, such as weather, recession, war, and technical innovation, that are not isolated to specific companies but can affect almost every one of them. Unexpected upswings in stock price can also cause a loss if one is planning to buy a stock, but the upswing raises the price before the order goes through.

Simple investment in the stock market by buying and selling stock, then, cannot give peace of mind against a sudden downturn in a stock price that can devastate a portfolio. This chapter is concerned with strategies involving "derivative" securities for reducing the risk. Consider the following two investors with their fortunes tied to the outcome of the stock market on March 31: Jack, an unemployed college student who on March 1 already owns 100 shares of Company X and needs to sell $1200 worth of shares to pay his rent on March 31, and Jill, a businessperson who intends to invest $1200 from her next paycheck on March 31 to buy shares in Company X for her retirement account. Company X's shares are expected to sell for $12 per share on March 31, so Jack's 100 shares would provide him the $1200 and Jill's $1200 will let her acquire 100 shares for her retirement account.

If the price on March 31 is higher (e.g., $20) than the expected $12 per share, Jack gets a windfall by being able to pay his rent by selling only 60 shares, while Jill will be hurt, since her $1200 will purchase only 60 shares. By con-

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

trast, if the price on March 31 is lower than expected, Jack gets evicted, and Jill lives well upon retirement. Both Jack and Jill have specific goals for investing, and they are both subject to risk from their investments but in opposite directions.

Since both Jack and Jill are expecting the price of Company X stock on March 31 to be $12, and would prefer to avoid market risk, the solution to their problem is a forward contract between the two. A *forward* contract is an agreement between two parties to trade an asset at a future time (maturity date) for a pre-specified price. If Jack and Jill enter into a forward contract to exchange 100 shares of Company X stock for $12 on March 31, then the risk of the stock price differing from the expected value is completely hedged. They are both certain the trade will go through at the agreed-upon $12.

Since there is no actual trade of an asset at the point the forward contract is agreed upon, forward contracts are called derivatives. The value of a forward contract does not come from its own intrinsic worth; instead, it *derives* from its ability to transform (postpone) the trade of Company X's shares (from Jack to Jill) into a future trade with a certain price. Company X stock would then be the underlying asset from which the forward contract derives its value.

In our simple example we have a forward contract with two fortunate individuals who coincidentally found an investor with the exact opposite needs. In the market for forward contracts, however, there may be a preponderance of investors on one side or another, which would push the forward price away from its expected value. A market consisting of sellers (like Jack) would be delighted to receive the full price of $12, but some sellers may also be willing to receive a little less (sell at a discount) in order to eliminate risk. How much less they would be willing to receive will depend on how fearful they are of risk of receiving much less than $12 on the maturity date. It is all a matter of how risk averse they are. (For a discussion of risk aversion, see Chapter 6.) There will be very few investors willing to take a substantial discount, but as the discount is reduced, more investors will be willing to trade. See the right-hand side panel of Figure 3.1.1.

On the other hand, a market consisting of buyers (like Jill), while happy with $12 per share, may be willing to pay a little more (premium) with the forward contract for making sure that the future price will not exceed $12. Since the $12 amount enters the calculation asymmetrically, the supply and demand curves for a forward contract are oddly shaped with a kink at $12. The left side of Figure 3.1.1 illustrates the demand curve and right side the supply curve.

Depending on the relative number of sellers and buyers for a forward contract, the price may be either above the expected value, called *contango*, when there are more buyers, or below the expected value, called *backwardation*, when there are more sellers. Forward contracts are actually more common in currency and commodity trades, where the same large buyers and sellers routinely trade with each other. For small investors in the stock market, forward contracts are a difficult proposition since to enter a forward contract one has

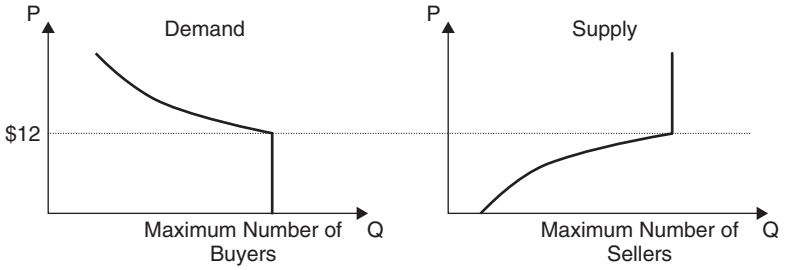



**Figure 3.1.1** Kink in the demand and supply curves in future markets

to first find an investor with an equal and opposite hedging need (a Jack for a Jill), and each contract is between two individuals. Such bilateral forward contracts are difficult for two reasons:

1. A bilateral forward contract in the stock market introduces default risk because one of the parties to the contract is certain to be a loser when the outcome is compared to the spot market price at the maturity date. On March 31, if the stock price in the spot market is above $12, Jack will be getting less money with his forward contract with Jill than what he would have gotten on the spot market. On the other hand, if the spot price is below $12, Jill would be the loser, since she has agreed to buy at $12 from Jack when she could have bought the same stock cheaper on the spot market. On the maturity date, the spot price is known, and the situation is not the same as it was when the forward contract was initiated: the risk is gone. Therefore with a forward contract this regret creates an incentive for one party to default. Since it is not common to buy the same stock again and again at regular intervals from the same seller, there is a smaller chance for repeat business between two parties. Forward contracts are thus feasible only if both the buyer and the seller have a clean reputation and an incentive not to default.

2. Individualized bilateral forward contracts are illiquid before the maturity date. If Jack enters into the forward contract, and then decides to sell the stock early, perhaps due to a family emergency, the forward contract still exists. He either needs to find another investor to take over the contract, or compensate Jill enough to void the contract. Of course, suing the defaulting party in a court of law is always feasible but costly and time-consuming. Either way, there are substantial transactions costs for enforcing or getting out of a forward contract that lead to the illiquidity of such bilateral contracts.

### 3.1.1 A Market for Trading in Futures

Since hedging is supposed to reduce risk, and for the average investor bilateral forward contracts just add new types of risks, the futures exchange was

introduced. A futures contract is a specific type of forward contract that is sold in standardized units in a centralized exchange, eliminating the bilateral aspect completely. These features also reduce the illiquidity because the standardized contracts give them a wider market, and the centralized exchange reduces search costs, and provides enforcement of the contract to eliminate default risk. In order to trade on a futures exchange, one must have a clean reputation, and provide enough collateral so that he or she will follow through even on losing contracts.

The standardization also creates an interesting new market where some of the underlying assets on which futures contracts are bought and sold do not exist, and actual delivery of the physical asset is not expected. For example, futures on the S&P 500 stock index are bought and sold on the Chicago Mercantile Exchange (CME). If one sells the S&P 500 future for a March contract, it is not expected that come March you actually buy shares of each of the 500 companies in the index in the correct proportion to exchange for the sales price. Purchasing that volume of shares would involve a huge transactions cost and considerable time and effort. Also many of the assets in standardized futures contracts do not even exist, since the contracts specify all the details of the product in order to have a homogeneous asset. For example, 30-year US Treasury bond futures traded at the Chicago Board of Trade are bonds with a 6% coupon payment. Good luck trying to find a 6% return on a currently issued Treasury bond in the current low-interest environment. Yet standardization keeps the underlying asset consistent throughout time, whereas returns on actual bonds will fluctuate with the market.

Rather than deliver the physical commodity or asset, which can be prohibitively costly, if at all feasible, the buyer or seller receives the price difference between the contracted price of the futures contract and the market price at maturity if the contracted price difference is in his or her favor. If the price difference is in the opposite direction, the buyer or seller pays the difference. For instance, in February Jill hedges the level of the stock market by buying a March S&P 500 futures contract for 900 on the CME. S&P 500 contracts have a value of $250 times the index value, so she is agreeing to pay $225,000 for the contract. If Jill holds the contract to maturity, the price on the settlement day (at maturity, the third Friday of the month for this contract) will determine what she owes. If the S&P 500 is at 950 on the settlement day, Jill has a contract to buy at 900, which is 50 points lower than the market. Therefore Jill will receive 50($250) = $12,500 to make up the difference between her contracted price and the market price. If the index in March is only at 850, Jill would owe $12,500 since the S&P 500 stocks would be cheaper at the market price than the contracted price. This money would then go to the individual in the opposite position of the contract who sold the March S&P 500 futures. Even more common is that an investor will carry out an offsetting transaction (buying futures to cover a sell, or selling to cover a buy) at some point before the settlement date so that the gain or loss is locked in.

The absence of physical assets in futures contracts make them somewhat mysterious, and it is tempting to liken the futures markets with gambling casinos. Accounting and tax treatment of some of futures contracts is not straightforward, requiring expertise and adds to their mystery. For example, a futures contract can enable an unscrupulous firm to treat future revenue as current with a view to defraud and mislead the stock market analysts, a strategy that the Enron Corporation used to an extreme measure before its bankruptcy in 2001.

***Margin Accounts.*** There must therefore be a mechanism to ensure that Jill will pay the extra $12,500 for the stocks when they are cheaper ex post in the open market (without the futures contract). The margin account is the mechanism that keeps the losers from walking away. For a S&P 500 futures contract, an investor must deposit an initial margin of $19,688 with the exchange. Each day this margin account is "marked to market," which means that if the contract loses value, the loss is deducted from the margin account, and if the contract gains value, the gain is added to the margin account. If the margin account has enough losses so that it dips below the maintenance margin ($15,750 for the S&P 500), then the investor must deposit additional funds to bring the account back to the initial margin. The difference between the initial and maintenance margin keeps investors from getting margin calls for minor losses, but the maintenance margin ensures that at least $15,750 is already deposited with the futures exchange.

Unless the contract drops more than $15,750, there is no default risk because there is enough collateral to cover a loss. The margin amounts are generally set so that it is extremely unlikely that a contract will have a loss exceeding the maintenance margin (using VaR). Furthermore, if an investor fails to meet the margin call for additional funds, the exchange simply puts through an offsetting contract (for a seller, they will put in a buy; for a buyer, a sell) so that the investor is contracted for a buy and sell price and can have no further losses. It is also common that in more volatile markets, the initial margin and the maintenance margin are equal to each other to further guard against a potential default.

***Marking to Market.*** The system of marking to market is a valuable service that the exchange provides to guard against default risk, but a pesky side effect of buying on margin is an increase in overall speculation. The predominant investor in the futures market is not the hedger wanting to secure a purchase or sales price but the speculator or arbitrager, the former betting on price movements and the latter exploiting minute price discrepancies. The futures market is ideal for these practices since the margin account is the only investment needed at the time of the purchase, but the speculator gains (or loses) the whole amount represented by the change in the price of the underlying asset. In our example above, the S&P 500 contract worth $225,000 could be

purchased by supplying the $19,688 initial margin. This is an investment of only 8.75% of the underlying asset value.

In the scenario where the index gains 50 points and Jill receives $12,500, which is a 63.5% return on the up-front money when S&P 500 only increased 5.56%. Buying on margin in effect magnifies the return. Of course, futures contracts can magnify returns downward also. In this contract, there is a potential for up to a $225,000 loss from the $19,688 initial investment, or a huge 1142.8% loss.

### 3.1.2 How Smart Money Can Lose Big

Private investment partnership designed to trade in securities and financial derivatives are called hedge funds. Despite the name, these can be more risky than other investments.

Unlike regular mutual funds, hedge funds are not subject to many Securities and Exchange Commission (SEC) regulations. They are not permitted to advertise, and their managers need not be registered as investment advisers. Considerable information about hedge funds is available on the Internet. They are available to a limited number of qualified or accredited investors, mostly those whose net worth is in millions. The General Partner of the fund often receives 20% of the profits and a 1% management fee. The Van U.S. Hedge Fund Index underperformed the Standard & Poor's 500 between 1995 and 1997 and outperformed it between 1999 and 2001. Among arbitrage strategies used by fund managers are (i) Some attempt to exploit anomalies in the prices of corporate convertible bonds, warrants and convertible preferred stock. (ii) Some managers look for price anomalies associated with specific events, such as bankruptcies, buybacks, mergers and acquisitions. (iii) Some managers exploit anomalies in pricing of bonds and shares of distressed securities for companies that are in or near bankruptcy. (iv) Some managers neutralize the exposure to market risk by identifying overvalued and undervalued stocks in conjunction with long and short positions.

The experience of Long-Term Capital Management (LTCM), the Greenwich, Connecticut, hedge fund illustrates the potential dangers of buying on margin in the derivatives markets. The objective of LTCM was to model the financial condition of countries around the world and to exploit differences between market prices and the predicted prices from their models. And who wouldn't trust their models? Behind the models were two Nobel Prize winners in economics, and several stars from Wall Street. Initially, LTCM made 30% to 40% returns, but the opportunities became scarce, so margins were increased to augment the return.

LTCM owned $125 billion in financial assets with only $4.8 billion in capital. This means that a 1% loss on their portfolio will be a $1.25 billion drop in capital, or over a quarter of their entire investment. Then the unthinkable happened. Several variables outside their models started to turn sour. Asian markets all fell; Russia devalued its currency, instantly making investments

worthless. To cover these losses, LTCM invested more heavily in derivatives to hedge, but at one point their positions totaled an unwieldy $1.25 trillion. Eventually the New York Federal Reserve Bank put a stop to the dangerous spiral of LTCM and organized a private bailout and increased oversight in 1999. LTCM experience gives a cautionary tale that even unlikely events happen sometimes. Also, when variables change that are not in your model, the predictions of the model should not be taken at face value.

### 3.1.3 Options Contracts

Options contracts are a form of financial derivative where the investor has the right to buy or sell the underlying asset, not the obligation as with futures contracts. Options eliminate the downside risk evident in futures contracts by making the trade optional at the choice of the investor holding the option (called the *long position*). If making the trade will result in a loss to the investor, then the investor simply chooses not to exercise the option and avoids any further loss.

There are two basic types of options: a *call option*, which is the option to buy the underlying asset, and a *put option*, which is the option to sell the underlying asset. Investors can also either buy the option (*long position*) or write the option, meaning they can sell another investor the option (*short position*).

For example, consider a call option on Company X stock for $12 per share (strike price) with a settlement date in one month. If the price of Company X in one month is less than $12, then it will be foolish to purchase the stock using the option since the stock can be directly purchased for less in the market. When the option is not exercised it is *out of the money*. If the price of Company X's stock in one month is more than $12, say $13, then the call option saves $1 for every share purchased. Therefore the option will be exercised, and the option is called *in the money*. It is also evident that for every higher dollar that Company X stock is priced above $12, there is an additional dollar of savings to the investor holding the call option. Figure 3.1.2 provides a graph of the value of an option of a long position in a call and put.

Since the investor in the long position has the choice of whether or not to exercise the option, there can be no further loss on the stock trade. The writer of the option, on the other hand, has a short position. He must comply by yielding the possession of the asset if the purchaser exercises the option, and the writer knows the option will only be exercised if it is not in the writer's favor. So the payouts to an investor writing an option are shown in Figure 3.1.3.

Why, then, would anyone write an option? The answer is that a buyer of an option must pay a premium for the benefit of always being on the wining side. The net payoffs are shown in Figure 3.1.4.

Now, if the price increases and the option is exercised, the writer gains the premium to offset some of the loss, and if the option is not exercised, the writer keeps the premium without any loss. With the premium, each position has the possibility of gain or loss, but the risk is different. For the long position, loss

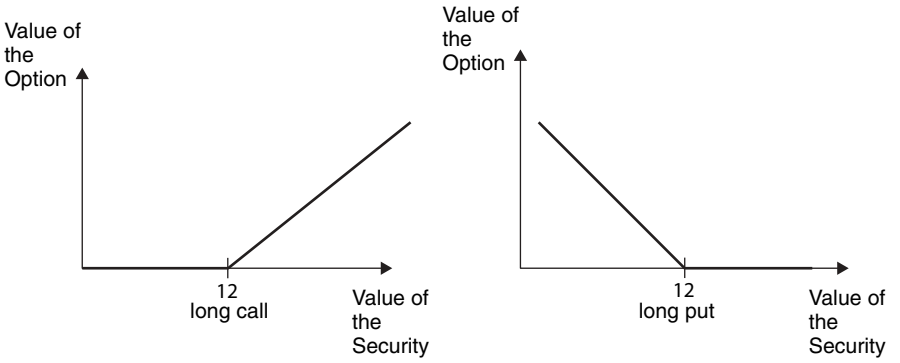



**Figure 3.1.2** Value of the long position in an option

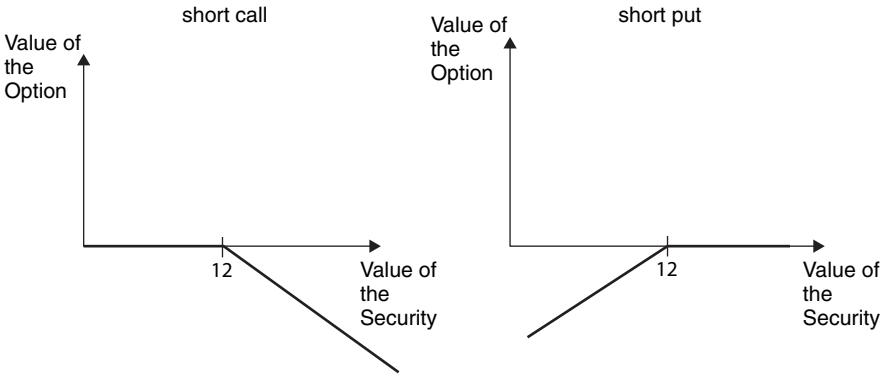



**Figure 3.1.3** Value of the short position of an option

is limited to the premium, but gains are virtually unlimited. For the short position, the most they will gain is the premium, but losses are limited only to all (100%) of the initial investment.

This separation between gains and losses foreshadows an important role options play when predicting downside risk. More than being a method to reduce risk by locking in a minimum or maximum price, the observable market value of these options can give valuable insight into market perception of downside risk of a company. This will be revisited in Chapter 5. For the moment, the rest of this chapter deals with the fundamentals of analyzing and pricing derivatives to understand this valuable tool better.

We have covered the basic structure of derivatives contracts. In practice, derivatives are combined according to the desired level of risk. But if one understands the pricing behind the simple derivative contract, pricing of more complicated derivatives just involves separating and combining various component parts.

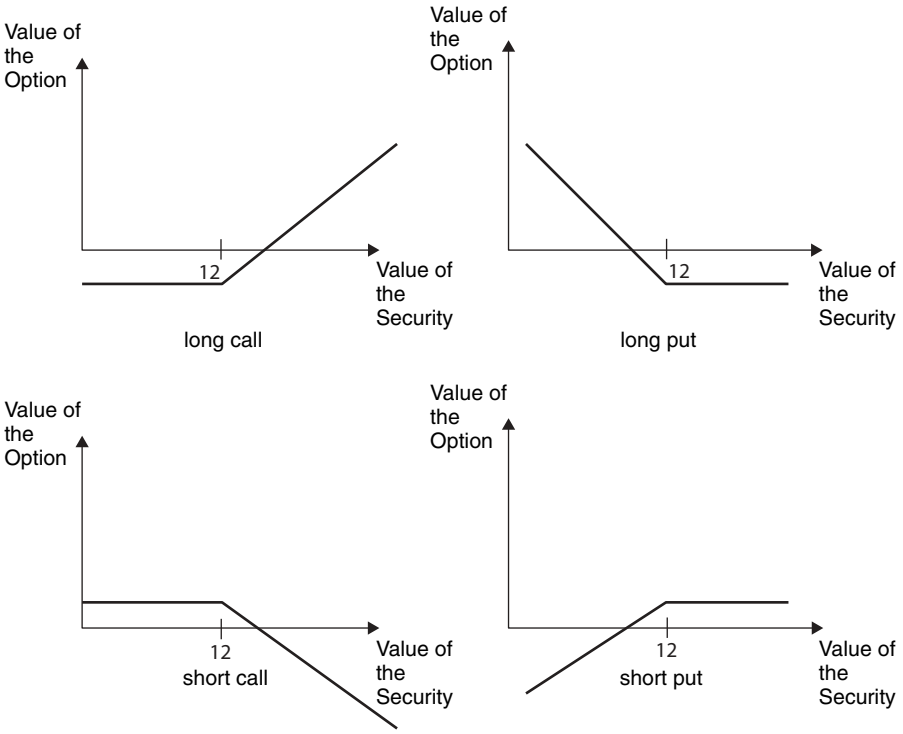



**Figure 3.1.4** Value of an option after the premium

## 3.2 VALUING DERIVATIVE SECURITIES

We will begin this section by talking specifically about option valuation, although the pricing techniques for options discussed here can be used for any derivative security. The basis of any derivative valuation is that the risk of the derivative can be completely hedged by holding a portfolio of the derivative and the underlying asset in such amounts that it guarantees the risk-free return.

When dealing with stocks, risk can never be completely diversified because of the presence of systematic market risk that affects all stocks. But with derivatives, there is an underlying asset that is subject to the exact same shocks, since the value of the derivative is based on the value of the stock. We will see that both can be played off each other to eliminate all risk, idiosyncratic risk and systematic market risk.

Let's begin by looking at a simpler world, and we will build upon the insights we find there. Assume there is a security that is currently priced at \$50 per share in a market where the return on a safe, risk-free investment is 5%. The future of this stock is risky, though. There is a 25% chance that the

company's new process will work by the year's end, and the stock will appreciate in value to $96 per share. But there is a 75% chance that the new process will not work, and the misallocated resources will cause the company's stock to decline to $40 per share. Note that we are making a very unrealistic assumption here that the stock price cannot be anything outside the set {$40, $50, $96}.

If we just look at the stock investment itself, the expected price of this stock in one year is 96(0.25) + 40(0.75) = $54. This yields an expected return of 8%, where the extra 3% compensates the investor for the additional risk associated with holding this stock instead of a risk-free security (bank deposit).

An investor who may not wish to live with the possibility that the stock will only be worth $40 may wish to purchase a put option (right to sell) for a one-year maturity with a strike price of $50. But how much should this put option cost? We know the price must be above zero because the holder of this option will not exercise the option for a loss.

The method used to price an option is known as *arbitrage pricing*. Securities with the same risk and same payoffs should have the same return. Otherwise, arbitragers would exploit the difference and drive the returns back to equality. We know that the risk-free security has a return of 5%, so if we can construct a risk-free portfolio of the stock and option, the return should still be 5% to discourage arbitrage.

Now let us collect in Table 3.1.1 what we know. First, look at a combination of one share of the stock and a put option contract for 5.6 shares. Why 5.6 shares? The following arithmetic needs this for creating a perfect hedge. In the prosperous scenario of the company where the company's stock is $96, the option contracts are not exercised and the pay is $0. In the unsuccessful scenario, the one stock is worth $40, and the put options contracts are exercised for an additional $10 on each of the 5.6 shares, or $56. So the total revenue in the unsuccessful scenario is $40 + $56 = $96, which is exactly the same as in the prosperous scenario! This portfolio is clearly risk free, since the value in one year is always $96. Therefore the initial price of this portfolio at the beginning of the year must yield a return of only 5% for the portfolio, similar to a risk-free bank deposit. Thus the price of the hedge portfolio is $96/1.05 = $91.43.

If the total portfolio costs $91.43, $50 of that price goes to buy one share of the stock, the remaining $41.43 goes to buy the put options for 5.6 shares.

**Table 3.1.1 Example of a $50 Stock Restricted to {$40, $96} Value in One Year**

| | Stock Value | Option Value | Value of the Risk-Free Asset |
|---|---|---|---|
| Now | 50 | ? | 100 |
| In One Year | | | |
| 25% Probability | 96 | 0 | 105 |
| 75% Probability | 40 | 10 | 105 |

If the put options for 5.6 shares are bought for \$41.43, each put option must cost \$41.43/5.6 = \$7.40. Thus we have magic. We are able to assign a price to the put option by using the perfect hedge, without having to know anything about the market prospect for the underlying stock.

While it is satisfying to have a solution, it is easy to see that we have used drastic simplifications. So far our option formula only works in a world with one time period for a stock that only has two possibilities. But we have two important insights:

1. The probability of a prosperous outcome has no effect on the value of the option. Since the timing is the same for both the option and the stock, the number of shares required to perfectly hedge does not depend on the probability of the stock increasing in value.

2. If the possible outcomes are more spread out (higher variance) it will take fewer puts to even out the payments and construct the perfect hedge. Therefore increased variance increases the value of the option.

### 3.2.1 Binomial Option Pricing Model

The binomial pricing model takes the simple scenario and expands it to multiple periods. In the binomial model we still assume that the stock is going to go up \$$k$, with probability $\pi$, or down \$$h$, with probability $(1 - \pi)$. But we extend the analysis repeatedly over $t$ periods. If we draw out the possible time paths, we see from Figure 3.2.1 that the number of permutations of upward and downward movements increases as the number of time periods increase.

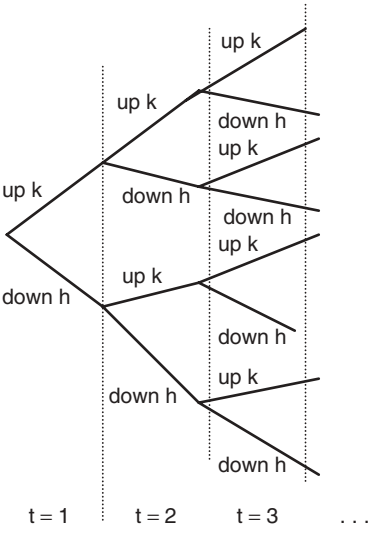



**Figure 3.2.1** Binomial options model

Even with so many possibilities, risk can be perfectly hedged just as with the simpler case to construct a risk-free portfolio, and the insights are similar: probability of success does not change the value of the option, while higher volatility increases the value of an option.

Both models clarify that for options prices volatility is not a bad thing. This insight takes some getting used to. After building up risk premiums for stocks for numerous factors, it turns out that in the world of options, risk in the form of volatility can be a profit opportunity.

The explanation for the positive effect of volatility on option prices emanates from the choice given to the buyer of an option: they do not have to exercise the option. Therefore for high-risk underlying assets, if there is a large swing upward, the buyer of a call option can exercise the option and redeem the increase, and if there is a similar large downswing in the price of the underlying asset, the buyer of a call option can walk away and not purchase the stock at all. Therefore call options act as insurance to eliminate downside risk. Of course, like any insurance policy, it requires the payment of an insurance premium. This suggest the following additional insight. We can use observable option premiums quoted in the market as one tangible measurement of downside risk.

### 3.2.2 Option Pricing from Diffusion Equations

In this subsection we consider a more technical method of options pricing that uses the diffusion equations from Section 1.4 to price options in continuous time (rather than discrete jumps in time) and for continuous changes in stock prices. This method is commonly referred to as the Black-Scholes pricing formula, named after the developers of the method.

While obviously more complicated, the Black-Scholes method uses the same technique of finding a portfolio that perfectly balances all risk, so that it may be priced at the risk-free rate of interest. Recall that in the example of Table 3.1.1 the certain year-end yield of \$96 was priced using the risk-free 5% interest rate.

Consider a stock following the diffusion equation from (1.4.3):

$$\frac{dS}{S} = \mu dt + \sigma dz. \tag{3.2.1}$$

Also consider a call option written on that stock. Let $C$ denote the price of the call option and insert the primes on parameters of (3.2.1). The call option diffusion equation is

$$\frac{dC}{C} = \mu' dt + \sigma' dz. \tag{3.2.2}$$

Although we consider a call option here, the method works the same for a put or any derivative contract. The option follows the same general diffusion

process, but with its own drift $\mu'$ (expected return) and standard deviation $\sigma'$ denoted by primes. Since it is written on the stock, the shocks $dz$ will be the same force affecting both securities. This means that if an investor shorts $[-\sigma'/(\sigma - \sigma')]$ shares of the stock and buys $Z = \sigma/(\sigma - \sigma')$ shares of the option, the combined portfolio will follow the diffusion:

$$Z\left(\frac{dC}{C}\right) - Z\left(\frac{dS}{S}\right). \tag{3.2.3}$$

If we substitute this in (3.2.1) and (3.2.2) there is cancellation:

$$\begin{aligned}
&\frac{[-\sigma'/(\sigma-\sigma')]dS}{S} + \frac{[\sigma/(\sigma-\sigma')]dC}{C} \\
&= \left[\frac{-\sigma'}{\sigma-\sigma'}\right]\mu dt + \left[\frac{-\sigma'}{\sigma-\sigma'}\right]\sigma dz + \left[\frac{\sigma}{\sigma-\sigma'}\right]\mu' dt + \left[\frac{\sigma}{\sigma-\sigma'}\right]\sigma' dz \\
&= \left(\left[\frac{-\sigma'}{\sigma-\sigma'}\right]\mu + \left[\frac{\sigma}{\sigma-\sigma'}\right]\mu'\right)dt.
\end{aligned}$$

The portfolio has no risk, only an expected return, and the expected return of a portfolio with no risk must be the risk-free rate:

$$\left[\frac{-\sigma'}{\sigma-\sigma'}\right]\mu + \left[\frac{\sigma}{\sigma-\sigma'}\right]\mu' = r_f \tag{3.2.4}$$

or

$$\mu' - r_f = \left(\frac{\sigma'}{\sigma}\right)(\mu - r_f). \tag{3.2.5}$$

This formula explains how the risk premium of the option is related to the underlying asset.

Ito's lemma, a mathematical identity that relates the drift of the option to the movements of the stock (this is derived in Chapter 8), lets us substitute for $\mu'$ and $\sigma'$ and condense the price of the option into a partial-differential equation where we denote the partials by subscripts for brevity:

$$0.5\sigma^2 S^2 C_{SS} + rSC_S - rC + C_t = 0, \tag{3.2.6}$$

where the change in the option value is identified with the change in the stock. $C_S$ is the first partial derivative of the call option price with respect to the stock price, $C_{SS}$ is the second partial derivative, and $C_t$ is partial derivative of the Call option price with respect to time.

Equation (3.2.6) identifies the price of an option since given the type of option (put or call), or even more complicated contracts such as swaps, only

one pricing formula will satisfy the equation. We give the solution below. For those with knowledge or interest in differential equations, we have traced through the solution in Chapter 8.

$$C = S\Phi(d) - X\exp\{-r(T-t)\}\Phi(u), \tag{3.2.7}$$

where $d = \ln(S/X) + (r + 0.5\sigma^2)(T - t)\left\{\sigma\sqrt{(T-t)}\right\}^{-1}$ and where $u = \ln(S/X) - (r + 0.5\sigma^2)(T - t)\left\{\sigma\sqrt{(T-t)}\right\}^{-1}$.

This is known as the Black-Scholes formula. While the Black-Scholes formula is mathematically more complex than the binomial, and perhaps more realistic, it offers most of the same insights as the simpler formulas.

Now we list some factors and indicate how they affect the premium for a call option with arrows:

1. ↑Stock price, ↑Premium. The higher the current stock price, the more likely a call option will be in the money.
2. ↑Strike price, ↓Premium. The higher the strike price of a call option, the more that must be paid if it is exercised, and the lower the likelihood the option will be exercised.
3. ↑Volatility (Standard deviation), ↑Premium. As with the binomial method, upswings are profits kept and downturns are losses ignored by simply not exercising the call option.
4. ↑Time to maturity, ↑Premium. The longer the time to maturity, the more that can happen. Remember that with a nonstationary process the longer the time period, the higher is the volatility.
5. ↑Risk-free interest rate, ↑Premium. This factor is a little more obscure. One way to think of this is that options act as a loan. You pay a small premium, and can receive the stock gains without purchasing the shares until later. Therefore, if the interest rate increases, this loan is more expensive.

One important thing that is missing from this list of factors is the expected return of the stock. Intuitively, it would seem that higher returning stock would be more likely to be in the money. But in a competitive world this return is simply a risk premium, and since the option is based on this stock, it must be discounted at a higher rate too. So any changes in the expected return are a wash, not affecting the premium for the option. This property of option pricing will come in handy in the next section when we attempt to value more complicated options for which the Black-Scholes formula does not apply.

## 3.3 OPTION PRICING UNDER JUMP DIFFUSION

The Black-Scholes formula for option pricing has been invaluable in pinning down the valuation of derivatives. There are several nice features about the

Black-Scholes formula that makes it a beautiful mathematical construct. For starters, there is no need to calculate risk premia since we are relying on a perfect hedge. Since the calculation of a risk premium implies that we have to know something about the investor preferences, it is a relief to eliminate it from the equation. The other feature that makes it a rare academic feat is that it is a closed-form solution. Once you know the parameters $\sigma$, $T$, $S$, and $x$, you just plug them in the formula (3.1.8) and you have a price for the call option. A similar formula is available for the put options. There is no need to solve additional equations, nor to integrate anything; the formula gives one number as the price.

The beauty of the mathematical solution of the Black-Scholes formula, however, is conditioned by some fine print that applies to any theoretical result: It only applies if the assumptions of the model are correct. If any of the assumptions do not represent a close approximation to reality, the formula cannot be applied to such a world, and the model needs to be either reformulated with a more realistic assumption or abandoned. So let's review a few of the assumptions of the Black-Scholes model:

1. Stock prices follow a random walk with a drift, and we have normally distributed random shocks. This makes stock returns lognormally distributed. The assumption of normality brings with it the assumption of symmetry of returns, and the assumption that extreme movements are unlikely.
2. The underlying stock is a traded asset and a hedge can be set up to be balanced at any moment with trades made instantaneously.
3. The volatility of the underlying asset can be measured accurately to predict the volatility over the life of the option.

We saw in Chapter 1 that there are diffusion processes other than the strict random walk that are used for stock prices. For example, one diffusion process that allows for downside risk to differ from upside risk is the jump diffusion process of equation (1.4.4) repeated here:

$$\frac{dS}{S} = (\mu - \lambda k)dt + \sigma dz + dq, \tag{3.1.1}$$

where $\lambda$ is the average number of jumps per unit of time, $k$ is the average percent a stock price changes at a jump, and $dq$ is a Poisson process, and where the total average return is $\mu$ after allowing for $+\lambda k$ from the jump process.

If the jump has a negative average return, then this can capture the likelihood of a market crash in a stock price. Merton (1976) worked with pricing an option on a stock that follows a jump diffusion process, but even this small addition to normality causes problems in finding a closed-form solution.

First, since the process combines two distributions, normal for $dz$ and Poisson for $dq$, it cannot be fully hedged against risk. If one hedges using the

same weights as with the Black-Scholes, the random walk portion is diversified, but the jump portion is not constant:

$$\frac{[-\sigma'/(\sigma-\sigma')]dS}{S}+\frac{[\sigma/(\sigma-\sigma')]dC}{C}=\left(\left[\frac{-\sigma'}{\sigma-\sigma'}\right]\mu+\left[\frac{\sigma}{\sigma-\sigma'}\right]\mu'\right)dt+dq. \quad (3.3.2)$$

Without a perfect hedge, the risk-free rate cannot be used and a closed-form answer cannot be found. Merton did identify some special cases where a solution can be found. The usual assumption in order to find a pricing solution is the assumption that the jump is entirely nonsystematic. Once that assumption is made, there is no need to price the jump portion of the distribution since it does not contain market risk.

In reality, however, downside jumps are often marketwide phenomena. Just observe the Asian crisis of the late 1990s, the technology fiasco of the early 2000s, or the accounting scandals of 2002, and notice the patterns of downside drops spreading throughout the domestic and worldwide economy. This makes the assumption of no market risk (zero beta) for the jump impractical for modeling downside risk in the stock market; the zero beta assumption is more useful for special cases when the jumps are uncorrelated random occurrences.

This is the difficulty of the Black-Scholes formula. If any of the assumptions are broken, the solution is no longer as elegant, and either a numerical calculation or a simulation needs to be run to approximate the option price, if it can be calculated at all. Numerical solutions and simulations of options prices will be covered in Chapter 9.

## 3.4 IMPLIED VOLATILITY AND THE GREEKS

The Black Scholes option pricing formula in its basic form gives an important tool for determining how the market quantifies the risk of a security. The formula itself is based on the current underlying asset price, time to maturity, the risk-free rate, the strike price, and the volatility of the underlying asset. Of these items the most difficult one to pin down is the volatility. The others (assuming the US T-bill is a good fit for the risk-free rate) can all be looked up in the market reports or are simply a part of the specification of the option contract. Volatility, however, is not quoted for stocks. It is behind the diffusion process but not directly observable. The closest thing we can get is a historical volatility, such as an estimated standard deviation in the past time series data.

To determine risk, one would prefer a forward-looking measure of volatility. Since option prices factor in volatility, investors can use the market quotes of option prices to back out volatility. It is the value implicitly used by options traders in their calculations based on the Black-Scholes formula. This is known as implied volatility.

The Chicago Board of Options Exchange (CBOE) lists a Volatility Index that is based on a weighted average of the implied volatility put and call options on the S&P 100 index (VIX) and the Nasdaq-100 index (VXN). The implied volatility gives an overall estimate of the volatility options traders are factoring into their prices.

### 3.4.1 The Role of Δ of an Option for Downside Risk

Besides implied volatility, other forward-looking measures of risk can be gleaned from options prices. The delta of an option, Δ, is defined as the change in the price of the option for a one dollar change in the price of the underlying asset. Call options have positive deltas since they increase in value as the underlying asset goes up in price, and conversely put options have negative deltas. For a call option $\Delta = 0.25$ means a quarter-point rise in premium for every dollar that the stock goes up. For a put option $\Delta = -0.25$ means a quarter-point rise in premium for every dollar that the stock goes down. Also options prices cannot increase by more than the price of the underlying asset because even if the option were guaranteed to be in the money, it would give only a dollar more gain for the option. Thus, for call options, $0 \leq \Delta \leq 1$, and for put options, $-1 \leq \Delta \leq 0$.

Delta becomes useful for measuring downside risk because it can be thought of as the probability that the option will end up in the money. Options that are deep in the money and almost assured of paying off will have a delta close to 1, and options that are not likely to pay off will have a delta close to zero. Therefore the delta measures the likelihood that for calls the price will exceed the strike price, and for puts it measures the likelihood that the downside will go lower that the strike price.

### 3.4.2 The Gamma (Γ) of an Option

Of course, as the price of the underlying asset changes, and the option moves closer or farther from the strike price, the delta changes also. An option's gamma (Γ) measures how quickly the delta changes with a change in the price of the underlying asset. Options that are at the money will have the largest gamma, and options deep in or out of the money will have low gammas.

### 3.4.3 The Omega, Theta, Vega, and Rho of an Option

A delta of an option expressed in terms of percentages (or an elasticity of the options price with respect to the underlying asset) is known as an omega (Ω). Since these sensitivity measures for options prices are all denoted by greek letters, they are known collectively as "the Greeks." The Greeks above use the price of the underlying to try to tease out downside risk, but additional common sensitivity numbers denoted by greek letters are theta (Θ), the absolute change in the option value for a reduction in time to expiration by

one unit (usually the unit is one week) vega (Ω), the change in option price with respect to a 10% change in volatility, and rho (ρ) the reaction of the option price to changes in the risk-free rate. For example, if a $\rho = 0.5$ the option's theoretical value will increase by 0.05 if the interest rate is decreased by 1%. The information about the Greeks, usually accompanied by the definitions of terms, is increasingly being made available by brokerage houses for the convenience of options market investors.

In this chapter we have examined derivative contracts as first a method of hedging risk, and then as a source of collecting information about the perceived volatility of the market. In later chapters these measures will be used as a means of improving our historical estimates of downside risk with forward-looking measures from the options market.

An example at http://www.cboe.com/LearnCenter/ProtectivePutsAsHedge.asp explains in practical terms how to use put options to insure against downside risk. Long-term Equity Anticipation Products (LEAPS) are put options with expirations longer than nine months. For example, on July 13, 2004 a put option LEAP (ticker WJJMD) provides complete protection against the price of JetBlue (ticker JBLU) falling below $20 by January 2006 (the farthest date available). For 100 shares this protection costs about $390, which is equivalent to paying an insurance premium.

A simple alternative to derivatives is to buy bonds to reduce the downside risk. The downside risks associated with corporate or local government revenue bonds include: (i) The entity which issues the bond (or the one that insures the bond) can go bankrupt. (ii) In times of rising interest rates or any other reason affecting the prospects for the issuer of bonds, the bonds can lose value. Bond price variations during the holding period become irrelevant if the bond is held till maturity. (iii) Inflation can cause the real purchasing power to decrease.

Some bonds linked to stocks are of particular interest for reducing the downside risk. The American Stock Exchange reports information about linked bonds under the structured-products tab at www.amex.com. These structured products are offered by various brokerage firms and have strange names given by the brokerage houses. Morgan Stanley calls them "SPARQS," or Stock Participation Accreting Redemption Securities. Merrill Lynch calls them "STRIDES," or Stock Return Income Debt Securities.

Instead of the LEAPS mentioned above for JetBlue Airlines, one can buy Merrill's 9% STRIDES (ticker CSJB). They are not endorsed by JetBlue Airlines. During the second quarter or 2004, historical range for JetBlue stock price was $12 to $24.222. The CSJB STRIDE was issued on May 25, 2004 at the par value of $25 and maturity date May 22, 2006 with a two year maturity. At maturity, STRIDE holders *must* exchange them for a predetermined number of shares of JetBlue common. The CSJB are "callable" at any date after May 23, 2005 at Merrill's discretion for cash value specified in prospectus for CSJB. If called, this works like a short term (one to two year) bond. The prospectus describes many hypothetical price scenarios: (1) If on the

maturity date May 22, 2006, the JetBlue stock declines to $5.18, much below the lower limit $12 of the historical range, then the annualized net loss to STRIDE holder would be annualized 43.85%, which is substantial, except that the owners of JetBlue would suffer an even larger comparable annualized loss of 55.23%. This shows that the downside loss is somewhat reduced but not eliminated. (2) If the maturity price is $36.27, owners of the JetBlue common stock would earn 18.29% annualized return, whereas CSJB owners will earn only 17%. Thus the upward potential can never exceed 17% per year to the owner of the STRIDE. The prospectus lists such scenarios for hypothetical prices and all risk factors.

STRIDES are useful for small investors for two reasons: (i) They allow a bullish investor to earn a high yield without incurring further costs (insurance premium, commissions, etc.) needed to use the options market, and (ii) STRIDES can have longer maturity dates than typical options and LEAPS.

## APPENDIX: DRIFT AND DIFFUSION

In this appendix we attempt to introduce the reader to the mathematical details related to the diffusion model discussed in Section 1.4 of Chapter 1 in continuous time:

$$dS = \mu(t, S)dt + \sigma(t, S)dz,$$

where $\mu(t, S)$ is the predictable drift part and $\sigma(t, S)$ is the random diffusion part. Both are functions of the two arguments, $S$ price of stock and $t$ for time.

We may rewrite the equation above as

$$\frac{dS}{S} = \mu dt + \sigma dz. \tag{3.A.1}$$

If the $S$ represents a risk-free asset paying a return of $\mu$ per unit time $dt$, $S_0$ as its value at $t = t_0$, the random component is zero causing the second part of (3.A.1) to disappear. Then we have a simpler differential equation: $dS/S = \mu dt$. The solution of the simpler ordinary differential equation is $S = S_0 \exp[\mu(t - t_0)]$. To verify this solution, differentiate both sides of the solution with respect to $t$ to yield $dS/dt = S_0 \exp[\mu(t - t_0)]$ times $\mu$. Now we replace $\{S_0 \exp[\mu(t - t_0)]\}$ by $S$ on the right hand side, to yield $dS/dt = S\mu$, which can be rearranged as the original "simpler" differential equation. This is a direct verification of the solution.

If the initial price is $S$, the subsequent price is $S + dS$ after one time step of $dt$. Now let $dt = 1$ for one year and $\mu = 0.05$ or 5% per year. Now the ratio of $S + dS$ to $S$ equals $1 + (dS/S) = 1 + \mu$, or the initial price 100 becomes 1.05 times the initial price 100, or 105, suggesting that the mathematics makes intuitive sense.

Next let us bring in the random part or the diffusion part $\sigma dz = \sigma u\sqrt{dt}$, where $u$ is the unit normal random variable N(0, 1) with mean zero, and variance unity. If $E$ denotes the mathematical expectation, $Eu = 0$ and $Eu^2 = 1$. Although $u \in (-\infty, \infty)$ is the original range of $u$, the range [–3.39, 3.39] as tabulated in the typical standard normal tables is all that is needed for all practical purposes. The scale factor $\sqrt{dt}$ is needed because without it, there will be technical difficulties later when we let $dt \to 0$. Now the initial price of a risky asset (stock) $S$ becomes $S + dS$, with $dS$ having both the drift and diffusion parts. The range of $S + dS$ will not be [–3.39, 3.39] but require a modification depending on the parameters $\sigma$ and $\mu$ and $dt$. As a first approximation, typical price of risky assets can be assumed to follow the transformed normal (lognormal) distribution. The lognormal distribution ranges over 0 to $\infty$ rather than $-\infty$ to $\infty$, which is obviously the appropriate range for prices, since prices of assets cannot be negative, by assumption.

The increment $dS = \mu S dt + \sigma S dz$, where $dz = u\sqrt{dt}$, where $u$ is known to be unit normal $u \sim N(0, 1)$. Hence $E(dS) = \mu S dt$, which depends exclusively on the drift parameter $\mu$. The variance of the increment $\text{var}(dS) = \sigma^2 S^2 E(dz)^2 = \sigma^2 S^2 dt$ by using $E(u^2) = 1$. Thus the variance depends exclusively on the diffusion parameter $\sigma$.

What happens when $dt \to 0$? Since $dz = u\sqrt{dt}$, we note that $dz \to O(\sqrt{dt})$, where the notation "$A$ is of order $B$" (written $A \to O(B)$) means that the ratio $(A/B)$ is even smaller than some constant. In our case, as $dt \to 0$ the value of $dz$ is of smaller order of magnitude than $\sqrt{dt}$. We will use this to cancel the term $dt\sqrt{dt}$ in Chapter 8. It seems all right to assume that the drift part goes to zero, but letting the variance become zero amounts to saying that we have a constant, or that the randomness goes away. Since complete disappearance of randomness is intuitively unacceptable, mathematicians have developed a toolkit involving stochastic integrals. For our purposes it is convenient to simply use the final result of that theory, that when $dt \to 0$, we have $(dz)^2 \to dt$ with probability 1, which does not have an intuitive proof.

This appendix has given a mostly intuitive short description of the role of drift and diffusion aspects in determining the dynamics of option prices. Chapter 8 gives further details on Ito's lemma, a fundamental result in stochastic calculus on par with Taylor series. We also consider what happens when two random walks move together or are correlated with each other based on a mixture function attributed to Merton (1976). We use it derive the Black-Scholes partial differential equation.

CHAPTER 4

# Monkey Wrench in the Works: When the Theory Fails

## 4.1 BUBBLES, REVERSION, AND PATTERNS

In the previous chapters we examined the way markets are supposed to work. We explained common modeling techniques and showed where the inclusion of downside risk may be relevant for the rational investor, and we will explore those additions in later chapters. But first, we should look at some elements of downside risk that have very little to do with a rational investor.

One of our main tools in creating a financial theory has been to assume that investors are paying a fair market price for a stock based on all information available. If all investors are doing this, then the market price for a stock should be fair also. Efficient markets hypothesis (EMH) says that the market price should be fair in the sense that it should not retain profit opportunities in being predictable from observable information. If one looks at the enormous amount of activity spent on stock market analysis, where is the room for profit opportunities? With all of the investors, brokerages, investment banks, research departments, stock analysts, all constantly crunching numbers and looking at the same reports, how can one expect to have any new information that is not already reflected in the stock price?

The EMH is supported by a very plausible and rational argument. It does not say that information has no effect on stock prices; it says only unexpected information will affect the price. Old news is not valuable. Some stocks will get higher returns than others, but that is rational if the stock with higher returns is also riskier. It just boils down to a risk premium, not that investors buying the low return stock are deficient of reasoning. This section discusses the apparent factual empirical problems with EMH. The data reveal seemingly

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

irrational patterns in the stock market that make prices appear to be predictable.

This is logically dangerous ground, and we should tread with care. If we lose our rational investor, then all bets are off. If the stock market is not efficient, then there is no rhyme nor reason behind the valuation of securities, and we lose all of our hard work of building a theory to price downside risk. We hear stories of market psychology, herd behavior, and illogical price movements, but if this becomes the rule rather than the exception, then stock investing becomes a form of gambling or astrology. So when taking a look at these anomalies, we should keep an open mind but be critical of brokers and sales people who make it seem that profits on Wall Street can be made too easily.

### 4.1.1 Calendar Effects

One of the early patterns discovered on Wall Street have to do with the timing of investments to achieve predictable extra returns, namely market timing based on the day of the week, the month of the year, the time of the day, and so on. Stories behind these effects are filled with market psychology and coincidence, but usually these explanations come after the discovery. The most famous example of a calendar effect is the Monday effect. From observations of historical returns it was noted that Mondays tended to yield lower stock returns than other days of the week. Besides, who cannot relate to this? The mind conjures up images of scowling, liveless stock traders dragging themselves to work Monday morning back from the suburbs. They just aren't up to buying anything, and prices wane until the brokers are back into their routine by Tuesday. The same can be said for rainy days, Mondays before a Tuesday holiday, Tuesdays after a Monday holiday, and so on.

It is harder to accept that stock traders don't like money on Mondays. If one can consistently predict that prices will be lower on Monday, there is a profit opportunity. That is enough to make any stock trader love Mondays and be bounding to work after the weekend. Because the psychological explanation is not rational, researchers have come up with some more rational-sounding explanations.

For example, low returns on Mondays have been explained by the corporate practice of releasing bad news on Friday evening after markets have closed so that there will not be an instant reaction in the market. If the Monday effect is in response to new information, it does not violate the efficient market hypothesis. Although it is still a mystery why traders do not anticipate bad news.

The January effect (high returns in January) is attributed to tax laws, which state that stock sales by individuals in December qualify for favorable tax treatment of capital gains and losses. Also on the corporate side postponing sales to January will put off the tax bill by a whole year. This explanation is somewhat justified by historical data, and as various firms have moved their

fiscal years for tax purposes to other months of the year, the January effect has lessened over time.

Calendar effects have also been a source of interest in international stock markets. In India, the Bombay Stock Exchange and the National Stock Exchange had delivery days for their shares of Monday and Wednesday, respectively, until 2002. This caused a calendar effect where prices were cyclical throughout the trading week: high when delivery was close, and lower when delivery was several days away. The pattern was deceptive. It would seem that one could make excess returns simply by buying when it is known that the price is low and selling at the end of the cycle when the price is higher. That would be true if this were an irrational cycle, but in India this cycle had a rational cause: when the delivery date was farther away, a higher risk level was associated with the purchase. Therefore the lower prices at the beginning of the weekly cycle were just compensation for waiting longer for delivery. In an effort to develop a more efficient stock market, India moved to a rolling delivery date (three days from purchase), and now the cyclical pattern is no longer evident.

### 4.1.2 Mean Reversion

The random walk hypothesis of stock prices in a simple form states that stock markets are efficient and therefore stock prices are not predictable from past prices. This was challenged by Fama and French (1988) who stated that "25–45% of the variation of 3–5 year stock returns is predictable from past returns." Poterba and Summers (1988) also found that that price movements for stock portfolios tend to offset over long horizons. If a stock price veers off too much from its long-term mean (or trend), one can safely predict that in the next few quarters it will revert to the trend value. Reverting to the trend is also described as reverting to the mean (e.g., average growth rate) or simply mean reversion.

There are two main types of mean reversion characterized by the persistence of their effect on returns: autoregression and moving average

***Autoregression.*** If the past value of returns affects the current value, then the returns show autoregression.

$$\text{AR(1)}: \quad r_t = \alpha + \rho r_{t-1} + \varepsilon_1, \tag{4.1.1}$$

where $|\rho| < 1$. As the equation shows, the first-order autoregression is denoted by AR(1) where current period returns are being influenced by the previous period's returns. The intercept $\alpha$ depends on the average return and $\rho$. As time passes, the effect of past time period just gets transmitted on to the next period. If $\alpha = 0$, and $\rho = 2$, it means that the return at time $t$ is double the return of the previous time $(t - 1)$. After just five time periods the return will be $2^5 = 32$ times the return at time $(t - 1)$. Such a series will soon explode or grow too

large to be realistic. For the data on returns, there is some evidence that the slope regression coefficient ρ in (4.1.1) is positive and fractional. When ρ is fractional, its higher powers become successively smaller. These effects are declining, since they are transmitted at a fraction ρ of their original value. Therefore after four periods there will only be $\rho^4$ of the effect remaining in the AR(1) case, bringing returns back to their long-run average.

In the AR(1) model, the regression coefficient also equals the autocorrelation coefficient of order 1 when we assume that we have n observations $r_t$ with average return of $\bar{r}$. However, in general, the autocorrelation coefficient of order $p$ denoted by $\rho_p$ is not a similar regression coefficient. It is defined by

$$\rho_p = \frac{\sum_{t=p+1}^{n}(r_t - \bar{r})(r_{t-p} - \bar{r})}{\sum_{p=1}^{n}(r_t - \bar{r})^2}, \tag{4.1.2}$$

where autocovariance of order $k$ is the numerator of autocorrelation. More formally, some authors define mean reversion as negative serial correlation at all values of $p$, that is, for all leads and lags. Negative correlation means that if the value goes above the mean, it goes back toward the mean in the next few time periods.

Equation (4.1.1) is AR(1) or first-order autoregression, In general, $p$th order autoregression follows the following regression equation:

$$\text{AR}(p): \quad r_t = \alpha + \beta_1 r_{t-1} + \beta_2 r_{t-2} + \ldots + \beta_p r_{t-p} + \varepsilon_t, \tag{4.1.3}$$

where $\Sigma\rho_p < 1$. Since the past return is in the next period's value, autoregressive effects never completely disappear. Realism requires that the series should not explode, or grow indefinitely.

The conditions for realism are formally called *stationarity conditions* in time series literature. Briefly, they require that the mean, variance, and autocovariance of the series all be finite and not changing over time. As time increases, any one of these quantities should *not* grow indefinitely (blow up). For the AR(1) stationarity condition is given by $|\rho| < 1$. When $p$ is larger than 1, these conditions for realism of the model become complicated. For example, when $p = 2$, we have AR(2) model with the following two (stationarity) conditions for realism:

$$\left|0.5\left[\beta_1 + \sqrt{\left\{(\beta_1)^2 + 4\beta_2\right\}}\right]\right| < 1 \quad \text{and} \quad \left|0.5\left[\beta_1 - \sqrt{\left\{(\beta_1)^2 + 4\beta_2\right\}}\right]\right| < 1. \tag{4.1.4}$$

In evaluating conditions (4.1.4), note that if $(\beta_1)^2 < 4\beta_2$ and $\beta_2 < 0$, the square root of $(\beta_1)^2 + 4\beta_2$ will involve imaginary numbers. However, the imaginary numbers simply mean that the time series of returns is subject to ups and downs similar to a sine or cosine curve. Since cyclical behavior of returns is quite realistic, we cannot rule imaginary numbers out. However, we do want

the sine and cosine curves not to keep getting wider and wider. They should be damped, if they are to be realistic. The realism conditions (4.1.4) do guarantee damped behavior. Ultimately the damping is strong enough and the series is mean reverting. In general, conditions are available for the series to be realistic return series for any AR($p$) model.

***Moving Average.*** Many technical trading rules are based on comparing short- and long-term moving averages. A moving average process is a transitory holdover from a previous period. A first-order moving average, MA(1), process is written

$$\text{MA(1):} \quad r_t = \alpha - \delta\varepsilon_{t-1} + \varepsilon_t. \tag{4.1.5}$$

Only the random, unexpected portion of returns from the last period ($\varepsilon_{t-1}$) is carried over in the moving average process, not the entire value. In the next period, when we replace $t$ by $t + 1$, $\varepsilon_{t-1}$ will have no affect on $\varepsilon_{t+1}$.

Expanding to the $q$th-order moving average process, we get

$$\text{MA}(q): \quad r_t = \alpha - \delta_1\varepsilon_{t-1} - \delta_2\varepsilon_{t-2} + \ldots - \delta_q\varepsilon_{t-q} + \varepsilon_t. \tag{4.1.6}$$

Combining the processes (4.1.3) and (4.1.6), we get an ARMA($p$, $q$) model:

$$\begin{aligned} r_t = \alpha + \beta_1 r_{t-1} + \beta_2 r_{t-2} + \ldots + \beta_p r_{t-p} - \delta_1\varepsilon_{t-1} \\ - \delta_2\varepsilon_{t-2} + \ldots - \delta_q\varepsilon_{t-q} + \varepsilon_t, \end{aligned} \tag{4.1.7}$$

which is a mixture of autoregressive (AR) of order $p$ and moving average (MA) of order $q$. A great deal is known about this class of models and their extensions as seen from a survey in Hamilton (1994). There are standard software tools available for estimation, testing and criticism, and further modification and testing of ARMA models.

Since mean reversion allows prediction of current returns based on past returns and errors, it does not fit with any for of the efficient market hypothesis. Richardson and Stock (1989), Kim et al. (1991), and Richardson (1993) found that the mean reversion is not statistically significant, but there are armies of day traders and technical analysts who firmly believe otherwise.

### 4.1.3 Bubbles

Stock market bubbles are the most troubling empirical phenomenon to justify with financial theory since they are, by definition, a sustained overvaluation of the stock market, until it bursts. Some famous examples of stock markets bubbles include the tulip bulb craze of the 1600s in Holland and the South Sea Company bubble of the 1700. More recently the 1990s are being termed the technology or "dot com" bubble. When a market bubble is present, the market

has unrealistic expectations about the growth prospects until the bubble bursts. Great riches can be earned by anyone who knows when a bubble condition exists and has the foresight to sell assets before the bubble bursts.

The trouble with finding bubbles, however, is that they don't seem like bubbles before they burst. The dot com experience now seems ludicrous, but before the bubble burst the technology stocks could have worked. There were several key variables, such as Internet security, consumer's willingness to buy online, and the speed of the Internet, that—had they turned out more beneficial—Internet companies could be bringing in barrels of cash right now. Ex post, these variables did not turn out favorable for dot coms, but this was more from bad luck than irrationality. Furthermore the macroeconomic theory has a vast literature (see Blanchard and Watson, 1982) on bubbles as a rational outcome.

## 4.2 MODELING VOLATILITY OR VARIANCE EXPLICITLY

The volatility or variance of market prices has a special importance in finance. Even a cursory look at financial time series reveals that some periods are more risky than others. Engle (1982) proposed the autoregressive conditional heteroscedasticity (ARCH) model, which lets the volatility be different in different ranges of the data. There are great many extensions of the ARCH model in econometric literature, with a virtual alphabet soup of names. Among them, generalized ARCH, or GARCH, and integrated ARCH, or IGARCH, stand out and will be considered later. We first focus on forecasting volatility.

Consider the standard multiple linear regression models for time series data. Let $y_t$ denote the value of the dependent variable at time $t$, and let $x_t$ denote a vector of $k$ values of $k$ regressors (including a column of ones for the intercept) at time $t$. Let the unknown population regression parameters be denoted by a vector $\beta$, and let the true unknown error in the regression equation at time $t$ be denoted by $\varepsilon_t$. The regression model then is

$$y_t = \beta' x_t + \varepsilon_t, \tag{4.2.1}$$

where it is customary to assume that the errors satisfy

$$E[\varepsilon_t] = 0, \quad \text{var}[\varepsilon_t] = \sigma^2, \tag{4.2.2}$$

where the mean of errors is constant (= 0) for all time periods and so is the variance $\sigma^2$. The latter is called the homoscedasticity assumption. In cross-sectional data, heteroscedasticity arises commonly when the error means are constant for all individuals in the cross section (e.g., 50 states in the United States), but the error variances are related to some (e.g., size) variable. In our time series data, if the variance changes over time in some prescribed way, we have time-dependent heteroscedasticity.

If heteroscedasticity is ignored, it can lead to a false sense of security regarding estimated least squares coefficients as the test statistics are biased toward finding significance. Instead, the ARCH-GARCH models discussed here try to treat this as a modeling challenge and hope to correct the deficiency of least squares and provide forecasts of variances (volatility). One of the simplest ways of modeling time-dependence is to propose a simple first-order autoregression AR(1) among error variances, usually estimated by the squared residuals of the regression.

The autoregressive conditional heteroscedasticity (ARCH) models of Engle (1982), and many others inspired by it, continue to assume zero mean of errors but relax the assumption of constant variance. Since the mean is zero, the usual (unconditional) variance is given by mathematical expectation of squared errors, $E[\varepsilon_t^2] = \sigma^2 h_t$, can have a part that is fixed over time, $\sigma^2$, and a part $h_t$ that changes over time. In the special case of stationarity of these variances, the unconditional variance does not blow up as time passes but remains finite.

The part of the variance that changes over time may be viewed as the variance conditional on the known past errors. For example, the set of the most recent $p$ past errors known at time $t$ is $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-p}\}$. Note that at time $t$, the past errors are known, and the unknown current error $\varepsilon_t$ may be related to past errors. Hence the "conditional" variance of the current error given past errors, $E[\varepsilon_t^2|$ past errors] can possess a well-defined time series model. For example, ARCH($p$) model is defined by the following auxiliary regression similar to AR($p$) of (4.1.3), except that now it is defined in squared errors, not returns. The expectation of the current time period variance given the values of previous error terms can be written as the auxiliary regression

$$\text{ARCH}(p): \quad E[\varepsilon_t^2|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-p}] = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_p \varepsilon_{t-p}^2. \quad (4.2.3)$$

Since the mean of errors is zero, the conditional variance is given by using the law of iterated expectations as

$$E[\varepsilon_t^2|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, +\varepsilon_{t-p}] = \alpha_0 + \alpha_1 E[\varepsilon_{t-1}^2] + \ldots + \alpha_p E[\varepsilon_{t-p}^2]. \quad (4.2.4)$$

Assuming stationarity mentioned above, the unconditional variance for all time subscripts equals a common constant $\sigma^2$; that is, $E[\varepsilon_{t-1}^2] = \sigma^2$, $E[\varepsilon_{t-p}^2] = \sigma^2$. Then we can rewrite the equation above, after collecting all expectation terms on the left side, as

$$\sigma^2(1 - \alpha_1 - \ldots - \alpha_p) = \alpha_0. \quad (4.2.5)$$

As with any autoregressive model of (4.1.3), AR($p$), determining the largest number $p$ of lag terms is subject to further testing, analysis, and debate. Researchers in finance have found that many lags are needed to track high frequency (hourly, daily) data. See Bollerslev and Ghysel (1974) study of daily

returns for mark–sterling exchange rate data that introduced generalized ARCH or GARCH models. It is found that instead of including a large number of terms, it is parametrically more parsimonious to introduce a few (= $q$) terms involving the lagged errors of the auxiliary regression (4.2.3).

The GARCH($p$, $q$) models are simply ARMA($p$, $q$) models applied to squared regression residuals. Denoting the conditional variance by a subscript on $\sigma^2$ we have the GARCH(1, 1) model defined by

$$E\left[\varepsilon_t^2|\varepsilon_{t-1}\right] = \sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \gamma_1\sigma_{t-1}^2, \tag{4.2.6}$$

where the $\varepsilon_{t-1}^2$ term on the right side is the lagged dependent variable and is the AR(1) part and the $\sigma_{t-1}^2$ term is the MA(1) part. It is clear that this model forecasts volatility one period ahead, that is, $\varepsilon_t^2$ from the past forecast and the past residual, as follows:

Start with the initial value of $\sigma_0^2$ for $t = 0$. Then $\sigma_1^2 = \alpha_0 + \alpha_1\varepsilon_0^2 + \gamma_1\sigma_0^2$ yields the one-step ahead forecast.

Now let the error in this forecast be defined as $(\varepsilon_1)^2$. The second period forecast is given by

$$\sigma_2^2 = \alpha_0 + \alpha_1\varepsilon_1^2 + \gamma_1\sigma_1^2.$$

Now use the error in this forecast to define $\varepsilon_2$.

Substituting this, obtain the third period forecast of volatility as

$$\sigma_3^2 = \alpha_0 + \alpha_1\varepsilon_2^2 + \gamma_1\sigma_2^2.$$

It is a simple matter to write an iteration to obtain a GARCH updating forecast for several periods in the future. The long-run forecast is called "unconditional volatility," and it is a constant, provided the regression coefficients satisfy the inequality: $\alpha_1 + \gamma_1 < 1$. Then conditionally heteroscedastic GARCH models are said to be a mean reverting with a constant unconditional variance.

Given data on financial returns it is customary to use the maximum likelihood (ML) method to estimate the GARCH(1, 1) model. For a discussion of the ML method, see Sections 4.4.4 and 6.5.1. Several computer programs are available for this purpose. McCullough and Renfro (1999) offer a benchmark computer program for testing if a computer program is correct. McCullough and Vinod (1999, 2003a, b) discuss the importance of numerical accuracy in these calculations, since the results are sensitive to starting values and various settings in the program. The authors note that different software packages give different results and emphasize a need for benchmarks. We will explore these numerical issues in greater depth in Chapter 9.

The "Ljung-Box test" with 15 time lags ($k' = 15$) is popularly used for testing the adequacy of the model. Ljung and Box (1978) show that under the null

hypothesis that the estimated model is adequate, the following $Q_{\text{L–B}}$ statistic gives a good approximation to the true critical region:

$$Q_{\text{L–B}} = T(T+2)\frac{\sum_{j=1}^{k'}(\rho_j)^2}{T-j}, \tag{4.2.7}$$

where $\rho_j$ are sample estimates of autocorrelations as in (4.1.2), among residuals of the auxiliary regression (4.2.6), of the GARCH model. For a GARCH($p$, $q$) model, the user would reject the null hypothesis if this statistic exceeds the standard chi-square distribution tabled value for ($k' - p - q$) degrees of freedom, for a given level of significance.

## 4.3 TESTING FOR NORMALITY

Since so many of the formulas in finance use the normal distribution, it is imperative that we determine if stock returns are, indeed, normally distributed. There are many reasons why the normal distribution is used in theory. It has several nice properties to it.

1. It is additive, meaning that if two normally distributed variables are summed, then the result is also normally distributed.
2. The normal distribution is symmetrical, meaning that the probability above and below the mean is identical. This makes the question of downside risk moot, since upside and downside risk is the same if returns follow a normal distribution.
3. The mathematical formula for the probability density function of the normal distribution, $N(\mu, \sigma^2)$,

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{4.3.1}$$

includes an exponential. Taking a natural logarithm of returns is often done to calculate a percentage change or log-likelihood function (see Chapter 9). Taking logs cancels the exponential, making the math for maximum likelihood easier.

### 4.3.1 The Logistic Distribution Compared with the Normal

The logistic distribution is also common since it is symmetrical, and there exists is a closed-form solution for the cumulative distribution function. The pdf of the logistic density is $f(x) = e^{-x}[1 + e^{-x}]^{-2}$. Figure 4.3.1 plots the logistic side by side with the standard normal N(0, 1) to show that it is also symmetric and

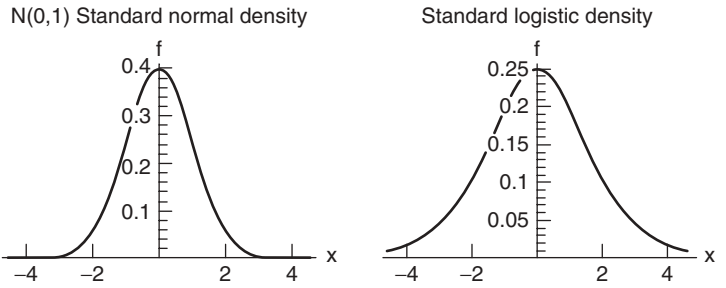



**Figure 4.3.1** Comparison of N(0, 1) standard normal density and standard logistic density

bell-shaped, but has a wider range. With most distributions, including the normal distribution, the probability of being below a specific value can only be calculated by taking the integral of the pdf. The cumulative distribution function (CDF) of the logistic distribution can be calculated without calculus by the probability statement

$$\Pr(x < z) = \frac{1}{(1+e^{-z})}. \tag{4.3.2}$$

Among further properties of the logistic, its variance is $\pi^2/3 \approx 3.2899$, kurtosis is 4.2, compared to the unit variance and kurtosis of 3 for the N(0, 1). It is often claimed that returns data exhibit excess kurtosis and fat tails compared to the normal. Both these properties are satisfied by the logistic. Hence the logistic may be preferable to N(0, 1) if one wants to fit a symmetric density to the returns data.

By changing the origin and scale of N(0, 1), one obtains a more general $N(\mu, \sigma^2)$ density where the mean $\mu$ and variance $\sigma^2$ are its two parameters. One fits the normal distribution to any given data by simply equating the sample mean and variance to the parameter values. A parametric form of logistic cdf with parameters for mean $\mu$ and variance $\beta^2\pi^2/3$ is simpler than the pdf. The cdf of two-parameter logistic is

$$F(x) = \left[1 + \exp\left(-\frac{x-\mu}{\beta}\right)\right]^{-1}. \tag{4.3.2}$$

This density too is estimated by simply equating the sample mean and variance to the appropriate parameter values.

Of course, with the availability of computers and statistical software, these calculation concerns are no longer the priority that they used to be. Even as recently as 1998, we were still teaching statistical programming on a mainframe computer. Storage space was at a premium, and programs would take considerable time to run. In that environment researchers had to do all they

could to make sure programs were as streamlined as possible. Current personal computers have thousands of times more storage and are several orders of magnitude faster than those mainframes. Therefore there is no need to sacrifice accuracy for simpler math.

### 4.3.2 Empirical cdf and Quantile–Quantile (Q–Q) Plots

Rather than disregard the normal distribution in all applications, we will give it the benefit of the doubt and let the numbers tell us if an approximation is good. A test of the goodness of fit is usually made for discrete pdf's like the binomial and Poisson by preparing a table of observed ($O_{\text{bs}}$) and fitted ($F_{\text{it}}$) values with say $j = 1, 2, \ldots, k$ rows. The Pearson goodness of fit statistic is simply

$$G_{\text{of}} = \sum_{j=1}^{k} \frac{[O_{\text{bs,j}} - F_{\text{it,j}}]^2}{F_{\text{it,j}}}, \tag{4.3.3}$$

where we have inserted an additional subscript $j$ for the $j$th row. We reject the null hypothesis of a good fit if the observable statistic $G_{\text{of}}$ exceeds the chi-square value from standard tables for $k - 1$ degrees of freedom for a given level of significance (type I error). A similar test can be performed for continuous distributions also by dividing the range of data into suitable intervals. Kendall and Stuart (1979) devote an entire chapter to the issue of tests of fit. A graphical feel for goodness of fit is discussed next.

Q–Q plotting is a graphical method of comparing the quantiles of the observed data with those of a theoretical parametric distribution. These are most used for plotting the residuals of a regression and checking if they are approximately normal. Consider our AAAYX mutual fund data from Chapter 2. First, we reorder the $T$ data points from the smallest to the largest. The ordered data are denoted as $x_{(1:T)}, x_{(2:T)}, \ldots, x_{(T:T)}$ and called order statistics $x_{(t:T)}$. Note that the probability mass between $(-\infty, x_{(1:T)}]$ is $(1/T)$, the probability between $[x_{(1:T)}, x_{(2:T)}]$ is also $(1/T)$ and similarly for each consecutive interval. Figure 4.3.2 shows an empirical CDF for a small dataset.

The cumulative probability $p_j$ is simply accumulation of probability masses. The cumulative probability starts at zero and monotonically increases in steps of $(1/T)$ until it is unity: $p_j \in [0, 1]$. The distribution of $p_j$ is called the empirical CDF (ecdf) and defined by $p_j = F_{\text{ecdf}}(x)$, based on the ordered data. It is also possible to interpolate among $p_j$ values and inverse ecdf is well defined for any cumulative probability $p$ as $F^{-1}_{\text{ecdf}}(p) = x$. Note that if $p = 0.5$, this relation gives the median, or the 0.5 quantile of the observed data.

To test normality, we want to compare the data with the unit normal density N(0, 1) and cumulative distribution function $\Phi(x)$. For the theoretical N(0, 1) density we first subdivide the entire area (= 1) into $T + 1$ intervals, each containing the probability mass $(1/T)$. Given the usual normal CDF tables or a

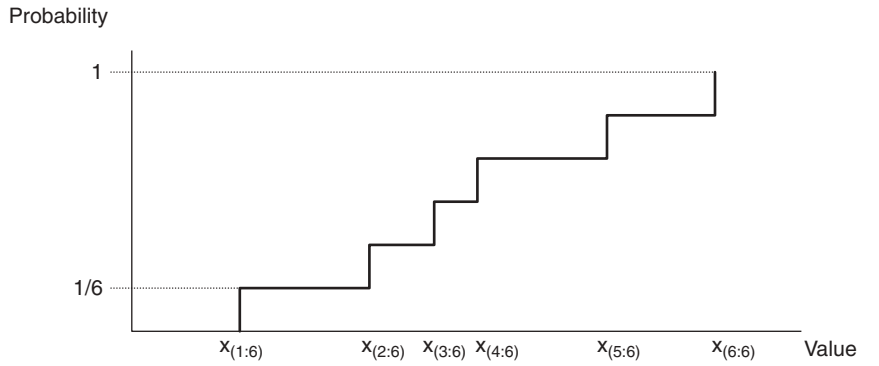



**Figure 4.3.2** Sample empirical CDF

computer program for inverse CDF of the standard normal, $\Phi^{-1}(p_j)$, it is possible to compute the quantiles at $p_j = j/(T + 1)$ for $j = 1, 2, \ldots, T$.

The Q–Q plot has normal quantiles on the horizontal axis and observed data quantiles that are simply the order statistics $x_{(j:T)}$ on the vertical axis. Probability theory texts, Spanos (1999, p. 243), explain that the theoretical basis for Q–Q plots is the "probability integral transformation" from math texts. The Q–Q plots can also be used to compare the data against other densities besides the normal. All one needs is a way to compute the inverse CDF of that distribution at points $p_j$. For formal statistical tests regarding the closeness of an ecdf to a theoretical density Shapiro-Wilk, Kolmogorov-Smirnov, and other tests, are discussed in statistical texts and monographs, including Kendall and Stuart (1979, ch. 30). We will discuss programming for selected normality tests in Chapter 9.

For our mutual fund AAAYX, note that the Q–Q plot shown in Figure 4.3.3 is dramatically far away from the 45 degree line there. If the data are close to normal, the Q–Q plot lies close to the 45 degree line, at least in the middle range. Hence it shows that the commonly assumed normal density will not be suitable for these data, and that is the monkey wrench. Thus the theme of this chapter is well illustrated by Figure 4.3.3, and it also confirms the Pearson estimates from Chapter 2.

### 4.3.3 Kernel Density Estimation

The empirical cdf with $p_j$ increasing in steps of $(1/T)$ from 0 to 1 is too jagged in appearance to directly compare to theoretical, smooth distribution. Statisticians have tried to remove the jump discontinuities among them by smoothing and averaging the nearby values, Silverman (1986). If we choose parametric densities such as $N(\mu, \sigma^2)$ or the logistic (4.3.2), we place observed data into a straightjacket of a particular form of the chosen density. The finan-

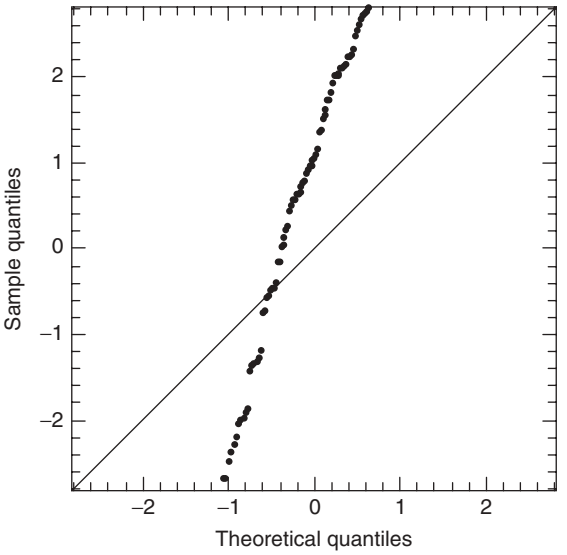



**Figure 4.3.3** Q-Q plot for AAAYX mutual fund

cial observed data (e.g., returns) may not fit into any parametric form and a parametric density may be a poor approximation.

Sometimes this problem is solved by choosing a parametric *family* of distributions (e.g., Pearson family), rather than only one member (e.g., normal) of the family. The appropriate family member is chosen by considering Pearson's plot of skewness versus kurtosis (see Figure 2.1.1). The problem of a parametric straightjacket may still persist and a flexible nonparametric density may be worth considering. Kernel density estimation is a popular method for this purpose.

The Kernel estimation begins with choosing a kernel weighting function to smooth the empirical cdf of the data. This weighting should yield a higher probability density for ranges where there are relatively more data points, and lower probability density where observations are sparse.

Let $K$ denote a kernel function whose integral equals unity to represent weights. Often this is a density function since it will have a total area of 100%. The biweight kernel is defined for $y \in [-1, 1]$ by $K_{\text{biw}}(y) = (15/16)(1 - y^2)^2$. For the same range, Epanechnikov kernel is $K_{\text{epa}} = (3/4)(1 - y^2)$. The Gaussian kernel is defined over the range of the entire real line $y \in (-\infty, \infty)$ and equals the density of $N(0, 1)$ or $K_{\text{gau}} = (1/\sqrt{(2\pi)})\exp(-y^2/2)$. These kernels provide the weight with which to multiply each observation to achieve smoothing of the data. Clearly, the weight suitable for a point $x$, whether observed or interpolated, should depend on the distance between $x$ and each observed data point $x_t$ for $t = 1, 2, \ldots, n$. The density at $x$, $f(x)$ then is given by

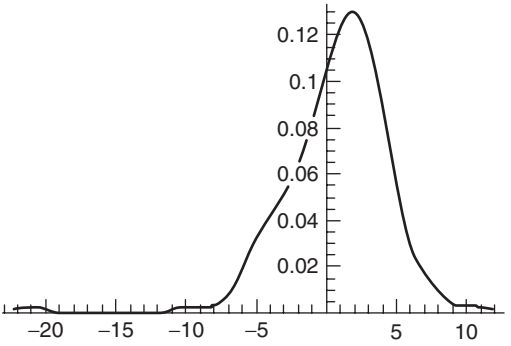



**Figure 4.3.4** Nonparametric kernel density for the mutual fund AAAYX

$$f(x) = \left(\frac{1}{nc}\right)\sum_{t=1}^{n} K(y_t), \quad y_t = \frac{(x - x_t)}{c}, \tag{4.3.4}$$

where the kernel function can be any suitable function similar to Gaussian or biweight and where $c$ is a smoothing bandwidth parameter. Since all distances of $x$ from observable data points $x_t$ are measured in units of this smoothing bandwidth parameter $c$, it appears in the denominator used in the definition of the argument $y_t$ of the kernel function. See Silverman (1986) for further details on Kernel estimation. Sheather and Jones (1991) propose an automatic method for bandwidth selection.

Consider our AAAYX mutual fund data from Chapter 2. Now Rose and Smith's (2002) software called "mathStatica" can be used for nonparametric kernel density estimation in two steps. First, we specify the kernel as the Gaussian kernel, and our second step is to choose the bandwidth denoted here by $c = 1.104$ based on the Sheather-Jones method (see Chapter 9). The nonparametric density for this dataset given in Figure 4.3.4 clearly indicates that the density is nonnormal with long left tail, suggesting skewness or the fact that downside risks are different from upside potential for growth.

## 4.4 ALTERNATIVE DISTRIBUTIONS

This section discusses some probability distributions that are potentially useful for studying financial data, specifically those that can account for downside risk and skewness. We provide graphics for these distributions so that the reader can assess their applicability in any particular situation. Recent finance literature is using many of these nonnormal distributions and their generalizations. For example, Rachev et al. (2001) recommend VaR calculations based on a stable Pareto variable, because these can be readily decomposed into the mean or centering part, skewness part, and dependence (autocorrelation)

structure. Research papers in finance often assume that the reader is familiar with many nonnormal distributions. Hence we provide some of the basic properties of the distributions listed above, with an emphasis on properties for which analytical expressions are available.

### 4.4.1 Pareto (Pareto-Levy, Stable Pareto) Distribution

Economists are familiar with the Pareto distribution as the one suitable for representing income distribution. In this section we discuss the simple Pareto density and postpone the discussion of its extensions, known as Pareto-Levy and stable Pareto, until Section 4.4.6. Longin and Solnik (2001) extend the stable Pareto to generalized Pareto and show with examples that cross-country equity market correlations increase in volatile times and become particularly important in bear markets.

1. The pdf is given by $f(x) = \theta a^{\theta} x^{-(\theta+1)}$. It has two parameters $a$ and $\theta$.
2. The cumulative distribution function, cdf, denoted by $F(x)$ has a convenient form for Pareto density: $F(x) = 1 - (a/x)^{\theta}$, where $x \geq a$ and $\theta > 0$.
3. The mean of Pareto distribution, assuming $\theta > 1$, is $(\theta a)/(\theta - 1)$.
4. The variance, assuming $\theta > 2$, is $(\theta a^2)/[(\theta - 1)^2(\theta - 2)]$.
5. The first absolute moment defined by $E|X - \mu|$ is called the mean deviation, and it measures the spread of the distribution. For the Pareto density, the mean deviation is given by $[2a(\theta - 1)^{\theta-2}]/[\theta^{\theta-1}]$.
6. The mode is defined as that value of the random variable that is observed with the maximum frequency. For the form of the Pareto distribution given above, the mode is simply $a$.
7. Given a number $q$ in the interval $[0, 1]$, the $q$th quantile $x_q$ of a random variable $X$ is defined as the solution of the equation $P(X \leq x_q) = q$. If $q = 0.5$, we have the median. The median of the Pareto distribution is $[a2^{(1/\theta)}]$.
8. The ratio of standard deviation to the mean is called the coefficient of variation and is a measure of variability of a random variable, which is not sensitive to the units of measurement. Note that the variance or the standard deviation is very sensitive to the units. Assuming $\theta > 2$, the coefficient of variation for the Pareto distribution is $\sqrt{[\theta^{-1}(\theta - 2)^{-1}]}$.
9. The $r$th raw moment about the origin, $\mu'_r$ is defined as $E(X^r)$. Assuming that $\theta > r$, for the Pareto distribution, $\mu'_r = \theta a^r/(\theta - r)$.

In Figure 4.4.1 for the Pareto distribution, we fix $\theta = 2$ and change $a$ to yield three lines: $a = 2$ (dark line), $a = 3$ (lighter line), and $a = 4$ (dashed line). It is clear from the figure and the properties discussed above that Pareto density is not suitable to represent a density of returns themselves. However, after appropriate adjustment for the direction, the Pareto density is useful in a study

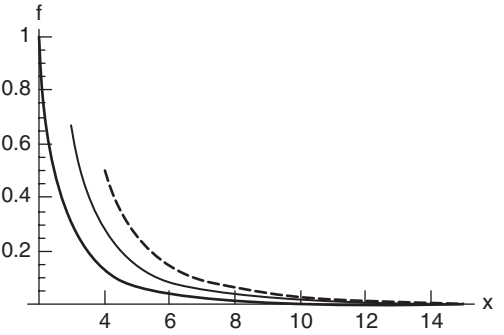



**Figure 4.4.1** Pareto density

of extreme (left) tails of the density of returns for the purpose of computing value at risk.

### 4.4.2 Inverse Gaussian

The inverse Gaussian is derived from the normal density. The density is given as follows:

1. The pdf is $f(x) = (\sqrt{\lambda/2\pi x^3})\exp[-\lambda(x-\mu)^2/2\mu^2 x]$; $x > 0$, $\lambda > 0$, $\mu > 0$.
2. The mean of the inverse Gaussian distribution is $\mu$.
3. The variance is $\mu^3/\lambda$.
4. Assuming $\varphi = (\lambda/\mu)$, the mode is $\mu\{\sqrt{[1+9/(4\varphi^2)]} - [3/(2\varphi)]\}$.
5. The coefficient of variation for the inverse Gaussian distribution is $\sqrt{[\mu/\lambda]}$.
6. If $\mu_r$ denotes the $r$th central moment, $E[X - E(X)]^r$, then we let the coefficient of skewness be defined as $\gamma_1 = \mu_3/(\mu_2)^{(3/2)}$. Since skewness is zero for the normal distribution, it can be negative if $\mu_3$ is negative. Thus, for the inverse Gaussian distribution, $\gamma_1 = 3\sqrt{[\mu/\lambda]}$.
7. Let the coefficient of excess kurtosis be defined as $\gamma_2 = [\mu_4/(\mu_2)^2 - 3]$. Clearly, it is zero for the normal distribution. For the inverse Gaussian it is $15\sqrt{[\mu/\lambda]}$.

In Figure 4.4.2 for the Inverse Gaussian distribution, we fix $\lambda = 2$ and change $\mu$ to yield three lines: $\mu = 1$ (dark line), $\mu = 2$ (lighter line), and $\mu = 9$ (dashed line). It is clear from the figure and the properties discussed above, that the inverse Gaussian density is not suitable to represent a density of excess returns unless it can be safely assumed that these are never negative or can be transformed to a new variable that is always positive. The inverse Gaussian does offer considerable flexibility of shapes.

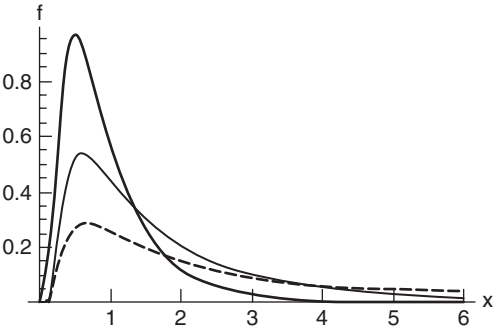



**Figure 4.4.2** Inverse Gaussian density

### 4.4.3 Laplace (Double-Exponential) Distribution

1. $f(x) = (1/2\sigma)\exp[-|x-\mu|/\sigma]$; $x$ and $\mu \in (-\infty, \infty)$ and $\sigma > 0$
2. Mean: $\mu$
3. Variance: $2\sigma^2$
4. Mean Deviation: $\sigma$
5. Median: $\mu$
6. Mode: $\mu$
7. Coefficient of variation: $(\sigma/\mu)\sqrt{2}$
8. Coefficient of skewness: $\gamma_1 = 0$
9. Coefficient of excess: $\gamma_2 = 3$
10. Odd moments: $\mu_{2r-1} = 0$ for $r = 1, 2, \ldots$; even moments: $\mu_{2r} = (2r)!\sigma^r$ for $r = 1, 2, \ldots$

In Figure 4.4.3 for the Laplace (double-exponential) distribution, we fix $\sigma = 2$ and change $\mu$ to yield three lines: $\mu = 1$ (dark line), $\mu = 2$ (lighter line), and $\mu = 3$ (dashed line). It is clear from the figure and the properties discussed above, that the Laplace density is suitable to represent a density of excess returns, except that the sharp peak in the middle and strict symmetry abound. So it may not be realistic, especially if the downside and upside are different from each other. Laplace does offer an interesting alternative to symmetric densities like the normal or student *t*.

### 4.4.4 Azzalini's Skew-Normal (SN) Distribution

Azzalini's simple SN density can give interesting new insights. The SN density is defined as

$$\mathrm{SN}(x,\lambda) = 2\phi(x)\Phi(\lambda x) = \left[1/\sqrt{(2\pi)}\right]\exp\left(\frac{-x^2}{2}\right)\left[1+\mathrm{Erf}\left(\frac{x\lambda}{\sqrt{2}}\right)\right], \qquad (4.4.1)$$

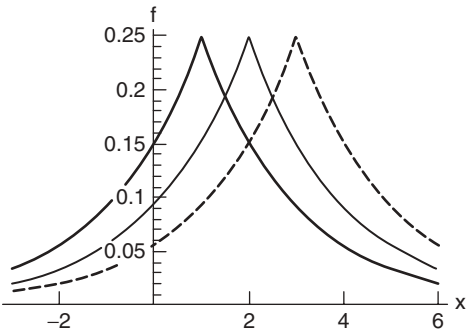



**Figure 4.4.3** Laplace or double-exponential distribution density

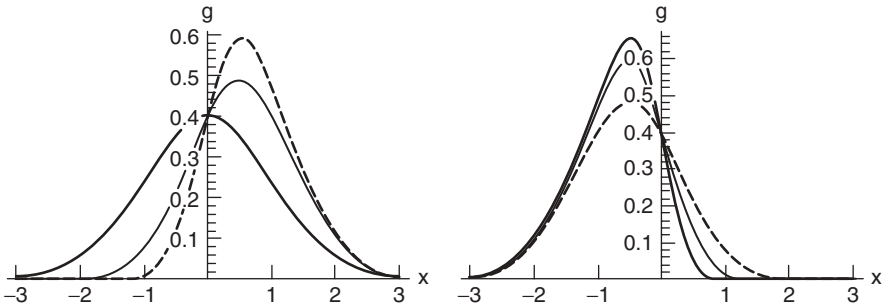



**Figure 4.4.4** Azzalini skew-normal density. *Left panel*: SN density for $\lambda = 0$ (dark line, unit normal), $\lambda = 1$ (dotted line), and $\lambda = 2$ (dashed line). *Right panel*: $\lambda = -3$ (dark line), $\lambda = -2$ (dotted line), and $\lambda = -1$ (dashed line)

where $\phi(x) \sim N(0, 1)$, the unit normal, $\Phi(\lambda x) = \int_{-\infty}^{\lambda x} \phi(t)dt$. Note that $\Phi(\lambda x)$ is the cumulative distribution function (CDF) of $N(0, 1)$, and $\lambda$ is any real number. Note that Erf, the Gaussian error function, and Erfc, its complement, are evaluated in many software programs. For $x \geq 0$, $\Phi(x) = 0.5[1 + \text{Erf}(x/\sqrt{2})]$, and for $x < 0$, $\Phi(x) = 0.5[\text{Erfc}(-x/\sqrt{2})]$. Figure 4.4.4 shows how this density behaves for various values of $\lambda$ including $\lambda = 0$ for $N(0, 1)$ and $\lambda = 1, 2, -3, -2$, and $-1$. Clearly, the SN density is positively (negatively) skewed for positive (negative) $\lambda$, and hence the parametric SN can cover a range of skew distributions.

The mean of SN can be shown to be $\lambda(\sqrt{2})[\pi + \pi\lambda^2]^{-0.5}$, which is positive (negative) only when $\lambda$ is positive (negative). The variance of SN can be shown to be $1 - 2\lambda^2[\pi + \pi\lambda^2]^{-1}$. Since $\lambda$ appears only as $\lambda^2$ in this expression, the variance is always smaller than unity, except that when $\lambda = 0$, we have special case of the unit variance of $N(0, 1)$. If we rescale $x$ and replace $x$ by $\eta x$, where $\eta$ is a real number, the variance expression becomes $\{\eta^2 - 2\eta^2\lambda^2[\pi + \pi\lambda^2]^{-1}\}$.

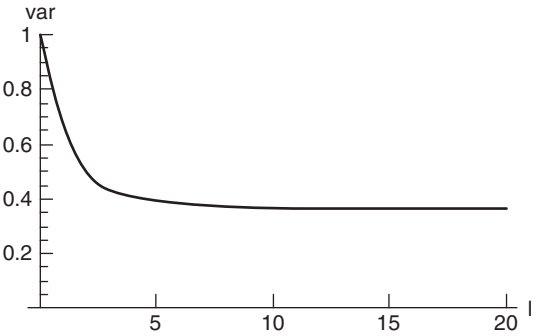



**Figure 4.4.5** Variance of Azzalini distribution as λ increases

Pearson's measures of skewness and kurtosis based on the third and fourth moments, respectively, have complicated expressions, that depend on λ.

Figure 4.4.5 plots the variance of the SN density, indicating that as λ increases, the variance stabilizes to a value near $(\pi - 2)/\pi$ or 0.36338. The behavior is similar for negative λ in the negative direction. An important lesson from this analysis is that reasonable values of the variance for skew-normal density will be in the range [1, 0.36338]. It can be shown that when $\lambda \to \infty$; the density converges to the so-called half-normal density or folded normal distribution.

The variance in financial data can be much larger than unity. Thus, if we wish to use the SN density in finance, we need to transform the centering and scaling as follows. For the example of AAAYX mutual fund of Chapter 2, the observed variance 14.1743 is much larger than unity, the basic SN is not suitable. We begin by introducing the scale term $w$ to $\phi$ appearing in (4.4.1) by writing it as $\phi' = [1/w\sqrt{(2\pi)}]\ \exp(-(x^2/2w^2)$. The corresponding cdf $\Phi$ becomes $\Phi' = 0.5[1 + \text{Erf}(x\lambda/w\sqrt{2})]$, and the revised two-parameter density $SN'[x, \lambda] = 2\phi'\ \Phi'$. Now we re-center by introducing a location parameter $\xi$ into the model and transforming $x$ to $y = x + \xi$. Then $h(y)$, the adjusted SN density (AdjSN), is

$$\begin{aligned}\text{AdjSN}(y,\lambda,w) &= 2\phi'(y)\Phi'(\lambda y)\\ &= \left[\frac{1}{w\sqrt{(2\pi)}}\right]\exp\left(-\frac{(x-\xi)^2}{2w^2}\right)\left[1+\text{Erf}\left(\frac{(y-\xi)\lambda}{w\sqrt{2}}\right)\right]. \quad (4.4.2)\end{aligned}$$

It can be verified that (4.4.2) integrates to unity, and that $\xi = 0$ and $w = 1$ yield the special case of (4.4.2) given in (4.4.1). Thus we have a version of SN suitable for estimation from financial data. MathStatica reports the log likelihood (LL) function for AdjSN and the score functions, which are the derivatives of LL with respect to the parameters. We maximize the likelihood by equivalently maximizing the LL, that is, by satisfying the first- and second-

order conditions for its maximum. The numerical solution is obtained by setting the score equations to zero and solving for the parameters $\lambda$, $\xi$, and $w$. These are called ML estimators.

For our data the maximized LL value using mathStatica is –361.792, where our first ML solutions are $\xi_m = 0.782569$, $\lambda_m = -0.0144832$, and $w_m = 3.75042$. By the first-order conditions we know from calculus that the gradients of the observed LL evaluated at the ML solution should be zero. Actually the respective gradients are –0.0049738, –0.019452, and 0.00794965. The second-order conditions for maximization from calculus in our context mean that the matrix of second-order partial derivatives of LL with respect to the three parameters (the Hessian matrix) should be negative definite, that is, all eigenvalues of the Hessian should be strictly negative. Otherwise, we end up with a minimum of the likelihood function rather than the maximum we seek. The actual eigenvalues of the Hessian matrix H are –92.8108, –18.7781, and 0.06612. Clearly, the eigenvalue 0.06612, although almost zero, raises questions about the validity of this ML solution.

In light of these problems with the first ML solution, we use mathStatica tools to search further by using the Newton search method. We simply provide new starting values near the first ML solution for this further search of the likelihood surface for a better solution. The further search yields a superior solution, since the new maximized LL = –354.229 exceeds a similar LL = –361.792 at the first ML solution. This means we are going higher on the likelihood surface. At the new ML solution the maximized LL value is –354.229, where our second ML solutions (denoted by the additional subscript 2) are $\xi_{m2} = 4.54772$, $\lambda_{m2} = -2.30187$, and $w_{m2} = 5.34561$. The new gradient vector is $(-1.98*(10)^{-6}, -3.16*(10)^{-6}, 2.19*(10)^{-7})$, where all elements are very close to zero. The eigenvalues of the Hessian (–16.0411, –12.9347, and –1.70855) are all negative, implying that H is negative definite, and the observed log likelihood is indeed concave near the solution. McCullough and Vinod (1999, 2003) recommend such searches near all ML solutions.

How good is the adjusted Azzalini skew normal AdjSN($y$, $\lambda$, $w$) fit? This may be seen in Figure 4.4.6 where the transformed variable y is on the horizontal axis and the solid line is the observed pdf. It has a flat top as in Figure 2.1.2. We compare the pdf to the dashed line representing AdjSN density fitted for these data. Since the fit is visually close we conclude that Azzalini density may be a valuable tool in finance.

Note that the negative skewness (long left tail) of the density is quantified by the negative estimate of the $\lambda$ parameter. For statistical inference the asymptotic covariance matrix is given by $(-H^{-1})$ evaluated at the estimates of the parameters. Thus the standard errors for $\xi$, $\lambda$, and $w$ are from $\sqrt{(-H^{-1})}$. We invert the Hessian matrix, evaluate it at the solution values, change the sign, and then compute the square root to get standard errors. The ratio of the estimated parameter value to the corresponding standard errors are asymptotic Student's $t$ statistics: $t$-stat($\xi$) = 9.78377, $t$-stat($\lambda$) = –4.27738, and $t$-stat($w$) = 11.4388. Since all are much larger than 2, the coefficients of AdjSN density are statistically significant. They confirm that the mean and skewness is nonzero.

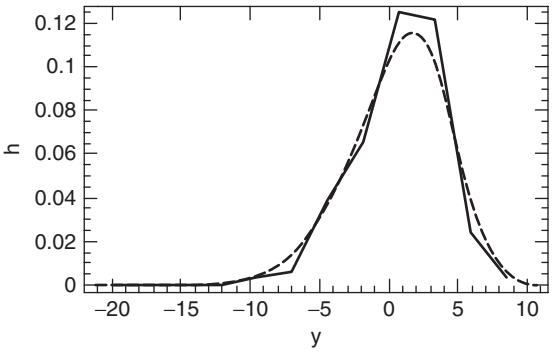



**Figure 4.4.6** Plot of empirical pdf (solid line) and fitted adjusted skew-normal density

If there were two funds with identical means and variances but two distinct skewnesses, the investor would be better off choosing the one with a larger (more positive) skewness.

The value at risk defined in Section 2.1 estimated from the location-scale SN density is \$9,222.00, rounded to the nearest dollar. If we had used the normal density, the VaR would be \$8,033.00. The advantage of SN density is revealed by the fact that the VaR is properly larger for the negatively skewed SN density for AAAYX fund.

### 4.4.5 Lognormal Distribution

The lognormal distribution is one of the most important distributions in finance. If returns are $r_t$, gross returns are $(1 + r_t)$. Let us consider log of gross returns, $\log(1 + r_t)$. Note that the Taylor series for $\log(1 + x) = x - x^2/2 + x^3/3$. . . . If we retain only the first term, we have $\log(1 + x) \approx x$. Hence in our case $\log(1 + r_t) = (r_t)$. The lognormal model says that log of gross returns is normally distributed. It has the advantage that it does not violate limited liability (Campbell et al. 1997). The expressions for the mean and variance of the lognormal will be useful in Black-Scholes option pricing formula in the next chapter.

1. $f(x) = (1/\sigma x \sqrt{(2\pi)})\exp[-(\log x - \mu)^2/(2\sigma^2)]$; $x > 0$, $\mu \in (-\infty, \infty)$ and $\sigma > 0$
2. Mean: $\exp[\mu + 0.5\,\sigma^2]$
3. Variance: $w(w - 1)\exp(2\mu)$, with $w = \exp(\sigma^2)$
4. Median: $\exp(\mu)$
5. Mode: $\exp(\mu - \sigma^2)$
6. Coefficient of variation: $\sqrt{(\exp(\sigma^2) - 1)}$
7. Coefficient of skewness: $(w + 2)\sqrt{(w - 1)}$ with $w = \exp(\sigma^2)$
8. Coefficient of excess: $w^4 + 2w^3 + 3w^2 - 6$
9. Raw moments from zero: $\mu'_r = \exp[r\mu + 0.5r^2\sigma^2]$

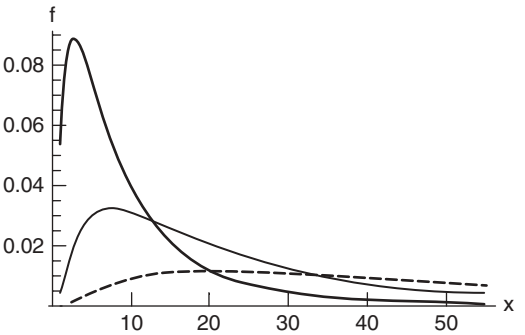



**Figure 4.4.7** Lognormal for $\sigma = 1$ and $\mu = 2$ (solid line), $\mu = 3$ (light line), and $\mu = 4$ (dashed line)

### 4.4.6 Stable Pareto and Pareto-Levy Densities

The stability property of distributions often refers to a shape parameter, which stays the same regardless of the scale. Rachev and Mittnik (2000) is a huge book devoted to a discussion of the properties of distributions include the estimation of stable distributions in finance. We discussed the simple Pareto density in Section 4.4.1. Here we consider its extensions, which are of interest in finance because they have various desirable properties of fat tails, excess kurtosis, and skewness, for example. The derivation of the stable Pareto density is done by starting with the so-called characteristic function of the Pareto, that its density is quite intricate. Since we are assuming that the reader is not familiar with mathematical statistics, we begin this subsection by explaining the moment-generating function, characteristic function, and related statistical concepts. Readers familiar with these concepts can skip to the characteristic function of stable Pareto in (4.4.6) and (4.4.7).

A function that is used to generate the moments of a random variable $X$ is called the moment-generating function (mgf). It is defined as the expected value

$$M_{\text{gf}}(t) = E\exp(tx) \tag{4.4.3}$$

provided that the expectation is a real number. If the mgf exists, the raw moments (measured from zero) are obtained by

$$\mu_r' = \left(\frac{d}{dt}\right) M_{\text{gf}}(t) \quad \text{evaluated at} \quad t = 0. \tag{4.4.4}$$

For example, the mgf for the normal density is given by $\exp[t\mu + 0.5t^2\sigma^2]$. The $(d/dt)$ of this is $D_1 = [\mu + t\sigma^2]\exp[t\mu + 0.5t^2\sigma^2]$, evaluated at $t = 0$, this equals $\mu$. Now the $(d/dt)$ of $D_1$ is $\mu D_1 + t\sigma^2 D_1 + \sigma^2 \exp[t\mu + 0.5t^2\sigma^2]$. Evaluated at $t = 0$, this second derivative becomes $\mu^2 + \sigma^2$. The well-known relations between

raw and central moments then give the second central moment as $\sigma^2$. The moment-generating function for the Pareto density does not exist, since the expectation of $\exp(tx)$ contains the imaginary number $i = \sqrt{-1}$. By contrast, characteristic functions (CFs) always exist, since they are defined to already include the imaginary number. One can obtain moments from the CF by evaluating its derivatives at $t = 0$:

$$C_{\mathrm{f}} = E\exp(itx), \mu'_r = i^{-r}\left(\frac{d}{dt}\right)C_{\mathrm{f}}(t) \quad \text{evaluated at} \quad t = 0. \tag{4.4.5}$$

The central limit theorem, which is of central importance in statistics, states that the sum of a large number (e.g., >30) of independent and identically distributed (iid) random variables, each having a finite variance, converges to a normal distribution. More generally, if the variance is allowed to be infinite, the limiting distribution is, in general, a stable density (not necessarily the normal density).

In finance there is reason to believe that some iid components (one of thousands of trades) may have very large (>30) variance. Hence stable density is an attractive model in finance. For the symmetric case Fama and Roll (1971) study the stable density. However, Hsu et al. (1974) show that the symmetric stable density does not offer much more than the normal density, and that instead of infinite variance, nonstationary volatility is better. We are attracted by the potential of stable density for asymmetric distributions, which are important for the theme of this book. Unfortunately, practical estimation tools are difficult to implement (Rachev and Mittnik, 2000).

The stable Pareto (or simply stable) density $S_{\mathrm{tab}}(\mu, v, \alpha, \beta)$ has four parameters: $\mu$ is for mean and $v$ is for the variance. Also $\alpha \in (0, 2]$, a half-open interval, is a kurtosis parameter and $\beta$ is a skewness parameter $\beta \in [-1, 1]$. Unfortunately, stable density does not have a closed form except in special cases (e.g., the normal density is a special case). However, its characteristic function does have a closed form. Since the data can be always standardized, there is no loss of generality in assuming that the mean is zero and variance is unity. Then there are only two parameters and the $C_{\mathrm{f}}$ expressions simplify but still have two parts depending on whether $\alpha$ equals unity or not. This is indicated by the additional subscripts 1 and 2:

$$C_{\mathrm{f},1}(t) = \exp\left\{-|t|^{\alpha}\left[1 + i\beta\left[\frac{t}{|t|}\right]\tan\left(\frac{\alpha\pi}{2}\right)\right]\right\} \quad \text{if } \alpha \neq 1, \tag{4.4.6}$$

$$C_{\mathrm{f},2}(t) = \exp\left\{-|t|\left[1 + i\beta\left[\frac{t}{|t|}\right]\left(\frac{2}{\pi}\right)\log|t|\right]\right\} \quad \text{if } \alpha = 1. \tag{4.4.7}$$

The underlying density $f_{\mathrm{stable}}(x)$ is obtained by inverting the characteristic function. The $f_{\mathrm{stable}}(x)$ has $x \in (-\infty, \infty)$, which is the entire real line $\Re$, only if we rule out the following two choices of parameters: (1) $\alpha < 1$ and $\beta = -1$, or

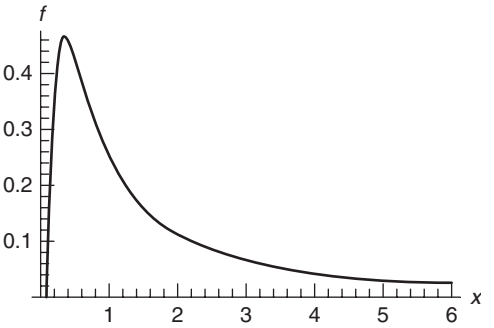



**Figure 4.4.8** Pareto-Levy density ($\alpha = 0.5$ and $\beta = -1$)

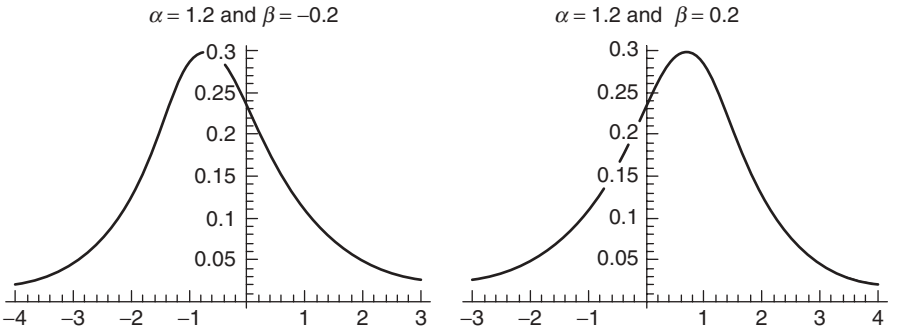



**Figure 4.4.9** Stable Pareto density under two choices of parameters (sign of β is sign of skewness)

(2) $\alpha < 1$ and $\beta = 1$. In case 1, $f_{\text{stable}}(x)$ covers only the positive part of the real line $\Re_+$. In case 2, $f_{\text{stable}}(x)$ covers only the negative part of the real line $\Re_-$. It is convenient to assume the $|\beta| \neq 1$ and focus on $C_{\text{f,1}}(t)$ in the sequel.

If $\beta = 0$, we have symmetric stable density and the characteristic function becomes $C_{\text{f}} = \exp\{-|t|^{\alpha}\}$, whose special case $\alpha = 2$ leads to the N(0, 1) density. It is well known that the CF of the N(0, 1) is $\exp(-t^2)$. Another well-known special case is when $\alpha = 1$ with the CF $\exp(-|t|)$, which is the standard Cauchy density. A third special case $\alpha = 0.5$ and $\beta = -1$, called the Levy density, is less well known with the density $f_{\text{Levy}}(x) = (2\pi)^{-0.5}\, x^{-3/2} \exp(-1/2x)$ illustrated by Figure 4.4.8.

Note that $\alpha = 2$ is the largest value of $\alpha$ possible and that as $\alpha \to 0$, the amount of probability mass in the tails (peakedness) of the stable density increases. Long left (right) tail or negative (positive) skewness is associated with negative (positive) values of $\beta$. Let us fix $\alpha = 1.2$, and let $\beta = -0.2$ or 0.2 to illustrate the behavior in Figure 4.4.9. Some additional shapes of the stable density are given in Figure 4.4.10.

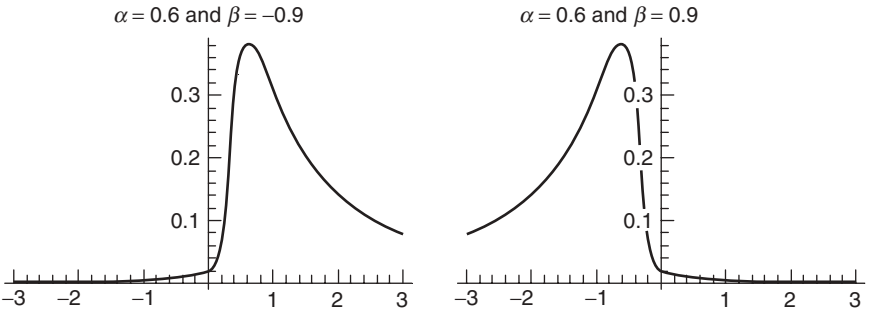



**Figure 4.4.10** Stable Pareto density (mass in tails increases as α decreases)

Since a stable distribution arises from any model that permits very large variance, it is of interest in theoretical finance. The older mean-variance optimization under normality assumption has been extended to stable distributions. Since there are a very wide variety of shapes feasible for the stable Pareto density, it remains of interest for empirical finance, except for the difficulties in estimating the parameters. Rachev and Mittnik (2000) suggest fast Fourier transform (FFT) methods for estimation, but these too remain difficult to implement. They also discuss generalizations of Makowitz mean-variance model using the stable density.

CHAPTER 5

# Downside Risk

## 5.1 VAR AND DOWNSIDE RISK

This chapter extends the discussion of Chapter 2 regarding value at risk (VaR) where we had implicitly assumed that the distribution of returns has the same time horizon as the banker or investor. For example, the artificial data example assumed that the time horizon $\tau$ was only one day. Extending the time horizon to the case where $\tau > 1$ requires a forecast of its volatility. Now that we have discussed GARCH volatility forecasts in Chapter 4 we are ready to relax the assumption of $\tau = 1$.

Recall that Wall Street investors and bankers often want a dollar figure on the potential loss in a worst-case scenario. Value at risk (VaR) is designed to assist these decision makers without burdening them with formal statistical theory. In some legal forums VaR dollar figures are used to assess if an entity is fulfilling its fiduciary obligations. For example, portfolio rebalancing rules are formulated to ensure that the banks satisfy their fiduciary obligations to the depositors. Although the computation of VaR is often a "black box" to some of these decision makers, VaR is computed quite simply from a low (mostly 1% or 5%) quantile of a parametric or nonparametric probability distribution $f(X)$ of excess returns.

Let $\Delta P_t(\tau) = P(t + \tau) - P(t)$ be the change in the value of a portfolio at time $t$ for horizon $\tau$ and let $\alpha$ (<0.5) denote a probability. The VaR is defined by the probability statement

$$\Pr[\Delta P(\tau) < -\mathrm{VaR}] = 1 - \alpha', \tag{5.1.1}$$

where the negative sign before the VaR is designed to measure losses in positive dollars. Intuitively, VaR measures a worst-case scenario loss associated

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

with "long" positions (buying side) of $\$K$. This definition is another way of stating the definition in (2.1.3), which defined $\text{VAR}(\alpha') = -R_{\alpha'}{}^{*}K$, where $R_{\alpha'}$ is the quantile of the distribution of returns.

For example, consider an investor with a time horizon $\tau$ of one year buys $K = \$100{,}000$ worth of mutual fund shares, then $P(t) = 100{,}000$ is the initial value of the portfolio. Now assume that the fund could lose 25% or more of its value in a year with probability $\alpha' = 0.01$. Then $\Pr[\Delta P(\tau) < 25{,}000] = 0.99$, implying that $\text{VaR}(0.01) = \$25{,}000$ is an upper bound on the loss.

One can compute the $\text{VaR}(\alpha')$ for any density $f(x)$ of excess returns, as well as for the empirical distribution of returns. Section 4.4 of Chapter 4 has a discussion of alternative probability distributions (Gaussian, Pareto, etc.), and a discussion of the empirical and theoretical cumulative distribution functions (CDF). For example, the CDF of the logistic density is given in (4.3.2). Recall that inverse CDF of the normal density and also of the empirical CDF were needed in the testing for normality by using the Q-Q plots described in Section 4.3.1. If the CDF is denoted by $F(x)$, its inverse evaluated at a given probability $\alpha'$ is denoted by $F^{-1}(\alpha')$. The inverse CDF of the standard normal density $N(0, 1)$ is readily known from normal tables to be $-2.33$ for $\alpha' = 0.01$ and 1.645 for $\alpha' = 0.05$. Assuming that the initial capital investment is $K$ (e.g., $K = \$100{,}000$), the mean of excess returns is $\bar{x}$, and the standard deviations is $s$, we have

$$\text{VaR}(0.01) = -K[\bar{x} - 2.33 * s], \quad \text{VaR}(0.05) = -K[\bar{x} - 1.645 * s]. \tag{5.1.2}$$

Because it relies on asymptotic normality, (5.1.2) is often used to compute $\text{VaR}(\alpha')$. Statisticians have worked out similar $F^{-1}(\alpha')$ values for a variety of densities analytically or numerically. It is clear from formula (5.1.2) that we need to change the location of the density to zero by using the observed mean and to change the scale by using the observed standard deviation.

In general, for parametric probability distributions $f(x, \theta)$ the density depends on parameters $\theta$. For example, the normal density depends only on the parameter vector $\theta = \{\mu, \sigma^2\}$ containing its mean and variance. Denote by $F^{-1}(\alpha', \theta)$ the inverse CDF of a distribution with a vector of parameters denoted by $\theta$. Then a general formula for VaR is

$$\text{VaR}(\alpha') = -KF^{-1}(\alpha', \theta). \tag{5.1.3}$$

This formula can be used only if the inverse CDF and all parameters are known. It is known only numerically (from tables) for the $N(0, 1)$ density, that is when the mean $\mu = 0$ and variance $\sigma^2 = 1$. The negative sign in (5.1.3) may be confusing at first sight if $F^{-1}(\alpha', \theta)$ happens to be positive. It simply represents profits, not losses at the low quantiles. Thus, whenever VaR is negative, the worst-case scenario is not a loss but a positive profit, which is, of course, a desirable outcome.

Note that a familiar nonstandardized transformation of location and scale is used to accommodate the case where the mean is nonzero and variance is different from unity. We state a general formula for VaR involving the destandardization as

$$\text{VaR}(\alpha') = -K[\mu - \sigma F^{-1}(\alpha', 0, 1)], \tag{5.1.4}$$

where $f(x, \mu, \sigma^2)$ is the underlying density for excess returns over a time period $\tau$, and $F^{-1}(.)$ represents the inverse CDF. Usual observable densities are over only one time period. Hence, if such data are used, there is an implicit assumption that the time horizon of the investor is also $\tau = 1$ period. We will relax this assumption later in Section 5.1.3.

### 5.1.1 VaR from General Parametric Densities

Recall the discussion of the Pearson family of densities from Chapter 2 and Appendix 1 of that chapter. These are readily used as a generalization of the normal density. The Pearson family remained a theoretical curiosity for the last century from the viewpoint of finance. However, thanks to modern computers and software tools, the Pearson members have become practical for applications in finance. Rose and Wood (2002, ch. 5) discuss computer tools for estimating in a fairly mechanical fashion any member of Pearson types I to VII from data on excess returns. The software first estimates the moments of the density and plots a graph of skewness parameter versus kurtosis parameter on the two axes. The software also indicates from that plot which Pearson type the data belongs to. The user specifies the type, and software fits it to the data. Chapter 2 gave an example of an estimator based on Pearson family for computation of $\text{VaR}(\alpha')$.

Rose and Wood (2002) also provide software tools for fitting the Johnson family of distributions, indicated as SL for lognormal, SU as unbounded, and SB as bounded densities. All parametric densities $f(x, \theta)$ are distinguished by the presence of the vector of parameters $\theta$, which is first estimated from data, usually from estimated moments of different orders. For example, the sample mean and variance are used to define the normal density.

Rachev et al. (2001) also recommend VaR calculations based on stable Pareto variable because these can be readily decomposed into the mean or centering part, skewness part, and dependence (autocorrelation) structure. Longin and Solnik (2001) extend the stable Pareto to generalized Pareto and show with examples that cross-country equity market correlations increase in volatile times and become particularly important in bear markets. Since these correlations are neither constant over time nor symmetric with respect to the bull and bear markets, these authors reject multivariate normal as well as multivariate GARCH with time-varying volatility.

The estimation of VaR is related to estimating the worst-case scenario in terms of the probability of very large losses in excess of a threshold $\theta$. The so-

called positive θ-exceedances correspond to all observed losses that exceed θ (e.g., 10%). Longin and Solnik (2001) estimate generalized Pareto distribution parameters using 38 years of monthly data. As θ increases, the correlation across markets of large losses does not converge to zero but *increases*. This is ominous for an investor who seeks to diversify across corrupt developing countries. It means that losses in one country will not cancel with gains in another country. We conclude that realistic VaR calculations show that similar countries might suffer extreme losses all at the same time.

### 5.1.2 VaR from Nonparametric Empirical Densities

Nonparametric densities use no parameters at all. We illustrate the nonparametric density by the example of the empirical CDF (ecdf) as follows. It is easier to visualize the ecdf which changes from 0 to 1, rather than the density $f(x)$. The observed data $x_t$ with $T$ observations can be first arranged in increasing order of magnitude to define the so-called order statistics, $x_{(t)}$. Note that $\min(x_t) = x_{(1)}$ and $\max(x_t) = x_{(T)}$. Next, one assumes that the empirical data dictate the underlying density for the random variable $X$ without any flexibility beyond the observed values themselves. Then the probability $\Pr(X < x_{(1)}) = 0 = \Pr(X = x_{(T)})$. The construction of the empirical CDF does not permit any intermediate values either. Thus the ecdf = 0 if $(X < x_{(1)})$. It jumps to the height of $(1/T)$ at $(X = x_{(1)})$ and remains flat until $(X = x_{(2)})$, at which time it jumps to the height $2/T$ and remains flat until $(X = x_{(3)})$, and so on. At $(X = x_{(T)})$ it attains its full height = $T/T$ or 1, and remains there for all $(X > x_{(T)})$. The value at risk for a nonparametric density is obtained by inverting the ecdf as

$$\text{VaR}(\alpha') = -K[\text{ecdf}]^{-1}(\alpha'), \tag{5.1.5}$$

where the inverse of ecdf is denoted by $[\text{ecdf}]^{-1}$, and we evaluate it numerically by linear interpolation at $\alpha'$ (e.g., $\alpha' = 0.01$) to represent the lower quantile as before.

As a practical example, consider a mutual fund named Alliance All-Asia Investment Advisors Fund, with the ticker symbol AAAYX, for the period of $T$ = 132 months from January 1987 to December 1997 from Morningstar (2000). Now we construct the eCDF for these data and evaluate its lower quantiles. For a $K$ = \$100,000 capital, the potential loss $\text{VaR}_{0.01}$ = \$9,629 is obtained by applying (5.1.5). Its interpretation is simply that we expect that the probability that the actual loss to be *less* than \$9,629 is 0.99.

There are semiparametric densities, which are in-between fully parametric ones from the Pearson or Johnson family, and fully nonparametric based on the ecdf. If a bandwidth parameter for smoothing of nearby observations is chosen, it is possible to devise a nonparametric VaR based on the kernel density from Chapter 4.

Vinod (2004) argues that in finance the ecdf has severe limitations. The assumption that $\Pr(X < x_{(1)}) = 0 = \Pr(X > x_{(T)})$ means one can never obtain

returns beyond the observed range is almost ludicrous. Also intermediate values within factions of the observed returns may well be realistic. One solution was the kernel density from Chapter 4. Alternatively, Vinod (2003b, 2004) uses the principle of minimum (prior) information or maximum entropy (ME) to propose an ME density and ME cumulative density. Along the left-hand tail of the distribution, the ME principle leads to the exponential density with one parameter based on the average of the $x_{(1)}$ and $x_{(2)}$. This is an example of a semiparametric density, since both tails involve one parameter. Clearly, one can apply (5.1.5) upon replacing eCDF by ME-CDF.

Vinod and Morey (2002) note that financial economists usually ignore the estimation risk. This is also true of the commonly calculated VaR, which focuses only on market investment risk, not VaR estimation risk. Hence there is a need to develop inference methods for VaR based on the sampling variability in an ensemble of time series $\Omega$ from which comes the observed $x_t$. Using the maximum entropy density, Vinod (2004) has proposed a maximum entropy algorithm (ME-alg) for this purpose and it creates a large number $J$ of plausible but distinct time series similar to $x_t$. The $J$ time series are then used to create $J$ estimates of VaR, which can then be ordered from the smallest to the largest and denoted by $\text{VaR}_{(j)}$ with $j = 1, 2, \ldots, 999$, viewed as order statistics. An obvious possibility is to choose the ordered ensemble estimate $\text{VaR}_{(10)}$ as the estimate of VaR. For the example of the mutual fund AAAYX discussed above, we have implemented this ME-alg and find that the VaR increases from \$9,629 to \$15,388, which incorporates both estimation risk and market investment risk.

To illustrate the effect of downside risk, we will go through the VaR steps for the S&P 500 with 10 years of monthly data from 1994 to 2003. Assuming normality, the S&P has an average monthly return of 0.8108% and a variance of 0.0021. (These numbers are rounded off. Calculations are done with eight decimal places.) The typical VaR for the first percentile would be nearly a 10% loss on the principal (–0.09843).

Incorporating downside risk, we can look at the nonparametric kernel density plot of the S&P depicted in Figure 5.1.1 seems to be skewed to the left, and should warn us that downside risk may be higher than when assuming normality.

We next estimate the Pearson family and plot in Figure 5.1.2 the S&P. As the figure shows, it falls squarely in the type I Pearson family.

The one percentile of returns for the Pearson estimated distribution of the S&P is –11.55% of principal. This exceeds the VaR with the normal distribution by 1.71% of principal, or 17% higher. This is an especially large discrepancy when considering that these are monthly returns (1/12 of the year), and that S&P index contains larger than average corporations on major exchanges. Therefore ignoring the downside risk could leave an investor with larger losses than VaR more than 1% of the time.

Moving to a more risky arena, we have the Russell 2000 index of small-cap firms, which has the Pearson plot and density shown in Figure 5.1.3. In looking

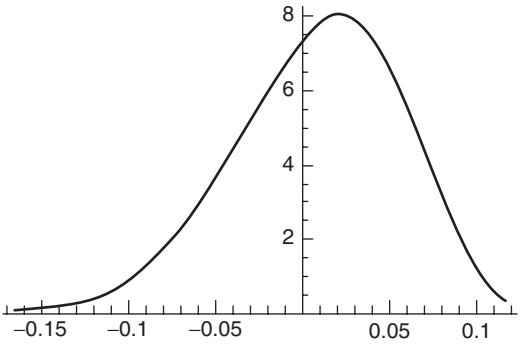



**Figure 5.1.1** Kernel density plot of the S&P 500 monthly returns, 1994 to 2003

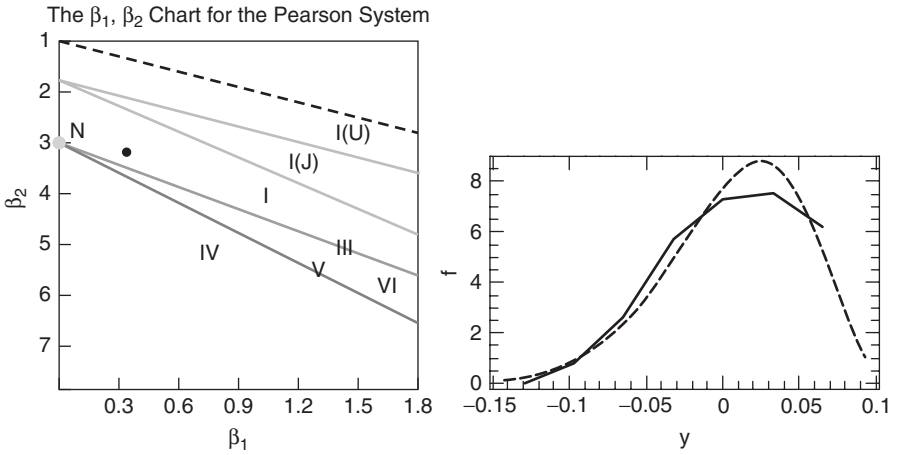



**Figure 5.1.2** Pearson plot choosing type I, the fitted type I density (dashed line), and density from raw monthly data (1994–2003) for the S&P 500 index

for the next new thing, investors often turn to smaller, growth companies. One must be careful, however, not to be burned by a hot new investment. Smaller companies are often newer, less tested, and they generally have less capital than the larger companies measured by the S&P 500. This means that their stocks are more volatile than stocks of larger, more capitalized companies. Note that during this period the Russell 2000 in Figure 5.1.3 has an average monthly return of 0.8046%, comparable to the S&P 500, and a variance of 0.0032, 50% higher than the relatively larger S&P companies. Also, compared with the Pearson plot, the Russell 2000 index is skewed negatively.

Calculating the one percentile VaR based on the normal distribution would lead us to believe a 12.4% loss was the largest we should expect. The Russell 2000 data indicates that it is in the type IV Pearson family, and not normally

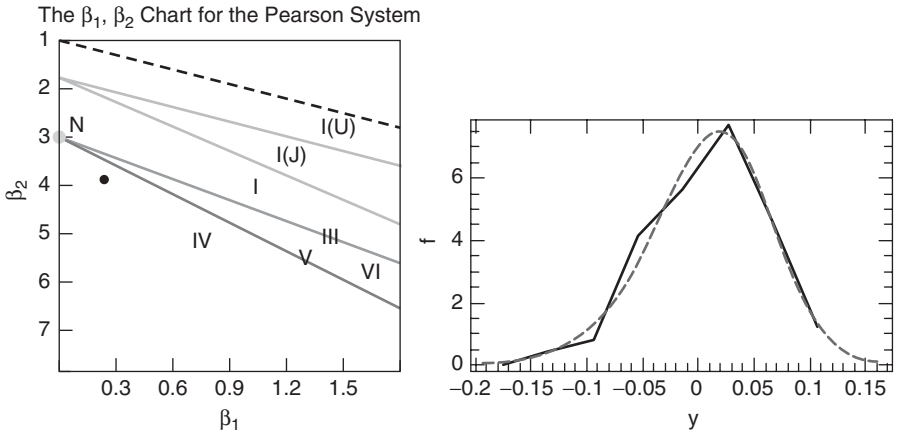



**Figure 5.1.3** Pearson plot choosing type IV, the fitted type IV density (dashed line), and density from raw monthly data (1994–2003) for the Russell 2000 index

distributed. Factoring in downside risk reduces this number to a 14.7% loss, over a 2% discrepancy, and more than the difference shown by the S&P 500.

### 5.1.3 VaR for Longer Time Horizon ($\tau > 1$) and the IGARCH Assumption

So far we have assumed that the time horizon $\tau$ for which the density $f(x)$ is available is also the time horizon for which the investor or Bank is interested in investing. Generally, one observes the excess returns $x_t$ for a reference time period (one-day, one-month, one-quarter, etc.) in annualized percentage terms. This time period need not coincide with the investment period of the investors and it is unrealistic to develop $f(x)$ for each conceivable time horizon used by the investors.

Consider an artificial example where $x_t = \{5, 4, 3, 5\}$ data are annualized percentage returns for $t = 1, 2, \ldots, \tau = 4$ periods. Accordingly the initial \$100 investment becomes 105 at the end of the first year. If the initial investment is \$1, it becomes $(1 + y_1)$, where $y_t = [x_t/100]$. Since we do not liquidate the investment at the end of the first year, we need to use compound interest type calculation. At the end of two years \$1 becomes $(1 + y_1)(1 + y_2)$. In general, for initial investment of \$1 we have

$$(1+R_\tau) = \text{Investment position after } \tau \text{ years} = (1+y_1)(1+y_2)\ldots(1+y_\tau), \quad (5.1.6)$$

where $y_t = [x_t/100]$. Since the Taylor series for $\log(1 + y)$ is $y - y^2/2! + \ldots$, a linear approximation of $\log(1 + y)$ is simply $y$. Now taking the log of both sides and using the linear Taylor approximation, we have

$$R_\tau = y_1 + y_2 \ldots + y_\tau. \quad (5.1.7)$$

That is the overall return after $\tau$ periods may be approximated as the sum of each period returns. In Section 4.4.5 we noted that the lognormal model says that the logs of gross returns, $\log(1 + y_t) \approx y_t$, may be assumed to be normally distributed. Assume that each realization $y_t \sim N(\mu', \sigma'^2)$, is the observable density for $\tau = 1$ time horizon. Since the mean of the sum of normals is sum of means, $E(y_1 + y_2 \ldots + y_\tau) = \tau\mu'$. If we further assume that these realizations are serially independent, the variance of the sum is the sum of variances. That is, $\text{var}(y_1 + y_2 \ldots + y_\tau) = \tau\sigma'^2$. Recall that $100y_t = x_t$ is a scale change, so we can let $\mu = 100\mu'$ and $\sigma' = 100\sigma$, obtaining $x_t \sim N(\mu, \sigma^2)$. Thus, upon allowing a time horizon of $\tau$ periods, we have the value at risk,

$$\text{VaR}(\alpha', \tau) = -K\left[\tau\mu - \left(\sqrt{\tau}\right)\sigma F^{-1}(\alpha', 0, 1)\right]. \tag{5.1.8}$$

RiskMetrics™ methodology for VaR computation assumes that $\mu = 0$ and proposes a simple relation between the VaR discussed so far and time horizon $\tau$ given by

$$\text{VaR}(\alpha', \tau) = \left(\sqrt{\tau}\right)\text{VaR}(\alpha'). \tag{5.1.9}$$

This is called the square root of time rule by RiskMetrics. We have shown the sense in which this is theoretically justified if the underlying density has zero mean ($\mu = 0$).

The derivation of (5.1.8) and (5.1.9) used the relation $\text{var}(x_1 + x_2 \ldots + x_\tau) = \tau\sigma^2$, which will not be true if the volatility of prices is allowed to change over time and if they are serially dependent. Our next task is to study this assumption in light of the GARCH(1, 1) model for variance of returns given in (4.2.6) with a special choice of parameters $\alpha_0 = 0$ and $\alpha_1 = 1 - \gamma_1$. This choice suggests that there is a unit root (driftless random walk) in the autoregressive AR side of the equation and $\alpha_1 + \gamma_1 = 1$. It is often called integrated GARCH or IGARCH(1,1) volatility process:

$$E\left[\varepsilon_t^2 | \varepsilon_{t-1}\right] = \sigma_t^2 = (1 - \gamma_1)\varepsilon_{t-1}^2 + \gamma_1\sigma_{t-1}^2. \tag{5.1.10}$$

Now repeated substitution yields the result that after $\tau$ time periods, the volatility $\sigma^2$ becomes $\tau\sigma^2$. This is necessary to say that standard deviation for horizon $\tau$ is $(\sqrt{\tau})$ times the original standard deviation and justifies (5.1.9). Of course, our point in explaining the derivation is that (5.1.9) is based on possibly unrealistic and strong assumptions to achieve simplicity. We suggest using modern computers to directly obtain more realistic estimates of future returns over $\tau$ time periods with actual compounding of gross returns to estimate what the value of $K$ will be at the end of the period in a worst-case proportion $\alpha'$ (1%) scenario and then estimate the potential loss as the value at risk $\text{VaR}(\alpha', \tau)$.

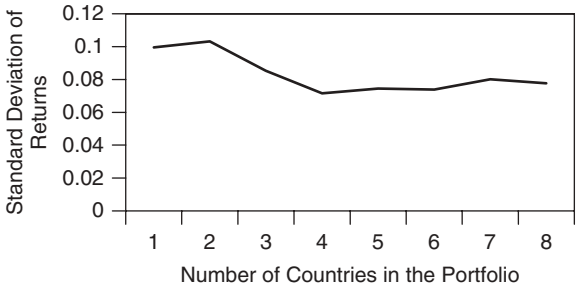



**Figure 5.1.4** Diversification of emerging market stocks

### 5.1.4 VaR in International Setting

We looked at diversification in Chapter 2 by plotting standard deviation of a US stock portfolio as more stocks are added to the portfolio. Figure 5.1.4 does the same exercise for emerging markets. In the figure we chose a stock index from Argentina, Brazil, Greece, India, Korea, Mexico, Thailand, and Zimbabwe, the countries for which there is a long span of stock market data. We started with Argentina and progressively added countries to our portfolio.

Note in the figure that while there is some diversification, standard deviation does not fall much, and it is still double that of the standard deviation for US stocks. Diversification alone is not enough for investing globally. We will continue examining the framework of downside risk to have all of our tools at hand, and come back to the solutions in Chapter 10.

As additional evidence, there is a vast and growing literature on VaR for global investing. In this section we first mention some recent work on VaR computation using generalized stable Pareto distributions for modeling market and credit risks and using international data (Omran, 2001). Later we will mention other recent work that discusses how common use of VaR by international investors and banks tends to discourage foreign direct investments (FDI) in poor countries that also suffer from corruption (Vinod, 2003a).

Stable Pareto random variables are commonly described by their characteristic functions and by four parameters: tail index $\alpha$, skewness $\beta$, location $\mu$, and scale $\sigma$. Modeling with such parameters can depict fat tails and skewness of realistic return distributions. Omran's (2001) estimates of $\alpha$ for Japan, Singapore, and Hong Kong markets are around 1.50, indicating high probability of large returns.

Vinod (2003a) considers an international investor whose portfolio consists of many assets. The VaR for any portfolio is computed by decomposing it into "building blocks" that depend on some risk factors. For example, currency fluctuation and corruption are risk factors with open economy international investments. Risk professionals first use risk categories for detailed separate analyses. They need complicated algorithms to obtain total portfolio risk by aggregating risk factors and their correlations.

Consider an example of an international investor who wishes to include some developing countries. A well-known barrier to foreign direct investment is that exchange rates (currency values) fluctuate over time, and this can mean a loss when the return is converted into the investor's home currency. The individual investor must always allow for potentially unfavorable timing of currency conversion. However, financial markets have derivative instruments including forward and future exchange rate markets to hedge against such risks. Hence derivative securities linked to exchange rates at a future date can mitigate, if not eliminate, the exchange rate risk, at least for large investors. Arbitrage activities by traders can be expected to price the foreign investments appropriately different from domestic investments to take account of exchange rate risk.

Of course, these hedging activities need free and open markets in target currencies. For developing countries like Nigeria and India, which have exchange control, there are black markets with a fluctuating premium over the official exchange rate. This means that in such countries corruption becomes a risk factor that enhances the exchange rate risk and risk associated with related costs.

Vinod (2003a) argues that corruption is an additional risk factor. Corruption can suddenly lead to a cancellation of a contract or sudden and unexpected increases in the cost of doing business. Depending on the magnitude of corrupt practices, the cost can vary considerably. When we consider almost the entire world, enforcing property rights, especially in developing countries, can be expensive and time-consuming. Furthermore both corruption and lack of transparency increase the cost of investing in corrupt countries, even if the investor does not bribe anyone. We claim that corruption increases the cost of enforcement of all property rights and becomes an additional burden.

The VaR calculations by major investment houses and banks are often extended to joint portfolios of mutually dependent but varied classes of assets. The traditional parametric VaR based on asymptotic theory does have some difficulty in handling joint dependencies and complicated return computations involving taxes and state-contingent estimates of future gains and expenses. Hence these computations often use simulations and computational power instead of purely parametric statistical tools. We will examine these tools in more depth in Chapter 9.

## 5.2 LOWER PARTIAL MOMENTS (STANDARD DEVIATION, BETA, SHARPE, AND TREYNOR)

Lower partial moments are a relatively old concept, whereby the moments are defined over a lower portion of the support of the distribution $f(x)$ of portfolio returns. In finance one is often interested in the portfolio risk of loss which obviously refers to the lower portion of the support of $f(x)$. Bawa (1975)

cites references and proves that "mean-lower partial variance" selection rule is optimal for some nonnormal $f(x)$. This section defines the lower partial moments including the variance.

In statistics, partial moments of $x$ are moments defined over a subset of the original range $[x_*, x^*]$. Lower partial moments are defined over the lower part of the range, $[x_*, 0]$, or the downside. If the actual returns are less than the benchmark return, there is a "loss" ($x < 0$) that is the risk.

The square root of the lower partial variance may be called the downside standard deviation (DSD). Note that DSD can be explicitly stated for the standard normal. If $f(x)$ is unit normal or N(0, 1) with $x \in (-\infty, \infty)$, the downside standard deviation (DSD) is the square root of lower partial variance. Thus

$$\text{DSD} = [-x\phi(x) + \Phi(x)]^{0.5}, \tag{5.2.1}$$

where $\phi(x)$ is the PDF and $\Phi(x)$ is the CDF of N(0, 1). Hence theoretical DSD can be found for any $x \in (-\infty, \infty)$ from N(0, 1) tables. In particular, if we consider $x^* = 0$, DSD $= (0.5)^{0.5}$. In general, there is a somewhat complicated monotonic nonlinear relation between $\sigma$ and DSD if $x$ is distributed exactly as N(0, $\sigma^2$).

In the context of portfolio analysis we can define the downside as whenever there is a loss to the investor. The loss can be defined to occur when the return for the asset $i$ at time $t$ is less than the risk-free return $x_{i,t} < x_{f,t}$, that is, when the excess return is negative. Alternative definitions could be if the return is below the market returns, or below some risk-adjusted target return as in CAPM and tracking error. It is convenient to suppress the subscript $i$ and denote the excess return by $x_t$.

Under exact normality of excess return distribution, it can be shown that the ranking of the portfolio risks by the usual standard deviation $\sigma$ and DSD is exactly the same. This is true because DSD is a monotonic nonlinear function of the usual $\sigma$. In practice, the excess return distribution is rarely, if ever, precisely normal. Hence it is not surprising that the ranking of risk by DSD does not coincide with the usual ranking of risk by $\sigma$. Xie (2002) shows how portfolio choices change when one considers downside risk.

It is convenient to consider $T$ observations for a given portfolio with known excess returns $x_t$, which are not classified into any intervals. Let $T'$ denote the number of observations in the downside range of values of $x_t$ where a "loss" occurs. Now DSD can be computed by giving a zero weight to all $x > 0$. We modify the usual definition of the variance to focus only on the downside by inserting a weight as

$$s_{\text{dn}}^2 = \frac{\sum_{t=1}^{T'} w_t (x_t - \bar{x})^2}{\sum_{t=1}^{T'} w_t}, \tag{5.2.2}$$

where $w_t = 0$ if $x > 0$, and where the subscript "dn" refers to the downside.

Next we define the weighted versions of higher central moments as $m_{jw}$ for $j \geq 3$:

$$m_{jw} = \frac{\sum_{t=1}^{T'} w_t (x_t - \bar{x})^j}{\sum_{t=1}^{T'} w_t}. \tag{5.2.3}$$

In particular, the weighted third central moment, which can be negative, is

$$m_{3w} = \frac{\sum_{t=1}^{T'} w_t (x_t - \bar{x})^3}{\sum_{t=1}^{T'} w_t}. \tag{5.2.4}$$

The estimated $m_{3w}$ (skewness) has a sampling distribution with high variance due to the third power of deviations from the mean needed in (5.2.4). Now the (possibly negative) skewness measure is defined by scaling with standard deviation as

$$S_k = \frac{m_{3w}}{s_{dn}^3}. \tag{5.2.5}$$

It is a common impression that $S_k = 0$ implies symmetry. However, Ord (1968) gives some asymmetric distributions with as many zero odd-order moments as one desires. One example, cited by Kendall and Stuart (1977, p. 96, ex. 3.26), has an asymmetric distribution with all odd-order moments zero. Although these results mean that $S_k = 0$ does not guarantee symmetry, $S_k \neq 0$ does mean asymmetry.

For example, consider a set of three portfolios with deterministic (fixed) returns $x_t = m^{(1)}$, $m^{(2)}$, and $m^{(3)}$ for all $t$, with means $m^{(1)}$, $m^{(2)}$, and $m^{(3)}$. Hence their standard deviations are all zero $s = s^{(i)} = 0$ for $i = 1, 2, 3$. Assume that $m^{(1)} > 0$, $m^{(2)} = 0$, $m^{(3)} < 0$. In common parlance, the third portfolio with a negative mean will lose money and is more risky compared to the others. The portfolio with positive $m, m^{(1)}$ has a profit potential. However, since $s$ is exactly same for the three, the summary measure average return already correctly ranks them as $m^{(1)} > m^{(2)} > m^{(3)}$. The summary measure of risk we seek to define for stochastic returns should be distinct from the mean and focus on losses arising from volatility, or changes in returns beyond differences in average returns.

We argue that a summary measure of financial risk (RSK) should rank the return variabilities in the reverse order of preference. In portfolio analysis a low RSK means high utility. The standard deviation is always nonnegative, but

the skewness $S_k$ can be negative and is defined on the entire real line $(-\infty, \infty)$. We propose a new measure of risk defined as

$$\text{RSK}(c) = (s - cS_k), \tag{5.2.6}$$

where $c > 0$ is a constant parameter for each individual, depending on his or her attitude toward risk. Note that RSK($c$) combines the standard deviation $s$ and $S_k$ into a single summary measure. The negative sign before $c$ recognizes that negative skewness implies greater preponderance of losses, $\Pr(x_t < 0) > \Pr(x_t > 0)$. The constant $c$ in (5.2.6) is set at $c = 1$ to define a benchmark RSK $= (s - S_k)$. It is close to the intuitive idea of risk that incorporates standard deviation and skewness. Formally, a meaningful RSK measure should satisfy the following properties:

- **R1.** If two portfolios have common mean, $m^{(1)} = m^{(2)}$, unknown $s$ and $\text{RSK}(c)^{(1)} < \text{RSK}(c)^{(2)}$, then $\text{RSK}(c)^{(2)}$ should have a higher probability of a loss, $\text{P}(x_t < 0)$.
- **R2.** If two portfolios have a common mean, common $s$, and distinct signed skewness measures $S_k^{(1)} < S_k^{(2)}$, then $\text{RSK}(c)^{(1)}$ should be higher. A more negative skewness should lead to a higher evaluation of risk (RSK).
- **R3.** It should be possible to map the RSK ranking on the real line $(-\infty, \infty)$ with a higher RSK suggesting a portfolio with a higher potential loss.

Note that we cannot rule out the possibility that a meaningful RSK measure is negative. If two portfolios have common $(\bar{x}, s)$ and distinct $S_k^{(1)}$ and $S_k^{(2)}$, the benchmark RSK $= s - S_k$ can be negative since $S_k > s$ cannot be ruled out. For example, if $S_k^{(1)} > s^{(1)} > 0$, we have $\text{RSK}(c)^{(1)} < 0$, though $s^{(1)}$ is nonnegative. The interpretation is simply that the variability of the returns of the first portfolio has a greater profit potential than loss potential. The negativity does not change the fact that the first portfolio dominates the second. Since we are familiar with positive volatility measured by $s$, the negative RSK may seem unfamiliar or puzzling to financial analysts. Fortunately, a simple shifting of the origin can readily remove the puzzling negativity.

### 5.2.1 Sharpe and Treynor Measures

The population value of Sharpe's (1966) performance measure for portfolio $i$ is defined as

$$\text{Sh}_i = \frac{\mu_i}{\sigma_i}, \qquad i = 1, 2, \ldots, n. \tag{5.2.7}$$

It is simply the mean excess return over the standard deviation of the excess returns for the portfolio.

The population value of Treynor's (1965) performance measure is

$$\mathrm{Tr}_i = \frac{\mu_i \sigma_m^2}{\sigma_{im}} = \frac{\mu_i}{\beta_i}, \qquad i = 1, 2, \ldots, n, \tag{5.2.8}$$

where the subscript "m" is used to indicate the market proxy portfolio often denoted by the S&P 500 index and $\beta_i = \sigma_{im}/\sigma_m^2$ is a regression coefficient (the familiar beta) from the capital asset pricing model (CAPM). It is based on the covariance of $i$th portfolio with the market portfolio and the variance of the market portfolio.

Now we turn to further modifications of Sharpe and Treynor measures based on downside standard deviation (DSD) of (5.2.2) to incorporate preference for positive skewness. The Sharpe ratio similar to (5.2.7) becomes

$$\mathrm{DSh}_i = \left[\frac{\bar{x}_w}{\mathrm{DSD}}\right]_i. \tag{5.2.9}$$

***A Sharpe Ratio Problem.*** An old unsolved problem with the Sharpe ratio has been that when its numerator is negative, it gives incorrect ranking of portfolios. For example, consider two portfolios with a common mean of (–1) and distinct standard deviations 1 and 10. Now the $\mathrm{Sh}^{(1)} = \bar{x}/s = -1$ and $\mathrm{Sh}^{(2)} = -1/10 = -0.1$ implying that $\mathrm{Sh}^{(1)} < \mathrm{Sh}^{(2)}$. However, it is obvious that the $\mathrm{Sh}^{(2)}$ with negative return and high standard deviation is doubly undesirable. We claim that considering the absolute values in the numerator $\bar{x}$ cannot solve this problem. Consider two portfolios with $\bar{x}^{(1)} = -1$, $\bar{x}^{(2)} = -3$, $s^{(1)} = s^{(2)} = 1$. Here the absolute values will obviously give the wrong ranking: $\mathrm{Sh}^{(1)} = 1 < \mathrm{Sh}^{(2)} = 3$.

***Add Factor for the Sharpe Ratio.*** The Sharpe ratio is not defined for zero variance portfolios, since the denominator cannot be zero. The sign of Sharpe ratio depends only on the sign of the numerator, since the standard deviation in the denominator must be positive. Adding a common positive constant to all $\bar{x}$ values in the numerator does not change the ranking of Sharpe ratios. Hence we propose a simple "add factor" to $\bar{x}$ that will solve the problem of misleading rankings by the traditional Sharpe ratio in the presence of negative returns. Of course, the add factor has to be large enough to make all numerators positive: If the add factor > $\max(|\bar{x}|)$, the ranking becomes correct, despite the presence of negative $\bar{x}$ values. Since these add factors are quite arbitrary, they should be used only if needed in the presence of negative means and only for ranking purposes.

Similar to (5.2.8) a new downside version of the Treynor measure is

$$\mathrm{DTr}_i = \left[\frac{\bar{x}_w}{\beta_w}\right]_i, \tag{5.2.10}$$

where the CAPM regression is replaced by a weighted regression, which uses nonzero weights only on the downside.

## 5.3 IMPLIED VOLATILITY AND OTHER MEASURES OF DOWNSIDE RISK

The up and down variability of prices is often referred as volatility on the Wall Street. Formally, it may be defined as the relative rate at which an asset's price $P_t$ at time $t$ moves up and down. The standard deviation $\sigma$ is the most common estimate of volatility, which is also the most common measure of scale in statistics. On Wall Street the volatility $\sigma$ is usually found by calculating the annualized standard deviation of daily *change* in price, not the price itself.

If the price of an asset fluctuates up and down rapidly over a wide range of values during a short time period, it has high volatility. If the price almost never changes, it has low volatility. For example, the price of a speculative stock of a small company often has high volatility, whereas the yield on a passbook savings account has low volatility.

The implied volatility requires a backward use of the famous and highly mathematical Black-Scholes (BS) option pricing formula. As explained elsewhere in this book (Chapter 3, Section 3.2), the BS formula has volatility $\sigma$ is an input and (put or call) option price as the output. Implied volatility uses the inverse formula in the sense that option price quoted by traders in the option exchange market is the input and $\sigma$ is the output.

Hence implied volatility is a theoretical value of volatility of the asset (usually a stock or similar security) price based on an underlying rule (e.g., BS formula) for determining the price of the option. We now list some factors which influence the implied volatility: (1) the price at which the option can be exercised, (2) the riskless rate of return, (3) the maturity date, and (4) the price of the option.

For $S$ the stock price (adjusted for dividends), let $\mu$ = instantaneous expected return of the proportional change in the stock price per unit of time and $\sigma^2$ = instantaneous variance of the proportional change in the stock price. Assume that a geometric Brownian motion (GBM) process gives the dynamics of the stock price:

$$dS = \mu S dt + \sigma S dz, \tag{5.3.1}$$

where $\mu$ is the drift parameter and $\sigma$ is the volatility. The option is called a derivative security since its price is derived from $S$. The drift parameter $\mu$ causes the price of the derivative to depend on the risk-free interest rate.

Let $X$ denote the strike price of a European call option, $r$ the risk-free rate, $C_{\text{BS}}$ the Black-Scholes price of the call option, and $\Phi(.)$ the cumulative density function of the standard normal variable. Then the Black and Scholes (1973) pricing equation for the call and put option is given by

$$\begin{aligned} C_{\text{BS}} &= S\Phi(d) - X\exp\{-r(T-t)\}\Phi(u), \\ P_{\text{bs}} &= X\exp\{-r(T-t)\}\Phi(-u) - S\Phi(-d), \end{aligned} \tag{5.3.2}$$

where

$$d = \ln(S/X) + (r + 0.5\,\sigma^2)(T - t)\{\sigma\sqrt{(T-t)}\}^{-1},$$
$$u = \ln(S/X) - (r + 0.5\,\sigma^2)(T - t)\{\sigma\sqrt{(T-t)}\}^{-1}.$$

See Chapter 8 for a detailed derivation of relevant formulas starting with the diffusion equation and derivation of the underlying BS differential equation for the price of a derivative security $f(S, t)$. The trick is to eliminate randomness and then use Ito's lemma. The formula above is a solution to the partial differential equation and bears a similarity to a solution to the heat equation in theoretical physics.

In the GBM model entire variability is reflected by a surprise, which is a realization of the standard normal, $dz = N(0, 1)$. There is a one-to-one relation between any option's price $C$ and $\sigma$ from (5.3.1). In the context of data analysis, note that data on $C$, $S$, $X$, risk-free rate, and $T$, $t$ are readily available and that the value of $\sigma$ can be uniquely inferred from the formula (5.3.2). The inferred value of $\sigma$ is called *implied volatility*, Campbell et al. (1997, p. 377). Options Clearing Corporation (OCC) owned by the exchanges that trade listed equity options, (*www.888options.com*) defines *implied volatility* as the volatility percentage that produces the "best fit" for all underlying option prices on that underlying stock. Data are available for implied volatility based on both call and put options denoted here as "callv" and "putv" respectively. Clearly, putv represents a forward-looking nonsymmetric option market measure of downside risk.

To illustrate a nonnormal model, consider jump diffusions. This model inserts an additional source of variability in (3.6). Merton (1976) discussed the jump diffusion process with a closed form solution for the call option prices $C_{JD}$ given below. It is a combination of GBM and a Poisson-distributed jump process PO($\lambda$), where $\lambda$ is both the mean and the variance of the number of Poisson jumps. Aït-Sahalia et al. (2001) reject the null hypothesis that S&P 500 options are efficiently priced and use a jump component to partly reconcile the difference.

Let $W = (\sqrt{dt})dz$,

$$\frac{dS}{S} = (\mu - \lambda k)dt + \sigma dW + dv, \tag{5.3.3}$$

where $dv = \Sigma_{i=0}^{i=x} s_i$ and where the lowercase $s_i \sim N(\theta, \delta)$ represents the size of the shock. If the number of jumps were fixed, the term dv will be a sum of a fixed number of normals. The complication in (5.3.3) arises because the number of jumps follows PO($\lambda$). For (5.3.3) the moments of the distribution of total return $\ln(S_T/S_0)$ are denoted here with a subscript JD for jump diffusion. The average size of the shock viewed as a jump is $\theta$ and the variance $\delta$ is yielding:

$\text{Mean}_{JD} = (\mu - 0.5\sigma^2)T$
$(\text{Standard deviation})_{JD} = \sigma_{JD} = [\sigma^2 + \lambda(\theta^2 + \delta^2)]^{0.5}\sqrt{T}$
$\text{Skewness}_{JD} = (\lambda/\sqrt{T})[\theta(\theta^2 + 3\delta^2)/\sigma_{JD}]^{1.5}$
$(\text{Excess kurtosis})_{JD} = (\lambda/T)[(3\delta^4 + 6\delta^2\theta^2 + \theta^4)/\sigma_{JD}]^2$

If the Poisson parameter $\lambda$ is zero, jump diffusion reduces to the simple diffusion with zero skewness and zero excess kurtosis. If the size of the jump shock is negative on an average ($\theta < 0$) the distribution of stock returns is skewed to the left. The call option price of the jump diffusion is a cumulative sum of $PO(\lambda)$ times the Black-Scholes price or

$$C_{JD} = \left\{\sum_{i=0}^{i=\infty} PO(\lambda)\right\} C_{BS}. \tag{5.3.3}$$

Chidambaran et al. (2000, p. 587) discuss (5.3.3) and some useful extensions. It is obvious that jump diffusions permit nonzero skewness and that one can use jump diffusions to indirectly incorporate the asymmetry. However, the fact remains that the pricing formulas (5.3.1) and (5.3.3) contain $\sigma^2$ as measures of volatility that are not directly observed. Merton (1976) discusses this as a source of "pricing error" and "misspecification." He also mentions nonstationarity of variance and studies the pattern and magnitude of errors in option pricing by using a simulation. We note that one source of misspecification is that $\sigma^2$ values symmetrically treat upside potential and downside risk, and that the pricing formula assumes the jumps are entirely nonsystematic. The second source is that underlying distributions, investors' loss functions, and psychological attitudes to the two sides are, in general, not symmetric.

Reagle and Vinod (2000) claim that the implied volatility of the put option (putv) places all the focus on the downside and therefore putv and callv have an asymmetric reaction to downside risk. Thus we have tools to incorporate more realistic asymmetry. Reagle and Vinod (2000) use implied putv for about 100 stocks and other balance sheet variables for the same companies.

For downside risk to play an important role in portfolio selection, there are three important factors that must be true:

1. The distribution of returns is nonsymmetrical.
2. Portfolio choice based on downside risk gives different rankings than portfolio choice based on symmetrical risk.
3. Investors need a higher risk premium for downside risk than for upside risk.

We will explore each of these issues in Chapter 7. First, we need to consider utility theory, which provides a basis for an investor reaction to downside risk.

CHAPTER 6

# Portfolio Valuation and Utility Theory

## 6.1 UTILITY THEORY

In earlier chapters we discussed financial theory in terms of risk measurement, hedging, diversification, and the like, and how the theory might fail. We avoided mention of the utility associated with high returns. Of course, it is implicit that for a given risk level, more is always better, but we have not considered "how much better," and at what cost. We have ignored the attitude of the investor toward risk; that is, some investors may get more upset with a $100 loss than they are happy with a $100 gain. This is called loss aversion, and it belongs to an extension of expected utility theory (EUT), called non-EUT. In short, in earlier chapters we emphasized only those financial theories that apply to everyone. The main appeal of the option-pricing model proposed by Black and Scholes (1973) is that it can apply to everyone because it is preference-free. Similarly various probability distributions $f(x)$ associated with excess returns are treated as if they apply to everyone.

This chapter brings in the utility function $U(x)$. Section 6.1 lays the groundwork with some insights on EUT, and non-EUT is discussed in Section 6.2. Although we have abandoned the aim of making the theory applicable to everyone, we still want to keep the theory applicable to hundreds of thousands of investors, if not millions. Accordingly, Section 6.2 introduces a parameter $\alpha \in [0, 1]$ that measures the extent of departure from EUT precepts, with $\alpha = 1$ suggesting no departure (perfect compliance) and $\alpha \to 0$ suggesting large departure. We customize the theory for given levels of $\alpha$, since everyone cannot be assumed to have $\alpha = 1$, permitting greater realism.

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

Often the bottom line in finance is to be able to compare portfolios, to be able to rank them and eventually choose the best. We assume that the comparison between portfolios can be formulated in terms of a comparison among probability distribution functions (PDFs) of excess returns, $f(x)$, relevant for each portfolio. Section 6.2 develops non-EUT weights to incorporate attitudes to risk (loss aversion) parameterized by $\alpha$. Section 6.3 explains the powerful and elegant concept of "stochastic dominance." It is useful because given some relevant properties of the investor's $U(x)$, whenever a portfolio A "dominates" another portfolio B, the investor obtains more utility in choosing the portfolio A over B.

Stochastic dominance involves the utility function $U(x)$, which is generally unique for each individual. Nevertheless, the theory attempts to satisfy hundreds of thousands of individual investors in admitting only the signs of the first four derivatives of $U(x)$. We will explain the interpretation of the signs of the first four derivative orders of $U(x)$ that lead to the four orders of stochastic dominance.

Section 6.4 discusses what is known about incorporating utility theory into lower partial moments (mean variance and other moments focusing on the downside) and into option pricing. Finally, Section 6.5 discusses forecasting of the all-important future returns. The forecasts will presumably require additional data on a set of relevant regressors, including macroeconomic, international, product-specific, and any other relevant variables to obtain better forecasts of future returns. We discuss nonlinear forecasting equations, mainly neural networks (NN) designed to use all such information. The NN models have been used in finance literature, Abhyankar et al. (1997). For example, in the consumer products, tourism, and entertainment industries, future profits (returns) can be affected by consumer tastes, opinion surveys, fashion ratings, or quality ratings. The success in investing money in these sectors may be directly linked to the success in forecasting future quality ratings. The quality ratings usually involve defining value in terms of categories. For example, four ratings might be "poor, fair, good, and excellent." or they may be in terms of one to five stars as with hotels and restaurants. Section 6.5 also describes some generalized linear models (GLM) for profit forecasts involving dependent variables limited by its number of possible values due to their categorical nature. The NN models are also applicable to categorical variables.

### 6.1.1 Expected Utility Theory (EUT)

Anyone who has gone to a financial planner knows the first question they ask is about preferences. Each investor has preferences about risk, timing of payments, and even the social responsibility of a company. Financial theories, such as CAPM, develop from a preference-free basis that is the same for all investors, regardless of their attitudes toward risk. Such an approach, while mathematically elegant, must allow many assumptions about the world that

by far simplify reality. However, CAPM has been around for over four decades, and not everyone has the same stock portfolio as a result.

Preference-free theories have given invaluable analytical tools for the market in general, but to make those fine-tuning choices for the individual, preferences must be respected and addressed. To help us put a number on our preferences in order to be able to rank our investment choices, we turn to utility theory. Utility theory uses a utility function to evaluate the importance of different levels of consumption of each good (commodity) and different combinations of such goods. In finance, we are fortunate that we generally deal (directly or indirectly) in only one commodity, money. Therefore we can just write utility as a function of returns, $U(r)$ or excess returns $U(x)$. There are some general statements that usually apply to all utility functions:

1. Monotonically increasing. Everyone likes money, so as $r$ increases, all people should have a higher utility level $U(r)$.
2. Diminishing marginal utility (MU). Those first dollars are very important, they pay rent, food, and clothing. But as people earn more, utility of each additional dollar increases by a smaller and smaller amount. We will initially make this assumption, but we will also point out cases where it is debatable.

Choosing the most appropriate functional form of the utility function has occupied researchers for over a century. Philosophers have long discussed hedonism and utilitarianism. In 1854 Gossen wrote the first mathematical utility function $U(x_i, i = 1, 2, \ldots, n)$ as a sum of quadratic polynomials of the quantity of goods $a_i + b_i x_i - c_i x_i^2$ with $b_i > 0$ and $c_i > 0$ for all $i$. Gossen's quadratic utility function satisfies the property of diminishing marginal utility (MU), since the derivative of $U$ with respect to $x_i$ denoted by $U' = \text{MU} = b_i - c_i x_i$ decreases as $x_i$ increases.

In 1881 Edgeworth argued that $U(x)$ should be a joint function of all arguments and not separable. About the same time Pareto questioned the very existence of cardinal utility defined as a numerical quantity representing the utility of anything. He and others argued that interpersonal comparisons of utility are impossible and that at best one can only hope to rank-order the utilities, but not attach numerical values (e.g., utils) suggested by cardinal utility theorists. Moreover, when several goods are involved, there is the proverbial problem of comparing apples and oranges.

If there is no uncertainty in a deterministic world, and the choice is among exactly three outcomes, such as gaining \$10, gaining nothing, or losing \$5, it is a matter of common sense that an economic agent will rank-order the utility from gains to losses as $U(\$10) > U(\$0) > U(-\$5)$, where $>$ denotes higher utility. When a decision involves uncertainty, we find that individuals do not simply look at the expected dollar gain. Orderings of expected utility can differ vastly from the expected value of returns. If it is once-in-a-lifetime type decision (to choose a spouse, to have a heart bypass operation, or to send a retaliatory

nuclear missile on the enemy), it is not clear that probabilistic choice concepts have relevance. However, in matters of finance, probabilistic concepts can provide practical and useful insights for rank-ordering among choices involving uncertain outcomes, called stochastic choice.

Let $x_i$ denote money gained with probability $p_i$. A random variable $X$ representing money gained has realizations denoted by $x_i$. We imagine a probability model called a probability distribution $f(x)$ that assigns a probability $p_i$ for each realized value $x_i$ of the random variable. Expected value or mathematical expectation of $X$ is defined as $E(X) = \Sigma_{i=1}^{n} x_i p_i$. In statistics the expected value is also called the first raw moment or the mean (or average) of the random variable. If $X$ is a continuous random variable, statisticians replace the summation by an integral, with no essential change in the concept.

The expected utility theory (EUT) states that the expected utility of money gained, $E[U(X)]$, may be simply measured by the mathematical expectation. Economists use EUT as a good first approximation for suggesting the appropriate choice under uncertainty. To see this, suppose that an investor has a choice between (a) a 5% return with certainty (think about a Treasury security or an insured CD where the final payoff is known and guaranteed by the government) and (b) a 50% chance of a return of 15% coupled with a 50% chance of a loss of 5% (think about a stock where the payoff could be large or small).

An individual simply looking at the expected outcome will see that both (a) and (b) are the same. Choice (a) provides an expected return of 5% with no uncertainty. Choice (b) provides the expected return of 5% [15%*0.5–5%*0.5] involving two uncertain outcomes. Since the mathematical expectations for both choices are the same ($E$(a) = $E$(b) = 0.05), it is customary to define the choice (b) as a fair gamble.

But most individuals will lean one way or the other between the two choices. This means that something is missing in using the mathematical expectation. It ignores the matter of downside risk, and neglects the fact that the investor choosing (b) will not know the outcome until the end. We saw before, when looking at interest rates in Chapter 1, that the sleepless nights from risky investments must be rewarded for most individuals, and that the downside of a lower return hurts more than the possibility of a higher return benefits.

Samuelson (1983, p. 511) showed, by specifying a nonlinear utility function subject to two axioms, that a perceived risk aversion can be consistent with EUT. Samuelson's (1983, app. C, §5–8) enlarged doctoral dissertation includes an extensive discussion useful to financial economists on the mean-variance model and its limitations, probabilistic choice, and portfolio analysis.

**Axiom 1** (Concavity). $U[E(X)]$ functionals are concave; that is, individuals are risk averse.

The geometry of concave and convex utility functions is presented in Figure 6.1.1 by graphing an individual's utility as a function of his or her wealth. If

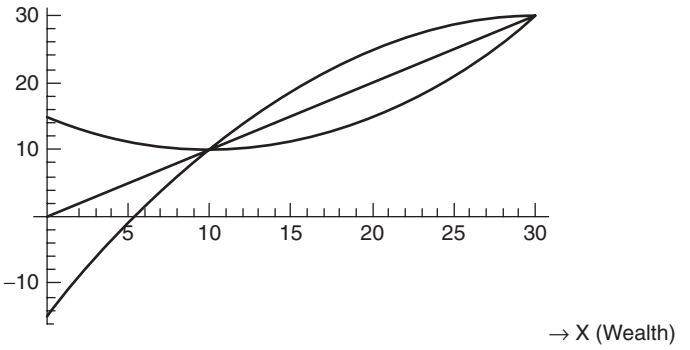



**Figure 6.1.1** Concave (bowed out) and convex (bowed in) utility function $U(X)$

the utility function is concave (bowed out) $U(X)$, we see that as wealth increases, utility increases more gradually. This is again the idea of decreasing marginal utility, but it will have another interpretation in the presence of uncertainty. Looking at our fair gamble, we can see that the 5% return with certainty will always have a higher expected utility than the risky choice. The 15% return does not increase utility as much as the 5% loss decreases utility, so the average utility will be lower than U(wealth +5%). This defines a risk averse individual satisfying the inequality $U[E(X)] > E[U(X)]$, where the effect of randomness is less severe on the left side where utility is computed after the expectation is evaluated.

Any concave curve has the definitional property (due to bowing out) that the points in the middle lie higher than a points along the the straight line joining the extreme points. Common examples of risk averse utility functions are quadratic utility functions with decreasing marginal utility, and log utility functions.

While risk aversion explains why the typical investor would require a risk premium, subsets of investors may not be as susceptible to risk. We define risk neutral individuals as those who are indifferent between a fair gamble and the equivalent certain payoff, that is, satisfying $U[E(X)] = E[U(X)]$. They do not expect any premium for certainty, and their utility function is along the 45 degree line in Figure 6.1.1. Many theoretical assertions in economics and choice theory assume risk neutrality. Institutional investors and corporations who are investing over a long time period can afford to be risk neutral, since over time, the highs and lows will average out and also investing for others can allow greater objectivity. Maximization of expected value is justified under the risk neutrality assumption.

We define risk loving individuals as those who prefer a fair gamble rather than the equivalent certain payoff, that is, satisfying $U[E(X)] < E[U(X)]$. They are willing to pay a premium for privilege of gambling, rather than settle for certainty along the 45 degree line. Geometrically their utility function is convex which has lower utility value for middle points. Bernoulli suggested

that convex utility cannot occur, but in the stock market certain situations of limited liability can cause investors to act risk loving and take extreme gambles since they are not responsible for the downside.

$$\begin{aligned} U[E(X)] &< E[U(X)] \quad \text{for risk loving,} \\ U[E(X)] &= E[U(X)] \quad \text{for risk neutral,} \\ U[E(X)] &> E[U(X)] \quad \text{for risk averse.} \end{aligned} \tag{6.1.1}$$

**Axiom 2** (Independence). If the utility of $X$ is no worse than that of $Y$, $U(X) \geq U(Y)$, there is a fraction $1 > q > 0$, and an additional choice $Z$, then $U[qX + (1 - q)Z] \geq U[qY + (1 - q)Z]$. If $X$ is equivalent to $Y$ in utility, $U(X) = U(Y)$, then $U(X) = U[qX + (1 - q)Y]$.

Axiom 2 ensures that the combinations of risky investments do not affect the expected utility of each. Otherwise, an investor would have to continually revalue the same stock depending on what other stocks are held in the portfolio.

For example, the utility function $U(X) = \ln X$ matches Samuelson's axioms. The marginal utility MU = $\partial U(X)/\partial X = (1/X)$ is diminishing, since as $X$ increases, MU diminishes. Therefore the function is concave, and evaluating the utility regardless of other choices assumes independence. We will start at a wealth of \$1 M. Taking natural logarithms, the starting utility is ln(1) = 0. Now let us move to our choice. The safe investment (a) would give a final wealth of \$1.05 M, or a utility of ln(1.05) = 0.0488. The risky investment (b) has a 50% chance of a utility of ln(1.15) = 0.1398 and a 50% chance of a utility ln(.95) = –0.0513. Choice (b) gives an expected utility of $E(U) = 0.5(0.1398) + 0.5(-0.0513) = 0.0443$. Since $0.0488 > 0.0443$, the commonsense choice in favor of the choice (a) can be justified simply by choosing a (nonlinear concave) logarithmic functional form of the utility function.

Von Neumann and Morgenstern (1953) use a clever way to elicit the entire utility curve from knowing the rank order of three or more choices, conveniently denoted by $C(15) > C(5) > C(-5)$. Here the numbers after $C$ do not necessarily measure the value of the choice, but are chosen to mimic the example given above. The Von Neumann-Morgenstern idea is to use EUT to determine the probability necessary to make the individual indifferent between the choice $C(5)$, on the one hand, and the uncertain outcomes similar to 50% chance of both $C(15)$ and $C(-5)$ with the expected return of 5, on the other. Remember, we only know the rank order not the numerical utility. By offering agents sets of gambles, one can elicit how much more they prefer $C(15)$ over $C(5)$, and so on. In finance we already have money values to compare, and we do not need such somewhat impractical but intellectually elegant theoretical arguments to convert ordinal measurements into cardinal numbers.

Friedman and Savage (1948) showed that human behavior includes both risk aversion and risk seeking at the same time. Markowitz wrote two impor-

tant papers in 1952, one of which (1952b) deals with portfolio selection and laid the foundations of modern finance. Markowitz (1952a) clarified that Friedman and Savage framework should not be applied in terms of absolute levels of wealth, but should refer to the status quo level fo wealth and the risky decisions should be viewed in terms of changes to the status quo level. More recent prospect theory discussed later in this chapter is an extension of these theories and does evaluate the choices with reference to the status quo, as urged by Markowitz.

### 6.1.2 A Digression: Derivation of Arrrow-Pratt Coefficient of Absolute Risk Aversion (CARA)

This subsection considers a derivation of a coefficient of absolute risk aversion (CARA). If we know the functional form of the utility function $U(X)$, we can compute this measure. Most investors are risk averse in some sense. Is there a formal measure of risk aversion related to the utility function? Some readers may consider an answer to this question to be a digression and skip this subsection. However, we remind the reader that CARA will be relevant in Section 6.3.5, which lays the groundwork for fourth-order stochastic dominance (4SD).

Consider a random variable $Z \sim \mathrm{N}(0, 1)$ similar to a unit normal, such that $E(Z) = 0$ and variance is unity. Thus $Z$ represents a zero expectation gamble, which yields on an average no gain or loss and $X$ represents current wealth. Note that $Z$ can be negative and we are offering the choice between a certain payoff of $X$ similar to choice (a) above and an uncertain payoff of $X + Z$, a fair gamble similar to choice (b) above.

The Risk Premium is the amount of side-payment $\pi$ necessary to make the risk averse individual treat the fair gamble as equivalent to the certain outcome. Thus $\pi(X, Z)$ depends on the current wealth and the zero expectation random variable $Z$. It satisfies the following equality: $E[U(X + Z)] + \pi = U(X)$. That is, the risk premium $\pi$ is positive, if the following inequality is satisfied: $U(X) - E[U(X + Z)] > 0$. Since individuals are assumed to prefer larger income, $U(X)$ is an increasing function of $X$, and we have $U'(X) > 0$, denoting derivatives by primes. Note that any inequality remains valid if we multiply both sides by a positive quantity. Hence we can divide both sides of the inequality by the positive quantity $U'(X)$ to yield the condition for positive risk premium to be

$$\frac{U(X)}{U'(X)} > \frac{E[U(X+Z)]}{U'(X)}. \tag{6.1.2}$$

Now we want to write the right side of (6.1.2) in such a way that the random part $Z$ in $U(X + Z)$ is separated from the nonrandom part $X$. Accordingly, a Taylor series expansion of $U(X + \mathrm{Z})$ yields

$$U(X+Z)=U(X)+ZU'(X)+0.5Z^2U''(X)+\text{higher order terms.} \quad (6.1.3)$$

Dividing both sides of (6.1.3) by $U'(X)$, we have

$$\frac{U(X+Z)}{U'(X)}=\frac{U(X)}{U'(X)}+Z+\frac{0.5Z^2U''(X)}{U'(X)}+\text{higher order terms.}$$

Then we take expectations of both sides and assume that $Z$ is "small," $|Z| < 1$, so that higher order terms involving $Z^3$ can be ignored. Recall that $X$ is not a random variable only $Z$ is random. So the expectation will affect only the terms involving $Z$. Since $E(Z) = 0$, $Z$ will disappear, and since $E(Z^2) = 1$ (variance of $Z$ is 1) is known, $Z^2$ is replaced by unity. Thus we have

$$\frac{E[U(X+Z)]}{U'(X)}=\frac{U(X)}{U'(X)}+\frac{0.5U''(X)}{U'(X)}. \quad (6.1.4)$$

Substituting this into (6.1.2), the condition for positive risk premium becomes $U(X)/U'(X) > U(X)/U'(X) + 0.5\ U''(X)/U'(X)$. Note that the first term on the right side of this inequality is the same as the left side of the inequality. Hence the condition becomes $0 > 0.5\ U''(X)/U'(X)$; that is, $0 > U''(X)$ since $U'(X) > 0$. So far we have assumed for simplicity that $Z$ is N(0, 1). More generally, we can permit the variance of $Z$ to be any suitable positive number, and still the inequality will be essentially unchanged. Hence the sufficient condition that a person with the utility function $U(X)$ will demand a positive risk premium (bribe) before he is willing to choose a risky alternative (portfolio) is

$$\text{CARA}=\frac{-U''(X)}{U'(X)}>0. \quad (6.1.5)$$

Condition (6.1.5) holds true for any $Z$ random variable with zero mean. Thus we have provided a simple derivation of Arrow-Pratt coefficient of (absolute) risk aversion as CARA = $[-U''(X)/U'(X)]$, which should be positive for risk averse individuals. The adjective "absolute" is used to distinguish it from CRRA = $[-XU''/U']$, which inserts $X$ in the numerator and denotes a coefficient of relative risk aversion. It is *relative* to the level of wealth X.

### 6.1.3 Size of the Risk Premium Needed to Encourage Risky Equity Investments

Expected utility theory gives some important insights about how individuals will react to uncertainty, but there are some situations in finance where the theory does not transfer readily into practice. If one studies long-term data on net returns from equity investments in United States and compares it to net

returns from fixed-income securities, it is clear that equity investors enjoy excess returns. Our earlier discussion of risk premium suggests that society must pay some risk premium for investing in risky assets (equities) that ultimately lead to growth and prosperity. However, one can question the size and long-term persistence of risk premia. This is the "equity premium puzzle" by Mehra and Prescott (1985). The size and persistence of equity premium cannot be explained, even if one follows the dictates of EUT and adjusts for risk. Explanations of the puzzle rely more on loss aversion and myopia than a rational choice by a risk-averse investor.

If the intersection point of the three curves in Figure 6.1.1 is placed at the status quo point, loss aversion means that the curve will be convex to the left of the status quo point. This means that the curve itself moves as an individual moves to different wealth levels. Starmer (2000) cites much experimental evidence that rejects EUT classified under three categories of inconsistent human behavior:

1. Violation of monotonicity is a catchall phrase to say that agents can make stupid choices if they are not aware of their stupidity. Starmer explains with the help of an example that 100% of the agents are right when faced with a simplified choice. However, if the same option is stated in a less obvious, complicated manner, they fail to choose the correct option 58% of times.

2. Splitting of good attributes into multiple subattributes can sometimes encourage humans to prefer the good choice a bit more strongly. For example, exposure to ultraviolet radiation (UVR) causes a generic skin cancer. The same risk appears worse to humans in lab experiments if the generic cancer is split into many different skin cancers, even if each bears a lower probability, and even if the aggregate probability is the same.

3. Transitivity of choices means that if a preference for choice A over B is stated as A > B and if we know that B > C, then transitivity requires that the individual must choose A > C. Violations of transitivity are indeed observed in experiments, such as when agents "regret" that they did not choose something. The behavior of agents in financial markets is also subject to similar violations. For example, agents may regret that they did not buy a stock in the past, and despite changed circumstances end up buying that stock even if it is obviously inferior to some other stock in the new situation. We conclude that non-EUT models cannot be ignored in finance.

### 6.1.4 Taylor Series Links EUT, Moments of $f(x)$, and Derivatives of $U(x)$

The phrase expected utility in the name EUT erroneously suggests that it might be concerned only with the expected value or the mean of the underlying probability distribution $f(x)$. The aim of this subsection is to show that EUT encompasses variance, skewness, and kurtosis (higher order moments)

also. This section also establishes a link between first four moments of $f(x)$ and the first four derivatives of the utility function. This link is relevant for stochastic dominance discussed in Section 6.3.

Let our utility function be viewed in terms of realizations of the random variable $X = x$ be written as $U(x)$, and let its derivatives be denoted by primes, as before. Let the probability distribution $f(x)$ be summarized not only by its expected value or mean $\bar{x}$, but also by variance $s^2$, and further $k$th sample moments around the mean denoted by $m_k$.

Now let us expand $U(x)$ around $\bar{x}$ by Taylor series and evaluate the expectation of both sides as

$$E(U(x)) = E(U(\bar{x})) + E(x-\bar{x})U' + \frac{E(x-\bar{x})^2 U''}{2!} + \frac{E(x-\bar{x})^3 U'''}{3!} + \ldots. \quad (6.1.6)$$

By definition, the second, and third moments of $f(x)$ involve $s^2 = E(x - \bar{x})^2$ for the variance and $m_3 = E(x - \bar{x})^3$, which in turn yields the skewness. Substituting these definitions in (6.1.6) yields the following elegant result from EUT:

$$E(U(x)) = U(\bar{x}) + \frac{s^2 U''}{2!} + \frac{m_3 U'''}{3!} + \frac{m_4 U''''}{4!} \ldots, \quad (6.1.7)$$

where the partials are evaluated at the sample mean. The key assumption for validity of Taylor series is $|x - \bar{x}| < 1$, which is rarely mentioned. Under ordinal utility theory only the ordering, not magnitudes of $x$ are relevant. Then the units of measurement of $x$ can be chosen to satisfy the condition $|x - \bar{x}| < 1$ for convergence of Taylor series. However, in portfolio theory, the magnitudes of $x$ are very relevant and the assumption is not generally satisfied.

Vinod (2001) discusses the following remarks. Equation (6.1.6) with $U' > 0$ states that the expected utility never becomes small when $\bar{x}$ becomes large. When we assume that $U'' < 0$, the contribution of the variance $s^2$ to $E(U(x))$ is negative, which makes intuitive sense, since high variance similar to risk is undesirable. If $U''' > 0$ in (6.1.6), then the contribution of $m_3$ to the expected utility has the same sign as $m_3$. Note that negatively skewed $f(x)$ have bigger left tail (losses) compared to the right tail (profits). In other words, $U''' > 0$ leads to a preference for positive skewness. Similarly $U'''' < 0$ leads to a preference for less probability mass in the tails (reject fat tails or prefer high peaked density or high kurtosis). We remind the reader that high powers cause poor sampling properties of higher moment estimates and that (6.1.6) is subject to $|x - \bar{x}| < 1$, which need not be true for our definition of $x$ as excess return over changing risk-free rate, since the absolute deviation may well exceed unity in the observed data.

## 6.2 NONEXPECTED UTILITY THEORY

In this section we are concerned with allowance for psychological human reaction to the downside risk in asset markets. We mentioned in Section 6.1.2 some failures of the expected utility theory (EUT) and that a realistic approach to reaction of human agents to market conditions should not ignore non-EUT behavior. Since the literature on non-EUT behavior is vast, this section will focus on those aspects of non-EUT, which can be readily handled. We will explain later that we can handle four desirable properties (4DP), including reflectivity, asymmetric reaction, loss aversion, and diminishing sensitivity. Of course, different individuals will exhibit different levels of departure from EUT. Hence any model for incorporating such behavior should be flexible enough to recognize such individual differences. In this section we develop a parameter $\alpha$ to measure the departure from EUT. The development of this apparatus is a tall order.

An investor uses past data on past excess returns to construct an ex ante distribution of excess returns $f(x)$. Chapter 4 considers a variety of parametric and nonparametric distributions $f(x)$, including $N(\mu, \sigma^2)$ and empirical CDF. The past data cannot truly represent the unknown future shape of $f(x)$, where the left side represents the potential losses. The psychology of loss aversion of agents means that their attitude to losses is different from their attitude to potential gains. Then, in terms of utility theory the utility of $x$, $U(x)$ on the gain side or the right-hand side of $f(x)$ is not matched by the disutility on the loss side. In our context, relevance of non-EUT is primarily due to loss aversion. Since the range $x \in (-\infty, \infty)$ for $f(x)$ is inconvenient, we will work with the empirical cumulative distribution function (eCDF), whereby the cumulative probability $p$ is in the $[0, 1]$ range. We will later place the eCDF of a market agent perfectly satisfying EUT on the horizontal axis with the EUT compliance parameter $\alpha = 1$. Section 6.2.2 will discuss construction of a weight function for incorporating the lessons of non-EUT. In our application in finance a mapping from an observed eCDF defined over the $[0, 1]$ range to a similar range is almost natural. However, the EUT researchers learned this mapping from Lorenz curves dealing with income distributions. It is perhaps instructive for historical reasons to consider Lorenz curve, although some readers may consider the details regarding the Lorenz curve as a distraction and may wish to skip the following subsection.

### 6.2.1 A Digression: Lorenz Curve Scaling over the Unit Square

Lorenz curves and Gini coefficients were initially developed to measure the degree of income inequality in a society. The proportion of the total population satisfying a certain characteristic is a well-known concept. In statistics these are called cumulative probabilities, which can be compared across income distributions for different societies and over time. In particular, one can compute the following five percentiles of any distribution at percent values

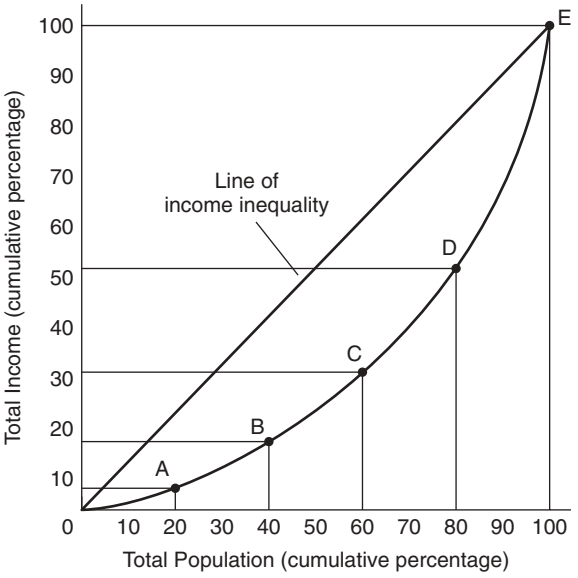



**Figure 6.2.1** Lorenz curve

of 20, 40, 60, 80, and 100, called quintiles. The important insight of Lorenz was to construct similar quintiles on both the population and income sides. Lorenz considered the proportion of "total income" enjoyed by the poorest 20%, 40%. 60%, 80%, and the richest 20% individuals and plotted the income quintiles on the vertical axis and the corresponding population quintiles on the horizontal axis.

Lorenz argued that if everyone enjoyed exactly the same income, all quintiles would fall along the 45 degree line in Figure 6.2.1. Assume that the poorest 20% enjoys only 7% of the income, and represent it by the point $A$. Similarly $B$ to $D$ represent 18, 32, and 52 percent values. The curve passing through the origin and points $A$ to $E$ for the five quintiles is called the Lorenz curve.

The Gini coefficient is defined as twice the area between the 45 degree line and the Lorenz curve in Figure 6.2.1. Clearly, the higher the Gini coefficient, the greater is income inequality. We will now try to represent the departure from EUT in terms of deviation from a similar 45 degree line. For greater generality, we will work with arbitrary quantiles rather than just five quintiles.

We begin with a formal definition of a quantile. Given a number $p \in [0, 1]$ a "quantile of $x$ of order $p$," or $(100p)$-th percentile of the distribution is defined as

$$x(p) = \inf\{x | F(x) \geq p\}. \tag{6.2.1}$$

As we explained in Chapter 2, a quantile is simply the inverse of the CDF (or inverse of empirical CDF). If $p = 0.10$, then quantile $x(p)$ is such that the CDF evaluated at $x$ is 0.10. In (6.2.1), the same is stated in the equivalent infimum notation that the corresponding quantile is the smallest value of the variable $x$ so that CDF evaluated at $x$ exceeds $p = 0.10$ and satisfies two probability relations: $Pr(x \leq x(p)) \geq p$ and $Pr(x \geq x(p)) \leq 1 - p$.

Figure 6.2.1 reveals that the Lorenz curve is the graph of one cumulative probability against another or one percentile against another. It ranges in the [0, 1] or [0, 100] interval along both axes. The Lorenz curve plot is preferred to quantile–quantile plot of Section 4.3.1, since the latter does not have such a fixed range. To draw the Lorenz curve in Figure 6.2.1, we fix the horizontal coordinate at five {20, 40, 60, 80, 100} percent values for a country's population. We let $X$ be the income variable, and we simply compute the vertical coordinates of the Lorenz curve from cumulative probabilities of the distribution of incomes.

If income follows the Pareto distribution (see Section 4.4.1) defined by $f(x) = \theta\ a^{\theta}x^{-(\theta+1)}$, with parameters $a$ and $\theta$, its expected value or average income is $E(X) = a\theta/(\theta - 1)$. The cumulative distribution function (CDF) of the Pareto is $F_{\text{pto}}(x) = 1 - (a/x)^{\theta}$. Given an income $x$, and the values of the two parameters, $F_{\text{pto}}(x) = p$ gives the proportion $p$ of persons earning less than or equal to $x$. If we are given $p = 0.20$, say, the same relation can be used to get the income as $x_{0.20} = F^{-1}_{\text{pto}}(p = 0.20)$ by way of the inverse of the CDF. We want to find $L(p)$, the proportion of "total income" enjoyed by the poorest 20%. It can be shown that $L(p)$ is the integral of $f(x)$ for values of $x$ from 0 to $x_{0.20}$. For the Pareto density, the vertical coordinate of the Lorenz curve $L(p)$ is analytically known by using integral calculus and algebra for any proportion p: $L(p) = 1 - (1 - p)^{[1-(1/\theta)]}$. Note that $L(p)$ depends only on the $\theta$ parameter and the parameter $a$ is absent in the $L(p)$ expression. The Gini coefficient, is twice the area between the 45 degree line. The Pareto density has a simple analytic expression for Gini coefficient given by $1/(2\theta - 1)$. The most useful idea from the Lorenz curve for our purposes is the advantage of working with two CDFs on the two axes.

### 6.2.2 Mapping from EUT to Non–EUT within the Unit Square

In this subsection we apply the mapping of one empirical cumulative distribution function (eCDF) on another, defined on a unit square, first used for the Lorenz curve. This subsection develops decision weights as a useful tool developed by researchers in non-EUT to make non-EUT results both practical and measurable. As we stated earlier, we place on the horizontal axis the cumulative probabilities $p$ when the agent perfectly follows the EUT norm ($\alpha = 1$). We place on the vertical axis the decision weights function denoted by $W(p, \alpha)$, which yields modified values of cumulative probabilities $p$, given the value of a parameter $\alpha$ that measures the extent of departure from the EUT norms.

We let $\alpha = 0$ denote a perfectly noncompliant agent, which is one who does not behave according to EUT in any sense.

Recall from Chapters 2 and 5 that certain indexes are useful for ranking portfolios, called CAPM beta, Sharpe, and Treynor performance measures. Once we develop decision weights, they can be used for developing these indexes for agents who do not behave according to the norms of the EUT. For example, if $\alpha$ for an agent is close to zero, she is extremely loss averse. The portfolio choice for her can then depend on the weighted estimates of CAPM beta, Sharpe, and Treynor performance measures.

Although we follow Lorenz in restricting the mapping to the range [0, 1] on both axes, the shape of $W(p, \alpha)$ function is not at all like that of the Lorenz curve. Once we restrict the range, it turns out that a great deal of the non-EUT literature is conveniently summarized by requiring that the mapping from $p$ to $W(p, \alpha)$ should satisfy four desirable properties as explained later.

Let $T$ denote the number of observations on $x_t$ for $t = 1, 2, \ldots, T$, the excess return from investing in a risky asset. If we rank-order $x_t$ values from the smallest to the largest, we have the order statistics $x_{(t)}$. The probability of each order statistic is $(1/T)$ and the empirical cumulative probability goes from $(1/T)$ to 1 in increments of $(1/T)$. Thus we have the cumulative probabilities as $p_t = 1/T$, $2/T, \ldots, 1$ on the horizontal axis. Note that these are in the range [0, 1]. See Figure 4.3.2 for a plot of eCDF.

Although these $p_t$ are not the proportions 0.20, 0.40, etc., used by Lorenz, his insight of using the same proportions for the vertical axis can be used in our application to evaluate some realistic modifications designed to generalize the EUT. Having placed the observable empirical CDF $p_t$ on the horizontal axis, we now place the perceived utility weight $W(p)$ associated with that $p_t$ on the vertical axis, as in Figure 6.2.2. Now we introduce a single parameter $\alpha$ to measure the extent of departure from EUT. As with Lorenz curve, we will use the 45 degree line to represent zero departure from perfect compliance with the EUT. That is, we assign points along the 45 degree line to $\alpha = 1$ for perfectly rational individuals, who are not fooled by departures from EUT.

Assume that we can ignore the joint distribution of portfolios being compared and focus on only one portfolio at a time. Prelec (1998) defines the following simple, powerful, and flexible function. Its slope is $\alpha$ at its inflexion point. When $\alpha = 1$ for the perfect EUT-compliant case, the slope is unity at the point of inflexion. Of course, this is the slope of any 45 degree line:

$$W(p,\alpha) = \exp\left(-(-\ln p)^{\alpha}\right), \quad \text{where } 0 < \alpha < 1. \tag{6.2.2}$$

If $\alpha \to 0$, we have a step function that is flat (zero slope) everywhere except at the end points, implying maximum departure from EUT. If $\alpha \to 1$, we have $W(p, \alpha) = p$, or perfect compliance with EUT. The smaller the $\alpha$, the greater is the departure from the expected utility theory. Hence we can indeed refer to $\alpha$ as EUT-compliance parameter. Starmer (2000) notes that the use of cumulative $p$ is useful in avoiding violations of monotonicity and transitivity

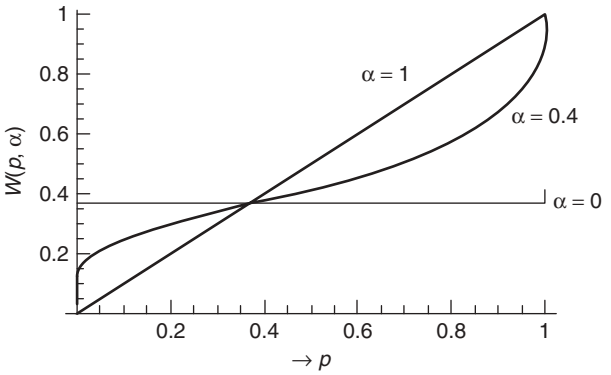



**Figure 6.2.2** Plot of $W(p, \alpha) = \exp(-(-\ln p)^{\alpha})$ of (6.2.3) for $\alpha = 0.001$(nearly flat line), $\alpha = 0.4$ (curved line), and $\alpha = 0.999$ (nearly the 45 degree line, nearly EUT compliant)

axioms. Thus $W(p, \alpha)$, or $W(p)$ for short, represents the perceived utility importance of that $p$.

Prelec (1998) provides a summary of the current knowledge about $W(p)$ in a graph of $W(p)$, against $p$ on the horizontal axis, similar to Figure 6.2.2. A realistic utility function (as indicated by experimental results on risk attitudes) leads to the following four desirable properties (4DPs) satisfied by (6.2.2): Regressive, asymmetry, inverted S-shaped, and reflective utility.

1. Regressive property means that $W(p)$ should intersect the 45 degree line from above. Then the weight $W(p)$ is high when $p$ is small. For example, eagerness to buy fire insurance against a low probability event (house burning) would result in $p$'s gathering to the left of the 45 degree line. Here risk aversion for small $p$ losses means $W(p)$ is much larger than $p$ itself.
2. Asymmetry property means that a fixed point $p = W(p)$ should occur at $p = 1/3$, and not at $p = 0.5$, the midpoint of the [0, 1] range. This means agents prefer positive skewness of $f(x)$ and start downweighting beyond $p = 1/3$. In Figure 6.2.2 each $W(p, \alpha)$ intersects the 45 degree line at $p = (1/e) = 0.37$ (which is assumed to be close enough to 1/3).
3. Inverted S-shaped property means that $W(p)$ is concave near the origin and then becomes convex. It is concave in gains and steeply convex in losses as shown by Starmer (2000). Kahneman and Tversky (1992) interpret this shape restriction to mean two characteristics: (a) "diminishing sensitivity" $u''(x) \leq 0$ for $x \geq 0$, and $u''(x) \geq 0$ for $x \leq 0$, and (b) "loss aversion" implying $u'(x) < u'(-x)$ for $x \geq 0$.
4. Reflective utility property means that low probability events (gains or losses) are treated similarly.

First, we arrange the dynamic excess return data $x_t$ for $t = 1, 2, \ldots, T$ in an increasing order of magnitude and denote the order statistics by $x_{(1)}$ to $x_{(T)}$. Next, we transform $W(p, \alpha)$ into decision weights $w_t(\alpha, p_t)$ applied to a particular observation as shown below. Since we seek weights on order statistics $x_{(t)}$, we "first-difference" the weights in (6.2.2). For any given $\alpha$, the appropriate weights for the ordered value $x_{(t)}$ are

$$w_t(\alpha, p_t) = \exp\left(-(-\ln p_t)^{\alpha}\right) \quad \text{only if } t = 1 \text{ and}$$
$$w_t(\alpha, p_t) = \exp\left(-(-\ln p_t)^{\alpha}\right) - \exp\left(-(-\ln p_{t-1})^{\alpha}\right) \quad \text{for } t > 1, \qquad (6.2.3)$$

where lower case $w_t$ is used and where $w_t$ for the first observation $t = 1$, is different from the $w_t$ for all others. We need to use (6.2.3) only once for each choice of $T$ to get the weights $w_t$. What is the interpretation of these weights? Note that when $T = 5$, all five weights for an EUT-compliant person having $\alpha = 1$ equal 0.2 (= 1/5). However, if $\alpha = 0.1$, the weights are (0.350, 0.0207, 0.0215, 0.0303, 0.577); if $\alpha = 0.5$, the weights become (0.281, 0.103, 0.105, 0.134, 0.376); and if $\alpha = 0.9$, the weights are (0.216, 0.181, 0.182, 0.193, 0.228). The smallest value (= $x_{(1)}$, possibly a loss) of $x_t$ is weighted more than the middle values. Also the largest value (= $x_{(T)}$, possibly a big gain) is weighted by a number larger than $1/T$.

Given $\alpha \in (0, 1)$, the weights $w_t$ are readily computed and then used in various summary measures discussed in the next section. Conversely, given sufficient data from the past behavior of an individual, we can estimate the implicit $\alpha$ appropriate for that individual. An individual with a small $\alpha$ adheres less closely to the norms of the expected utility theory, and is less EUT compliant. Vinod (2001) shows with a simple example that the rankings of portfolios are not too sensitive to the choice of $\alpha$. We will use an example from Vinod (2004) to compare 1281 international mutual funds and show how we can help an investor focus on the best five funds and how the choice is affected by the choice of $\alpha = 0.1, 0.5, 1$.

Lattimore, Baker, and Witte (1992) suggest a weight function with two parameters

$$W(p_t, \alpha, \beta) = \frac{\alpha(p_t)^{\beta}}{\alpha(p_t)^{\beta} \sum_{k=1, k \neq t}^{T} \alpha(p_t)^{\beta}}. \qquad (6.2.4)$$

The EUT is a special case where $\alpha = \beta = 1$. When $\beta < 1$, the curve (6.2.4) is shaped as inverted S, and when $\alpha < 1$, the weights do not add up to unity and are called prospect pessimism. We claim that the weights in (6.2.3) or (6.2.4) are important tools in solving the portfolio selection problem in a more flexible and realistic setting of non-EUT compliant agents. It is appealing to be able to measure EUT compliance by $\alpha$ in (6.2.3).

For a discussion of the non-EUT literature and prospect theory, see Kahneman and Tversky (1979), and Prelec (1998), among others, surveyed

in Starmer (2000). Vinod (2001) proposed weighted estimates of Sharpe and Treynor type performance measures used in portfolio selection. This is useful for the customization of a portfolio selection to suit the attitudes toward risk of a particular client of a brokerage firm depending on his $\alpha$, such that if $\alpha$ is close to unity, the agent is close to perfect compliance with the norms of EUT.

## 6.3 INCORPORATING UTILITY THEORY INTO RISK MEASUREMENT AND STOCHASTIC DOMINANCE

Thus far utility theory has only been able to make choices by inserting possible outcomes into the investor's utility function and ordering the expected utility. This gives us some insight on how individual investors react to risk, but it is of limited value for those of us who have not sat down and teased out all the parameters of our own personal utility function. The purpose of this section is to review the insights from stochastic dominance literature, which can give some straightforward rules for portfolio choice that apply to general classes of investors. We explain how to conclude that a particular portfolio A is superior to another portfolio B (A dominates B) by knowing some general characteristics of the utility functions $U(x)$ of the agents (buyers) of these portfolios. The beauty of this theory is that we do not need to know the form of the actual parametric utility functions of agents. We simply need to know the signs of certain derivatives of utility functions, such as $U' > 0$, $U'' < 0$, $U''' > 0$, and $U'''' < 0$, where the number of primes measures the order of derivatives.

### 6.3.1 Class D1 of Utility Functions and Investors

The utility functions in this class require that the marginal utility for investors be positive, that is, $U' > 0$. This means that the investor always gets a higher utility by having extra $x$ (more money). As long as the investor is insatiable with respect to wealth, he fits into this category.

### 6.3.2 Class D2 of Utility Functions and Investors

The class D2 is more restrictive than class D1. Here the $U(x)$ satisfies two inequalities on the first two derivatives of $U(x)$. This class requires that the marginal utility (MU) evaluated at higher levels of its argument (wealth) should be smaller than when evaluated at lower levels; that is, MU for the rich is lower. The requirement is that the marginal utility $U'$ be not only positive but also decrease as $x$ increases: $(d/dx)\ U' < 0$ simply means that $U'' < 0$. The class D2 satisfies the "law of diminishing MU" in economics texts.

Investors in this class get less of an increase in utility as wealth gets higher, leading to aversion to variance, or risk aversion. Those first few dollars give a

big boost in utility. But after that, further extra money, while still wanted, adds less to the enjoyment. Therefore in a risky venture a large increase will definitely increase the enjoyment only if it comes without an increase in volatility or variance. Agents whose $U(x) \in$ D2 are usually called risk averse.

### 6.3.3 Explicit Utility Functions and Arrow-Pratt Measures of Risk Aversion

In this section we consider examples of $U(x)$ that are popular among economic theorists. Logarithmic utility says that $U(x) = \log x$, therefore the satisfaction or utility of $\$x$ does not grow linearly but more slowly as the log of $x$ grows. Since $x > 0$ is a reasonable assumption and since the derivatives of $\log x$ are well known, we know the signs of the first four derivatives of the logarithmic utility function as $U' = (1/x) > 0$, $U'' = (-1/x^2) < 0$, $U''' = 2x^{-3} > 0$, and $U'''' = -6x^{-4} < 0$. The signs of all these derivatives are correct and consistent with economic theory.

Another popular utility function is called the power utility, $U(x) = x^{1-\gamma}/(1-\gamma)$, where the power is fractional or negative since $\gamma > 0$. Again, the derivatives are well known, and we have $U' = x^{-\gamma} > 0$, $U'' = -\gamma x^{-\gamma-1} < 0$, $U''' = \gamma(\gamma+1)x^{-\gamma-2} > 0$, and $U'''' = -\gamma(\gamma+1)(\gamma+2)x^{-\gamma-3} < 0$. Again, all signs are as desired.

Recall from (6.1.5) that the sufficient condition that a person with the utility function $U(X)$ will demand a positive risk premium (bribe) before he is willing to choose a risky alternative (portfolio) is that the Arrow-Pratt coefficient of absolute risk aversion CARA $= -U''/U'$ is positive. For the logarithmic utility, CARA $= (1/x^2)/(1/x) = 1/x$ is obviously positive, since $x > 0$, and decreases as $x$ increases. In economics white bread is considered an inferior good in the sense that as the income of the individual increases, she prefers a better bread or cake. Arrow (1971) showed that risky assets are not inferior goods and investors continue to buy them even if their income increases. He further showed that $U(x)$ must satisfy the condition of nonincreasing a*bsolute* risk aversion (NIARA) in order to prevent a good like $x$ from becoming inferior. Thus the logarithmic utility function satisfies NIARA.

The coefficient of *relative* risk aversion (CRRA) is simply $x$ times CARA. It is unity for the logarithmic utility function. For the power utility function, CARA $= (\gamma x^{-\gamma-1})/(x^{-\gamma}) = \gamma/x$, which likewise is both positive and decreases as $x$ increases. Thus power utility satisfies NIARA property. Now, by definition, CRRA $= x *$ CARA $= \gamma$. Thus CRRA is constant for both the logarithmic and power utility functions.

If we do not know the functional form of $U(x)$ as logarithmic or power, how can we ensure that the NIARA property holds true? We want the CARA to decrease as $x$ increases without knowing the form of $U(x)$. That is, we want $(d/dx)$ CARA $< 0$, where CARA $= -U''/U'$. A necessary condition for NIARA then is that $U''' > (U'')^2/U'$.

### 6.3.4 Class D3 of Utility Functions and Investors

The class D3 is more restrictive than both classes D1 and D2. Here the $U(x)$ satisfies three inequalities involving the first three derivatives of $U(x)$. The D3 class requires not only that the MU evaluated at higher levels of its argument (wealth) should decline, but also that the decline should become more pronounced. We saw in the previous subsection that the desirable NIARA property is satisfied if $U''' > (U'')^2/U'$ holds true. If $U(x)$ belongs to class D2, we know that $U' > 0$ and $U'' < 0$, that is, $(U'')^2/U' > 0$. After substituting this on the right-hand side of the inequality for NIARA, note that the third inequality involving the third derivative is simply $U''' > 0$. Thus the necessary requirement for a $U(x)$ to belong to class D3 is that all three inequalities: $U' > 0$, $U'' < 0$, and $U''' > 0$ are satisfied.

### 6.3.5 Class D4 of Utility Functions and Investors

The class D4 of utility functions is more restrictive than the class D3 of utility functions. Here the $U(x)$ satisfies four inequalities involving the first four derivatives of $U(x)$. Where does the fourth derivative come from? What is its interpretation?

Kimball (1990) notes that Arrow-Pratt risk aversion measures need to be extended before they can realistically represent the typical investor's choice problem. He introduces a new measure called "prudence" $= (-U''')/U''$, which measures the "propensity to prepare or forearm oneself in the face of uncertainty." As $x$ increases, this propensity is expected to decrease. The prudence for the power utility function equals $(\gamma + 1)/x$, which does decrease as $x$ increases. This is obviously an advantage of the power utility function, which may be balanced against constancy of relative risk aversion (CRRA), and this result will be seen later to be a disadvantage. In conclusion, it is realistic to assume diminishing MU as well as diminishing prudence as $x$ (wealth) increases. Vinod (2001) argues that any realistic $U(x)$ should exhibit Kimball's nonincreasing prudence (KNIP) and proves the following:

**Lemma 6.3.1.** The requirement $U'''' < 0$ is necessary for nonincreasing prudence.

*Proof.* Nonincreasing prudence requires $(d/dx)\{(-U''')/U''\} < 0$, that is, $\{(-U'''')U'' - (-U''')(U''')/(U'')^2 < 0$. Therefore prudence requires that $U''''U'' > (U''')^2 > 0$. Since $U'' < 0$, the necessary condition for the left side $U''''U'' > 0$, is $U'''' < 0$, as claimed by the lemma. □

In macroeconomics, the consumption function is a relation between aggregate consumption in an economy and aggregate income. For example, $C = a + bW$, is a linear consumption function relating the consumption $C$ to investment income or wealth $W$. The marginal propensity to consume (MPC) is defined as the derivative of consumption with respect to income, that is,

$dC/dW$. The microeconomic MPC can also be defined at the level of the individual consumer. Carroll and Kimball (1996) note that if the only form of uncertainty is in labor income, CRRA utility implies the linear consumption function: $C = a + bW$, which has constant MPC = $b$, even if wealth $W$ increases. As Carroll and Kimball argue any realistic MPC should change as investor's wealth changes. This could be a direct result of Markowitz's (1952a) assertion that the utility to the consumer from additional consumption depends on the status quo level of wealth before the change.

Recall that the logarithmic and power utility functions have constant CRRA of unity and $\gamma$, respectively. Since the implication of constancy of CRRA that the MPC is constant is unrealistic, the power utility or logarithmic utility specifications need to be replaced by a more general utility function specification. The hyperbolic absolute risk aversion (HARA) utility function is defined by the constancy of a new parameter $\kappa = U'''U'/(U'')^2$. A HARA utility function satisfies the desirable nonincreasing absolute risk aversion (NIARA) property only if $\kappa > 1$. We conclude that HARA utility functions with $\kappa > 1$ offer adequate generalization of the power or logarithmic utility functions for our purposes.

Huang and Litzenberger (1988) have further references and a review of the literature on utility functions and risk aversion. Vinod (1997) discusses estimation of HARA utility functions in the context of "consumption" capital asset pricing model (C-CAPM). See Campbell et al. (1997) for related macro econometrics involving Euler equations. Kraus and Litzenberger (1976) proposed a modification of the CAPM to incorporate the "market gamma" or systematic skewness.

In light of Kimball's results noted above, it would be useful for the investor's utility function to satisfy both the NIARA ($U'''U'/(U'')^2 > 1$), as well as diminishing prudence ($-U'''/U''$ properties. Unfortunately, the known conditions for achieving these properties are necessary but not sufficient. Hence we cannot guarantee that NIARA and diminishing prudence will always be achieved.

In summary, a class of utility functions satisfying $U' > 0$, $U'' < 0$, $U''' > 0$, and $U'''' < 0$, is called D4. The last condition, $U'''' < 0$, is necessary (not sufficient) for Kimball's nonincreasing prudence (KNIP). The classes D1 to D4 are sequentially more restrictive, which means that the smallest percentage of the population of investors can be reasonably expected to satisfy all requirements of D4. A larger percentage will belong to the D3 class, even larger will belong to the D2 class, and the largest will belong to the D1 class. The next four subsections deal with four orders of stochastic dominance that apply to investors in classes D1 to D4, respectively. They are potentially useful in rejecting dominated portfolios outright. Since choosing among assets is a time-consuming and expensive task, investors can save some of these costs if they can limit the focus of their attention on a few dominant portfolios. Since data reported by asset managers can be misleading, as revealed by recent Enron, Worldcom, and other scandals, there is a need for in-depth look before com-

mitting large funds to risky portfolios, even if stochastic dominance reveals them to be dominant.

### 6.3.6 First-Order Stochastic Dominance (1SD)

If the probability distribution of returns for one portfolio first-order stochastically dominates another portfolio, then all D1 investors will prefer it. We evaluate 1SD by first restricting $x \in [x_*, x^*]$, a closed interval. Assume that we have two uncertain prospects $A$ and $B$ defined by their PDFs $f_a(x)$ and $f_b(x)$ with corresponding CDFs $F_a(x)$ and $F_b(x)$. For example, $f_a(x)$ and $f_b(x)$ can be the distributions of two competing portfolios $A$ and $B$. An important question in financial economics is to compare the portfolios despite the uncertainty associated with both. Denote the difference as $F_{ab}(x) = F_a(x) - F_b(x)$. Portfolio $A$ dominates $B$ if we have

$$F_{ab}(x) \le 0, \quad \text{or} \quad F_a(x) \le F_b(x) \quad \text{for all } x \in [x_*, x^*]. \tag{6.3.1}$$

Since the CDF for choice $A$ is always less than that for choice $B$, the probability of being below any specified return is always lower for choice $A$. Any investor who prefers a higher return will therefore prefer choice $A$.

The geometric representation in terms of PDFs is more intuitive, it says that $f_a(x)$ for the superior prospect will be to the right-hand side of $f_b(x)$. For illustration, Figure 6.3.1 uses the well-known beta density, which was mentioned as Pearson's type I density in Chapter 2. The PDF is $f_{\text{beta}}(x, r, s) = x^{(r-1)}(1-x)^{(s-1)}/\text{Beta}(s, r)$, where $r > 0$ and $s > 0$ are the two shape parameters, Beta$(r, s)$ denotes the beta function, and $x \in [0, 1]$. The mean of the beta density is known to be $r(r+s)^{-1}$. The dashed line in Figure 6.3.1 is $f_{\text{beta}}(x, 3, 6)$ with the mean $3/9 = 0.33$. The dashed line stochastically dominates the solid line based on $f_{\text{beta}}(x, 2, 5)$ with the mean $2/7 = 0.286$. As can be seen in the figure, first-order stochastic dominance also implies that choice $A$ with a dashed line has a higher average return ($= 0.33$) than the average for the choice $B$($= 0.286$) represented by a solid line.

When evaluated at a fixed $x$, the CDF for the distribution on the right will be lower and will reach its highest value of unity at a higher value of $x$. In Figure 6.3.1 note that the CDF for the dashed line is lower at all levels of $x$. The first-order stochastic dominance is written $A \succeq_1 B$. In sum, $A \succeq_1 B$ means that the expected utility of $A$ is no smaller than the expected utility of $B$ for every nondecreasing $U(x) \in D_1$.

Due to random variation, $f_a(x)$ may exceed $f_b(x)$ in most of the relevant range but may fail for some points of the range. Do we then still prefer $A$ to $B$? This becomes a question of statistical testing discussed later in Chapter 9. The definition of class $D_1$ involves only one restriction on the sign: $U' \ge 0$, which means positive MU. However, economic theory requires diminishing, not just positive MU, implying the second restriction on the sign: $U'' \le 0$. Hence

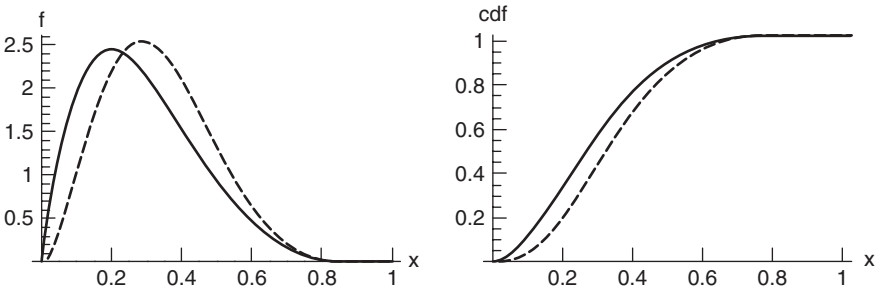



**Figure 6.3.1** First-order stochastic dominance of dashed density over solid density

the preference ordering under 1SD may not be economically meaningful, so it leads us to consider second-order dominance next.

Stochastic dominance result can also be easily applied to data. Given two securities with $T$ observed returns, since each return comprises $1/T$ percent of the data, $1/T$ can be used as an empirical probability. This probability may be used to draw an empirical CDF, or the returns can be ordered from smallest to largest. Letting $r_i(1)$ indicate the smallest return, and $r_i(T)$ the largest return for firm $i$, the following inequality may be used:

**First-Order Stochastic Dominance.** If $r_1(j) > r_2(j)$ for all $j = 1, \ldots, T$, then security 1 first-order stochastically dominates security 2. Therefore all investors will prefer security 1 regardless of risk preference.

### 6.3.7 Second-Order Stochastic Dominance (2SD)

The prospect $A$ dominates prospect $B$ in 2SD; that is, $A \leq_2 B$ for the class $D_2$ utility functions if we have

$$\int_{x_*}^{x} F_{ab}(y)dy \leq 0 \quad \text{for } x \in [x_*, x^*]. \tag{6.3.2}$$

Figure 6.3.2 presents graphs similar to Figure 6.3.1 for 2SD. Again the dashed line represents $f_a(x) = f_{\text{beta}}(x, 3, 6)$, and dominates the solid line representing $f_b(x) = f_{\text{beta}}(x, 2, 4)$. Here the shape parameters $s$ and $r$ of the beta are chosen so that the means $3/(3 + 6)$ and $2/(2 + 4)$ are each equal to 1/3. The variance of the beta density is analytically known to be $rs(r + s + 1)^{-1}(r + s)^{-2}$. Here the shape parameters are chosen in such a way that the variance of the dashed line (= 0.022) is less than the variance of the solid line (= 0.031). It is intuitively obvious in Figure 6.3.2 that the dashed line with a narrower spread (lower risk) dominates the solid line with a wider spread. The CDFs for the domi-

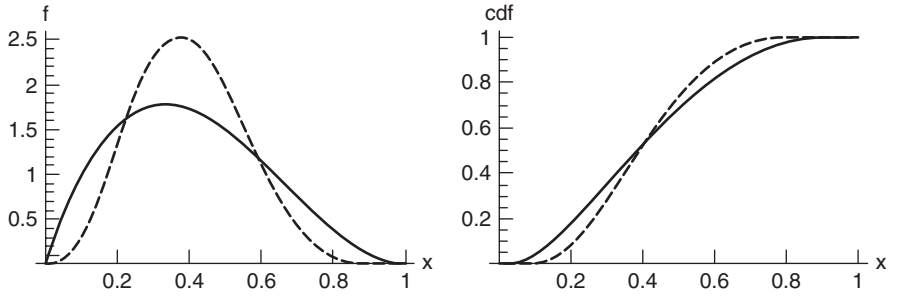



**Figure 6.3.2** Second-order stochastic dominance of dashed density (shape parameters 3, 6) over solid density (shape parameters 2, 4)

nated line do not clearly indicate the dominance. In Figure 6.3.1 the dashed CDF curve in the right-hand panel is obviously to the right of the solid curve, but the CDF's in the second-order case cross each other in Figure 6.3.2. Since the mean returns are fixed to be equal to 1/3 for the two cases and only the variances differ, we should not expect a visually clear-cut dominance here. Researchers have proved that we need to find the area under the CDF curves themselves to assess dominance in this second-order dominance case. Since the graphs do not show the dominance visually, we use numerical calculations at eleven values in the [0, 1] range in increments of 0.1 (deciles). For $f_{\text{beta}}(x, 3, 6)$, the cumulative sum is 7.16712, which slightly exceed a comparable cumulative sum for $f_{\text{beta}}(x, 2, 4)$; that is, when $s = 2$ and $r = 4$, it is 7.1665.

Equation (6.3.2) involves an integral of the difference between two CDFs ($F_{ab}$), which are already integrals of PDFs. Anderson's (1996) numerical algorithm discussed in Section 6.3.1 converts the integrations of (6.3.1) and (6.3.2) into simple matrix premultiplications. Recall that the third condition $U''' > 0$ on the third derivative is necessary for nonincreasing absolute risk aversion (NIARA) and leads us to consider the third-order next.

In the context of return data, we have 2SD if $\Sigma_{w=1}^{j} r_1(w) > \Sigma_{w=1}^{j} r_2(w)$ for all $j = 1, \ldots, T$, then security 1 second-order stochastically dominates security 2. Therefore all risk averse investors will prefer security 1 over security 2.

### 6.3.8 Third-Order Stochastic Dominance (3SD)

The prospect $A$ dominates prospect $B$ in 3SD; that is, $A \geq_3 B$ for the class $D_3$ utility functions if in addition to (6.3.2) we have

$$\int_{x_*}^{x} \int_{x_*}^{y} F_{ab}(z)dzdy \leq 0 \quad \text{for all } x \in [x_*, x^*]. \tag{6.3.3}$$

The inequality in (6.3.2) involves an integral of the difference between two CDFs ($F_{ab}$). The inequality in (6.3.3) involves a further integral of the integral of $F_{ab}$. Anderson's (1996) numerical algorithm based on modified trapezoidal rule converts the integration in (6.3.3) into a further matrix multiplication. For convincing graphs of 3SD we need analytical densities with common mean and variance, and showing that density with a larger skewness dominates. Instead of burdening with such graphs of four parameter density functions, we ask the reader to intuitively extrapolate from the earlier figures. However, numerical returns satisfying these properties are illustrated in Vinod (2001).

### 6.3.9 Fourth-Order Stochastic Dominance (4SD)

The fund $A$ dominates fund $B$ in 4SD; that is, $A \leq_4 B$ for the class $D_4$ utility functions if in addition to (6.3.3) we have

$$\int_{x_*}^{x}\int_{x_*}^{y}\int_{x_*}^{z} F_{ab}(w)dwdzdy \leq 0 \quad \text{for all } x \in [x_*, x^*]. \tag{6.3.4}$$

As with other stochastic dominance checks, the fourth order has an empirical counterpart involving cumulative sums of cumulative sums. It is possible to consider fifth and higher orders of stochastic dominance with conditions on derivatives of $U$. However, these conditions cannot be justified on economic grounds. Similarly conditions on fifth and higher moments of $f(x)$ cannot be justified. Using moment-generating functions, Thistle (1993) develops infinite degree stochastic dominance ($\infty$SD) for the special case when the support of $f(x)$, the PDF of $x$, is for $x > 0$. Since zero or negative returns are a common unpleasant fact in finance, we find that ($\infty$SD) is not a viable candidate to replace or extend 4SD.

### 6.3.10 Empirical Checking of Stochastic Dominance Using Matrix Multiplications and Incorporation of 4DPs of Non-EUT

Recall the 4DPs satisfied by the transformation (6.2.2). That distribution helps us find the weights that map the cumulative probability $p \in [0, 1]$ for perfectly EUT-compliant investor on the weighted $p$ for less compliant individuals also in the same closed interval [0, 1]. In this subsection we discuss a numerical algorithm for checking stochastic dominance of orders 1 through 4. The CDF represents the area under the PDF. A trapezoidal algorithm was suggested in Vinod (1985) for computing areas. Here we use a similar algorithm from Anderson (1996) which involves simple matrix multiplications.

Although $k$ will be large in practice, we illustrate a useful matrix for $k = 3$ used by Anderson (1996):

$$I_f = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Thus $I_f$ denote a $k \times k$ matrix of ones along and below the main diagonal and zeros above the main diagonal. Let $p^A$ denote the $k \times 1$ vector of individual probabilities for $f_a$. Anderson (1996) showed that the matrix multiplication $I_f\, p^A$ computes the cumulative probabilities and provides an approximation to $F_a$, the empirical CDF. Anderson (1996) provides nonparametric statistical tests for the three orders of stochastic dominance in the context of a comparison of income distributions. In this subsection we extend his algorithm to incorporate the four desirable properties of $U(x)$ from the prospect theory in the context of the portfolio selection problem. We also discuss a nonparametric test using Fisher's permutation distribution recently revived in the bootstrap literature.

If two portfolio distributions are being compared over the same variable $x$, it is necessary to create a common partitioning of the range space of $f_a$ and $f_b$ into $j = 1, 2, \ldots, k$ mutually exclusive and exhaustive class intervals with possibly variable widths $d_j$. Denote by $x^j$ the cumulated widths $d_j$. Since widths should be positive, we choose a suitable starting value $x^j$ when $j = 0$. It should be smaller than the smallest value of $x$ in the data, $x^0 < \min(x)$. Next we compute the areas with reference to $x^j$ as the variable of integration. The relative frequencies of the two sets are then computed and cumulated to yield the CDFs $F_a$ and $F_b$.

At this point we incorporate the lessons of prospect theory and its 4DPs. We transform the cumulative relative frequencies by (6.2.2) using $\alpha$, the EUT-compliance parameter, for the individual investor. For example, $W(F_a) = \exp(-(-\ln F_a)^{\alpha})$. More generally we can use two parameters $\alpha$ and $\beta$ in (6.2.4). Now we "first-difference" these revised cumulative relative frequencies as in (6.2.3) to yield $p_{aj}, p_{bj}$ by which we explicitly identify the underlying distributions $A$ or $B$. Let $p^A$ denote the $k \times 1$ vector of $p_{aj}$ and $p^B$ denote the $k \times 1$ vector of $p_{bj}$. For brevity, let us suppress the subscripts $a$ and $b$ and write $p_j = p_{aj}, p_{bj}$. Now, by definition of cumulative probability, suppressing the subscripts $a$ and $b$ of $F$, we have $F(x^j) = \Sigma_{i=1}^{j} p_i$. Our next task is to integrate this CDF by the modified trapezoidal rule to accommodate unequal interval widths. Denote

$$C(x^j) = \int_0^{x^j} F(z)dz \approx 0.5\left\{F(x^j)d_j + \sum_{i=1}^{j-1}(d_i + d_{i+1})F(x^i)\right\}. \tag{6.3.5}$$

Recall that we have illustrated the $k \times k$ matrix denoted by $I_f$ for $k = 3$ above. Similarly we now illustrate another matrix

$$I_F = 0.5\begin{bmatrix} d_1 & 0 & 0 \\ d_1 + d_2 & d_2 & 0 \\ d_1 + d_2 & d_2 + d_3 & d_3 \end{bmatrix}.$$

Note that this $I_F$ matrix has the uppercase $F$ as the subscript, and it is intended to be a useful matrix for implementation of the numerical integration needed for stochastic dominance of orders 2, 3, and 4. See Anderson (1996), who shows that matrix multiplication $I_F I_f(p^A - p^B)$ computes (6.3.5) for assessing 2SD.

We need to further integrate this $C(x^j)$ if we want 3SD as

$$\int_0^{x^j} C(z)dz \approx 0.5\left\{C(x^j)d_j + \sum_{i=1}^{j-1}(d_i + d_{i+1})C(x^i)\right\}. \tag{6.3.6}$$

Again, a similar matrix multiplication $I_F\ I_F\ I_f(p^A - p^B)$ yields an approximation to the integral in (6.3.6). All we have done to go from 2SD to 3SD is to premultiply by the $I_F$ matrix one more time.

The null hypothesis for 1SD in matrix notation is $\mathrm{H}_0: I_f(p^A - p^B) = 0$ and the alternate hypothesis is $\mathrm{H}_a: I_f(p^A - p^B) \leq 0$. If the last inequality is indeterminate, the conclusion of the test is also indeterminate. The statistical problem is simply whether the computed $I_f(p^A - p^B)$ is statistically significantly negative to conclude that $A \geq_1 B$. Then the investor can safely choose the portfolio $A$. This is a symmetric one-tail test procedure in the sense that if the observed $I_f(p^A - p^B)$ is statistically significantly positive, we conclude the opposite result that $B \geq_1 A$ implying that the portfolio $B$ is better than $A$.

The null hypothesis for 2SD in matrix notation is $\mathrm{H}_0: I_F I_f(p^A - p^B) = 0$ and the alternate hypothesis is $\mathrm{H}_a: I_F I_f(p^A - p^B) \leq 0$. If the last inequality is indeterminate, the conclusion of the test is also indeterminate.

The null hypothesis for 3SD in matrix notation is $\mathrm{H}_0: I_F I_F I_f(p^A - p^B) = 0$ and the alternate hypothesis is $H_a: I_F I_F I_f(p^A - p^B) \leq 0$. This is similar to 2SD except for the premultiplication by $I_F$. The null for 4SD has a premultiplication by $I_F$ to the 3SD null. If the last inequality is indeterminate, the conclusion of the test is also indeterminate.

Anderson (1996) shows that these direct nonparametric tests avoid the use of generalized Lorenz curves and belong to the family of Pearson goodness-of-fit tests. His nonparametric tests start with the multinomial allocation of units to $k$ class intervals. He assumes independence and that the cell sizes satisfy $T_{pi} > 5$. The test statistic is a quadratic form in normal variables leading to a $\chi^2$ with $k - 1$ degrees of freedom. Since our application to portfolio return data does not satisfy his (independence) assumptions, we seek an extension of his tests. We do not recommend his $\chi^2$ goodness-of-fit test.

Deshpande and Singh (1985) and Schmid and Trede (1998) take a different approach based on Kolmogorov-type test statistics. They use the one-sample problem whereby it is assumed that one of two CDFs is completely known.

Schmid and Trede (1996) propose a nonparametric test with a more realistic two-sample formulation where the observed data are from two independent $f_a$ and $f_b$. Again, the independence of $f_a$ and $f_b$ is unrealistic, since there will be correlation between two portfolios over time.

The CDFs describing the past performance data for two portfolios $A$ and $B$ are somewhat similar to survival curves for two drug therapies. There is considerable interest in biometrics in solving this old problem of "matched pairs." Fisher's permutation principle suggested in 1935 was conceived as a theoretical argument supporting $t$-test. For details, the reader should refer to Efron and Tibshirani's (1993, ch. 15) application of permutation tests to the matched pairs problem. We do not use Schmid and Trede's (1998) algorithm, which uses $2^T$ possible pairs.

Under Fisher's permutation approach, the null hypothesis $H_0$ is that there is no stochastic dominance between $f_a$ and $f_b$ or $F_a = F_b$. If the $H_0$ is true, any of the returns $f_a$ or $f_b$ can equally well come from either of the distributions. Fisher's approach permits different number of observations $T_a$ and $T_b$ from $F_a$ and $F_b$. Then, if we select the first sample of $T_a$ for $f_a$ *without* replacement from the combined data on $T_a + T_b$ returns, the remaining $T_b$ returns will form the sample from $f_b$. There are $J = (T_a + T_b)!/(T_a!)(T_b!)$ ways of selecting such samples.

The distribution that puts probability mass $J^{-1}$ on each of these samples is called the permutation distribution of a statistic. The sampling distribution under the null is completely determined if we can evaluate all possible selections ($= J$) of the two samples. Then we can construct the exact sampling distribution by a computer intensive method. If $T_a = 9 = T_b$, then $J = 48{,}620$. One can reduce the computational burden without much sacrifice in accuracy by using Monte Carlo methods to select some $J' = 999$, say, from $J$. On a home PC with 500 Mhz machine a Gauss computer package needs only about 20 minutes for $J = 48{,}620$ computations, and $J' = 999$ computations are done almost instantly. We claim that the sampling distribution under the $H_0$ can be reasonably approximated.

In our context, we first transform $F_a$ and $F_b$ to satisfy the 4DPs and then evaluate the following statistics for the four orders of stochastic dominance, respectively:

$$\begin{aligned}
\hat{\theta}_1 &= \max(I_f(p^A - p^B)),\\
\hat{\theta}_2 &= \max(I_{Fa}I_f p^A - I_{Fb}I_f p^B),\\
\hat{\theta}_3 &= \max(I_{Fa}I_{Fa}I_f p^A - I_{Fb}I_{Fb}I_f p^B),\\
\hat{\theta}_4 &= \max(I_{Fa}I_{Fa}I_{Fa}I_f p^A - I_{Fb}I_{Fb}I_{Fb}I_f p^B)
\end{aligned} \tag{6.3.7}$$

As in a bootstrap, this evaluation is done a large number ($J' = 999$) of times. Special care is needed to implement (6.3.7) since the number of elements selected from the two samples can be different under random selections. If

$J' = 999$, we simply order the $\hat{\theta}$ values from the smallest to the largest, and $C_r$ the critical values for a 5% type I error one-sided test is the 950th value. We reject $H_0$ if the observed $\hat{\theta} \geq C_r$.

The stochastic dominance literature appeals to the mathematical economists, because it can address intricate propositions. If $A$ dominates $B$ in 4SD, then the superiority of $A$ over $B$ does not depend on the functional form of the utility function, as long as $U(x) \in$ D4. This is a powerful statement, since we cannot generally estimate the exact form of $U(x)$. $U'''' < 0$ is obviously a useful condition, so we see no reason to settle for 2SD or 3SD. The four desirable properties (4DPs) satisfied by (6.2.2) and listed in Section 6.2. We have shown how to incorporate these properties and still evaluate the conditions for 4SD, avoiding parametric distributional assumptions. Moreover we can use 4SD methods to suggest heuristic tools and summary measures for selecting an admissible set of portfolios satisfying the NIARA and Kimball's nonincreasing prudence (KNIP) properties from EUT and also insert 4DPs related to non-EUT.

We illustrate this section with a Table 6.3.1 reproduced from Vinod (2004). It shows that our theory is quite practical to implement. To illustrate our model, we use data from the January 2001 Morningstar Principia CD ROM Data Disk, which covers January 1991 to December 2000 period. We study all 1281 funds included in Morningstar's International Fund Category. We show that our analysis can help investor focus on a few (e.g., 5) attractive funds from among the maze of 1281 funds.

Stochastic dominance concepts check if fund $A$ dominates fund $B$. We designate S&P 500 (ticker symbol SPX) as our fund $B$. One by one each of the 1281 funds becomes our fund $A$. We use data on monthly returns from January 1991 to December 2000. For SPX (from Yahoo) and these $j = 1, 2, \ldots, 1281$ funds we find their respective *excess* returns by subtracting the three-month treasury bill rate (Tb3) and mix them. We measure $d_j$ as distances from the minimum of the mixed set. While matrix $I_f$ is simple, matrix $I_F$ and vectors $p^A$ and $p^B$ require considerable work. We use these $d_j$ in the definition of the matrix $I_F$. similar to the one exhibited after (6.3.5). Next we use matrix multiplications 1281 times for each $i$ and $j$ to get ΣiSDj for $i = 1$ to 4. The cumulative sum is used as the overall indicator of dominance. The fund with the most negative value of the cumulative sum dominates the reference fund the most.

We also compute the statistical significance of the difference among the top five funds. Only when $\alpha = 0.1$ for individuals not complying very much with the precepts of EUT is there a statistically significant difference among the mutual funds as indicated in Table 6.3.1. The Σ1SDj for the fund ranked 1 is significantly different from the similar sum for funds ranked 3, 4, and 5. The Σ2SDj for ranks 3, 4, and 5 is significantly different from rank 2. The Σ3SDj for funds ranked 1 and 2 is significantly different from fund 4 and also significantly different from fund 5. The Σ4SDj for funds ranked 1 and 2 is significantly different from that for fund ranked 4. Vinod (2004) also reports 95% so-called bootstrap-$t$ confidence intervals for ΣiSDj when $i = 1, \ldots, 4$ and many

**Table 6.3.1 Best Five Funds Sorted by Σ4SDj for α = 0.1, 0.5, 1**

| α | Fund Number | Rank | Σ1SDj | Σ2SDj | Σ3SDj | Σ4SDj |
|---|---|---|---|---|---|---|
| 0.1 | 734 | 1 | –0.609 | –9.695 | –84.8 | –536.9 |
| 0.1 | 735 | 2 | –0.609 | –9.692 | –84.76 | –536.7 |
| 0.1 | 547 | 3 | –0.596 | –9.485 | –83.57 | –533.5 |
| 0.1 | 1046 | 4 | –0.593 | –9.407 | –83.36 | –532.6 |
| 0.1 | 1268 | 5 | –0.591 | –9.392 | –82.91 | –531.6 |
| 0.5 | 722 | 1 | –0.272 | –4.255 | –33.97 | –195.4 |
| 0.5 | 931 | 2 | –0.27 | –4.346 | –34.25 | –195.1 |
| 0.5 | 161 | 3 | –0.287 | –4.296 | –33.39 | –188.9 |
| 0.5 | 1241 | 4 | –0.304 | –4.629 | –34.5 | –188.5 |
| 0.5 | 929 | 5 | –0.262 | –4.254 | –33.27 | –188.1 |
| 1 | 931 | 1 | 0.06 | –0.688 | –5.608 | –27.04 |
| 1 | 161 | 2 | 0.052 | –0.553 | –5.028 | –25.87 |
| 1 | 722 | 3 | 0.056 | –0.504 | –4.834 | –25.11 |
| 1 | 929 | 4 | 0.07 | –0.636 | –5.192 | –24.6 |
| 1 | 439 | 5 | 0.049 | –0.567 | –4.886 | –24.37 |

more columns. The main point is that this methodology is a practical tool for finance.

## 6.4 INCORPORATING UTILITY THEORY INTO OPTION VALUATION

As mentioned at the beginning of this chapter, a great appeal of the option pricing model proposed by Black and Scholes (1973) was that it was free of preference. Unfortunately, their assumptions of complete markets with continuous trading is not realistic, and this forces us to consider what happens when these assumptions are relaxed. Rubenstein (1976) assumes that the underlying stock price as well as the aggregate utility function are bivariate lognormally distributed, and permits discrete instead of continuous time to prove that Black-Scholes price formulas can still be obtained. This result requires a restrictive assumption of constant proportional risk aversion (CPRA) utility, satisfied, for example, by the logarithmic utility functions. In general, riskless arbitrage is impossible when the trading opportunities are discrete, and therefore Black-Scholes prices do not equal equilibrium values.

Ritchken and Chen (1987) show that once we enter the realm of discrete trading, it is not possible to use arbitrage-based option pricing, so one must use utility functions. In continuous time modeling one needs to determine the only value of the stock option *relative* to the price of the stock by appealing to arbitrage opportunities. Once discrete trading is accepted, one needs to

determine the absolute price of the option *and* equilibrium price of the underlying stock. After we expand the opportunities available to investors to include the buying and selling of options, it becomes impossible to assume that the distribution of returns based on the expanded opportunity set is normal, student's $t$, or stable. How do we determine the equilibrium in the stock market? This requires some assumptions about the preferences or utility functions of agents. Also, since stocks are capital assets, we need a flexible capital asset pricing model (see CAPM discussed in Chapter 2) that permits flexible utility functions and flexible probability distributions of stock returns.

Ritchken and Chen (1987) propose a novel way to introduce the needed flexibility. They argue that instead of mean-variance framework, we should consider mean lower partial moment (MLPM) framework. Hogan and Warren (1974) call the lower partial variance the semivariance. It is defined as variance below some value of the random variable and is a superior measure of risk than the usual variance, which was recognized by Markowitz (1959). It is interesting that Markowitz considered decreasing marginal utility up to the threshold value and maintaining constant MU after that point, which is where the MLPM framework is appropriate.

Ritchken and Chen derive impressive list of option price formulas for flexible distribution of returns and flexible utility functions by exploiting the MLPM framework. The relevant variables are current price, strike price, time to expiration, interest rate, and volatility, as in Black-Scholes (BS). One distinction of MLPM is that the volatility depends on market expectations. Note that non-EUT modifications to this literature remain an open problem for researchers. The relative value of options decreases as market expectations increase. It is interesting that the highest discrepancy between BS and MLPM prices occurs when market expectations are very high. For example, if the market is in a bubble situation, the discrepancy will be large. They admit, however, that additional effort in implementing the complicated formulas may not be worthwhile in practice. In any case MLPM prices are close to Black-Scholes (BS) prices, which in turn are not biased in any direction, provided that dividends are not ignored.

## 6.5 FORECASTING RETURNS USING NONLINEAR STRUCTURES AND NEURAL NETWORKS

The pdf of excess returns $f(x)$ based on past data was used above to choose among a set of portfolios. Since conditions change, new technologies get developed, populations move, international crises including wars occur, government policies change, and the observed $f(x)$ and individual values of excess returns at different points in time also change. What we really need are forecasts of future returns $x_t$ and forecasts of related summary statistics, ranks, and so on. Chapter 4 dealt with autoregressive moving average (ARMA) and generalized autoregressive conditional heteroscedasticity (GARCH) and related

models for forecasting of returns and volatilities. In Section 6.5.1, we describe standard linear and nonlinear statistical forecasting tools. These may be useful for forecasting profits or returns in a business sector using past data on related macroeconomic or other variables. In Section 6.5.2, we consider forecasts of quality ratings of a product or business using categorical or limited dependent variables. In Section 6.5.3, we describe the neural network models designed to incorporate changing structures over time in forecasting models. We retain the notation from related statistical literature where the dependent variable is denoted by $y$ and explanatory variables are denoted by $x$. In other words, we do not denote the dependent variable as $x$, even if we are forecasting excess returns.

### 6.5.1 Forecasting with Multiple Regression Models

Regression methods for forecasting are well known. We describe them here for completeness using standard notation where $y$ is the dependent variable. The forecast will yield values of $y$ given values of regressors $x$. This meaning of $x$ should not be confused with $x$ used to represent excess returns in earlier sections. Consider the standard linear regression model with $y$ as output or dependent variable, $x_1$ to $x_p$ as inputs or regressor variables, $u$ as the error term, and $\beta_0$, $\beta_1$, to $\beta_p$ as the regression coefficients:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon = X\beta + \varepsilon. \tag{6.5.1}$$

Assuming that there are $T$ observations, $X$ is the $T \times (p + 1)$ matrix of data on all regressor variables including the first column of ones to represent the intercept $\beta_0$, $y$ is the $T \times 1$ vector of data on the dependent variable, $\beta$ is the $(p + 1) \times 1$ vector of regression coefficients, and $\varepsilon$ is the $T \times 1$ vector of unknown true errors. It is customary to incorporate the intercept into the column vector of coefficients and treat $\beta$ as a $p \times 1$ vector without loss of generality. Also any nonlinear model that is linear in parameters can be accommodated in (6.5.1) by having the nonlinear functions (e.g., logs, powers) as regressors in the $X$ matrix. We assume the following probabilistic structure for $y$ and errors:

$$y = X\beta + \varepsilon, \quad E(\varepsilon) = 0, \quad E\varepsilon\varepsilon' = \sigma^2\Omega, \tag{6.5.2}$$

where $\Omega$ is a $T \times T$ matrix of variances and covariances among errors. The variances are along the diagonal and parameterize the heteroscedasticity, and the covariances are off diagonal and often measure the autocorrelation structure. From (6.5.2) verity that $E(y) = X\beta$.

If we further assume that $y$ (and $\varepsilon$) are multivariate normal, $y \sim N(\mu, \sigma^2\Omega)$, we can write down the log likelihood (LL) function relatively simply. The likelihood function is obtained by simply multiplying the densities for each observation and interpreting the product as a function of the unknown parameters.

The *score vector* is defined as the derivative of LL with respect to the parameters. The first-order condition for a maximum of LL is that the score should be zero. The second-order condition for a maximum from calculus is that the matrix of second-order partial derivatives should be negative definite. The maximum likelihood (ML) estimator of the β vector satisfies both conditions.

The generalized least squares (GLS) estimator is obtained by solving the following score functions or "normal equations" for β. If the existence of likelihood function is not assumed, it is called a "quasi-" score function (QSF):

$$g(y, X, \beta) = X'\Omega^{-1}X\beta - X'\Omega^{-1}y = 0. \tag{6.5.3}$$

The GLS estimator

$$\beta_{GLS} = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}y \tag{6.5.4}$$

is also the maximum likelihood (ML) or quasi-ML estimator. The regression model is workhorse for structural estimation, forecasting, classification, and learning models and has been generalized in various ways.

### 6.5.2 Qualitative Dependent Variable Forecasting Models

A well-known generalization of standard regression models in (6.5.1) is to allow for dummy variables as dependent variables used in forecasting demand for durable goods. For example, when a person buys a car, the dependent variable is unity, and when she fails to buy, it is zero. The probability of purchasing a car can depend on income, sentiment, cost of gasoline, and myriad other variables. Since the automobile industry has a great deal of impact on business returns in related sectors, we should not ignore the statistical tools relating dummy dependent variable $y$ to a nonlinear function $f(x, \beta)$ of regressors $x$ on the right side. Some forecast applications have the $y$ variable restricted to three or more qualitative categories where $y$ takes only a limited set of usually discrete values.

For example, $y$ may be a dummy variable taking values 0 or 1, or represent three categories: "good, better, and best," or number of stars 1 to 5, as with Morningstar's classification of mutual funds reported in popular magazines for investors. The so-called probit, Tobit, and logit models are popular in econometrics texts (see Greene, 2000, for dealing with categorical data). Since neural networks belong to biostatistics, let us use a slightly different generalization of (6.5.1) for categorical data.

A popular biostatistics tool is the generalized linear model (GLM) method. The GLM method is flexible and distinct from the logits or fixed/random effects approach popularized by Balestra, Nerlove, and others, and covered in econometrics texts (Greene 2000). GLM is mostly nonlinear, but linear in parameters. McCullagh and Nelder (1989) describe GLM in three steps:

1. Instead of $y \sim N(\mu, \sigma^2\Omega)$, we admit any distribution from the exponential family of distributions with a flexible choice of relations between mean and variance functions. Nonnormality permits the expectation $E(y) = \mu$ to take on values only in a meaningful restricted range (e.g., nonnegative integer counts or [0, 1] for binary outcomes).
2. Define the systematic component $\eta = X\beta = \Sigma_{i=1}^{p} x_j\beta_j$, $\eta \in (-\infty, \infty)$, as a linear predictor.
3. A monotonic differentiable link function $\eta = h(\mu)$ relates $E(y)$ to the systematic component $X\beta$. The $t$th observation satisfies $\eta_t = h(\mu_t)$:

$$E(y) = h\left(\sum_{i=1}^{p} x_i\beta_i\right). \tag{6.5.5}$$

For the GLS estimator of (6.5.4), the link function $h$ is identity, or $\eta = \mu$, since GLS assumes no restriction on the range of the dependent variable: $y \in (-\infty, \infty)$. When $y$ data are counts of something, we obviously cannot allow negative counts. Then we need a link function that makes sure that even if $\mu = X\beta \in (-\infty, \infty)$, $h(\mu) > 0$. The link function has the job of mapping an infinite interval on $[0, \infty)$. Similarly, for $y$ as binary (dummy variable) outcomes, $y \in [0, 1]$, we need a link function $h(\mu)$ that maps the interval $(-\infty, \infty)$ interval for $X\beta$ on [0, 1] for the binary dependent variable $y$. A typical link function $h(\mu)$ is chosen to be the smoothly increasing function, such as any CDF. The most common software packages use the logistic $h(\mu) = \exp(\mu)/[1 + \exp(\mu)]$.

Which CDF is chosen as a link function between the categorical variable $y$ and $X\beta$? It is convenient to choose it from the exponential family of distributions, which includes Poisson, binomial, gamma, inverse-Gaussian, and so on. It is well known that "sufficient statistics" are available for the exponential family. In our context $X'y$, which is a $p \times 1$ vector similar to $\beta$, is a sufficient statistic. A "canonical" link function is one for which a sufficient statistic of $p \times 1$ dimension exists. Some well-known canonical link functions for distributions in the exponential family are $h(\mu) = \mu$ for the normal, $h(\mu) = \log \mu$ for the Poisson, and $h(\mu) = \log[\mu/(1 - \mu)]$ for the binomial; $h(\mu) = -1/\mu$ is negative for the gamma distribution. The GLM theorists ask us to choose the link function depending on the relationship between the mean and variance. For example, if the mean and variance are identical to each other, the Poisson link is appropriate.

For GLS computations Newton-Raphson method is commonly used to solve (6.5.3) even if $g(y, X, \beta)$ is highly nonlinear. This method can yield, often as a by-product of computations, the actual Hessian matrix based on second-order partial derivative vectors of quasi-score functions (or outer products or gradients of scores). The Hessians yield standard errors of estimated coefficients from square roots of diagonal terms. Iterative algorithms based on Fisher scoring are developed for finding the standard errors GLM coefficients. Fisher scoring uses the "expected value" of the Hessian matrix.

### 6.5.3 Neural Network Models

In the 1940s when little was known about how human brain processes information, McCulloch and Pitts proposed a simple neural network (NN) model. It allows for massive parallel processing of neural input in the brain at multiple layers and feedbacks leading to nonlinear output. Mathematically, NN involves weighted sums of inputs in layers with several nodes linked to an intermediate hidden layer and then output to fewer nodes. Biologists have since learned that human neurons are far more complex. However, the mathematical model has been applied with modifications to nonparametric estimation, medical diagnoses, and different kinds of learning. Financial economists and those in related fields may simply think of feed-forward NN as a method for generalizing linear regression functions (Kuan and White, 1994).

In early versions of neural network models the link functions similar to $h$ in (6.5.5) described earlier were heaviside or unit step function: $h(m) = 1$ if $m \geq 0$ and $h(m) = 0$ if $m < 0$. The neuron is turned on when $h(m) = 1$, say, and remains off otherwise. More generally it is a threshold function: $h(m) = 1$ if $m \geq m_0$ and $h(m) = 0$ if $m < m_0$, where $m_0$ is the threshold value. For further generality the link function $h(\mu)$ is chosen to be smoothly increasing CDF as in (6.5.5).

Neural networks, inputs are signals $x_i$, $i = 1, \ldots, p$, assembled in a column vector $x$. When such an $x$ vector is augmented by inserting the number 1 at the start, we denote it by $\hat{x}$ with $p + 1$ elements. The signal goes to output units $y_j, j = 1, \ldots, r$. $x_i\gamma_{ji}$ reaches the $j$th output. If $x_i = 1$ for the intercept, the output is $\gamma_{0j}$. We assemble $p + 1$ coefficients $\gamma_{j0}$ to $\gamma_{jp}$ in a $(p + 1) \times 1$ column vector denoted by $\gamma_j$. Since both $\gamma_j$ and $\hat{x}$ are $(p + 1) \times 1$ vectors, $(\hat{x}'\gamma_j)$ is a scalar, where the prime denotes a transpose. There are $r$ such scalars comprising $r$ equations of so-called seemingly unrelated regressions (SUR) model. Econometrics texts describe the SUR model as one of the generalizations of single equation regression model.

Now let us introduce one hidden layer of activations with $k = 1, \ldots, q$ additional sets of parameters assembled in $\beta_{kj}$, a $q \times r$ matrix. This matrix links the $q$ hidden layers to $r$ outputs.

In Figure 6.5.1 we illustrate a [4, 2, 3] neural network with four input nodes, two hidden nodes, and three output nodes. Figure 6.5.2 shows direct links from inputs to outputs that are permitted in the general setup. To include the direct links, let $\alpha$ denote a $p \times 1$ vector of parameters for directly linking the $p$ inputs to $r$ outputs without using the hidden layer. The $\alpha$ in this section is obviously different from the $\alpha$ used to denote departure from expected utility theory in most of this chapter. Since $x$ is a $p \times 1$ vector of inputs, $x'\alpha$ is a scalar. Let $\psi$ denote the function that operates on the scalar $(\hat{x}'\gamma_j)$ for each $j$, and let $\Sigma_j$ denote the summation from $j = 1$ to $j = r$ for the $r$ outputs. Now define the NN model in terms of the link function $h$ as

$$E(y) = h\left[x'\alpha + \beta_0 + \sum\nolimits_j \psi(\hat{x}'\gamma_j)\beta_j\right]. \tag{6.5.6}$$

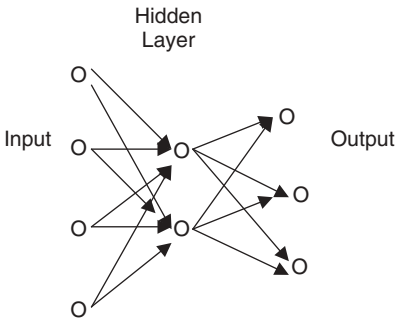



**Figure 6.5.1** Neural network

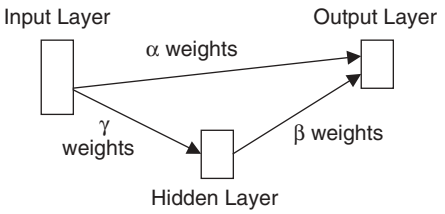



**Figure 6.5.2** Neural network with hidden and direct links

Ultimately the best-fitting coefficients in (6.5.6) must be found by computer programs, already available in many software packages including S-Plus, GAUSS, and others. There are difficulties and some unresolved issues about statistical testing of the fitted coefficients. A particular case of (6.5.6) when $\alpha = 0$ and $h$ is the identity function is known as projection pursuit (PP) regression pioneered by Friedman and Stuetzle (1981).

This generalization of the regression model was applied in Vinod (1998) for cointegration estimation, where Vinod shows how PP regression does not suffer from the curse of dimensionality, suffered by other nonparametric regression methods. Tsey (2002) provides an example of [2, 3, 1] feed-forward neural network to forecast the price of IBM stock.

It is clear from the large number of coefficients needed for NN models that they are less suitable for structural estimation and more useful for forecasting. In financial practice, NN models can help in choosing portfolios based on market forecasts. However, there is a danger of overfitting the in-sample data leading to poor out-of-sample forecasts.

Qi (1999) gives a good discussion of NN models for one-step-ahead forecasting and whose data are available on the website of the journal (JBES). The author fits the following version of NN:

$$x_t = h(x_{t-1}, \alpha, \beta, \gamma), \tag{6.5.7}$$

where the output $x_t$ is excess returns at time t, the inputs are financial and economic variables at time $t-1$ collected in $x_{t-1}$. Several references useful for relevant theory and for using the NN in financial forecasting problems are available in Kuan and White (1994), Abhyankar et al. (1997), and Qi (1999). The recursive formulation of evolution over time permits implicit incorporation of structural changes into the model. They conclude that profit opportunities do exist out of sample with these models. Does this mean markets are not efficient? Maasoumi and Racine (2002) discuss entropy measures to study the profit opportunity issue with NN models. The entropy NN models suggest small nonlinear *unconditional* autocorrelations but do not provide profit opportunities when they compare relative merits of churning of portfolio versus buy-and-hold strategy.

This section shows how NN methods generalize the simple regression to allow for nonlinearities (1) by allowing for limited dependent variables through link functions and (2) by considering several sets of equations as in seemingly unrelated regression (SUR) equations framework. We conclude that neural network methods can be a powerful tool for stock market forecasting, provided that care is taken to avoid overfitting. Since interpretation of fitted NN coefficients and their standard errors are not readily available, neural network technique has serious limitations if used for structural estimation.

Abhyankar et al. (1997) review various research results regarding nonlinear structure in four major stock market indexes including S&P 500, DAX, Nikkei 225, and FTSE 100 indexes in major countries plus two indexes based on S&P 500 futures and FTSE 100 futures. They confirm the presence of nonlinearities using Brock-Dechert-Scheinkman (BDS) test, Lee-White-Granger test based on neural networks, generalized autoregressive (moving average) conditional heteroscedasticity (GARCH), and nearest neighbor tests (see also Racine (2001) on nonlinear predictability of stock returns). Are stock markets returns best represented by a stochastic process or by a nonlinear deterministic process plus a noise term? Abhyankar et al. find by estimating Lyapunov exponent that market indexes do not follow "deterministic chaos." Those interested in chaos theory may refer to Brock et al. (1991) and surrounding vast literature. This section focuses on the potential applications of neural networks in finance.

CHAPTER 7

# Incorporating Downside Risk

## 7.1 INVESTOR REACTIONS

The preceding chapter gives a rational method for dealing with risky events. This chapter begins by giving us a foothold for understanding human behavior and indicates how it is asymmetric with reference to upside and downside. Hence there is a need to measure the downside risk by focusing on the down side directly. Accordingly we suggest measuring the downside risk by contemporaneous VaR, down standard deviation, down beta, and "put implied volatility," where put refers to the option to sell. Since one should be interested not just in the historical downside risk but the risk in the future, we consider the matter of forecasting downside risk and argue that "put implied volatility" is the only measure that is forward looking in Section 7.2. When we consider the boom and bust cycles, the same seeds of an impending bust are included in the data for boom periods. In Section 7.3 we briefly mention growth rate of cash flows as one potentially useful indicator consistent with economic theory, because it can be used to determine allocation of assets to equity. For choosing individual stocks, one must rely on expert analysts, but we show evidence that as a group they exhibit herd behavior and can be overoptimistic. In Section 7.3 we consider in much greater detail the downside risk in international investing, including currency devaluation risk and other risks when investing in third world developing countries. Note that international investments including some in third world countries are often strongly recommended by financial advisers to counter aging and slowing opportunities in the developed world. In Section 7.4 we discuss fraud, corruption, and other risk factors often ignored in most finance books.

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

### 7.1.1 Irrational Investor Reactions

The theory of risk aversion primarily provides a rationale for investors perceiving downside and upside risk differently. But this is certainly not the final word on investor behavior. There are many market phenomena that are persistent, but rational behavior simply cannot explain them.

Lamont and Thaler (2003) discuss several peculiar examples where the financial markets failed to follow even the very basic law of one price, which says that identical goods should have identical price. Their main examples are as follows:

1. Closed-end country fund for Germany had a premium of 100% in January 1990.
2. American depository receipts (ADRs) for a stock for a company called Infosys from India had a premium of 136% over the Bombay price.
3. Pricing of Royal Dutch compared to Shell should have a ratio of 1.5, but was 30% too low in 1981 and 15% too high in 1996.
4. Two classes of shares based on voting rights should not have wildly different market prices.
5. The ratio of Palm and 3Com shares should not differ from 1.5, and yet it did. Investors wanted to pay some 2.5 billion to buy expensive shares of Palm. The problem that arbitragers faced was that they could not sell Palm short.

### 7.1.2 Prospect Theory

One area of study that provides a rationale for seemingly irrational behavior is prospect theory developed by Kahneman and Tversky in the late 1970s, and this theory later won Kahneman the Nobel Prize in 2002. Prospect theory exploits the "emotional" part of human behavior by using psychology to examine individual's actions. There are many actions in the market that seem to reflect human emotion. Market analysts speak of herd behavior, irrational exuberance, and capital flight that are not easy to reconcile with our picture of an investor at a desk crunching numbers to obtain the optimal portfolio.

Prospect theory has been used to identify several anomalies of human behavior that call into question our rationality. For example, Kahneman and Tuersky pose a hypothetical situation where 600 people are infected with a virus for which the standard inoculation will save 200 people. The subject then has the choice between staying with the traditional inoculation, or trying an experimental treatment that has a one-third probability of saving everyone, and a two-thirds probability of not saving anyone. This scenario, or a similar scenario, has been presented to many different groups of people: students, businesspeople, doctors, and by and large most people choose to take the risk with the possibility of saving all.

This is understandable. We all want to be a hero. In this case there is a chance for the daring doctor to go out on a limb and perhaps cure everyone. If the treatment doesn't work, there will be no witnesses to the failure. Something similar happens in the stock market, where initial public offerings (IPOs) receive a substantial premium on their first day of trading (Ibbotson and Ritter, 1995). One explanation is that all investors want to be in on the next big thing and be that person who discovered Apple Computers, or Nike, before they were big. It is even rational since the expected value of people saved is the same whether you use the experimental treatment or not.

The anomaly surfaces when the same types of people are given the choice between using the standard inoculation that will kill 400 people, or the experimental treatment that has a one-third probability of killing no one and a two-thirds probability of killing everyone. Very few people will take the risk of killing the entire population. But if we compare the choices to the previous scenario, they are identical. The only difference is that the wording has changed from being in terms of saving lives to in terms of killing. Changes in wording should not affect the rational investor, but people are affected by emotion. This is how marketing can exist. No matter how much we wish it wasn't true, people are influenced by the color of the box, by the quality of paper of the company newsletter, or by who is dating the CEO.

Tenorio and Cason (2002) find similar deviations from rationality by recording choices made on the television game show "The Price Is Right." When it comes to spinning the big wheel, which determines who gets to go to the showcase showdown and compete for the big prizes, three contestants spin the wheel and the one closest to $1.00 without going over wins. They spin once, and then they can choose whether or not to spin again. There is an optimal strategy to this game, but Tenorio and Cason show that contestants spin more than the mathematically optimal number of times.

Pollsters need to constantly keep in mind the following when taking opinion polls. It is not just the questions that are asked, but the order in which they are asked, that will influence the results. Consider the question poll that asks: "Are you confident in our local treasurer?" It is different from the two-question poll that asks: "Are you aware that our local treasurer was indicted for tax evasion?" "Are you confident in our local treasurer?"

The last question in each poll is identical, but the answers one will receive will be drastically different. Past experience, new information, and even the gender of the pollster will change the results of the poll. We therefore cannot expect the same people to be rational automatons when they become stock market investors.

To compensate for the errors humans can make, a branch of literature on bounded rationality and learning models has emerged. These researchers try to model the individual who is not able to do complex mathematical calculations instantaneously when making decisions but is able to either approximate the solution, or learn the solution over time. Some papers that have tried to

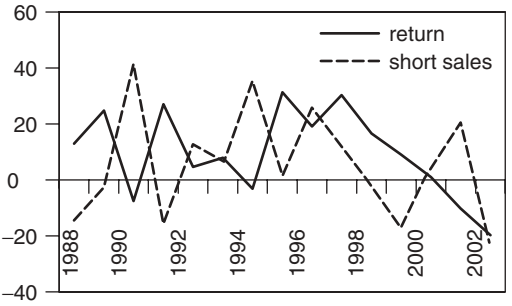



**Figure 7.1.1** Short sales and stock returns in the NYSE

model an imperfect learning behavior are Mookherjee and Sopher (1994), Cheung and Friedman (1997), and Camerer and Ho (1999).

### 7.1.3 Investor Reaction to Shocks

One sign of "learning while investing" in the stock market is the flow of capital in and out of the market following large price changes. Figure 7.1.1 above shows the reaction of "short sales" on stocks in the NYSE by year. Short sales are one indication of pessimism in the market, since the seller makes money in a down market. A rational investor in an efficient market with a diversified portfolio would not be changing her amount of short sales based on market fluctuations. But, as can be seen in Figure 7.1.1, short sales seem to mirror stock movements, with up years having fewer and down years having greater number of short sales. This indicates that either investors are initiating short sales before the downturns, and predicting the movements, or reacting to down movements with increased pessimism. Neither of these actions are consistent with our ideal of an efficient market.

When investing according to portfolio selection rules, one takes into account risks and acknowledges that the market can drop. The rational investor then diversifies her portfolio to balance out these drops. But consider the movement of assets out of mutual funds in down years (Kovaleski, 2002). The extreme action following a downturn implies that investors are surprised by the drop, learn from this new information, and change their investments accordingly. This reaction is not encompassed in any of the models we have studied. Rational investors should not be shocked or panicked when the market goes down, since they should have studied all they could beforehand.

We can also get an idea of investor reactions by exploring the risk premium required for the downside risk as compared to the upside risk. Figure 7.1.2 compares the relation between the CAPM beta and stock returns and our "down beta" and stock returns for the S&P 100 stocks in 1999. Both measures are correlated with the risk premium received by investors. There is a difference between the two graphs:

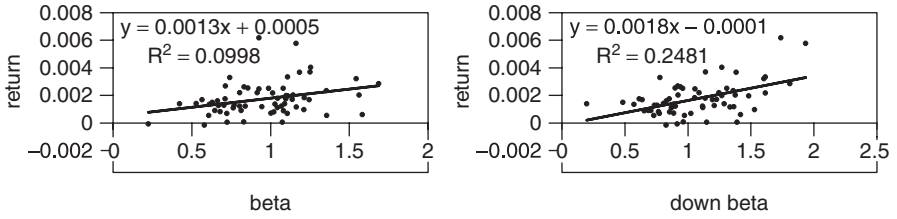



**Figure 7.1.2** Beta and down beta versus returns

1. The risk premium goes up much more quickly with down beta (0.0018 compared with 0.0013, or 38.4% higher), indicating that down beta is rewarded by a higher risk premium than the general beta that ignores the difference between upside and downside risk.
2. The $R^2$ measure of the goodness of fit of the regression of down beta is much higher than for beta, indicating that down beta fits risk premiums better.

By looking at downside risk, we have refined our measure of risk. Some of the noise and inaccuracy of the original CAPM beta is made up simply by compensating only for downside movements and focusing on the down beta. After all, upside movements are profits and do not need any compensation at all.

Another way to rationalize compensating downside movements more than upside movements is by revisiting the idea that only unexpected and non-diversifiable changes should receive a risk premium. In finance, however, good news is rarely unexpected. Companies simply do not hide positive news. In fact, as seen in Section 4.1.1, the Monday effect seen in the stock market, whereby Mondays receive lower returns than other days, has been attributed to companies timing their bad news releases after the weekend begins.

When companies have good news, breakthroughs, or potential increased earnings, they broadcast it as soon as possible. Company projections have been found to be eternally optimistic (Easterwood and Nutt, 1999; Espahbodi, Dugar, and Tehranian, 2001). Humans seem to have an innate need to sweep bad news under the carpet, however. Enron, Worldcom, Arthur Andersen, among others, highlight how companies can take initially small indiscretions and hold on to them until they compound to an astounding degree. This tells us that if we are looking for unexpected changes, downside movements are going to be a more useful and important source.

Brooks, Patel, and Su (2003) study the effect of unanticipated announcements on stock prices; they searched news resources for large announcements. It should be no surprise that all the news stories they focus on are bad news. Some of the keywords they use are "unexpected," "unanticipated," "surprised," and "shocked." They find that price reactions for unanticipated bad

news take over 20 minutes, much longer than for partially anticipated financial announcements. Another finding is that in the longer term, prices tend to reverse, indicating an initial overreaction.

The asymmetric reaction of investors can also be seen in the headlines following market trends. During the downturn of 2001 headlines pointed to the technology "bubble" and the market "revaluation." This terminology implies that the downturn is moving to a true value, and that everyone knew that the soaring market of the late 1990s was too good to be true.

During the recovery of 2003, however, the headlines were more cautionary, warning that the upswing could be a new bubble. While investors all want to be the first in on the next new thing, they tend not to trust high-flying stock prices when they happen. In terms of the main theme of this chapter, we show in this section more explicitly how to incorporate the downside in our risk assessments by recognizing the asymmetry of investor reactions. For example, we suggest using down beta instead of the usual CAPM beta and the striking behavior of data on short sales, which can contain potentially useful information about future downturns.

## 7.2 PATTERNS OF DOWNSIDE RISK

In the preceding section, we argued that downside risk should receive a higher risk premium than upside risk. In this section we will look at the source and magnitude of this difference. The first thing to observe is how close the traditional measures of downside risk are to the standard deviation proxy among our recommended measures of downside risk. There will, of course, be some correlation because traditional measures of risk include the downside as well as the upside. The degree of asymmetry, and general quirkiness of the return distribution, will determine how well traditional (general) measures of risk will follow downside measures.

It also may not always be the case that downside risk is higher than upside risk. For large stable corporations, there may be enough reserves and low-risk projects to weather any storm. Consider the case of IBM and Cisco Systems. Both firms are in the technology industry, but IBM is the elder and less volatile of the two. Taking monthly returns from 1993 to 2003 for both firms, their monthly returns are comparable: 2.26% for IBM and 3.04% for Cisco. Variance for Cisco, however, is much higher: 0.018 compared to 0.010. Value at risk for Cisco is also higher: –28.2% compared to –21.0%, assuming normality. Plotting the Pearson family for the two, however, adds to the picture.

As can be seen in Figure 7.2.1, the IBM is somewhat symmetrical, if not skewed positively, while Cisco is obviously skewed negatively. Both securities are Pearson's type IV distributions, and using Pearson's CDF gives a one percentile VaR of –18.8% for IBM and –32.3% for Cisco. So what we see is that not only does the assumption of normality understate VaR for Cisco by over 4%, but it actually overstates downside risk for IBM by over 2%. It is not

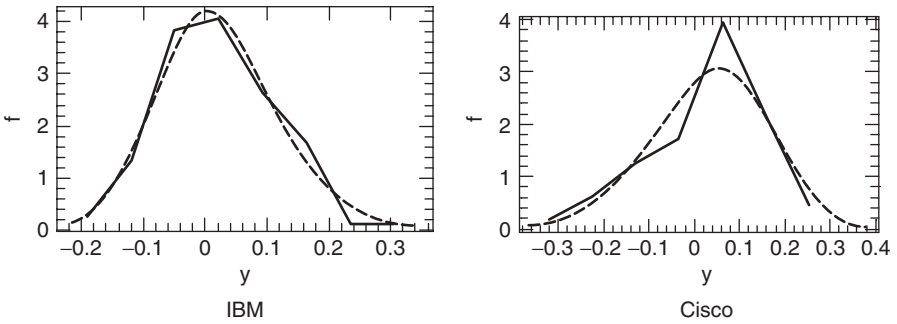



**Figure 7.2.1** Peasron's density plots for IBM and Cisco

**Table 7.2.1 Correlations between Volatility Measures and Downside Risk**

| | Next-Period Down Standard Deviation |
|---|---|
| Down beta | 0.52736 |
| Put-implied volatility | 0.566961 |
| Down standard deviation | 0.663702 |
| Beta | 0.356883 |
| Standard deviation | 0.374641 |

always bad news to look into the presence of downside risk, just as going to the doctor does not make you sick. For the case of Cisco, ignoring downside risk could be like hiding potential losses. For IBM ignoring downside risk would overstate risk, causing the investor to be unnecessarily conservative.

For our measures of downside risk, we will use contemporaneous VaR, down standard deviation, down beta, and put implied volatility. More important, however, we will look at how well these measures predict the next period's downside risk. Only one of those measures, put-implied volatility, is forward looking. Therefore we need to assess the ability of any measure to predict downside risk during the holding period of our stock.

Table 7.2.1 reports the correlation between various risk measures and the next-period down standard deviation. As can be seen, all of the downside measures predict the next-period down standard deviation better than the traditional general measures. For example the correlation between standard deviation and next period's down standard deviation is only 0.374641 compared to 0.663704 between consecutive down standard deviations. This indicates that asymmetry in volatility is persistent.

Since separating out downside risk requires additional work on our part, if traditional general measures of risk do just as well then we can save some

**Table 7.2.2 Correlation between Risk Measures and Their Downside Counterparts**

| Risk measure | Counterpart | Correlation |
|---|---|---|
| Standard deviation | Down standard deviation | 0.865589 |
| Beta | Down beta | 0.715209 |

efforts. As can be seen in Table 7.2.2, the correlation is reasonably high, but we have considered only some of the larger and more stable companies out there. Smaller companies that are still evolving may have larger asymmetries between upside and downside risk.

Predicting one-period ahead is yet another step removed for the traditional general measures of risk. For those of us that remember music on cassette tapes, this is the duplication problem. If copies are made from copies with magnetic media, the new copies tend to sound less like the original, and more static and distortion makes it into the mix. The same is true of estimation. In the first place, historical downside risk is used to predict future downside risk. Then the traditional general risk measures are used to predict the historical downside risk measures. Since we must use historical data anyway, it is better to use downside risk measures than traditional measures.

In the previous section we saw psychological reasons for why investors care about downside risk, and in previous chapters we saw statistical methods for measuring downside risk. However, is there a rational reason why an investor should care about downside risk? The only rational reason why downside risk should earn a higher risk premium would be that it is less diversifiable than upside risk. Pedersen (1998) finds that (for small companies in the United Kingdom) CAPM fails to explain portfolio selection, whereas downside measures provide a better selection criterion to fit the actions of investors.

To get an indication of how well we can diversify downside risk, we will replicate our exercise from Chapter 2 and graph down standard deviation as we add more stocks to our portfolio, as seen in Figure 7.2.2.

Note that the downside risk measure does not go away as quickly or as much as the traditional general risk measure. The standard deviation is reduced to a quarter of its initial level, while the down standard deviation only loses half of its original magnitude. This means that a systematic risk may be understated by looking only at traditional general risk measures.

When we look at CAPM and beta, we are getting an indication of how our stock moves with the market as a whole. Down beta denoted by $\beta_w$ is computed from a weighted CAPM regression giving nonzero weights only on the downside. It was used in (5.2.10) in the definition of the downside Treynor measure in order to give us a finer measure of how our stock moves with downturns in the market. Therefore the trend of the market can give us an indication of how large a role downside risk should play in our portfolio selection.

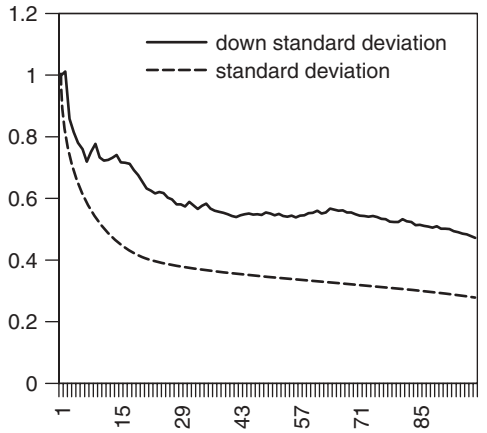



**Figure 7.2.2** Diversification as companies are added to a portfolio

## 7.3 DOWNSIDE RISK IN STOCK VALUATIONS AND WORLDWIDE INVESTING

This section first considers the stock market price valuations by comparing them with macroeconomic time series to look for signs of potential downturns. Since stock markets are known to cycle through boom and bust, it is plausible that seeds of a bust are included in the data during a market boom. We discuss a few such signs, without attempting to be exhaustive. The population in the developed world is aging and annual growth rates are stagnating compared to much younger population and high growth rates in some developing countries. Many investment advisers are recommending diversification of individual portfolios by investing abroad, including in developing countries. Hence in this section we discuss downside risk in foreign direct investments in much greater detail, and especially the notion of "home bias" and corruption.

### 7.3.1 Detecting Potential Downturns from Growth Rates of Cash Flows

Although everyone knows that sooner or later, hot markets must cool off and cool markets must heat up, no one knows when the pendulum will swing in the opposite direction. The S&P 500 stock index went up 37.4%, 22.9%, 33.4%, 28.6%, and 21% during 1995 to 1999 bull market. During this time the pendulum of the market kept going up. Since long-term gains from stocks were around 10%, such extraordinary or above-average gains were unsustainable in the year 2000. Then the market pendulum swung down and there began the bear market. The S&P 500 index went down 9.1%, 11.9%, and 22.1% during the three-year time period of 2000 to 2002. The market started the upswing of the pendulum in 2003. Thus market evaluations of some corporations are

sensitive to the timing of a boom or bust. The hard task is to identify seeds of bust in the data from the boom period.

Robert Hall's (2001) paper on struggling to understand the stock market provides an excellent review of the state of knowledge regarding market cycles in the context of macroeconomic indicators, such as GDP, productivity, capital stock, and intangible assets owned by corporations. He considers a long time horizon from 1947 to 2000 and compares the following ratio of market valuation to GDP:

$$\text{M2G} = \frac{\text{Equity claims on nonfarm nonfinancial corporations}}{\text{GDP}}. \quad (7.3.1)$$

Hall's graph shows that M2G ratio of market evaluation to GDP cycles over the boom and bust cycles of the market. It starts at 0.4 in 1947, goes to about 0.8 in late 1960s and falls back to around 0.4 in 1980; then it goes up over a long period through 1990s, and reaches over 1.4 at the start of the burst bubble in 2000. The true valuation of underlying real assets in relation to the GDP cannot possibly fluctuate between 0.4 to 1.4. This suggests that the stock market is not tied to production closely, and may thus be irrational.

Going deeper into the data, Hall (2001) derives further ratios, after including the effect of corporate debt, variations in hard asset values, intangibles, cash flows, and security values and discount rates. He argues that the collapse of stock market value by 50% in 1973 to 1974 can be attributed to a sudden reversal in cash flow growth. Similarly the enormous appreciation in 1990s was due to consistently high growth rates for cash flows of corporations. According to Hall, even the 2000 burst in the bubble is traced to sudden shortfall in the growth rate of cash flow.

### 7.3.2 Overoptimistic Consensus Forecasts by Analysts

Any company's data on past performance involve many factual series including gross earnings, EBITDA (earnings before interest, taxes, depreciation, and amortization), net profits, and its price-to-earnings ratio ($P/E$ ratio). Stock market research services, such as Value Line, publish reports on individual companies available at most public libraries. Yahoo finance and other places also provide such information on line. The S&P 500 stock index is often regarded as a leading indicator of market behavior. The long-term average of the $P/E$ ratio is about 15. When the $P/E$ ratio for any stock exceeds this long-term average, there is a risk that the stock is overpriced. In bull markets a higher than average $P/E$ ratio is often considered acceptable—not so in a bear market.

Price Earnings ($P/E$) ratios are a fundamental tool for assessing an overall evaluation of a stock. For example, in February 2004 the ($P/E$) ratio for the S&P 500 stocks was 18.2. Then the reciprocal of ($P/E$), known as market's earning yield, was 0.0549, which translates to about 5.5%. Stock market invest-

ment is thought to be for the long term and hence this yield is compared to 10-year US Treasury notes. In February, the treasuries were yielding only 4.09%. Using Federal Reserve bank's Monetary Policy Report to Congress dated July 1997 Edward Yardani developed so-called Fed model, suggesting that the market was over-valued when the reciprocal of the $(P/E)$ ratio exceeds the yield on 10-year treasury notes. Both earnings of corporations and treasury yields are in nominal dollars affected by inflation. Since treasury yields are highly correlated with inflation, the Fed model suggests that stock yields are also highly correlated with inflation. Although the Fed model has been a good descriptive model for the past 20 years, it also suggests that stock market participants are somewhat irrational (subject to inflation illusion) and pay too much attention to nominal rather than inflation-adjusted real interest rates. Modigliani and other economists argue that stock yields should be adjusted for inflation illusion, but do not agree on how to do the adjustment. The growth rate of real earnings of corporations is clearly related to growth in productivity. Fed chairman Alan Greenspan refers to the productivity growth due to the Internet and PC revolution in many sectors. A comprehensive model is not yet available. A related issue is whether stock market investing is a hedge against inflation. In times of relatively low inflation, it is believed to be a hedge, but it is also known that very large inflation rates hurt the stock prices.

We now list some common strategies for dealing with market risk without any particular attention to the downside risk:

1. A stop loss strategy says that the investor obeys some discipline to sell a stock on a declining trend when the loss is, say, 30% of the original price. A strategic percent like the 30% must be chosen before the outcome of the market prices is known. Otherwise, it is tempting to assess whether an individual stock has reached a bottom and is about to rebound. The stop loss strategies are designed for bear markets.

2. Realize gain strategy is the opposite of the stop loss. It asks the investor to sell a rising stock and realize the paper capital gain after a predetermined percentage (e.g., 200%) upside move. Once the gain is realized, the investor is asked to do the needed research so it is appropriate to buy at the new price. This strategy is designed for bull market.

3. Buy and hold strategy is often advisable for diversified portfolios similar to mutual funds or index funds. The underlying notions is to let the market go through its bull and bear cycle. The gains come because, even if some individual stocks never recover, the market as a whole usually does recover.

4. Dollar cost averaging strategy is to invest a fixed amount in a particular stock or mutual fund irrespective of market price fluctuations. That is not to try to time the market and invest more before the boom since these times are essentially unknowable. The fixed money buys more shares when the market is down, and less shares when the market is up, evening out their investment cost basis.

These are mostly passive strategies used by busy people with little or no interest in the stock market and by those with small holdings including orphans and widows. These strategies have proved to be quite effective for the long-term investors who happen to pick the right security. It appears that the first strategy called 'stop loss strategy' is specific for the downside risk.

### 7.3.3 Further Evidence Regarding Overoptimism of Analysts

The median rate of growth of earnings over a 48-year (1951–1988) period of publicly traded companies was about 6% per year. Chan et al. (2003) check the database looking for companies that exceeded the 6% mark consistently over any five-year period. They found that very few companies exceeded this modest benchmark and the ones that did were rarely the ones that research analysts had predicted. This means that very high $P/E$ ratios are very rarely justified. If we look at consensus forecasts of research analysts in 2004, more than 100 companies are expected to grow at an annual rate of over 40% over the next five years. This has simply never happened in the past data and shows that relying completely on consensus forecasts can lead to disappointing losses. Hence any measure of downside risk should allow for the existence of herd behavior leading to excessive optimism implicit in professional analysts' consensus estimates. We recommend a healthy grain of salt before believing all that the research analysts have to say in their so-called consensus estimates of future earnings.

A lesson from Hall's (2001) paper involving the aggregate macroeconomic data for the market as a whole is that we should look at the growth rates of cash flows to assess future stock market downturns. The aggregate macroeconomic results do not apply to individual stocks, but consensus forecasts by analysts do. The macroeconomic measure is not systematically biased and as an indicator of downside risk it has considerable economic theory behind it. We recommend using the macroeconomic measure to adjust overall exposure to equity markets and use analysts' recommendations for choosing among different equities. It is not clear if the growth rate of cash flow for individual stocks is a reliable indicator of its future price. There are additional purely statistical measures using long time series based on a study of past "turning points" in stock prices. These time series are obviously nonstationary, drifting upward, and there are not a large enough number of turning points in the available data to make valid statistical forecasting inference of nonstationary series.

### 7.3.4 Downside Risk in International Investing

If the stock market is efficient, then all movements in stock prices represent some new information, previously unknown and unexpected. While there is enough uncertainty in domestic capital markets, investing in international markets leads to new areas of uncertainty. Differences in legal systems,

accounting standards, and market oversight can lead to disasters that may not even be on the radar if one thinks according to domestic trading rules.

### 7.3.5 Legal Loophole in Russia

The *Financial Times* of July 25, 2000, reports a legal loophole in the Russian capital market that can leave stockholders with nothing if the company has a cash flow problem. In the United States, if a company is not able to make debt payments, the company can first try to work out the payments privately. If that does not work, then the company enters bankruptcy proceedings, where the legal system decides the allocation of the companies remaining assets. The US legal system, with its large army of expensive lawyers, is often not seen as the most efficient method of resolving conflict, but one can see some benefits when compared to the risky Russian system, which fails to protect investors (who can afford lawyers).

The Russian system reported on in the *Financial Times* article, "Russian Oligarchs Take the Bankruptcy Route to Expansion," is a simpler bankruptcy proceeding. When a Russian company cannot make debt payments, the debt-holders take the company to court and receive ownership of the company. Even in viable, profitable companies, if the cash flow is not there to make the debt payments, stockholders can end up with nothing.

### 7.3.6 Currency Devaluation Risk

One major concern when investing internationally is the value of the foreign currency when the stock is sold. No matter how well your stock has done in the past year, if the currency it is valued in suffers a depreciation of 50%, your position is suddenly worth half of its previous value overnight. Furthermore, if the currency is worth less, it does not matter which company you own, or how diversified your portfolio is, all stocks will lose value.

The Chinese stock market initially solved the currency issue by having a special class of stock for foreign investors denominated in dollars. The Chinese stock market was plagued by other problems, though, such as the small number of dollar-denominated shares available and the large portion of major corporations owned by the government. Currently the two-class system of shares is breaking down. Chinese citizens are now allowed to buy the dollar-denominated shares, and foreign investors are beginning to be allowed to trade in the domestic shares. The uncertainty of what accounting standards mean in a foreign country still make foreign investment merely trickle, compared to its potential in offering diversification against local market risk.

Some countries such as India use derivatives markets to smooth out foreign investor's risk. Futures and options markets can be used to hedge currency and downside risk. However, if expanded derivatives markets are not complemented by increased market oversight, the speculation or many other potential abuses can bring down the market, leading to obvious downside risk.

The downside dangers of currency devaluation had led to a wide literature trying to deal with it and predict future currency crises. Since currency values are determined at the macroeconomic level, one must trace the supply and demand for a country's currency to look for warning indicators that the currency is becoming overvalued. In international trade there are two accounts to consider: (1) the current account, which accounts for imports and exports of goods and services, and (2) the capital account, which includes financial transactions accounting for investments flowing into and out of the country. Once both accounts are considered, by definition, the two sides of the combined accounts must match. However, the underlying supply and demand for a country's exports are always changing over time and so is the demand for currencies. If the economic forces behind supply and demand and transfer of financial investments are out of balance, the value of a currency's currency will change to restore the equilibrium. If currency values are artificially manipulated by central banks, currency markets can be pushed out of equilibrium.

### 7.3.7 Forecast of Currency Devaluations

Some of the common indicators that are used to predict a currency crisis (see Salvatore, 1999; Salvatore and Reagle, 2000) come from these macroeconomic indicators:

1. A country's savings rate to determine if domestic capital is sufficient for investment.
2. Current account deficit to determine the pressure on currency.
3. Foreign debt (particularly short-term foreign debt) to predict future currency pressure when interest and capital payments are made.
4. Budget deficit as an indicator of government borrowing, either crowding out domestic borrowing or leading to foreign debt.
5. International reserves indicate how long a government can defend a temporary imbalance in the exchange rate.
6. Foreign direct investment versus portfolio investment. This gives an indication of how quickly capital can leave a market. Foreign direct investment is a relatively permanent source of investment that entails a foreign investor purchasing a substantial portion of a company directly and taking on an (equity) ownership role. Portfolio investment is simply a foreign investor purchasing securities though the public market. Portfolio investment in debt, rather than equity, is much quicker to purchase but can flee a country just as quickly if there is trouble, creating further selling pressure on the currency. These trigger contagions.
7. Currency contagion in recent history. Kaminsky et al. (2003) have an excellent table for explaining various such episodes. They identify the common characteristics across several countries and the name of the

common creditor country lending the affected countries. Their term is "unholy trinity" for the capital inflows, surprises, and common creditors responsible for currency crises.

These indicators are just a few of the many offered to predict currency crises. Kaminsky, Lizondo, and Reinhart (1998) compile an exhaustive list of indicators to predict currency and banking crises in emerging economies. These indicators are not surprising, since they arise from common macroeconomic data. Yet currency crises continue to come as a surprise. Take, for instance, the Asian "flu" of the late 1990s. Southeast Asia was the hottest market going in the early 1990s. Their manufacturing industry was booming and creating exports; the capital account demand for their bonds was high. These strengths deteriorated as labor costs went up, and their currency pegged to the dollar became overvalued as the dollar gained strength. Eventually these pressures led to a currency devaluation that left the financial market shattered. Why didn't anyone see it coming?

### 7.3.8 Time Lags and Data Revisions

One difficulty in using macroeconomic indicators is the time lag in the collection of the data. Even in the United States, we don't know that a recession has happened until almost six months later. Macroeconomic data for developing countries reported by the World Bank or the IMF usually come after a minimum two-year lag. These lags mean that if one is going to predict a crisis with these data, the crisis must be correctly predicted two years out-of-sample to be really useful. The underlying data are frequently revised so that values of the macroeconomic indicators, which looked dangerous one year, may turn out to be quite benign just as a result of official revisions. Similarly new information arrives from time to time, which can change the prospects for a country very significantly (Santangelo 2004).

Consider Figure 7.3.1, which graphs the number of indicators (out of six) in Southeast Asia that were in the danger zone according to Salvatore and Reagle (2000). The upper line is calculated with data from the World Bank's World Development Indicators for 2001. It clearly shows an increase in warning indicators prior to the crisis. The lower line, however, is the exact same data source, the same indicators, the same methodology, except that it was collected from the 2003 edition of the World Development Indicators. After the data were revised, there is no longer the same run-up in warning indicators to predict the crisis.

Even with perfect data most macroeconomic indicators are collected annually or quarterly at best. Most monthly data have been found to not aid in prediction. Therefore crisis prediction can never be fine-tuned. At best, these indicators show country locations that may be susceptible for crisis but cannot pinpoint the timing, depth, or other specifics of a particular currency crisis.

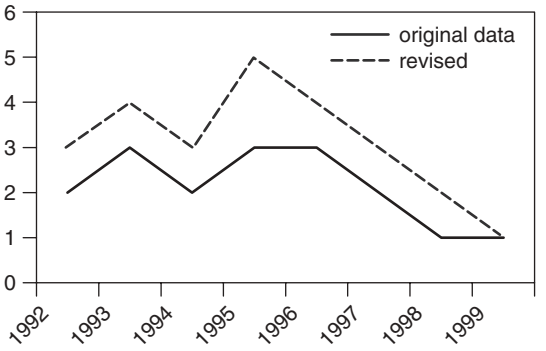



**Figure 7.3.1** East Asia crisis indicators before and after data revision

### 7.3.9 Government Interventions Make Forecasting Currency Markets Difficult

As with any forecasting, the historical data used to predict future currency movements can be problematic. These predictions are particularly difficult in currency markets, since the timing of currency market devaluation is closely controlled by central banks and is subject to local political considerations. Some dollar-zone countries simply use the US dollar as their currency, and do not permit much control by the local central banks. Many countries have target zones, and when currency fluctuates outside these zones, central banks intervene and devaluate or revaluate. Historical data are good for estimating an average level or average movement in currency values, but currency crises are an atypical occurrence. Predicting when a government will decide not to defend the currency any longer is a task that has little history to go by. Usually currency crises come one at a time (Mexico, 1994, and Brazil, 1998) or spread out in time under different regimes (five countries in Asia, 1997, pegged to the dollar). The minimum for any type of good statistical estimation is availability of 30 observations. This means that one has to go back decades to find records of past 30 currency crises in order to forecast future crises.

Even if we had such extensive data, each currency crisis is different. Mexico had a large budget deficit and a low savings rate. Asia, on the other hand, had a high savings rate but a large current account deficit and high foreign debt. Some countries can tolerate budget deficits better than others. A longstanding deficit that is compensated for in another area is more sustainable than a deficit that increases dramatically in a short period of time. To further complicate matters, contagion of the crisis can cause a previous healthy country to go into crisis simply because it is linked economically or geographically to another country with more serious problems.

This leads to a good news and bad news dichotomy regarding any method of predicting currency crises. The good news is that the indicators of a crisis

are easily identified. The bad news is that the measurement of the indicators is not fine enough to make a precise prediction. The best that can be done is to use extreme values and changes in these indicators as a sign that downside risk is higher in that country's financial markets. In any case we have indicated some tools and strategies available in the literature and outlined almost all risks in international investing, except for what is discussed in the next section.

## 7.4 DOWNSIDE RISK ARISING FROM FRAUD, CORRUPTION, AND INTERNATIONAL CONTAGION

This chapter has considered the downside risk from various angles including some additional risks associated with worldwide investing. In Section 7.3 we analyzed how international financial crises are caused by extreme movements in macroeconomic variables to unsustainable levels. No one was blamed; it just looked as if the economy got out of equilibrium and needed an adjustment to get back. In this section we assign some blame. When one considers incorporating downside risk, it is important to recognize that fraud and criminal behavior are a fact of life in any country and hence vigilance is needed everywhere. If a country takes no action to correct problems when they are discovered, the leadership must be blamed for incompetence or worse. Once the recent abuses by Enron, WorldCom, Tyco, and others, were recognized, the Securities and Exchange Commission (SEC), and other regulators and lawmakers, attempted to protect the individual investor from such abuses. For example, the Sarbanes-Oxley Act of 2002 is an excellent example of new rules on corporate governance. It is always difficult to strike a right balance between regulation and freedom to entrepreneurs to create wealth. In our opinion, there are provisions in Sarbanes-Oxley Act that deserve worldwide emulation. Clearly, the downside risk is lower in countries that follow proper international accounting standards and have mechanisms in place to prevent abuses by managers.

Now consider the role of fraud and corruption in the currency market contagions. For example, the Asian currency crisis started with the collapse of the Thai Baht in 1997. As we saw, for three decades before 1997, East Asian economies grew remarkably fast due to (1) low inflation and high investment in human capital, (2) high savings rate and protection for individual savers, (3) few price distortions and imbalances between agricultural and industrial sectors, and (4) adoption of foreign technology.

There were certainly some *structural problems* behind the financial crisis of 1997, including (1) tendency to finance long-term debt with short-term paper due to a lack of well-developed derivatives market, (2) inadequate attention to social safety net for the unemployed and the poor during boom times, (3) excess exploitation of natural resources as forests, fisheries, and so on, and (4) poor regulation of domestic capital markets that did not check global push toward a financial bubble. The global capital flows grew too fast (30%

annually), and foreign investors' ignorance of local conditions and greed for high returns were important. The investors expected to be bailed out (moral hazard) if things went wrong. The Asian contagion was fueled by slowing of export growth followed by competitive devaluations, and by strong trade and financial links among the East Asian countries. In addition, crony and corrupt capitalism played a critical role in worsening all the structural problems listed above.

Financial crises are characterized by (1) falling growth rates of real GDP, (2) large fluctuations of inflation rate, credit expansion rate, and capital inflow rate, (3) sharp declines in exchange rates, (4) adverse shock in international trade, and (5) sharp rise in real interest rate (Hardy and Pazarbasioglu, 1999). Vinod (1999) indicates evidence showing that corruption reduces economic growth and discourages savings. This creates a fundamental link between corruption and financial crises, since savings was one of our key indicators for crisis. Clearly, a reduced saving rate can lead to higher price in the form of higher interest rates. Governmental waste caused by corruption and inefficient policies for fighting inflation can lead to large swings in inflation rate and exchange rate. If bribe demands lead to interruption or cancellation of major international trade transactions, they can create trade shocks. Thus corruption contributes to each of the factors behind the Asian crises and behind financial crises in general. Vinod (2003) in the *Journal of Asian Economics* explains how the financial burden of corruption exists in an open economy beyond the effects of reduced savings and reduced capital formation on the domestic front.

### 7.4.1 Role of Periodic Portfolio Rebalancing

Once a currency crisis begins in one country, portfolio *rebalancing* by investors in response to a loss can cause ripples that can bring down otherwise healthy economies. We discuss some portfolio theory regarding risk and hedging in Section 7.4.2. If two countries in the same geographical region are correlated, and an adverse shock causes us to lower our investment in one, it makes sense that investment in the correlated country will also decrease. Schinasi and Smith (2000) note that such behavior is particularly strong if the investor is leveraged (uses borrowed funds) and that several traditional portfolio-rebalancing rules, including VaR rules, can induce such behavior. Of course, rebalancing is not a bad move if one updates his perception of downside risk, but the effects are multiplied by each investor that does so. Chang and Velasco (1998) attribute the contagion to international banking illiquidity and self-fulfilling pessimism, both of which are made worse by corruption.

### 7.4.2 Role of International Financial Institutions in Fighting Contagions

The IMF, the World Bank, and major multinational banks try to fight financial contagion before it occurs. They have begun to focus on removing the distress

factors, including corruption. To promote worldwide prosperity, the IMF identified "promoting good governance in all its aspects, including ensuring the rule of law, improving the efficiency and accountability of the public sector, and tackling corruption" (September, 1996). The World Bank also has indicated similar initiatives against corruption. More recently the IMF is working to eliminate the opportunity for bribery, corruption, and fraudulent activity in the management of public resources. The IMF directives seek to limit the scope for ad hoc decision making, for rent seeking, and for undesirable preferential treatment of individuals or organizations. However, the IMF needs to be more proactive against misuse of official foreign exchange reserves, abuse of power by bank supervisors, and similar areas where information should be available at the IMF. However, unless consensus is developed on the two fronts mentioned above, strong action by IMF or World Bank employees against corrupt practices cannot be expected. After all, these can be sensitive issues of local laws and personal safety of World Bank/IMF staff and their families.

### 7.4.3 Corruption and Slow Growth of Derivative Markets

Very large amounts of money (over a trillion US dollars) transfer between various countries on a daily basis. These short-term funds are called "hot money" transfers. Since these transfers create new kinds of risks, financial institutions need new tools (e.g., interest swaps) for managing them. Since these tools are not widely available throughout the developing world, currency market contagions become more widespread than they need to be. Below we show that corruption slows the growth of these tools and thereby hurts developing countries.

Consider a common situation where the hot money lenders want to lend only for a short term, say, 1 year, while the borrowers want to borrow for, say, 10 years. The market needs a way to resolve the mismatch between demand and supply, which is traditionally done by overnight loans and commercial paper. In developed countries where the newer market for interest rate derivatives is available, the solution to the mismatch mentioned above is simply to float 10-year paper with fixed rate and swap it into a floating interest rate. The market for swaps rewards the participants for assuming the risk associated with the difference between fixed and floating rate. Since this eliminates interest rate risk for those who wish to eliminate it (for a price), the global market for swaps is near five trillion dollars. Moreover, since the *interest swaps* have a market value reflecting the interest rate risk, both assets and liabilities move up and down correctly. While derivatives may contribute to greater volatility of markets, they do create new avenues for people interested in different forms of risk taking (gambling?). For sophisticated players they provide an opportunity to keep the cost of borrowing low.

Thinness of the market, lack of scale economies, lack of trust in local financial institutions and corruption are among the reasons for inadequate

derivative markets in developing countries. If Mr. Harshad Mehta in India can bribe bank officials to defraud thousands of Indian investors in the standard stock market in 2001, the potential for fraudulent manipulation looms even larger in thin derivative markets. Thus corruption risk is difficult to manage because corruption itself prevents access to the modern tools of managing this risk by preventing growth of derivative markets. Accordingly foreign direct investment (FDI) is reduced in corrupt developing countries due to the uncertainty of receiving the returns and keeping ownership of capital.

The derivative markets have focused on three different types of risk. (1) Credit risk refers to the ability of the borrower to generate revenue to pay back the debt. (2) Default risk is with reference to collecting when default occurs. (3) Transaction risk arises from possible problems with international electronic transfer of funds and enforcement of foreign exchange contracts. Although corruption risk is currently considered as a part each of these three, it may be helpful to have a separate category for corruption. La Porta et al. (1998) find that common-law countries (United States, United Kingdom, and India) provide superior legal protection to investors compared to French civil law countries. However, the corruption perception index (CPI) data from Transparency International suggests that common-law countries providing good legal protection of investors do not enjoy generally less corruption. France and her former colonies do not, in general, have any greater corruption than the United Kingdom and her former colonies.

The downside risk associated with international hot money transfers can be incorporated or managed to some extent by using derivatives markets. However, recent currency contagions have shown that corruption imposes a somewhat unique type of risk burden with a distinct probability distribution than for the risk associated with currency fluctuations, taxes, and investments. Along with sophisticated derivatives, we also need greater accountability, better surveillance to avoid abuses of these tools, and reduced corruption to avoid future currency market contagions.

### 7.4.4 Corruption and Private Credit Rating Agencies

Private rating agencies, which need to maintain their reputation by making right calls, can be a positive force against corruption. Their term "country risk" refers to creditworthiness of sovereign governments who have a larger leeway in structuring debts and payment schedules than private businesses. However, global money traders (hot money transfer agents) gather background information from rating agencies (S&P and Moody's) regarding actual deficits in different countries. Since rating agencies can potentially negate domestic fiscal policies of sovereign countries, policy makers often resent the power of money traders and such rating agencies. On the other hand, this creates a balance of power and forces some fiscal discipline on various countries without involving the IMF. For example, Thailand took on debt to build super highways and was punished for it by declining Thai Baht. In short, the positive contribution

of rating agencies and hot money is that they create a countervailing power and they discourage corruption.

We must also mention the negative role of hot money transfers when the investors rely on private rating agencies. A trillion dollars going in and out of different countries daily can obviously increase the volatility of markets. Changes in credit ratings lead to herd behavior by investors, even if the ratings themselves are in response to valid new information. The herd behavior creates intrinsically self-fulfilling short-run instabilities. When ratings go down, cost of borrowing for that country increases, precisely at the time when a country in trouble can least afford it. Sometimes investors lose money even before the country ratings are lowered. If so, they try to sell investments in other similar countries (same geographic region or similarly situated). Investors need only a limited amount of information about impending trouble, not detailed information. This can create a rush to beat the rating agencies. In any case, pessimism is self-fulfilling and can lead to further declines in balance of payments.

On balance and in the long run, rating agencies play a mildly positive role by inviting market discipline and thereby reducing corruption. Since markets treat everyone equally, market discipline is politically more bearable than edicts from IMF bureaucrats, who can sometimes make wrong policy recommendations. Since the IMF has little expertise in tax deficits, corruption, and weaknesses of local banks in all countries, the IMF did make some wrong calls in response to the Asian contagion.

### 7.4.5 Corruption and the Home Bias

When the presence of international trade is introduced in a traditional study of the trade-off between risk and return, we note that the trade-off does not smoothly extend to investments abroad. There appears to be a market failure, since investors in developed countries overwhelmingly exhibit a so-called home bias. More important, the impact of home bias is asymmetric with respect to rich and poor countries. The underlying incentives encourages flight of capital out of poor and corrupt countries. Hence policy makers in poor countries often impose capital controls to prevent flight of capital, which in turn encourages inefficient domestic monopolies and corruption.

French and Poterba (1991) were the first to discuss home bias and estimate its size. They assumed that a representative agent in each country has one of the simplest exponential decay constant relative risk aversion (CRRA) utility functions. We discussed it in Section 6.1. Here it is defined with respect to wealth $W$ as the argument (instead of consumption or returns)

$$U(W) = -\exp\left(-\lambda \frac{W}{W_0}\right). \tag{7.4.1}$$

where $W_0$ is initial wealth. The wealth increases by investing in portfolios consisting of possibly risky international assets. Each portfoilio is defined by a

vector $w$ of weights. If the probability distribution of asset returns is normal, only the mean and variance matter. Denote the mean vector by $\mu$ and covariance matrix by $V$. Now the expected utility is

$$E(U) = -\exp(-\lambda\{w\mu - 0.5\lambda w'Vw\}). \tag{7.4.2}$$

By expected utility theory (EUT), maximization of (7.4.2) yields a first-order condition involving a simple analytic formula. If $w^*$ denotes optimal portfolio weights, we have the necessary condition:

$$\mu = \lambda w^{*\prime} V. \tag{7.4.3}$$

French and Poterba bypass the need for historical data on international equity returns $\mu$ and compute estimates of $V$ for the United States, Japan, the United Kingdom, France, Germany, and Canada. They ask what set of "optimal" expected equity returns $\mu^*$ will justify the observed pattern of international holdings. They find, for example, that UK investors need about 500 basis points higher return in domestic market to justify not investing in US stocks. In general, investors in developed countries hold too high a proportion of their portfolio in domestic securities (94% for the United States and in excess of 85% for the United Kingdom and Japan). French and Poterba (1991) argue that within the set of six rich countries, there are few institutional barriers against international investments. Yet the "puzzle" lies in the empirically observed lack of adequate international diversification. Baxter and Jermann (1997) argue that the puzzle is "worse than you think." If we consider that all of our own human capital is concentrated in the home country, hedging will require us to invest even a larger proportion of our wealth abroad.

***Theoretical Explanations for Home Bias.*** How do we solve the puzzle or explain the home bias? Let us first consider theoretical explanations and then institutional ones. The CRRA utility function (7.4.1) used by French and Poterba is unrealistic. In the consumption context, Carroll and Kimball (1996) and Vinod (1998), among others, argue that CRRA utility should be rejected because it does not lead to a concave consumption function, as it should. These authors propose using a hyperbolic or diminishing absolute risk aversion (HyDARA) utility function. Another way of stating this issue is that CRRA is unrealistic because it gives inadequate importance to uncertainty.

A second theoretical explanation of home bias is that EUT assumed above is unrealistic. A need for non-EUT models is explained in the survey by Starmer (2000). Vinod (2001) considers risk aversion, stochastic dominance, and non-EUT in the context of portfolio theory. These modifications lead to a revision of (7.4.3) involving a nonlinear transformation of weights for greater realism. A third theoretical explanation is that the model lumps all risk in the variance term based on the distribution of returns. Since the currency devaluation risk applies only for foreign investments, it can be large enough to

cancel the hedging term. A fourth theoretical explanation of home bias is that it is derived by assuming that return distribution is normal. Nonnormal distributions require skewness, kurtosis, and higher moments that, when included, can reduce the home bias.

***Institutional Explanations for Home Bias.*** Now we note some institutional explanations for home bias in investments. First, investing in home country avoids the risk associated with enforcement of *property rights* in a foreign country. Even if no language barrier exists and the property laws are essentially similar, such as between the United States and the United Kingdom, enforcement costs can be large and thus a source of downside risk. The observed difference of 500 basis points noted by French and Poterba (1991) may be reflecting such costs including attorney fees. Second, corruption levels can be different in different countries at different times for different industrial sectors. This can mean uncertainty regarding size of bribes and financial and time costs of getting things done abroad can be larger.

Serrat (2001) develops a two-country exchange economy dynamic equilibrium model to attribute the home bias to the role of nontraded goods. He shows that risk-adjusted expected growth rates of home and foreign endowments of tradable and nontradable goods are relevant. Since the mean and variance of home and foreign endowments of tradables need not be constant, hedging demands causing home bias become less important in their dynamic setting with intertemporal substitution. In corrupt countries the hedging motive is likely to retain the home bias if the variance of home endowments is larger.

Corruption cannot fully explain both home bias and its asymmetry. Institutional explanations of the home bias do not apply symmetrically for investors in rich and poor countries. With lower corruption and better legal and insurance protections for small investors in rich countries, the investors in poor countries are not as reluctant to invest abroad; namely they are less subject to home bias. It is better to think of the home bias puzzle not as something to be solved but as a pedagogical device to learn economic behavior. A similar use of six consumption puzzles is advocated in Vinod's (1996) appendix. The useful insight from the home bias puzzle is summarized as follows: Other things remaining the same, all investors in different countries will be generally better off in terms of diversification and hedging if they invest a large proportion of their wealth in foreign assets. In practice, investors exhibit home bias. Various theoretical generalizations of the model and institutional barriers partially explain the home bias. However, the bias against investment in poor and corrupt countries is justifiably stronger than the bias against investing in rich and less corrupt countries due to the downside potential of corruption.

Another institutional explanation for home bias comes from policy decisions by central banks of developing countries. From the viewpoint of the developing country, it is desirable to let as much of the domestic savings fund

domestic investments. Accordingly developing countries often impose controls on the outflow of capital to prevent flight of scarce capital resources to foreign countries, thereby helping domestic capital accumulation. Clearly, the government can readily identify the domestic resident investor and punish any export of capital by outlawing capital account transfers.

Ironically, various exchange controls imposed by central banks themselves can encourage corruption and reduce "economic freedom" of entrepreneurs to create wealth around the world. Often, the first immediate effect of controls on out-going capital is to discourage in-coming investment in the form of FDI. The policy makers in developing countries have to consider quantitative information on the extent of the following related issues: (1) Can the domestic saver illegally invest abroad anyway? (2) Do these exchange controls influence the foreign nonresident owner of capital? (3) Are the investments in the form of equity or debt? (4) Do the controls encourage corruption and encourage inefficient allocation of resources? (5) If domestic industries rely solely on domestic capital, do they strive to become competitive in global marketplace? (6) Are the foreign investors likely to be too fickle with a very short-term focus? (7) Does the international flow of funds (hot money) destabilize the exchange rate?

### 7.4.6 Value at Risk (VaR) Calculations Worsen Effect of Corruption

As we learned in Chapter 2, value at risk has become a popular tool for practical portfolio choices. Here we consider the role of VaR methods in discouraging FDI in corrupt countries. Recall that VaR computations require estimation of low quantiles in the portfolio return distributions. Recall from Chapter 6 that expected utility theory (EUT) leads to equal weight on all returns in computation of risks, but the weights are high on extreme observations when an investor behaves according to the non-EUT. This means that worst-case scenarios are more important to a non-EUT investor leading to greater risk assigned to perceived costs of fraud and corruption.

In our context many authors have used VaR calculations of market risk and credit risk under the assumption of stable Pareto distribution. Since stable Pareto or related distributions do not have closed-form expressions for the density and distribution functions, they are described by their characteristic functions and by four parameters: tail index $\tau$, skewness $\beta$, location $\mu$, and scale $\sigma$. Modeling with such parameters helps depict fat tails and skewness of the distributions. Omran (2001) estimates that the tail index $\tau$ for Japan, Singapore, and Hong Kong markets are around 1.50, indicating that these markets have high probability of large returns.

A linear combination of independent stable (or jointly stable) random variables with tail index $\tau$ is again a stable random variable with the same $\tau$. Hence any stable random variable can be decomposed into the "symmetry" and "skewness" parts. If $\sigma$ is the volatility, $X^* = (z_{1-\alpha'} \sigma)$ is the limiting return at a

given confidence level $\alpha'$, and if $Y_0$ denotes the initial value of the portfolio, then value at risk is $\text{VaR} = -Y_0X^*$. Thus modeling value at risk (VaR) by stable Pareto distributions permits convenient decomposition of the distribution into three parts: the mean or centering part, skewness part, and dependence (autocorrelation) structure (Rachev et al., 2001).

Longin and Solnik (2001) extend stable Pareto to generalized Pareto and show, with examples, that cross-country equity market correlations increase in volatile times and become particularly important in bear markets. These correlations are neither constant over time nor symmetric with respect to the bull and bear markets. This means that one must reject multivariate normal as well as a multivariate econometric tool called GARCH with time-varying volatility. The estimation of VaR is related to estimating the worst-case scenario in terms of the probability of very large losses in excess of a threshold θ. The so-called positive θ-exceedances correspond to all observed losses that exceed θ (e.g., 10%). Login and Solnik show that the generalized Pareto distribution is a suitable probability distribution, and estimate its parameters using 38 years of monthly data. As θ increases, the correlation across markets of large losses does not converge to zero but *increases*. This is ominous for an investor who seeks to diversify across various countries, including corrupt developing countries. It means that losses in one country will not cancel with gains in another country. Rather, the data show that all similar countries might suffer extreme losses at the same time. This ominous observation is important to remember in incorporating the downside risk in investment decisions, especially for small investors.

Now consider a large international investor whose portfolio consists of many assets. The computation of VaR for any portfolio is computed by decomposing it into "building blocks" that depend on some risk factors. For example, currency fluctuation and corruption are risk factors with open economy international investments. Risk professionals first use risk categories for detailed separate analysis. They need complicated algorithms to obtain total portfolio risk by aggregating risk factors and their correlations.

The main risk in foreign investment is that exchange rates (currency values) fluctuate over time and can mean a loss when the return is converted into the investor's home currency. The individual investor would need to allow for potentially unfavorable timing of currency conversion. However, financial markets have derivative instruments including forward and future exchange rate markets to hedge against such risks. Hence derivative securities linked to exchange rates at a future date can mitigate, if not eliminate, the exchange rate risk. Arbitrage activities by traders can be expected to price the foreign investments appropriately different from domestic investments to take account of exchange rate risk. However, these hedging activities need free and open markets in target currencies. For developing countries like India, which have exchange control, there are black markets with a fluctuating premium over the official exchange rate. This increases the exchange rate risk and related costs even higher due to corruption.

Fraud and corruption are additional risk factors in both domestic and international investing, although our discussion focused much more on the latter. Corruption can suddenly lead to a cancellation of a contract or sudden and unexpected increases in the cost of doing business. Depending on the magnitude of corrupt practices, the cost can vary considerably. Enforcing property rights, especially in developing countries, can be expensive and time-consuming. Furthermore the presence of corruption and lack of transparency among public institutions add to the cost of investing abroad, even if the foreign investor is not directly affected by corruption. Since corruption increases the cost of enforcement of all property rights, it is obviously an additional burden.

We conclude this chapter by noting that fraud, corruption, and international contagion in currency markets increase the downside risk. Whether we are considering a large or small investor, empirical data and VaR calculations correctly warn us against large simultaneous losses in different countries that need not cancel each other. Since downside risks are high in corrupt countries, this implies a need for a higher compensation (risk premium). When interest rates are low in Western European countries, the United States, and Japan, it is tempting to diversify and invest in some fast-growing developing countries. The expected return from these foreign investments needs to be much higher to compensate for the probability of downside risk arising from unpredictable corruption.

CHAPTER 8

# Mathematical Techniques

## 8.1 MATRIX ALGEBRA

This chapter discusses six mathematical tools of special importance in finance, starting with matrix algebra, and moving to its application to finance. The latter sections highlight specialized mathematical techniques for solving some of the formulas discussed.

### 8.1.1 Sum, Product, and Transpose of Matrices

Matrix algebra is a mathematical technique that organizes the information of several equations so that a solution can be obtained for complex systems that cannot efficiently be solved otherwise. In this section we go over the rules of matrix algebra, which at times are different, or at least more strict, than general algebraic techniques. We then apply these techniques to the portfolio problem where the sheer numbers of securities make it necessary use matrix algebra.

If $a_{ij}$ with $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, n$, are a set of real numbers arranged in $m$ rows and $n$ columns, we have an $m$ by $n$ matrix $A = (a_{ij})$. For example, when $m = 2$ and $n = 3$, we have a $2 \times 3$ matrix $A$ defined as

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} 15 & -12 & 53 \\ 25 & 22 & -2 \end{bmatrix}. \tag{8.1.1}$$

If in the same example we have three rows instead of two, the matrix will be a square matrix. A square matrix has as many rows as columns, or $m = n$. Vectors are matrices with only one column. Ordinary numbers can also be thought to be matrices of dimension $1 \times 1$, and are called scalars. Similar to

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

most authors we use only column vectors, unless stated otherwise. Both vectors and matrices follow some simple rules of algebra.

1. Sum. If $A = (a_{ij})$ and $B = (b_{ij})$, then $C = (c_{ij})$ is defined by $c_{ij} = a_{ij} + b_{ij}$. This is defined only when both $A$ and $B$ are of exactly the same dimension, and obtained by simply adding the corresponding terms. For example, $A + B$ is a matrix of all zeros when $A$ is given by (8.1.1) and

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} = \begin{bmatrix} -15 & 12 & -53 \\ -25 & -22 & 2 \end{bmatrix}. \tag{8.1.2}$$

2. Product. If $A = (a_{ij})$ is $m \times n$ and $B = (b_{ij})$ has $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, p$, then the product matrix $C = (c_{ij})$ is of dimension $m \times p$, and is obtained by "row-column" multiplication as follows: $c_{ij} = \Sigma_{t=1}^{n} a_{it} b_{tj}$, where $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, p$. Note that $a_{it}$ represent the elements of $i$th row and $b_{tj}$ represent elements of $j$th column. Therefore the product above can be defined only when the range of $t$ is the same for both matrices $A$ and $B$, meaning $A$ should have as many columns as $B$ has rows. An easy way to remember this is that for our "row-column" multiplication, "column-rows" must match. For example, the product $AB$ does not exist or is not defined for $A$ and $B$ defined by (8.1.1) and (8.1.2). Since $A$ has 3 columns $B$ must have exactly 3 rows and any number of columns for the product $AB$ to exist.

In general, we do not expect $AB = BA$. In fact $BA$ may not even exist even when $AB$ is well defined. In the rare case when $AB = BA$, we say that matrices $A$ and $B$ commute.

3. Multiplication of a matrix by a scalar element. If $h$ is an element and $A$ is a matrix as above, $hA = Ah = C$ defined by $C = (c_{ij})$, where $c_{ij} = ha_{ij}$. So, to obtain the matrix $C = 2A$, simply double the value in each element of matrix $A$.

4. Multiplication of a matrix by a vector. If $A$ is $m \times n$ as above, $x = (x_1, x_2, \ldots, x_n)$ is a $1 \times n$ row vector. Now we simply treat multiplication of a matrix by a vector as a product of two matrices above. For example $Ax = c$ is an $m \times 1$ column vector.

5. Rules for sum and product of more than two matrices:

a. Associativity: $A + B + C = A + (B + C) = (A + B) + C$, or $ABC = A(BC) = (AB)C$.

b. Distributivity: $A(B + C) = AB + AC$, and $(B + C)A = BA + CA$.

c. Identity: If $Om,n$ is an $m \times n$ matrix of all zeroes, then $A + Om,n = A$. If $I_n$ is an $n \times n$ matrix $(\delta_{ij})$, where $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ otherwise, it is called the identity matrix, $AI_n = A$. Note that cancellation and simplification of matrices in complicated matrix expressions is usually obtained by making some expressions equal to the identity matrix.

6. Transpose of a matrix. The transpose is usually denoted by a prime, and is obtained by interchanging rows and columns, $A' = (a_{ji})$. The transpose of a

transpose gives the original matrix. If a matrix equals its transpose, it is called a symmetric matrix. For example, the transpose of $A$ in (8.1.1) is

$$A' = \begin{bmatrix} 15 & 25 \\ -12 & 22 \\ 53 & -2 \end{bmatrix}. \tag{8.1.3}$$

The product $A'B$ where $B$ is from (8.1.2) is now well defined. The transpose $A'$ has 2 columns matching the 2 rows of $B$, giving the column-row match mentioned above.

$$\begin{aligned} C &= A'B \\ &= \begin{bmatrix} 15 & 25 \\ -12 & 22 \\ 53 & -2 \end{bmatrix} \begin{bmatrix} -15 & 12 & -53 \\ -25 & -22 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 15(-15)+25(-25) & 15(12)+25(-22) & 15(-53)+25(2) \\ (-12)(-15)+22(-25) & (-12)(12)+22(-22) & (-12)(-53)+22(2) \\ 53(-15)+(-2)(-25) & 53(12)+(-2)(-22) & 53(-53)+(-2)(2) \end{bmatrix} \\ &= \begin{bmatrix} -850 & -370 & -745 \\ -370 & -628 & 680 \\ -745 & 680 & -2813 \end{bmatrix} \end{aligned}$$

A useful rule for the transpose of a product of two matrices $AB$ (not necessarily square matrices) is $(AB)' = B'A'$, where the order of multiplication reverses.

### 8.1.2 Determinant of a Square Matrix and Singularity

Let the matrix $A$ be a square matrix $A = (a_{ij})$ with $i$ and $j = 1, 2, \ldots, n$, in this section. The determinant is a scalar number. When $n = 2$, we have a simple calculation of the determinant as follows:

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{8.1.4}$$

When $n = 3$, we need to expand it in terms of three $2 \times 2$ determinants and the first row as follows:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}|, \tag{8.1.5}$$

where the uppercase $A$ with subscripts is the new notation. $A_{ij}$ denotes a submatrix formed by erasing $i$th row and $j$th column from the original $A$. For example,

$$A_{13} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}. \tag{8.1.6}$$

This is also called a minor. $|A_{ij}|$ in (8.1.5) denotes the determinant of the $2 \times 2$ matrix. The expansion of $|A|$ of (8.1.5), upon substituting for determinants of minors, will be

$$|A| = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \tag{8.1.7}$$

Signed determinants of minors $A_{ij}$ are called cofactors. Denote

$$c_{ij} = (-1)^{i+j}|A_{ij}|. \tag{8.1.7}$$

In general, for any $n$, one can expand the determinant of $A$, denoted by $|A|$ or $\det(A)$, in terms of cofactors as

$$\begin{aligned} \det(A) &= \sum_{j=1}^{n} a_{ij}c_{ij} \quad \text{for any one row from out of} \quad i = 1, 2, \ldots, n, \\ &= \sum_{j=1}^{n} a_{ij}c_{ij} \quad \text{for any one column from out of} \quad j = 1, 2, \ldots, n. \end{aligned}$$

An alternative definition of a determinant is

$$\det(A) = \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}, \tag{8.1.8}$$

where $\sigma$ runs over all permutations of integers $1, 2, \ldots, n$ and the function $\operatorname{sgn}(\sigma)$ is sometimes called signature function. Here it is 1 or –1, depending on whether a permutation is odd or even. For example, if $n = 2$, the even permutation of $\{1, 2\}$ is $\{1, 2\}$, whereby $\sigma(1) = 1$ and $\sigma(2) = 2$. Now the odd permutation of $\{1, 2\}$ is $\{2, 1\}$ having $\sigma(1) = 2$ and $\sigma(2) = 1$.

The following properties of determinants are useful:

1. $\det(A) = \det(A')$, where $A'$ is the transpose of $A$.
2. $\det(AB) = \det(A)\det(B)$ for the determinant of a product of two matrices. Unfortunately, for the determinant of a sum of two matrices there is no simple equality or inequality, $\det(A + B) \neq \det(A) + \det(B)$.
3. If $B$ is obtained from $A$ by interchanging a pair of rows (or columns), then $\det(B) = -\det(A)$.

4. If $B$ is obtained from $A$ by multiplying the elements of a row (or column), by constant $k$, then $\det(B) = k\det(A)$.
5. If $B$ is obtained from $A$ by multiplying the elements of $i$th row (or column), of $A$ by constant $k$, and adding the result of $j$th row of $A$ then $\det(B) = \det(A)$.
6. Determinant of a diagonal matrix is simply the product of diagonal entries.
7. Determinant is a product of eigenvalues, $\det(A) = \Pi_{i=1}^{n}\lambda_i(A)$. The eigenvalues are defined in Section 8.4 below.
8. Zero determinant and singularity. When is the $\det(A) = 0$? (a) If two rows of $A$ are identical, (b) if two columns of $A$ are identical, (c) if a row or column has all zeros, and (d) one eigenvalue is zero. Matrix $A$ is called nonsingular if $\det(A) \neq 0$.

### 8.1.3 The Rank and Trace of a Matrix

A $T \times p$ matrix $X$ is said to be of rank $p$ if the dimension of the largest nonsingular square submatrix is $p$. (Recall that nonsingular means nonzero determinant, and $X$ is often the matrix of regressors with $T$ observations.) The idea of the rank is related to linear independence as follows: We have

$$\text{rank}(A) = \min[\text{Row rank}(A),\ \text{Column rank}(A)], \tag{8.1.9}$$

where the row rank is the largest number of linearly independent rows, and where the column rank is the largest number of linearly independent columns. In the example

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} 15 & 30 & 53 \\ 25 & 50 & -2 \end{bmatrix} \tag{8.1.10}$$

note that $a_{12} = 2a_{11}$ and $a_{22} = 2a_{21}$. Hence the first two columns are linearly dependent, and there are only two linearly independent columns. The column rank is only 2. Recall that a set of vectors $a_1, a_2, \ldots, a_n$, is linearly dependent if a set of scalars $c_i$ exists and the scalars are not all zero and satisfy $\Sigma_{i=1}^{n} c_i a_i = 0$. In (8.1.10) we can let $a_1 = \begin{bmatrix} 15 \\ 25 \end{bmatrix}$ and $a_2 = \begin{bmatrix} 30 \\ 50 \end{bmatrix}$ and choose $c_1 = 2$ and $c_2 = -1$ to verify that the summation is indeed zero.

The following properties of rank are useful:

1. $\text{rank}(X) \leq \min(T, p)$. This says that the rank of a matrix is no greater than the smaller of the two dimensions of rows ($T$) and columns ($p$). In statistics and econometrics texts one often encounters the expression that "the matrix $X$ of regressors is assumed to be of full (column) rank."

Since the number of observations in a regression problem should exceed the number of variables, $T > p$ should hold, which means that $\min(T, p) = p$. If $\text{rank}(X) = p$, the rank is the largest it can be, and hence we say that $X$ is of full rank.

2. $\text{rank}(A) = \text{rank}(A')$.
3. $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$.
4. Sylvester's law. If $A$ is $m \times n$ and $B$ is $n \times q$, $\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB) \leq \min[\text{rank}(A), \text{rank}(B)]$.
5. If $B$ is nonsingular $\text{rank}(AB) = \text{rank}(BA) = \text{rank}(A)$.
6. If $A$ is a matrix of real numbers, $\text{rank}(A) = \text{rank}(A'A) = \text{rank}(AA')$. Assuming that the matrix products $AB$, $BC$, and $ABC$ are well defined, we have $\text{rank}(AB) + \text{rank}(BC) \leq \text{rank}(B) + \text{rank}(ABC)$.
7. The rank equals the number of nonzero eigenvalues (defined below) of a matrix. With the advent of computer programs for eigenvalue computation, it is sometimes easier to determine the rank by merely counting the number of nonzero eigenvalues.

Trace is simply the summation of the diagonal elements. It is also the sum of eigenvalues.

$$\text{Tr}(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i(A).$$

Also note that $\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B)$, and $\text{Tr}(AB) = \text{Tr}(BA)$. We use the latter to simplify matrix expressions below.

### 8.1.4 Matrix Inverse, Partitioned Matrices, and Their Inverse

The inverse matrix is an important tool in matrix algebra since any matrix multiplied by its inverse gives the identity matrix. Let $A_{ij}$ denote an $(n-1) \times (n-1)$ submatrix (called a minor) obtained by deleting $i$th row and $j$th column of an $n \times n$ matrix $A = (a_{ij})$. The cofactor of $A$ is defined in equation (8.1.7) above as $C = (c_{ij})$, where $i, j$ the element is $c_{ij} = (-1)^{i+j}\det(A_{ij})$. The adjoint of $A$ is defined as $\text{Adj}(A) = C' = (c_{ji})$, which involves interchanging the rows and columns of the matrix of cofactors-*indicated* by the subscript $ji$ instead of the usual $ij$. The operation of interchange is the transpose (above).

The inverse matrix is defined for only square matrices and denoted by superscript –1. If $A = (a_{ij})$ with $i, j = 1, 2, \ldots, n$, then its inverse

$$A^{-1} = (a^{ij}) = \frac{\text{Adj}(A)}{\det(A)} = (-1)^{i+j}\frac{\det(A_{ji})}{\det(A)}, \tag{8.1.11}$$

where the $i, j$ element of $A^{-1}$ is denoted by $a^{ij}$, and where $A_{ji}$ is the submatrix of $A$ obtained by eliminating $j$th row and $i$th column. Since the denominator

of the inverse has its determinant, the inverse of a matrix is not defined unless the matrix is nonsingular, meaning it has nonzero determinant. Otherwise, one has the problem of dividing by a zero.

***Inverse of a Product of Two Matrices.*** $(AB)^{-1} = B^{-1}A^{-1}$, where the order is reversed. This is similar to the transpose of a product of two or more matrices. However, we do not apply a similar rule to the inverse of a sum. In general, $(A + B)^{-1} \neq A^{-1} + B^{-1}$. In fact, even if $A^{-1}$ and $B^{-1}$ exist, their sum may not have an inverse. For example, if $B = -A$, the sum $A + B$ is the null matrix, which does not have an inverse because one cannot divide by a zero.

***Solution of a Set of Linear Equations.*** $S\beta = y$. Let $S$ be a $2 \times 2$ matrix $S = \begin{bmatrix} 5 & 3 \\ 4 & 2 \end{bmatrix}$, and let $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$, and $y = \begin{bmatrix} 19 \\ 14 \end{bmatrix}$ be two column vectors. Multiplying out the matrix form $S\beta = y$ yields the algebraic form of the two equations:

$$\begin{aligned} 5\beta_1 + 3\beta_2 &= 19, \\ 4\beta_1 + 2\beta_2 &= 14. \end{aligned} \tag{8.1.12}$$

This system of equations can be solved by hand, but let us look at the matrix solution:

| | |
|---|---|
| $S\beta = y.$ | |
| $S^{-1}S\beta = S^{-1}y$ | Multiplying each side by $S^{-1}$. |
| $I\beta = S^{-1}y$ | Since $S^{-1}S = I$. |
| $\beta = S^{-1}y.$ | Since any matrix times the identity is itself. |

This solution works for any number of equations in the system, and therefore will be of value when the equations get too numerous to solve by hand. In the example above,

$$S = \begin{bmatrix} 5 & 3 \\ 4 & 2 \end{bmatrix}, \ \text{Cofactor}(S) = \begin{bmatrix} 2 & -4 \\ -3 & 5 \end{bmatrix}, \det(S) = 10 - 12 = -2.$$

$$\text{Adj}(S) = \begin{bmatrix} 2 & -3 \\ -4 & 5 \end{bmatrix}, S^{-1} = \begin{bmatrix} -1 & 1.5 \\ 2 & -2.5 \end{bmatrix}, S^{-1}y = \begin{bmatrix} -1 & 1.5 \\ 2 & -2.5 \end{bmatrix}\begin{bmatrix} 19 \\ 14 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

So $\beta_1 = 2$ and $\beta_2 = 3$ is the solution.

### 8.1.5 Characteristic Polynomial, Eigenvalues, and Eigenvectors

There are many concepts in econometrics, such as multicollinearity, that are better understood with the help of characteristic polynomials, eigenvalues, and eigenvectors. Given a matrix $A$, it is interesting to note that one can define a characteristic polynomial in $\lambda$ as

$$f(\lambda) = \det(A - \lambda I), \tag{8.1.13}$$

where $A - \lambda I$ is sometimes called the characteristic matrix of $A$. How does a determinant become a polynomial? The best way to see this is with an example of a symmetric $2 \times 2$ matrix:

$$A - \lambda I = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1-\lambda & 0.7 \\ 0.7 & 1-\lambda \end{bmatrix}.$$

Now the determinant $\det(A - \lambda I) = (1 - \lambda)^2 - 0.49$, which is seen to be a quadratic (polynomial) in $\lambda$. Note that the highest power of $\lambda$ is 2, which is also the dimension of $A$. For $n$-dimensional matrix $A$ we would have an $n$th degree polynomial in $\lambda$.

The eigenvalues are also called characteristic roots, proper values, latent roots, and so on. These have fundamental importance in understanding the properties of a matrix. For arbitrary square matrices, eigenvalues are complex roots of the characteristic polynomial. After all, not all polynomials have real roots. For example, the polynomial $\lambda^2 + 1 = 0$ has roots $\lambda_1 = i$ and $\lambda_2 = -i$ where $i$ denotes $\sqrt{-1}$. The eigenvalues are denoted by $\lambda_i$ where $i = 1, 2, \ldots, n$, are in nonincreasing order of absolute values. For complex numbers, the absolute value is defined as the square root of the product of the number and its complex conjugate. The complex conjugate of $i$ is $-i$, their product is 1, with square root also 1. Thus $|i| = 1$ and $|-i| = 1$, and the polynomial $\lambda^2 + 1 = (\lambda - \lambda_1)(\lambda - \lambda_2)$ holds true. As in the example above, the roots of a polynomial need not be distinct.

By the fundamental theorem of algebra, an $n$th degree polynomial defined over the field of complex numbers has $n$ roots. Hence we have to count each eigenvalue with its proper multiplicity when we write $|\lambda_1(A)| \geq |\lambda_2(A)| \geq, \ldots, \geq |\lambda_{n-1}(A)| \geq |\lambda_n(A)|$. If eigenvalues are all real, we distinguish between positive and negative eigenvalues when ordering them by

$$\lambda_1(A) \geq \lambda_2(A) \geq, \ldots, \geq \lambda_{n-1}(A) \geq \lambda_n(A). \tag{8.1.13}$$

For the cases of interest in statistical applications we usually consider eigenvalues of only those matrices that can be proved to have only real eigenvalues.

***Eigenvectors.*** eigenvector $z$ of an $n \times n$ matrix $A$ is of dimension $n \times 1$. These are defined by the relationship. $Az = \lambda z$, which can be written as $Az = \lambda Iz$. Now moving $Iz$ to the left-hand side, we have $(A - \lambda I)z = 0$. In the example above where $n = 2$ this relation is a system of two equations when the matrices are explicitly written out. For the $2 \times 2$ symmetric matrix above the two equations are

$$(1-\lambda)z_1 + 0.7z_2 = 0,$$
$$0.7z_1 + (1-\lambda)z_2 = 0. \tag{8.1.14}$$

Note that a so-called trivial solution is to choose $z = 0$; that is, both elements of $z$ are zero. We ignore this valid solution because it is not interesting or useful. This system of $n$ equations is called degenerate, and it has no solution unless additional restrictions are imposed.

If the vector $z$ has two elements $z_1$ and $z_2$ to be solved from two equations, then why is it degenerate? The problem is that both equations yield the same solution. The two equations are not linearly independent. We can impose the additional restriction that the $z$ vector must lie on the unit circle, that is, $z_1^2 + z_2^2 = 1$. Now show that there are two ways of solving the system of two equations. The choice $\lambda_1 = 1 + 0.7$ and $\lambda_2 = 1 - 0.7$ yields the two solutions. Associated with each of the two solutions there are two eigenvectors that satisfy the defining equations $Az = \lambda_1 Iz$ and $Az = \lambda_2 Iz$.

The characteristic equation for the example above ($n = 2$) is

$$\det(A - \lambda I) = 0 = \prod_{i=1}^{n} (\lambda - \lambda_i).$$

A remarkable result known as Cayley-Hamilton theorem states that the matrix $A$ satisfies its own characteristic equation in the sense that if we replace $\lambda$ by $A$, we still have

$$0 = \prod_{i=1}^{n} (A - \lambda_i I) = (A - 1.7I)(A - 0.3I) = \begin{bmatrix} -0.07 & 0.7 \\ 0.7 & -0.07 \end{bmatrix} \begin{bmatrix} 0.7 & 0.7 \\ 0.7 & 0.7 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

### 8.1.6 Orthogonal Vectors and Matrices

Geometrically, orthogonal means perpendicular. Two nonnull vectors $a$ and $b$ are perpendicular to each other if the angle $\theta$ between them is 90 degrees or $\pi/2$ radians. Geometrically, the vector $a$ is represented as a line from the origin to the point $A$ (say) whose coordinates are $a_1$ to $a_n$ in an $n$-dimensional Euclidian space. Similarly the vector $b$ is represented by a line from the origin to a point $B$ with coordinates $b_1$ to $b_n$. From analytic geometry, the line joining the points $A$ and $B$ is represented by $|a - b|$, and we know the following fact about the angle $\theta$:

$$\cos\theta = \frac{|a|^2 + |b|^2 - |a-b|^2}{2|a||b|},$$

where $|a|$ is the Euclidian length of the vector $a$, that is, $|a| = \sqrt{(\Sigma_{i=1}^{n} a_i^2)}$, and similarly for $b$. The length of the line joining the points $A$ and $B$ is

$$|a-b| = \sqrt{\left(\sum_{i=1}^{n}[a_i - b_i]^2\right)} = \sqrt{\left(\sum_{i=1}^{n}[a_i^2 + b_i^2 - 2a_i b_i]\right)}.$$

The vectors are perpendicular or orthogonal if $\theta = \pi/2$, and $\cos\theta = 0$. Thus the left side of the equation above for $\cos\theta$ is zero if the numerator is zero, since the denominator is definitely nonzero. Thus we require that $|a|^2 + |b|^2 = |a - b|^2$, which amounts to requiring the cross-product term to be zero, or $-\Sigma_{i=1}^{n} 2a_i b_i = 0$, that is, $\Sigma_{i=1}^{n} a_i b_i = 0 = a'b$, with the prime used to denote the transpose of the column vector $a$. The reader should note the replacement of summation $\Sigma_{i=1}^{n} a_i b_i$ by the simpler $a'b$. In particular, if $b'a$, one has the squared Euclidian length $|a|^2 = a'a$ as the sum of squares of the elements of the vector $a$. If the Euclidian length of both $a$ and $b$ is unity, and $a'b = 0$, meaning the vectors are orthogonal to each other, they are called orthonormal vectors. To see this, let the two vectors defining the usual two-dimensional axes be $e_1 = (1, 0)$ and $e_2 = (0, 1)$ respectively. Now $e_1'e_2 = (0 + 0) = 0$; that is, $e_1$ and $e_2$ are orthogonal. Since $|e_1|^2 = e_1'e_1 = 1$ and similarly $|e_2| = 1$, they are orthonormal also.

A remarkable result called Gram-Schmidt orthogonalization is a process that assures that any given $m$ linearly independent vectors can be transformed into a set of $m$ orthonormal vectors by a set of linear equations.

***Orthogonal Matrices.*** The distinction between orthogonal and orthonormal is usually not made for matrices. A matrix $A$ is said to be orthogonal if its transpose equals its inverse. $A' = A^{-1}$, which also means that $A'A = I$.

The determinant of orthogonal matrix is either +1 or –1, and its eigenvalues are also +1 or –1. To make these results plausible, recall that $\det(AB) = \det(A)\det(B)$. Hence applying this to the relation $A'A = I$ for orthogonal $A$ we have $\det(A)\det(A') = \det(I)$. Now $\det(I) = 1$ is obvious.

### 8.1.7 Idempotent Matrices

In ordinary algebra the familiar numbers whose square is itself are 1 or 0. Any integer power of 1 is 1. In matrix algebra, the identity matrix $I$ plays the role of the number 1, and I certainly has the property $I^n = I$. There are other nonnull matrices that are not identity and yet have this property, namely $A^2 = A$. Hence a new name "idempotent" is needed to describe them. For example, the hat matrix is defined as $\hat{H} = X(X'X)^{-1}X'$.

Writing $\hat{H}^2 = \hat{H}\,\hat{H} = X(X'X)^{-1}X'\,X(X'X)^{-1}X' = X(X'X)^{-1}X' = \hat{H}$ means that $\hat{H}$ matrix is idempotent. Also $\hat{H}^n = \hat{H}$ for any integer power $n$. If eigenvalues are $\lambda_i(\hat{H})$, the relation $\lambda_n = \hat{H}$ means that $\lambda_i(\hat{H}^n) = \lambda_i(\hat{H}) = [\lambda_i(\hat{H})]^n$. Since the only numbers whose $n$th power is itself are unity and zero, it is plausible that the eigenvalues of an idempotent matrix are 1 or 0.

Recall that the rank of a matrix equals the number of nonzero eigenvalues and that the trace of matrix is the sum of eigenvalues. Using these two results

it is clear that for idempotent matrices $G$ whose eigenvalues are 0 or 1 we have the interesting relation trace($G$) = rank($G$).

If the matrix elements are functions of parameters, the derivative of the matrix may be taken as follows:

$A' = [a_1\ a_2\ a_3]$

$$\partial A/\partial x = \begin{bmatrix} \partial a_1/\partial x_1 & \partial a_1/\partial x_2 \\ \partial a_2/\partial x_1 & \partial a_2/\partial x_2 \\ \partial a_3/\partial x_1 & \partial a_3/\partial x_3 \end{bmatrix} \tag{8.1.15}$$

The first derivative of $x'B$ with respect to $x$ is simply $B$. For the first derivative of a quadratic form, let $Q = x'Bx$, where $x$ is $n \times 1$, and $B$ is $n \times n$. Then the $1 \times n$ derivative is:

$$\partial Q/\partial x = x'(B + B'). \tag{8.1.16}$$

The second derivative of a quadratic form is:

$$\partial^2 Q/\partial x \partial x' = B' + B (= 2B \text{ if } B \text{ is symmetric}) \tag{8.1.17}$$

### 8.1.8 Quadratic and Bilinear Forms

If $x$ is an $n \times 1$ vector and $A$ is an $n \times n$ matrix the expression $x'Ax$ is of dimension $1 \times 1$ or a scalar. A scalar does not mean that it is just one term in an expression, only that when it is evaluated it is a number. For example, when $n = 2$, and $A = (a_{ij})$, note that

$$x'Ax = [x_1 \quad x_2]\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2$$

is a quadratic expression. Hence $x'Ax$ is called a quadratic form, which is a scalar and has an expression having four terms. Note that the largest power of any element of $x$ is 2, even for the case where $A$ is a $3 \times 3$ matrix. When the $a_{ij}$ and $x_i$ are replaced by numbers, it is clear that the quadratic form is just a number, which is called a scalar to be distinguished from a matrix. A scalar cannot be a matrix, but a matrix can be a scalar.

***Definite Quadratic Forms.*** Just as a number can be negative, zero, or positive, a quadratic form also can be negative, zero, or positive. Since the sign of zero can be negative or positive, it is customary in matrix algebra to say that a quadratic form is negative or positive definite, meaning that $x'Ax < 0$ and $x'Ax > 0$, respectively. The expression "definite" reminds us that it is not zero. If $x'Ax \geq 0$, it is called nonnegative definite (nnd), and if $x'Ax \leq 0$, it is called nonpositive definite (npd). Remember that the matrix $A$ must be a square matrix for it to be the matrix of a quadratic form, but it need not be a symmetric matrix (i.e., $A' = A$ is not necessary).

If one has a symmetric matrix, the cross-product terms can be merged. For example, the symmetry of $A$ means that $a_{12} = a_{21}$ in the $2 \times 2$ illustration above:

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2 = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2.$$

If one has an asymmetric matrix $A$ and still wishes to simplify the quadratic form this way, she can redefine the matrix of the quadratic form as $B = (\frac{1}{2})(A + A')$ and use $x'Bx$ as the quadratic form, which can be proved to be equal to $x'Ax$.

A simple practical method of determining whether a given quadratic form $A$ is positive definite in the modern era of computers is to find its eigenvalues, and concentrate on the smallest eigenvalue. If the smallest eigenvalue $\lambda\min(A) > 0$, is strictly positive, the quadratic form $x'Ax$ is said to be positive definite. Similarly, if the smallest eigenvalue $\lambda\min(A) \geq 0$, which can be zero, the quadratic form $x'Ax$ is said to be nonnegative definite. In terms of determinants, there is a sequence of determinants that should be checked to be positive definite for the quadratic form to be positive definite. We check.

$$a_{11} > 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0, \ldots,$$

where we have indicated the principal determinants from 1, 2, 3. To see the intuitive reason for these relations, consider $n = 2$ and the quadratic form

$$Q = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2.$$

Complete the square by adding and subtracting $(a_{12}/a_{11})^2x_2^2$. Now write

$$Q = a_{11}\left[x_1 + \left(\frac{a_{12}}{a_{11}}\right)x_2\right]^2 + \left[a_{22} - \left(\frac{a_{12}^2}{a_{11}}\right)\right]x_2^2.$$

Observe that $Q > 0$ requires that $a_{11} > 0$ from the first term, and $a_{11}a_{22} - a_{12}^2 > 0$ from the second term. The second term is positive if and only if the $2 \times 2$ determinant above is positive.

### 8.1.9 Further Study of Eigenvalues and Eigenvectors

There are two fundamental results in the analysis of real symmetric matrices.

1. The eigenvalues of a real symmetric matrix are real. This is proved by using the fact that if the eigenvalues were complex numbers, they must come in conjugate pairs. $Ax = \lambda x$ and $A\bar{x} = \bar{\lambda}\bar{x}$, where the bar indicates the conjugate complex number. Premultiply $Ax = \lambda x$ by the $\bar{x}'$ vector to yield $\bar{x}'Ax = \lambda\bar{x}'x$ (is a scalar). Now premultiply $A\bar{x} = \bar{\lambda}\bar{x}$ by $x'$ to yield $x'A\bar{x} = \bar{\lambda}x'\bar{x}$. Now

the quadratic forms on the left sides of these relations must be equal to each other by the symmetry of $A$ (i.e., $\bar{x}'Ax = x'A\bar{x}$). Hence we have $(\lambda - \bar{\lambda})x'\bar{x} = 0$, which can only be true if $\lambda = \bar{\lambda}$, a contradiction. The complex conjugate cannot be equal to the original.

2. Eigenvectors associated with distinct eigenvalues are orthogonal to each other. The proof involves an argument similar to the one above, except that instead of $\bar{\lambda}$, we use a distinct root $\mu$ and instead of $\bar{x}$, we use the vector $y$. The relation $(\lambda - \mu)x'y = 0$ now implies that $x'y = 0$, since the roots $\lambda$ and $\mu$ are assumed to be distinct. Hence the two eigenvectors $x$ and $y$ must be orthogonal.

***Reduction of a Real Symmetric Matrix to the Diagonal Form.*** Let $A$ be a real symmetric matrix of order $n \times n$ having distinct eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ ordered in a sequence from the largest $\lambda_1 = \max^i\{\lambda_i, i = 1, 2, \ldots, n\}$ to the smallest $\lambda_n = \min^i\{\lambda_i, i = 1, 2, \ldots, n\}$. Denote by $G$ the $n \times n$ orthogonal matrix of corresponding eigenvectors $G = [x_{.1}, x_{.2}, \ldots, x_{.n}]$, where the dot notation is used to remind us that the first subscript is suppressed and each $x_{.i}$ is an $n \times 1$ column vector. Orthogonality of $G$ is verified by the property $G'G = GG'$. Recall the fundamental relation defining the $i$th eigenvalue and eigenvector $Ax_{.i} = \lambda_i x_{.i}$, in terms of the dot notation. Denote by $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, a diagonal matrix containing the eigenvalues in the nonincreasing order. Combining all eigenvalues, and eigenvectors, we write $AG = G\Lambda$. The relation $G'AG = G'G\Lambda = \Lambda$ is called eigenvalue-eigenvector decomposition of the real symmetric matrix A. For example, let

$$A = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} w & w \\ -w & w \end{bmatrix},$$

where

$$w = \frac{1}{\sqrt{2}} \quad \text{and} \quad A = \begin{bmatrix} 1.7 & 0 \\ 0 & 0.3 \end{bmatrix}.$$

It is easy to verify that $G'G = I = GG'$, and $G'AG = \Lambda$ by direct multiplication. In the dot notation, note that $x_{.1} = \begin{bmatrix} w \\ -w \end{bmatrix}$, that $x_{.2} = \begin{bmatrix} w \\ w \end{bmatrix}$, and that $G = [x_{.1}, x_{.2}]$. We will verify $A = G\Lambda G'$:

$$G\Lambda = \begin{bmatrix} w & w \\ -w & w \end{bmatrix}\begin{bmatrix} 1.7 & 0 \\ 0 & 0.3 \end{bmatrix} = \begin{bmatrix} 1.7w & 0.3w \\ -1.7w & 0.3w \end{bmatrix},$$

$$G\Lambda G' = \begin{bmatrix} 1.7w^2 + 0.3w^2 & -1.7w^2 + 0.3w^2 \\ -1.7w^2 + 0.3w^2 & 1.7w^2 + 0.3w^2 \end{bmatrix}$$

$$= \begin{bmatrix} 2w^2 & -1.4w^2 \\ -1.4w^2 & 2w^2 \end{bmatrix} = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix} = A,$$

since $w^2 = (1/2)$ by the definition of $w$. This completes the verification of the eigenvalue eigenvector decomposition in the simple $2 \times 2$ case.

## 8.2 MATRIX-BASED DERIVATION OF THE EFFICIENT PORTFOLIO

Since matrix algebra can be used to solve a system of several equations as easily as a system of two or three equations, the portfolio selection problem is an obvious application. Portfolio selection requires the investor to find the weights on each individual stock that give the lowest variance for a required return. The millions of possible combinations would be intractable by hand, but with the use of computers and the organization of the system that comes with matrix algebra, the solution becomes streamlined. A refined derivation of the tangency solution is based on quadratic programming methods used by Markowitz. The objective function is formulated as a quadratic function involving the minimization of variance subject to receiving a specified expected return. The portfolio selection problem is a generalization of mean-variance optimizing where the agent wishes to maximize the expected utility of some function of wealth $E(u(W))$ where the expected utility is a function of two arguments $f(w'\mu, w'\Sigma w)$ with $f$ being a nonlinear function increasing in the first argument and decreasing in the second argument. The vector $w$ denotes the portfolio weights. We assume that there are $n$ assets from which the investor chooses her portfolio. The constraint on the weights is that they must add up to unity: $w'\iota = 1$, where $\iota$ denotes an $n \times 1$ vector of ones. This is called the adding-up constraint. The problem can be stated as a maximization of a Lagrangian function of (8.2.1) below, involving three terms: a positive term for the mean return, a negative term for the variance or risk, and the last term for the budget constraint.

A compact and intuitive expression for the Lagrangian objective function and its solutions needs the following matrix expressions. Following Ingersoll (1987 p. 84), consider three observable quadratic forms $A$, $B$, and $C$ defined as

$$A = \iota' \sum^{-1} \iota, \quad B = \mu_R' \sum^{-1} \iota, \quad C = \mu_R' \sum^{-1} \mu_R, \tag{8.2.1}$$

where $\Sigma$ is the $n \times n$ covariance matrix among returns for the $n$ assets, $\iota$ is a $n \times 1$ vector of ones, and $\mu_R$ denotes an $n \times 1$ vector of average returns.

Our basic objective function is to maximize the return subject to conditions on the variance of returns and the adding-up constraint. Hence, combining with the conditions, our Lagrangian objective function in matrix algebra terms is

$$\max_w L = w'\mu_R - 0.5\gamma w' \sum w - \eta(w'\iota - 1). \tag{8.2.2}$$

The Lagrangian coefficient $\gamma$ attached to the variance term in (8.2.2) is called the coefficient of risk aversion, and the Lagrangian coefficient $\eta$ for the adding-up constraint does not have a particular name.

Using matrix algebra calculus (8.1.16), we can write the solution to the portfolio problem $w^*$ as

$$w^* = \gamma^{-1} \sum^{-1} (\mu_R - \eta). \tag{8.2.3}$$

There is an interesting uniqueness result here. As soon as either $\gamma$ or $\eta$ are determined, the solution is known uniquely, since the Lagrangian coefficients $\gamma$ and $\eta$ must satisfy the relation: $\gamma = B - A\eta$. There is a singularity when $\eta = B/A$, which must be assumed away.

Consider the mean standard deviation space with the mean on the vertical axis. For a given efficient portfolio $w^*$, consider the tangent line to the mean-variance frontier at the solution $w^*$, and note the intercept point at which this line meets the vertical axis. In the CAPM framework the zero-beta portfolio of $w^*$ is where the risk is zero, namely where the standard deviation is zero along the vertical axis. The expected return on such a zero-beta intercept portfolio can be interpreted as the Langrangian coefficient $\eta$.

Over time the mean-variance frontier is expected to change. In such a dynamic setting one is interested in the following questions: What is the effect of introducing an additional asset on the mean-variance frontier? How does the benchmark asset change with the presence of new assets? How does the frontier of the benchmark asset change? Huberman and Kandel (1987) introduced the concepts of spanning and intersection to deal with the answers to these questions.

In practice, if some agents believe that they have superior hunch or better information, they will choose asset proportions different from those in the efficient tangency portfolio. However, Markowitz's theory suggests that competitive markets should converge toward a consensus portfolio held by everyone. When faced with the reality of vast differences between individual portfolios, one may regard mean-variance theory as too constricting. Many authors have modified the preceding theory to make it more realistic. We have already noted that in 1960s Markowitz's mean-variance model was extended by Sharpe, Treynor, and Lintner into the CAPM. Both models rely on the assumption of independently and identically distributed (iid) normal distribution of asset returns. Otherwise, higher moments have to be considered.

## 8.3 PRINCIPAL COMPONENTS ANALYSIS, FACTOR ANALYSIS, AND SINGLULAR VALUE DECOMPOSITION

This section gives a brief introduction to certain matrix algebra techniques commonly employed in multivariate statistical analysis with the help of a simple example. Table 8.3.1 gives artificial data regarding excess returns on $i = 1$ to $i = n = 4$ assets at $t = 1$ to $t = T = 5$ time periods. We place the data in a $T \times n$ matrix denoted by $X$. There is no loss of generality even if the real world $X$ can have data on thousands of assets and hundreds of time periods.

**Table 8.3.1 Artificial Data for Four Assets and Five Time Periods in a Matrix**

| | Asset $i = 1$ | Asset $i = 2$ | Asset $i = 3$ | Asset $i = 4 = n$ |
|---|---|---|---|---|
| $t = 1$ | 1 | 1.2 | 0.99 | 1.3 |
| $t = 2$ | 1.3 | 2.1 | 1 | 1.2 |
| $t = 3$ | 2 | 3 | 0.99 | 2 |
| $t = 4$ | 1.5 | 2 | 1.1 | 2.2 |
| $t = 5 = T$ | 1.6 | 3 | 2 | 1.4 |

**Table 8.3.2 Matrix $U$ of Singular Value Decomposition**

| $U$ matrix | | | |
|---|---|---|---|
| $u_1$ | $u_2$ | $u_3$ | $u_4$ |
| –0.28657 | 0.198906 | –0.53663 | –0.75692 |
| –0.37696 | –0.14144 | 0.207701 | –0.19341 |
| –0.54379 | 0.230496 | 0.686777 | –0.15148 |
| –0.4449 | 0.588292 | –0.36106 | 0.559251 |
| –0.53116 | –0.73567 | –0.25857 | 0.232283 |

Singular value decomposition (SVD) is a matrix algebra tool used in multivariate statistics for extracting the information content from any such data set. In SVD one writes

$$X = USV' \tag{8.3.1}$$

as a product of three matrices $U$, $S$, and transpose of $V$. We use GAUSS software routine called svd2 in our implementation reported here. In the sequel we explain the important point that $U$, $S$, and $V$, the three component matrices, reveal the fundamental aspects of all information present in the $X$ data. For our $5 \times 4$ example, the $X$ matrix is full column rank, that is $p = n = 4$ and the numerical values of the three matrices are given here. The matrix $U$ is of the same dimension as $X$ and contains "principal coordinates" of $X$ standardized in the sense that $U'U = I$, the identity matrix. Note that columns of $U$ are denoted as $u_1$ to $u_p$ and that $U$ is not orthonormal, because $UU' \neq I$. The original coordinates are rotated to new coordinate system defined on the unit circle by this standardization. See Vinod and Ullah (1981, pp. 18–23). Table 8.3.2 contains our $U$ matrix for this simple example.

The matrix $S$ is a $p \times p$ diagonal matrix, where $p$ is the rank of $X$. If $n < T$, $p \leq n$, and $T < n$, we have $p \leq T$; that is, the rank cannot exceed the smaller of the two dimensions of $X$. The diagonal matrix $S$ contains ordered "singular values" of $X$ from the largest to the smallest $\{s_1 \geq s_2 \geq \ldots \geq s_p\}$. For our data, $S$ is given in Table 8.3.3. The squares of singular values are called "eigenvalues" $\{\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p\}$, and for our data they are {59.77705, 1.145765, 0.513569, 0.013815}. The ratio of the largest singular value to the smallest singular value

**Table 8.3.3 Diagonal Matrix of Singular Values**

| *S* matrix | | | |
|---|---|---|---|
| 7.731562 | 0 | 0 | 0 |
| 0 | 1.070404 | 0 | 0 |
| 0 | 0 | 0.716638 | 0 |
| 0 | 0 | 0 | 0.117539 |

**Table 8.3.4 Orthogonal Matrix of Eigenvectors**

| $v_1$ | $v_2$ | $v_3$ | $v_4$ |
|---|---|---|---|
| –0.43735 | 0.169461 | 0.211587 | –0.85746 |
| –0.67906 | –0.37115 | 0.494969 | 0.395144 |
| –0.35578 | –0.505 | –0.77858 | –0.11046 |
| –0.47014 | 0.760597 | –0.32256 | 0.310519 |

**Table 8.3.5 First Component from SVD with Weights from the Largest Singular Value**

| | | | |
|---|---|---|---|
| 0.969004 | 1.504527 | 0.78827 | 1.041642 |
| 1.274669 | 1.979119 | 1.036924 | 1.37022 |
| 1.838786 | 2.854998 | 1.495825 | 1.976624 |
| 1.504401 | 2.335813 | 1.223808 | 1.617172 |
| 1.796063 | 2.788664 | 1.461071 | 1.930698 |

is called the "condition number," and large number suggests an ill-conditioned matrix in the sense that its inverse is numerically difficult to compute.

The third matrix $V$ of SVD is also $p \times p$ and contains the eigenvectors (also called characteristic vectors) $\{v_1, v_2, \ldots, v_p\}$ given in the columns of $V$. For our data they are given in Table 8.3.4. They are directly linked with the corresponding eigenvalues. Geometrically $v_i$ vectors contain direction cosines that tell us how to orient the $i$th principal vector with respect to the original coordinate system. The matrix $V$ is said to be orthonormal in the sense that its inverse equals its transpose, that is, $V'V = I = VV'$.

Another way to express SVD is as follows:

$$X = (s_1 u_1 v_1') + (s_2 u_2 v_2') + \ldots + (s_p u_p v_p'). \tag{8.3.2}$$

In this decomposition note $u_1 v_1'$ is a $T \times p$ matrix being multiplied by the scalar $s_1 = 7.731562$. This scalar is much larger than the last scalar multiplier $s_p = 0.117539$ of $u_p v_p'$, which shrinks all elements of that matrix. We report the first term (8.3.2) in Table 8.3.5, where each element is seen to be relatively large in numerical magnitude.

**Table 8.3.6 Last Component from SVD with Weights from the Smallest Singular Value**

| | | | |
|---|---|---|---|
| 0.076286 | –0.03516 | 0.009827 | –0.02763 |
| 0.019492 | –0.00898 | 0.002511 | –0.00706 |
| 0.015267 | –0.00704 | 0.001967 | –0.00553 |
| –0.05636 | 0.025974 | –0.00726 | 0.020412 |
| –0.02341 | 0.010788 | –0.00302 | 0.008478 |

We report the last term in (8.3.2) in Table 8.3.6, where each element is seen to be small in numerical magnitude. Hence this artificial example clarifies that one can ignore trailing eigenvalues and eigenvectors, without much loss of information.

So far we have not omitted any columns from the SVD, implying that we have not achieved any dimension reduction. The dimension reduction can be visualized when we delete trailing singular values, that is, when we replace the trailing diagonal terms of $S$ by zeros. For example, if we want to keep two columns ($k = 2$) to approximate the information in $X$ matrix, we simply replace the remaining trailing $s_i$ by zeros and denote the revised $S$ matrix by $S_{1:k}$. It is easy to verify that for our example the matrix multiplication $US_{1:2}$ has the first two columns with nonzero elements and the last two columns have all zeros. Ignoring the last two columns of all zeros, we have reduced the column dimension from 4 to 2. The $T \times k$ matrix $P_{1:k}$ of selected $k$ principal components is given by the SVD from the relation: $US_{1:k} = P_{1:k}$. Exactly the same principal components can be obtained from the eigenvalues-eigenvector decomposition, although the latter fails to reveal the direct link with principal coordinates.

If we postmultiply both sides of $X = USV'$ by $V$, the matrix of eigenvectors, we have $XV = US$, since $V'V = I$. Hence the product $US$ can also be computed as

$$XV = X(v_1, v_2, v_3, v_4) = (p_1, p_2, p_3, p_4) = P_{1:k}, \tag{8.3.3}$$

which defines the factors $p_i$ as weighted sums of the original data at each $t$ by weights given by the matrix $V$ of eigenvectors. These weighted sums, akin to index numbers, are called the "principal components." Clearly, the dimension reduction is obtained by using first few $p_j$, where $j = 1, \ldots, k$.

The factors in factor analysis are obtained upon premultiplying $P_{1:k}$ by an appropriately chosen matrix $L$ of factor loadings. Factor analysis theory provides many alternative rotations leading to choices of $L$ to enhance the interpretation. The eigenvalues represent the geometrical "spread" or variance in the original $X$ data along each principal dimension. The principal components represent uncorrelated linear combinations of the variables in the columns of $X$. The amount of variability in the original data captured by the $i$th eigenvector is related to the corresponding eigenvalues. The total variability is

captured by the sum of eigenvalues, which must equal the trace of the matrix, $\Sigma\lambda_j = \text{Trace}(X'X)$, where the trace is also the sum of diagonals. In our example, $\lambda_1/\Sigma\lambda_j$ is 0.80235 and $(\lambda_1 + \lambda_2)/\Sigma\lambda_j$ is 0.91343. Thus more than 91% of variance is captured by the first two components. For large data sets the number of components needed is generally much smaller than the rank of $X$.

The somewhat more familiar eigenvalues-eigenvector decomposition from Section 8.1.9 can be directly obtained from SVD as follows: Define a diagonal $p \times p$ matrix $\Lambda$ containing the eigenvalues $\lambda_i = (s_i)^2$. Since $X = USV'$, its transpose is $X' = VS'U' = VSU'$, since $S$ is symmetric. Assume that $T \geq n$ and the rank of $X$ is $n$. Now the matrix multiplication $X'X$ gives

$$X'X = VSU'(USV') = VS^2V' = V\Lambda V', \tag{8.3.4}$$

where we have used the fact that $U'U = I$, and that $S$ being diagonal, $SS = S^2 = \Lambda$. This then is the eigenvalues-eigenvector decomposition of $X'X$, which is always a real and symmetric matrix. For each $i = 1, 2, \ldots, p$ $(= n)$ we have $X'Xv_i = \lambda_i v_i$. This equation is solved by considering the determinantal equation upon inserting the identity matrix:

$$|X'X - \lambda_i I|v_j = 0.$$

Since we started with SVD, we have already solved this equation, in the sense that we know the eigenvalues and eigenvectors. A number of computer algorithms are available for SVD, PCA, and other computations. From the viewpoint of numerical accuracy, SVD is generally believed to be most accurate, McCullough and Vinod (1999, 2003). Direct interpretation of eigenvectors as meaningful multivariate (regression-type) coefficients is problematic.

For a rigorous discussion of multivariate "factor analysis" and statistical inference, see Mardia et al. (1979) and similar texts. For references to recent research with financial applications, see Connor and Korajczyk (1986, 1988, 1993), Tsay (2002), and their references. Since SVD and PCA are sensitive to scaling of the data, if the columns of $X$ are not in comparable units, it is better to standardize the $X$ data so that $X'X$ is a correlation matrix, as in Vinod and Ullah (1981). Correlation matrices have ones along the diagonal; hence the sum of diagonals is simply $n$, and the proportion of variation included in the first $k$ principal components is $\Sigma_{j=1}^{k}\lambda_j/n$. For statistical inference when some regressors are principal components, the traditional methods involving normality assumption are no longer necessary if one is willing to use computer resources for the bootstrap.

## 8.4 ITO'S LEMMA

Ito's lemma is a fundamental result in stochastic calculus on par with Taylor series. If $f(S)$ is a continuous and differential function, its Taylor series expansion is

$$df = \frac{\partial f}{\partial S} dS + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} (dS)^2 + \ldots, \tag{8.4.1}$$

where

$$dS = \mu S dt + \sigma S dz$$

and

$$(dS)^2 = (\mu S dt + \sigma S dz)^2 = \mu^2 S^2 (dt)^2 + \sigma^2 S^2 (dz)^2 + 2\mu S dt \sigma S dz. \tag{8.4.2}$$

Now we let $dt \to 0$ and eliminate all terms smaller than $dt$ (in orders of magnitude); that is, we eliminate terms containing $(dt)^a$ when the power $a > 1$, but keeping terms involving $a = 1$. First we consider the limit of $dS = \mu S dt + \sigma S dz$, where the $dz$ in the last term tends to be of order $\sqrt{dt}$ or $(dt)^{0.5}$ where the power is no larger than unity. Hence the limit for $dS$ remains $\mu S dt + \sigma S dz$. When we evaluate the three terms in the expansion of $(dS)^2$, there is cancellation. The first term cancels since it involves $(dt)^2$ where the power exceeds unity. The last term has $dtdz$ of which $dz$ is of order 0.5, implying that this term is of order $(dt)^{1.5}$. Since the power of $dt$ exceeds 1, this last term cancels. The limit as $dt \to 0$ involves only the middle term $\sigma^2 S^2 (dz)^2$, where $(dz)^2 \to dt$, which we do retain. Thus, we conclude that $(dS)^2 \to \sigma^2 S^2 dt$. Now we are ready to substitute these limits in the Taylor series expansion and adjust Ito's lemma of (8.4.1) to become

$$df = \frac{\partial f}{\partial S} [\mu S dt + \sigma S dz] + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2 dt + \ldots. \tag{8.4.3}$$

Note that this lemma relates a small change in a function $f(S)$ of the spot price $S$ of an asset to a small change in the random shock $dz = u\sqrt{dt}$.

In general, if we consider $f(S, t)$ where the function is time dependent, the Taylor series will be

$$df(S,t) = \frac{\partial f}{\partial S} dS + \frac{\partial f}{\partial t} dt + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} dS^2 + \ldots.$$

Upon evaluating this limit, we have a more general version of Ito's lemma as

$$df = \frac{\partial f}{\partial S} \sigma S dz + \left[ \frac{\partial f}{\partial S} \mu S + \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} \sigma^2 S^2 \right] dt + \ldots. \tag{8.4.4}$$

As a first application of the lemma above consider $f(S) = \ln S$, the natural log of $S$, so that $\partial f/\partial S = 1/S$, $\partial f/\partial t = 0$, and $\partial^2 f/\partial S^2 = -1/S^2$. Let $f_0$ denote the initial value of $f(S)$ when the initial price is $S = S_0$ and $df$ is the change from the initial value $f - f_0$. Now Ito's lemma (8.4.4) yields

$$\frac{df}{f} = \sigma dz + \left[\mu - \frac{1}{2}\sigma^2\right]dt + \ldots.$$

This is a stochastic differential equation whose coefficients involve parameters $\mu$ and $\sigma^2$, which are assumed to be constant.

The $df = f - f_0$ when $f = \ln S - \ln S_0 = \ln(S/S_0)$ is the relative change in the price $S$ computed from the initial price $S_0$ over the time interval ($t_0$ to $t$). The time interval itself will generally be partitioned into infinitesimal (much smaller) intervals and the ultimate price $S$ consists of a sum of several small changes in pirce over those smaller intervals. That is, $S$ is a sum of price jumps or a stochastic integral. The basic model states $df$ is normally distributed with mean $t[\mu - 0.5\sigma^2]$, which is the drift part, where we replace $dt$ by $t$ due to accumulation of jumps (integration). The variance of $df$ is proportional to the time interval $\sigma^2 t$.

## 8.5 CREATION OF RISK-FREE NONRANDOM $g(S, t)$ AS A HEDGE PORTFOLIO

We have so far considered two random walks $df(S, t)$ and $dS$, both containing $dz$ as the random term. This means that the two random walks move together or are correlated with each other. Hence it is plausible that there is some mixture function that links the two. If we could choose the mixture function $g(S, t)$ that will eliminate this randomness, we will have achieved a zero-risk portfolio. The idea of creating such a mixture function is attributed to Merton (1976). Let us choose $g = f - bS$ as the linear function with the coefficient $b$ (to be determined) and apply Ito's lemma to the function $g$ to evaluate its dynamics from $dg = df - bdS$. All we have to do is substitute the expressions for $dS$ and $df$ from the previous section. The diffusion part of $df$ is $\partial f/\partial S\, \sigma S dz$, which is the one with the random part $dz$. We have to subtract $b$ times the diffusion part $\sigma S dz$ of $dS$ from this. We then have $[(\partial f/\partial S) - b]\sigma S dz$. Thus elimination of diffusion is possible if we choose $[(\partial f/\partial S) - b] = 0$.

The drift part of $df$ is

$$\left[\frac{\partial f}{\partial S}\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right]dt. \tag{8.5.1}$$

Note that we subtract $b$ times the drift part of $dS$ or $b\mu S dt$ to yield $[(\partial f/\partial S) - b]$ times $\mu S dt$ and some additional terms. Clearly, if we make $[(\partial f/\partial S) - b] = 0$ to remove the diffusion part altogether from $g(S, t)$, we have the additional bonus of deleting one term from the drift part. Thus we have

$$dg = \left[\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right]dt, \tag{8.5.2}$$

if we choose $b = \partial f/\partial S$.

An option is called a derivative because the price of an option depends on the price $S$ of the asset and time to expiry. Hence the price of an option is $f(S, t)$, which is subject to Ito's lemma with its own drift and diffusion parts. The change in the price of the underlying asset $dS$ is also subject to its own drift and diffusion. Clearly, $df$ and $dS$ are correlated over time. A linear combination $g(S, t) = f - bS$ is simply the cost of owning a portfolio consisting of buying one option at the price $f(S, t)$ and selling (due to the negative sign) $b$ units of the underlying stock. Constructing portfolios (usually involving simultaneous buying and selling of options and assets) to reduce or eliminate risk is called hedging. Hence let us interpret $g(S, t)$ as the hedge portfolio. As time passes, the price of the hedge portfolio changes by $dg = df - bdS$. The diffusion parts of $df$ and $dS$ alone have the randomness $dz$ in them. We have just seen how to eliminate risk arising from randomness by choosing $b = \partial f/\partial S$. We conclude that the hedge portfolio will totally eliminate risk if it consists of one unit of the option bought at price $f$ and simultaneously selling $\partial f/\partial S$ units of the underlying stock.

The hedge portfolio should not be confused with a "risk-neutral" portfolio, even though it neutralizes or eliminates risk from the diffusion part. The expression risk neutral actually refers to what we do with the drift part. Recall that $dS = \mu S dt + \sigma S dz$ has a drift part that depends on the drift parameter $\mu$, which represents the growth rate of the asset price. The Black-Scholes formula does not have any reference to this $\mu$ parameter at all. Even if different investors disagree about the growth prospects of the underlying asset, they can agree on the same price of the option $f(S, t)$. Now consider a fictional world where the growth rate $\mu$ of the asset price $S$ is exactly same as the Bank return $r$. If we are comparing discounted present values of different portfolios, the discount rate $r$ will neutralize the drift growth rate $\mu$. Such a fictional world is called "risk-neutral" world. The derivation of B-S equation without any risk-neutrality assumption is provided in the next section.

## 8.6 DERIVATION OF BLACK-SCHOLES PARTIAL DIFFERENTIAL EQUATION

The previous section derived the risk-free or risk-neutral portfolio consisting of both options and stocks. It eliminates the random component altogether under the following assumptions:

1. The price of option is consistent with dynamic behavior of $f(S, t)$, starting at $f$ and growing in increments of the size $df$.
2. The underlying stock price starts at $S$ and grows in increments of $dS$.
3. There are zero transaction costs (brokerage commissions for buying or selling) and the functions are differentiable so that all needed derivatives exist.

4. The portfolio consists of one unit of option $f(S, t)$ and $(-\Delta)$ units of the stock (remember the notation change from $b$ to $\Delta$ to conform with the Wall Street Greek), where exactly $\Delta = \partial f/\partial S$ units of stock are needed to eliminate the risk.
5. The investor has the option to invest in a government-guaranteed (risk-free) instrument (e.g., an FDIC insured bank) that pays a return $100 * r$ percent over the period $t_0$ to $t$. For example, if $r = 0.05$ and \$100 dollars are invested, the investor will get $\$105 = 100(1 + r)$ dollars over the time period $dt$ from $t_0$ to $t$. If only one dollar is invested, the investor will earn $(1 + r)dt$, where we insert the $dt$ in the expression instead of assuming $dt = 1$ to retain generality in the choice of the time interval $t_0$ to $t$.
6. There are no limits on borrowing, buying, and selling from the bank or from the market for options and stocks.
7. There are no trading curbs when the price moves too much in a short time period.
8. An investor can take a long (buy) or short (sell) position in any market without limit.

Although, these assumptions may seem unrealistic, they do not prevent practical use of the solution equation. Now think of as many dollars as it takes to buy one unit of the mixed portfolio as $g(S, t) = f(S, t) - \Delta S$. Let us compare the \$$g$ invested in the mixed portfolio with the investment of \$$g$ in the risk-free government-backed asset, say, a bank. Since a bank yields $r$ dollars for each dollar, \$$g$ in the bank will earn $rg$ dollars. This investment of \$$g$ in the mixed portfolio grows in increments of $dg$ per period from $t_0$ to $t$. Thanks to some cancellations, the expression for the return of \$$dg$ derived in the previous section is

$$dg = \left[\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right]dt. \tag{8.6.1}$$

The earning of \$$dg$ is risk free by construction of (8.6.1) and should be comparable to the $\$(rgdt)$ in the bank. If \$$dg$ exceeds $\$(rg(S, t)dt)$, a potential arbitrager (a savvy investor who profits by selling in one market and buying in another) will invest in the mixed portfolio of options and stocks (go long on the mixed portfolio) and borrow money in the bank (go short on the bank) to come out ahead, without bearing any risk of loss. Conversely, if $(rgdt)$ exceeds $dg$, a savvy investor can profit by going short on the mixed portfolio and transferring the funds to the bank. Clearly, if the market sets the price of options $f(S, t)$ correctly, there should be no such opportunity for profiting by simply transferring assets from the bank to the mixed portfolio, and vice versa. In other words, an efficient market will be characterized by equating the two risk-free returns, $\$(rgdt) = \$dg$.

The real unknown here is the price of the option $f(S, t)$. The solution to the option pricing problem is obtained by equating the return in the bank of $\$rg$ with the return in the mixed portfolio $g(S, t)$ of one option $f(S, t)$ and $-\Delta$ units $(= -\partial f/\partial S)$ of the stock. Thus the expression for $dg$ noted above should equal $rg(S, t)dt$, which can be written as $[rf(S, t)dt - rS(\partial f/\partial S)dt]$ by its very definition. Thus the efficient price of option $f(S, t)$ should satisfy

$$\left[\frac{\partial f}{\partial t}+\frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right]dt = rf(S,t)dt - rS\left(\frac{\partial f}{\partial S}\right)dt, \tag{8.6.2}$$

where $r$ is the return per dollar from the bank deposit in the time period $dt$. Since $dt$ is present in all terms, it cancels and we do some minor rearranging to yield the celebrated Black-Scholes partial differential equation (PDE) for efficient pricing of options or any derivative security $f(S, t)$, which depends on the price of the underlying asset $S$ and the time period $t$ for which the asset is held:

$$\left[\frac{\partial f}{\partial t}+\frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2 + rS\left(\frac{\partial f}{\partial S}\right)\right] = rf(S,t). \tag{8.6.3}$$

Formula (8.6.3) won Professors Black and Merton the 1997 Nobel memorial Prize in Economics and has proved to be enormously valuable. Since the highest order of the partial derivative with respect to $S$ is 2, it is a second-order PDE. Note that it also involves a first-order partial with respect to time. The "solution" of the PDE will be the formula for the price of the "derivative" security $f(S, t)$ (5.3.2). Even if finding the solution requires considerable mathematical skill, verifying whether the two sides of PDE agree merely involves ability to differentiate $f(S, t)$ twice with respect to $S$ and once with respect to $t$.

Once a PDE is defined, one can use the great mathematical machinery developed over the past two centuries for solving it explicitly and analytically. With the advent of computers this field has seen much progress in techniques for numerical solutions. Physicists have been interested in a similar PDE for estimating the diffusion of heat on a linear rod away from the heat source over time as heat flows from hot to cold areas. Under certain standard conditions, the solution of the heat equation involves the Gaussian or normal density. The solution to the Black-Scholes (B-S) equation also involves the normal density and cumulative normal.

The Black-Scholes pricing formula (5.3.2) is particularly appealing because it does not involve any need to forecast the price of the stock, attitude of market participants to the stock's future prospects, utility and risk aversions of the market participants, and so on. There can be much controversy if the price depends on utility and risk attitudes on which general agreement is impossible to achieve. The PDE is completely clean and depends on the

assumptions listed above. Considerable effort has gone in relaxing the assumptions in the recent past.

In general, the PDE can be written as $af_{SS} + bf_{St} + cf_{tt} + pf_S + qS_t + u(S, t) = 0$, where the subscripts indicate the partials and the coefficients $a$, $b$, $c$, $p$, $q$, and $u$ are all unspecified functions of both $S$ and $t$ to achieve flexibility. For example, subscript $St$ means partial with respect to $S$ and then with respect to $t$, or $(= \partial S \partial t)$. The coefficients $a$, $b$, and $c$ are of a quadratic with two roots. If $b^2 - 4ac > 0$, the roots are real and the PDE is called hyperbolic, as is often used for the study of ocean waves in physics (wave equations). If $b^2 - 4ac = 0$, the solutions represent an exact square. Such PDE is called parabolic and B-S equation is of that type. Finally, if $b^2 - 4ac < 0$, PDE are called elliptic.

## 8.7 RISK-NEUTRAL CASE

Under the risk-neutrality assumption the drift parameter μ equals the risk-free bank return $r$. Then the risk-neutral diffusion equation becomes $dS = rSdt + \sigma Sdz$, where $z$ is the Wiener process or Brownian motion. Now consider $f(S, t) = \ln S$, where ln is the natural log. Recall Ito's lemma (8.4.4) with μ replaced by $r$:

$$df = \frac{\partial f}{\partial S}[rSdt + \sigma Sdz] + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2 dt + \ldots. \tag{8.6.4}$$

Applying this to $f = \ln S$, we need $\partial f/\partial S = (1/S)$ and $\partial^2 f/\partial S^2 = -1/S^2$. Now we have

$$d\ln S = (r - 0.5\sigma^2)dt + \sigma dz. \tag{8.6.5}$$

Thus starting at time $t_0$ until time $t$, the probability distribution function of future prices $S$ is given by the lognormal distribution with mean $\ln[S_0 + (r - 0.5\sigma^2)t]$ and standard deviation $\sigma t$.

We illustrate some lognormal densities graphically in Chapter 4. See Section 4.4.5 for a discussion of the basic properties of lognormal. The lognormal graphs reveal that as the mean increases, the density shifts to the right and the spread is sensitive to the variance. Thus the diffusion by the lognormal assumption leads to a particular shape. Of course, there are several possible choices of the probability distributions. Section 4.4 discusses various possible distributions. The solution to the Black-Scholes partial differential equation is stated in (5.3.2).

CHAPTER 9

# Computational Issues

## 9.1 SAMPLING, COMPOUNDING, AND OTHER DATA ISSUES IN FINANCE

The intrinsic uncertainty in finance arises because we do not know the future prices and quantities of goods, services, and financial instruments. We have to make educated guesses about future time paths of these items and invest with limited resources, using the past data in making these educated guesses.

The data in financial economics is usually a cross section of portfolios at a fixed point in time, or a time series of returns over a certain period of time. A mixture combining of both kinds of data, that is, a cross section repeated over a time period, is known as panel (longitudinal) data. Statistical analysis of any data set assumes that there is an underlying population and observed data are assumed to be a sample from that population. Thus financial economics is also subject to uncertainty arising from statistical estimation of parameters from observable samples.

In time series analysis, the entire time series represents only one realization $\{x_t\}$ of a stochastic process. In 1930s Wiener, Kolmogorov, and Khintchine (WKK) developed a highly mathematical tool based on the assumption of stationary time series to conceptualize an ensemble $\Omega$ of a very large number of time series $x_t$ for $t = 1, 2, \ldots, T$, to represent the underlying population. That theory is not applicable unless we transform the available time series to a *covariance stationary* series. The definition of covariance stationary requires that (1) both the mean and variance are finite and do not depend on time and (2) covariance at various time lags depends only on the lag itself, not on time. The word stationary is often used synonymous with "covariance stationary," and represents a steady state behavior, whereby the time series reverts to its own normal steady state fluctuations in the long run. In other words, if a sta-

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

tionary time series departs from its steady state in a temporary big swing, it reverts back to its own pattern.

Financial time series usually possess permanent upward or downward shocks. Hence their means and variances change over time, making them nonstationary. WKK theory was developed in the 1930s, under the assumption of stationarity. For example, limiting distributions for time series statistics are available since that time. One can also convert nonstationary series into stationary series by spectral decomposition and differencing of various orders. The tools of the WKK theory have since been extended and developed for certain nonstationary series. Suffice it to say that under sufficiently strong assumptions inference tools for time series are quite similar to the tools for cross-sectional data.

Now let us discuss the inference methods for cross-sectional data. Let the data consist of prices, net returns, or some quantity of interest to the financial community. A cross section of $n$ portfolios or companies, $x_i$ ($i = 1, 2, \ldots, n$) may be thought of as a sequence of independent and identically distributed (iid) random variables with a common expectation $\mu$ and variance $\sigma^2 \in (0, \infty)$. Then the central limit theorem (CLT) refers to the behavior of the sum $S_n = \Sigma_{i=1}^{n} x_i$ as the sample size increases without bound, or as $n \to \infty$. It says that when the sum is standardized by subtracting its mean and dividing by its standard deviation as $[S_n - n\mu]/[\sigma \sqrt{n}]$ it converges to N(0, 1), the unit normal random variable. Most elementary statistics textbooks suggest that $n > 30$ is a reasonable number of observations for the CLT to kick in.

As the name suggests, CLT is of central importance, since it permits statistical inference about the population parameter values such as $\mu$ and $\sigma$ from the corresponding sample statistics $\bar{x}$ and $s^2$. Confidence intervals and student's $t$ statistics are familiar tools for inference. Assume that the observed statistic is computed from a *random* sample defined with reference to a well-defined population. Section 9.3.1 discusses random number generation used for random sampling. The probability distribution of such a statistic is called the *sampling distribution*, which represents its behavior over all possible samples of $x_t$ from the population. Since all possible samples can involve billions of possibilities, sampling distribution is hard to visualize and rarely available directly. However, any inference about parameters needs knowledge of the sampling distribution. Fortunately it is possible to approximate limiting density of the statistic. The CLT justifies the use of the normal density (or its transformations) as the limiting density provided the statistic is a transformation of the sum of $n > 30$ values of $x_t$. A rigorous discussion of statistical inference about the population parameters is available in various statistics and econometrics texts, including Spanos (1999). Usually the estimation uncertainty is expressed by constructing 95% confidence intervals for unknown parameters.

### 9.1.1 Random Sampling Implicit in the Available Data

Clearly, the 95% confidence intervals and other statistical inference procedures based on the probability theory are valid only if sampling of observations on returns $R$ from the underlying population is random, not subjective or judgment based. One kind of random sampling is similar to a lottery whereby each element in the population has an equal chance of being included in the sample. In selecting a random sample in finance, care is needed to make sure that the researcher fully understands the sampling mechanism to avoid selection bias, even if it does not strictly follow the selection rules of a lottery.

### 9.1.2 Sampling Distribution of the Sample Mean

Assume that (1) We have $n > 30$ observations and (2) the returns data is a random sample. Now, when $x_i$ are financial returns from a dollar of investment in a portfolio for a certain fixed time period, the CLT assures us that the average return $\bar{x} = [S_n/n]$ is distributed normally with mean $\mu$ and variance $(\sigma^2/n)$, that is,

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \tag{9.1.1}$$

which is the sampling distribution of the sample mean. The estimation uncertainty associated with sample means can be readily known by constructing, say, 95% confidence intervals $[\bar{x} - 1.96(s^2/n), \bar{x} + 1.96(s^2/n)]$, where the multiplier 1.96 is obtained from the normal distribution tables.

### 9.1.3 Compounding of Returns and the Lognormal Distribution

In the preceding text, we have often considered returns $R$ as "net" percent return over time period $\tau$, without compounding in our definition in (2.1.3) of Chapter 2. The absence of compounding is no problem if $\tau = 1$, such as it was one day as in our artificial example of Chapter 2. If $\tau > 1$ is extended over several unit time intervals, it is not appropriate to compute the simple sum of returns over the set of individual time intervals. When we considered the pdf for $R$ as $f(R)$, we have aggregated all $R$ values in one pile without regard to the time when it occurred and computed its pdf. When we assume the net returns are standard normal, we are considering an independent and identically distributed (iid) random variable $R \sim N(0, 1)$.

Now let us consider compounding with unit investment of one dollar. This means all percent return numbers have to be divided by 100 to get the return for one dollar. In other words, 1.2% return means that \$1 becomes \$1.012 at the end of one period. It is useful to define "gross" return per unit as the gross receipt at the end of the time interval as $(1 + R_1)$ if $R_1$ is the unit (not percentage) return in the first interval. Note that it is generally safe to assume

that gross return will not be negative and $\log(1 + R_1)$ is a well-defined quantity. If $R_2$ is the second period return, upon compounding over two periods, the gross return per dollar at the end of two time periods becomes $(1 + R_1)(1 + R_2)$. Then $\log[(1 + R_1)(1 + R_2)] = \log(1 + R_1) + \log(1 + R_2)$. Let us denote multiplication by $\Pi_t$ and the natural log of gross return by the lower case $g_t$. So we have

$$\begin{aligned} g_t &= \log(1 + R_t), \\ \exp(g_t) &= 1 + R_t, \\ \log\left[\prod_t (1 + R_t)\right] &= \sum g_t, \end{aligned} \tag{9.1.2}$$

The compound gross returns over a time interval of $\tau$ periods per dollar of investment is simply the sum of individual $g_t$ values over that interval. In other words, despite compounding we can use the log transformation to define $g_t$ and still work with summations of $g_t$ instead of multiplications. The log transformation greatly simplifies the mathematics of evaluating the resulting portfolio returns despite often mandatory compounding of returns in financial time series. If, however, we assume that net returns themselves are iid random normal, $R_t \sim N(0, 1)$ variables, we cannot assume at the same time that log of gross return $[\log(1 + R_t)] = g_t \sim N(0, 1)$. Fortunately there is a ready-made density for the purpose at hand called lognormal. The density of the lognormal is

$$f(g) = \frac{1}{g\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log(g) - \mu)^2}{2\sigma^2}\right]. \tag{9.1.3}$$

If log of gross returns $g_t$ is a lognormal random variable, its mean equals $E(g) = \exp[\mu + (\sigma^2/2)]$ and the variance equals $V(g) = \exp[2\mu + \sigma^2]\,(\exp[\sigma^2] - 1])$. See also a derivation by Hull (1997) involving lognormals and Section 4.4.5.

In general, from any distribution similar to the lognormal $f(g)$ of (9.1.3) financial economists may wish to compute the value at risk. For VaR we need the formula for the appropriate quantile (1 percentile) and the loss associated with the outcome represented by that quantile. If one has price data instead of return data, it is necessary to first convert it to return data before computing the VaR estimates. Similarly care is needed in computing the net returns into log gross returns if compounding is present.

The mutual fund with the ticker symbol AAAYX studied in Chapter 2 has some characteristics of the Pearson type IV density from (2.1.9). The VaR for this density will need computation of its quantiles (e.g., one percentile), which is difficult analytically. Recall that one percentile is simply the value of the random variable such that 1% of the data lie below it and 99% of the data lie above it. Hence, if we have a numerical evaluation of the cumulative distribution function (CDF), we can readily compute the needed percentiles. Since

CDF is the area under the density curve, it needs numerical integration software, such as Gaussian numerical quadrature methods for integration. Vinod and Shenton (1996) and Shenton and Vinod (1995) illustrate the use of such quadratures. From a practical viewpoint these represent computer programs (in Gauss) for evaluation of CDFs. Hence value at risk (VaR) defined as a low percentile of (2.1.9) has also become feasible, with great potential in applied finance.

### 9.1.4 Relevance of Geometric Mean of Returns in Compounding

The *Wall Street Journal* recently reported that the arithmetic mean (AM) of returns can be 25% a year, when in fact the true return correctly given by the geometric mean (GM) is zero. It is a mathematical fact that AM ≥ GM. For example, $(a + b)/2 \geq \sqrt{ab}$, unless $a = b$, when the two sides are equal. We will see that there are situations where AM is a good approximation to the GM. Suffice it to say that it is important for every investor to know the difference between arithmetic mean and geometric mean of investment returns. Before we discuss that example, and indicate why it holds, let us develop the relevant compounding concepts in a slightly different notation than above.

For clarity, let us start with a simple example where the investment in a local bank lasts $n = 3$ years and the bank yields 5%, 4%, and 6% interest per year with annual compounding. The per dollar returns are $R_i = \{0.05, 0.04, 0.06\}$ for $i = 1, 2, 3$. Now a deposit holding of \$1 becomes $H_1 = \$1.05$ at the end of one year, $H_2 = \$(1.05)(1.04)$ at the end of two years, and $H_3 = (1.05)(1.04)(1.06)$ at the end of three years. Thus one dollar of investment becomes a holding of $\$H_n = \Pi_{i=1}^{n}(1+R_i)$ at the end of $n$ years. Note that we may want to replace the simple average of returns $R_i$ by a product of n gross returns $H_n = \Pi_{i=1}^{n}(1+R_i)$ to allow for the compounding of returns over $n$ time periods to inject greater realism.

What is the average annual return? A simple average (arithmetic mean) of 5%, 4%, and 6% yields $\overline{R} = 0.05$. In general, the arithmetic mean is an overestimate. The correct estimate of the average return is the geometric mean of gross returns minus unity, given by $(H_n)^{(1/n)} - 1 = 0.04999$. Since this result 0.04999 is very close to 0.05 and since finding the $n$th root generally requires a calculator, financial advisors have been traditionally using the arithmetic average of net returns.

One way to ease the computational burden is to use the log transformation to convert the product into a sum (log of the product is the sum of logs) and then taking the anti-log or raising the answer to the power of $e$. The log transform permits us to write the geometric mean as $(H_n)^{(1/n)} = \exp[(1/n)\Sigma_{i=1}^{n}\log(1+R_i)]$. The final expression for the average net yield based on the value of one dollar at the end of $n$ periods is obtained after subtracting the initial investment of one dollar, or average compound return (ACR), over $n$ periods:

$$\text{ACR} = \exp\left[\left(\frac{1}{n}\right)\sum_{i=1}^{n}\log(1+R_i)\right] - 1. \tag{9.1.4}$$

Recall that $\log(1 + R_t)$ was denoted by $g_t$ in (9.1.2).

Finally, the *Wall Street Journal*'s (October 8, 2003, p. D2) dramatic example used to motivate this section is as follows: A two-year investment has two returns of 100% in the first year and (–50%) in the second year. If the investor invested \$100 in this portfolio, it would be worth \$200 at the end of the first year and back to \$100 at the end of the second year, which was the original amount invested. Hence the correct return after two years is zero. The arithmetic average of two returns is vastly overestimated as (100 – 50)/2 or 25% annual return. Here $R_1 = 1$ and $R_2 = -0.5$, $n = 2$, $(H_n)^{(1/n)}$ is simple square root. Note that the geometric mean of gross returns (i.e., after adding 1 to unit investment returns) is $\sqrt{(1+1)(1-0.5)} = \sqrt{1} = 1$, and after the initial investment of \$1 is subtracted, the return is indeed zero. It is useful to check that the log formula also works. Now $(1/n)\Sigma_{i=1}^{n}\log(1+R_i)$ is the simple average of log(2) = 0.6931 and log(0.5) = –0.6931, which is zero. Hence $(H_n)^{(1/n)} = \exp[(1/n)\Sigma_{i=1}^{n}\log(1+R_i)] = \exp(0) = 1$. Thus the formula in (9.1.4) correctly yields average compound return of zero.

It is important to recognize that the choice of the arithmetic mean over geometric mean is both simpler and self-serving to the investment professionals. Since AM ≥ GM always holds, security salespeople cannot go wrong by choosing the AM. Securities professionals often allege that if an investor placed her money in S&P 500 index stocks or Dow Jones Industrials from 1927, the return will be phenomenal, with the suggestion that everyone should invest in the stock market. There are three factors that tend to overstate the case for investing in the stock market. (1) Income taxes on earnings are not included in the calculations. (2) When a corporation goes bankrupt or loses its business profitability and ceases to be a major player, they are de-listed from the stock exchange and replaced by more successful corporations in the stock indexes. This practice causes an upward bias in returns. (3) The averaging is done with the arithmetic mean and not geometric mean of gross returns minus unity.

Similar to investment professionals, statisticians also prefer to work with arithmetic means because the sampling distribution of $\overline{R} \sim \text{N}(\mu, \sigma^2/n)$, given above in (9.1.1), is so convenient, having only two moments to consider. Hence it appears to be a conspiracy between statisticians and investment professionals to rely on the arithmetic mean even though it can be so wrong. There are assumptions under which the arithmetic can be a good approximation to the geometric mean. They are discussed next.

### 9.1.5 Sampling Distribution of Average Compound Return

We now explain how the CLT can still apply after compounding, provided that the data satisfy some assumptions and allow the log transform to be used to approximate the geometric mean by the arithmetic mean. Let us further

assume that per dollar returns are small numbers, $|R_i| < 1$. This rules out the 100% return or the per dollar return of $R_i = 1$, as in the *Wall Street Journal* example above. Consider the Taylor series expansion of $\log(1 + R_i) = R_i - (R_i)^2/2 + (R_i)^3/3 - (R_i)^4/4 + \ldots$. Since $|R_i| < 1$, we can ignore terms with higher powers of $R_i$ and approximate it by the linear term $R_i$. Now $\Sigma_{i=1}^{n} \log(1+R_i)$ is approximately $S_n = \Sigma_{i=1}^{n} R_i$. Substituting in (9.1.2), we have ACR = $\exp(\overline{R}) - 1$, as the average compound return based on $n$ periods. Now recall from (9.1.1) that the sampling distribution of the mean can be written as $\overline{R} \sim N(\mu, \sigma^2/n)$. The CLT permits us to say that this same sampling distribution remains a good approximation even if the underlying variable $R$ is not normally distributed, as long as the arithmetic average $\overline{R}$ is calculated from $n > 30$ observations. Recall that (9.1.3) gives the functional form of the lognormal. This functional form is often derived by using the following well-known transformation: If $\overline{R} \sim N(\mu, \sigma^2/n)$, $\exp(\overline{R})$ is lognormally distributed. Note that it is tempting to use $\log \bar{x}$ as the transformation instead of the correct $\exp(\overline{R})$. Thus we write $\exp(\overline{R}) \sim LN(\mu, \sigma^2/n)$. The functional form of the lognormal in (9.1.3) is only slightly more complicated than that of the normal. The lognormal is well described by the first two moments. Thus CLT lets us approximate the functional form of the sampling distribution of average compound return (ACR). In short, statistical inference about population parameters is possible even if we are working with geometric means needed for compound returns.

## 9.2 NUMERICAL PROCEDURES

This section focuses on the numerical methods, including important numerical accuracy issues, that are not widely known. Computer intensive methods are attractive in empirical research due to exponentially declining costs of computing. There is a natural tendency to simply assume that if an answer is found by using a computer, it must be correct. This myth needs to be debunked, since numerical errors can mean losses in millions of dollars. Our discussion here follows McCullough and Vinod (1999, p. 638), who write: "Computers are very precise and can make mistakes with exquisite precision." In general, if large sums of money are involved, our advice is to check everything carefully and use two or more computing environments as a check on one another.

### 9.2.1 Faulty Rounding Story

A story in the *Wall Street Journal* (November 8, 1983, p. 37) about the Vancouver Stock Exchange index (similar to the Dow-Jones Index) is a case in point. The index began with a nominal value of 1000.000, recalculated after rounding transactions to four decimal places, the last being truncated so that three decimal places were reported. Intuitively, truncation to the fourth decimal of numbers measured to $10^3$ is innocuous. Yet within a few months the

index had fallen to 520, without any recession but due to faulty rounding. The correct value should have been 1098.892 according to *Toronto Star* (November 29, 1983). Consider a number 12.005, if we always round up to 12.01, it creates an obvious bias. An old solution is to round to the nearest even number, whereby 12.005 becomes 12.00 (zero is an even number) but 12.015 becomes 12.02. It is not easy to program this sophisticated rounding, and in the absence of demand for more accurate arithmetic by the common users, most software programs simply ignore the rounding problem. Ideally, math coprocessors inside a computer's box should directly round correctly without any need for human intervention through software. Use of alternative computing environments can reduce if not eliminate this and other problems discussed here.

### 9.2.2 Numerical Errors in Excel Software

Consider the familiar Microsoft Excel 2000 (v. 9.0.2720) package. Place numbers 1, 2, and 3 in the first column, and use STDDEV function to compute the standard deviation of these numbers. The correct answer is 1. Now add 99,999,999 to each number. Clearly, the standard deviation should not change by addition of any number. However, Excel gives the answer of zero! It appears that as long as the numbers are smaller than 100 million, Excel works all right. This is a serious problem with such popular software, which was pointed out to Microsoft years ago and still has not been fixed. One explanation is the monopoly that Microsoft enjoys, and perhaps a more potent explanation is that the commercial users of the software do not care enough about numerical accuracy.

### 9.2.3 Binary Arithmetic

While we humans calculate using decimal base 10 arithmetic, computers map all numbers to base 2 (binary) arithmetic (denoted by subscript (2)) where only 0 and 1 are valid numbers. For example $111_{(2)} = 1+ 2 + 4 = 7 = 2^3 - 1$, $1111_{(2)} = 7 + 8 = 2^4 - 1 = 15$, $10000000000_{(2)} = 2^{10} = 1024$. An eight bit binary number $11111111_{(2)} = 2^8 - 1 = 255$, a sixteen bit binary is $2^{16} - 1 = 65535$. The number of bits stored in computer memory limit the largest binary number possible inside a computer. The largest integer number correctly represented without rounding inside a sixteen bit computer is only 65535. The numerical inaccuracies arise because base 2 must be combined, as a practical matter, with finite storage and hence finite precision.

Consider a computer "word" with 32 bits, where a floating point binary number is often partitioned as $s \times M \times B^E$, where s is the sign, M is the Mantissa, B is the base (= 2), and E is the exponent. One bit is for the sign s, eight bits are for E, the signed exponent, which permit $2^7 - 1$ or 127 or –126 as the exponent. A positive integer Mantissa is a string of ones and zeros, which can be 24 bits long.

It can be shown that real numbers between $(1.2 \times 10^{-38}, 3.4 \times 10^{38})$ alone can be represented in a computer (McCullough and Vinod, 1999, p. 640). Numbers outside this range often lead to underflow or overflow errors. It is tempting to avoid the underflow by simply replacing the small number by a zero. The newer computers usually do this by saving it into computer registers as zero, while keeping in "mind" the correct number. The trick (not recommended by experts) is to use zero "as is" quickly, and in place, without "saving" it in computer memory registers. In general, the trick fails because, after it is saved as a zero, we can neither prevent that zero from becoming a denominator somewhere later nor rule out other unpredictable errors.

Another similar problem called log-of-zero problem arises when logarithms of a variable that can become zero are computed. In maximum likelihood estimation, hiding the log-of-zero problem by simply replacing numbers close to zero by a very small number is also not recommended. McCullough and Vinod (2003) note that this introduces a kink in the likelihood (objective) function where derivatives do not exist, vitiating the standard calculus methods of finding the maxima of a function by setting its derivatives equal to zero.

### 9.2.4 Round-off and Cancellation Errors in Floating Point Arithmetic

The best representation of the floating decimal number 0.1 in binary format is 0.09999999403953, which does round to 0.1, but the reader is warned that it is only an approximation. By contrast, the decimal $0.5 = 2^{-1}$ needs no approximation, since the binary method readily deals with powers of 2. The computers are hard wired such that when two floating point numbers are added, the number that is smaller in magnitude is first adjusted by shifting the decimal point until it matches the decimal points of the larger number and then added to efficiently use all available bits. Unfortunately, this right-shifting of the decimal introduces so-called round-off errors. A particularly pernicious "cancellation error" is introduced when nearly equal numbers are subtracted (the answer should be close to zero) and the error can be of the same magnitude as the answer. Yet another error called "truncation error" arises because software uses finite-term approximations.

Consider a long summation of $(1/n)^2$ over a range of $n = 1$ to 10,000 by computer software. The result does not necessarily equal the same sum added backward from $n = 10{,}000$ to 1. The experts know that the latter is more accurate, since it starts adding small numbers, where the truncation error is known to be less severe.

### 9.2.5 Algorithms Matter

Ling (1974) compared algorithms for calculating the variance for a large data set. Letting $\Sigma$ denote the sum from $i = 1$ to $i = n$, the numerically worst is the "calculator" formula:

$$V_1 = \left[\frac{1}{n-1}\right]\left[\sum (x_i)^2 - \left(\frac{1}{n}\right)\sum (x_i)^2\right]. \tag{9.2.1}$$

The standard formula

$$V_2 = \left[\frac{1}{n-1}\right]\left[\sum (x_i - \bar{x})^2\right] \tag{9.2.2}$$

is more precise because $V_1$ squares the observations themselves rather than their deviations from mean and, in doing so, loses more of the smaller bits than $V_2$.

Ling shows, however, that following a less familiar formula is more accurate. The formula allows one to explicitly calculate $[\Sigma(x_i - \bar{x})]$, the sum of deviations from the mean, which can be nonzero only due to cancellation, rounding, or truncation errors. The idea is to use the error in this calculation to fix the error in the variance as follows:

$$V_3 = \frac{\left[\sum (x_i - \bar{x})^2\right] - [1/n]\left[\sum (x_i - \bar{x})\right]}{n-1}. \tag{9.2.3}$$

Many students find it hard to believe that three algebraically equivalent formulas under paper and pencil algebra methods can be numerically so different, and some even laugh at the idea of adding $[\Sigma(x_i - \bar{x})]$, which should be zero.

### 9.2.6 Benchmarks and Self-testing of Software Accuracy

In light of all these problems with computer arithmetic it is useful to have a way to compare algorithms. For this we need explicit standards called benchmarks with which to compare algorithms. Longley (1967) showed that software reliability cannot be taken for granted. With the advent of the Internet, the Statistical and Engineering Division of the National Institute for Standards and Technology (NIST) has released useful benchmarks for statistical software called Statistical Reference Datasets (StRD) at: *http://www.nist.gov/itl/div898/strd* (see Rogers et al., 1998). To circumvent rounding error problems, NIST used multiple precision calculations, carrying 500 digits for linear procedures and using quadruple precision for nonlinear procedures to come up with a set of "certified" solutions.

Certain software products are known not to meet NIST benchmarks and such findings are published in software evaluation sections of the *Journal of Applied Econometrics*, *American Statistician*, and elsewhere. Unless the software vendors have solved the problems in later versions, one should avoid any software that does not meet the NIST standards. Any software user can easily test whether a particular software computes the results that are close enough

to the NIST certified values. McCullough and Renfro (1999) provides benchmarks for GARCH estimation that may be of interest to financial economists. For much of software commonly used in finance, it would be useful to have more such benchmarks with certified correct values.

### 9.2.7 Value at Risk Implications

Let $z$ be the standard normal variable, N(0, 1), and let the return $R$ be normally distributed with mean $\mu$ and standard deviation $\sigma$, $R \sim N(\mu, \sigma^2)$. Let $\alpha' = 0.025 \in [0, 1]$, where the associated quantile $Z_{\alpha'}$ is defined by the following probability statement $\Pr(z \leq Z_{\alpha'}) = \alpha'$. Equation (2.1.3) in Chapter 2 defined value at risk as $\text{VaR}(\alpha') = -R_{\alpha'} * K$, where $R_{\alpha'} = (\mu + Z_{\alpha'}\sigma)$ and $K$ denotes the capital invested. Note that $\alpha'$ is a low percent quantiles of a probability distribution. The quantiles $Z_{\alpha'}$ involve looking up the normal tables backward. With the wide availability of the Microsoft Excel program, the inverse CDF is simply one of the available functions, so the table lookup is no longer needed. Table 9.2.1 reports the quantiles from the inverse of the CDF of unit normal (NORMINV) in the column entitled quantiles. The last row of the table suggests an Excel command. For example, $\alpha' = 0.01$ level estimate in Table 9.2.1 based on Excel is $Z_{\alpha'} = -2.326347$ compared to the more accurate

**Table 9.2.1 Normal Quantiles from Excel and Their Interpretation**

| Numerator | Denominator | Proportion or ($\alpha'$) | Percent or ($\alpha'$) * 100 | Quantile $Z_{\alpha'} = \Phi^{-1}(\alpha')$ | pdf of N(0, 1) |
|---|---|---|---|---|---|
| 1 | 1000 | 0.001 | 0.1 | –3.090253 | 0.003367 |
| 1 | 500 | 0.002 | 0.2 | –2.878172 | 0.003367 |
| 1 | 100 | 0.01 | 1 | –2.326347 | 0.00634 |
| 1 | 75 | 0.013333 | 1.333333 | –2.216361 | 0.026652 |
| 1 | 50 | 0.02 | 2 | –2.053748 | 0.034216 |
| 1 | 25 | 0.04 | 4 | –1.750686 | 0.048418 |
| 1 | 20 | 0.05 | 5 | –1.644853 | 0.086174 |
| 1 | 10 | 0.1 | 10 | –1.281552 | 0.103136 |
| 1 | 5 | 0.2 | 20 | –0.841621 | 0.175498 |
| 1 | 2 | 0.5 | 50 | 0 | 0.279962 |
| 1 | 1.25 | 0.8 | 80 | 0.841621 | 0.398942 |
| 1 | 1.111111 | 0.9 | 90 | 1.281552 | 0.279962 |
| 1 | 1.052632 | 0.95 | 95 | 1.644853 | 0.175498 |
| 1 | 1.041667 | 0.96 | 96 | 1.750686 | 0.103136 |
| 1 | 1.016949 | 0.983333 | 98.333333 | 2.128044 | 0.086174 |
| 1 | 1.010101 | 0.99 | 99 | 2.326347 | 0.041452 |
| 1 | 1.001001 | 0.999 | 99.9 | 3.090253 | 0.026652 |
| 1 | 1.002004 | 0.998 | 99.8 | 2.878172 | 0.003367 |
| 1 | 1 | 1 | 100 | Infinity | 0 |
| | | = (Ai/Bi) | = Ci * 100 | NORMINV | NORMDIST |

$Z_{\alpha'} = -2.326348$ by the S-Plus software. The Excel result is wrong only in the last digit, but this can mean thousands of dollars.

In fact the Excel function NORMINV accepts arbitrary values of mean and variance to directly yield $R_{\alpha'}$. Although we have not inserted the subscript $\tau$ on the mean and standard deviation, it is understood that they are true for some specific time horizon $\tau$. Note that it makes sense to multiply the $K$ by 0.01, as in formula (2.1.4), because our data refer to the percentage returned for the time period of one day. Thus, for our artificial daily data, VaR($\tau$ = one day, $\alpha' = 0.01$) = $[2.326348(0.820537) - 0.322333] * (K * 0.01) = 1.586522$ if $K = 100$. We estimate that VaR = \$1.59 to the nearest penny.

RiskMetrics Inc. is a popular source of VaR wisdom to finance practitioners. However, it is important to note that the quantile for $\alpha' = 0.025$, from N(0, 1) in RiskMetrics publications is 1.65, which is incorrect compared to the correct number 1.644853. Again, the accuracy of these quantiles matters in financial applications, since the trailing digits can mean error of thousands of dollars in the estimate of VaR.

Any estimate of VaR obtained from historical (sample) data is subject to sampling variation, and it is not difficult to construct a confidence interval using the estimated VaR. Finance practitioners can achieve greater precision if they learn to use such intervals around several similar sample quantities such as the Sharpe ratio from Section 5.2.1. This will be discussed later in this chapter in the context of bootstrap resampling techniques for finding the limits of these intervals.

## 9.3 SIMULATIONS AND BOOTSTRAPPING

We assume that the reader is aware of randomization and computer simulations. For example, one encounters charitable events having a raffle lottery where donors buy a two-stub ticket bearing a common number. One of the stubs is placed in a barrel, which is rolled in front of the audience and usually a cute child picks up the winning number(s) from the barrel. This is a form of random selection whereby each ticket gets an equal chance of being a winner. Formally, this amounts to selecting a number from the uniform density.

### 9.3.1 Random Number Generation

More generally, random numbers can be designed to give not one but a set of items an equal chance for being selected. In those situations where repeat selection is allowed, a simple way to keep the selection chance the same is to return the item back to the barrel before each selection. This is called selection *with replacement* and will play an important role in the context of the bootstrap.

In our computer age the barrel is readily replaced by computer programs. Transcendental numbers like $e$ and $\pi$ are limits of something and can be

written out by computers to any number of digits of accuracy. The sequence of digits can be in the thousands, appears to be random, and can perhaps be used to write software for random number generation. Although there are better methods to use in practice, we use it to help explain some of the principles involved in computer-generated random numbers. For example, consider $\pi$ expanded up to 50 digits as

$$\pi = 3.14159265358979323846264338327950288419716939937 51, \quad (9.3.1.1)$$

where the number of repeats of different digits from 0 to 9 are (1, 5, 5, 8, 4, 5, 4, 4, 5, 8), respectively. If we focus on digits 1, 2, 5, and 8 only, they repeat exactly five times, implying a uniform distribution for these digits. Recall from Section 9.2 that any integer can be considered as a member of the binary system. By analogy, it can be a member of a base 4 system (modulo 4). Now we can ignore all numbers other than 1, 2, 5, and 8 from the expansion of $\pi$, change them to 0, 1, 2, and 3, and develop a computer program to generate uniform random integers.

***Four Properties of Pseudorandom Numbers.*** Although such a scheme based on $\pi$ is quite inconvenient to use in practice, we use it here to illustrate four key points about computer-generated random numbers: (1) After obtaining any set of alleged uniform random numbers, one must check whether they satisfy all properties of the underlying (uniform) distribution. The checking itself can be a complicated task. (2) The random numbers may be selected to start after ignoring a user-specified number (e.g., 6) of initial digits on the right side of the decimal in (9.3.1.1). We call the number of digits ignored *the seed* for the subsequent random numbers. (3) Note that once the seed is decided the "random" numbers in the sequence (9.3.1.1) are completely deterministic and known in advance. So we have what are accurately described as pseudo-random numbers. (4) Although the numbers from the simple sequence in (9.3.1.1) do not seem to cycle or repeat themselves, absence of cyclical behavior cannot be guaranteed when one needs billions of random numbers from different software tools. In fact it is found that most algorithms yield pseudo-random numbers that do cycle eventually. A preferred algorithm should have a cycling period equal to a very large number ($>2^{31}$).

In 1960s IBM's developed RANU, a FORTRAN computer language generator of pseudorandom numbers from the uniform density. RNDU was later criticized, and new generations of random numbers are still being developed that improve on earlier attempts. All computer-generated random numbers are necessarily sensitive to the chosen seed. A decent generator should start with a user-specified seed so that replication is possible. Some older packages keep the seed implicit (rather than user-specified) and get it from the system clock. If one is attempting serious replicable research, it is imperative that the unknown seed from the system clock not be used. The new computer programs

usually have the option to choose the seed, but they involve a few extra programming steps.

***DIEHARD Tests.*** Under a grant from the National Science Foundation, G. Marsaglia of the Florida State University developed a battery of tests for checking whether the alleged uniform random numbers are truly uniform. Marsaglia (1996) posts 18 tests called DIEHARD on the Internet to reveal hidden defects among the pseudouniform random numbers produced by a software package. Details regarding such tests are too technical for our purposes and best omitted for brevity. It is possible to test whether any given software satisfies the DIEHARD tests. For example, Vinod (2000a) found that an older (then current) version of the Gauss software passed eight tests called (1) Sparse occupancy overlapping pairs, (2) Sparse occupancy overlapping quadruples, (3) DNA, (4) Count the stream of bytes of ones, (5) Squeeze, (6) Overlapping sums, (7) Runs, and (8) Craps. However, Gauss software failed the following 10 tests, called (1) Birthday spacings, (2) Overlapping 5-permutation, (3 to 5) Binary rank for $31 \times 31$, $32 \times 32$, and $6 \times 8$ matrices, respectively, (6) Bitstream, (7) Count the individual one bytes, (8) Parking lot, (9) 3-D spheres, and (10) Minimum distance.

***Random Numbers from Arbitrary Densities.*** The next natural question is: How to generate random numbers from any well-defined probability distribution besides the uniform? The answer is surprisingly not difficult. Assuming that the inverse cumulative probability distribution function (ICDF) is well defined, one can simply start with uniform random numbers and use the ICDF to find the quantiles as random numbers from the relevant distribution. This works because the CDF covers a finite range from a probability of zero to one. Areas of the distribution with higher density will cover a wider range of the CDF, and thus be mapped with a higher likelihood using the ICDF.

There are other methods that do not use the ICDF applicable for certain densities such as the normal or Cauchy. Although this brief introduction cannot explain the complete details, it is intended to give the reader a flavor of what is involved. Since finance applications can involve large sums of money, it is important that the reader knows about publicly available tests to protect oneself from bad random number generators.

### 9.3.2 An Elementary Description of Simulations

A Monte Carlo simulation is a technique designed to generate a large number of trials of any random phenomenon. The trials are individual realizations in a simulation study. The random phenomenon from the real world often involves complex systems and randomness arising at different stages in complicated ways having complicated interactions. A large collection of simulation trials is used to separately create the random situations and eventually combined them using complicated interactions to create a (mathematical) model

of the random phenomenon. These models can have many mathematical equations and require realistic statistical assumptions regarding the underlying random probabilities. The aim of any simulation is to find an approximate probabilistic representation of the entire range of possible outcomes from a random phenomenon. In many complex situations this is the only way to understand the phenomenon and the only way to make statistical inference about error probabilities.

***Simulation Pitfalls.*** One danger in using simulation methods is that unscrupulous persons can use complex simulation models to sell a particular viewpoint or product. Besides numerical problems with random number generators, a simulation study is very sensitive to the accuracy of the assumed model, assumed random patterns, and accuracy of random number generators. Professor John Tukey of Princeton once said "Only good simulation is a dead simulation." Other critics have called it "Garbage in, garbage out." Hence we warn the reader to take a careful look at the assumptions and empirical factual basis of the simulations before accepting the conclusions of a simulation study.

***Simulating VaR.*** In Chapter 2 we discussed the value at risk calculation needed by finance practitioners. A Monte Carlo simulation can be used to know the low quantile of the key formula (2.1.4) for VaR for any complicated density as follows: We have noted elsewhere that Pearson type I density and similar complicated densities are more realistic compared to the normal density in finance applications. Often the ICDF for these densities is only available in the form of a computer algorithm (subroutine), since analytical formulas do not exist. The simulation method here first generates uniform random numbers. Then use the ICDF algorithm for the desired density to generate a large random sample from the type I or some other complicated density. Finally it is not difficult to compute the percentiles of interest from the simulation. Although early developers of VaR relied on the normality assumption, it can be readily relaxed by using the simulation steps outlined above.

***Out-of-Sample Checking of Simulations.*** In finance we can use Monte Carlo simulations for more ambitious applications than creating the percentiles of complicated nonnormal densities. We can use simulations to represent the reality of the complex and interacting financial world. For example, one can simulate asset prices and trading volumes based on historical data and/or assumptions. Large multinational corporations have several subsidiaries that deal with worldwide clients including other partly or fully owned subsidiaries. They may be selling something in one market and buying a similar product in another market. They can also use myriad sophisticated derivative securities for hedging.

All of these things are subject to market conditions, which are random. Hence no simple mathematical model can describe such multinational operations. Often the only feasible way to understand and incorporate the complexities is to create a large brute force input–output type simulation model. One needs to spell out each buying, selling, input, and output component in great detail. At the minimum any simulation model should be able to recreate the known past sample data fairly accurately within random error bands and also remain accurate for few out-of-sample time periods. A careful simulation study should only use a subset of available time periods and reserve a few known data points for out-of-sample calibration and checks of the model.

VaR can be computed by a careful listing of what can happen to (1) prices, (2) quantities bought and sold, (3) financial instruments bought and sold, (4) the firm itself through deaths of key personnel, mergers and acquisitions, natural disasters, terrorism, litigation, corruption and other unforeseen events. We recommend that these lists should be quantitative and be accompanied by corresponding probabilities, mostly based on past data. Of course, in the absence of data, expert opinions and guesses can be used. If the probabilities are conditional probabilities (given the values of related items), it is nowadays possible to create the unconditional (also known as marginal) density by using Gibbs sampling computer algorithms. Casella and George (1992) give an excellent introduction to the idea behind Gibbs sampler. It is often best to use marginal densities and let the simulation model represent the interactions.

Depending on the complexity of a portfolio, it is clearly possible to create a computer software tool (e.g., an Excel workbook) listing hundreds of possible outcomes and work through the scenarios to determine potential losses with the associated probabilities. The next step is to rank-order the exhaustive list of scenario loss numbers from the worst case to the best case. VaR is simply the low percentile of these loss numbers. Clearly, VaR computations are a useful management tool. Large entities (multinationals) need to first create a simulation model to assemble various implications of major investment decisions. They then need to create the underlying relevant probability density associated with the investment decision in conjunction with the simulation model. The next step is to focus on the loss side by computing the losses net of all gains and simulate a range of "worst-case loss scenarios," each with a corresponding probability. Finally focus on the potential loss associated with probability $\alpha'$ (= 0.01) given by the quantile is the VaR for investment decisions by large and complex entities (multinationals).

Instead of using random numbers to generate prices or quantities, in some cases it is more realistic to make the "changes" in prices or quantities random. Time series forecasting methods are often useful here. Modern software tools already create realistic simulations of airplane flights with randomness provided by several things including the weather. We claim that detailed simulations of all subcomponents can indeed handle complex business conditions. However, we warn against unnecessary or ill-understood complexity. The simulation design should provide an option to obtain results for special cases

when some nonlinear relations become linear and some nonnormal densities become normal. It is best to simplify as much as possible and avoid intercorrelations among prospects. For handling unavoidable covariances methods based on extensions to (2.1.10) can be used.

### 9.3.3 Simple iid Bootstrap in Finance

Here we review Efron's (1979) method, called the bootstrap. The bootstrap literature has made tremendous progress in solving an old statistical problem: making reliable confidence statements in complicated small sample, multiple-step, dependent, nonnormal cases. The appeal of bootstrap is that it substitutes raw computing power for intricate statistical or econometric theory, without sacrificing the quality of inference.

Some of the early applications of the bootstrap in finance are worth mentioning. Stock market data often need estimators of stable law parameters, whose sampling distribution can be studied with the help of the bootstrap, as in Akgiray and Lamoureux (1989). Schwartz and Torous (1989) estimate standard errors of maximum likelihood estimates in a study of prepayment and valuation of mortgage-backed securities. Affleck-Graves and McDonald (1990) report an application to multivariate testing of the capital asset-pricing model (CAPM). Their tests include maximum entropy methods, designed to deal with the singularity when the number of assets included in the CAPM exceeds the number of time series data points. Bartow and Sun (1987) report confidence intervals for relative risk models. Next we explain these older and many new applications of the bootstrap in finance in the context of a simple regression application.

### 9.3.4 A Description of the iid Bootstrap for Regression

Here we describe the bootstrap for the multiple regression model. We refer the reader to Appendix A of this chapter for a full description of the model in matrix notation. Let us consider the usual regression model:

$$y = X\beta + \mu, \quad E(u) = 0, \quad E(uu') = \sigma^2 I, \tag{9.3.1}$$

where $y$ is a $T \times 1$ vector of the dependent variable, $X$ is a $T \times p$ matrix of $p$ regressors where the first column consists of all ones if an intercept is present in the model. The matrix $X$ has rank $p$ if the regressors are not collinear. If the rank of $X$ is deficient, the regression coefficients $\beta_1, \beta_2, \ldots, \beta_p$, which are the individual elements of the $p \times 1$ vector $\beta$, cannot be estimated. The vector $u$ is a $T \times 1$ vector of independent and identically distributed (iid) true unknown errors with elements $u_t$ and unknown possibly nonnormal true distribution function $F$ satisfies $E(u) = 0$, $E(uu') = \sigma^2 I$; that is, they have mean zero and variance $\sigma^2$.

The ordinary least squares (OLS) estimator of β minimizes the error sum of squares $u'u$ and the solution is $b = (X'X)^{-1}X'y$. If we plug in these values in the model (9.3.1), we can obtain the fitted values of the dependent variable denoted by $\hat{y} = Xb$. Hence the vector of estimated values of true unknown errors $u$, denoted by $e$ is simply $y - \hat{y}$, that is, $e = y - Xb$. The statistical properties of the OLS estimator $b$ depend crucially on the covarience matrix of $u$ given by $E(uu')$, which determines the covariance matrix of $b$:

$$\text{cov}(b) = s^2(X'X)^{-1}, \quad s^2 = \frac{e'e}{T-p}. \tag{9.3.2}$$

An empirical cumulative distribution function ECDF of OLS residuals puts probability mass $1/T$ at each $e_t$. Let the ECDF of observable residuals be denoted here by $F_e$. Now a basic bootstrap idea is to use $F_e$ which has mean zero and variance $\sigma_e^2$ as a feasible, approximate, nonparametric estimate of the unobservable CDF of the true unknown errors denoted by $F_u$.

***Shuffling of Observed Residuals.*** Let $J$ be a suitably large number (e.g., = 999). We draw $J$ sets of bootstrap samples of size $T$, with elements denoted by $e_{jt}^*$ ($j = 1, 2, \ldots, J$ and $t = 1, \ldots, T$) from $F_e$ using random sampling with replacement. To understand this sampling, imagine $T$ tickets marked with $e_1$ to $e_T$, the observed OLS residuals. Next, imagine drawing a random sample $j = 1$ of size $T$ from a box containing the $T$ tickets having elements $e_{jt}^*$ ($j = 1$, $t = 1, 2, \ldots, T$). We imagine shaking the box vigorously and replacing the selected residual back to the box to make sure that each residual $e_t$ has exactly the same chance of being the next element, that is,

$$P\left(e_{jt}^* = e_t\right) = \frac{1}{T}. \tag{9.3.3}$$

The shuffling is *with replacement* because we replace the selected residuals back into the box. It makes sure that (9.3.3) holds true, even though it may intuitively seem peculiar. Although each sample has $T$ elements, any given resample could have some of the original $e_t$ represented more than once and some not at all.

Now select a second random sample again of size $T$ from the same box, to yield new elements $e_{jt}^*$ ($j = 2$, $t = 1, 2, \ldots, T$). Continuing the sampling with replacement $J$ times generates $J$ sets of $T \times 1$ vectors denoted by $e_j^*$ having elements $e_{jt}^*$ ($j = 1, 2, \ldots, J$ and $t = 1, 2, \ldots, T$). In practice, there is no box but a computer program that shuffles the observed residuals $e_t$ with replacement to create $J$ sets of $T \times 1$ vectors.

It is important to recognize that the time series properties of residuals arising from time sequence associated with the numbers are completely lost in the iid bootstrap shuffle. Formally this means that the autocorrelation function (ACF) and partial autocorrelations (PACF) of the original time series (of

residuals) are not retained in bootstrap shuffle. In other words, all information contained in the time subscript is lost, and the shuffling treats them as independent and identically distributed without verifying that they are so. Moving blocks bootstrap (MBB) attempts to solve this problem with time series bootstrap by keeping blocks of time series together in the shuffle. It assumes that as long as the time subscript is more than $m$ lags apart, it can be shuffled around.

There is vast theoretical statistical literature on MBB, although the practical experience with MBB has not been very encouraging. Davison and Hinkley (1997, ch. 8) provide detailed explanations of MBB and related ideas in a chapter dealing with complex dependence in time series. Despite being rather difficult to implement, MBB assumes that history repeats itself by joining the time series into a continuous circle of outcomes and that the range of outcomes is restricted to the observed range. In finance it is unrealistic to assume that the future behavior of markets will be restricted to the observed range of outcomes. Appendix C provides a new bootstrap that extends the range of data beyond the observed range and remains applicable to dependent time series. Vinod (2004) has proposed a novel method using maximum entropy density to deal with this problem and shown an application in the *Journal of Empirical Finance*. We briefly describe the ME bootstrap in Appendix C at the end of this chapter.

***Creation of* J *Regression Problems from Shuffled Residuals.*** Recall that the shuffling has created vectors $e_j^*$ ($j = 1, 2, \ldots, J$). It is convenient to use the notation * to refer to a bootstrap shuffled value in this section. Now we create $J$ regression problems by creating $J$ sets of data for the dependent variable, called pseudo $y$ data denoted by $y_j^*$ by simple addition of the shuffled residuals to the fixed vector of OLS fitted values $\hat{y} = Xb$. We write

$$y_j^* = Xb + e_j^*, \qquad j = 1,2,\ldots,J, \tag{9.3.4}$$

yielding a large number $J$ of regression problems. Note that there is no * for $\hat{y} = Xb$ in (9.3.4), since these values remain fixed for all shuffles. This implicitly assumes validity of the linear regression model leading to the fitted values $\hat{y}$. Each new regression problem has a new vector of resampled residuals. Each such regression problem yields a new estimate $b^*$ of the vector of regression coefficients $\beta$.

Now shuffling and re-estimation has given us a large set $b_j^*, j = 1, 2, \ldots, J$, for such estimates. The individual elements $(b_1, b_2, \ldots, b_p)$ of the $b_j^*$ vector can be ordered from the smallest to the largest to give an approximate, brute force computer-intensive, sampling distribution of the statistic $b$. This is what is used for the bootstrap inference described below.

By definition, the variance of any random variable z is computed as $\text{var}(z) = \Sigma[z - E(z)]^2 P(z)$. If its mean is zero, we have $\text{var}(z) = \Sigma(z)^2 P(z)$. The regression residuals have mean zero, $E(e_t) = 0$, and variance $s^2$, $\text{var}(e_t) = s^2$, by the

properties of least squares estimation. Hence, if we consider an expectation of a sample of $T$ observations over the bootstrap shuffles, $E(e^*_{jt}) = 0$ also holds because from (9.3.3) the probability $P(e^*_{jt} = e_t)$ is $1/T$. So the expectation over the bootstrap shuffles of the variance of shuffled residuals is proportional to the variance $s^2$ of the original regression residuals:

$$\sigma^2_{e^*} = E\left(e^*_{jt}\right)^2 = \sum_{t=1}^{T}\left(\frac{1}{T}\right)\mathrm{var}(e_t) = \left(\frac{1}{T}\right)\sum_{t=1}^{T}(e_t)^2 = \frac{s^2(T-p)}{T}, \qquad (9.3.5)$$

where the factor of proportionality is $(T - p)/T$. Sometimes the residuals are rescaled by the square root of $T/(T - p)$, that is, $(e^*_{jt}) \rightarrow \sqrt{[T/(T-p)]}\,(e^*_{jt})$. Now we see from (9.3.5) that the variance of the rescaled $e^*_{jt}$ becomes $s^2$. Since $T$ is usually large and the number of regressors $p$ is usually small, $(T - p)/T \approx 1$, the rescaling is usually unnecessary. Some simulation studies in the literature have shown that rescaling can be omitted in practice.

In Efron's (1979) construction, apart from the eye-catching name, keeping the mean zero and variance $s^2$ was the critical selling point of the bootstrap. The bootstrap is sometimes also called resampling (with replacement) because it uses the same sample data over and over again. Note that once we have a brute force representation of thousands of reincarnations of the observed statistic from the resamples, we need not assume or use complicated asymptotic theory for statistical inference about the range of possible values of $b$ due to random sampling. The usual asymptotic theory exploits the following property: when $z \sim \mathrm{N}(0, 1)$ is standard normal, the 95% confidence interval for the population mean $\mu$ is $[-1.96, 1.96]$. Once we have the bootstrap, we do not have to have the shape of the sampling distribution of the statistic $b$ to be normal to write convenient confidence intervals. In fact the sampling distribution of the statistic need not even be from any well-known family at all. We can directly try to empirically approximate the entire sampling distribution of $b$ by investigating the variation of $b$ over a large number $J$ of bootstrap reincarnations.

The initial idea behind bootstrapping was that a relative frequency distribution of $b$ calculated from the resamples can be a good approximation to the sampling distribution of $\beta$. This idea has since been extended to conditional models and one-step conditional moments. Recall that $b^*$ denotes resampled regression coefficients, $b$ the original coefficients, and $\beta$ the unknown parameters. The extended bootstrap approximates the unknown sampling distribution and other properties of $(b - \beta)$ from the $J$ reincarnations of the observable $(b^* - b)$. See Davison and Hinkley (1997) for recent references and Vinod (1993) for older references, including some applications in finance.

***Downside Risk and Estimation Risk.*** An important point made in this section is to show that besides risk associated with the ups and downs of the market, there is also some risk associated with the fact that the characteristics

of the probability distribution of market returns is not known exactly but estimated from random data. We suggest a new way of incorporating estimation risk with an additional denominator to the estimated Sharpe ratio.

We start with some notation from financial economics. Let $r_{i,t}$ represent the excess return from the $i$th portfolio in period $t$, where $i = 1, 2, \ldots, n$. A random sample of $T$ excess returns on the $n$ portfolios is then illustrated by $r'_t = [r_{1t}, r_{2t}, \ldots, r_{nt}]$, where $t = 1, 2, \ldots, T$, and where $r_t$ is assumed to be multivariate normal random variable, with mean $\mu = \{\mu_i\}$, $i = 1, 2, \ldots, n$, and a covariance matrix $\Sigma = \{\sigma_{ij}\}$, where $i, j = 1, 2, \ldots, n$. It is well known that the unbiased estimators of the $(n \times 1)$ mean vector and the $(n \times n)$ covariance matrix are, respectively,

$$\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t,$$
$$S = \{s_{ij}\} = \frac{1}{T-1}\sum_{t=1}^{T}(r_t - \bar{r})(r_t - \bar{r})'. \tag{9.3.6}$$

These two estimators are then used to form the estimators of the traditional Sharpe performance measure.

The population value of the Sharpe (1966) performance measure for portfolio $i$ is defined as $\mathrm{Sh}_i = \mu_i/\sigma_i$, $i = 1, 2, \ldots, n$. It is simply the mean excess return over the standard deviation of the excess returns for the portfolio. The conventional sample-based point estimates of the Sharpe performance measure are

$$\hat{\mathrm{Sh}}_i = \frac{\bar{r}_i}{s_i} \quad \text{for } i = 1, 2, \ldots, n. \tag{9.3.7}$$

The (sample) Sharpe ratio is defined in equation (9.3.7) as the ratio of the mean excess return to their standard deviation. Vinod and Morey (2001) define a "double" Sharpe ratio to incorporate the estimation risk as

$$\mathrm{DSh}_i = \frac{\hat{\mathrm{Sh}}_i}{s_i^{\mathrm{Sh}}}, \tag{9.3.8}$$

where the denominator $s_i^{\mathrm{sh}}$ is the standard deviation in the sample of the Sharpe ratio estimate, or the estimation risk. As is clear in (9.3.8), the double Sharpe penalizes a portfolio for higher estimation risk by placing it in the denominator.

Because of the presence of the random denominator $s_i$ in the definition of (9.3.7), the original Sharpe ratio does not permit an easy method for evaluating the estimation risk in the point estimate itself. This is because the

finite-sample distribution of the Sharpe measure is nonnormal, and hence the usual method based on the ratio of the statistic to its standard errors is biased and unreliable. The same problem continues in (9.3.8) for formal statistical inference based on the double Sharpe ratio.

Vinod and Morey (2000) report a bootstrap resampling for the Sharpe ratio discussed earlier in Section 5.2.1 based on resampling of original excess returns themselves for $j = 1, 2, \ldots, J$ ($J = 999$), times. This means we have $J$ distinct Sharpe ratios from the single original excess return series. The choice of the odd number 999 is convenient, since the rank-ordered 25th and 975th values of estimated Sharpe ratios arranged from the smallest to the largest readily yield a so-called naïve 95% confidence interval. There is considerable literature on improving these naïve intervals reviewed in Davison and Hinkley (1997), and others. Unfortunately, a typical 95% confidence interval may not cover the true parameter with the coverage probability of 0.95.

Before we consider a Sharpe ratio application, we need to explain the idea of "first-order correct" confidence intervals in the general context. Consider a standard statistic similar to the sample mean or a regression coefficient $b$ that estimates $\beta$ with the standard error $S$. Now the quantity $q = (b - \beta)/S$ is pivotal, in the sense that its distribution does not depend on the unknown parameter $\beta$. Under normality of regression errors, $q$ has the well-known $t$ distribution leading to the usual confidence intervals. We wish to relax the normality assumption, the simple bootstrap confidence intervals that are "first-order correct," meaning correct to order $(1/\sqrt{T})$ when $T$ is the sample size. For further improvements, we need further adjustments to make sure that it covers the unknown parameter with the correct coverage probability of 0.95.

Since the Sharpe ratio of (9.3.7) and Treynor index have random denominators, Vinod and Morey (2000) argue that their sampling distribution is nonstandard and recommend studentized bootstrap and double bootstrap methods. The following two subsections briefly describe the key ideas behind them.

***Studentized Bootstrap (boot-t).*** Studentized bootstrap (boot-$t$) is implemented as follows: The first step is to convert the $J = 999$ bootstrap resampled values into a studentized scale by subtracting the mean and dividing by the standard deviation of the J resampled values. Next we select the 25th and 975th ordered values from the studentized transformed scale. Then we undo the studentization by multiplying 25th and 975th values by the standard deviation and adding the mean. Davison and Hinkley (1997) offer an elegant proof that the boot-$t$ is second-order correct, in the sense that the error in coverage by the estimated confidence interval of the true parameter is of order $(1/T)$. This second-order accuracy result depends on availability of a reliable estimate of the standard deviation used as the denominator of the studentization transformation. The method described below is more computer intensive and used sparingly when the estimated standard deviation is unreliable and the sampling distribution is highly nonnormal.

***Double Bootstrap* (d-*boot*).** Beran (1987) invented the idea of bootstrapping the bootstrap or double bootstrap. The first econometric application (to shrinkage estimators) is found in Vinod (1995). The *d*-boot involves $K$ further replications of each $j = 1, 2, \ldots, J$. If $J = 999$, the optimal choice is $K = 174$, as described in McCullough and Vinod (1998) who give a step-by-step description of the *d*-boot. Clearly, the *d*-boot is computationally intensive, since it will require 174 times 999 computations of the underlying statistic.

Roughly speaking, the method used in the *d*-boot is the following: If the original estimate is $\hat{\theta}$ and the single bootstrap estimate is denoted by $\theta_j$, we resample each $\theta_j$ as many as $K$ times and denote the resampled estimates as $\theta_{jk}$. The software keeps track of the proportion of times $\theta_{jk}$ is less than or equal to $\hat{\theta}$. The theory of *d*-boot proves that under ideal conditions, the probability distribution of the proportions in which $\theta_{jk}$ is less than or equal to $\hat{\theta}$ must be a uniform random variable. The appeal of the double-bootstrap method is that when conditions are not ideal, the resulting nonuniform distribution remains easy to handle. All that happens is that the probabilities 0.025 and 0.975 occur at some nearby numbers (e.g., 0.028 and 0.977). For adjusting the single bootstrap (*s*-boot) confidence interval the *d*-boot involves choosing the nearby (e.g., 28th and 977th) order statistics instead of the 25th and 975th values indicated by *s*-boot. The point in using the *d*-boot is that one does not need to try to know the form of the nonnormal sampling distribution of the statistic. In fact the distribution need not have any known form. Vinod and Morey (2000) report an application of *d*-boot to confidence intervals on the Sharpe and Treynor performance measures for several well-known mutual funds.

We have taken specific examples to show that simulation and bootstrap are important tools for studying the uncertainties in the markets as well as studying the uncertainties in our estimates of market indicators and our favorite tools for future changes in the market. Bootstrap is a particular kind of simulation designed for statistical inference without assuming functional forms of probability distributions involved. Bootstrap lets the data help decide the appropriate probability distribution, which is rarely normal as the classical theory would have us believe.

We conclude the section on simulation and bootstrap with following observations. In the olden days the only way to deal with market uncertainty was to simply accept it and call it good or bad luck. It was as if we are playing dice with nature. With the advent of computers and probability theory, we can pretend to play that game of dice with nature hundreds of times without actually risking any money. The main advantage, of course, is that the gains and losses in the make-believe simulation games played with the help of computers are not real, and yet the risk taker can learn a lot about potential losses and gains.

## APPENDIX A: REGRESSION SPECIFICATION, ESTIMATION, AND SOFTWARE ISSUES

In the Appendix to Chapter 1 we discussed the estimation issues related to the simple regression model $y = \alpha + \beta x + \varepsilon$ without using the matrix notation, since there was only one regressor variable $x$. A great many software products are available for regression, and there are very important numerical accuracy issues with many of them.

To run this regression in an Excel workbook, with the understanding that its numerical reliability is questionable, we use the following steps:

### Microsoft Excel Steps for Computing Regressions

**Step 1.** Enter the $y$ and $x$ data in workbook columns. Note if column labels are present.

**Step 2.** *Preliminary step*. This step might not be needed. In Excel "Tools" menu see if you see "data analysis" as an option, if not, click on "Add-ins . . ." of the "Tools" menu. If you click into the box on the left side of "Analysis ToolPak," it places a check mark there. Similarly, click along the box on the left side of "Analysis ToolPak—VBA." Finally, click on the OK button to get out of the menu.

**Step 3.** Click on "Tools" menu, then on "Data Analysis. . . ." Next, highlight "Regression" and click OK.

**Step 4.** Fill in the cell range for "Input-Y range:" Next, fill in the cell range for the regressor in "Input-X range:" area. Check the box for "Labels" if data labels are present. Choose starting cell location where the output should appear, making sure that there is enough blank space to the right. Excel output gives most of the needed estimates: regression coefficients, their standard errors and confidence intervals, adjusted $R^2$ for assessing the overall regression fit, and so on.

Consider a multiple regression model $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ where there are two regressors $x_1$ and $x_2$. This is a simple extension of the simple regression model. Its Excel implementation is also simple. In step 1 above, it is necessary to line up all regressor columns next to each other and in step 4 the range for the entire set of regressor columns needs to be indicated. As before, the coefficient estimates and most relevant estimates are readily printed out by Excel software. Note that Excel can handle several regressors without difficulty.

### Rescaling of Variables for Numerical Accuracy

One important step for avoiding serious numerical problems is to make sure that the units of measurement are similar for $y$ and for all regressors. This may mean rescaling some regressors by multiplying them by a constant such as 0.01,

0.1, 10, 100, and 1000. It can be shown that numerical problems are less severe if all variables are standardized so that they are measured from their means in units of their standard deviations. This scaling is achieved by subtracting the mean and dividing by the standard deviation.

***Collinearity and Near-collinearity of Regressors.*** If one regressor is nearly proportional to another, the estimated regression coefficients will not be reliable at all. More generally, if one regressor is a weighed sum of two or more regressors, then also the regression coefficients are unreliable. Many numerical problems can arise in these situations. Good software programs warn the user about the presence of collinearity.

When the investment of large sums is being considered with the help of a regression equation, getting numerically most accurate software and checking the results on more than one software platforms becomes important. Free software called R and its commercial cousin S-Plus appears to have good numerical reliability, as does TSP. Excel has important limitations and is generally not recommended for serious work. However, Excel can provide good initial estimates of regression coefficients (subject to care in avoiding bad scaling of variables and collinearity issues) to a researcher in finance who may not be familiar with professional statistical software. Since the Microsoft Excel is almost universally available, it offers great convenience for our pedagogical descriptive purposes here. Our use of Excel here should not be interpreted as our endorsement of it in serious research due to known numerical inaccuracies.

***Column of Ones Instead of the Intercept.*** For some theoretical and pedagogical purposes it is convenient to change the notation slightly and rewrite the two regressor model $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ without explicit reference to the intercept $\alpha$ by merging it into just another regression coefficient. We completely change the notation and write it as $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$, where the old $\alpha$ is now $\beta_1$, a new artificial column of regressor data where all elements are unity is now called $x_1$. The old $x_1$ now becomes $x_2$ and the old $x_2$ becomes $x_3$. Verify within Excel that this change gives the same results provided one clicks on a box left of "Constant is zero" in the regression menu. This is next to the Labels box of the regression menu of the "Data analysis . . ." option.

The column of $T$ data points (say) on the dependent variable $y$ is denoted by the $y$ vector. Note that the software already works with several columns of regressor data as a whole and treats the entire collection of numbers regarded as regressors in a prespecified columnwise format. The set of numbers for regressors (without the column headings) is called the $X$ matrix. Matrix algebra deals with the rules for operating on (adding, multiplying, dividing) matrices and vectors. The rules depend on the number of rows and number of columns of each matrix or vector object.

### Multiple Regression in Matrix Notation

Regression model remains the main work horse in many branches of finance. We have used the matrix notation in (9.3.1) for multiple regression models. For the convenience of some readers who may not be comfortable with it, this section of the Appendix gives further details of multiple regression in matrix notation. First, it is useful to consider $y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$, stated above, in the vector notation. We let each regressor variable $x_i$ for $i = 1, 2, \ldots, p$, with $T$ observations be a $T \times 1$ vector. Similarly we let $y$ and error term $\varepsilon$ also be $T \times 1$ vectors, and note that the regression coefficients $\beta_1$ to $\beta_p$ are constants, known as $1 \times 1$ scalars. The expression $y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$ can be written in matrix notation as follows, with matrix dimensions indicated as subscripts:

$$y_{(T\times 1)} = X_{(T\times p)}\beta_{(p\times 1)} + \varepsilon_{(T\times 1)}. \tag{9.A.1}$$

The matrix expression $X_{(T\times p)}\beta_{(p\times 1)}$ represents $\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$ due to the so-called row-column multiplication. Note that after multiplication the resultant matrix is of dimension $T \times 1$. This is critical, since when one adds matrices on the right hand side of (9.A.1), each part of the addition has to have exactly the same dimension $T \times 1$.

For statistical inference, it is customary to assume that regression errors have mean zero, $E\varepsilon = 0$. Let $\varepsilon'$ denote the $1 \times T$ matrix representing the transpose of $\varepsilon$ so that $\varepsilon\varepsilon'$ is a $T \times T$ matrix from the outer product of errors. Upon taking expectation, we have $\Omega$, representing the $T \times T$ variance covariance matrix of errors, that is, $E\varepsilon\varepsilon' = \sigma^2\Omega$. If the errors are heteroscedastic, the diagonal elements of $\Omega$ are distinct from each other, and if there is autocorrelation among the errors, $\Omega$ has nonzero off-diagonal elements.

***CAPM from Data.*** When one tries to represent theoretical finance with the help of available data, there are always some practical problems, many of which have been already mentioned in the text above. For example, the representation of the market portfolio by a market index (S&P 500 or Vanguard 500 index fund) for the capital asset pricing model (CAPM) remains controversial. Some of the objections are as follows:

1. The index composition changes over time because removing failed corporations implies a bias in favor of successful companies.
2. There are many small corporations excluded, so the index is not truly comprehensive.
3. It ignores many classes of asset markets, among these bonds and real estate.

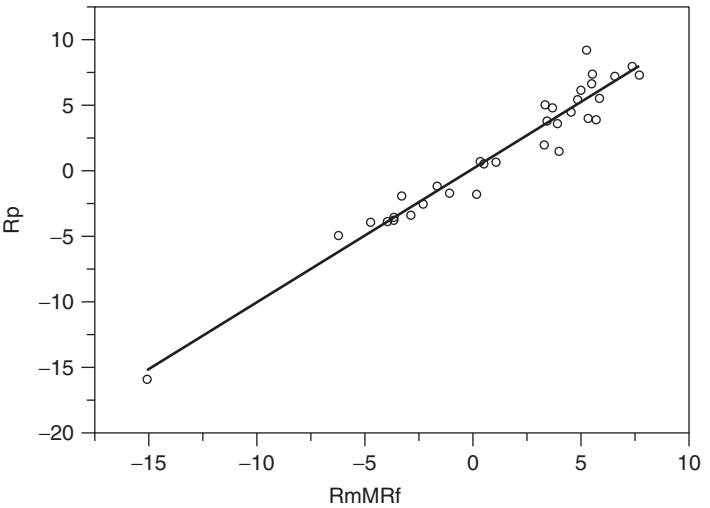



**Figure 9.A.1** Scatter plot of excess return for Magellan over TB3 against excess return for the market as a whole (Van 500 *minus* TB3), with the line of regression forced to go through the origin; beta is simply the slope of the straight line in this figure

### Illustration of CAPM Estimation of Wall Street Beta from Data

The data from Table 2.1.4 can be used to estimate the CAPM model and to compute the beta for Fidelity's Magellan fund provided that we make following assumptions. We assume that a good proxy for "market" return is the Vanguard 500 index fund (Van 500) containing 500 stocks in the S&P 500 index with proportions of stock quantities close to the proportions in S&P 500. We also assume that interest rates on three-month treasury bills (TB3) are a good proxy for risk-free yield. The simplest method of finding the beta for the risky portfolio represented by Fidelity's Magellan fund is to regress (Magellan fund *minus* TB3) on a column of ones for the intercept and (Van 500 *minus* TB3). This expands the right-hand side of (2.2.9) by a regressor for the intercept. Now the beta for the risky asset, Magellan fund is simply the regression coefficient of the second regressor (Van 500 *minus* TB3). This regression equation can be estimated in an Excel workbook itself.

The theoretical formula does not include any intercept (constant term) in the regression equation. Its estimation is accomplished by checking the box "Constant is zero" in the "Data analysis" part of Excel's "Tools" menu. One may need to use "Add-ins" to make sure that the "Data analysis" option is present in the "Tools" menu. Forcing the intercept to be zero is also described as forcing the regression line through the origin in some statistics textbooks. It is well known that zero intercept regressions are subject to numerical accuracy problems. For example, $R^2$, the measure of overall fit called squared multiple correlation, can sometimes become negative and be unreliable. Hence it is advisable to run two regressions, one with and one without the intercept.

Next, we suggest comparing the two estimates of beta the regression coefficient of (Van 500 *minus* TB3) and the two estimates of the $R^2$. If these two estimates differ too much, we can assume the presence of serious numerical accuracy problems needing further investigation by software tools not readily available in Excel.

For our data, if the extra nonzero intercept is included, $R^2 = 0.948$, and the estimate of $\beta$ denoted by $\hat{\beta} = 1.0145$, with the Student's $t$ statistic 23.04. The estimate of the intercept is 4.7734 with the $t$ statistic of 1.4392, which is statistically insignificant. Since one of the claims of CAPM is that the intercept will be zero, if the data yield an insignificant intercept, this supports CAPM.

If the intercept (constant term) is forced to be zero, $R^2 = 0.945$ and $\hat{\beta} = 1.0243$ with the student's t statistic 23.15. We conclude that the regression fit is excellent; the high $t$ values mean that the beta is statistically highly significant, so we reject the null hypothesis $\beta = 0$. Since the two estimates of $\beta$ and $R^2$ do not differ much, we can reliably accept the estimate $\hat{\beta} = 1.0243$.

In light of the discussion of equation (2.2.9) we note that in this application we are more interested in testing whether the null hypothesis $H_0$: $\beta = 1$. That is, we want to know whether the risk of the Magellan fund is as much as the risk of the market portfolio. Since hypothesis testing is more reliable for the model with the intercept, we use the Student's $t$ statistic revised as $[\hat{\beta} - 1]/\text{SE} = 0.3295$, where the standard error (SE) is 0.0440. This is statistically insignificant, suggesting that the beta for Magellan is close enough to unity and the observed small difference could have arisen from random variation. The simple average of the returns in Table 2.1.4 given along the bottom row suggest that the risk premium of the Magellan over TB3 (= 1.948 – 0.411 = 1.537) is not small. However, the Van 500 average return 1.927 is close to 1.948 for the Magellan. The usual risk measured by the standard deviation for the Magellan fund ($s_m = 5.2201$) exceeds the risk of Van 500 ($s_v = 4.9609$).

However, are the two standard deviations statistically significantly different? Could the $s_m$ arise from a population with standard deviation $\sigma_0 = 4.9609$ (based on the Van 500) estimate? The test statistic $T_m = (n-1)(s_m)^2/(\sigma_0)^2$, where $n = 33$, the number of observations in the data. Now $T_m = 35.43014$ follows a $\chi^2$ distribution with $(n - 1)$ degrees of freedom with a tabulated value of 46.19426 for a 5% tail. We conclude that the Magellan fund's standard deviation is not significantly different from 4.9609. Similarly we can define $T_v = (n-1)(s_v)^2/(\sigma_0)^2 = 28.90195 < 46.19426$ the tabulated value. We conclude that the two variances are not statistically significantly different. The point to remember is that the risk measured by beta is distinct from the risk measured by the standard deviation of the distribution of returns.

## APPENDIX B: MAXIMUM LIKELIHOOD ESTIMATION ISSUES

Sir R.A. Fisher invented the following idea. When the same joint density is viewed as a function of parameters for given data, it is called the likelihood

function. Fisher also invented the maximum likelihood (ML) estimator of the parameter vector by maximizing the likelihood function. We begin with a short review of the likelihood function itself. Let $Y$ represent height in inches. Let $y_1 = 63$, be a single observation from $y \sim N(\theta, v)$, where $v$ equals the usual $\sigma^2$. Now the likelihood function is $L(\theta, v|y) = (2\pi v)^{-1/2}\exp(-[y_i - \theta]^2/2v)$ for $i = 1$. The natural log of the above likelihood function is $LL = -(1/2)\log(2\pi) - (1/2)\log v - [y_i - \theta]^2/2v$.

This example shows that the likelihood function is simply the original normal density function written as a function of the unknown parameters given the data. If there are $y_1$, $y_2$, and $y_3$ as three observations from the same density, and the observations are independent of each other, the joint density is simply a product of individual densities. Then

$$L(\theta, v|y) = (2\pi v)^{-3/2}\exp\left(-\frac{[y_1-\theta]^2}{2v}\right)^*\exp\left(-\frac{[y_2-\theta]^2}{2v}\right)^*\exp\left(-\frac{[y_3-\theta]^2}{2v}\right). \quad (9.B.1)$$

Clearly, we can generalize it to any positive integer $T$. The aim of maximum likelihood estimation is to maximize this function with respect to the parameters. Since a logarithm is a monotonic transformation, it can be shown that maximizing $L$ and $LL$ give exactly the same answer. For certain densities from the exponential family, maximizing $LL$ is more convenient.

Since exp($a$)*exp($b$)*exp($c$) is exp($a + b + c$), the log likelihood with $T$ (= 3) observations is

$$LL = -\left(\frac{T}{2}\right)\log(2\pi) - \left(\frac{T}{2}\right)\log v - \sum_{i=1}^{T}\frac{[y_i-\theta]^2}{2v}.$$

The partial derivative function is called "the score" function and is given by

$$\text{Score}(\theta) = \frac{\partial LL}{\partial \theta} = \sum_{i=1}^{T}\frac{2[y_i-\theta]}{2v} = \sum_{i=1}^{T}\frac{[y_i-\theta]}{v}, \quad (9.B.2)$$

$$\text{Score}(v) = \frac{\partial LL}{\partial v} = -\left(\frac{T}{2}\right)\left(\frac{1}{v}\right) + \sum_{i=1}^{T}\frac{[y_i-\theta]^2}{2v^2}. \quad (9.B.3)$$

The first-order condition (FOC) for maximizing the likelihood function is that the score functions are both zero. That is we obtain ML estimates (denoted by the subscript ML) by solving two equations:

$$\sum_{i=1}^{T}\frac{[y_i-\theta]}{v} = 0 \quad \text{and} \quad -\left(\frac{T}{2}\right)\left(\frac{1}{v}\right) + \sum_{i=1}^{T}\frac{[y_i-\theta]^2}{2v^2} = 0$$

for $\theta$ and $v$. We need to simplify them to find the ML estimates. First consider

$$\text{Score}(\theta) = \sum_{i=1}^{T} \frac{[y_i - \theta]}{v} = 0 = \sum_{i=1}^{T} [y_i - \theta] = \sum_{i=1}^{T} [y_i] - \sum_{i=1}^{T} [\theta] = 0.$$

Clearly, the solution is $\theta_{ML} = \Sigma_{i=1}^{T} y_i/T$, which is the sample mean. Next, consider Score($v$), where $2v$ cancels from the denominator, leaving $-T + \Sigma_{i=1}^{T}[y_i - \theta]^2/v = 0$. Substituting $\theta_{ML}$ solution of the first score equation, we have the ML estimate of v given by $v_{ML} = (1/T)\ \Sigma_{i=1}^{T}[y_i - \theta_{ML}]^2$. Thus the ML estimators of $\theta$ and $v = \sigma^2$ are

$$\theta_{ML} = \sum_{i=1}^{T} \frac{y_i}{T},$$
$$\sigma^2_{ML} = \left(\frac{1}{T}\right) \sum_{i=1}^{T} [y_i - \theta_{ML}]^2. \tag{9.B.4}$$

The second-order condition is needed to ensure that we have obtained a maximum rather than a minimum of LL. The second-order partial derivatives are conveniently placed in a $2 \times 2$ matrix with elements $h_{11}$ and $h_{12}$ along the first row, and $h_{21}$ and $h_{22}$ along the second row. By Young's theorem from elementary calculus, $h_{12} = h_{21}$. Verify that

$$h_{11} = \frac{\partial^2 LL}{\partial \theta^2} = -\sum_{i=1}^{T} \left(\frac{1}{v}\right),$$
$$h_{22} = \frac{\partial^2 LL}{\partial v^2} = \left(\frac{T}{2}\right)\left(\frac{1}{v^2}\right) - \sum_{i=1}^{T} \frac{[y_i - \theta]^2}{2v^3},$$
$$h_{12} = h_{21} = \frac{\partial^2 LL}{\partial \theta \partial v} = \sum_{i=1}^{T} \frac{[y_i - \theta]}{v^2}.$$

Consider the mathematical expectation of these terms. $E(h_{11}) = h_{11}$, since $h_{11}$ already contains population value $\sigma^2$ and no sample estimates. since $y \sim N(\theta, \sigma^2)$, we have $E(y) = \theta$, or $E(y - \theta) = 0$. Hence $E(h_{12}) = 0 = E(h_{21})$. Note that the term $h_{22}$ does contain sample $y$ values, and from the properties of a chi-square random variable it is well known that $E\Sigma_{i=1}^{T}[y_i - \theta]^2 = (T-1)v$. Hence

$$E(h_{22}) = \left(\frac{T}{2v^2}\right) - \frac{(T-1)v}{2v^3} = \left(\frac{1}{2v^2}\right)[T - (T-1)] = \frac{1}{2v^2}. \tag{9.B.5}$$

The matrix of expectations $E(\{h_{ij}\})$ is called the Hessian matrix. In our simple case of the normal distribution it is diagonal with $\{1/v, 1/2v^2\}$ as elements, and can be written as

$$\text{Hessian} = \text{diag}\{1/\sigma^2, 1/(2\sigma^4)\}. \tag{9.B.6}$$

For this diagonal matrix the eigenvalues are the same as the diagonal terms. Since both are positive, the Hessian is positive definite. A clear understanding

of the theory above for a set of $T$ observations from a normal density is useful, because essentially the same concepts and names like Hessian and score carry over to more interesting regression problems needed in finance.

### Normal Regression Model

Consider the model (9.3.1), $y = X\beta + u$, with multivariate normal errors $u \sim N(0, \sigma^2 I)$. Now the $LL$ is written in matrix notation as

$$LL = -\left(\frac{T}{2}\right)\log 2\pi - \left(\frac{T}{2}\right)(\log \sigma^2) - \frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}. \tag{9.B.7}$$

Hence its partials yield

$$\text{Score}(\beta) = X'(y - X\beta) \quad \text{and} \quad \text{Hessian} = (-X'X). \tag{9.B.8}$$

A solution of $\text{Score}(\beta) = X'(y - X\beta) = 0$ is simply $\beta_{ML} = (X'X)^{-1}X'y$. Hence the ordinary least squares (OLS) estimator can be viewed as a solution of the ML estimation problem. Note that "multicollinearity" means the inverse of $(X'X)$ matrix cannot be found and "near multicollinearity" means that standard errors of regression coefficients are large, the inverse of $(X'X)$ matrix cannot be reliably found, and the numerical inverse given by the computer cannot be trusted. Hence statistical inference based on usual $t$ values cannot be trusted.

### Nonlinear Regression

Now we consider estimation methods more general than multiple regression. Nonlinear regression simply means that $y$ is related to the regressors by a nonlinear relationship: $y = g(x, \theta) + \varepsilon$, where $x$ contains $x_i$ for $i = 1, 2, \ldots, p$, regressors and $\theta$ represents the nonlinear regression parameters, which need not equal $p$ in number. Assume that there are $t = 1, 2, \ldots, T$, observations for $y$ and each regressor $x_i$. If the vector of errors $\varepsilon$ follows the multivariate normal density, only the mean and variance matter. Let the density function of errors be denoted by $f(\varepsilon) = f[y - g(x, \theta)]$. Clearly, the density function is a function of the data on $y$ and $x$ and of the parameter vector $\theta$:

$$LL = \sum_{t=1}^{T} \log f(\varepsilon_t) = \sum_{t=1}^{T} \log f(y_t - g(x_t, \theta)). \tag{9.B.9}$$

Even if we are not sure that the functional form of the density $f(\varepsilon)$ is indeed normal, the likelihood function of the normal density can still be viewed as the quasi-likelihood function. Maximizing the log likelihood function can be a demanding numerical problem when the likelihood function is complicated

nonlinear function of several variables. Then the solution needs to be found by the user supplying some starting values and the software searching the parameter space of θ (usually having a large number of dimensions) iteratively by some form of extensions to the Newtonian iterative scheme.

It is useful to begin with the intuition from elementary calculus to understand the intricacies of nonlinear optimization algorithms (software). The slope of a function $f(\theta)$ is also known as the tangent gradient or the partial derivative $\partial f/\partial\theta$. Its sign is positive, $\partial f/\partial\theta > 0$ for increasing functions. We begin exploring the function with some initial starting values and a choice of step size. Clearly, one is on an upward trajectory of $f(\theta)$ if the gradient is positive for each chosen direction at the starting value. A local maximum of $f(\theta)$ is reached when the gradient stops increasing, becomes zero at the maximum, and decreases if any steps are taken beyond the maximum. Since the sign of $\partial f/\partial\theta$ is changing from positive to zero to negative, it must be a decreasing function, that is, the second derivative must be negative, $\partial^2 f/\partial\theta^2 < 0$, when evaluated at the maximum. Thus the first-order condition (FOC) for a maximum is that the gradient or the first derivative is zero, which makes intuitive sense. Similarly the second-order condition that the second derivative is negative at the maximum also makes intuitive sense.

Maximum likelihood estimation for nonlinear regression problems is essentially an extended highly nonlinear multiple parameter version of the simple calculus problem of finding the maximum (see McCullough and Renfro, 1999, 2000; McCullough, 2004). In fact any nonlinear solver software must be consistent with the basic calculus principles and intuition given above.

### Tools for Verification of the Solution by a Nonlinear Solver

It is important to actively verify the solution found by the software to make sure that the solution is valid. This involves performing the following checks known in numerical analysis (Gill et al., 1981; McCullough and Vinod, 2004):

1. The gradient vector containing the first derivatives with respect to each parameter should be zero ( a suitably defined very small number, say, $10^{-16}$, at the solution.

2. Numerical analysts, Gill et al. (1981), recommend looking at the path taken by the algorithm in reaching the solution. The path should show appropriate (quadratic) rate of convergence. If not, the likelihood surface may be ill-behaved or flat near the solution. McCullough and Vinod (2003) use an example to show how to do this by considering not only the value of the LL at consecutive evaluations but successive differences. The latter should decline fast for quadratic convergence.

3. The matrix of second-order partials, known as the Hessian matrix, should be negative definite. It is useful to conduct an eigensystem analysis of the Hessian to find the condition number $K^{\#}$ defined as the ratio of the largest eigenvalue $\lambda_{max}$ and the smallest eigenvalue $\lambda_{min}$ of the Hessian. Some authors

define $K^{\#}$ in terms of a ratio of square roots of eigenvalues, also known as singular values. Let us define

$$K^{\#} = \frac{\lambda_{\max}}{\lambda_{\min}}. \tag{9.B.10}$$

If the Hessian is ill-conditioned similar to a multicollinear data matrix, $\lambda_{\min}$ is close to zero and the ratio $K^{\#}$ is very large. Recall from (9.B.8) that the Hessian for the normal regression model is $(-X'X)$, which is directly related to the multicollinearity issue. For a deeper analysis it is useful to look at the eigenvectors associated with the largest and smallest eigenvalues of $(X'X)$ denoted by $\lambda_{\max}$ and $\lambda_{\min}$ as needed. One needs to identify the name of the corresponding regressor variable from the set $\{x_i\}$ as follows: As a first step, look for the largest absolute magnitudes in these particular eigenvectors. In the second step, check the name of the particular offending regressor attached to that absolute magnitude. This regressor variable is most likely the cause of the numerical problems such as multicollinearity or unstable nonlinear optimization. In the third step, change the model specification with respect to the offending regressor. For example, mere change of scale for the offending variable can remove the numerical problems. Sometimes the offending variable needs to be omitted and another selected. In general, it is possible to have a set of variables as the offending set. In any case, it is useful to avoid models that have Hessians with large $K^{\#}$ values.

4. Profiling the likelihood is a method developed by Box and Tiao (1973), Bates and Watts (1988), among others. Let $\theta_{\mathrm{ML}}$ be the ML estimate of the parameter $\theta$ with estimated standard error $s$, which corresponds to a value $LL_{\mathrm{ML}}$. Now fix $\theta = \theta_0$ and re-estimate the model with one less free parameter, obtaining a value $LL_0$. Next allow $\theta_0$ to vary about $\theta_{\mathrm{ML}}$, and plot these values of $\theta_0$ against the associated values of $\log L_0$. This plot over a typical range $(\theta_0 - 4s, \theta_0 + 4s)$ constitutes the likelihood profile for the parameter $\theta$. Ideally, the line plotted should have a mountain-like symmetric shape with a peak near $\theta_0$.

This fourth check was recommended by McCullough and Vinod (2003), who give an economics example. Calculus methods are used to find the maximum of a function such as the likelihood function if the function locally approximately goes up and comes down, or is quadratic. More important, the quadratic approximation is crucial to compute the standard errors (SE) of coefficients and confidence intervals around parameters using Wald-type statistics (e.g., Student's $t$ statistic). Wald-type statistics are ratios of coefficient estimates to their standard errors. Computationally more demanding methods involve evaluation of the likelihood at the null and alternative hypotheses leading to likelihood ratio statistics. If the quadratic approximation is not valid, it will be revealed by profiling the likelihood. If the likelihood profile suggests that a quadratic approximation is not valid, any division by SE is not

appropriate. Alternative inference methods using bootstraps including d-boot discussed in Section 9.3 and Appendix C of this chapter are relevant here.

### Verification Tools Based on Stopping Rules, Step Sizes, and Problems Too Demanding for a PC

Some additional checks are needed in the context of maximum likelihood (ML) estimation arising from the choice of starting values and step sizes. We want a large variety of starting values and step sizes, so that we have a good chance of reaching the global maximum and not let the algorithm be stuck near some local maximum or in some flat region of the parameter space.

1. We want a good stopping algorithm when the optimum is reached. Clearly, one should stop taking further steps in the parameter space when the value of the log likelihood ($LL$) itself does not increase any further.
2. If consecutive $LL$ values are changing only in the sixteenth significant digit, we know we have reached the limits of PC computing and the problem at hand is too large or too demanding for a PC with 32 bit words (i.e., 16 digits).
3. Another stopping rule is to stop when the consecutive evaluations of the function do not make any difference to the estimated gradient or the estimated parameters. The stopping rules can have a profound impact on the nonlinear solver.
4. The search for the solution is not restricted to a small region of the parameter space so that the local solution obtained by calculus type methods does not miss the global solution. This is sometimes accomplished by using a large variety of starting values.
5. The scaling of the data can have a profound effect on the numerical accuracy of the solution. One possibility is to standardize all data by subtracting the mean and dividing by its standard deviation. This transformation creates some statistical inference problems since division by the random standard deviation changes the sampling distribution. However, from a numerical viewpoint it will reveal if the chosen scaling is giving a badly conditioned Hessian matrix.

Since financial economists may be dealing with millions of dollars, it is important to have numerically as accurate and reliable answers as possible. In traditional economics the consequences of making small errors need not be so large.

## APPENDIX C: MAXIMUM ENTROPY (ME) BOOTSTRAP FOR STATE-DEPENDENT TIME SERIES OF RETURNS

The traditional bootstrap based on independent and identically distributed (iid) randomness is discussed in Section 9.3.3. There we discuss how regres-

sion residuals are resampled to create a large number $J$ of regression problems and how confidence intervals are created by rank-ordering the $J$ estimates of regression coefficients. The aim of this appendix is to explain some recent solutions to some of the problems associated with the traditional bootstrap for applications in finance.

The iid bootstrap suffers from the following three properties, which are especially unsuitable when $x_t$ represents financial returns in excess of a benchmark (S&P 500). Vinod's (2004) new ME bootstrap is designed to avoid all three properties simultaneously.

**P1.** The traditional bootstrap sample repeats some $x_t$ values while deleting as many others. There is no reason to believe that excess returns a few basis points away from observed $x_t$ are impossible and do not belong in the resample.

**P2.** The traditional bootstrap sample lies in the closed interval $[\min(x_t), \max(x_t)]$. The vast finance literature dealing with value at risk (VaR) proves that investors do worry about potentially large losses from returns lower than $\min(x_t)$ in the data.

**P3.** The traditional bootstrap sample shuffles $x_t$ such that any dependence information in the sequence $x_{t+1}$ follows $x_t$, which in turn follows $x_{t-1}$, and so forth, is lost. This was discussed in Section 9.3.3. If we try to restore the original order to a shuffled resample, we end up with essentially the original set $\{x_t\}$, except that some dropped $x_t$ values are replaced by the repeats of adjacent values. Hence it is impossible to generate a large number ($J = 999$) of sensibly distinct resamples, if we wish to restore the time series properties to the traditional bootstrap.

The method discussed here is designed to avoid the properties P1 to P3 listed above without introducing unnecessary arbitrariness. We rely extensively on the maximum entropy density. For example, to work around the property P2, we use the method described in the appendix of Vinod (1985) to extend the range of income data. The entropy has been widely used in both social and natural sciences as a measure of "ignorance," or more accurately, noninformativeness in a system.

The observed time series is viewed as the random variable $x_t$ taking $T$ possible values and let $f(x)$ denote its density. The bootstrap will create a large ensemble consisting of $J$ time series with elements $x_{jt}$, where $j = 1, \ldots, J$, and $t = 1, \ldots, T$.

The entropy is defined by the mathematical expectation of Shannon information:

$$H = E(-\log f(x)) = \text{Entropy} \tag{9.C.1}$$

We impose mass and mean-preserving constraint and select the particular density $f(x)$, which maximizes $H$. Such an $f(x)$ is called the ME distribution and consists of a combination of $T$ pieces joined together.

First, we sort the data in an increasing order of magnitude, denote the order statistics by $x_{(t)}$, and compute new "intermediate points" defined as the average of successive order statistics:

$$z_t = 0.5(x_{(t)} + x_{(t+1)}), \qquad t = 1, \ldots, T-1. \tag{9.C.2}$$

The intermediate points help define our intervals defined by pairs of successive $z$'s. For $t = 2$ to $t = T - 1$ we have $I_2 = (z_1, z_2), \ldots, I_{T-1} = (z_{T-2}, z_{T-1})$. The key novelty is to also include open-ended intervals $I_1\ (-\infty, z_1)$ and $I_T\ (z_{T-1}, \infty)$ at the two end points for $t = 1$ and $t = T$. Thus we have exactly $T$ intervals, each of which contains exactly one $x_{(t)}$.

The mass-preserving constraint (on moments of order zero) says that a fraction $1/T$ of the mass of the probability distribution must lie in each interval. This is achieved by the ME bootstrap by giving an equal chance to each interval $I_1\ (-\infty, z_1), I_2(z_1, z_2), \ldots, I_T\ (z_{T-1}, \infty)$ of being included in the resample. The traditional bootstrap selects uniform pseudo random numbers between [0, 1], rounds them into a sequence of random integers in [1, T] to define the shuffled selection from the set $\{x_t\}$. It gives each observation an equal chance $(1/T)$ of being included in the resample. An advantage of the ME bootstrap is that it uses the true pseudorandom numbers, and not rounded integers.

The mean-preserving constraint (on order 1 moments of $f(x)$) is $\Sigma x_t = \Sigma\, x_{(t)} = \Sigma m_t$, where $m_t$ denote the mean of $f(x)$ within intervals $I_t$. This property is satisfied by the traditional bootstraps. Recall from (9.3.3) that the traditional bootstrap selects the observations $x_t$ with replacement, such that the probability of selection remains $(1/T)$. Formally, the $f(x)$ of the traditional iid bootstrap at each $x_t$ is a so-called delta function with the entire mass $(1/T)$ at $x_t$ in $I_t$.

The mean-preserving requirement is too complicated for the ME bootstrap. It requires that the mean $m_t$ in the interval $I_t$ be equal to a weighted sum of the order statistic $x_{(t)}$ with weights from the set {0.75, 0.50, 0.25} as explained below.

Only if the $f(x)$ maximizes our "ignorance" defined by the entropy (9.C.1), we can call it the ME distribution. The problem of finding such $f(x)$ was solved long ago in the statistical literature dealing with so-called characterization problems (Kagan et al., 1973). A reference for economists in a different context is Theil and Laitinen (1980). The solution states that the form of the density is exponential in the two open-ended intervals and uniform in all intermediate intervals with a fixed range with parameters given as follows:

**1.** $f(x) = (1/\theta)\exp([x - z_1]/\theta)$, with $\theta = m_1 == 0.75x_{(1)} + 0.25x_{(2)}$ in $I_1$.
**2.** $f(x) = (1/\theta)\exp([z_{T-1} - x]/\theta)$ with $\theta = m_T = 0.25x_{(T-1)} + 0.75x_{(T)}$ in $I_T$.

**3.** $f(x) = 1/(z_k - z_{k-1})$ is a uniform density where the mean of the uniform $(z_{k-1} + z_k)/2$ equals $m_k = 0.25x_{(k-1)} + 0.50x_{(k)} + 0.25x_{(k+1)}$ in the intervals $I_k$ $(z_{k-1}, z_k)$ for $k = 2, \ldots, T - 1$.

The weights {0.75, 0.50, 0.25} help satisfy the mean-preserving constraint $\Sigma x_t = \Sigma m_t$. Thus the ME distribution avoids arbitrary choices and is completely known from the data.

For further clarity let us consider a simple example of $T = 5$ data points where $x_{(t)} = \{4, 8, 12, 20, 36\}$. From consecutive averages we have $z_{(t)} = \{6, 10, 16, 28\}$, which determine the five intervals upon inclusion of the open-ended infinite intervals at the two extremes. At the left extreme, the open-ended interval is $I_1$ $(-\infty, 6)$ and the ME distribution in open-ended intervals has been established by statisticians to be exponential given by $(1/\theta)\exp([x - 6]/\theta)]$. We have the option to choose $\theta$, the mean of the exponential density. We choose $\theta$ to satisfy the mean-preserving constraint as $\theta = m_1 = 0.75 * 4 + 0.25 * 8 = 5$. The second interval $I_2$ is $(6, 10)$ is not open ended but has fixed limits at 6 and 10. Theoretical statisticians have established that the ME distribution in it is uniform. Again, we have the option to choose $\theta$, the mean of the uniform density. We choose $\theta$ to satisfy the mean-preserving constraint as $m_2 = 0.25 * 4 + 0.5 * 8 + 0.25 * 12 = 8$. Similarly for the intervals $(10, 16)$ and $(16, 28)$ the means of the uniform are chosen to be $m_3 = 13$ and $m_4 = 22$, respectively. The last open-ended interval is $I_T$ $(28, \infty)$, wherein the ME density is exponential: $(1/\theta)\exp([28 - x]/\theta)$. With our choice, $\theta = m_T = m_5 = 0.25 * 20 + 0.75 * 36 = 32$.

Now the mean-preserving constraint $\Sigma x_t = 80 = \Sigma m_k = 5 + 8 + 13 + 22 + 32$, based on the weights 0.75, 0.50, and 0.25, is indeed satisfied. Thus the ME bootstrap distribution consists of $T$ attached parts over the entire real line $(-\infty, \infty)$. If we select a bootstrap resample from the ME density, we clearly avoid the property P2 of the traditional bootstrap.

It is fortunate that the ME density involves the exponential, since its CDF is $[\exp(-\theta x) - 1]$ and inverse CDF at any probability $p_r$ is $[\theta \ln(1 - p_r)]$. Thus pseudorandom numbers from the ME density are easy to obtain on a computer. Note that the inverse CDF here is obviously a positive number when $\theta > 0$, since the logarithm of a fraction is always negative. However, the mean $\theta$ in the tails need not be always positive. Hence we use the absolute value $|\theta|$ in the inverse CDF. The pseudorandom numbers from the ME density are easy to obtain on a computer, despite the presence of the exponential. Since the extrapolation in the left and right tails needs some care, we explain it in a separate section below.

### Tail Extrapolations

First consider the right tail of the ME density similar to the one plotted in Figure 9.C.1, which begins at $z_{T-1}$ and goes to infinity. The tail extrapolation involves mapping the interval $[0, \infty)$ of the usual (standard) exponential with mean $\theta = m_T = 0.25x_{(T-1)} + 0.75x_{(T)}$ on the interval $[z_{T-1}, \infty)$ needed here and

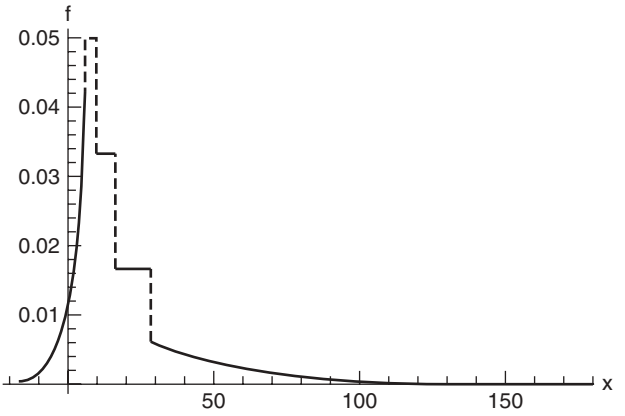



**Figure 9.C.1** Plot of the maximum entropy density when $x_{(t)} = \{4, 8, 12, 20, 36\}$

assigned the probability mass $(1/T)$ not 1. That is, we shift the starting point of the exponential to $z_{T-1}$. Consider $T$ uniform pseudorandom draws denoted by $p_s$ lying in the [0, 1] interval. They yield $j$th ensemble elements $x_{jt}$ for $t = 1, \ldots, T$, as quantiles $x_{jt}$ of the ME density. If the random draw exceeds $(T - 1)/T$, the quantile $x_{jt}$ will belong to the right tail and is given by

$$z_{T-1} - |\theta| \ln(1 - p_r). \tag{9.C.3}$$

The left tail extrapolation is similar to that for the right tail, except that the exponential is in the reverse direction so that $(1 - p_r)$ in (9.C.3) will become $(p_r)$. The left tail begins at $-\infty$ and goes to $z_1$. Let us first compute the quantile as if it were in the right direction and then adjust for the direction. The quantile of the ME density will belong to the left tail only if the random draw $(p_r)$ is less than $1/T$. Since the mean of the exponential in the left tail is $\theta = m_1 = 0.75x_{(1)} + 0.25x_{(2)}$, which may well be negative, we use $|\theta|$. The desired quantile similar to (9.C.3) is

$$z_1 - |\theta| |\ln(p_r)|. \tag{9.C.4}$$

**List of Steps in the Seven-Step Algorithm**

The entire set of steps needed to create a random realization of $x_t$ is as follows:

1. Define a $T \times 2$ sorting matrix called $S_1$, and place the observed time series $x_t$ in the first column and the set of integers $a = \{1, 2, \ldots, T\}$ in the second column of $S_1$.

2. Sort the matrix $S_1$ with respect to the numbers in its first column. This sort yields the order statistics $x_{(t)}$ in the first column and a vector $a^s$ of sorted

$a$ in the second column will be needed later. From $x_{(t)}$ construct the intervals $I_t$ defined from $z_t$, from (9.C.2), and $m_t$ with certain weights on the order statistics $x_{(t)}$ defined above.

3. Choose a seed, create $T$ uniform pseudorandom numbers $p_s$ in the [0, 1] interval, and identify the range $R_t = (t/T, (t + 1)/T]$ for $t = 0$ to $T - 1$, wherein each $p_s$ falls.

4. Match each $R_t$ with $I_t$ . If the pseudorandom number lies in the left tail, $p_s \in R_0$ use equation (9.C.4) and if it falls in the right tail, $p_s \in R_{T-1}$, use (9.C.3). Otherwise, use linear interpolation and obtain a set of $T$ values $\{x_{jt}\}$ as the $j$th resample. Make sure that each mean of the uniform equals the correct mean $m_t$. If $|(1/T)\Sigma_t\, x_{jt} - \bar{x}| > T_{ol}$, where $T_{ol}$ is a tolerance number (defined by the user), the entire set $\{x_{jt}\}$ is an outlier. Some users may reject outlier sets based on $Q_1 - 1.5 * \text{IQR}$ or $Q_3 + 1.5 * \text{IQR}$, where $Q_1$ is the first quartile, $Q_3$ is third quartile, and $\text{IQR} = Q_3 - Q_1$. If $\{x_{jt}\}$ is judged to be an outlier, go back to step 3 and select a new $\{x_{jt}\}$.

5. Define another $T \times 2$ sorting matrix $S_2$. Reorder the $T$ members of the set $\{x_{jt}\}$ for the $j$th resample obtained in step 4 in an increasing order of magnitude, and place them in column 1. Also place the sorted set $a^s$ of step 2 in column 2 of $S_2$.

6. Sort the $S_2$ matrix with respect to the second column to restore the order $\{1, 2, \ldots, T\}$ there. Denote the jointly sorted column 1 of the elements by $\{x_{sjt}\}$. These represent the ME resample where the additional subscript $s$ reminds us of the sorting step, which has restored the time dimension to correspond with the original data.

7. Repeat steps 1 to 6 a large number of times for $j = 1, 2, \ldots, J$ (e.g., $J = 999$).

We conclude that we have removed all three undesirable properties of the traditional bootstrap. The bootstrap literature is vast, and there are piecemeal attempts to remove one of P1 to P3. For example, the so-called moving blocks bootstrap deals with P3 by assuming $m$-dependent time series. Since no other bootstrap removes P1, P2, and P3 simultaneously, any comparison with them will not be fair to the ME bootstrap.

Vinod (2003b, 2004) illustrated this method with examples from finance. He considers the top ranked mutual fund (No. 734 from 1281 funds) studied over the boom period of January 1991 to December 2000. The seven steps above yield $J = 999$ resamples and hence $J$ values of various statistics discussed in the text. The 25th- and 975th-order statistics for any statistic yield the so-called naïve 95% confidence interval. He concludes that the fund 734 will dominate the S&P index fund with approximately 95% chance. The maximum entropy bootstrap methods described in this appendix are applicable elsewhere for time series inference.

CHAPTER 10

# What Does It All Mean?

The bottom line is that the stock market is risky, but it is not a gamble. The obvious difference is that gambling, even with fair odds, is a zero-sum game. Someone wins and someone loses. The stock market, on the other hand, is the exchange of investment vehicles. The buyer and seller can both accomplish their goals at once. The less obvious difference is that stock market risk is not symmetrical. When flipping a coin, the odds of heads or tails are equal. If you call head, your likelihood of winning and losing are the same. As we have seen in the stock market, however, upside and downside risk are two very different concepts. Investment analysis that ignores this difference can be too conservative if downside risk is low, or understate the risk if downside risk is high.

In the course of this book, we have started with the basics of financial asset valuation and risk measurement to determine where and how downside risk will play a role. Financial valuation relies on being able to discount future earnings with the appropriate risk premium. If this risk premium is affected by downside risk differently than upside risk, its importance is secured. Furthermore, if downside risk is not included in value at risk calculations, the true worst-case scenario is not being shown. To that end we covered four main areas that must be recognized before we incorporate downside risk into portfolio selection.

1. Distribution of returns. No matter how afraid of downside risk an investor is, if returns are symmetrically distributed around the mean, it doesn't matter as far as measurement is concerned. For a symmetric distribution, upside and downside risks are equal. Stock returns are known, however, not to be symmetric. While the normal distribution is theoretically elegant, and it is commonplace enough that most software packages and even hand calcula-

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

tors will compute it, it is not appropriate for most investment portfolios. The mere fact that returns are truncated at –100% indicates that symmetry is not likely for stocks.

One adjustment may be to choose another distribution that matches the data better. We have suggested a number of distributions that allow more flexibility than the normal distribution, including skewness and excess kurtosis. Once a candidate distribution is chosen, a goodness-of-fit test can be done to double-check that it is an appropriate choice. From the new distribution, simulations, value at risk, and other measures can be calculated.

If the correct distribution is not clear, or you don't want to be pinned down to the rigor of a particular distribution, there are still alternatives. Historical data can be used as an empirical cdf. This does not require complicated estimations, but it will be sensitive to extreme values in the data that can be smoothed out with a distribution. A nonparametric density function can be used, such as a kernel density. Finally, and the choice that we would recommend, the data can be classified into a Pearson family of distributions. Pearson's estimation gives the distribution as many moment parameters as the data indicate. This estimation is flexible, while yielding a density function and equation that allow for calculations and simulations that cannot be done with an empirical CDF. Even if the data do fit a common distribution, the worst that will happen with Pearson's estimation is that it will return that distribution. Performing the one extra step of verifying the distribution through the Pearson estimation is cheap insurance when there are potentially millions of dollars of hidden downside risk.

2. Diversification of downside risk. CAPM has gotten kicked around a lot as a theory of asset pricing. It makes many grandiose assumptions and does not fit the data perfectly. However, as a theory CAPM gives one of the fundamental insights of finance: you only get compensated for systematic risk. You don't get rewarded for taking stupid risks, although investors must be compensated in order to get them to buy a security with risk they cannot avoid.

Extreme downside movements are hard to avoid. Downside movements are often unexpected, coming from natural disasters, R&D foul-ups, and so on. Let's face it, companies don't plan to have losses, and don't publicize that things may go badly. We illustrated in Chapter 7 that only half as much downside risk can be avoided compared to general risk.

Further complicating matters, downside risk is contagious. Accounting scandals, trading scandals, international currency crises, and corruption are not isolated to a single company. How do you diversify a risk that is endemic to the foundation of the entire economy? This means that even after a portfolio is compiled there are some extra questions to ask:

Are my investments too close geographically?

Do my investments rely on the same technology?

Are there political, legal structures that could adversely affect my investment?

Is there underlying fraud or corruption in the investment process?

If the answer to any of these questions is yes, downside risk is going to be larger than the financial reports of the company indicate.

3. Investors' attitude toward downside risk. Stock market movements are shown on the television, but they are not entertainment. Investors have to take the consequences of what happens to their portfolio, good or bad. In Chapter 6 we explained the theory behind risk aversion, and methods to model human behavior. More important is how you feel about risk, personally. The stock market assigns a risk premium, but that does not constrain your preferences; it is only the average risk premium that clears the market. If you are considering an investment and the value at risk is too high for your tastes, walking away may be the better option than laying awake at night reminding yourself that market prices should be efficient.

4. Accurate measure of downside risk. Computer output such as financial statements and portfolio simulations tend to look believable to humans. Maybe it has to do with observing routine calculations that computers can do many times faster, or maybe it comes from the movies where robots never make mistakes, and computer operating systems never crash. In the real world, computers need as much help as anyone to come up with reliable results. There are some common sources of errors that should be examined before any output is believed, much less before your money is put at stake.

The first source of error is always the data. The numbers put into the calculations must be current and appropriate to your holding period. Even today's numbers may not be appropriate for calculations if a company is about to go through some type of restructuring. The data must be accurate. Corruption, data revision, or just plain mistakes can cause inaccuracy in calculations that make them unreliable. Some data sources will be more prone to inaccuracy than others. Typically the more times data are moved, the more errors are possible. Data straight from the NYSE is more accurate than that recorded from slips of paper. Also consider the incentives one has to misrepresent information, and the penalties for falsification.

David Bianco of UBS analyzed S&P 500 companies for the quality of earnings. He found that the index as a whole returned only 3.3% annually during the period 1988 to 2003, but the 50 companies with the best quality earnings generated a return of 12.3% including dividends, whereas shares of bottom fifty companies in terms of quality earnings returned zero percent. Bianco's measure of quality takes account of

A. One-time or special charges: In Bianco's view, companies who take such charges tend to do so year after year. Also, these charges may actually include recurring operational costs that should be deducted from revenue.

B. Choosing not to subtract the costs of stock options: Employee stock ownership plans (ESOPs) especially stock options is a way to compensate the employees when the price of stock goes higher than a set limit, but they dilute earnings per share. Stock options as a percent of outstanding shares have not declined between year 2000 and 2002, even after the abuses at Enron and other companies created uproar and some accounting reforms were in the news. For examples the percentages between 2000 and 2002 are: (Cisco, 14, 19), (Hewlett Packard, 8, 15), (IBM, 9, 13), (Intel, 9, 13).

C. Rosy pension assumptions: inflating pension earnings or masking pension costs that improve corporate earnings.

D. Off the book loans to key employees (CEOs and CFOs): These loans can be forgiven at the discretion of the board of directors.

E. Contingent convertible bonds (CoCo's): CoCo's are a financial instrument somewhat more complex than ESOPs. Under current accounting rules their impact on cash flow need not be shown on the company books until the price reaches the contingency price of the stock. However, the impact can be substantial. For example, at Interpublic Group Companies, the additional CoCo shares would cut earnings by 6.7% (N.Y. Times, Feb 15, 2004)

These items are not reflected in pro-forma earnings. When the stock prices rise, the employees with stock options exercise them placing a drain on the cash flow of companies. If stock options as a percent of shares is increasing, the pro-forma earnings are that much more unreliable. Similarly, if CoCo bonds are converted into shares, the drain on cash flow is not reflected in pro-forma reports. Ideally the Financial Accounting Standards Board (FASB) and/or the London-based International Accounting Standards Board's (IASB) should ask public companies to report these quality adjusted earnings in addition to pro-forma earnings. The clever accountants come up with ever newer tools to make it appear that the earnings are higher than they really are whereas FASB is not keeping up with these schemes. In September 2003 FASB agreed to look into stock options.

The perks granted to CEO's and other key employees of large US corporations are sometimes revealed in company filings, and sometimes become public because of divorce or other litigation involving these individuals. The divorce of Citigroup's Mr. Weill shocked many by the excessive perks mostly hidden from the shareholders. Dennis Kozlowski, former chief executive of Tyco International bought $29 million worth land and property in Florida and $16.8 million apartment on Fifth Avenue in Manhattan. He also spent $14 million for renovating and furnishing the apartment. Jet airplanes for personal use have been granted to Michael Eisner, CEO of Disney and Richard M. Scrushy, former CEO of HealthSouth. Martha Stewart, the founder of Martha

Stewart Living Omnimedia billed her company $17,000 per year for a weekend driver and other costs (New York Times, Feb 22, 2004).

Another source of inaccuracy is in the computer calculation. The same question, asked two different ways to a computer can get two different answers. Chapter 9 gave a glimpse into the logic used by computers. We provided some tricks to getting more precise output, and verifying the results. If numbers look too good to be true, get a second opinion. Use a different software, rescale the data, or do a back-of-the-envelope double check. At the end of the day, the computer is not going to refund your money for its calculation errors.

Once these four areas are confirmed, it is vital to incorporate downside risk into investment decisions. Since we devoted a chapter to mathematical tools, it may be useful to add an example discussed by J.A. Paulos in a book called *A Mathematician Plays the Stock Market*. Consider a simple weekly strategy to buy stock on Monday and sell it on Friday under the assumption that half the time stock (e.g., initial public offering, IPO) gains 80% and half the time it loses 60%. That is, one dollar can become $1.80 or $0.40. Consider a two-week time horizon with four possible outcomes: (1) gains of 80% both weeks $(1.8)^2 = 3.24$, (2) losses of 60% both weeks $(0.4)^2 = 0.16$, or (3) a gain in one week followed by a loss in the next week $(1.8)(0.4) = 0.72$, or (4) a loss in the first week followed by a gain the next week $(0.4)(1.8) = 0.72$. If $10,000 is invested, the four outcomes are $32,400, $1,600, $7,200, or $7,200 respectively. Thus the average amount after two weeks is a huge return of $12,100. It is tempting to look back at the end of just two weeks and assume that a similar *average* performance will repeat. That is, it is tempting to assume that this average rate of return experienced in first two weeks will hold for the whole year (over 26 two-week periods). This is called *law of small numbers*, that is, the human temptation to generalize from a small sample. The investor will expect to end up with $(1.21)^{26} = 142.0429$, that is, $10,000 will become over $1.42 million in just one year. Unfortunately, working with the average return is invalid. It is more appropriate to assume that the investment will rise for 26 weeks and fall for 26 weeks in any order. Then the investor ends up with $(1.8)^{26}(0.4)^{26} = 0.0001952742$ or only $1.95.

In this book we have shown that there is much more to investing than just compound interest mathematics. It is important to use the past data, statistics, market psychology, forecasts of government monetary and fiscal policy, forecasts of international financial trends, and many related tools in combination to come up with a strategy suitable for one's own risk tolerance. We have noted that implied volatility in the put options is a forward-looking measure for forecasting downside risk. We have also discussed the use of (put and call) option markets to buy insurance (hedge) against downside risk at a price.

Investors must also be careful not to ignore markets other than the stock market. Investors generally do have nonfinancial assets, such as real estate. Falling financial wealth due to the recent fall in stock prices and the 2001 recession was coupled with rise in personal debt. Recent history shows that there has been a property market decline with all recent recessions since the 1960s.

However, the 2001 recession actually caused a shift toward property markets, and this along with the record low interests could have contributed to a bubble in property prices instead of a recession. Karakitos (2002) has developed a model that links the stock market with the property markets in the United States. It is well known that the property market crash in Tokyo started the long-term decline in the Japanese stock market. Realistically, corporate bond markets, real estate markets, government securities markets, and currency markets dwarf the stock market. Those markets are just not as sexy to write about since most of their valuation techniques are relatively straightforward.

Real Estate Investment Trust (REIT) is a useful vehicle for investing in real estate. REIT pools the capital of many investors to purchase and manage properties. There are REIT's devoted to income streams or mortgage loans and are traded on major exchanges just like stocks, although they are often granted special tax treatments. Unlike owning actual real estate, these are liquid investments since the owner can readily buy or sell them. An individual owner is rarely able to own large malls or commercial properties, but the REIT can do so. REITs need not be correlated with the broader market. For example, Morgan Stanley real estate investment trust REIT index often goes up when S&P500 went down. In 2003, it continued to go up even when S&P went up. This means REITs offer a possible vehicle for diversification. The rating of REITs is a specialized task involving assessment of the company with respect to (i) Market position, (ii) Asset quality, (iii) Diversification and stability of operations, and (iv) Management Operations including cash flow, management structure and so forth.

Whatever market is chosen to invest in, if the risk is identified, proper steps can be taken, such as increased capital reserves, shifting more assets toward safer investments, and hedging. Given two portfolios with the same average return and same standard deviation, we cannot consider them identical until we examine their downside risk. We can expect the unexpected insofar as while we don't know exactly what will happen, we know that extreme downside movements do happen. Putting downside risk into our calculations beforehand means that when losses do occur, we are prepared.

# Glossary of Greek Symbols

| Spoken | Symbol | Short Description | Presentation |
|---|---|---|---|
| Alpha | $\alpha$ | Average return | Section 1.5 |
| Alpha | $\alpha$ | Intercept | Appendix (chapter 1) |
| Alpha | $\alpha$ | $\alpha \in [0, 1]$ measures extent of departure from EUT precepts; $\alpha = 1$ is perfect compliance | Chapter 6 |
| Alpha dash | $\alpha'$ | Fraction in [0, 1] closed interval | Section 2.1 |
| Beta | $\beta$ | Beta of the CAPM | Section 2.2 |
| Beta sub 1 | $\beta_1$ | Pearson skewness | Equation (2.1.4) |
| Beta sub 2 | $\beta_2$ | Pearson kurtosis | Equation (2.1.4) |
| Beta sub w | $\beta_w$ | Weighted beta or downside beta | Equation (5.2.10) |
| Delta squared | $\delta^2$ | Variance of the jump | Section 1.4 |
| Delta | $\delta$ | Excess return for anomaly | Equation (1.5.2) |
| Delta | $\Delta$ | Difference operator $\Delta S_t = S_t - S_{t-1}$ | Section 1.4 |
| Delta | $\Delta$ | Change in price of an option | Section 3.4.1 |
| Delta t | $\Delta t$ | Time duration | Section 1.4 |
| Delta Pt for tau | $\Delta P_t(\tau)$ | Change in value of portfolio $P(t + \tau) - P(t)$ for horizon $\tau$ | Section 5.1 |
| Epsilon | $\varepsilon$ | Error term | Chapter 1 |

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

| Spoken | Symbol | Short Description | Presentation |
|---|---|---|---|
| Epsilon for set theory (or In) | $\in$ | In the set | Section 2.4 |
| Eta | $\eta$ | Speed at which asset returns to the mean | Section 1.4 |
| Gamma | $\Gamma$ | Absolute change in $\Delta$ of anoption | Section 3.4.2 |
| Kappa | $\kappa$ | Roots of quadratic indicates Pearson types | Equation (2.1.7) |
| Lambda | $\lambda$ | Average number of jumps per unit time | Equation (1.4.4) |
| Mu | $\mu$ | Average return | Section 1.3, 7.4 |
| Mu sub j ($j = 1, \ldots, 4$) | $\mu_j$ | Central moments of order $j$ (e.g., variance is $j = 2$) | Section 2.1 |
| Omega | $\Omega$ | Ensemble of time series | Section 9.1 |
| Omega | $\Omega$ | $\Delta$ of option in percentages | Section 3.4.3 |
| Pi | $\pi$ | Risk premium | Section 6.1 |
| PO(lambda) | PO($\lambda$) | Poisson density with parameter $\lambda$ | Equation (5.3.3) |
| PHI | $\Phi$ | Cumulative distribution function for normal | Section 2.1 |
| Pi sub T sup e | $\pi_T^e$ | Expected inflation rate over time 0 to $T$ | Section 1.2 |
| Pi sup s ($s = 1, \ldots, S$) | $\pi^s$ | Probability in state $s$ | Section 1.1 |
| Rho | $\rho$ | Absolute change in option value for 1% change in the interest rate | Section 3.4.3 |
| Rho | $\rho$ | Slope parameter in AR(1) model | Equations (4.1.1), (4.1.3) |
| Rho sub p | $\rho_p$ | Autocorrelation of lag order $p$ | Equation (4.1.2) |
| S | $s$ | Sample standard deviation of excesss returns; sample counterpart of $\sigma$ | Section 5.1 |
| S squared | $s^2$ | Sample variance of portfolios; sample counterpart of $\sigma^2$ | Equation (2.1.10) |
| Sigma | $\sigma$ | Standard deviation | Section 1.2 |
| Sigma squared | $\sigma^2$ | Variance | Equation (1.3.3) |
| Sigma sub i | $\sigma_i$ | Standard deviation, Portfolio $i$ | Equation (5.2.7) |

| Spoken | Symbol | Short Description | Presentation |
|---|---|---|---|
| Tau | $\tau$ | Tax rate | Section 1.2 |
| Tau | $\tau$ | Time horizon | Section 2.1 |
| Tau | $\tau$ | Tail index for stable Pareto density | Section 7.4 |
| Theta | $\theta$ | Risk premium | Section 1.2 |
| Theta | $\theta$ | Threshold for losses, "$\theta$ exceedances" of a generalized stable Pareto density | Section 7.4 |
| Theta | $\Theta$ | Absolute change in the option value for a reduction in time to expiration | Section 3.4.3 |
| Vega | V | Option price change with a 1% change in volatility | Section 3.4.3 |

# Glossary of Notations

| Spoken | Notation | Short Description | Presentation |
|---|---|---|---|
| A | $a$ | Intercept | Equation (1.A.5) |
| B | $b$ | Slope | Equation (1.A.5) |
| Cap-em | CAPM | Capital asset pricing model | Section 2.2 |
| C sub BS | $C_{BS}$ | Black-Scholes option pricing formula, call option | Equation (5.3.2) |
| C sub JD | $C_{JD}$ | Jump diffusion option pricing formula, call option | Equation (5.3.3) |
| Dq | $dq$ | Poisson process PO($\lambda$) | Section 1.4 |
| Dz | $dz$ | Wiener process | Section 1.4 |
| D sub j ($j = 1, \ldots, 4$) | $D_j$ ($j = 1, \ldots, 4$) | Class $D_j$ of utility functions | Section 6.3 |
| DSh sub i | $\mathrm{DSh}_i$ | Downside Sharpe ratio | Equation (9.3.8) |
| D sup s | $D^s$ | Dividends in state $s$ | Section 1.1 |
| DSh sub i | $\mathrm{DSh}_i$ | Downside Sharpe ratio for portfolio $i$ | Equation (5.2.9) |
| DTr sub i | $\mathrm{DTr}_i$ | Downside Treynor's performance measure, portfolio $i$ | Equation (5.2.10) |
| E sub 1 to e sub T | $e_1$ to $e_T$ | Regression residuals | Section 9.3 |
| E* sub jt | $e_{jt}^*$ ($j = 1, 2, \ldots, J$ and $t = 1, 2, \ldots, T$) | Shuffled residuals for $j$th replications and $t$th observation | Section 9.3 |

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

| Spoken | Notation | Short Description | Presentation |
|---|---|---|---|
| F of R | $f(R)$ | pdf of returns or vexcess returns | Section 2.1 |
| F of x | $f(x)$ | pdf of excess returns | Section 2.1 |
| cumulative distribution of x | $F(x)$ | CDF for $f(x)$ | Section 5.1 |
| F of x for parameters θ | $f(x, \theta)$ | pdf of excess returns $f(x)$ having a parameter vector θ | Section 5.1 |
| F inverse at α′ for parameters θ | $F^{-1}(\alpha', \theta)$ | Inverse CDF evaluated at α′ when the pdf depends on θ vector of parameters | Section 5.1 |
| G sub t | $g_t = \log(1 + R_t)$ | Natural log of gross return $g_t$ | Section 9.1 |
| Greater than or equal to sub $j$ for $j = 1$ to 4 | $\geq_j$ for $j = 1, \ldots, 4$ | If $A \geq_j B$ means that $A$ dominates $B$ in $j$th order | Section 6.3 |
| G sub of | $G_{\text{of}}$ | Goodness-of-fit statistic | Equation (4.3.3) |
| H sub n | $H_n = \Pi_{i=1}^{n}(1 + R_i)$ | Product of $n$ gross returns | Section 9.1 |
| H sub t | $h_t$ | Changing variance in ARCH models | Section 4.2 |
| I(j), $j = 0, 1, 2$ | $I(j)$ | Integrated of order $j$ | Section 1.4 |
| Infimum | inf | Smallest value in a set | Equation (6.2.1) |
| I sub lowercase f | $I_f$ | Matrix having all ones on and below the diagonal | Section 6.3.10 |
| I sub uppercase f | $I_F$ | Matrix having nonzero terms on and below the diagonal to approximate integration | 6.3.10, below Equation (6.3.5) |
| K dash | $K'$ | Radical of a quadratic | Section 2.1 |
| K sup # | $K^{\#}$ | Condition number | Section 9.A2 |
| M sub j for $j = 1, \ldots, 4$ | $m_j$ | Sample moments | Section 2.1 |
| M sub jw for $j = 1, \ldots, 4$ | $m_{jw}$ | Sample weighted moments | Equation (5.2.4) |
| P sub i | $p_i$ | Proportions in a portfolio | Section 2.1 |
| P sub j | $p_j$ | Principal component | Equation (2.4.3) |
| P sub t ($t = 1, \ldots, T$) | $P_t$ | $P_0$ = initial price, $P_t$ price at future date | Section 1.1 |
| P sub v | $P_v$ | Present value | Section 1.1 |

| Spoken | Notation | Short Description | Presentation |
|---|---|---|---|
| P sup A | $p^A$ | $k \times 1$ vector of individual probabilities for prospect $A$ | Section 6.3.10 |
| PE ratio | $(P/E)$ | Price-to-earnings ratio ($P/E$ ratio) | Section 7.3 |
| Q sub LB | $Q_{\text{L-B}}$ | Ljung and Box test | Section 4.2.1 |
| R bar | $\bar{r}$ | Average return | Section 4.1 |
| R sub t ($t = 1, \ldots, T$) | $r_t$ | Return after $t$ years | Section 1.1 |
| R sub t ($t = 1, \ldots, T$) | $r_t$ | (Percent) daily returns | Section 2.1 |
| R sub f | $r_f$ | Risk-free return | Section 2.2 |
| R sub i paren j | $r_i(j)$ | Returns for firm $i$ ordered from the smallest to the largest over $j = 1, \ldots, T$ | Section 6.3.6 |
| R sub m | $r_m$ | Market return | Section 2.2 |
| R sub T sup f | $r_T^{\text{f}}$ | Risk-free rate | Section 1.2 |
| R sub T sup n | $r_T^{\text{n}}$ | Nominal interest rate | Section 1.2 |
| R sub T supat | $r_T^{\text{at}}$ | Return after tax | Section 1.2 |
| S bar | $\bar{S}$ | Average price of financial asset | Section 1.4 |
| S sub n | $S_n$ | $S_n = \Sigma_{i=1}^{n} x_i$ | Section 9.1 |
| Sh sub i | $\text{Sh}_i$ | Sharpe ratio for $i$ portfolio | Equation (5.2.7), Section 9.3 |
| S sub k | $S_k$ | Sample skewness with weights $w_t$ | Equation (5.2.5) |
| Sum 4SD over all j data points | $\Sigma 4\text{SDj}$ | Sum 4th order stochastic dominance terms over all $j$ data points | Section 6.3 |
| Tr sub i | $\text{Tr}_i$ | Treynor's performance measure, portfolio $i$ | Equation (5.2.8) |
| T sub a or b | $T_a$ or $T_b$ | Number of returns from prospect $A$ or $B$ | Section 6.3.10 |
| T uppercase | $T$ | Future date | Section 1.1 |
| U($x_i$, $i = 1, 2, \ldots, n$) | $U(x_i, i = 1, 2, \ldots, n)$ | Utility of ($x_i$) when $i = 1, 2, \ldots, n$ | Section 6.1 |
| U prime | $U'$ | Marginal utility | Section 6.1 |
| U two dashes | $U''$ | Second derivative of $U$ | Section 6.1 |
| U three dashes | $U'''$ | Third derivative of $U$ | Section 6.1 |
| U four dashes | $U''''$ | Fourth derivative of $U$ | Section 6.1 |
| V sub j | $v_j$ | Eigenvectors | Section 2.4 |
| V sub 1 (or 2 or 3) | $V_1$ or $V_2$ or $V_3$ | Alternative (numerically distinct) formulas for variance | Section 9.2 |

| Spoken | Notation | Short Description | Presentation |
|---|---|---|---|
| VaR at alpha dash | $\mathrm{VaR}(\alpha')$ | Value at risk evaluated at a small proportion $\alpha'$ (e.g., 0.01 or 0.05) | Sections 2.1, 5.1 |
| W zero | $W_0$ | Initial wealth | Section 7.4 |
| W(p, alpha) | $W(p, \alpha)$ | Decision weight function of proportion $p$ and $\alpha$, the EUT compliance parameter | Section 6.2.2 |
| X | $x$ | Regressor | Appendix (Chapter 1) |
| Xbar | $\bar{x}$ | Sample mean of excess returns | Section 5.1 |
| X sub it ($i = 1, 2$), ($t = 1, \ldots, T$) | $x_{it}$ | Excess return on asset $i$ at time $t$ | Section 2.2 |
| X sub t colon T | $x_{(t:T)}$ | Ordered values of $x_t$ when $t = 1, \ldots, T$ | Section 4.3 |
| Y | $y$ | Dependent variable | Appendix (Chapter 1) |
| Y hat | $\bar{y}$ | Fitted $y$ | Appendix (Chapter 1) |
| Z | $z$ | Standard normal N(0, 1) variable | Section 1.4 |
| Z bar sub v | $\bar{Z}_v$ | Vector of average returns from Table 2.1.4 | Section 2.1 |
| Z sub alpha dash | $Z_{\alpha'}$ | Quantile of N(0, 1) | Section 2.1 |

# Glossary of Abbreviations

| Symbol | Short Description | Presentation |
|---|---|---|
| 1SD | First-order stochastic dominance | Section 6.3.6 |
| 2SD | Second-order stochastic dominance | Section 6.3.7 |
| 3SD | Third-order stochastic dominance | Section 6.3.8 |
| 4DP | Four desirable properties of utility functions to be consistent with non-EUT insights (reflectivity, asymmetric reaction, loss aversion, and diminishing sensitivity). | Section 6.2 |
| 4SD | Fourth-order stochastic dominance | Section 6.3.9 |
| AAAYX | AAAYX stock market ticker symbol mutual fund | Section 4.3 |
| ACR | Average compound return over $n$ periods | Section 9.1 |
| ACF | Autocorrelation function | Section 9.3 |
| APT | Arbitrage pricing theory | Section 2.4.2 |
| Add factor | Number added to Sharpe ratio for correct ranking | Section 5.2 |
| AdjSN | Adjusted Azzalini skew-normal density | Section 4.4 |
| AR($p$) | Autoregression of order $p$ | Section 1.4.1, Equation (4.1.1) |
| ARCH | Autoregressive conditional heteroscedasticity | Section 4.2 |
| ARMA($p, q$) | AR and MA mixed model of orders $p$ and $q$ | Equation (4.1.7) |
| ARIMA($p, d, q$) | AR and MA mixed model of orders $p$ and $q$ applied after data are differenced $d$ times | Section 1.4.1 |
| BS | Black-Scholes option pricing formula, call option | Equation (5.3.2) |
| CARA | Coefficient of absolute risk aversion, $-U''/U'$ | Equation (6.1.5) |
| CBOE | Chicago board of options exchange | Section 3.4 |

*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

| Symbol | Short Description | Presentation |
|---|---|---|
| CDF or cdf | Cumulative distribution function | Sections 2.1, 4.3.1 |
| CEV | Constant elasticity of variance | Section 1.4 |
| CIR SR | Cox, Ingersoll, and Ross square root | Section 1.4 |
| CIR V | Cox, Ingersoll, and Ross highly volatile | Section 1.4 |
| CLT | Central limit theorem | Section 9.1 |
| Cov1, 2 | Covariance of portfolio for diversification | Equation (2.2.2) |
| CPI | Corruption perception index | Section 7.4 |
| CPRM | Constant proportional risk aversion utility | Section 6.4 |
| CRRA | Coefficient of relative risk aversion $-xU''/U'$ | Section 6.1.1, 7.4 |
| D-boot | Double bootstrap | Section 9.3 |
| DSD | Downside standard deviation | Equation (5.2.1) |
| EBITDA | Earnings before interest, taxes, depreciation, and amortization | Section 7.3 |
| eCDF or ECDF | Empirical CDF | Section 4.3.2, Figure 4.3.2 |
| EMH | Efficient markets hypothesis | Sections 1.5, 4.1 |
| $E(P)$ | Expected price | Section 1.5 |
| EUT | Expected utility theory | Section 6.1 |
| $E(X)$ | Expected value of $X$ | Section 6.1 |
| Fed | Federal Reserve Bank (controls US monetarypolicy) | Section 1.5 |
| FDI | Foreign direct investment | |
| GARCH | Generalized autoregressive conditional heteroscedasticity | Section 4.2 |
| GBM | Geometric Brownian motion | Section 1.4 |
| H | Entropy or expectation of Shannon information | Equation (9.C.1) |
| HyDARA | Hyperbolic or diminishing absolute risk aversion utility | Section 7.4 |
| ICDF | Inverse cumulative probability distribution function | Section 9.2 |
| IQR | Inter-quartile range = $Q_3 - Q_1$. | Appendix C (Chapter 9) |
| iid | Independent and identically distributed | Section 9.1 |
| JD | Jump diffusion | Section 5.3 |
| $K$ | Total investment | Section 2.1 |
| KNIP | Kimball's nonincreasing prudence | Section 6.3.10 |
| LL | Log likelihood | Section 4.4 |
| $L(p)$ | Lorenz curve for proportion $p$ on unit square | Section 6.2.1 |
| MAD | Mean absolute deviation | Equation (1.3.2) |
| MA($q$) | Moving average of order $q$ | Equations (4.1.6), (4.1.6) |
| MBB | Moving blocks bootstrap | Section 9.3 |
| ME | Maximum entropy | Appendix C (Chapter 9) |
| ML | Maximum likelihood | Section 4.4 |
| MLPM | Mean lower partial moment | Section 6.4 |

| Symbol | Short Description | Presentation |
|---|---|---|
| MPC | Marginal propensity to consume | Section 6.3.5 |
| MU | Marginal utility | |
| NIARA | Nonincreasing absolute risk aversion | Section 6.3.4 |
| non-EUT | Nonexpected utility theory | Section 6.1 |
| PACF | Partial autocorrelation function | Section 9.3 |
| pdf or PDF | Probability distribution functions | Section 1.2 |
| Prudence | Kimball's prudence $-U''''/U''$ | Section 6.3.5 |
| $Q$–$Q$ | Quantile to quantile | Section 4.3.2 |
| $R$ | Interest rate, net percentage return over | Section 1.19.1 |
| $r$ | time period $\tau$ returns or excess returns | Section 2.1 |
| $r_f$ | Risk-free rate | Section 2.2 |
| $r_m$ | Excess return of the market portfolio in CAPM model | Section 2.2 |
| RSK($c$) | Measure of risk $(s - cS_k)$, $s$ = standard deviation, $S_k$ is skewness, $c$ is a constant | Equation (5.2.6) |
| $s$ | State | Section 1.1 |
| $S$ | Stock price | Section 1.4 |
| SEC | Securities and Exchange Commission | Section 7.4 |
| SN | Azzalini skew-normal density | Section 4.4 |
| SPX | Wall Street ticker symbol for S&P 500 | Section 6.3.10 |
| $t$ | Time subscript | Section 1.1 |
| TB3 | Interest rates on 3-month treasury bills | Appendix A (Chapter 9) |
| $V$ | Variance-covariance matrix | Equation (2.1.10), Section 7.4 |
| Van 500 | Vanguard 500 index fund | Appendix A (Chapter 9) |
| VaR | Value at risk | Section 2.1 |
| $W$ | Wealth | Section 7.4 |
| $w$ | Vector of weights | Section 7.4 |
| WKK | Wiener, Kolmogorov, and Khintchine | Section 9.1 |
| $x$ | Excess returns | Section 2.1 |
| $X$ | $n \times T$ matrix of excess returns | Section 2.4 |

# References


Abhyankar, A., L. Copeland, and W. Wong. (1997). Uncovering nonlinear structure in real-time stock-market indexes: The S&P 500, the Dax, the Nikkei 225, and the FTSE-100. *J. Bus. Econ. Stat.* 15(1): 1–14.

Ades, A., and R. Di Tella. (1999). Rents competition and corruption. *Am. Econ. Rev.* 89: 982–993.

Affleck-Graves, J., and B. McDonald. (1989). Non-normalities and tests of asset pricing theories. *J. Fin.* 44: 889–908.

Affleck-Graves, J., and B. McDonald. (1990). Multivariate tests of asset pricing: The comparative power of alternative statistics. *J. Finan. Quant. Anal.* 25: 163–185.

Ahn, S. C., Y. H. Lee, and P. Schmidt. (2001). GMM estimation of linear panel data models with time-varying individual effects. *J. Econometr.* 101: 219–255.

Akgiray, V., and C. G. Lamoureux. (1989). Estimation of stable law parameters: A comparative study. *J. Bus. Econ. Stat.* 7: 85–93.

Allen, F., and R. Michaely. (1995). Dividend policy. In R. Jarrow et al., eds., *Handbooks in OR & MS*, Vol. 9. Amsterdam, North Holland, pp. 963–992.

Anderson, G. (1996). Nonparametric tests of stochastic dominance in income distributions. *Econometrica* 64(5): 1183–1193.

Anderson, S., and W. Hauck. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Commun. Statist.* 12: 2663–2692.

Aït-Sahalia, Y., Y. Wang, and F. Yared. (2001). Do option markets correctly price the probabilities of movement of the underlying asset? *J. Econometr.* 102(1): 67–110.

Arrow, K. (1971). *Essays in the Theory of Risk Bearing*. Chicago: Markham Publishing.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scan. J. Stat.* 12: 171–178. *http://gwen.stat.unipd.it/SN/Intro/intro.html* has a brief introduction.

Bai, J., and S. Ng. (2002). Determining the number of factors in approximate factor models. *Econometrica* 70: 191–221.



*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7

Banz, R. W. (1981). The relationship between return and market value of common stocks. *J. Fin. Econ.* 9(1): 3–18.

Bardhan, P. (1997). Corruption and development: A review of issues. *J. Econ. Lit.* 35: 1320–1346.

Bartow, W. E., and W. H. Sun. (1987). Bootstrapped confidence intervals for linear relative risk models. *ASA Proc. Statist. Comput. Sect.* 260–265.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *J. Fin.* 32: 763–782.

Basu, S. (1983). The relationship between earnings yield, market value, and return for NYSE common stocks. *J. Fin. Econ.* 12(1): 129–156.

Bates, D., and D. G. Watts. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.

Bauman, W. S., S. M. Conover, and R. E. Miller. (1998). Growth stocks versus value stocks and large versus small cap stocks in international markets. *Fin. Anal. J.* 54: 75–89.

Bauman, W. S., and R. E. Miller. (1997). Investor expectations and the performance of value stocks versus growth stocks. *J. Portfolio Manag.* 23: 57–68.

Baxter, M., and U. J. Jermann. (1997). The international diversification puzzle is worse than you think. *Am. Econ. Rev.* 87(1): 170–180.

Bawa, V. S. (1975). Optimal rules for ordering uncertain prospects. *J. Fin. Econ.* 2: 95–121.

Berger, J. O., and M. Delampady. (1987). Testing precise hypotheses. *Stat. Sci.* 2: 317–352.

Beran, R. (1987). Prepivoting to reduce level error in confidence sets. *Biometrika* 74: 457–468.

Bhattacharya, S. (1979). Imperfect information, dividend policy, and the bird in the hand fallacy. *Bell J. Econ.* 10: 259–270.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *J. Bus.* (July): 444–455.

Black, F., and M. Scholes. (1973). The pricing of options and corporate liabilities. *J. Pol. Econ.* 81: 637–654.

Blanchard, O., and M. Watson. (1982). Bubbles, rational expectations and financial markets. In P. Wachtel, ed., *Crises in the Economic and Financial Structure*. Lexington, MA: Lexington Books.

Blaug, M. (1962). *Economic Theory in Retrospect*. Homewood IL: Irwin.

Bollerslev, T., R. Engle, and J. Woolridge. (1988). Capital asset pricing model with time-verying covariances. *J. Pol. Econ.* 96: 116–131.

Box, G. E. P., and G. C. Tiao. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison Wesley.

Brennan, M. J., and E. S. Schwartz. (1980). Analyzing convertible bonds. *J. Fin. Quantit. Anal.* 15: 907–929.

Brock, W. A., D. Hsieh, and B. LaBaron. (1991). *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Empirical Evidence*. Cambridge: MIT Press.

Brown, S. J., and W. N. Goetzmann. (1995). Performance persistence. *J. Fin.* 50: 679–698.

Camerer, C., and T. Ho. (1999). Experience-weighted attraction learning in normal form games. *Econometrica* 67: 827–874.

Campbell, J. Y. (1987). Stock returns and the term structure. *J. Fin. Econ.* 18: 373–399.

Campbell, J. Y., A. W. Lo, and A. C. MacKinley. (1997). *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Campbell, J. Y., and L. Hentschel. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *J. Fin. Econ.* 31: 281–318.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. (1997). *The Econometrics of Financial Markets.* Princeton: Princeton University Press.

Capaul, C., J. Rowley, and W. F. Sharpe. (1993). Intertemporal value and growth stock returns. *Fin. Anal. J.* 49: 27–41.

Carpenter, J. F., and A. Lynch. (1999). Survivorship bias and attrition effects in measures of performance persistence. *J. Fin. Econ.* 54: 337–374.

Carroll, C. D., and M. S. Kimball. (1996). On the concavity of the consumption function. *Econometrica* 64(4): 981–992.

Casella, G., and E. I. George. (1992). Explaining the Gibbs sampler. *Am. Statist.* 46(3): 167–185.

Cecchetti, S. G., P. Lam, and N. C. Mark. (1990). Mean reversion in equilibrium asset prices. *Am. Econ. Rev.* 80: 398–418.

Chamberlain, G., and M. R. Rothschild. (1983). Arbitrage and mean-variance analysis on large asset markets. *Econometrica* 51: 1281–1304.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders. (1992). An empirical comparison of alternate models of the short-term interest rate. *J. Fin.* 57: 1209–1227.

Chang, J.-K., J. C. Morley, C. R. Nelson, N. Chen, R. Roll, and S. A. Ross. (1986). Economic forces and the stock market. *J. Bus.* 59: 383–403.

Chang, R., and A. Velasco. (1998). Financial crises in emerging markets: A canonical model. Working Paper 98-21, C. V. Starr Center. Department of Economics, New York University.

Cheung, Y., and D. Griedman. (1997). Individual learning in normal form games: Some laboratory results. *Games and Econ. Beh.* 19: 46–76.

Chidambaran, N. K., C.-W. Jevons, and J. R. Trigueros. (2000). Option pricing via genetic programming. In Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, and A. S. Weigend, eds., *Computational Finance 1999*. Cambridge: MIT Press, pp. 583–598.

Chou, P. H., and G. Zhou. (1997). Using bootstrap to test portfolio efficiency. Paper presented to the fourth Asia–Pacific Finance Association Conference, Kuala Lumpur.

Clover, C. (2000). Russian oligarchs take the bankruptcy route to expansion. *Fin. Times*, July 25.

Connor, G., and R. A. Korajczyk. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *J. Fin. Econ.* 15: 373–394.

Connor, G., and R. A. Korajczyk. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *J. Fin. Econ.* 21: 255–289.

Connor, G., and R. A. Korajczyk. (1993). A test for the number of factors in an approximate factor model. *J. Fin.* 48: 1263–1291.

Constantinides, G. M., and J. E. Ingersoll. (1984). Optimal bond trading with personal taxes. *J. Fin. Econ.* 13: 299–335.

Copeland, T. E., and J. F. Weston. (1992). Financial theory and corporate policy, 3rd ed. New York: Addison Wesley.

Cox, J. C. (1975). Notes on option pricing I: Constant elasticity of variance diffusions. Working Paper. Stanford University.

Cox, D. R., and D. Hinkley. (1982). *Theoretical Statistics*. New York: Chapman and Hall.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. (1980). An analysis of variable rate loan contracts. *J. Fin.* 35: 389–403.

Cox, J. C., J. E. Ingersoll, and S. A. Ross. (1985). A theory of the term structure of interest rates. *Econometrica* 53: 385–407.

Cox, J. C., and S. Ross. (1976). The valuation of options for alternative stochastic processes. *J. Fin. Econ.* 3: 145–166.

Cross, F. (1973). The behavior of stock prices on Fridays and Mondays, *Fin. Anal. J.* 29: 67–69.

Cumby, R. E., and J. D. Glen. (1990). Evaluating the performance of international mutual funds. *J. Fin.* 45: 497–521.

Davison, A. C., and D. V. Hinkley. (1997). *Bootstrap Methods and Their Applications*. New York: Cambridge University Press.

DeRoon, F. A., and T. E. Nijman. (2001). Testing for mean variance spanning: A survey. *J. Empir. Fin.* 8(2): 111–155.

Deshpande, J. V., and H. Singh. (1985). Testing for second order stochastic dominance. *Commun. Stat. Theory Meth.* 14: 887–897.

D'Souza, M. I. (2002). Emerging countries stock markets: An analysis using an extension of the three-factor CAPM. Ph.D. dissertation. Economics Department, Fordham University.

Dothan, U. L. (1978). On the term structure of interest rates. *J. Fin. Econ.* 6: 59–69.

Easterwood, J. C., and S. R. Nutt. (1999). Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *J. Fin.* 54: 1777–1797.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *An. Stat.* 7: 1–26.

Efron, B., and R. Tibshirani. (1993). *An Intoduction to the Bootstrap.* New York: Chapman and Hall.

Elton, E. J., M. J. Gruber, and C. R. Blake. (1995). Survivorship bias and mutual fund performance. *Rev. Fin. Stud.* 9: 1097–1120.

Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *J. Econ. Persp.* 15(4): 157–168.

Erlich, I., and F. T. Lui. (1999). Bureaucratic corruption and endogenous economic growth. *J. Pol. Econ.* 107(6): s270–s293.

Espahbodi, R., A. Dugar, and H. Tehranian. (2001). Further evidence on optimism and underreaction in analysts' forecasts. *Rev. Fin. Econ.* 10: 1–21.

Fama, E. F., and K. R. French. (1988). Permanent and temporary components of stock prices. *J. Pol. Econ.* 96: 246–273.

Fama, E. F., and K. R. French. (1992). The cross section of expected stock returns. *J. Fin.* 47(2): 427–465.

Fama, E. F., and K. R. French. (1995). Size and book-to-market factors in earnings and returns. *J. Fin.* 50: 131–155.

Fama, E. F., and K. R. French. (1996). Multifactor explanations of asset pricing anomalies. *J. Fin.* 51: 55–84.

Fama, E. F., and J. D. MacBeth. (1973). Risk, return and equilibrium: Empirical test. *J. Pol. Econ.* 81: 607–636.

Fama, E. F., and R. Roll. (1971). Parameter estimates for symmetric stable distributions. *J. Am. Stat. Assoc.* 66: 331–338.

French, K. R., G. W. Schwert, and R. F. Stambaugh. (1987). Expected stock returns and volatility. *J. Fin. Econ.* 19: 3–29.

French, K. R., and J. M. Poterba. (1991). Investor diversification and international equity markets. *Am. Econ. Rev.* 81(2): 222–226.

Friedman, M., and L. J. Savage. (1948). The utility analysis of choices involving risk. *J. Pol. Economy* 56(4): 279–304.

Friedman, J., and W. Stuetzle. (1981). Projection pursuit regression. *J. Am. Stat. Assoc.* 76: 817–823.

GAUSS, Aptech Corp. Maple Valley, WA (info@aptech.com).

Geddes, R., and H. D. Vinod. (2002). CEO tenure, board composition, and regulation. *J. Reg. Econ.* 21: 217–235.

Gibbons, M., and W. Ferson. (1985). Tests of asset pricing models with changing expectations and an unobservable market portfolio. *J. Fin. Econ.* 14: 217–236.

Gill, P. E., W. Murray, and M. H. Wright. (1981). *Practical Optimization.* New York: Academic Press.

Gordon, M. G., M. G. Paradis, and C. Rorke. (1972). Experimental evidence on alternative portfolio decision rules. *Am. Econ. Rev.* 62: 107–118.

Greene, W. H. (2000). *Econometric Analysis*, 4th ed. Upper Saddle River, NJ: Prentice Hall.

Groenewold, N., and P. Fraser. (1997). Share prices and macroeconomic factors. *J. Bus. Fin. Account.* 24: 1367–1383.

Gruber, M. J. (1996). Another puzzle: The growth in actively managed mutual funds. *J. Fin.* 51: 783–811.

Hansen, L. P., and R. Jeganathan. (1991). Implications of security market data for models of dynamic economies. *J. Pol. Econ.* 99: 225–262.

Hardy, D. C., and C. Pazarbasioglu. (1999). Determinants and leading indicators of banking crises: Further evidence. *IMF Staff Papers* 46(3): 247–258.

Hendricks, D., J. Patel, and R. Zeckhauser. (1997). The J-shape of performance: Persistence given survivorship bias. *Rev. Econ. Stat.* 79: 161–170.

Heyde, C. C. (1997). *Quasi-likelihood and Its Applications.* New York: Springer Verlag.

Hogan, W., and J. Warren. (1974). Towards the development of an equilibrium capital market model based on semivariance. *J. Fin. Quantit. Anal.* 9: 1–11.

Hsu, D.-A., R. B. Miller, and D. W. Wichern. (1974). On the stable Paretian behaviour of stock-market prices. *J. Am. Stat. Assoc.* 69(345): 108–113.

Huberman, G., and S. Kandel. (1987). Mean variance spanning. *J. Fin.* 42: 873–888.

Hull, J. C. (1997). *Options, Futures, and Other Derivatives*, 4th ed. Upper Saddle River, NJ: Prentice Hall.

Huang, C., and R. H. Litzenberger. (1988). *Foundations for Financial Economics*. Upper Saddle River, NJ: Prentice Hall.

Ibbotson, R. G., and J. R. Ritter. (1995). Initial Public Offerings. In R. Jarrow et al., eds., *Handbooks in OR & MS*, 9: 993–1016.

IMF. (2001). International monetary fund. Annual report on exchange restrictions.

Ingersoll, J. E. (1997). *Theory of Financial Decision Making*. Savage, MD: Rowman and Littlefield.

Jegadeesh, N. (1991). Seasonality in stock price mean reversion: Evidence from the U.S. and the U.K. *J. Fin.* 46: 1427–1444.

Jobson, J. D., and B. M. Korkie. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *J. Fin.* 36(4): 889–908.

Jog, V. (1998). Canadian stock pricing anomalies: Revisited. *Can. Invest. Rev.* (Winter): 28–33.

Kagan, A. M., Y. V. Linnik, and C. R. Rao. (1973). *Characterization Problems in Mathemataical Statistics*. New York: Wiley.

Kahneman, D., and A. Tversky. (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2): 263–291.

Kaminski, G., S. Lizondo, and C. Reinhart. (1998). Leading indicators of currency crises. *IMF Staff Papers* 45: 1–45.

Kaminski, G., C. Reinhart, and C. Vegh. (2003). The unholy trinity of financial contagion. *J. Econ. Persp.* 17: 51–74.

Karakitos, E. (2002). *A Macro and Financial Model of G4.* London: Trafalgar Asset Managers Ltd.

Keim, D. (1983). Size-related anomalies and stock return seasonality. *J. Fin. Econ.* 12: 13–22.

Kendall, M. G., and A. Stuart. (1979). *The Advanced Theory of Statistics*, 4th ed. Vol. 2. New York: Macmillan.

Kim, C.-J., and C. R. Nelson. (1998). Testing for mean reversion is heteroskedastic data II: Autoregression tests based on Gibbs-sampling-augmented randomization. *J. Empir. Fin.* 5: 385–396.

Kim, C.-J., J. C. Morley, and C. R. Nelson. (2001). Is there a positive relationship between stock market volatility and the equity premium? Manuscript. Washington University.

Kim, M.-J., C. R. Nelson, and R. Startz. (1991). Mean reversion in stock prices? A reappraisal of the empirical evidence. *Rev. Econ. Stud.* 48: 515–528.

Koenig, E. F., T. F. Siems, and M. A. Wynne. (2002). New economy, new recession? *Southwest Econ. Fed. Res. Bank of Dallas* 2: 11–16.

Kovaleski, D. (2002). Falling down: Assets sink 7.2% in a bad year for DC mutual funds; managers follow market downturn; assets in top funds fall to $282 billion. *Pensions and Invest.* (April) 29: 21.

Kuan, C.-M., and H. White. (1994). Artificial neural networks: An econometric perspective. *Economet. Rev.* 13(1): 1–91.

Kraus, A., and R. H. Litzenberger. (1976). Skewness preference and the valuation of risk assets. *J. Fin.* 31(4): 1085–1100.

Lakonishok, J., A. Shleifer, and R. W. Vishny. (1994). Contrarian investment, extrapolation, and risk. *J. Fin.* 49: 1541–1578.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer, and R. W. Vishny. (1998). Law and finance. *J. Pol. Econ*. 106(6): 1113–1155.

Lamont, O. A., and R. H. Thaler. (2003). Anomalies: The law of one price in financial markets. *J. Econ. Persp.* 17(4): 191–202.

Lattimore, P. K., P. Baker, and A. D. Witte. (1992). The influence of probability on risky choice: A parametric examination. *J. Econ. Beh. Org.* 17: 377–400.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. New York: Wiley.

Lehmann, E. L. (1993). Theories of testing hypotheses: One theory or two? *J. Am. Stat. Assoc.* 88: 1242–1249.

Leinweber, D. (2001). Internet information for transaction cost control. In B. R. Bruce, ed., *Transaction Costs: A Cutting Edge to Best Execution*, Institutional Investors, Inc.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev. Econ. Stat.* 47: 13–37.

Lintner, J. (1969). Aggregation of investor's diverse judgments and preferences in a purely competitive security markets. *J. Fin. Quantit. Anal.* (Dec.) 4: 347–400.

Longin, F., and B. Solnik. (2001). Extreme correlation of international equity markets. *J. Fin.* 56(2): 649–676.

Maasoumi, E., and J. Racine. (2002). Entropy and predictability of stock market returns. *J. Econometr.* 107: 291–312.

MacKinlay, A. C., and M. P. Richardson. (1991). Using generalized methods of moments to test mean-variance efficiency. *J. Fin.* 46: 511–527.

Maheu, J. M., and T. H. McCurdy. (2000). Identifying bull and bear markets in stock returns. *J. Bus. Econ. Stat.* 18: 100–112.

Malkiel, B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. *J. Fin.* 50: 549–572.

Mardia, K. V., J. T. Kent, and J. M. Bibby. (1979). *Multivariate Analysis*. New York: Academic Press.

Markowitz, H. (1952a). The utility of wealth. *J. Pol. Economy* 60: 151–158.

Markowitz, H. (1952b). Portfolio selection. *J. Fin.* 7: 77–91.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York: Wiley.

Marsaglia, G. (1996). DIEHARD: A battery of tests of randomness. *stat.fsu.edu/~geo/. E-mail: geo@stat.fsu.edu*.

Mayers, D. (1972). Non-marketaable assets and the capital market equilibrium under uncertainty. In M. Jensen, ed., *Studies in the Theory of Capital Markets*. New York: Praeger, pp. 223–248.

McCullagh, P., and J. A. Nelder. (1989). *Generalized Linear Models*, 2nd ed. New York: Chapman and Hall.

McCullough, B. D. (2004). Some details of nonlinear estimation. In M. Altman, J. Gill, and M. P. McDonald, eds., *Numerical Methods in Statistical Computing for the Social Sciences*. New York: Wiley.

McCullough, B. D., and C. G. Renfro. (1999). Benchmarks and software standards: A case study of GARCH procedures. *J. Econ. Soc. Meas.* 25(2): 59–71.

McCullough, B. D., and C. G. Renfro. (2000). Some numerical aspects of nonlinear estimation. *J. Econ. Soc. Meas.* 26(1): 63–77.

McCullough, B. D., and H. D. Vinod. (1998). Implementing the double bootstrap. *Comput. Econ.* 12: 79–95.

McCullough, B. D., and H. D. Vinod. (1999). The numerical reliability of econometric software. *J. Econ. Lit.* 37(2): 633–665.

McCullough, B. D., and H. D. Vinod. (2003a). Verifying the solution from a nonlinear solver. *Am. Econ. Rev.* 93(3): 873–892.

McCullough, B. D., and H. D. Vinod. (2003b). Comments: Econometrics and software. *J. Econ. Persp.* 17(1): 223–224.

McCullough, B. D., and H. D. Vinod. (2004). Verifying the solution from a nonlinear solver. Reply. *Amer. Econ. Rev.* 94(1): 391–396 & 400–403.

McElroy, M. B., and E. Burmeister. (1988). Arbitrage pricing theory as a restricted nonlinear multivariate regression model: Iterated nonlinear seemingly unrelated regression estimates. *J. Bus. Econ. Stat.* 6: 29–42.

Mehra, R., and E. Prescott. (1985). The equity premium puzzle. *J. Monetary Econ.* 15: 145–161.

Merton, R. (1973a). An intertemporal capital asset pricing model. *Econometrica* (September) 41: 867–888.

Merton, R. C. (1973b). Theory of rational option pricing. *Bell J. Econ. Manag. Sci.* 4: 141–183.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *J. Fin. Econ.* 3: 125–144.

Merton, R. C. (1990). *Continuous Time Finance*. Cambridge, MA: Blackwell.

Mookherjee, D., and B. Sopher. (1994). Learning behavior in an experimental matching pennies game. *Games and Econ. Beh.* 7: 62–91.

Morey, M. R., and H. D. Vinod. (2001). A double Sharpe ratio. In C. F. Lee, ed., *Advances in Investment Analysis and Portfolio Management*, Vol. 8. New York: JAI-Elsevier Science, pp. 57–65.

Morningstar. (2000). *On-Disk, Principia, Principia-Plus, and Principia-Pro Manuals* (Various editions from 1991–2000). Chicago: Morningstar Publications.

NIST. (1999). National Institute of Standards and Technology. US Dept. of Commerce, Technology Administration, Gaithersburg, MD 20899-001. Statistical Reference Data Sets (StRD). *www.itl.nist.gov/div898/strd/general/dataarchive.html*.

Omran, M. F. (2001). International evidence on extreme values and value at risk of stock returns. *Proc. Business and Economic Statistics Section of the American Statistical Association.* Alexandria, VA: ASA.

Ord, J. K. (1968). The discrete Student's distribution. *An. Math. Stat.* 39: 1513.

Pedersen, C. S. (1998). Empirical tests for differences in equilibrium risk measures with application to downside risk in small and large UK companies. U. of Cambridge Discussion Papers in Accounting and Finance: AF41.

Poterba, J. M., and L. H. Summers. (1988). Mean reversion in stock prices: Evidence and implications. *J. Fin. Econ.* 22: 27–59.

Prelec, D. (1998). The probability weighting function. *Econometrica* 66(3): 497–527.

Qi, M. (1999). Nonlinear predictability of stock returns using financial and economic variables. *J. Bus. Econ. Stat.* 17(4): 419–429.

Rachev, S. T., and S. Mittnik. (2000). *Stable Paretian Models in Finance*. New York: Wiley.

Rachev, S., E. Schwartz, and I. Khindanova. (2001). Stable modeling of market and credit value at risk. *Proc. Business and Economic Statistics Section of the American Statistical Association.* Alexandria, VA: ASA.

Rachev, S. T., and S. Mittnik. (2000). Paretian models in finance. New York: Wiley.

Racine, J. (2001). On the nonlinear predictability of stock returns using financial and economic variables. *J. Bus. Econ. Stat.* 19(3): 380–382.

Reagle D., and H. D. Vinod. (2003). Inference for negativist theory using numerically computed rejection regions. *Computa. Statis. Data Ana.* 42: 491–512.

Reagle, D., and H. D. Vinod. (2002). Measuring the asymmetry of stock market risk. 2001 *Proc. Bus. Econ. Statist. Sec. Amer. Stat. Assoc.* Alexandria, VA: ASA.

Richardson, M. (1993). Temporary components of stock prices: A skeptic's view. *J. Bus. Econ. Stat.* 11: 199–207.

Richardson, M., and J. H. Stock. (1989). Drawing inferences from statistics based on multiyear asset returns. *J. Fin. Econ.* 25: 323–348.

Richardson, M. P., and T. Smith. (1993). A test for multivariate normality in stock returns. *J. Bus.* 66: 295–321.

Ritchken, P., and W. K. Chen. (1987). Downside risk option pricing models. In S. Khoury, and A. Ghosh, eds., *Recent Developments in International Banking and Finance*, Vol. 2. Lexington, MA: D. C. Heath, pp. 205–225.

Rogers, J., J. Filliben, L. Gill, W. Guthrie, E. Lagergren, and M. Vangel (1998). StRD: Statstical reference datasets for testing the numerical accuracy of statistical software. NIST\#1396. National Institute of Standards and Technology.

Roll, R. (1976). A critique of the asset pricing theory's tests. *J. Fin. Econ.* (March) 4: 129–176.

Roll, R., and S. Ross. (1980). An empirical investigation of the arbitrage pricing theory. *J. Fin.* (December) 35: 1073–1103.

Rose, C., and M. D. Smith. (2002). *Mathematical Statistics with Mathematica*. New York: Springer Verlag.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *J. Econ. Theory* 13: 341–360.

Rubinstein, M. (1973). A mean-variance synthesis of corporate finance theory. *J. Fin.* (March) 28: 167–182.

Rubinstein, M. (1976). The valuation of uncertain income streams and the pricing of options. *Bell J. Econ. Manag. Sci.* 7: 407–425.

Salvatore, D. (1999). Lessons from the financial crisis in East Asia. *J. Pol. Modeling* 21: 283–347.

Salvatore, D., and D. Reagle. (2000). Forecasting financial crises in emerging market economies. *Open Econ. Rev.* 11: 247–259.

Samuelson, P. (1983). *Foundations of Economic Analysis*. Cambridge: Harvard University Press.

Santangelo, G. (2004). Analysis of international data revision: Theory and application. Ph.D. dissertation. Economics Department, Fordham University.

Schaller, H., and S. van Norden. (1997). Regime switching in stock market returns. *Appl. Fin. Econ.* 7: 177–191.

Schinasi, G. J., and R. T. Smith. (2000). Portfolio diversification, leverage and financial contagion. *IMF Staff Papers* 47(2): 159–176.

Schmid, F., and M. Trede. (1996). Nonparametric tests for second-order stochastic dominance from paired observations. In V. D. Lippe et al., eds., *Wirtschafts und Sozialstatistik Heute, Theorie und Praxis*, Festschrift für Walter Krug. Köln, Germany: Verlag Wissenschaft & Praxis.

Schmid, F., and M. Trede. (1998). A Kolmogorov-type test for second-order stochastic dominance. *Stat. Prob. Lett.* 37: 183–193.

Schuirmann, D. L. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokin. Biopharmac.* 15: 657–680.

Schwartz, E. S., and W. N. Torous. (1989). Prepayment and the valuation of mortgage-backed securities. *J. Fin.* 45: 375–392.

Schwert, G. W. (1989). Business cycles, financial crises, and stock volatility. *Carnegie-Rochester Conference Series on Public Policy* 31: 83–126.

Serrat, A. (2001). A dynamic equilibrium model of international portfolio holdings *Econometrica* 69(6): 1467–1490.

Shanken, J. (1985a). Multi-beta CAPM or equilibrium APT? A reply. *J. Fin.* (September) 40: 1189–1196.

Shanken, J. (1985b). Multivariate tests of the zero-beta CAPM. *J. Fin. Econ.* (September) 14: 327–348.

Sharpe, W. F. (1966). Mutual fund performance. *J. Bus.* 39(P2): 119–138.

Sheather, S. J., and M. C. Jones. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Stat. Methodol.: J. Royal Stat. Soc. Ser. B* 53: 683–690.

Shenton, L. R., and H. D. Vinod. (1995). Closed forms for asymptotic bias and variance in autoregressive models with unit roots. *J. Comput. Appl. Math.* 61: 231–243.

Silverman, B. W. (1986). Density estimation for statistics and data analysis. New York: Chapman and Hall.

S-Plus. Insightful Corp. 1700 Westlake Avenue North. Seattle, WA. *info@insightful.com*.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *J. Econ. Lit.* 38: 332–382.

Summers, L. H. (1986). Does the stock market rationally reflect fundamental values? *J. Fin.* 41: 591–601.

Tenorio, R., and T. N. Cason. (2002). To spin or not to spin? Natural and laboratory experiments from "the price is right". *Econ. J.* 112: 170–195.

ter Horst, J. R., T. E. Nijman, and M. Verbeek. (2001). Eliminating look-ahead bias in evaluating persistence in mutual fund performance. *J. Emp. Fin.* 8: 345–373.

Theil, H., and K. Laitinen. (1980). Singular moment matrices in applied econometrics. In P. R. Krishnaiah, ed., *Multivariate Analysis—V*. New York: North-Holland, pp. 629–649.

Thompson, J., M. K. Yucel, and J. V. Duca. (2002). Economic impact of biotechnology, southwest economy. *Federal Reserve Bank of Dallas* (2): 1–20.

Treynor, J. L. (1965). How to rate management of investment funds. *Harvard Bus. Rev.* 43: 63–75.

Tse, R. Y. C. (2001). Impact of property prices on stock prices in Hong Kong. *Rev. Pacific Basin Fin. Markets Policies* 4(1): 29–43.

Tsay, R. S. (2002). *Analysis of Financial Time Series*. New York: Wiley.

Turner, C. M., R. Startz, and C. R. Nelson. (1989). A Markov model of heteroskedasticity, risk, and learning in the stock market. *J. Fin. Econ.* 25: 3–22.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *J. Fin. Econ.* 5: 177–188.

Vinod, H. D. (1982). Maximum entropy measurement error estimator of singular covariance matrices. *J. Econ.* 20: 163–174.

Vinod, H. D. (1985). Measurement of economic distance between blacks and whites. *J. Bus. Econ. Stat.* 3: 78–88.

Vinod, H. D. (1993). Bootstrap, jackknife resampling and simulation: Applications in econometrics. In G. S. Maddala, C. R. Rao, and H. D. Vinod, eds., *Handbook of Statistics: Econometrics*, Vol. 11. New York: North Holland, pp. 629–661.

Vinod, H. D. (1995). Double bootstrap for shrinkage estimators. *J. Econom.* 68: 287–302.

Vinod, H. D. (1996). Consumer behavior and a target seeking supply side model for income determination. In M. Ahsanullah, and D. Bhoj, eds., *Applied Statistical Science, I*. Carbondale, IL: Nova Science, pp. 219–233.

Vinod, H. D. (1997). Concave consumption, Euler equation and inference using estimating functions. *Proc. Business and Economic Statistics Section of the American Statistical Association*. Alexandria, VA: ASA, pp. 118–123.

Vinod, H. D. (1998a). Nonparametric estimation of nonlinear money demand cointegration equation by projection pursuit methods. In S. Weisberg, ed., *30th Symp. on the Interface, Computing Science and Statistics*, Vol. 30. Fairfax Station, VA: Interface Foundation, p. 174.

Vinod, H. D. (1998b). Foundations of statistical inference based on numerical roots of robust pivot functions (Fellow's corner). *J. Econom.* 86: 387–396.

Vinod, H. D. (1999). Statistical analysis of corruption data and using the internet to reduce corruption. *J. Asian Econ.* 10: 591–603.

Vinod, H. D. (2000a). Review of Gauss for windows, including its numerical accuracy. *J. Appl. Econom.* 15(2): 211–220.

Vinod, H. D. (2000b). Ranking mutual funds using unconventional utility theory and stochastic dominance. Discussion Paper. Fordham University.

Vinod, H. D. (2000c). Foundations of multivariate inference using modern computers. *Linear Algebra Appl.* 321: 365–385.

Vinod, H. D. (2000d). Non-expected utility, stochastic dominance and new summary measures for risk. Presented at Joint Statistical Meetings of International Indian Statistical Association. *J. Quanti. Econ.* 17(1): 1–24.

Vinod, H. D. (2001). Non-expected utility, stochastic dominance and new summary measures for risk. *J. Quanti. Econ.* 17(1): 1–24.

Vinod, H. D. (2003a). Open economy and financial burden of corruption: Theory and application to Asia. *J. Asian Econ.* 12: 873–890.

Vinod, H. D. (2003b). Constructive ensembles for time series in econometrics and finance. *Proc. 35th Symp. on Interface between Computing Science and Statistics* [CD-ROM]. Interface Foundation of North America, PO Box 7460, Fairfax Station, VA 22039. Presented in Salt Lake City, March 14, 2003.

Vinod, H. D. (2004). Ranking mutual funds using unconventional utility theory and stochastic dominance. *J. Empir. Fin.* 11(3): 353–377.

Vinod, H. D., and M. R. Morey. (2000). Confidence intervals and hypothesis testing for the Sharpe and Treynor performance measures: A bootstrap approach. In Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, and A. S. Weigend, eds., *Computational Finance 1999*. Cambridge: MIT Press, pp. 25–39.

Vinod, H. D., and M. R. Morey. (2001). A double Sharpe ratio. In C. F. Lee, ed., *Advances in Investment Analysis and Portfolio Management*, Vol. 8. New York: JAI-Elsevier Science, pp. 57–65.

Vinod, H. D., and L. R. Shenton. (1996). Exact moments for autoregressive and random walk models for a zero or stationary initial value. *Economet. Theory* 12: 481–499.

Xie, A. (2002). Downside Risk Measures. Ph.D. dissertation. Economics Department, Fordham University.

Yu, E. S. H., I. C. H. Chin, H. F. Y. Hang, and G. L. C. Wai. (2001). Managing risk by using derivatives: The case of Hong Kong firms. *Rev. Pacific Basin Fin. Markets Policies* 4(4): 417–425.

Zheng, L. (1999). Is money smart? A study of mutual fund investors' fund selection ability. *J. Fin.* 54: 901–933.

# Name Index



*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

# Index



*Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments*, by Hrishikesh D. Vinod and Derrick P. Reagle
ISBN 0-471-23442-7 

## WILEY SERIES IN PROBABILITY AND STATISTICS

ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS



The ***Wiley Series in Probability and Statistics*** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches.

This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

ABRAHAM and LEDOLTER · Statistical Methods for Forecasting
AGRESTI · Analysis of Ordinal Categorical Data
AGRESTI · An Introduction to Categorical Data Analysis
AGRESTI · Categorical Data Analysis, *Second Edition*
ALTMAN, GILL, and McDONALD · Numerical Issues in Statistical Computing for the Social Scientist
AMARATUNGA and CABRERA · Exploration and Analysis of DNA Microarray and Protein Array Data
ANDEĚL · Mathematics of Chance
ANDERSON · An Introduction to Multivariate Statistical Analysis, *Third Edition*
*ANDERSON · The Statistical Analysis of Time Series
ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG · Statistical Methods for Comparative Studies
ANDERSON and LOYNES · The Teaching of Practical Statistics
ARMITAGE and DAVID (editors) · Advances in Biometry
ARNOLD, BALAKRISHNAN, and NAGARAJA · Records
*ARTHANARI and DODGE · Mathematical Programming in Statistics
*BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences
BALAKRISHNAN and KOUTRAS · Runs and Scans with Applications
BARNETT · Comparative Statistical Inference, *Third Edition*
BARNETT and LEWIS · Outliers in Statistical Data, *Third Edition*
BARTOSZYNSKI and NIEWIADOMSKA-BUGAJ · Probability and Statistical Inference
BASILEVSKY · Statistical Factor Analysis and Related Methods: Theory and Applications
BASU and RIGDON · Statistical Methods for the Reliability of Repairable Systems
BATES and WATTS · Nonlinear Regression Analysis and Its Applications
BECHHOFER, SANTNER, and GOLDSMAN · Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons
BELSLEY · Conditioning Diagnostics: Collinearity and Weak Data in Regression

*Now available in a lower priced paperback edition in the Wiley Classics Library.

*BELSLEY, KUH, and WELSCH · Regression Diagnostics: Identifying Influential Data and Sources of Collinearity
BENDAT and PIERSOL · Random Data: Analysis and Measurement Procedures, *Third Edition*
BERRY, CHALONER, and GEWEKE · Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner
BERNARDO and SMITH · Bayesian Theory
BHAT and MILLER · Elements of Applied Stochastic Processes, *Third Edition*
BHATTACHARYA and WAYMIRE · Stochastic Processes with Applications
BILLINGSLEY · Convergence of Probability Measures, *Second Edition*
BILLINGSLEY · Probability and Measure, *Third Edition*
BIRKES and DODGE · Alternative Methods of Regression
BLISCHKE AND MURTHY (editors) · Case Studies in Reliability and Maintenance
BLISCHKE AND MURTHY · Reliability: Modeling, Prediction, and Optimization
BLOOMFIELD · Fourier Analysis of Time Series: An Introduction, *Second Edition*
BOLLEN · Structural Equations with Latent Variables
BOROVKOV · Ergodicity and Stability of Stochastic Processes
BOULEAU · Numerical Methods for Stochastic Processes
BOX · Bayesian Inference in Statistical Analysis
BOX · R. A. Fisher, the Life of a Scientist
BOX and DRAPER · Empirical Model-Building and Response Surfaces
*BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process Improvement
BOX, HUNTER, and HUNTER · Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building
BOX and LUCEÑO · Statistical Control by Monitoring and Feedback Adjustment
BRANDIMARTE · Numerical Methods in Finance: A MATLAB-Based Introduction
BROWN and HOLLANDER · Statistics: A Biomedical Introduction
BRUNNER, DOMHOF, and LANGER · Nonparametric Analysis of Longitudinal Data in Factorial Experiments
BUCKLEW · Large Deviation Techniques in Decision, Simulation, and Estimation
CAIROLI and DALANG · Sequential Stochastic Optimization
CASTILLO, HADI, BALAKRISHNAN, and SARABIA · Extreme Value and Related Models with Applications in Engineering and Science
CHAN · Time Series: Applications to Finance
CHATTERJEE and HADI · Sensitivity Analysis in Linear Regression
CHATTERJEE and PRICE · Regression Analysis by Example, *Third Edition*
CHERNICK · Bootstrap Methods: A Practitioner's Guide
CHERNICK and FRIIS · Introductory Biostatistics for the Health Sciences
CHILÈS and DELFINER · Geostatistics: Modeling Spatial Uncertainty
CHOW and LIU · Design and Analysis of Clinical Trials: Concepts and Methodologies, *Second Edition*
CLARKE and DISNEY · Probability and Random Processes: A First Course with Applications, *Second Edition*
*COCHRAN and COX · Experimental Designs, *Second Edition*
CONGDON · Applied Bayesian Modelling
CONGDON · Bayesian Statistical Modelling
CONOVER · Practical Nonparametric Statistics, *Third Edition*
COOK · Regression Graphics
COOK and WEISBERG · Applied Regression Including Computing and Graphics
COOK and WEISBERG · An Introduction to Regression Graphics
CORNELL · Experiments with Mixtures, Designs, Models, and the Analysis of Mixture Data, *Third Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

COVER and THOMAS · Elements of Information Theory
COX · A Handbook of Introductory Statistical Methods
*COX · Planning of Experiments
CRESSIE · Statistics for Spatial Data, *Revised Edition*
CSÖRGŐ and HORVÁTH · Limit Theorems in Change Point Analysis
DANIEL · Applications of Statistics to Industrial Experimentation
DANIEL · Biostatistics: A Foundation for Analysis in the Health Sciences, *Eighth Edition*
*DANIEL · Fitting Equations to Data: Computer Analysis of Multifactor Data, *Second Edition*
DASU and JOHNSON · Exploratory Data Mining and Data Cleaning
DAVID and NAGARAJA · Order Statistics, *Third Edition*
*DEGROOT, FIENBERG, and KADANE · Statistics and the Law
DEL CASTILLO · Statistical Process Adjustment for Quality Control
DEMARIS · Regression with Social Data: Modeling Continuous and Limited Response Variables
DEMIDENKO · Mixed Models: Theory and Applications
DENISON, HOLMES, MALLICK and SMITH · Bayesian Methods for Nonlinear Classification and Regression
DETTE and STUDDEN · The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis
DEY and MUKERJEE · Fractional Factorial Plans
DILLON and GOLDSTEIN · Multivariate Analysis: Methods and Applications
DODGE · Alternative Methods of Regression
*DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
*DOOB · Stochastic Processes
DOWDY, WEARDEN, and CHILKO · Statistics for Research, *Third Edition*
DRAPER and SMITH · Applied Regression Analysis, *Third Edition*
DRYDEN and MARDIA · Statistical Shape Analysis
DUDEWICZ and MISHRA · Modern Mathematical Statistics
DUNN and CLARK · Basic Statistics: A Primer for the Biomedical Sciences, *Third Edition*
DUPUIS and ELLIS · A Weak Convergence Approach to the Theory of Large Deviations
*ELANDT-JOHNSON and JOHNSON · Survival Models and Data Analysis
ENDERS · Applied Econometric Time Series
ETHIER and KURTZ · Markov Processes: Characterization and Convergence
EVANS, HASTINGS, and PEACOCK · Statistical Distributions, *Third Edition*
FELLER · An Introduction to Probability Theory and Its Applications, Volume I, *Third Edition*, Revised; Volume II, *Second Edition*
FISHER and VAN BELLE · Biostatistics: A Methodology for the Health Sciences
FITZMAURICE, LAIRD, and WARE · Applied Longitudinal Analysis
*FLEISS · The Design and Analysis of Clinical Experiments
FLEISS · Statistical Methods for Rates and Proportions, *Third Edition*
FLEMING and HARRINGTON · Counting Processes and Survival Analysis
FULLER · Introduction to Statistical Time Series, *Second Edition*
FULLER · Measurement Error Models
GALLANT · Nonlinear Statistical Models
GHOSH, MUKHOPADHYAY, and SEN · Sequential Estimation
GIESBRECHT and GUMPERTZ · Planning, Construction, and Statistical Analysis of Comparative Experiments
GIFI · Nonlinear Multivariate Analysis
GLASSERMAN and YAO · Monotone Structure in Discrete-Event Systems
GNANADESIKAN · Methods for Statistical Data Analysis of Multivariate Observations, *Second Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

GOLDSTEIN and LEWIS · Assessment: Problems, Development, and Statistical Issues
GREENWOOD and NIKULIN · A Guide to Chi-Squared Testing
GROSS and HARRIS · Fundamentals of Queueing Theory, *Third Edition*
*HAHN and SHAPIRO · Statistical Models in Engineering
HAHN and MEEKER · Statistical Intervals: A Guide for Practitioners
HALD · A History of Probability and Statistics and their Applications Before 1750
HALD · A History of Mathematical Statistics from 1750 to 1930
HAMPEL · Robust Statistics: The Approach Based on Influence Functions
HANNAN and DEISTLER · The Statistical Theory of Linear Systems
HEIBERGER · Computation for the Analysis of Designed Experiments
HEDAYAT and SINHA · Design and Inference in Finite Population Sampling
HELLER · MACSYMA for Statisticians
HINKELMAN and KEMPTHORNE: · Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design
HOAGLIN, MOSTELLER, and TUKEY · Exploratory Approach to Analysis of Variance
HOAGLIN, MOSTELLER, and TUKEY · Exploring Data Tables, Trends and Shapes
*HOAGLIN, MOSTELLER, and TUKEY · Understanding Robust and Exploratory Data Analysis
HOCHBERG and TAMHANE · Multiple Comparison Procedures
HOCKING · Methods and Applications of Linear Models: Regression and the Analysis of Variance, *Second Edition*
HOEL · Introduction to Mathematical Statistics, *Fifth Edition*
HOGG and KLUGMAN · Loss Distributions
HOLLANDER and WOLFE · Nonparametric Statistical Methods, *Second Edition*
HOSMER and LEMESHOW · Applied Logistic Regression, *Second Edition*
HOSMER and LEMESHOW · Applied Survival Analysis: Regression Modeling of Time to Event Data
HUBER · Robust Statistics
HUBERTY · Applied Discriminant Analysis
HUNT and KENNEDY · Financial Derivatives in Theory and Practice
HUSKOVA, BERAN, and DUPAC · Collected Works of Jaroslav Hajek—with Commentary
HUZURBAZAR · Flowgraph Models for Multistate Time-to-Event Data
IMAN and CONOVER · A Modern Approach to Statistics
JACKSON · A User's Guide to Principle Components
JOHN · Statistical Methods in Engineering and Quality Assurance
JOHNSON · Multivariate Statistical Simulation
JOHNSON and BALAKRISHNAN · Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz
JOHNSON and BHATTACHARYYA · Statistics: Principles and Methods, *Fifth Edition*
JOHNSON and KOTZ · Distributions in Statistics
JOHNSON and KOTZ (editors) · Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 1, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 2, *Second Edition*
JOHNSON, KOTZ, and BALAKRISHNAN · Discrete Multivariate Distributions
JOHNSON, KOTZ, and KEMP · Univariate Discrete Distributions, *Second Edition*
JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of Econometrics, *Second Edition*
JUREČKOVÁ and SEN · Robust Statistical Procedures: Aymptotics and Interrelations

*Now available in a lower priced paperback edition in the Wiley Classics Library.

JUREK and MASON · Operator-Limit Distributions in Probability Theory
KADANE · Bayesian Methods and Ethics in a Clinical Trial Design
KADANE AND SCHUM · A Probabilistic Analysis of the Sacco and Vanzetti Evidence
KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data, *Second Edition*
KASS and VOS · Geometrical Foundations of Asymptotic Inference
KAUFMAN and ROUSSEEUW · Finding Groups in Data: An Introduction to Cluster Analysis
KEDEM and FOKIANOS · Regression Models for Time Series Analysis
KENDALL, BARDEN, CARNE, and LE · Shape and Shape Theory
KHURI · Advanced Calculus with Applications in Statistics, *Second Edition*
KHURI, MATHEW, and SINHA · Statistical Tests for Mixed Linear Models
*KISH · Statistical Design for Research
KLEIBER and KOTZ · Statistical Size Distributions in Economics and Actuarial Sciences
KLUGMAN, PANJER, and WILLMOT · Loss Models: From Data to Decisions, *Second Edition*
KLUGMAN, PANJER, and WILLMOT · Solutions Manual to Accompany Loss Models: From Data to Decisions, *Second Edition*
KOTZ, BALAKRISHNAN, and JOHNSON · Continuous Multivariate Distributions, Volume 1, *Second Edition*
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Volumes 1 to 9 with Index
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Supplement Volume
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 1
KOTZ, READ, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update Volume 2
KOVALENKO, KUZNETZOV, and PEGG · Mathematical Theory of Reliability of Time-Dependent Systems with Practical Applications
LACHIN · Biostatistical Methods: The Assessment of Relative Risks
LAD · Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction
LAMPERTI · Probability: A Survey of the Mathematical Theory, *Second Edition*
LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE · Case Studies in Biometry
LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*
LAWLESS · Statistical Models and Methods for Lifetime Data, *Second Edition*
LAWSON · Statistical Methods in Spatial Epidemiology
LE · Applied Categorical Data Analysis
LE · Applied Survival Analysis
LEE and WANG · Statistical Methods for Survival Data Analysis, *Third Edition*
LEPAGE and BILLARD · Exploring the Limits of Bootstrap
LEYLAND and GOLDSTEIN (editors) · Multilevel Modelling of Health Statistics
LIAO · Statistical Group Comparison
LINDVALL · Lectures on the Coupling Method
LINHART and ZUCCHINI · Model Selection
LITTLE and RUBIN · Statistical Analysis with Missing Data, *Second Edition*
LLOYD · The Statistical Analysis of Categorical Data
MAGNUS and NEUDECKER · Matrix Differential Calculus with Applications in Statistics and Econometrics, *Revised Edition*
MALLER and ZHOU · Survival Analysis with Long Term Survivors
MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel
MANN, SCHAFER, and SINGPURWALLA · Methods for Statistical Analysis of Reliability and Life Data

*Now available in a lower priced paperback edition in the Wiley Classics Library.

MANTON, WOODBURY, and TOLLEY · Statistical Applications Using Fuzzy Sets
MARCHETTE · Random Graphs for Statistical Pattern Recognition
MARDIA and JUPP · Directional Statistics
MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with Applications to Engineering and Science, *Second Edition*
McCULLOCH and SEARLE · Generalized, Linear, and Mixed Models
McFADDEN · Management of Data in Clinical Trials
*McLACHLAN · Discriminant Analysis and Statistical Pattern Recognition
McLACHLAN, DO, and AMBROISE · Analyzing Microarray Gene Expression Data
McLACHLAN and KRISHNAN · The EM Algorithm and Extensions
McLACHLAN and PEEL · Finite Mixture Models
McNEIL · Epidemiological Research Methods
MEEKER and ESCOBAR · Statistical Methods for Reliability Data
MEERSCHAERT and SCHEFFLER · Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice
MICKEY, DUNN, and CLARK · Applied Statistics: Analysis of Variance and Regression, *Third Edition*
*MILLER · Survival Analysis, *Second Edition*
MONTGOMERY, PECK, and VINING · Introduction to Linear Regression Analysis, *Third Edition*
MORGENTHALER and TUKEY · Configural Polysampling: A Route to Practical Robustness
MUIRHEAD · Aspects of Multivariate Statistical Theory
MULLER and STOYAN · Comparison Methods for Stochastic Models and Risks
MURRAY · X-STAT 2.0 Statistical Experimentation, Design Data Analysis, and Nonlinear Optimization
MURTHY, XIE, and JIANG · Weibull Models
MYERS and MONTGOMERY · Response Surface Methodology: Process and Product Optimization Using Designed Experiments, *Second Edition*
MYERS, MONTGOMERY, and VINING · Generalized Linear Models. With Applications in Engineering and the Sciences
*NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
NELSON · Applied Life Data Analysis
NEWMAN · Biostatistical Methods in Epidemiology
OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences
OKABE, BOOTS, SUGIHARA, and CHIU · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams, *Second Edition*
OLIVER and SMITH · Influence Diagrams, Belief Nets and Decision Analysis
PALTA · Quantitative Methods in Population Health: Extensions of Ordinary Regressions
PANKRATZ · Forecasting with Dynamic Regression Models
PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
*PARZEN · Modern Probability Theory and Its Applications
PEÑA, TIAO, and TSAY · A Course in Time Series Analysis
PIANTADOSI · Clinical Trials: A Methodologic Perspective
PORT · Theoretical Probability for Applications
POURAHMADI · Foundations of Time Series Analysis and Prediction Theory
PRESS · Bayesian Statistics: Principles, Models, and Applications
PRESS · Subjective and Objective Bayesian Statistics, *Second Edition*
PRESS and TANUR · The Subjectivity of Scientists and the Bayesian Approach
PUKELSHEIM · Optimal Experimental Design
PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied Statistics
PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming
*RAO · Linear Statistical Inference and Its Applications, *Second Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

RAUSAND and HØYLAND · System Reliability Theory: Models, Statistical Methods, and Applications, *Second Edition*
RENCHER · Linear Models in Statistics
RENCHER · Methods of Multivariate Analysis, *Second Edition*
RENCHER · Multivariate Statistical Inference with Applications
*RIPLEY · Spatial Statistics
RIPLEY · Stochastic Simulation
ROBINSON · Practical Strategies for Experimenting
ROHATGI and SALEH · An Introduction to Probability and Statistics, *Second Edition*
ROLSKI, SCHMIDLI, SCHMIDT, and TEUGELS · Stochastic Processes for Insurance and Finance
ROSENBERGER and LACHIN · Randomization in Clinical Trials: Theory and Practice
ROSS · Introduction to Probability and Statistics for Engineers and Scientists
ROUSSEEUW and LEROY · Robust Regression and Outlier Detection
RUBIN · Multiple Imputation for Nonresponse in Surveys
RUBINSTEIN · Simulation and the Monte Carlo Method
RUBINSTEIN and MELAMED · Modern Simulation and Modeling
RYAN · Modern Regression Methods
RYAN · Statistical Methods for Quality Improvement, *Second Edition*
SALTELLI, CHAN, and SCOTT (editors) · Sensitivity Analysis
*SCHEFFE · The Analysis of Variance
SCHIMEK · Smoothing and Regression: Approaches, Computation, and Application
SCHOTT · Matrix Analysis for Statistics
SCHOUTENS · Levy Processes in Finance: Pricing Financial Derivatives
SCHUSS · Theory and Applications of Stochastic Differential Equations
SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization
*SEARLE · Linear Models
SEARLE · Linear Models for Unbalanced Data
SEARLE · Matrix Algebra Useful for Statistics
SEARLE, CASELLA, and McCULLOCH · Variance Components
SEARLE and WILLETT · Matrix Algebra for Applied Economics
SEBER and LEE · Linear Regression Analysis, *Second Edition*
*SEBER · Multivariate Observations
SEBER and WILD · Nonlinear Regression
SENNOTT · Stochastic Dynamic Programming and the Control of Queueing Systems
*SERFLING · Approximation Theorems of Mathematical Statistics
SHAFER and VOVK · Probability and Finance: It's Only a Game!
SILVAPULLE and SEN · Constrained Statistical Inference: Inequality, Order, and Shape Restrictions
SMALL and MCLEISH · Hilbert Space Methods in Probability and Statistical Inference
SRIVASTAVA · Methods of Multivariate Statistics
STAPLETON · Linear Statistical Models
STAUDTE and SHEATHER · Robust Estimation and Testing
STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications, *Second Edition*
STOYAN and STOYAN · Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics
STYAN · The Collected Papers of T. W. Anderson: 1943–1985
SUTTON, ABRAMS, JONES, SHELDON, and SONG · Methods for Meta-Analysis in Medical Research
TANAKA · Time Series Analysis: Nonstationary and Noninvertible Distribution Theory
THOMPSON · Empirical Model Building
THOMPSON · Sampling, *Second Edition*
THOMPSON · Simulation: A Modeler's Approach

*Now available in a lower priced paperback edition in the Wiley Classics Library.

THOMPSON and SEBER · Adaptive Sampling
THOMPSON, WILLIAMS, and FINDLAY · Models for Investors in Real World Markets
TIAO, BISGAARD, HILL, PEÑA, and STIGLER (editors) · Box on Quality and Discovery: with Design, Control, and Robustness
TIERNEY · LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics
TSAY · Analysis of Financial Time Series
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
VAN BELLE · Statistical Rules of Thumb
VAN BELLE, FISHER, HEAGERTY, and LUMLEY · Biostatistics: A Methodology for the Health Sciences, *Second Edition*
VESTRUP · The Theory of Measures and Integration
VIDAKOVIC · Statistical Modeling by Wavelets
VINOD and REAGLE · Preparing for the Worst: Incorporating Downside Risk in Stock Market Investments
WALLER and GOTWAY · Applied Spatial Statistics for Public Health Data
WEERAHANDI · Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models
WEISBERG · Applied Linear Regression, *Third Edition*
WELSH · Aspects of Statistical Inference
WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and Methods for *p*-Value Adjustment
WHITTAKER · Graphical Models in Applied Multivariate Statistics
WINKER · Optimization Heuristics in Economics: Applications of Threshold Accepting
WONNACOTT and WONNACOTT · Econometrics, *Second Edition*
WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
WOODWORTH · Biostatistics: A Bayesian Introduction
WOOLSON and CLARKE · Statistical Methods for the Analysis of Biomedical Data, *Second Edition*
WU and HAMADA · Experiments: Planning, Analysis, and Parameter Design Optimization
YANG · The Construction Theory of Denumerable Markov Processes
*ZELLNER · An Introduction to Bayesian Inference in Econometrics
ZHOU, OBUCHOWSKI, and MCCLISH · Statistical Methods in Diagnostic Medicine

*Now available in a lower priced paperback edition in the Wiley Classics Library.