Spring 2006

Modeling and simulation of value -at -risk in the financial market area

Xiangyin Zheng
Louisiana Tech University

MODELING AND SIMULATION OF VALUE-AT-RISK

IN THE FINANCIAL MARKET AREA

by

Xiangyin Zheng, M.S.

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

COLLEGE OF ENGINEERING AND SCIENCE
LOUISIANA TECH UNIVERSITY

May 2006

LOUISIANA TECH UNIVERSITY

THE GRADUATE SCHOOL

05/03/06
_____
Date

We hereby recommend that the dissertation prepared under our supervision

by_____XIANGYIN ZHENG_____

entitled_____MODELING AND SIMULATION OF VALUE-AT-RISK_____

_____IN THE FINANCIAL MARKET AREA_____

_____

be accepted in partial fulfillment of the requirements for the Degree of

_____DOCTOR OF PHILOSOPHY_____

_____
Supervisor of Dissertation Research

_____
Head of Department

CAM
_____
Department

Recommendation concurred in:

_____

_____   Advisory Committee

_____

Approved:
_____
Director of Graduate Studies

_____
Dean of the College

Approved:
_____
Dean of the Graduate School

ABSTRACT

Value-at-Risk (VaR) is a statistical approach to measure market risk. It is widely used by banks, securities firms, commodity and energy merchants, and other trading organizations. The main focus of this research is measuring and analyzing market risk by modeling and simulation of Value-at-Risk for portfolios in the financial market area. The objectives are (1) predicting possible future loss for a financial portfolio from VaR measurement, and (2) identifying how the distributions of the risk factors affect the distribution of the portfolio. Results from (1) and (2) provide valuable information for portfolio optimization and risk management.

The model systems chosen for this study are multi-factor models that relate risk factors to the portfolio's value. Regression analysis techniques are applied to derive linear and quadratic multifactor models for the assets in the portfolio. Time series models, such as ARIMA and state-space, are used to forecast the risk factors of the portfolio. The Monte Carlo simulation process is developed to comprehensively simulate the risk factors according to the four major distributions used to describe data in the financial market. These distributions are: multivariate normal, multivariate t, multivariate skew-normal, and multivariate skew t. The distribution of the portfolio is characterized by combining the multifactor models with the Monte Carlo simulation process. Based on the characterization of the portfolio distribution, any VaR measure of the portfolio can be calculated.
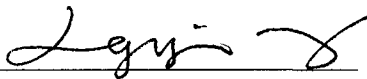
The results of the modeling and simulation show that (1) a portfolio may not have the same kind of distribution as the risk factors if the relationship between the portfolio and the risk factors is expressed as a quadratic function; (2) the normal distribution underestimates risk if the real data have a heavy tail and a high peak; and (3) diversification is the best strategy of investment since it reduces the VaR by combining assets together.

The computational approach developed in this dissertation can be used for any VaR measurement in any area as long as the relationship between an asset and risk factors can be modeled and the joint distribution of risk factors can be characterized.

APPROVAL FOR SCHOLARLY DISSEMINATION

The author grants to the Prescott Memorial Library of Louisiana Tech University the right to reproduce, by appropriate methods, upon request, any or all portions of this Dissertation. It is understood that "proper request" consists of the agreement, on the part of the requesting party, that said reproduction is for his personal use and that subsequent reproduction will not occur without written approval of the author of this Dissertation. Further, any portions of the Dissertation used in books, papers, and other works must be appropriately referenced to this Dissertation.

Finally, the author of this Dissertation reserves the right to publish freely, in the literature, at any time, any or all portions of this Dissertation.

Author _____

Date _____ 05/11/06 _____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and appreciation to my advisor, Professor Raja Nassar, for providing me the unique opportunity to work in the research area of applying statistics to financial market, and for his guidance, encouragement, patience, and support. As a statistician, he amazes me with his comprehensive statistical knowledge and unique insights into statistical problems. As an advisor, he encouraged and supported me in my research and taught me how to be a good scientific researcher. I appreciated all his time and advice.

I would also like to thank Dr. Weizhong Dai, Dr. Richard J. Greechie, and Dr. Andrei Paun, not only for their participation as Advisory Committee members but also for their advice and support.

I also want to thank all of the professors who have taught me in the past few years. I learned a lot from them. Although I do not mention their names here, I keep them in mind. I am very grateful for all of the support I obtained from all of them. I know it is only because of such support that I could stand higher and see further.

CHAPTER 1

INTRODUCTION

1.1 Background of Value-at-Risk (VaR)

Value-at-Risk (VaR) is a method of assessing risk that uses standard statistical techniques routinely used in other technical fields. It is one of the most important developments in risk management over the past 10 years. It implements a new class of risk measures that are specifically designed to measure and aggregate diverse risky positions across an entire institution using a common conceptual framework. VaR is generically defined as the worst expected loss for a given position or portfolio within a known confidence interval over a specific time horizon under normal market conditions. VaR provides users with a summary measure of market risk. It is truly a forward-looking risk measure. VaR has become the market standard for measuring, managing, and reporting market risk. Its influence has already spread into the domain of other types of financial risks, such as credit, insurance, operational, business volume, and behavioral risks.

The VaR revolution has been brought about by a convergence of factors [1]. These include (1) the pressure from regulators for better control of financial risks; (2) the globalization of financial markets, which has led to exposure to more sources of risk; and

1

(3) technological advances, which have made enterprise-wide risk management a not-so-distant reality.

1.1.1 Lessons from Financial Disasters

There have been a number of massive financial losses in finance history, including the following three famous ones.

(1) Orange County (December 1994): Orange County, California has an investment pool that supports various pension liabilities. The pool lost $1.7 billion through high-risk investments in derivatives. This loss was the largest ever recorded by a local government investment pool and led to the bankruptcy of the county on December 6, 1994 [2]. This loss was the result of unsupervised investment activity of Robert Citron, the county treasurer, who was entrusted with a $7.5 billion portfolio belonging to county schools, cities, special districts, and the county itself. Citron's mistake was to report his portfolio at cost instead of report the market value of the portfolio. Members of the board of supervisors claim that they did not receive critical information that would have indicated the risks that Citron was taking.

(2) Barings Bank (February 1995): Barings, Britain's oldest merchant Bank lost $1.4 billion because a Singapore-based trader, Nick Leeson, took unauthorized futures and options positions linked to the Nikkei 225 and Japanese government bonds (JGBs). The Barings' board and management claim to have been unaware of Leeson's activities. The disaster has revealed an amazing lack of control at Barings, and the collapse of Barings served as a wakeup call for financial institutions all over the world.

(3) Metallgesellschaft (December 1993): Metallgesellschaft AG was a German company that had over 20,000 employees and 10 billion US dollars in revenue. In 1993, the company lost over 1.4 billion dollars after speculating an increase in oil prices in the oil futures market. A subsequent drop in oil prices left the company buying the oil at a higher price than the market price. In the oil futures market, the spot price is normally greater than the futures price. When this occurs, the market is said to be in backwardation. When, however, the market shifts and futures prices are greater than the spot price, the market is said to be in contango. In the contango market, the spot decreased more than the futures prices. As long as the market stayed in contango, the company continued to lose money and experienced a massive cash flow crisis.

Barings, the famous investment bank, collapsed after risky derivatives trading due to the lack of effective control systems. The Orange County crisis in California was the result of poor risk measurement as well as ineffective communication of the risks involved to the investors. It has been widely reported in the press that the contango market was the key to Metallgesellschaft's downfall; however, the disaster in the oil markets should be seen as a reminder to the corporate community to understand the nature of their position in financial markets and to understand the consequences of market movements on their financial positions.

The main cause of these financial disasters is the lack of proper risk management. In virtually all cases management did not know what risks the institution was taking. The following quotation from the book about Barings applies as well to most of the other fiascoes [3]:

What Barings showed in the clearest terms was that the real issue was not derivatives in general or over-the-counter derivatives in particular. The real issue was about the quality of management and control required in modern financial markets.

1.1.2 Importance of Risk Management

The term "risk management" originated in the 1950s [4]. It had long been used to describe techniques for addressing property and casualty contingencies. However, it was not until 1990s, after a series of financial disasters, that financial institutions came to realize the importance of financial risk management as a discipline.

The new "risk management" that evolved during the 1990s had a new meaning -- the entire process of identifying, evaluating, controlling, and reviewing risks to make sure that the organization is exposed to only those risks that it needs to take to achieve its primary objectives. Risk cannot be eliminated. However, through good risk management, it can be (1) Transferred to another party, who is willing to take risk; (2) Reduced, by having good internal controls; (3) Avoided, by not entering into risky businesses; (4) Retained, to either avoid the cost of trying to reduce risk or anticipate higher profits by taking on more risk, and; (5) Shared, by following a middle path between retaining and transferring risk.

The benefits of risk management are [5] (1) Risk management helps to increase the value of the firm in the presence of bankruptcy costs, because it makes bankruptcy less possible. (2) The presence of informational asymmetries means that external finance is more costly than internal finance, and good investment opportunities can be lost. Risk management helps alleviate these problems by reducing the variability of the corporate

cash flow. (3) Risk management helps investors achieve a better allocation of risks, because financial institutions would typically have better access to capital markets.

1.1.3 Evolution of VaR

1.1.3.1 The difficulties of
  traditional risk measures

Traditional risk measures came from the trading floor originally and are consequently designed to make sense to the trading community. They reflect the way different financial instruments are traded. Taken one-by-one each risk measure is easy to understand. Since each product group may have one or more risk measures, this combination leads to a large number of risk measures being required to measure all the market risks a trading operation is exposed to.

Imagine the risk manager reporting the exposure of the bank to the chief executive officer (CEO) by providing the numbers associated with perhaps a dozen key risk measures. The CEO may have difficulty in assessing the overall riskiness of the bank's positions. He may ask, "How much money can I lose in total? Is the bank safe?" The risk manager armed only with traditional risk measures will have difficulty answering these questions.

In addition to the difficulty with aggregating risk across trading areas with sensitivity-based measures, it is also not possible to use traditional risk measures to compare the riskiness of one trading activity with another. Another significant problem with traditional risk measures is that they give no perception for the probability that the amount identified will be lost. It is left to the trader, or manager, to use his or her experience and judgment–something that gets more difficult for the senior management to understand.

One aspect of portfolio management that is just as important to senior management as to traders is diversification. Diversification is the extent to which risk is reduced in a portfolio by the investment being spread over a number of assets, or asset classes. Traditional risk measures neither describe nor quantify diversification within a portfolio.

There are many questions that traditional sensitivity-based risk measures can not easily answer [6]. Here are a few: How much could I lose on a normal day–and with what probability? How much could I lose in extreme circumstances if the stock market crashes? What is my overall exposure across all products, asset classes, and currencies? What diversification benefits does the bank obtain as a whole from having an equities and an interest rate area? Which trading area is taking the most risk? Do our existing limits allow the bank to take more risk than we can afford? Are our trading operations making sufficient return when compared to the risks they are taking? Value-at-Risk will provide us tools to answer these questions.

1.1.3.2 The importance of VaR

VaR can potentially have a significant business impact in three areas [7]:

(1) Risk Comparability: VaR defines a common metric that can be applied universally across all risk positions or portfolios: the maximum possible loss within a known confidence level over a given holding period. It allows the relative importance of each position or portfolio to be directly compared and aggregated. Besides being able to be applied universally across all risk categories, including market, credit, operational, and insurance risks, this metric is also expressed in units that are meaningful at all levels of

management: dollars. It therefore serves as a relevant focal point for discussing risks at all levels within the institution, creating a risk dialogue and culture that is otherwise difficult to achieve given the otherwise technical nature of the issues [1].

(2) Determinant of Capital Adequacy: VaR is calculated in currency units and is designed to cover most of the losses that a business risk might face. Therefore, it also has the intuitive interpretation as the amount of economic or equity capital that must be held to support that particular level of risky business activity. The philosophy that economically-determined VaR is the relevant measure for determining capital requirements for risk businesses is also being increasingly adopted by regulators and supervisors [7].

(3) Performance Measurement: The final important reason for calculating VaR is to help management to evaluate the performance of business units and strategies on a risk-adjusted basis. Given the interpretation of VaR as the minimum equity required to support a risky business, it is natural to use this measure to evaluate the relative performance of different businesses.

1.1.3.3 Banking regulatory initiatives for VaR

The landmark Basel Capital Accord of 1988 provided the first step toward tighter risk management. The so-called Basel Accord sets minimum capital requirements that must be met by commercial banks to guard against credit risks. This agreement was later amended to incorporate market risks. In this amendment, central bankers implicitly recognized that risk-management models in use by major banks were far more advanced than anything they could propose. Banks now have the option to use their own VaR risk

management model as the basis for required capital ratios. Thus VaR is being officially promoted as good risk-management practice.

1.1.3.4 A Brief history of VaR

Value-at-risk became popular with trading organizations during the 1990s. It was during this period that the name "Value-at-Risk" entered the financial dictionary. However, VaR measures had been used long before this.

An early user was Harry Markowitz. In his groundbreaking (1952) paper "Portfolio Selection" [8], he adopted a VaR metric of single period variance of return and used this to develop techniques of portfolio optimization. In the early 1980s, the United States Securities and Exchange Commission (SEC) adopted a crude VaR measure for use in assessing the capital adequacy of broker-dealer's trading non-exempt securities. A few years later, Bankers Trust implemented a VaR measure for use with its risk-adjusted return on capital (RAROC) allocation system. During the late 1980s and early 1990s, a number of institutions implemented VaR measures to support capital allocation or market risk limits.

In the early 1990s, three events popularized value-at-risk as a practical tool for use on trading floors:

(1) In 1993, the Group of 30 published a groundbreaking report on derivatives practices. It was influential and helped shape the emerging field of financial risk management. It promoted the use of value-at-risk by derivatives dealers and appears to be the first publication to use the phrase "value-at-risk."

(2) In 1994, J.P. Morgan launched its free RiskMetrics service. This was intended to promote the use of value-at-risk among the firm's institutional clients. The service comprised a technical document describing how to implement a VaR measure and a covariance matrix for several hundred key factors updated daily on the internet.

(3) In 1995, the Basel Committee on Banking Supervision implemented market risk capital requirements for banks. These were based upon a crude VaR measure, but the committee also approved, as an alternative, the use of banks' own proprietary VaR measures in certain circumstances.

These three initiatives came during a period of heightened concern about systemic risks due to the emerging—and largely unregulated—OTC derivatives market. It was also a period when a number of organizations—including Orange County, Barings Bank, and Metallgesellschaft—suffered staggering losses due to speculative trading, failed hedging programs or derivatives. Financial risk management was a priority for institutions, and Value-at-Risk was rapidly embraced as the tool of choice for quantifying market risk. It was implemented by financial firms, corporate treasuries, commodities merchants, and energy merchants.

1.2 Risk Management and VaR

1.2.1 Risk Management and Risk Measurement

Risk has two components: uncertainty and exposure to that uncertainty. To characterize the risk, we need to describe the uncertainty as well as the exposure to the uncertainty. The term "risk" is used in finance in two different but related ways: as the magnitude of (a) potential loss or (b) the standard deviation of the potential revenue of a

trading or investment portfolio over some period of time. The quantitative relationship between risk as potential loss and risk as uncertainty in future revenue is a function of the estimated probability distribution of future revenue.

There are many kinds of financial risks. Market risk is the risk that the value of an investment will decrease due to moves in market factors. Credit risk is the risk of financial loss suffered when a company that the bank has dealt with defaults [6]. Operational risk is a broad category of risk that can result from inadequate or failed internal processes, people and systems, or from external events [9]. This research focuses on market risk.

Risk management is the process of measuring, or assessing risk and then developing strategies to manage the risk. The objective of risk management is twofold: (1) to improve a bank's financial performance and (2) to ensure an institution does not suffer unacceptable losses. Therefore, risk management consists of the following basic activities [6]: (1) Understanding the risks being taken by an institution, (2) Measuring the risks, (3) Controlling the risks, and (4) Communicating the risks.

Financial risks can be quantified in a four step process [10]: (1) Define the risk to be measured; (2) Agree on a model for that risk; (3) Specify a risk measure that is compatible with that model; and (4) Estimate the value of that measure implied by the model. For example, the process might be as follows: (1) Risk: the market risk of a specified portfolio; (2) Risk model: market variables are assumed to be jointly normally distributed with specified volatilities and correlations; (3) Risk measure: one-day 90%

VaR; and (4) Risk estimate: being achieved with Monte Carlo simulation using 5,000 quasi-randomly generated scenarios.

<u>1.2.2 Risk Measurement before VaR</u>

<u>1.2.2.1 Scenario analysis</u>

Scenario analysis is a process of analyzing possible future events by considering alternative possible outcomes (scenarios). This sort of "what if" analysis is designed for better decision making since it allows more complete consideration of outcomes and their implications. In scenario analysis, we set out different scenarios and investigate what we stand to gain or loss under them. Scenario analyses can be more or less sophisticated, and early scenario analyses were inevitably crude given the limited computing power available. However, advances in computer technology have changed this. Networked computers and centralized databases make it easy to gather inputs. Large numbers of scenarios can be considered, and analyses of those scenarios can be elaborate.

Scenario analysis has several shortcomings. It only addresses risk due to the specific scenarios considered. Scenario analysis is highly dependent on assumptions. The output of scenario analysis tends to be cumbersome—multiple tables summarizing results in stead of a single number. Scenario analysis is not easy to carry out. A lot depends on our ability to identify the "right" scenarios, and there are relatively few rules to guide us when selecting them. Scenario analysis also tells us nothing about the likelihood of different scenarios, so we need to use our judgement when assessing the practical significance of different scenarios. Therefore, the results of scenario analyses are highly subjective and depend to a large extent on the skill of the analyst.

<u>1.2.2.2 Portfolio theory</u>

A different approach to risk measurement is provided by portfolio theory. In finance, a portfolio is a collection of investments held by an institution or a private individual. Portfolio theory starts from the premise that investors choose between portfolios on the basis of their expected return and the standard deviation (or variance) of their return. The standard deviation of the portfolio return can be regarded as a measure of the portfolio's risk. Other things being equal, an investor wants a portfolio whose return has a high expected return and a low standard deviation. These objectives imply that the investor should choose a portfolio that maximizes expected return for any given portfolio standard deviation. A portfolio that meets these conditions is efficient, and a rational investor will always choose an efficient portfolio. When faced with an investment decision, the investor must therefore determine the set of efficient portfolios and rule out the rest. Some efficient portfolios will have more risk than others, but the more risky ones will also have higher expected returns. Faced with the set of efficient portfolios, the investor chooses one particular portfolio on the basis of his or her own preferred trade-off between risk and expected returns. An investor who is very averse to risk will choose a safe portfolio with a low standard deviation and a low expected return, and an investor who is less risk averse will choose a riskier portfolio with a higher expected return.

One of the key insights of portfolio theory is that the risk of any individual asset is not the standard deviation of the return to that asset but the extent to which that asset contributes to overall portfolio risk. The risk in a portfolio of diverse individual stocks will be less than the risk inherent in holding any single one of the individual stocks

(provided the risks of the various stocks are not directly related). An asset might be very risky (i.e., have a high standard deviation) when considered on its own, and yet have a return that correlates with the returns to other assets in our portfolio in such a way that acquiring the new asset does not increase the overall portfolio standard deviation. The lower the correlation, other things being equal, the less the asset contributes to overall risk. If the correlation is negative, it will offset existing risks and lower the portfolio standard deviation.

Portfolio theory provides a broad context for understanding the interactions of systematic risk and reward. It has profoundly shaped how institutional portfolios are managed and motivated the use of passive investment management techniques. The mathematics of portfolio theory is used extensively in financial risk management and was a theoretical precursor for today's Value-at-Risk measures.

1.2.3 Basic Concepts of VaR

To better understand VaR, let's review the following descriptions from financial institutions and regulators [11]:

"Chase's two principle risk management tools are VaR and stress testing. VaR measures risk in an everyday environment, while stress testing measures market risk in an abnormal market environment. The VaR, a dollar amount, is a forward looking estimate of the potential for loss. The VaR looks forward one trading day, and is calculated as the loss level expected to be exceeded with a 1 in 100 chance."--The 1998 Chase annual report

"Value at risk is a measure of a point in the distribution of possible outcomes. It has two parameters: a horizon and a probability. For example, a common regulatory definition of VaR is the amount of capital that you should expect to lose no more than once in a hundred two-week intervals, given your current positions. At Goldman Sachs, we commonly focus on an amount of capital that we should expect to lose no more than once per year in a given day. We think of this not as a "worst case", but rather as a regularly occurring event with which we should be comfortable."--A 1996 Goldman Sachs research report

"The VaR measure represents an estimate of the amount by which an institution's position in a risk category could decline due to general market movements during a given holding period." -- A joint report of the Department of the Treasury, the Federal Reserve, and the Federal Deposit Insurance Corporation

The following definition of VaR, considered being more precise, is from J.P. Morgan's 1996 RiskMetrics Technical Document [12]: *"Value at risk is a measure of the maximum potential change in value of a portfolio of financial instruments over a pre-set horizon. VaR answers the question: how much can I lose with x% probability over a given horizon."*

In summary, VaR is a statistical measure of the risk that estimates the maximum loss that may be experienced on a portfolio with a given level of confidence. It is typically calculated for a one day time period – known as the holding period – and is often calculated with 95% confidence. Ninety-five percent confidence means that there is (on average) a 95% chance of the loss on the portfolio being lower than the VaR

calculated. Thus the typical definition of VaR becomes the maximum amount of money that may be lost on a portfolio in 24 hours, with 95% confidence.

VaR has three parameters: (1) The time horizon (period); i.e., the length of time over which we plan to hold the assets in the portfolio, also called the "holding period." The typical holding period is 1 day, or 1 week. (2) The confidence level at which we plan to make the estimate. Popular confidence levels usually are 99% and 95%. (3) The unit of the currency which will be used to denominate the value at risk, for example, dollars.

VaR has been defined as the loss (stated with a specified probability) from adverse market movements over a fixed time horizon, assuming the portfolio is not managed during this time. So VaR is measured as a lower percentile of a distribution for theorectical profit and loss that arises from possible movements of the market risk factors over a fixed risk horizon. To see this, first note that the loss (or profit) for a portfolio that is left unmanaged over a risk horizon of h days is

$$\Delta_h P_t = P_{t+h} - P_t \ .$$
(1.1)

In other words, $\Delta_h P_t$ is the forward-looking h-day theoretical (or 'unrealized') P&L, that is, the P&L obtained by simply marking the portfolio to market today and then leaving it unchanged and marking it to market again at the risk horizon. We do not know exactly how the underlying risk factors are going to move over the next h days, but we do have some ideas. For example, we might expect that historical volatilities and correlations would remain much the same. The possibilities for movements in risk factors can be summarized in a multivariate distribution, and this in turn will generate a distribution of

$\Delta_h P_t$, as each set of possible values for the risk factors at the risk horizon are entered into the pricing model for the portfolio, weighted by their joint probabilities.

The significance level of VaR, that is, the probability that is associated with a VaR measurement, corresponds to the frequency with which a given level of loss is expected to occur. Thus a 5% 1-day VaR corresponds to a loss level that one expects to exceed, in normal market circumstances, one day in 20. And a 1% 1-day VaR is the loss level that might be seen one day in 100. Now the definition of VaR above can be rephrased as follows: the 100α% h-day VaR is that number x such that the probability of losing x, or more, over the next h days equals 100α%. It can be written in mathematical terms as:

100α% h-day VaR is that number x such that Prob($\Delta_h P_t < -x$) = α.

Sometimes we use the notation $\mathrm{VaR}_{\alpha,h}$ to emphasize the dependency of the VaR measurement on the two parameters α, the significance level, and h, the holding period. Thus,

$$\mathrm{Prob}(\Delta_h P_t < -\mathrm{VaR}_{\alpha,h}) = \alpha \qquad (1.2)$$

is a mathematical statement that is equivalent to saying that the 100α% h-day VaR measurement x is the lower α quantile of the unrealized P&L distribution, as depicted in Figure 1.1.

If the shaded area is α
then the cut-off point is
$-VaR_{\alpha,h}$

$Prob(P_{t+h} - P_t < -VaR_{\alpha,h}) = \alpha$

$- VaR_{\alpha,h}$

$P_{t+h} - P_t$
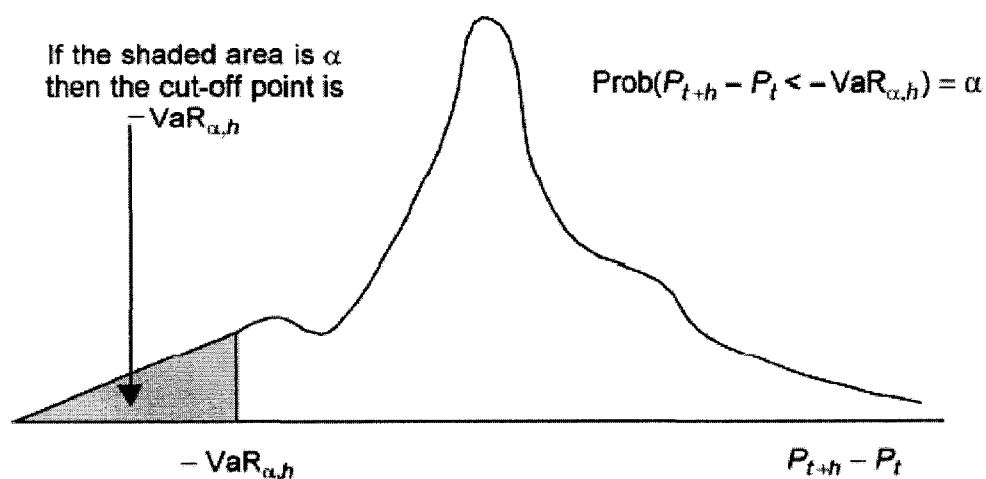
Figure 1.1   The P&L density and value-at-risk. [13]

We could measure value-at-risk as the standard deviation of portfolio value or the standard deviation of portfolio return. Essentially, any parameter of the distribution of a portfolio's future value can be used to measure value-at-risk.

1.2.4 Portfolio Theory and VaR

In some aspects VaR is a natural progression from earlier portfolio theory (PT). Yet there are also important differences between them [5]: (1) PT interprets risk in terms of the standard deviation of the return, while VaR approaches interpret it in terms of the maximum likely loss. The VaR notion of risk is easier to understand. (2) PT presupposes that P/L or returns are normally (or nearly normally) distributed, whereas VaR approaches can accommodate a very wide range of possible distributions. VaR approached are therefore much more flexible. (3) VaR approaches can be plausibly applied to a much broader range of risk problems: PT theory is limited to market risks, while VaR approaches can be applied much more to credit, liquidity, and other risks as well. (4) The variance-covariance approach to VaR has the same theoretical basis as PT – in fact, its theoretical basis is portfolio theory – but the other two main approaches to

VaR (i.e., the historical simulation and Monte Carlo simulation approaches) do not. VaR systems can be based on a wider range of estimation methods.

1.3 Overview of VaR Methodology

Basically, there are two types of methods: parametric and non-parametric. Parametric methods will include the variance-covariance approach and some analytical methods. The non-parametric model includes historical simulation and the Monte Carlo approach. All VaR measurement approaches use a similar scheme: (1) Selection of basic parameters (time horizon, confidence level, time of measurement); (2) Selection of relevant market factors; (3) Risk mapping; and (4) VaR calculation.

For step (1) we define the relevant parameters according to our goals and resources. The next two steps, (2) and (3), assume some kind of model, either just a set of relevant factors or a completely specified pricing model. In any case the relatively small set of relevant parameters should be defined, and some method for portfolio valuation based on this set should be established. Step (4) includes the calculation itself. This step can be very time consuming, especially when Monte Carlo methods are used. There are numerous techniques for speeding the calculation. The following are the different types of techniques to calculate VaR.

1.3.1 Historical Simulation

Historical simulation is probably the simplest non-parametric method. There is no assumption of a complex structure of the market. The historical simulation methodology repeatedly values current holdings based on the market conditions that existed over a specific historical period of time. The most useful version of this approach is illustrated when the risk mapping procedure defines the price of the whole portfolio as a

deterministic function of the market parameter $P(\vec{p})$. Here P is the pricing function and $\vec{p}$ is the vector of all relevant market parameters. Then today's (day t) price is $P(\vec{p}_t)$. The market parameters at some day j were $\vec{p}_j$ and on day $j+1$ the parameter was $\vec{p}_{j+1}$. Then we can model the possible changes in today's parameters in the following ways. We can use the relative change, where each market parameter is multiplied by the ratio of the same parameter at day $j+1$ and day j. Another approach is seen when we add today's value to the difference between the values at day $j+1$ and day j for each parameter. The multiplicative method is applicable when the volatility increases with the level of the parameter. This method is useful for stock indexes, exchange rates, etc. The additive approach assumes that the volatility is level independent. For example, for the additive approach we would take as a possible price tomorrow $P(\vec{p}_t + (\vec{p}_{j+1} - p_j))$. More complex combinations of both methods can be used as well [14], for example, the modeling of interest rates.

Using a moving window, we calculate the profits and loss for each h-day period. After ordering all the resulting data, we get the level of VaR at the 5% quantile of worst outcomes (assuming that we are working with a 95% confidence interval).

This approach has the advantage of being very intuitive. Unlike the parametric method, no assumption on the distribution of changes in market factors is required (the parametric method assumes normally distributed market returns) and, therefore, the historical simulation can better handle fat tails (kurtosis), i.e., extreme event risk, and asymmetric distributions (skewness), as are experienced in relatively illiquid markets such as emerging markets. Furthermore, the historical simulation methodology explicitly

understands the characteristics of instruments with non-linear behavior and analyzes based on historic market performance. Most sell-side organizations have moved to historical simulation because it has the advantage that in a complex organization each business can perform its own VaR calculation, and the results for each historical period (normally daily for one to three years) can be simply summed and the aggregate risk (with correlations embedded based on history) determined.

A typical problem with this approach is that there is not enough data. The further we go into the past for data, the less relevant this information is to today's market. This is not a simple trade-off. On the one hand, we would like to have more data in order to observe the rare events, especially the heavy loss. On the other hand, we do not want to build our current risk estimates on very old market data. Let's assume that we have agreed to take the last five years of data for our VaR estimate. If there is a big loss on a particular day, then exactly five years later the big jump will not appear in the set of data we use. This approach will lead to a jump in our VaR estimate from one day to the next, which demonstrates that the results are not stable when using the historical simulation.

1.3.2 Variance-Covariance Approach

The variance-covariance approach is a parametric method, based on the assumption that the returns are normally distributed (the underlying market factors have a multivariate normal distribution). Under this assumption, it is possible to determine the distribution of mark-to-market portfolio profits and losses, which is also normal. Historical data is used to measure the major parameters: means, standard deviations, correlations. The overall distribution of the market parameters is constructed from this data. Using the risk mapping technique, the distribution of the profits and losses over the

time horizon (typically one day) can be found. When the market value of the portfolio is a linear function of the underlying parameters, the distribution of the profits is normal as well. Therefore, the 5% quantile corresponding to VaR can be calculated at $1.65 \cdot \sigma$ below the mean ($2.33 \cdot \sigma$ will give the 1% level). One significant advantage of this scheme is that for many market parameters all of the relevant data is well known. The J. P. Morgan's RiskMetrics^{TM} is probably the best source for this type of data in many markets.

When VaR was first developed by the sell-side over a decade ago, the parametric approach was the standard because it was computationally extremely efficient. The efficiency results from the fact that this is an "analytic" approach, which directly calculates a solution, rather than the alternative approaches that determine a solution by iteratively simulating potential scenarios.

The strong side of this approach is that it is flexible, simple, and widely used. It also enables the addition of specific scenarios and enables the analysis of the sensitivity of the results with respect to the parameters. However, it relies heavily on the important assumption that all the major market parameters are normally distributed. In fact, historical distributions of market returns are far from being normal. Therefore when a significant portion of the portfolio is not linear (with options for example), this method can not be used directly.

1.3.3 Monte Carlo Simulation

The Monte Carlo simulation is another non-parametric method. It is probably one of the most popular methods among sophisticated users. It does not assume a specific form of the distributions. The first step is to identify the important market factors. Next,

one should build a joint distribution of these factors based on one of the following historical data: data implicitly implied by observed prices or data based on specific economic scenarios. Finally, the simulation is performed, typically with a large number of scenarios. The profit and losses at the end of the period are measured for each scenario. As in the other methods, these numbers should be ordered. The 5% quantile of the worst results is the VaR estimate.

This method has several important advantages. First, it does not assume a specific model and can be easily adjusted to economic forecasts. The results can be improved by taking a larger number of simulated scenarios. Options and other non-linear instruments can be easily included in a portfolio. In addition, one can track path-dependence because the whole market process is simulated rather than the final result alone.

Another advantage of the Monte Carlo method is that it allows the use of the preliminary results of all of the methods mentioned above. The historical simulations can give a first approximation to the distribution functions. The variance covariance shows which connections between variables are important and which can be neglected.

In addition, one can easily perform stress testing on the Monte Carlo simulation or perform a more detailed analysis of a specific set of scenarios, including dynamic strategies, such as prepayments or partial recoveries.

One important disadvantage of a Monte Carlo simulation is it converges very slowly. Any Monte Carlo type simulation converges to the true value as $\frac{1}{\sqrt{N}}$, where N is the total number of simulated trajectories. This result means that in order to increase the

precision by a factor of 10 one must perform 100 times more simulations. This problem is the most serious disadvantage of this method.

An additional problem with Monte Carlo simulations is that one needs to know the joint distribution of many market parameters. When there are more than three to four important parameters, it is not easy to clean all the data and to build this multidimensional distribution, especially true in cases where the variables are strongly correlated.

1.3.4 VaR Related Methods

1.3.4.1 Backtesting

The Basel standard requires backtesting–a procedure in which one checks (a posteriori) how often the actual losses have exceeded the level predicted by VaR. As soon as a 99% confidence interval and the 10 day time horizon are used, there should not be too many cases in which the actual losses are greater than the predicted ones.

There are three zones. If during the last year (approximately 250 business days) there are four or less exceptions (losses that exceed the VaR level), the model is said to be a green zone, and it is acceptable. If there are five to nine exceptions it is in the yellow zone and certain actions (such as increase of the safety multiplier from 3 to 3.65) are recommended [15]. When there are 10 or more exceptions the whole model should be revised. This mechanism prevents banks from setting the VaR too low.

1.3.4.2 Expected tail loss

Another VaR metric is expected tail loss (ETL), which is sometimes also known as expected shortfall, conditional VaR, worst conditional expectation, and various other names. The ETL is the expected value of our losses, L, if we get a loss in excess of VaR:

ETL = E[$L|L$ > VaR]. The VaR tells us the most we can expect to lose if a bad (i. e., tail) event does not occur, and the ETL tells us what we can expect to lose if a tail event does occur. For example, a 90% ETL VaR metric indicates the expected loss conditional on that loss exceeding its own .90-quantile.

1.3.4.3 Stress testing

If VaR covers so called "normal" market behavior, then the impact of extreme price changes can be ascertained by using stress testing. Stress testing consisting of applying pre-determined price changes to the assets making up a portfolio and working out the value change of the portfolio as a result of these price changes [6].

The main job of stress testing is to identify scenarios which cause the bank a significant loss. The job of the risk manager is then to get senior and trading management to think about these scenarios and to decide whether they are willing to accept the level of risk implied by the losses that would result from the scenarios identified given current market conditions.

1.3.5 An Assessment of VaR

The major advantages of VaR is that it is a risk measure that can be applied to all traded products, it can be used to compare the market risks of all types of activities in the firm, and it provides a single number that is easily understood by senior management. It takes into account the correlation and cross-hedging between various asset categories or risk factors, and it can be calculated according to a number of different methods. VaR may also take into account specific risks by including individual equities among risk factors [13]. Therefore, it is a standard risk benchmark which allows the risk being taken by different trading areas to be compared directly. As VaR can be used to measure the

risk on any product, it can be combined across different trading areas to give a single figure for the risk being taken by all trading areas combined [6].

There are two other things that VaR can do that traditional risk measures cannot. The first is that VaR gives an estimate of the likelihood of a loss greater than a given figure occurring; i.e., VaR has a probability associated with it. The second thing is that VaR takes into account how price changes of different assets are related to each other. This analysis allows the reduction in risk through diversification (i.e. holding positions in a number of assets) to be measured.

However, VaR only effectively measures market risk when the market is behaving "normally." This means that VaR is a measure of the day-to-day, or "business-as-usual" risk on the portfolio, with a given level of confidence. VaR does not deal adequately with the fairly frequent extreme price moves observed in the financial markets. Therefore, VaR must be combined with stress testing to provide a more comprehensive market risk measurement framework.

<u>1.3.6 Current Applications of VaR</u>

Currently, experience is of a revolution in risk management to become more quantitative in its approach to all risks. The new, more constraining, regulatory environment has prompted the rapid development of new methods for measuring and modeling financial risk. Financial institutions are setting new standards for risk control that require better pricing models and more stringent validation of all trading models. And recent changes in the rules for calculating risk capital charges have promoted the development of new risk systems, from data management to the internal models for measuring market and credit risk capital [16].

VaR has emerged as a major tool for measuring market risk, and it is used internally by banks for risk management and as a regulatory tool for ensuring the soundness of the financial system. A large amount of research work into VaR has emerged, and various aspects of VaR have been extensively documented. There are two areas of VaR-related research that have been relatively neglected: the relationship of VaR to statistical theory and the financial-economic foundation of VaR. Most VaR methods are based on normality, however, as stated by Alan Greenspan in 1997 [17], "the biggest problems we now have with the whole evaluation of risk is the fat-tailed problem, which is really creating very large conceptual difficulties."

1.4 The Objectives of this Dissertation

Most of the research in financial area has been based on the assumption: risk factors are multivariate normal distribution. Few mentioned about multivariate t-distribution which describes fat-tails. In the real world, the distribution of financial data is usually skewed and fat-tailed. Therefore, multivariate skew-normal or multivariate skew t-distribution could be a better choice for the real data. This research will characterize risk factors by using four kinds of distributions: multivariate normal distribution, multivariate skew-normal distribution, multivariate t-distribution, and multivariate skew t-distribution. The purpose is to get a deeper understanding of how the distributions of the risk factors affect the distribution of the portfolio. Based on this, the distribution of the portfolio can be more accurately characterized, and VaR measures can be more accurately calculated. In addition, by comparing the results of the simulation of

different distributions, we can get enough information to select what distribution to use in practical application.

1.5 Organization of this Dissertation

This dissertation combines statistical and mathematical modeling, Monte Carlo simulation, time series analysis and forecasting, regression analysis and the knowledge of risk management and Value-at-Risk methodology to measure and analyze market risk for a financial portfolio (a collection of investments held by an institution or an individual), to predict future loss for the portfolio, and to identify how the possible changes (distributions) of the risk factors affect the possible change (distribution) of the portfolio. The results of the modeling and simulation in this research can provide valuable information for portfolio optimization and risk management.

There are six chapters in this dissertation. Chapter 1, the current chapter, is an introduction to this research covering the background and basic concepts of risk management and VaR measurement. Chapter 2 is a detailed description of VaR methodologies and related models. Chapter 3 is a discussion of time series methods used in VaR inference procedure. Time series models, such as ARIMA and state-space, will be used to forecast the risk factors of the portfolio. Chapter 4 is a discussion of regression analysis techniques and VaR mapping procedure. A detailed step-by-step procedure will be developed to derive linear and quadratic multi-factor models for the portfolio. Chapter 5 is a discussion of Monte Carlo simulation and VaR transformation procedure. Four different multivariate distributions (multivariate normal distributions, multivariate t-distributions, multivariate skew-normal distributions and multivariate skew

t-distributions) will be extensively discussed and simulated. The simulation method will be applied to the linear and quadratic multifactor models to characterize the distribution of the portfolio and to calculate any VaR measures of the portfolio. Chapter 6 concludes the dissertation and addresses some areas for further study.

Through the whole dissertation, SAS software package is used for statistical analysis and modeling. All the Monte Carlo simulation processes are written in C language and the source codes are in the Appendix.

CHAPTER 2

VALUE-AT-RISK METHODOLOGIES AND

RELATED MODELS

2.1 Introduction

There are three sections in this chapter. The current section serves as an introduction. The second section describes three main procedures in VaR methodology: mapping procedure, inference procedure, and transformation procedure. The third section discusses the common VaR calculation models.

In economics and finance, the Value-at-Risk, or VaR, is a measure used to estimate how the value of an asset or of a portfolio of assets will decrease over a certain time period (usually over one day or 10 days) under usual conditions. VaR is based on the probability distribution for a portfolio's market value. It is typically used by securities houses or investment banks to measure the market risk or volatility risk of their asset portfolios, but it is actually a very general concept that has broad applications. VaR has three parameters: (1) time horizon: the length of time over which the assets in the portfolio will be held, also called holding period or forcast horizon; (2) the confidence level at which we plan to make the estimate and (3) base currency. The typical holding period is one day for a stock fund and one month for a mortgage company. For some

29

problems, even a holding period of one year is appropriate. Popular confidence levels usually are 99% and 95%. Table 2.1 gives explanations for the three parameters of VaR.

Table 2.1 Three parameters of Value-at-Risk

Confidence level	Probability of loss associated with VaR measurement.
Forecast horizon	Financial institutions (e.g., banks, hedge funds) consistently use a 1-day forecast horizon for VaR analysis of all market risk positions. Investment managers often use a 1-month forecast window, while corporations may apply quarterly or even annual projections of risk.
Base currency	The base currency for calculating VaR is typically the currency of equity capital and reporting currency of a company. For example, Bank of America would use USD to calculate and report its worldwide risks, while the United Bank of Switzerland would use Swiss francs.

Measure time in trading days. Let 0 be the current time and a portfolio's current market value ^{0}p. Its market value ^{1}P in one trading day is unknown. It is a random variable. The target is to ascribe to ^{1}P a probability distribution at time 1. That is to say, to calculate VaR, we need to characterize the distribution of ^{1}P at time 1 conditional on information available at time 0.

Suppose the portfolio has holdings ω_i in m assets. The assets' accumulated market values at time 1 are random variables, which are denoted as $^{1}S_i$. Then

$$^{1}P = \omega_1{}^{1}S_1 + \omega_2{}^{1}S_2 + \ldots + \omega_m{}^{1}S_m \tag{2.1}$$

A more manageable approach may be to model the portfolio's behavior, not in terms of individual assets, but in terms of specific risk factors. Depending upon the

composition of the portfolio, risk factors might include exchange rates, interest rates, commodity prices, spreads, and implied volatilities. These modeled risk factors are called key factors. Denote their values at time 1 as ^1R_i. The key factors comprise an ordered set (or "vector"), which are called the key vector and are denoted as $^1\vec{R}$:

$$^1\vec{R} = \begin{pmatrix} ^1R_1 \\ ^1R_2 \\ ^1R_3 \\ \vdots \\ ^1R_n \end{pmatrix} \sim \begin{pmatrix} first - key - factor \\ second - key - factor \\ third - key - factor \\ \vdots \\ n^{th} - key - factor \end{pmatrix} \tag{2.2}$$

Generally, the number n of key factors need to model will be substantially less than the number m of assets held by the portfolio. That is $n < m$. The advantages of risk-factor modeling and the criteria of risk factors selection were extensively discussed in [18], [19], and [20].

Selecting which key factors to model is as simple—or complex!—as choosing a set of market variables such that a pricing formula φ_i for each asset ^1S_i held by the portfolio can be expressed in terms of those variables. That is, for each asset, there must be a valuation function φ_i such that:

$$^1S_i = \varphi_i(^1\vec{R}). \tag{2.3}$$

Because the value of the portfolio ^1P is a linear polynomial of the asset values ^1R_i, ^1P can be expressed in terms of the key factors:

$$^1P = \sum_{i=1}^{m} \omega_i \, {}^1S_i = \sum_{i=1}^{m} \omega_i \varphi_i ({}^1R) \tag{2.4}$$

This formula illustrates a functional relationship that specifies the portfolio's market value ^1P in terms of the key factors ^1R_i. Shorthand notation for the relationship is

$$^1P = \theta({}^1\vec{R}) \tag{2.5}$$

Relationship (2.5) is called a portfolio mapping. The function θ is called the portfolio mapping function.

To characterize the entire distribution of ^1P, we need to apply the portfolio mapping function θ to the entire joint distribution of $^1\vec{R}$ to obtain the entire distribution of ^1P.

Suppose a portfolio were to remain untraded from the current time 0 to some future time 1. The portfolio's market value ^0p at the start of the period is known. Its market value ^1P at the end of the period is unknown. It is a random variable, which may be assigned a probability distribution conditional upon information available at time 0. The portfolio's market risk might be quantified with some real-valued parameter of that conditional distribution.

A VaR metric is a real-valued function of (1) the distribution of ^1P conditional on information available at time 0, and (2) the portfolio's current value ^0p. For example:

(1) $^0std({}^1P)$—Standard deviation of ^1P, conditional on information available at time 0;

(2) $^0std({}^1Z)$—Conditional standard deviation of a portfolio's simple return $^0std({}^1Z)$;

$$
{}^{0}std({}^{1}Z) = {}^{0}std\left(\frac{{}^{1}P - {}^{0}p}{{}^{0}p}\right) = \frac{{}^{0}std({}^{1}P - {}^{0}p)}{{}^{0}p} = \frac{{}^{0}std({}^{1}P)}{{}^{0}p}. \tag{2.6}
$$

If portfolio loss is defined as ${}^{1}L = {}^{0}p - {}^{1}P$, then the conditional standard deviation of ${}^{1}L$ is also a VaR metric: ${}^{0}std({}^{1}L) = {}^{0}std({}^{0}p - {}^{1}P) = {}^{0}std({}^{1}P)$.

Quantiles of portfolio loss make intuitively appealing VaR metrics. An expected tail loss VaR metric indicates a portfolio's expected loss conditional on that loss exceeding some specified quantile of loss.

To fully specify a VaR metric, we must indicate three things: (1) horizon – the time period; (2) function – the function of ${}^{0}p$ and the conditional distribution of ${}^{1}P$; (3) currency – the currency in which ${}^{0}p$ and ${}^{1}P$ are denominated.

A VaR measure is just an operation–some set of computations–designed to support a VaR metric. To design a VaR measure, we generally have some financial model in mind. Models take many forms, embracing certain assumptions and drawing on fields such as portfolio theory, financial engineering, or time series analysis. Such models are the assumptions and logic that motivate a VaR measure. They are called VaR models. To use a VaR measure, we must implement it. We must get necessary inputs, code the measure as software, and run on computers. The result is a VaR implementation.

A risk factor is any random variable ${}^{1}Q_{i}$ whose value will be realized during the interval $(0,1]$ and will affect the market value of a portfolio at time 1. A risk vector ${}^{1}\vec{Q}$ is a random vector of risk factors.

One particular risk factor and two risk vectors play important roles in VaR measures. These are (1) the portfolio's future value ^1P, (2) the asset vector $^1\vec{S}$, and (3) the key vector $^1\vec{R}$. The portfolio's future value ^1P represents the market value at time 1 of the portfolio for which VaR is to be measured. Asset vector $^1\vec{S}$ has asset values ^1S_i as components. These represent accumulated values of specific assets that may make up a portfolio. Accumulated value is denominated in the base currency employed by the VaR metric.

Mathematically, we define a portfolio as a pair ($^0p, ^1P$), where the constant ^0p is the portfolio's current value, and the random variable ^1P is the portfolio's future value. Similarly, we define an asset as a pair $\left(^0s_i, ^1S_i\right)$, where ^0s_i is the asset's current value, and ^1S_i is the asset's future value.

Every VaR measure must directly characterize a conditional probability distribution for some vector of risk factors, such as prices, interest rates, spreads, or implied volatilities. Those risk factors ^1R_i are called key factors.

An inference procedure characterizes a conditional distribution for $^1\vec{R}$. It is complete if it fully specifies a conditional distribution for $^1\vec{R}$. Otherwise it is incomplete. In practice, techniques of time series analysis are employed–in conjunction with financial theory – to obtain a reasonable characterization.

A mapping procedure characterizes a portfolio's exposures by expressing the portfolio's value ^1P as a function of risk factors $^1\vec{R}$, as shown in Equation (2.5).

A transformation procedure characterizes a conditional distribution for ^1P and uses this characterization to value a desired VaR metric. It is complete if its characterization of the conditional distribution for ^1P is sufficiently general to support any practical VaR metric.

Figure 2.1 illustrates how VaR measures work.

2.2 Value-at-Risk Procedures

All practical VaR measures accept portfolio data and historical market data as inputs. They process these with a mapping procedure, inference procedure, and transformation procedure. Output comprises the value of a VaR metric. That value is the VaR measurement. According to the schematic in Figure 2.1, any practical VaR measure must include these three procedures.

2.2.1 Mapping Procedure

The purpose of a mapping procedure is to characterize a portfolio's exposures. It does so by expressing the portfolio's value as a function of applicable market variables, such as stock prices, exchange rates, commodity prices, or interest rates. In the context of VaR, the word "mapping" is reserved for functions relating specific risk vectors to one another.

A portfolio mapping is a mapping that defines a portfolio's value ^1P as a function of some risk vector $^1\vec{R}$ as in Equation (2.5).
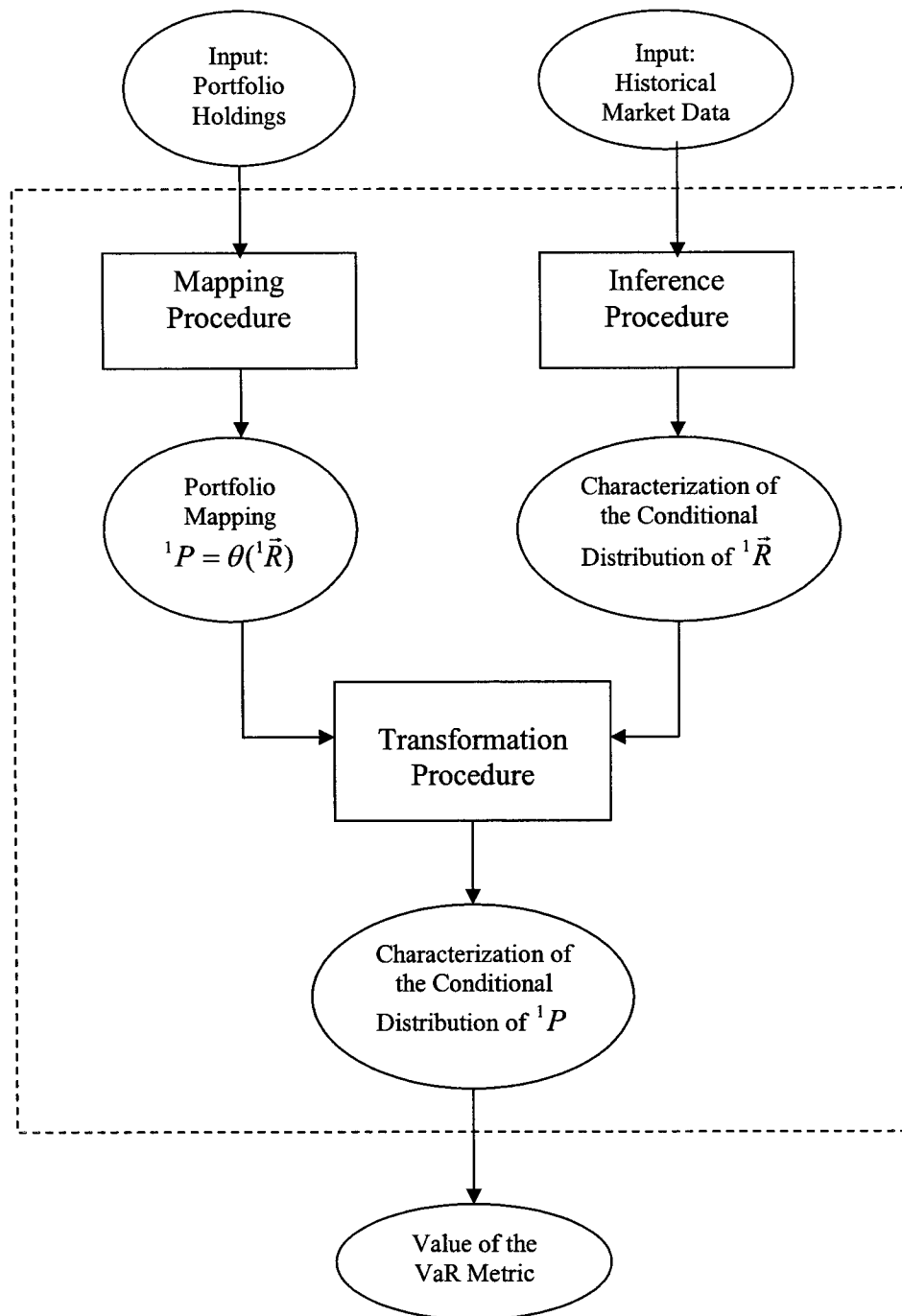
Figure 2.1 Schematic of how VaR measures work [21]

To calculate a portfolio's VaR, we must calculate the value of some function – VaR metric – of ^0p and the conditional distribution of ^1P. ^1P is the portfolio's market value at time 1. Mathematically, there are two ways we may define the random variable ^1P:

1. Directly specify a conditional distribution for ^1P.

2. Define ^1P as a function of some random vector.

The first approach is hardly feasible. Portfolios and financial markets tend to be complicated, so it is difficult to directly specify a conditional distribution for ^1P. We choose to define ^1P using the second approach – which leads to portfolio mapping. VaR measures define ^1P as a function of some asset vector $^1\vec{S}$:

$$^1P = \vec{\omega}^1\vec{S}.$$

(2.7)

$^1\vec{S}$ is interpreted as a vector of accumulated values. To complete our definition of ^1P, we must mathematically define $^1\vec{S}$. As with ^1P, there are two ways to define $^1\vec{S}$:

1. Directly specify a conditional distribution for $^1\vec{S}$.

2. Define $^1\vec{S}$ as a function of some other random vector.

According to Equation (2.1), the portfolio's value ^1P in terms of the asset vector $^1\vec{S}$ is

$$^1P = \omega_1{}^1S_1 + \omega_2{}^1S_2 + \ldots + \omega_m{}^1S_m.$$

(2.8)

The mapping for the first approach is represented schematically as

$$^1P \xleftarrow{\vec{\omega}} {}^1\vec{S}.$$

(2.9)

Rather than directly specify a joint distribution for $^1\vec{S}$, we define asset values ^1S_i as functions of $^1\vec{R}$ as shown by Equation (2.3). Combining Equations (2.3) and (2.10), we get Equation (2.5).

The mapping for the second approach is represented schematically as

$$
\overbrace{{}^{1}P \xleftarrow{\vec{\omega}} {}^{1}\vec{S} \xleftarrow{\vec{\varphi}} {}^{1}\vec{R}}^{\theta} \tag{2.10}
$$

Portfolio mappings constructed in this manner—starting with asset vector ${}^{1}\vec{S}$ and holdings $\vec{\omega}$, and perhaps mapping ${}^{1}\vec{S}$ to some key factors ${}^{1}\vec{R}$-are called primary mappings. The mapping function $\vec{\varphi}$ is a vector of component function φ_{i}, each of which values some asset according to Equation (2.3).

From an implementation standpoint, the mapping $\vec{\varphi}$ corresponds to a library of financial engineering models – a model library. The model library that defines $\vec{\varphi}$ requires various inputs in order to value assets, and these are the key factors ${}^{1}R_{i}$. It is desirable that key factors be linearly independent to avoid ${}^{1}\vec{R}$ having a singular covariance matrix. To avoid multicollinearity, it is also desirable to avoid key factors that are highly correlated or are in some other sense "almost" linearly dependent. It is desirable that the dimensionality of ${}^{1}\vec{R}$ not be too great.

For many VaR measures, output of the mapping procedure is a primary mapping, but not for all. Use of primary mappings can pose certain problems. Applying a transformation procedure to a complicated portfolio mapping can be computationally expensive. For this reason, many mapping procedures replace primary mappings with simpler approximations. Those approximations are called portfolio remappings.

Formally, a remapping is an approximation of a risk vector $^1\vec{R}$ with some other risk vector $^1\widetilde{\vec{R}}$. If we have a portfolio mapping $^1P = \theta(^1\vec{R})$, the remappings may take three forms:

(1) A function remapping approximates $^1P = \theta(^1\vec{R})$ by replacing θ with an approximate mapping function $\widetilde{\theta}$, so $^1\widetilde{P} = \widetilde{\theta}(^1\vec{R})$.

(2) A variable remapping approximates $^1P = \theta(^1\vec{R})$ by replacing $^1\vec{R}$ with alternative key vector $^1\widetilde{\vec{R}}$, so $^1\widetilde{P} = \theta(^1\widetilde{\vec{R}})$.

(3) A dual remapping approximates $^1P = \theta(^1\vec{R})$ by replacing both θ and $^1\vec{R}$, so $^1\widetilde{P} = \widetilde{\theta}(^1\widetilde{\vec{R}})$.

2.2.2 Inference Procedure

The purpose of an inference procedure is to characterize the joint probability distribution of the key vector $^1\vec{R}$ at time 1 conditional on information available at time 0. It generally accepts historical market data as an input and applies techniques of time series analysis to characterize the joint distribution conditional on information available at time 0. It is not always necessary to fully specify a distribution. Only the information sufficient to value our chosen VaR metric is required. Some inference procedure characterizes the conditional distribution of $^1\vec{R}$ with just a covariance matrix. An inference procedure is complete if it fully specifies a conditional distribution for $^1\vec{R}$. Otherwise, it is incomplete.

Inference procedures take various forms. One can simply make up a distribution suitable for a particular case. In practice, techniques of time series analysis are employed

-- in conjunction with financial theory--to obtain a reasonable characterization. The most common are those of uniformly-weighted moving averages (UWMA) and exponentially-weighted moving averages (EWMA). What are needed are time-series methods that can address conditional heteroskedasticity in high dimensions. While research is ongoing, such methods are not yet perfected. Even though there are many sophisticated techniques available to support mapping and transformation procedures, techniques for inference procedures are less developed. Researchers are studying ways to extend traditional methods of time series analysis to the needs of VaR measures, but techniques currently used are largely ad hoc [21]. In finance, a variety of models are used for conditionally heteroskedastic processes. These include autoregressive conditional heteroskedastic (ARCH) models, generalized ARCH (GARCH) models, regime-switching models, and stochastic volatility models. Chapter 3 will discuss inference procedure in detail.

2.2.3 Transformation Procedure

A transformation procedure combines the outputs from the mapping and inference procedures and uses them to characterize the distribution of ^1P, conditional on information available at time 0. Based on that characterization, and perhaps the portfolio's current value ^0p, the transformation procedure (or "transformation") represents risk with a characterization of the conditional distribution of ^1P. Using this conditional distribution, the value of the desired VaR metric can be determined.

Basically, there are four kinds of transformations:

- linear transformations,
- quadratic transformations,
- historical transformations, and

- Monte Carlo transformations.

Linear transformations are simple and run in real time. They apply only if a portfolio mapping function θ is a linear polynomial. Quadratic transformations are slightly more complicated but also run in real time (or near-real time). They apply only if a portfolio mapping function θ is a quadratic polynomial and $^1\vec{R}$ is multivariate normal. Monte Carlo and historical transformations are widely applicable but tend to run slowly (run times of an hour or more are common). Both employ the Monte Carlo method. Both generate a large number of realizations $^1\vec{r}^{[k]}$ for $^1\vec{R}$ and value ^1P for each. The histogram of realizations $^1p^{[k]}$ for ^1P provides a discrete approximation for the conditional distribution of ^1P. From this, any VaR metric can be calculated. Monte Carlo and historical transformations differ only in how they generate the realizations $^1\vec{r}^{[k]}$. Monte Carlo transformations generate them with pseudorandom number generators. Historical transformations draw them from historical market data.

Traditionally, VaR measures have been categorized according to the transformation procedures they employ:

(1) Linear VaR measures (other names include: parametric, variance-covariance, closed form, or delta normal VaR measures).

(2) Quadratic VaR measures (also called delta-gamma VaR measures).

(3) Historical VaR measures.

(4) Monte Carlo VaR measures.

2.3 Common Value-at-Risk Calculation Models

<u>2.3.1 Linear Value-at-Risk</u>

A portfolio is linear if its portfolio mapping function θ is a linear polynomial. Using matrix notation this means

$$^1P = \theta(^1\vec{R}) = \vec{b}\,^1\vec{R} + a \tag{2.11}$$

where $\vec{b}$ is a row vector and a is a scalar.

Linear transformations are based upon an important result from probability theory related to linear polynomials of random vectors. Let $^{1|0}\vec{\mu}$ and $^{1|0}\Sigma$ be the mean vector and covariance matrix of $^1\vec{R}$. (The superscripts $^{1|0}$ indicate that both parameters are for time 1, conditional on information available at time 0.) Let $^0E(^1P)$ and $^0Std(^1P)$ be the mean and standard deviation of ^1P conditional on information available at time 0. Then probability theory tells us that

$$^0E(^1P) = \vec{b}\,^{1/0}\vec{\mu} + a \tag{2.12}$$

$$^0Std(^1P) = \sqrt{\vec{b}\,^{1/0}\Sigma\vec{b}'} \tag{2.13}$$

where a prime $'$ indicates transposition. Note that these formulas are general. They require that ^1P be a linear polynomial of $^1\vec{R}$, but they make no assumptions about the distribution of $^1\vec{R}$. With these results, it is possible to value a variety of VaR metrics, including standard deviation of loss

$$^0Std(^1L) = {}^0Std(^0p - {}^1P) = {}^0Std(^1P), \tag{2.14}$$

and standard deviation of return

$$^0Std\left(\frac{^1P-^0p}{^0p}\right)=\frac{^0Std(^1P)}{^0p}.$$ (2.15)

However, results from Equations (2.14) and (2.15) are not sufficient to value a quantile of loss VaR metric on their own. Without additional information about a distribution, a mean and standard deviation do not determine the distribution's quantiles. A standard solution is to assume ^1P is normally distributed. Because a normal distribution is fully determined by its mean and standard deviation, this assumption, together with Equations (2.14) and (2.15), fully specifies the distribution of ^1P. It should be possible to value any VaR metric. For example, if our VaR metric is one-day 95% USD VaR, we can calculate VaR as

$$95\%VaR = 1.645\,^0Std(^1P) + (^0p - ^{1|0}\mu)$$ (2.16)

where $^{1|0}\mu = ^0E(^1P)$.

Equation (2.16) is based on the fact that the 5%-quantile of a normal distribution always occurs 1.645 standard deviations below its mean. See Figure 2.2 for illustration.
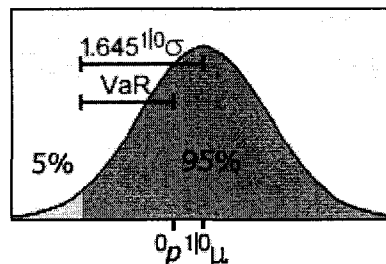
Figure 2.2 One-day 95% VaR [22]

Because the computations for a linear transformation are so modest, implementations typically run in real time. Simple linear VaR measures can even be implemented on a spreadsheet.

<u>2.3.2 Quadratic Value-at-Risk</u>

A portfolio is quadratic if its portfolio mapping function θ is a quadratic polynomial

$$^{1}P = \theta(^{1}\vec{R}) = ^{1}R'c^{1}\vec{R} + \vec{b}^{1}\vec{R} + a. \tag{2.17}$$

Here, c is a symmetric square matrix, $\vec{b}$ is a row vector and a is a scalar. If $^{1}\vec{R}$ is a joint normal random vector with mean vector $^{1|0}\vec{\mu}$ and covariance matrix $^{1|0}\Sigma$. Then by a change of variable technique, one obtains

$$^{1}\vec{R} = h^{-1}\,^{1}\dot{\vec{R}} + ^{1|0}\vec{\mu}, \tag{2.18}$$

where $h = uz^{-1}$, z is the Cholesky matrix of $^{1|0}\Sigma$, and u is a matrix whose rows are the orthonormal eigenvectors of $z'cz$. A new expression for ^{1}P is [21]:

$$^{1}P = ^{1}\dot{\vec{R}}'\dot{c}\,^{1}\dot{\vec{R}} + \dot{\vec{b}}\,^{1}\dot{\vec{R}} + \dot{a}, \tag{2.19}$$

where $\dot{c} = h'^{-1}ch^{-1}$, $\dot{\vec{b}} = (2^{1|0}\vec{\mu}'c + \vec{b})h^{-1}$, and $\dot{a} = ^{1|0}\vec{\mu}'c^{1|0}\vec{\mu} + \vec{b}^{1|0}\vec{\mu} + a$.

This change of variables achieves four conditions:

(1) $^{1}\dot{\vec{R}}$ is joint normal;

(2) the mean vector of $^{1}\dot{\vec{R}}$ is the $\vec{0}$ vector;

(3) the covariance matrix of $^{1}\dot{\vec{R}}$ is the identity matrix I;

(4) $\dot{c}$ is diagonal matrix.

The fourth item means that ^1P can be depend upon each of the variables $^1\dot{R}$ in one of the four ways:

1. No dependence: $\dot{c}_{i,i} = 0$ and $\dot{b}_i = 0$;

2. Linear dependence: $\dot{c}_{i,i} = 0$ and $\dot{b}_i \neq 0$;

3. Central quadratic dependence: $\dot{c}_{i,i} \neq 0$ and $\dot{b}_i = 0$; or

4. Non-central quadratic dependence: $\dot{c}_{i,i} \neq 0$ and $\dot{b}_i \neq 0$.

In the last case ^1P has a dependence of the form $\dot{c}_{i,i}\dot{R}_i^2 + \dot{b}_i{}^1\dot{R}_i$. Completing the squares, this becomes $\dot{c}_{i,i}({}^1\dot{R}_i + \dot{b}_i / 2\dot{c}_{i,i})^2$.

Consequently, ^1P is a linear polynomial of independent random variables, each of which is either standard normal, central chi-squared with one degree of freedom, or non-central chi-squared with one degree of freedom and non-centrality parameter $(\dot{b}_i / 2\dot{c}_{i,i})^2$.

Since a linear polynomial of independent normal random variables is itself normal, all normal terms can be combined into one. A general expression for ^1P is

$$^1P = \left(\sum_{k=1}^{m} \gamma_k X_k \right) + \beta X_0 + \alpha \qquad (2.20)$$

where the X_k are chi-squared with one degree of freedom and non-centrality parameter δ_k^2 and X_0 is standard normal.

Quantile-based VaR metrics are more difficult to calculate. Various solutions have been proposed [21], four of which are listed below.

2.3.2.1 Quantile of loss from fast
Fourier transformation

The inversion theorem of probability theory provides the following expression for the probability density function ϕ of a random variable in terms of its characteristic function Ψ,

$$\phi(x) = \int_{-\infty}^{\infty} \Psi(2\pi t)e^{-i2\pi xt} dt \qquad (2.21)$$

which is a Fourier transformation that can be approximated with the FFT.

The characteristic function of a random variable of the form (2.20) is

$$\Phi(^1 p) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{e^A \sin(B+C)}{D} dw , \qquad (2.22)$$

where

$$A = -\frac{w^2}{2}\left(\beta^2 + 4\sum_{k=1}^{m} \frac{\gamma_k^2 \delta_k^2}{1 + 4\gamma_k^2 w^2}\right), \qquad (2.23)$$

$$B = w\left(\alpha - ^1 p + \sum_{k=1}^{m} \frac{\gamma_k \delta_k^2}{1 + 4\gamma_k^2 w^2}\right), \qquad (2.24)$$

$$C = \frac{1}{2}\sum_{k=1}^{m} \tan^{-1}(2\gamma_k w) , \qquad (2.25)$$

$$D = w\left(\prod_{k=1}^{m}(1 + 4\gamma_k^2 w^2)\right)^{1/4} \qquad (2.26)$$

and $\tan^{-1}$ denotes the inverse tangent function with output in radians.

2.3.2.2 Quantile of loss with Johnson curves

When faced with a body of statistical data, researchers often try to fit some standard probability distribution to the data. For this purpose, various families of

probability distributions-called families of curves-have been defined. Johnson curves are constructed through translation of variables. A more general Johnson family of curves is defined with the family of translation functions comprising any of the forms

$$Z = \gamma + \delta \log\left(\frac{X - \xi}{\xi + \lambda - x}\right), \qquad \xi < x < \xi + \lambda \tag{2.27}$$

$$Z = \gamma + \delta \log\left(\frac{X - \xi}{\lambda}\right) \tag{2.28}$$

$$Z = \gamma + \delta \log\left(\frac{X - \xi}{\lambda} + \sqrt{\left(\frac{X - \xi}{\lambda}\right) + 1}\right) \tag{2.29}$$

where γ, δ, ξ and λ are parameters.

For normal distributions, given a quadratic portfolio of form (2.20),

$$Z = \varphi(^1 P) \tag{2.30}$$

where $Z \sim N(0,1)$ and φ is a translation function. Since translation functions are monotone, they are invertible. From Equation (2.30), one obtains the inverse

$$^1 P = \varphi^{-1}(Z). \tag{2.31}$$

Because φ is monotone, φ^{-1} is also monotone. Monotone functions map quantiles to quantiles, so you can calculate any quantile of $^1 P$ from the corresponding quantile of Z. For example, the 0.05-quantile of a standard normal random variable is -1.645. Accordingly, the 0.005-quantile of $^1 P$ is approximately $\varphi^{-1}(-1.645)$.

2.3.2.3 Cornish-Fisher expansion

The Cornish-Fisher approach constructs approximation to quantiles from estimates of skewness and kurtosis [23].

The cumulants of a random variable X are conceptually similar to its moments. They are defined, as those values κ_r such that the identity

$$\exp\left(\sum_{r=1}^{\infty}\frac{\kappa_r t^r}{r!}\right) = \sum_{r=0}^{\infty}\frac{E(X^r)t^r}{r!} \tag{2.32}$$

holds for all t. Cumulants of a random variable X can be expressed in terms of its mean $\mu = E(X)$ and central moments $\mu_r = E[(X-\mu)^r]$. Expressions for the first cumulants are

$$\kappa_1 = \mu, \tag{2.33}$$

$$\kappa_2 = \mu_2, \tag{2.34}$$

$$\kappa_3 = \mu_3, \tag{2.35}$$

$$\kappa_4 = \mu_4 - 3\mu_2^2, \tag{2.36}$$

$$\kappa_5 = \mu_5 - 10\mu_3\mu_2. \tag{2.37}$$

Suppose X has mean 0 and standard deviation 1. Cornish and Fisher [24] provide an expansion for approximating the q-quantile, $\Phi_X^{-1}(q)$, of X based upon its cumulants. Using the first five cumulants, the expansion is

$$\Phi_X^{-1}(q) \approx \Phi_Z^{-1}(q) + \frac{\Phi_Z^{-1}(q)^2 - 1}{6}\kappa_3 + \frac{\Phi_Z^{-1}(q)^3 - 3\Phi_X^{-1}(q)}{24}\kappa_4 - \frac{2\Phi_Z^{-1}(q)^3 - 5\Phi_Z^{-1}(q)}{36}\kappa_3^2$$

$$+ \frac{\Phi_Z^{-1}(q)^4 - 6\Phi_Z^{-1}(q)^2 + 3}{120}\kappa_5 - \frac{\Phi_Z^{-1}(q)^4 - 5\Phi_Z^{-1}(q)^2 + 2}{24}\kappa_3\kappa_4 + \frac{12\Phi_Z^{-1}(q)^4 - 53\Phi_Z^{-1}(q)^2 + 17}{324}\kappa_3^3$$

$$\tag{2.38}$$

where $\Phi_Z^{-1}(q)$ is the q-quantile of $Z \sim N(0,1)$. Although (2.38) applies only if X has mean 0 and standard deviation 1, we can still use it to approximate quantiles if X has some other mean μ and standard deviation σ. Simply define the normalized X as

$$X^* = \frac{X - \mu}{\sigma}, \tag{2.39}$$

where $\sigma = \sqrt{\mu_2}$ is the standard deviation of X. Apply the Cornish-Fisher expansion to

obtain the q-quantile x^* of X^*. The corresponding q-quantile x of X is then

$$x = x^* \sigma + \mu. \tag{2.40}$$

2.3.2.4 The inversion theorem

The inversion theory is primarily of theoretical interest. Practically, it can be used

for the purpose of evaluating the CDF of a linear polynomial of independent random

variables.

To define characteristic functions, we must extend the notion of random variables

into the complex plane. Let U_1 and U_2 be real random variables, and let $i = \sqrt{-1}$. Then

$$U = U_1 + iU_2 \tag{2.41}$$

is a complex random variable. Its expectation is defined as

$$E(U) = E(U_1) + iE(U_2). \tag{2.42}$$

The characteristic function of a random variable X is defined as

$$\Psi(w) = E(e^{iwX}), \tag{2.43}$$

where w is real and $i = \sqrt{-1}$. If X is continuous,

$$\Psi(w) = E(e^{iwX}) = \int_{-\infty}^{\infty} \phi(x) e^{iwx} dx. \tag{2.44}$$

If $\vec{X}$ is a random vector with independent components X_i, and Y is a linear polynomial of

$\vec{X}$, written as

$$Y = \vec{b}\vec{X} + a \tag{2.45}$$

with $\bar{b}$ is a real row vector and a is a scalar, then,

$$\Psi_Y(w) = e^{aiw}\Psi_{X_1}(b_1 w)\Psi_{X_2}(b_2 w)\cdots\Psi_{X_n}(b_n w).$$ (2.46)

Characteristic functions for $U(a,b), N(\mu,\sigma^2)$, and $\chi^2(v,\delta^2)$ random variables are,

respectively,

$$\Psi(w) = \frac{e^{ibw} - e^{iaw}}{iw(b-a)},$$ (2.47)

$$\Psi(w) = \exp\left(i\mu w - \frac{\sigma^2 w^2}{2}\right),$$ (2.48)

$$\Psi(w) = \frac{\exp\left[\delta^2 iw/(1-2iw)\right]}{(1-2iw)^{v/2}}.$$ (2.49)

Inversion theorem states that the CDF of a random variable is uniquely determined by its characteristic function [21]. If two random variables have the same characteristic function, they have the same CDF. An inversion theorem provides the CDF of a random variable X in terms of its characteristic function:

$$\Phi(x) = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{\Psi(-w)e^{ixw} - \Psi(w)e^{-ixw}}{iw}dw.$$ (2.50)

2.3.3 Historical Value-at-Risk

Historical simulation is the simplest and most transparent method of calculation. This involves running the current portfolio across a set of historical price changes to yield a distribution of changes in portfolio value, and computing a percentile (the VaR). The benefits of this method are it is simple to implement, and the fact that it does not assume a normal distribution of asset's returns. Drawbacks are the requirement for a large market database and the computationally intensive calculation.

2.3.3.1 Single instrument portfolios

Historical simulation can be described in terms of five steps to perform analysis on a single instrument portfolio:

1. Identify the basic market factors, and obtain a formula expressing the mark-to-market value of the portfolio in terms of the market factors.

2. Obtain historical values of the market factors for the last N periods.

3. Take into account the current portfolio to the changes in market rates and prices experienced on each of the most recent 100 business days; calculate the daily profits and losses that would occur if comparable daily changes in the market factors are experienced, and the current portfolio is mark-to-market.

4. Order the mark-to-market profits and losses from the largest profit to the largest loss.

5. Select the loss which is equaled or exceeded five percent of the time. This loss is the value at risk at 95% confidence level.

2.3.3.2 Multiple instrument portfolios

An extension of the above methodology to handle realistic, multiple instrument portfolios can be accomplished in three steps. First, market factors must be identified, and pricing formulas expressing the instruments' values in terms of the market factors must be obtained. Second, the historical values of all the market factors must be collected. Third, it is crucial that the mark-to-market profits and losses on every instrument be computed and then summed for each period, before they are ordered from lowest loss to highest profit.

2.3.4 Monte Carlo Value-at-Risk

Based on the characterization of the joint distribution of $^1\vec{R}$, a Monte Carlo transformation procedure generates several thousand pseudorandom realizations $^1r^{[k]}$ for $^1\vec{R}$, and then calculates the corresponding realizations, $^1p^{[k]} = \theta(^1r^{[k]})$. The histogram of these realizations $^1p^{[k]}$ provides a discrete approximation for the distribution of ^1P. Based on the histogram, any reasonable VaR can be determined.

The Monte Carlo simulation methodology has a number of similarities to historical simulation. The main difference is that rather than carrying out the simulation using the observed changes in the market factors over the last N periods to generate N hypothetical portfolio profits and losses, one chooses a statistical distribution that is believed to adequately capture or approximate the possible changes in the market factors. Then, a pseudo-random number generator is used to generate thousands or tens of thousands of hypothetical changes in the market factors. These are then used to construct thousands of hypothetical portfolio profits and losses on the current portfolio, and the distribution of possible portfolio profits or losses. Finally, the value at risk is determined from the distribution.

2.3.4.1 Single instrument portfolios

Similar to historical simulation, the Monte Carlo simulation can also be described in five steps:

1. Identify the basic market factors, and obtain a formula expressing the mark-to-market value of the portfolio in terms of the market factors.

2. Determine or assume a specific distribution for changes in the basic market factors, and estimate the parameters of that distribution. The ability to pick the

distribution is the feature that distinguishes the Monte Carlo simulation from the other approaches, for in the other methods the distribution of changes in the market factors is specified as part of the method. The designers of the risk management system are free to choose any distribution that they think reasonably describes possible future changes in the market factors.

3. Use a pseudo-random generator to generate N hypothetical values of changes in the market factors based on the selected distribution, where N is almost certainly greater than 1,000 and perhaps greater than 10,000. These hypothetical market factors are then used to calculate N hypothetical mark-to-market portfolio values. Then from each of the hypothetical portfolio values we subtract the actual mark-to-market portfolio value to obtain N hypothetical daily profits and losses.

Steps 4 and 5 are the same as in historical simulation. The mark-to-market profits and losses are ordered from the largest loss to the largest profit, and the value at risk is defined as the loss which is equaled or exceeded five percent of the time.

<u>2.3.4.2 Multiple instrument portfolios</u>

Just as with historical simulation, extending the methodology to handle realistic, multiple instrument portfolios requires only that a bit of additional work be performed in three steps. First, market factors must be identified, and pricing formulas expressing the instruments' values in terms of the market factors must be obtained. Second, the joint distribution of possible changes in the values of all the market factors must be determined. Third, similar to historical simulation, to reflect accurately the correlations of

market rates and prices it is necessary that the mark-to-market profits and losses on every instrument be computed and then summed for each period, before they are ordered from lowest loss to highest profit.

The magnitude of the standard error depends upon several factors, such as:

- the VaR metric

- the portfolio, and

- the sample size used in the Monte Carlo analysis.

A solution to this problem employs variance techniques, namely, control variates and stratified sampling. Both of these techniques employ a quadratic remapping $^1\tilde{P}$ for ^1P. The remapping $^1\tilde{P}$ does not replace ^1P. Instead, it is used to facilitate variance reduction so VaR can be more easily calculated for ^1P.

With the method of control variates, $^1\tilde{P}$ is used as a control variate for ^1P. Since it is easy to calculate the VaR of $^1\tilde{P}$ with a quadratic transformation procedure, variance reduction is excellent for most portfolios and VaR metrics.

With the method of stratified sampling, $^1\tilde{P}$ is used to construct stratification. The methodology varies depending upon the VaR metric. For a quantile of loss VaR metric, realizations $^1r^{[k]}$ of ^1R are stratified into two regions:

- one comprising realization $^1r^{[k]}$ such that $^1\tilde{p}^{[k]} = \tilde{\theta}(^1r^{[k]})$ exceeds the VaR of $^1\tilde{P}$, and

- the other comprises realizations $^1r^{[k]}$ such that $^1\tilde{p}^{[k]} = \tilde{\theta}(^1r^{[k]})$ is less than or equal to the VaR of $^1\tilde{P}$.

2.4 Comparison of Value-at-Risk Methodologies

Two tables are listed below to compare the above three VaR methodologies. Table 2.2 summarizes the three methodologies. Table 2.3 lists the advantages and disadvantages for each of them.

Table 2.2 Summary of three VaR methodologies

Methodology	Description	Applications
Parametric	Estimates VaR with equation that specifies parameters such as volatility, correlation, delta, and gamma.	Accurate for traditional assets and linear derivatives, but less accurate for nonlinear
Monte Carlo simulation	Estimates VaR by simulating random scenarios and revaluing positions in the portfolio.	Appropriate for all types of instruments, linear and nonlinear
Historical simulation	Estimates VaR by reliving history; takes actual historical rates and revalues positions for each change in the market.	

Table 2.3 Advantages and disadvantages of each methodology

Methodology	Advantage	Disadvantage
Parametric	• Fast and simple calculation • No need for extensive historical data (only volatility and correlation matrix are required)	• Less accurate for nonlinear portfolios, or for skewed distributions
Monte Carlo simulation	• Accurate* for all instruments • Provides a full distribution of potential portfolio values (not just a specific percentile) • Permits use of various distributional assumptions (normal, t-distribution, normal mixture, etc.), and therefore has potential to address the issue of fat tails (formally known as "leptokurtosis") • No need for extensive historical data	• Computationally intensive and time-consuming (involves revaluing the portfolio under each scenario) • Quantifies fat-tailed risk only if market scenarios are generated from the appropriate distributions
Historical simulation	• Accurate* for all instruments • Provides a full distribution of potential portfolio values (not just a specific percentile) • No need to make distributional assumptions (although parameter fitting may be performed on the resulting distribution) • Faster than Monte Carlo simulation because less scenarios are used to current conditions,	• Requires a significant amount of daily rate history (note, however, that sampling far back may be a problem when data is irrelevant to current conditions, e.g., currencies that have already devalued) • Difficult to scale far into the future (long horizons) • Coarse at high confidence levels (e.g., 99% and beyond) • Somewhat computationally intensive and time-consuming

CHAPTER 3

TIME-SERIES MODELS AND

INFERENCE PROCEDURE

In Chapter 2, we described three components of a Value-at-Risk measure: an inference procedure, a mapping procedure, and a transformation procedure. In this chapter, we will focus on the inference procedure. There are five sections of this chapter. The first section is an introduction to stochastic processes and time series analysis. The second section summarizes the five features of time series data. The third section is an overview about a variety of time series models. The fourth section is a description of the models used in this research. The last section is a discussion of results.

A model is a relationship between variables. It is a simplification of the reality. It is a construct to help us explain and better understand the system under study. In general all models have an information input, an information processor, and an output of expected results.

Modeling is a process to build a model [25]. "*Model building forces the individual to think clearly about, and account for, all the important interrelationships involved in a problem. The reliance on intuition can be dangerous at times because of the possibility that important relationships will be ignored or improperly used. In addition, it is important that individual relationships be tested or validated in some way or another.*"

In other words, modeling is more than the process of constructing a model. One needs to examine the system, select the important variables, find the relationships between the variables, build a model, solve the mathematical problems that are set up for the model, verify and validate the model, draw conclusions from the model, and forecast based on the model.

Models are of many types. A model could be a theory, a law, a hypothesis, an equation, or even a structured idea. A mathematical model is a model that can be described by a set of mathematical equations. A statistical model is a model that characterizes a system based upon its statistical parameters such as mean, mode, variance or regression coefficients. Statistical models are useful in helping identify patterns and underlying relationships between data sets. A time series model, a subcategory of statistical models, is one that postulates a relationship among a number of temporal sequences or time series – a sequence of observations ordered in time.

Inference means (1) the process of deriving logical conclusions from a hypothesis known or assumed to be true, or (2) the act of reasoning from fact-based knowledge or evidence. In time series models, we assume that we know nothing about the real world cause-and-effect relationships that affect the variable we are trying to forecast. Instead, we study and analyze the past behavior of a time series in order to infer something about its probable future behavior. In VaR, the traditional methods of time series analysis have been applied and extended to inference procedures. The research in this area is still ongoing.

Figure 3.1, a reproduction of Figure 2.1, highlights the position of inference procedure.
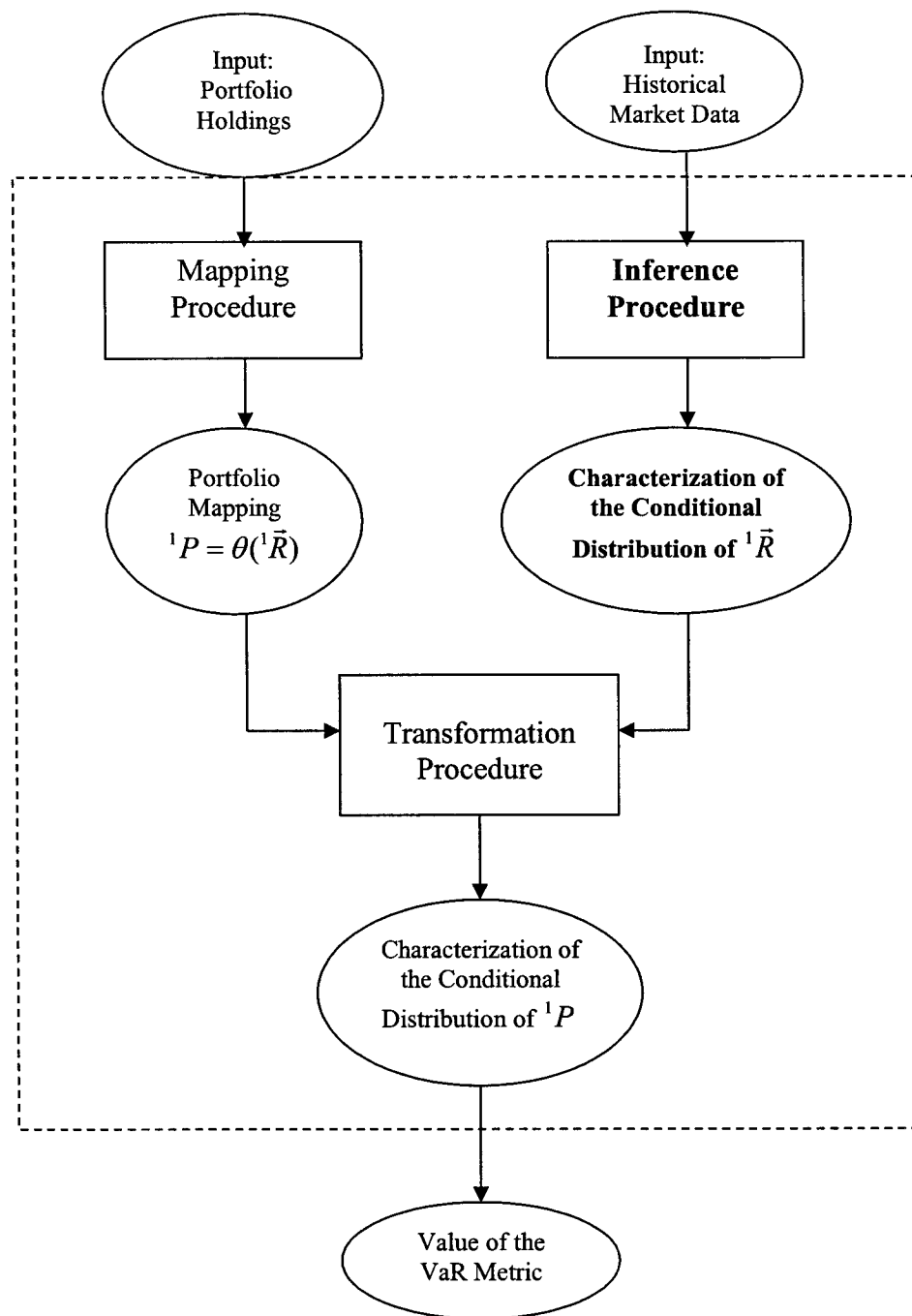
Figure 3.1 A reproduction of Figure 2.1 [21]

3.1 Introduction to Stochastic Processes and Time-Series Analysis

3.1.1 Random Variables and Probability Distributions

3.1.1.1 Random variables

A random variable is itself a function of a statistical experiment in which each outcome has a definite probability of occurrence. The value of the random variable will vary from trial to trial as the experiment is repeated. For example, the outcome when a coin is tossed can be "heads" or "tails". There are two types of random variable-discrete and continuous. A random variable has either an associated probability distribution (discrete random variable) or probability density function (continuous random variable).

3.1.1.2 Expected value

The expected value of a random variable X indicates its average value and is symbolized by $E(X)$ or μ.

If X is a discrete random variable with possible values x_1, x_2, x_3, ..., x_n, and $p(x_i)$ denotes $P(X = x_i)$, then the expected value of X is defined by

$$\mu = E(X) = \sum x_i p(x_i) \tag{3.1}$$

where the elements are summed over all values of the random variable X.

If X is a continuous random variable with probability density function $f(x)$, then the expected value of X is defined by

$$\mu = E(X) = \int x f(x) dx . \tag{3.2}$$

3.1.1.3 Variance and standard deviation

The variance of a random variable is a non-negative number that gives an idea of how widely spread the values of the random variable are likely to be; the larger the variance, the more scattered the observations on average. Variance is symbolized by $V(X)$ or $Var(X)$ or σ^2. The variance and the standard deviation of the random variable X are defined respectively as:

$$V(X) = \sigma^2 = E[X - E(X)]^2 = E(X^2) - E(X)^2, \tag{3.3}$$

and

$$STD(X) = \sqrt{V(X)} = \sqrt{\sigma^2} = \sigma, \tag{3.4}$$

where $E(X)$ is the expected value of the random variable X.

3.1.2 Stochastic Processes

A stochastic process is an ordered collection of random variables, indexed by a set T. That is, for each t in the index set T, $X(t)$ is a random variable. Mathematically, it can be expressed as $X = \{X(t), t \in T\}$. Given any t, the possible values of $X(t)$ are called the *states* of the process at t. The set of all values $X(t)$ can ever take for all t is the state space, denoted by S. The set T is usually composed of either the non-negative integers or the non-negative real values. In financial applications, the parameter t is often interpreted as time and T is the collection of possible times. If T is an interval of real values, the process $X = \{X(t), t \in T\}$ is a continuous-time process, if T is a sequence of integers, the process $X = \{X(t), t \in T\}$ is a discrete-time process and is represented by X_t in this case.

In the mathematics of probability, a stochastic process is a random function. Its behavior is governed by a random mechanism and its future values can only be predicted

with a certain amount of probability. Therefore, our main interest is the probability distribution of $X(t)$ in the continuous time case and X_t in the discrete time case.

3.1.3 Time Series

A time series is a sequence of observations that are ordered in time. Time series analysis uses statistical techniques to investigate and study the patterns of the variation of time series data. The goals of time series analysis are to understand the structure of the time series (how it depends on time, itself, and other time series variables) and to forecast or predict the future values of the time series. Time series analysis is critical for real-world risk models and can help risk models better emulate actual situations. It is widely used in most business applications, such as economic forecasting, sales forecasting, stock market analysis, and process and quality control.

Time series data, x_t, for $t = 1, 2, \ldots, n$, generally consists of both deterministic and stochastic components. The deterministic component gives rise to trends, seasonal patterns, and cycles, while the stochastic component causes statistical fluctuations that have a short term correlation structure. The time-series method for forecast involves the use of deterministic models such as extrapolation techniques or the use of complex stochastic models for adaptive forecasting [25].

3.2 Features of Time-Series Data

There are five key features for economic and business time series data [20]: trends, seasonality, aberrant observation, conditional heteroskedasticity, and nonlinearity.

3.2.1 Trends

One of the dominant features of many economic and business time series is the trend. Trend is a long term movement in a time series. It is the underlying direction and

rate of change in a time series. Such a trend can be upward or downward, it can be steep

or not, and it can be exponential or approximately linear. A trend is usually predictable

and modeled by regression methods. Here is an example of a simple regression model

being used to quantify a trend:

$$y_t = \alpha + \delta t + e_t \quad t = 1, 2, \ldots, n,$$ (3.5)

where α and δ are unknown parameters and where e_t is an unknown residual error time

series.

The simple regression model in Equation (3.5) assumes that the trend in y_t can be

represented by a linear trend $t = 1, 2, 3, \ldots$. An alternative method to obtain insight into the

trend pattern is to consider the growth rate of the variable. If the raw data are denoted by

w_t, we usually model and forecast $y_t = \log(w_t)$. In this case, it follows that

$$y_t - y_{t-1} = \log(w_t / w_{t-1})$$

$$= \log[1 + (w_t - w_{t-1})/w_{t-1}]$$

$$\approx (w_t - w_{t-1})/w_{t-1} \qquad \text{when } (w_t - w_{t-1})/w_{t-1} \text{ is small.}$$

Hence, the first difference of y_t corresponds approximately to the growth rate of w_t. If a

forecast is needed for w_t, it is usual to re-transform using $w_{n+h} = \exp(y_{n+h})$, preferably

with some correction. One of the reasons for the log-transformation is that an exponential

trend becomes linear. However, there are recent empirical and theoretical studies

question the validity of this automatism [26].

A trend pattern in economic data will be reflected by a significant average growth

rate. Alternative to Equation (3.5), we can therefore consider the regression

$$y_t - y_{t-1} = \mu + e_t \qquad t = 2, 3, \ldots, n.$$ (3.6)

Notice that we have effectively removed the trend in the regression Equation (3.6) by differencing the data. A result of this operation is that there are now n-1 observations that can be used to estimate μ. When it is possible to get rid of the trend by differencing the data, we say that y_t has a stochastic trend. If regression Equation (3.5) is more adequate, we say that y_t has a deterministic trend.

There are several different approaches to describe a trend, and each of these has a different impact on forecasting. It may not be easy to make a proper choice between the different versions of trend descriptions.

3.2.2 Seasonality

Many time series display seasonality. By seasonality, we mean periodic fluctuations. For example, retail sales tend to peak for the Christmas season and then decline after the holidays. So time series of retail sales will typically show increasing sales from September through December and declining sales in January and February. Seasonality is quite common in economic time series. It is less common in engineering and scientific data. Seasonality can be detected by graphical techniques, such as the run sequence plot, seasonal subseries plot, multiple box plots, and the autocorrelation plot. It is short term and predictable and refers to those relatively small and predictable ups and downs that are found at regular intervals in many time series.

The following regression model [26], which is a modification of Equation (3.6), is used to model seasonality:

$$y_t - y_{t-1} = \mu_1 D_{1,t} + \mu_2 D_{2,t} + \ldots + \mu_S D_{S,t} + e_t \quad t = 2,3,\ldots,n \qquad (3.7)$$

where $D_{s,t}$ is a seasonal dummy variable with

$$D_{s,t} = 1 \quad \text{when } t = (T-t)S - s, \quad \text{with } s = 1,2,\ldots\ldots S \qquad (3.8)$$

$$\text{and} \quad T = 1,2,\ldots,N$$

$$D_{s,t} = 0 \quad \text{otherwise.}$$

3.2.3 Aberrant Observations

Aberrant observations are also called outliers. They are observations that are strikingly different from the other observations on the same variable. For time series data, aberrant observations can be defined as unusual and surprising values with respect to the rest of the time series analyzed.

3.2.4 Conditional Heteroskedasticity

3.2.4.1 Homoskedasticity vs. heteroskedasticity

A univariate stochastic process X is said to be homoskedastic if standard deviations of terms X_t are constant for all time t. Otherwise, it is said to be heteroskedastic. Heteroskedasticity can take two forms. A process is unconditionally heteroskedastic if unconditional standard deviation σ_t is not constant. It is conditionally heteroskedastic if conditional standard deviations $\sigma_{t|t-1}$ are not constant. Homoskedasticity means constant variance and heteroscedasticity means changing variance.

3.2.4.2 Models for conditional heteroskedasticity

There are several kinds of models being used for a conditional heteroskedastic process.

- Autoregressive Conditional Heteroskedasticity (ARCH) model is a model of dynamic heteroskedasticity where the variance of the error term, given past information, depends linearly on the past squared errors.

- Generalized ARCH models.

- Regime-switching models.

- Stochastic volatility models.

3.2.5 Non-Linearity

Mathematically, a pure stochastic time series model for x_t is a function of an independent and identically distributed (iid) sequence consisting of the current and past shocks:

$$x_t = f(a_t, a_{t-1}, \ldots).$$ (3.9)

Here, $\{a_t\}$ is a sequence of iid random variables with a well-defined distribution function. If $f(\cdot)$ is a linear function of its arguments, Equation (3.9) is a linear model. Otherwise, Equation (3.9) is a non-linear model. Here, non-linearity means the function is non-linear.

3.3 Time-Series Models

3.3.1 Stationarity

A common assumption in many time series techniques is that the data are stationary. A time series is said to be strictly stationary if the statistical properties remain constant over time. A time series is said to be weakly stationary if the mean and covariance functions remain constant over time.

Any stochastic time series $x_1, x_2, \ldots, x_T$ can be regarded as having been generated by a set of jointly distributed random variables $\{X_t, t = 1, 2, \ldots, T\}$ [14][21]; that is to say,

the set of data points $x_1, x_2, \ldots, x_T$ represents a particular realization of the joint probability distribution function $p(x_1, x_2, \ldots, x_T)$. A future observation x_{T+1} can be thought of as being generated by a conditional distribution function $p(x_{T+1} \mid x_1, x_2, \ldots, x_T)$, a probability distribution for x_{T+1} given the past observations $x_1, x_2, \ldots, x_T$. A strictly stationary process is defined as one whose joint distribution and conditional distribution are constant with respect to the displacement in time.

If the series $\{x_t\}$ is weakly stationary, the mean of the series, $\mu_x = E(x_t)$, must be stationary, so that $E(x_t) = E(x_{t+m})$, for any t and m. Furthermore, the variance of the series, $\sigma_x^2 = E[(x_t - \mu_x)^2]$, must be stationary, so that $E[(x_t - \mu_x)^2] = E[(x_{t+m} - \mu_x)^2]$, and finally, for any lag k, the covariance of the series,

$$\gamma_k = Cov(x_t, x_{t+k}) = E[(x_t - \mu_x)(x_{t+k} - \mu_x)]$$

must be stationary, so that $Cov(x_t, x_{t+k}) = Cov(x_{t+m}, x_{t+k+m})$.

Stationarity is used as a tool in time series analysis, where the raw data are often transformed to become stationary. Processes are described as trend stationary if they are a linear combination of a stationary process and one or more processes exhibiting a trend. Transforming this data to leave a stationary data set for analysis is referred to as de-trending.

For practical purposes, stationarity can usually be determined from a run sequence plot. If the time series is not stationary, it can be transformed to stationarity with one of the following techniques [27]:

- Difference the data. That is, given the series X_t, we create the new series $Y_t = X_t - X_{t-1}$. The differenced data will contain one less point than the

original data. Although you can difference the data more than once, one difference is usually sufficient.

- If the data contain a trend, we can fit some type of curve to the data and then model the residuals from that fit. Since the purpose of the fit is to simply remove long term trend, a simple fit, such as a straight line, is typically used.

- For non-constant variance, taking the logarithm or square root of the series may stabilize the variance. For negative data, you can add a suitable constant to make the entire data positive before applying the transformation. This constant can then be subtracted from the model to obtain predicted (i.e., the fitted) values and forecasts for future points.

The above techniques are intended to generate series with constant location and scale. Although seasonality also violates stationarity, this is usually explicitly incorporated into the time series model.

3.3.2 Correlation and Autocorrelation

In probability theory and statistics, correlation, also called correlation coefficient, indicates the strength and direction of a linear relationship between two random variables. In general statistical usage, correlation or co-relation refers to the departure of two variables from independence.

3.3.2.1 Correlation

The correlation $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is defined as:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}. \tag{3.10}$$

Since $\mu_X = E(X)$, $\sigma_X^2 = E(X^2) - E^2(X)$ and likewise for Y, we may also write

$$\rho_{X,Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}}. \tag{3.11}$$

The correlation is defined only if both standard deviations are finite and both of them are

nonzero.

3.3.2.2 Autocorrelation

The autocorrelation ρ_k is the correlation between the neighboring data points in

the same series $\{x_t\}$. The autocorrelation with lag k is defined as

$$\rho_k = \frac{E[(x_t - \mu_x)(x_{t+k} - \mu_x)]}{\sqrt{E[(x_t - \mu_x)^2] \cdot E[(x_{t+k} - \mu_x)^2]}} = \frac{Cov(x_t, x_{t+k})}{\sigma_{x_t} \cdot \sigma_{x_{t+k}}}. \tag{3.12}$$

For a stationary process the variance at time t is the same as the variance at time $t + k$.

This means, $\sigma_{x_t} = \sigma_{x_{t+k}} = \sigma_x$, and $\gamma_k = Cov(x_t, x_{t+k}) = E[(x_t - \mu_x)(x_{t+k} - \mu_x)]$, and ρ_k

can be expressed as

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{3.13}$$

where $\gamma_0 = \sigma_x^2$. For any stochastic process, ρ_0 equals to 1.

3.3.3 Linear Time-Series Models

3.3.3.1 Autoregressive (AR) models

In an autoregressive process of order p the current observation x_t is generated by

a weighted average of past observations going back p periods, together with a random

disturbance in the current period. This process is denoted as AR(p) and is written as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \delta + \varepsilon_t, \qquad (3.14)$$

where δ is a constant term which relates to the mean of the stochastic process and ε_t is

an independently distributed random variable with zero mean.

3.3.3.2 Moving-average (MA) models

In the moving average process of order q each observation x_t is generated by a

weighted average of random disturbances going back q periods. This process is denoted

as MA(q) and is written as

$$x_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}, \qquad (3.15)$$

where the parameters $\theta_1, \theta_2, \ldots, \theta_q$ may be positive or negative.

In the moving average model and the autoregressive model, the random

disturbances are assumed to be independently distributed across time, i.e., generated by a

white noise process. In this case, $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, and $E(\varepsilon_t \varepsilon_{t-k}) = 0$ for $k \neq 0$.

3.3.3.3 Autoregressive-moving average (ARMA) models

In the autoregressive-moving average process of order (p, q) each observation x_t

is generated by a weighted average of past observations going back p periods, together

with a weighted average of random disturbances going back q periods. This process is

denoted as ARMA (p, q) and is written as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}. \qquad (3.16)$$

If the process is stationary, its mean is constant over time and is given by

$$\mu = \phi_1 \mu + \phi_2 \mu + \cdots + \phi_p \mu + \delta$$

or

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}.$$ (3.17)

This gives a necessary condition for the stationarity of the process:

$$\phi_1 + \phi_2 + \cdots + \phi_p < 1.$$ (3.18)

The process ARMA (1, 1) is the simplest autoregressive-moving average process:

$$x_t = \phi_1 x_{t-1} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}.$$ (3.19)

The variance of this process is given by

$$\gamma_0 = \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2} \sigma_\varepsilon^2.$$ (3.20)

The covariances $\gamma_1, \gamma_2, \ldots, \gamma_k, \ldots$ can be determined recursively as follows:

$$\gamma_1 = E[x_{t-1}(\phi_1 x_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] = \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2$$

$$= \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 - \phi_1^2} \sigma_\varepsilon^2,$$ (3.21)

$$\gamma_2 = E[x_{t-2}(\phi_1 x_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] = \phi_1 \gamma_1,$$ (3.22)

and

$$\gamma_k = \phi_1 \gamma_{k-1} \text{ for } k \geq 2.$$

The autocorrelation function is given by

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}$$ (3.23)

and

$$\rho_k = \phi_1 \rho_{k-1} \text{ for } k \geq 2.$$ (3.24)

Thus, the autocorrelation function begins at its starting value ρ_1 and then decays geometrically from this starting value.

3.3.3.4 Integrated ARMA (ARIMA) models

AR, MA and ARMA models are built on stationary (constant mean and variance) data. Most business and economic time series are not stationary because they contain trends or random shifts in level. Thus they must be transformed to stationarity before they can be fitted to ARMA models. A nonstationary series implies that the distant past has as much or more weight than the recent past. Box and Jenkins suggest differencing to obtain stationarity. When a series is differenced, each value is replaced by its incremental change from the last value. The degree of differencing is the number of times the data transformation is executed. We continue to difference until the data are stationary. If the original series is growing at an increasing rate (exponential trend), the first difference (1) will still show a trend. A second difference (2) will be stationary.

After differencing the series x_t to produce the stationary series w_t, we can model w_t as a ARMA process. If $w_t = \Delta^d x_t$, where Δ denotes differencing and d is the order of differencing, and w_t is an ARMA (p, q) process, then x_t is an integrated autoregressive-moving average process of order (p, d, q). This process is denoted as ARIMA $(p, d. q)$ and is written as

$$\phi(B)\Delta^d x_t = \delta + \theta(B)\varepsilon_t \qquad (3.25)$$

with

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \qquad (3.26)$$

and

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q . \qquad (3.27)$$

Here B is the backward shift operator which imposes a one-period time lag each time it is applied to a variable. Thus, $Bx_t = x_{t-1}$, $B^2 x_t = x_{t-2}, \ldots$, and $B^n x_t = x_{t-n}$. $\phi(B)$ is called

the autoregressive operator and $\theta(B)$ is called the moving average operator. The mean of w_t is given by

$$\mu_w = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}.$$ (3.38)

If δ is not equal to 0, the integrated series x_t will have a built-in deterministic trend.

3.3.3.5 Seasonal models

For a seasonal time series x_t with periodicity s, seasonal differencing means $\Delta_s x_t = x_t - x_{t-s} = (1 - B^s)x_t$. The seasonal time series model is defined mathematically as [23]

$$(1 - B^s)(1 - B)x_t = (1 - \theta)(1 - \Theta B^s)\varepsilon_t,$$ (3.39)

where s is the periodicity of the series, $|\theta| < 1$, and $|\Theta| < 1$.

3.3.4 Non-Linear Time-Series Models

3.3.4.1 Conditional heteroskedastic models

Consider the information set F_{t-1}, which contains all the information on the variable x_t until $t - 1$. The conditional mean and variance of x_t, given F_{t-1}, are

$$\begin{aligned}
\mu_t &= E(x_t \mid F_{t-1}) \\
\sigma_t^2 &= Var(x_t \mid F_{t-1}) = E[(x_t - \mu_t)^2 \mid F_{t-1}].
\end{aligned}$$ (3.40)

Assume that x_t follows a stationary ARMA(p, q) model,

$$x_t = \mu_t + a_t, \qquad \mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} - \sum_{i=1}^{q} \theta_i a_{t-i}.$$ (3.41)

Combining Equations (3.40) and (3.41), we get

$$\sigma_t^2 = Var(x_t \mid F_{t-1}) = Var(a_t \mid F_{t-1}).$$ (3.42)

There are two categories of conditional heteroscedastic models according to how σ_t^2 is defined. The models in the first category use an exact function to govern the evolution of σ_t^2, e.g, the GARCH model. The models in the second category use a stochastic equation to describe σ_t^2, e.g., the stochastic volatility model.

(1) ARCH Models. ARCH stands for Autoregressive Conditional Heteroskedasticity. It is a technique in finance to model asset price volatility over time. It is observed in much time series data on asset prices that there are periods when the variance is high and periods where the variance is low. The ARCH economic model for this (introduced by Engle (1982) [28]) is that the variance of the series itself is an AR (autoregressive) time series, often a linear one. ARCH (autoregressive conditional heteroskedasticity) models recognize the presence of successive periods of relative volatility and stability. The error variance, conditional on past information, evolves over time as a function of past errors.

In the linear ARCH (m) model originally introduced by Engle (1982) [28], the time varying conditional variance is postulated to be a linear function of the past m squared innovations.

$$a_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \cdots + \alpha_m a_{t-m}^2, \qquad (3.43)$$

where $\{\varepsilon_t\}$ is a sequence of iid random variables with mean zero and variance 1, $\alpha_0 > 0$, and $\alpha_i \geq 0$ for $i > 0$. In practice, ε_t is often assumed to have the standard normal or a standardized Student t-distribution.

According to Equation (3.43), the ARCH(1) model can be written as

$$a_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2, \qquad (3.44)$$

where $\alpha_0 > 0$ and $\alpha_1 \geq 0$. Usually, the ARCH(1) model is carefully studied to understand the properties of ARCH models.

(2) GARCH Models. Bollerslev (1986) [29] proposed a GARCH(m, s) (generalized ARCH) model as

$$a_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 , \tag{3.45}$$

where again $\{\varepsilon_t\}$ is a sequence of iid random variables with mean zero and variance 1, $\alpha_0 > 0$, and $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$.

Rearranging the GARCH(m, s) model by defining $\eta_t = a_t^2 - \sigma_t^2$ so that $\sigma_t^2 = a_t^2 - \eta_t$, it follows that

$$a_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) a_{t-i}^2 + \eta_t - \sum_{j=1}^{s} \beta_j \eta_{t-j} , \tag{3.46}$$

which defines an ARMA (Max(m, s), s) model for a_t^2.

Only low-order GARCH models are used in most applications, such as GARCH(1,1), GARCH(2,1) and GARCH(1,2). According to Equation (3.45), the GARCH(1,1) model can be written as

$$a_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{3.47}$$

where $0 \leq \alpha_1, \beta_1 \leq 1$, and $(\alpha_1 + \beta_1) < 1$.

3.3.4.2 Regime-switching models

A stochastic process is regime-switching if its behavior is determined by different models, different regimes, during different periods.

(1) Threshold Autoregressive (TAR) Models. Consider a simple AR(p) model for a time series x_t:

$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \sigma \varepsilon_t \tag{3.48}$$

where ϕ_i ($i = 1, 2, \cdots, p$) are the AR coefficients, $\varepsilon_t \sim WN(0,1)$ and $\sigma > 0$ is the standard deviation of the random error ε_t. The model parameters μ, ϕ_i ($i = 1, 2, \cdots, p$) and σ are independent of the time t and remain constant. To capture non-linear dynamics, parameters are assumed to change according to the value of a weakly exogenous threshold variable z_t:

$$x_t = X_t \Phi^{(j)} + \sigma^{(j)} \varepsilon_t \quad \text{if } r_{j-1} < z_t \leq r_j \tag{3.49}$$

where $X_t = (1, x_{t-1}, x_{t-2}, \cdots, x_{t-p})$, $j = 1, 2, \cdots, k$, and $-\infty = r_0 < r_1 < \cdots < r_{k-1} < r_k = \infty$. In essence, the $k - 1$ non-trivial thresholds $(r_1, r_2, \cdots, r_{k-1})$ divide the domain of the threshold variable z_t into k different regimes. In each different regime, the time series x_t follows a different AR(p) model.

When the threshold variable $z_t = x_{t-d}$, with the delay parameter d being a positive integer, the dynamics or regime of x_t is determined by its own lagged value x_{t-d} and the TAR model is called a self-exciting TAR (SETAR) model.

The TAR model is simple and easy to understand, but rich enough to generate complex non-linear dynamics.

(2) Smooth Transition Autoregressive Models. In the TAR models, a regime switch is discontinuous. If the regime switch happens gradually in a smooth fashion and the discontinuity of the thresholds is replaced by a smooth transition function, TAR

models can be generalized to smooth autoregressive (STAR) models. A systematic modeling approach of STAR models is proposed by Terasvirta [30] and a recent development of STAR is reviewed in [31].

(3) Markov Switching Models. For TAR, SETAR and STAR models, regimes determined by observable variables. In a Markov switching model, regimes are determined by an unobserved state or regime variable which follows a discrete Markov process, also called a Markov chain.

In a paper published in 1989 [32], James Hamilton developed an extremely useful tool for statistically modeling regime shifts in autoregressive time series models. In order to unstandard this model, it is useful to begin with a simple linear time-series framework for the growth rate of some measure of economic activity, x_t.

$$(x_t - \mu) = \rho(x_{t-1} - \mu) + \varepsilon_t \qquad (3.50)$$

with $\varepsilon_t \sim N(0, \sigma^2)$.

In this model, the growth rate of economic activity has a mean denoted by μ. Deviations from this mean growth rate are created by the stochastic disturbance ε_t. These deviations are serially correlated, modeled as an AR(1) time series process with parameter ρ.

Hamilton's innovation was to allow the parameters of the model in Equation (3.50) to switch between two regimes, where the switching is governed by a state variable, $S_t = \{0,1\}$. When $S_t = 0$, the parameters of the model are different from those when $S_t = 1$. The probability driving S_t is captured by the following four transition probabilities:

$$P(S_t = 1 \mid S_{t-1} = 1) = p$$
$$P(S_t = 0 \mid S_{t-1} = 1) = 1 - p$$
$$P(S_t = 0 \mid S_{t-1} = 0) = q \qquad (3.51)$$
$$P(S_t = 1 \mid S_{t-1} = 0) = 1 - q$$

Hamilton specified the following augmented version of Equation (3.50):

$$(y_t - \mu_{S_t}) = \rho(y_{t-1} - \mu_{S_{t-1}}) + \varepsilon_t$$
$$\varepsilon_t \sim N(0, \sigma^2)$$
$$\mu_{S_t} = \mu_0 + \mu_1 S_t \qquad (3.52)$$
$$\mu_1 < 0$$

where S_t depends on the transition probabilities in Equation (3.53). Here, when S_t

switches from 0 to 1, the growth rate of the economic activity switches from μ_0 to

$\mu_0 + \mu_1$. Since $\mu_1 < 0$, the model will estimate these switches at times when economic

activity switches from high growth to low growth states.

3.3.4.3 Stochastic volatility models

A stochastic volatility model models both the underlier's value and its volatility as

stochastic processes. The aim of a stochastic volatility model is to incorporate the

empirical observation that volatility appears to vary, at least in part, randomly. The idea

is to make the volatility itself a stochastic process. To ensure positiveness of the

conditional variance, SV models use $\ln(\sigma_t^2)$ instead of σ_t^2. An SV model is defined as

[33]

$$a_t = \sigma_t \varepsilon_t$$
$$(1 - \alpha_1 B - \cdots - \alpha_m B^m) \ln(\sigma_t^2) = \alpha_0 + v_t \qquad (3.53)$$

where ε_t's are iid N(0, 1), v_t's are iid $N(0, \sigma_v^2)$, $\{\varepsilon_t\}$ and $\{v_t\}$ are independent, and α_0

is a constant.

An alternative lognormal SV model is given by

$$X_t = \sigma_t e_t,$$ (3.54)

$$\ln \sigma_t^2 = \mu + \phi(\ln \sigma_{t-1}^2 - \mu) + \sigma v_t,$$ (3.55)

where X_t is a continuously compounded return and e_t, v_t are two uncorrelated iid

$N(0, 1)$ sequences.

3.3.5 State-Space Models

The state-space model represents a multivariate time series through auxiliary variables, some of which may not be directly observable. These auxiliary variables are called the state vector. The state vector summarizes all the information from the present and past values of the time series relevant to the prediction of future values of the series. The observed time series are expressed as linear combinations of the state variables plus error terms [34]. The state variables are generally unknown. The dynamics of the state variables, and their relationships are to be inferred from the observed time series. The state space model is also called a Markovian representation, or a canonical representation, of a multivariate time series process.

The state-space form encompasses a very rich class of models. Any Gaussian multivariate stationary time series can be written in a state-space form, provided that the dimension of the predictor space is finite. In particular, any autoregressive moving average (ARMA) process has a state-space representation and, conversely, any state-space process can be expressed in an ARMA form.

State-space models are a flexible family of models which can be used for the modeling of many scenarios. The strongest feature of state-space models is the existence of a very general algorithm, kalman filter, for filtering, smoothing and predicting [35].

The Kalman filter is a technique that can be used to recursively estimate unobservable quantities called state variables from an observed time series. It is used in different areas, such as control theory and control systems, computer graphics, and economics, because it is efficient and easy to implement.

The state-space model is a two-layer model [36]. The external layer expresses the observed time series $\vec{X}_t$ as a function of the state variables $\vec{S}_t$ plus noise. The internal layer determines the state $\vec{S}_{t+1}$ at time $t + 1$ in terms of the previous state $\vec{S}_t$ and a noise term. Depending on the mapping function between the observed time series and the state variables and the mapping function between the state variables, the state-space models are divided into linear and non-linear categories.

3.3.5.1 Linear state-space models

The basic model can be described by a pair of equations:

$$\vec{S}_{t+1} = F_t\vec{S}_t + \vec{V}_t, \quad t = 1,2,\cdots, \tag{3.56}$$

$$\vec{X}_t = G_t\vec{S}_t + \vec{W}_t, \quad t = 1,2,\cdots, \tag{3.57}$$

where $\vec{X}_t \in \Re^n, t = 1,2,\ldots,T$ are the observations and $\vec{S}_t \in \Re^m, t = 1,2,\ldots,T$ are the hidden internal states of the dynamic system. These internal states are in the so called state-space, hence the name state-space model. The vector $\vec{V}_t$ is the process noise which is assumed to be drawn from a zero mean multivariate normal distribution with covariance Q_t, written as $\vec{V}_t \sim N(0,Q_k)$. Similarly, $\vec{W}_t$ is the observed noise which is assumed to be a Gaussian white noise with covariance R_t, written as $\vec{W}_t \sim N(0,R_t)$. The mappings F_t and G_t are the linear observation and prediction mappings. Equation

(3.56) is called the state equation which determines the state $\vec{S}_{t+1}$ at time $t + 1$ in terms of

the previous state $\vec{S}_t$ and a noise term. Equation (3.57) is called the observation equation

which expresses the n-dimensional observation $\vec{X}_t$ as a linear function of an m-

dimensional state variable $\vec{S}_t$ plus noise.

Linear state-space models (SSMs) are used widely because they are very efficient

and provide a good enough short-term approximation for many cases even if the true

dynamics are nonlinear. The most famous variant of linear SSMs is the Kalman filter. It

gives a learning algorithm for the model defined by Equations (3.56) and (3.57) assuming

all the parameters are Gaussian and that certain independence assumptions are met.

There are three fundamental problems associated with the state-space model,

defined by Equations (3.56) and (3.57). These are all concerned with finding best

(minimum mean square error) linear estimates of the state-vector $\vec{S}_t$ in terms of the

observations $\{\vec{X}_t\}$ and a random vector $\vec{X}_0$ that is orthogonal to $\vec{V}_t$ and $\vec{W}_t$ for all $t \geq 1$.

In many cases $\vec{X}_0$ will be the constant vector $(1,1,\ldots,1)'$. Estimation of $\vec{S}_t$ in terms of

$\{\vec{X}_t, t = 0,1,2,\ldots,t-1\}$ is defined as the prediction problem. Estimating of $\vec{S}_t$ in terms of

$\{\vec{X}_t, t = 0,1,2,\ldots,t-1,t\}$ is defined as the filtering problem. Estimating of $\vec{S}_t$ in terms of

$\{\vec{X}_t, t = 0,1,2,\ldots,t-1,t,t+1,\ldots,n\}$ is defined as the smoothing problem. See reference

[37] for detailed formulations and proofs.

3.3.5.2 Non-linear state-space models

The nonlinear state-space model is in principle a very simple extension of the

linear model. All that needs to be done is to replace the linear mappings F_t and G_t of

Equations (3.56) and (3.57) with general non-linear mappings to get

$$\vec{S}_{t+1} = f(\vec{S}_t) + V_t, \quad t = 1,2,\cdots, \tag{3.58}$$

$$\vec{X}_t = g(\vec{S}_t) + W_t, \quad t = 1,2,\cdots, \tag{3.59}$$

where f and g are sufficiently smooth nonlinear mappings. One of the greatest difficulties with the nonlinear model is that while a linear mapping $G: \mathfrak{R}^m \rightarrow \mathfrak{R}^n$ can be uniquely determined by $m \cdot n$ real numbers (the elements of the corresponding matrix), there is no way to do the same for nonlinear mappings. Representing an arbitrary nonlinear mapping with even moderate accuracy requires many more parameters.

3.4 Methods in this Research

In this research, ARIMA and state-space models will be used to forecast the three risk factors for a portfolio of three stocks with three risk factors.

3.4.1 Selected Risk Factors

3.4.1.1 Candidate factors

The type and number of factors is extremely important to build a reasonable model. Three risk factors are selected based on a five-factor variation of the Fama and French model [38]. The five candidate factors are

(1) The weekly yield on 13-week T-Bills as of the beginning of the week. This is a proxy for the risk-free rate (RFR), which is used to capture changes in interest rates over the 5-year period, and the effects thereof on security values.

(2) The difference between the weekly yield on 30-year T-Bonds, as of the beginning of the week, and the risk-free rate from risk-factor (1). This provides a proxy for the Maturity Risk Premium (MRP) in security values.

(3)   The difference between the weekly value on the S&P 500 and beginning-of-the-week's yield on 30-year T-Bonds (from risk-factor (2), above). This provides a proxy for the Market-Risk Premium (MktRP).

(4)   The difference between the weekly values for the S&P 600 Small-Cap Index and the S&P 500 Index (which is comprised of large-cap stocks). This provides a proxy for the "Small-Firm" or Size Premium (SizeP).

(5)   The difference between the weekly values for the Dow-Jones Industrial Average (DJIA) and the Nasdaq Composite Index. This is chosen as a proxy for the "Value-versus-Growth Premium" (VvsGP) in security values, under the assumption that DJIA stocks are more likely to be value stocks than growth stocks, while NASDAQ stocks are more likely to be growth than value stocks.

To construct the five candidate risk factor, we need historical data for the following six indices: (1) 13-week Treasury Bill Index (^IRX), (2) 30-year T-Bond yield Index (^TYX), (3) S&P 500 Index (^GSPC), (4) S&P Small Cap Index (^SML), (5) Down Jones Index (^DJI), and (6) NASDAQ Composite Index (^IXIC).

In this research, 10-year weekly data of the six indices were downloaded from Yahoo! Finance, starting from Jan. 2, 1996 and ending at Jan. 17, 2006.

3.4.1.2 Selected factors

After collecting the historical data and constructing the five candidate risk factors, the correlation matrix of the five risk factors is computed and the result is in Table 3.1.

Table 3.1 Correlation matrix for the five candidate risk factors

	R_{RFR}	R_{MRP}	R_{MktRP}	R_{SizeP}	R_{VvsGP}
R_{RFR}	1				
R_{MRP}	-0.91586	1			
R_{MktRP}	0.137099	-0.35094	1		
R_{SizeP}	-0.30068	0.456678	-0.96329	1	
R_{VvsGP}	-0.49502	0.2572	0.681609	-0.51768	1

According to Table 3.1, it is obvious that R_{RFR} and R_{MRP} are highly correlated, and

R_{MktRP} and R_{SizeP} are highly correlated. In the effort to choose risk factors that are near

independent so as to avoid multicollinearity, two risk factors, R_{RFR} and R_{SizeP}, are

deleted. The correlation matrix for the remaining three factors is calculated and shown in

Table 3.2.

Table 3.2 Correlation matrix for the three risk factors

	R_{MRP}	R_{MktRP}	R_{VvsGP}
R_{MRP}	1		
R_{MktRP}	-0.35094	1	
R_{VvsGP}	0.2572	0.681609	1

3.4.2 Models and SAS PROCs

3.4.2.1 ARIMA in SAS

PROC ARIMA performs analysis in three distinct phases and produces separate

output for each phase [39]. Table 3.3 illustrates the three phases.

Table 3.3 Illustration of the three phases in PROC ARIMA

Identification Phase	Produces plots for identifying the process that underlies a time series
Estimation Phase	Produces model parameter estimates and associated significance tests, goodness-of-fit statistics, and other estimation diagnostics.
Forecasting Phase	Produces forecasts, standard errors, and confidence limits for the specific number of time periods.

3.4.2.2 STATESPACE in SAS

PROC STATESPACE analyzes and forecasts multivariate time series using the state space model. The STATESPACE procedure is appropriate for jointly forecasting several related time series that have dynamic interactions. By taking into account the autocorrelations among the whole set of variables, the STATESPACE procedure may give better forecasts than methods that model each series separately. The STATESPACE procedure automatically selects the best state-space model appropriate for the time series. Alternatively, you can specify the state-space model by giving the form of the state vector and the state transition and innovation matrices. The methods used by the STATESPACE procedure assume that the time series are jointly stationary. Nonstationary series must be made stationary by some preliminary transformation, usually by differencing.

PROC STATESPACE is based on Akaike's method of estimating state-space models, which uses canonical correlations. The model is defined by the following state transition equation:

$$z_{t+1} = Fz_t + Ge_{t+1}.$$
(3.60)

In addition to the state transition equation, state-space models usually include a measurement equation or observation equation that gives the observed values $\vec{x}_t$ as a function of the state vector $\vec{z}_t$.

$$\vec{x}_t = [I_r \quad 0]\vec{z}_t \, . \tag{3.61}$$

Table 3.4 lists all the variables in the Equations (3.60) and (3.61) and gives explanation for each of them.

Table 3.4 Information on the variables in Equations (3.60) and (3.61)

$\vec{x}_t$	an observation vector of dimension r, after differencing and subtracting the sample mean.
$\vec{z}_t$	A state vector of dimension s ($s > r$), whose first r elements are $\vec{x}_t$ and whose last s-r elements are conditional prediction of future $\vec{x}_{t+k\mid t}$.
F	An $s \times s$ transition matrix which determines the dynamic properties of the model.
G	An $s \times r$ transition matrix, with the identity matrix I_r forming the first r rows and columns. It is called the *input matrix* and determines the variance structure of the transition equation.
$\vec{e}_t$	A sequence of independent normally distributed random vectors of dimension r with mean 0 and covariance matrix Σ_{ee}. It is called the innovation vector or shock vector.
I_r	An $r \times r$ identity matrix.

3.5 Results and Discussions

3.5.1 MA(1) Model for Risk Factor MRP

The PROC ARIMA is used to identify, estimate and forecast risk factor R_{MRP}.

First, take the first differences of the time series of R_{MRP}.

The output in Figure 3.2 lists the periods of differencing. The sample ACF drops to 0 at lag 1, which indicates a random process. Also, the Autocorrelation Check for White Noise has non-significant Q-statistics. This anomaly indicates that the differenced series is white noise. In addition, The Autocorrelation Check for Residuals shows that one of the Q-statistics is significant. Thus the residuals are not white noise. It is concluded that first differencing is not adequate to make the time series stationary.

Second, take second differences of the time series of R_{MRP}.

The output in Figure 3.3 lists the periods of differencing. The sample ACF drops to 0 after lag 1, which indicates an MA (1) process. The Autocorrelation Check for White Noise has significant Q-statistics. This indicates that the differenced series is not white noise. In addition, the Autocorrelation Check for Residuals shows that none of the Q-statistics are significant. Thus the residuals are white noise. It is concluded that an MA(1) model provides an adequate fit for the second differenced R_{MRP} time series. The model information is shown at the bottom of the Figure 3.3. The estimated MA(1) model for R_{MRP} is written as

$$(1-B)^2 R_{MRP,t} = -0.00026 + \varepsilon_t - 0.95251\varepsilon_{t-1}. \tag{3.62}$$

Figure 3.4 displays the forecasts for the R_{MRP} time series. It begins at 0.8448 for the first future week , then they decrease each week and reach a value of 0.4414 in the tenth

future week. Table 3.5 lists the historical values for R_{MRP} and the corresponding

forecasting values.

```
                        Name of Variable = MRP

        Period(s) of Differencing                       1
        Mean of Working Series                    -0.00138
        Standard Deviation                        0.137179
        Number of Observations                         523
        Observation(s) eliminated by differencing        1
```

```
                              Autocorrelations

Lag    Covariance    Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1    Std Error

 0     0.018818       1.00000       |                    |*******************|         0
 1    -0.0008984      -.04774       |                 .*|.                   |     0.043727
 2     0.0011569      0.06148       |                 . |*.                  |     0.043826
 3     0.0014566      0.07741       |                 . |**                  |     0.043991
 4     0.00061854     0.03287       |                 . |*.                  |     0.044251
 5     0.00007711     0.00410       |                 . | .                  |     0.044297
 6     0.00008359     0.00444       |                 . | .                  |     0.044298
 7     0.00008270     0.00439       |                 . | .                  |     0.044299
 8     0.00065680     0.03490       |                 . |*.                  |     0.044300
 9    -0.0012768      -.06785       |                 .*| .                  |     0.044352
10     0.00015716     0.00835       |                 . | .                  |     0.044550
11     0.0012540      0.06664       |                 . |*.                  |     0.044553
12     0.0010726      0.05700       |                 . |*.                  |     0.044744
13     0.0010197      0.05419       |                 . |*.                  |     0.044882
14    -0.0001916      -.01018       |                 . | .                  |     0.045007
15     0.0012520      0.06653       |                 . |*.                  |     0.045011
16     0.00050837     0.02702       |                 . |*.                  |     0.045199
17     0.0014166      0.07528       |                 . |**                  |     0.045230
18     0.00054030     0.02871       |                 . |*.                  |     0.045469
19    -0.0002080      -.01105       |                 . | .                  |     0.045504
20     0.00096661     0.05137       |                 . |*.                  |     0.045509
21     0.00001917     0.00102       |                 . | .                  |     0.045619
22    -0.0000324      -.00172       |                 . | .                  |     0.045619
23     0.0017022      0.09046       |                 . |**                  |     0.045620
24     0.00083250     0.04424       |                 . |*.                  |     0.045961
```

```
                  Autocorrelation Check for White Noise

 To     Chi-            Pr >
Lag    Square    DF    ChiSq    --------------------Autocorrelations--------------------

  6     6.95      6    0.3259   -0.048   0.061    0.077    0.033    0.004    0.004
 12    14.23     12    0.2864    0.004   0.035   -0.068    0.008    0.067    0.057
 18    22.18     18    0.2242    0.054  -0.010    0.067    0.027    0.075    0.029
 24    29.26     24    0.2107   -0.011   0.051    0.001   -0.002    0.090    0.044
 30    32.12     30    0.3621    0.014   0.044   -0.039    0.015   -0.004    0.034
 36    47.86     36    0.0894    0.023  -0.027    0.090   -0.012    0.067    0.119
```

```
                  Autocorrelation Check of Residuals

 To     Chi-            Pr >
Lag    Square    DF    ChiSq    --------------------Autocorrelations--------------------

  6     6.48      5    0.2626   -0.003   0.065    0.082    0.037    0.006    0.005
 12    14.12     11    0.2266    0.006   0.032   -0.066    0.008    0.070    0.062
 18    22.78     17    0.1565    0.057  -0.005    0.068    0.033    0.078    0.032
 24    30.28     23    0.1415   -0.008   0.051    0.003    0.002    0.093    0.049
 30    33.07     29    0.2748    0.018   0.044   -0.037    0.014   -0.002    0.035
 36    49.02     35    0.0581    0.024  -0.022    0.089   -0.005    0.071    0.120
 42    62.87     41    0.0156   -0.049   0.090   -0.020   -0.008    0.037   -0.110
```

Figure 3.2 Output of ARIMA: Identifying the first differenced R_{MRP} time series

Name of Variable = MRP

Period(s) of Differencing   1,1
Mean of Working Series   -0.00033
Standard Deviation   0.198666
Number of Observations   522
Observation(s) eliminated by differencing   2

Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	0.039468	1.00000	|********************|	0
1	-0.021747	-.55099	***********| .	0.043769
2	0.0017095	0.04331	. |* .	0.055488
3	0.0010921	0.02767	. |* .	0.055553
4	-0.0002334	-.00591	. | .	0.055579
5	-0.0005908	-.01497	. | .	0.055580
6	0.00002640	0.00067	. | .	0.055588
7	-0.0005300	-.01343	. | .	0.055588
8	0.0024153	0.06120	. |* .	0.055594
9	-0.0032782	-.08306	**| .	0.055723
10	0.00033504	0.00849	. | .	0.055960
11	0.0012667	0.03209	. |* .	0.055962
12	-0.0001227	-.00311	. | .	0.055997
13	0.0010942	0.02772	. |* .	0.055998
14	-0.0026117	-.06617	. *| .	0.056024
15	0.0022249	0.05637	. |* .	0.056174
16	-0.0017559	-.04449	. *| .	0.056282
17	0.0019154	0.04853	. |* .	0.056349
18	-0.0001384	-.00351	. | .	0.056429
19	-0.0019536	-.04950	. *| .	0.056430
20	0.0020750	0.05257	. |* .	0.056513
21	-0.0008533	-.02162	. | .	0.056606
22	-0.0018215	-.04615	. *| .	0.056622
23	0.0026411	0.06692	. |* .	0.056694

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	----------------Autocorrelations----------------
6	160.92	6	<.0001	-0.551 0.043 0.028 -0.006 -0.015 0.001
12	167.28	12	<.0001	-0.013 0.061 -0.083 0.008 0.032 -0.003
18	174.12	18	<.0001	0.028 -0.066 0.056 -0.044 0.049 -0.004
24	180.86	24	<.0001	-0.049 0.053 -0.022 -0.046 0.067 -0.007
30	186.96	30	<.0001	-0.031 0.057 -0.067 0.035 -0.026 0.022
36	207.74	36	<.0001	0.019 -0.080 0.105 -0.085 0.013 0.109

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	----------------Autocorrelations----------------
6	5.76	5	0.3302	-0.083 0.029 0.047 0.005 -0.024 -0.021
12	12.41	11	0.3336	-0.020 0.011 -0.092 -0.014 0.046 0.036
18	16.95	17	0.4580	0.032 -0.031 0.049 0.007 0.062 0.011
24	22.15	23	0.5111	-0.032 0.032 -0.018 -0.021 0.077 0.029
30	24.90	29	0.6833	-0.003 0.031 -0.057 -0.001 -0.021 0.017
36	38.79	35	0.3028	0.007 -0.044 0.079 -0.024 0.058 0.112
42	55.88	41	0.0606	-0.069 0.088 -0.033 -0.017 0.037 -0.122

Figure 3.3 Output of ARIMA: Identifying the second differenced R_{MRP} time series

```
Forecasts for variable MRP
```

Obs	Forecast	Std Error	95% Confidence Limits	
515	0.8448	0.1394	0.5715	1.1180
516	0.8093	0.2020	0.4134	1.2051
517	0.7736	0.2532	0.2772	1.2699
518	0.7376	0.2992	0.1511	1.3241
519	0.7014	0.3423	0.0306	1.3722
520	0.6650	0.3834	-0.0865	1.4164
521	0.6283	0.4233	-0.2013	1.4580
522	0.5914	0.4624	-0.3149	1.4977
523	0.5543	0.5010	-0.4276	1.5362
524	0.5169	0.5392	-0.5399	1.5738
525	0.4793	0.5773	-0.6521	1.6107
526	0.4414	0.6152	-0.7644	1.6473

Figure 3.4 Output of ARIMA: Forecasts of R_{MRP} time series over 10 weeks

Table 3.5 Comparison of real and forecast values for the R_{MRP} time series over 10 weeks

R_{MRP}	0.78	0.81	0.83	0.9	0.82	0.66	0.57	0.46	0.31	0.28
R_{MRP_f}	0.84	0.81	0.77	0.70	0.67	0.63	0.59	0.52	0.48	0.44

3.5.2 State-Space Model for Risk Factors MktRP and VvsGP

PROC STATESPACE requires that the series be stationary. The PROC ARIMA is used to identify time series R_{MktRP} and R_{VvsGP}, and the result is that the second differenced R_{MktRP} and R_{VvsGP} series are stationary. This is reasonable since single differencing is used to de-trend linear trends and double differencing is used to de-trend quadratic trends. These risk factors show quadratic trends and have to be doubly differenced to make them stationary.

Figure 3.5 shows descriptive statistics for each of the two risk factors: number of observations used for the analysis, the mean (after differencing), the standard deviation and the specified order of differencing.

$$\vec{x}_t = \begin{pmatrix} r_{MktRP,t} \\ r_{VvsGP,t} \end{pmatrix} = \begin{pmatrix} (1-B)^2 R_{MktRP,t} + 0.02126 \\ (1-B)^2 R_{VvsGP,t} + 0.24534 \end{pmatrix}, \tag{3.63}$$

where B represents the backshift operator.

```
            Number of Observations      522

                          Standard
Variable        Mean       Error
MktRP         -0.02126    40.14533    Has been differenced.
                                      With period(s) = 1,1.
VvsGP         -0.24534    268.1544    Has been differenced.
                                      With period(s) = 1,1.
```

Figure 3.5 Descriptive statistics for R_{MktRP} and R_{VvsGP}

Figure 3.6 shows preliminary estimates of the state-space model: the state vector $\vec{z}_t$, preliminary estimate of the transition matrix F, preliminary estimate of the input matrix G (an identity matrix), and preliminary estimate of the variance matrix Σ_{ee}.

Figure 3.7 shows the final estimates of the state-space model: the state vector $\vec{z}_t$, the final estimate of the transition matrix F, the final estimate of the input matrix G, and the final estimate of the variance matrix Σ_{ee}.

Figure 3.8 displays the parameter estimates, standard errors, and t-values of the F and G matrices.

Selected Statespace Form and Preliminary Estimates

State Vector

| MktRP(T;T) | VvsGP(T;T) | MktRP(T+1;T) | VvsGP(T+1;T) |

Estimate of Transition Matrix

0	0	1	0
0	0	0	1
0.076303	-0.0053	-0.13883	0.011249
0.385688	0.060963	-1.04448	0.134561

Input Matrix for Innovation

1	0
0	1
-1.07532	0.017088
-0.77694	-0.80731

Variance Matrix for Innovation

743.5361	3372.236
3372.236	36995.57

Figure 3.6 Preliminary estimates of the state-space model

Selected Statespace Form and Fitted Model

State Vector

| MktRP(T;T) | VvsGP(T;T) | MktRP(T+1;T) | VvsGP(T+1;T) |

Estimate of Transition Matrix

0	0	1	0
0	0	0	1
0.098524	-0.00693	-0.16869	0.011998
0.353225	0.073266	-1.02562	0.117066

Input Matrix for Innovation

1	0
0	1
-1.08607	0.018846
-0.73011	-0.80503

Variance Matrix for Innovation

840.4586	3363.17
3363.17	39514.18

Figure 3.7 Final estimates of the state-space model

Parameter Estimates

Parameter	Estimate	Standard Error	t Value
F(3,1)	0.098524	0.063184	1.56
F(3,2)	-0.00693	0.009033	-0.77
F(3,3)	-0.16869	0.056791	-2.97
F(3,4)	0.011998	0.008618	1.39
F(4,1)	0.353225	0.415197	0.85
F(4,2)	0.073266	0.060365	1.21
F(4,3)	-1.02562	0.369686	-2.77
F(4,4)	0.117066	0.057398	2.04
G(3,1)	-1.08607	0.052434	-20.71
G(3,2)	0.018846	0.007638	2.47
G(4,1)	-0.73011	0.342831	-2.13
G(4,2)	-0.80503	0.050998	-15.79

Figure 3.8 The parameter estimates of the state-space model

According to the PROC STATESPACE result, the state vector $\vec{z}_t$ is

$$\vec{z}_t = \begin{pmatrix} r_{MktRP,t|t} \\ r_{VvsGP,t|t} \\ r_{MktRp,t+1|t} \\ r_{VvsGP,t+1|t} \end{pmatrix}. \tag{3.64}$$

The transition matrix F is

$$F = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.0985 & -0.0069 & -0.1687 & 0.012 \\ 0.3532 & 0.0733 & -1.0256 & 0.1171 \end{pmatrix}. \tag{3.65}$$

The input matrix for innovation is

$$G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1.0861 & 0.01885 \\ -0.7301 & -0.8050 \end{pmatrix}. \tag{3.66}$$

The estimated state-space model is written as

$$\begin{pmatrix} r_{MktRP,t+1|t+1} \\ r_{VvsGP,t+1|t+1} \\ r_{MktRp,t+2|t+1} \\ r_{VvsGP,t+2|t+1} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.099 & -0.007 & -0.169 & 0.012 \\ 0.353 & 0.073 & -1.026 & 0.117 \end{pmatrix} \begin{pmatrix} r_{MktRP,t|t} \\ r_{VvsGP,t|t} \\ r_{MktRP,t+1|t} \\ r_{VvsGP,t+1|t} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1.086 & 0.019 \\ -0.73 & -0.81 \end{pmatrix} \begin{pmatrix} e_{t+1} \\ n_{t+1} \end{pmatrix},$$

$$\tag{3.67}$$

$$Var\begin{pmatrix} e_{t+1} \\ n_{t+1} \end{pmatrix} = \begin{pmatrix} 840.4586 & 3363.17 \\ 3363.17 & 39514.18 \end{pmatrix}.$$

$$\tag{3.68}$$

Figure 3.9 displays the forecasts for the R_{MktRP} and R_{VvsGP} time series. Table 3.6

lists the real values for R_{MktRP} and R_{VvsGP} and their corresponding forecast values.

State Space Model Forcasts

Obs	MktRP	MktRP_ f1	VvsGP	VvsGP_ f1
500	1221.84	1234.39	8380.12	8470.01
501	1225.95	1225.23	8443.40	8383.57
502	1215.29	1229.05	8423.67	8448.26
503	1200.72	1219.03	8276.52	8443.75
504	1213.73	1200.22	8306.30	8256.52
505	1237.08	1212.57	8503.05	8275.47
506	1233.36	1241.21	8481.59	8529.26
507	1210.78	1238.52	8302.75	8521.12
508	1224.24	1212.44	8417.01	8292.76
509	1191.33	1224.23	8201.96	8398.35
510	1181.86	1193.22	8222.51	8204.73
511	1174.98	1179.37	8133.01	8184.98
512	1193.64	1170.98	8312.89	8101.80
513	1215.29	1192.65	8361.33	8302.09
514	1229.97	1215.21	8483.57	8373.02
515	.	1234.96	.	8514.96
516	.	1239.36	.	8548.97
517	.	1243.54	.	8573.66
518	.	1247.62	.	8597.84
519	.	1251.81	.	8621.62
520	.	1256.03	.	8645.56
521	.	1260.31	.	8669.87
522	.	1264.65	.	8694.56
523	.	1269.05	.	8719.67
524	.	1273.51	.	8745.20

Figure 3.9 Forecasts of R_{MktRP} and R_{VvsGP} time series over 10 weeks

They begin at 1234.96 for R_{MktRP} and 8514.96 for R_{VvsGP} for the first future week,

then they increase each week and reach 1273.51 for R_{MktRP} and 8745.20 for R_{VvsGP} in the

tenth future week.

Table 3.6 Comparison of real and forecast values for the R_{MktRP} and R_{VvsGP} time series over 10 weeks

R_{MktRP}	R_{MktRP_f}	R_{VvsGP}	R_{VvsGP_f}
1243.58	1234.96	8539.26	8514.96
1263.59	1239.36	8668.61	8548.97
1260.36	1243.54	8604.14	8573.66
1254.63	1247.62	8521.85	8597.84
1262.67	1251.81	8623.11	8621.62
1264.11	1256.03	8633.85	8645.56
1243.74	1260.31	8512.18	8669.87
1280.88	1264.65	8653.69	8694.56
1283.08	1269.05	8642.83	8719.67
1256.96	1273.51	8419.69	8745.20

CHAPTER 4

FACTOR MODELS, REGRESSION ANALYSIS IN

VaR MAPPING PROCEDURE

4.1 Introduction

There are six sections of this chapter. This section is an introduction—a review of mapping procedure in VaR measure. The second section is a discussion about the factor models. The third section is an overview about regression analysis techniques. The fourth section is a description on the selection of the risk factors, stocks, and the data source in this research. The fifth section is a detailed discussion about the selection of the time series regression models. The last section is the discussion of results.

Figure 4.1 is a reproduction of Figure 2.1 which shows how VaR measures work [21]. All practical VaR measures accept portfolio data and historical market data as inputs. They process these with a mapping procedure, inference procedure, and transformation procedure. Output comprises the value of a VaR metric. That value is the VaR measurement. The position of mapping procedure is highlighted in Figure 4.1.

95
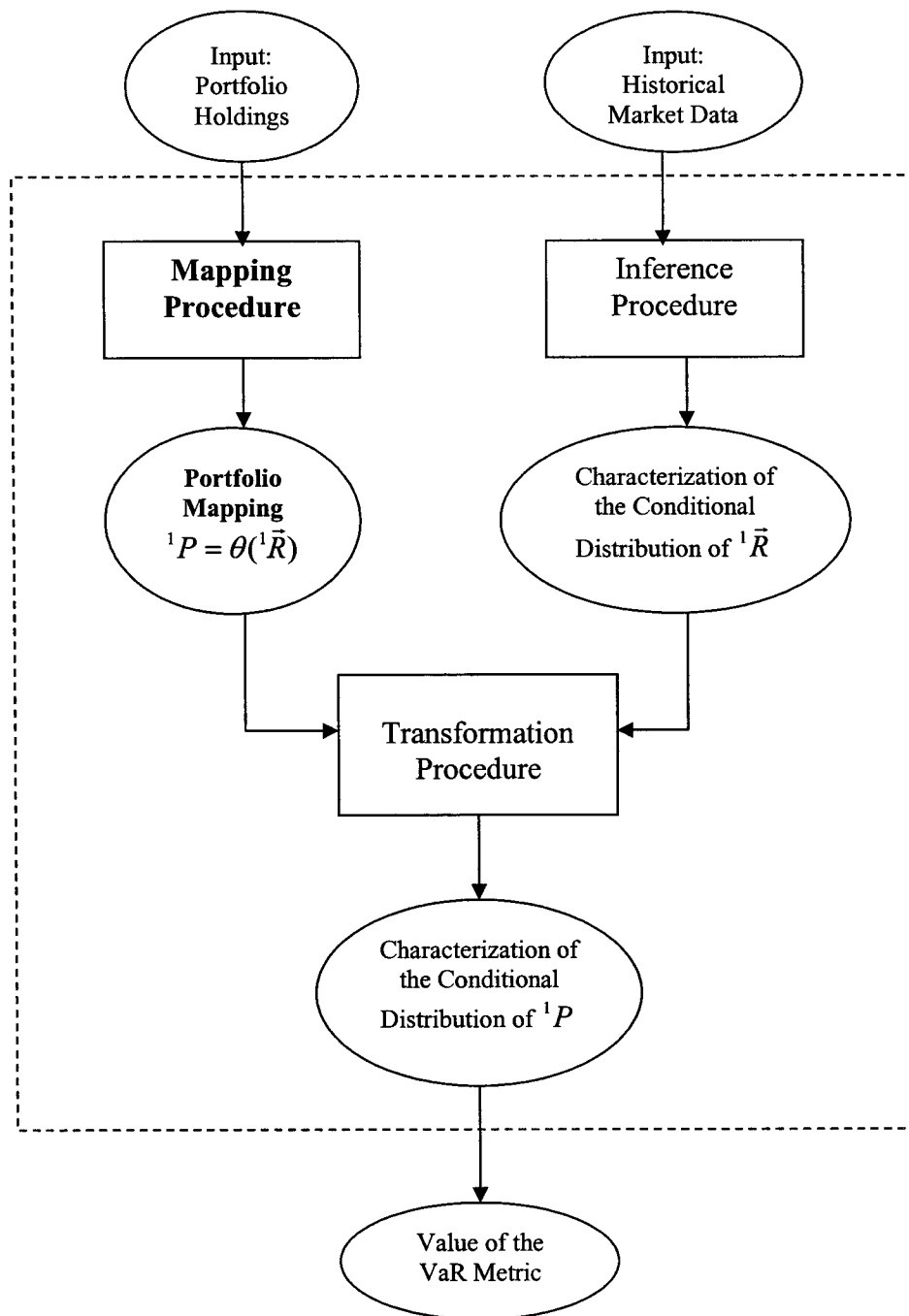
Figure 4.1 A reproduction of Figure 2.1 [21]

Risk has two components [21], uncertainty and exposure. An inference procedure describes uncertainty by characterizing the joint distribution for $^1\vec{R}$. A mapping procedure describes exposure by specifying a portfolio mapping. A transformation procedure describes risk by combining exposure and uncertainty to characterize the distribution of ^1P and then use the distribution of ^1P to summarize the value of some VaR metric.

The term "risk factor" was first coined by the heart researcher Dr. Thomas R. Dawber in a landmark scientific paper in 1961[40], where he isolated the major factors associated with heart disease: high blood pressure, high cholesterol levels, and smoking.

Generally, a risk factor is defined as something that is likely to increase the chances that a particular event will occur. However, risk factors are not necessarily causal. In VaR, a risk factor is a random variable whose value will affect the value of a portfolio.

Let time 0 be the current time, and let time 1 be the end of the VaR horizon (Value-at-risk considers a portfolio's performance over a specific horizon—a trading day, two weeks, a month, etc. This is called the VaR horizon.). A risk factor is any random variable ^1R_i whose value will be realized during the interval $(0,1]$ and will affect the market value of a portfolio at time 1. A risk vector $^1\vec{R}$ is a random vector of risk factors. If a risk vector reflects a future value of some time series, its current value is represented as $^1\vec{r}$ and historical values are represented as $^0\vec{r}, {}^{-1}\vec{r}, {}^{-2}\vec{r}, \ldots$. Those risk factors ^1R_i are called key factors. They are the components of the key vector $^1\vec{R}$. Occasionally, we use asset values ^1S_i as key factors. More often, it is convenient to use more basic financial variables as key factors [21].

A mapping procedure accepts a portfolio's composition as an input. Its output is a portfolio mapping function θ that defines ^1P as a function of $^1\vec{R}$.

For each asset, there must exist a valuation function φ_i such that

$$^1S_i = \varphi_i(^1\vec{R}).$$ (4.1)

If the value of the portfolio ^1P is a linear polynomial of the asset values ^1S_i, we can express ^1P in terms of the key factors:

$$^1P = \sum_{i=1}^{m} \omega_i \, ^1S_i = \sum_{i=1}^{m} \omega_i \varphi_i(^1\vec{R}),$$ (4.2)

where ω_i is the weight of asset ^1S_i.

This is a functional relationship that specifies the portfolio's market value ^1P in terms of the key factors $\{^1R_i\}$. Combining Equations (4.1) and (4.2), we have

$$^1P = \theta(^1\vec{R}),$$ (4.3)

which is called a portfolio mapping. The function θ is called the portfolio mapping function. Schematically, the mapping can be represented as

$$^1P \xleftarrow{\quad \vec{\omega} \quad} {}^1\vec{S}$$ (4.4)

and

$$\overbrace{{}^1P \xleftarrow{\quad \vec{\omega} \quad} {}^1\vec{S} \xleftarrow{\quad \vec{\varphi} \quad} {}^1\vec{R}}^{\theta}$$ (4.5)

where $\vec{\omega}$ is the portfolio's holdings.

The portfolio mapping relates ^1P to $^1\vec{R}$: the mapping function θ maps the n-dimensional space of the key factors to the one-dimensional space of the portfolio's market value. Given a realization for $^1\vec{R}$, θ gives us the corresponding value of ^1P. The

mapping function θ could be linear or non-linear. Specifying θ is largely an exercise in financial engineering. Since θ must value an entire portfolio, it can be complicated.

Portfolio mappings of Equations (4.4) or (4.5) are called primary mappings. Mathematically, a primary mapping works by valuing each asset held by the portfolio, multiplying the values by the portfolio's holdings in each asset, and summing. For many VaR measures, output of the mapping procedure is a primary mapping, but not for all. Since the portfolio in this research is a portfolio of three stocks and three risk factors, primary mapping is appropriate for this portfolio.

The purpose of a mapping procedure is to characterize a portfolio's exposure to risk by expressing the portfolio's value as a function of appropriate market variables – risk factors. This can be done by building multi-factor models.

A multi-factor model is a method of decomposing an asset's return into factors common to all assets and an asset specific factor. Often the common factors are interpreted as capturing fundamental risk components, and the multi-factor model isolates an asset's sensitivities to these risk factors.

Multifactor models are widely used in three ways [42]: (1) predicting future volatility and understanding their risk exposures, (2) optimizing portfolios, and (3) understanding past performance.

A multifactor model is an important tool that helps portfolio managers identify, quantify, and control their risk exposures [42] and makes portfolio risk and yield evaluations precise, quick, and reliable. The reasons are [43]

- a number of factors is taken into account during the asset yield evaluation;

- in addition to the standard method of regression analysis, multidimensional data analysis is used to research how the price of every asset reacts to each factor change;

- assets are grouped by a similar risk factor with cluster analysis methods;

- the model takes into account the lag effect. A change in macroeconomic parameters may influence asset price with a time lag of days to months;

- risk factors, such as, for example, market index and macroeconomic parameters, are forecasted within the timeframe required for a portfolio or asset future yield evaluation;

- the model dynamically takes into account each factor of influence, based on its forecast.

In Chapter 3, we focused on inference procedure. We applied time series techniques, ARIMA and state-space models, to forecast the joint distribution of the three risk factors based on their historical data. In this chapter, we'll focus on mapping procedure to build a multi-factor model for our stock portfolio in terms of three risk factors.

4.2 Multi-Factor Models for Market Risk

The Bank for International Settlement (BIS) defines market risk as [44] *"the risk of loss in on- and off-balance-sheet positions arising from movements in market prices."* The major factors contributing to market risk are equity, interest rate, foreign exchange, and commodity risk. In addition to market risk, the price of a financial instrument may be influenced by the following residual risk: spread risk, basis risk, specific risk and

volatility risk. To determine the total price risk of a financial instrument, market risk and residual risk have to be aggregated.

4.2.1 Modern Portfolio Theory and VaR

Modern portfolio theory (MPT)—or portfolio theory—was introduced by Harry Markowitz with his paper "Portfolio Selection," which appeared in the 1952 *Journal of Finance*. It is the philosophical opposite of traditional stock picking. Prior to Markowitz's work, investors focused on assessing the risks and rewards of individual stocks in constructing their portfolios. Standard investment advice was to identify those stocks that offered the best opportunities for gain with the least risk and then construct a portfolio from these. Following this advice, an investor might conclude that stocks in a certain industry all offered good risk-reward characteristics and compile a portfolio entirely from these, which is against the common sense saying "Don't put all your eggs in one basket." One example is the high-tech bubble. A lot of people who invested in only high-tech stocks lost their fortune when the high-tech sector fell dramatically in 2001–2002. Markowitz formalized this intuition by detailing mathematics of diversification. He proposed that investors focus on selecting portfolios based on their overall risk-reward characteristics instead of only compiling portfolios from securities that each individually has attractive risk-reward characteristics. In other words, investors should select portfolios, not individual securities.

MPT helps us understand the market as a whole. Investments are described statistically in terms of their expected long-term return rate and their expected short-term volatility. The volatility is equated with "risk," measuring how much worse than average an investment's bad years are likely to be. The goal is to identify your acceptable level of

risk tolerance, and then to find a portfolio with the maximum expected return for that level of risk [45].

A portfolio is a collection of investments held by an institution or a private individual. Portfolio theory starts from the premise that investors choose between portfolios on the basis of their expected return and the standard deviation (or variance) of their return. The standard deviation of the portfolio return can be regarded as a measure of the portfolio risk,

$$\sigma_p^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq 1}}^{N} w_i w_j \rho_{ij} \sigma_i \sigma_j ,$$ (4.6)

where σ_p is the portfolio variance, σ_i is the variance of asset i, ρ_{ij} is the correlation coefficient of asset i and j, and w_i is the proportion of asset i in the portfolio. The portfolio return is expressed as,

$$\overline{R}_p = \sum_{i=1}^{n} w_i r_i$$ (4.7)

where $\overline{R}_p$ is the portfolio return and r_i is the return of asset i.

Equations (4.6) and (4.7) show that the portfolio return is the weighted average of the returns of the individual assets in the portfolio, and the portfolio risk is the weighted sum of the variances of the individual assets and the covariance between those assets. One of the key insights of portfolio theory is that risk is not additive. The risk of any individual asset contributing to the portfolio is not the standard deviation of the asset return, but the extent to which that asset contributes to overall portfolio risk. The lower the correlation to other assets, other things being equal, the less the asset contributes to overall risk. If the correlation is negative, it will offset existing risks and lower the

portfolio standard deviation. Total risk is less than the sum of its parts. This effect is described as diversification effect.

Portfolio optimization requires minimal risk with certain expected return. According to Equations (4.6) and (4.7), we can construct efficient portfolio frontier by varing weights of the assets w_i in the portfolio. There may be multiple portfolios that have the same standard deviation. Modern portfolio theory assumes that for a specified standard deviation, a rational investor would choose the portfolio with the greatest return. Similarly, there may be multiple portfolios that have the same return and modern portfolio theory assumes that, for a specified level of return, a rational investor would choose the portfolio having the lowest standard deviation. A portfolio is said to be efficient if there is no other portfolio having the same standard deviation with a greater expected return, and there is no portfolio having the same return with a lesser standard deviation. The efficient frontier is the collection of all efficient portfolios. Key to the efficient frontier is that it represents the highest level of a portfolio's return for any given level of risk. Figure 4.2 below displays a typical efficient frontier chart. The horizontal axis represents risk (typically in the form of standard deviation of return in dollars), while the vertical axis represents return (typically in the form of mean return expected of the portfolio). The line, which is the efficient frontier itself, represents the highest level of expected return for any given level of risk for any possible combination of assets in this portfolio. The area under the curve represents every other "non-efficient" portfolio that can exist.
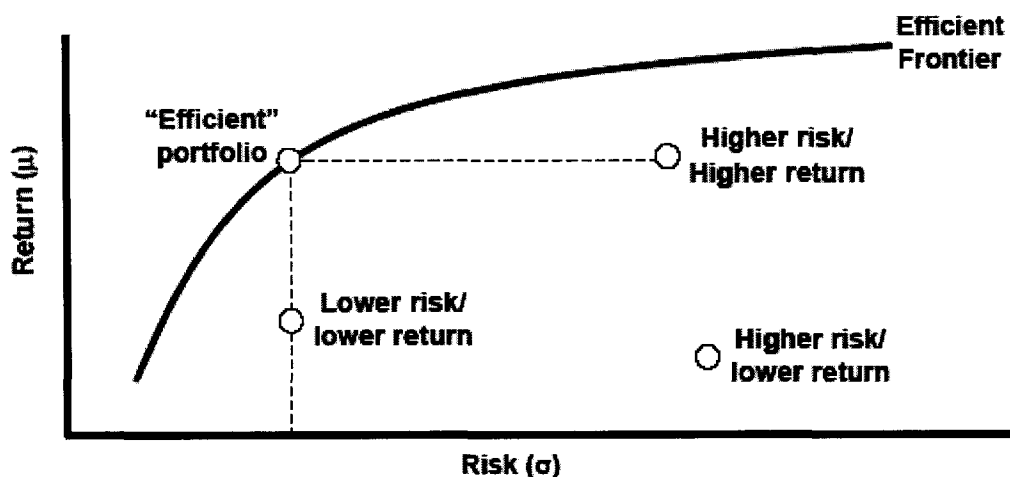
Figure 4.2 Efficient frontier diagram [46]

In practice, some efficient portfolios will have more risk than others, but the more risky ones will also have higher expected returns. Faced with the set of efficient portfolios, the investor chooses one particular portfolio based on his or her own preferred trade-off between risk and expected returns. An investor who is very averse to risk will choose a safe portfolio with a low standard deviation and a low expected return, and an investor who is less risk averse will choose a riskier portfolio with a higher expected return.

According to Equations (4.6) and (4.7), however, for a portfolio with N assets, portfolio optimization requires estimates of N expected returns, N standard deviations, and $N(N-1)/2$ correlations to perform mean-variance analysis. When N is increased to an extremely large number, not only the data-gathering procedure becomes difficult, but also the construction of the correlation matrix is an overwhelming burden. This is known as curse of dimensionality. To overcome these difficulties, factor models are used.

Portfolio theory provides a broad context for understanding the interactions of systematic risk and reward. It has profoundly shaped how institutional portfolios are

managed, and motivated the use of passive investment management techniques. The mathematics of portfolio theory is used extensively in financial risk management and was a theoretical precursor for today's value-at-risk measures.

Estimating the standard deviation $^{10}\sigma$ of the portfolio's market value is analogous to the task of estimating the standard deviation of a portfolio's return, a task one may be familiar with from portfolio theory. Except for the fact that VaR deals with market values instead of returns, we may adopt this familiar mathematics of portfolio theory for estimating VaR [46].

We use a general result from probability. Suppose $X_1, X_2 ...., X_n$ are random variables having standard deviations σ_i and correlations $\rho_{i,j}$. Suppose another random variable Y is defined as a linear polynomial of the X_i:

$$Y = b_1 X_1 + b_2 X_2 + \cdots + b_n X_n + a. \tag{4.8}$$

Then the standard deviation of Y is given by

$$\sigma_Y = \sqrt{\sum_i (b_i \sigma_i)^2 + 2 \sum_{j>i} (b_i \sigma_i)(b_j \sigma_j) \rho_{i,j}} \quad . \tag{4.9}$$

Equation (4.9) is completely general. So long as Y is a linear polynomial of the X_i, we can use Equation (4.9). We need no other assumptions or information about the random variables X_i.

We can apply Equation (4.9) to estimate the standard deviation $^{1|0}\sigma$ of our portfolio's market value. Suppose the portfolio has holdings ω_i in m assets. The assets' accumulated market values at time 1 are random variables, which we denote ^1S_i. Then

$$^1P = w_1{}^1S_1 + w_2{}^1S_2 + \cdots + w_m{}^1S_m. \tag{4.10}$$

Based on Equation (4.10), we can apply Equation (4.9) to obtain $^{1|0}\sigma$. All we need as inputs are standard deviations and correlations for the ^1S_i. These might be inferred by applying methods of time series analysis to historical price data for the assets. In some cases, this is feasible. In others, it is not. Collecting historical price data for every asset held by a portfolio may be a daunting task. Besides, it also has the same problems of MPT as we mentioned above since it is an extension of MPT. A more manageable approach may be to model the portfolio's behavior, not in terms of individual assets, but in terms of specific risk factors.

4.2.2 The Single-Factor Model

Portfolio theory provides a useful framework for handling multiple risks taking account of how those risks interact with each other. It is therefore of obvious use to—and is widely used by—portfolio managers, mutual fund managers and other investors. However, it tends to run into estimation and data problems. The estimation of the expected returns and the covariance matrix becomes difficult for a large portfolio. We also need a sufficiently long data set to make our statistical estimation techniques reliable. To get around these problems, William Sharpe [47] and others in the 1960s suggested Capital Asset Pricing Model (CAPM)—an equilibrium model that specifies the relationship between risk and required rate of return for assets held in well-diversified portfolios based on the premise that only one factor affects risk.

Capital Asset Pricing Model (CAPM) forms a theoretical basis for a number of various financial technologies for yield and risk management, used for medium and long-term investing in stocks. CAPM considers the stock yield depending on the market behavior as a whole. Another primitive conjecture of CAPM is that investors make decisions, taking into account only two factors: expected yield and risk. Though this model is a simplified presentation of the financial market, many large investment businesses use it in their activities: for example, Merrill Lynch and Value Line. The expected return of a specific stock is calculated as follows:

$$r_i - R_f = alpha_i + beta_i \times (R_m - R_f) + \varepsilon_i,$$ (4.11)

$$beta_i = Cov(r_i - R_f, R_m - R_f) / Var(R_m - R_f),$$ (4.12)

where r_i is the expected return of stock i, R_f is the rate of a "risk-free" investment (cash), R_m is the expected return of the market, $beta_i$ is the sensitivity of asset i relative to market movement m, and $alpha_i$ is the intercept.

The quality of the fit is given by the statistical number r-squared which measures the percent of a stock's variance that is explained by the market. An r-squared of 1.0 would mean that the model fit the data perfectly, with the line going right through every data point. More realistically, with real data one would get an r-squared of around .85. From that you would conclude that 85% of the fund's performance is explained by its risk exposure, as measured by beta.

According to Equation (4.11), the expected return of an individual stock can be expressed as

$$E(r_i - R_f) = alpha_i + beta_i \times E(R_m - R_f).$$ (4.13)

The risk of the stock i is

$$Var(r_i - R_f) = beta_i^2 \times Var(R_m - R_f) + Var(\varepsilon_i),$$ (4.14)

According to the model, the risk associated with stock i, could be broken into two types: systematic and non-systematic. Systematic risk is caused by the general market movements and economic changes influencing all the stocks. Non-systematic risk is associated with the specific stock i and can be reduced with creating a diversified portfolio from sufficiently great number of stocks, or even from a smaller number of stocks, negatively correlated among themselves. Therefore, non-systematic risk is also referred as diversifiable risk [47].

Systematic risk cannot be reduced, but influence of the market on financial assets' yield can be measured. CAPM uses the ß (beta) parameter as a systematic risk measure. ß describes the financial asset sensitivity to changes in market yield. Knowing asset's ß parameter, one can quantitatively estimate the amount of risk associated with price changes in the entire market. The more ß value, the higher the stock's price advances during general market growth, and vice versa. Stocks with greater positive b value drop lower when the entire market falls. Financial managers need a precise calculation of the ß parameter to select assets which match their investment strategy best. Using ß, one can form investment portfolios of most different types-conservative, aggressive, and balanced ones.

The CAPM made portfolio theory much more practical given the data and computational power available. However, it is a simplified view of risk and can not perfectly capture the movement of the stock market [41][44][47][48]. The logical generalization is the multi-factor models. A multi-factor takes into account other sources of risk that might affect the expected return.

4.2.3 Multi-Factor Models

4.2.3.1 Arbitrage pricing theory (APT)

APT assumes that a portfolio or asset yield (or expected return) can be modeled as a linear function of a number of systematic risk factors, where sensitivity to changes in each factor is represented by a factor specific beta coefficient. Such factors could be various macroeconomic parameters, prices of oil and other major resources, currency exchange rates, and so forth. APT yields a multi-factor equation:

$$R_{it} - R_t^f = c_i + \beta_{im}(R_{mt} - R_t^f) + \beta_{iA}F_{At} + \beta_{iB}F_{Bt} + \ldots + e_{it}, \qquad (4.15)$$

where $t = 1,2,\ldots,T$ and $i = 1,2,\ldots,N$. Time-series multiple regression can be used to quantify an asset's tendency to move with multiple risk factors F_A, F_B, etc.

APT specifies neither the number of factors nor the identification of the factors. Therefore, to estimate and test the model, we need to determine the factors which may be observed and unobserved.

4.2.3.2 Fama-French three-factor model

CAPM uses a single factor, beta, to compare a portfolio with the market as a whole. But more generally, you can add factors to a regression model to give a better r-squared fit. The best known approach in this regard is the three factor model developed by Fama and French [49].

Fama and French noted that, averaging over many years, big firms appeared to have lower returns than small firms. Therefore, they decided to include a variable that would account for the difference in the market equity size, namely the difference between the return on a portfolio of small stocks minus the return on a portfolio of big stocks. This variable is called R_{SMB}, for S minus B. In addition, since there are differences between what the market thinks a company is worth (market equity) and the amount accountants say that a company is worth (book equity), Fama and French also included another variable to represent the Book-to-Market-Equity, namely the return on a portfolio of high book-to-market stocks (customarily called "value" stocks), minus the return on a portfolio of low book-to-market stocks (called "growth" stocks). This variable is called R_{HML}, for H minus L.

The Fama-French three-factor model is an extension of CAPM by adding these two factors to CAPM:

$$R_{stock} = R_f + \beta * (R_m - R_f) + \beta_{SIZE} * (R_{SMB}) + \beta_{BM} * (R_{HML}) \qquad (4.16)$$

According to the definitions of *SMB* and *HML*, the corresponding coefficients β_{SIZE} and β_{BM} take values on a scale of roughly 0 to 1: β_{SIZE} = 1 would be a small cap portfolio, β_{SIZE} = 0 would be a large cap, β_{BM} = 1 would be a portfolio with a high book/price ratio, etc.

Like CAPM, the Fama and French model is used to explain the performance of portfolios via linear regression; only now the two extra factors give you two additional axes, so instead of a simple line the regression is in the fourth dimension.

4.2.3.3 General multi-factor models

The general form of the multi-factor model is [50]

$$S_{it} = \alpha_i + \beta_{1i}R_{1t} + \beta_{2i}R_{2t} + \cdots + \beta_{Ki}R_{Kt} + \varepsilon_{it} \tag{4.17}$$

$$= \alpha_i + \vec{\beta}_i'\vec{R}_t + \varepsilon_{it}, \tag{4.18}$$

where S_{it} = return or excess return on asset i ($i = 1, 2, \ldots, n$)

$R_{kt} = k^{th}$ common risk factor ($k = 1, 2, \ldots, K$)

β_{ki} = factor loading on asset i for k^{th} factor

ε_{it} = random error for asset i

4.2.3.4 Risk factor selection

Multi-factor models are important tools in portfolio optimization and risk management. Selecting risk factors is a key stage in developing any multi-factor model. It is important to realize that Equation (4.18) does not require any assumptions about the nature of the factors. Any set of factors will fit into to Equation (4.18)–of course, with varying success, as measured by r-squared [42].

There are two approaches to characterize risk factors. The first one is to pre-specify risk factors from market data–from fundamental factors such as size and book-to-market to macroeconomic factors, such as treasury interest rates and changes in industrial production. For the multi-factor models based on pre-specified risk factors, factor variables R_{kt} are observed but factor loadings β_{ki} must be estimated. The second one is to compute the factors from asset historical data using statistical methods such as Principal Components Analysis. For the models based on the statistical factors, both factor variables R_{kt} and factor loadings β_{ki} must be estimated.

No matter which approach is chosen to select risk factors, there is a guiding principle: the number of factors should be small, usually 3 to 5.

4.3 Methodology of Multi-Factor Regression

4.3.1 Overview on Regression Analysis

In statistics, regression analysis is used to model the relationship between a random variable Y called response variable (sometimes called the output variable or dependent variable) and other random variables $X_1, X_2, \cdots, X_p$ called predictors (sometimes called the input variables, independent variables or explanatory variables). A regression analysis has several objectives:

- Prediction of future observations

- Determining how closely the response can be predicted by the predictors

- Assessing the relationship between the predictors

When building a regression model, if more than one independent variable is being considered, we call it a multiple regression analysis, if only one independent variable is being considered, the analysis is a simple linear regression.

4.3.2 Simple Linear Regression

A simple regression model is a regression model with a single independent variable x that has a linear relationship with a dependent variable y [51]. The model is

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{4.19}$$

where β_0 is the intercept, β_1 is the slope and ε is the error term. The errors are assumed to have mean zero and unknown variance σ^2.

To determine the equation of the model, what we are looking for is the values of β_1 and β_0. The method used for that purpose is called the least squares method. Suppose that we have n pairs of sample data $(y_1,x_1),(y_2,x_2),\ldots,(y_n,x_n)$, from Equation (4.18) we can write the sample regression model as

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i = 1,2,\ldots,n.$$ (4.20)

The least-squares estimators for β_1 and β_0 are obtained from the following Equation [49]:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$ (4.21)

and

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} y_i x_i - \dfrac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}},$$ (4.22)

where $\bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ are the averages of y_i and x_i, respectively. $\hat{\beta}_1$ can be rewritten as:

$$\hat{\beta}_1 = \frac{Cov(X,Y)}{Var(X)}.$$ (4.23)

Simple linear regression method can be used to estimate the coefficients of a single-factor model such as CAPM.

4.3.3 Multiple Regression

Multiple regression is a regression analysis with more than one independent variable. It is used in multi-factor models to quantitatively measure the impact of two or more independent variables. The general multiple linear regression model with k independent variables can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad (4.24)$$

where $\beta_j, j = 0,1,2,\ldots,k$ are called the regression coefficients and the error term satisfies $E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2$. The parameter β_j represents the expected change in the response y per unit change in x_j when all the other independent variables $x_i (i \neq j)$ are held constant [51].

The least squares method can be used to estimate the regression coefficients in Equation (4.24). Suppose there are $n > k$ observations, the sample regression model corresponding to Equation (4.24) can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

$$= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i, \qquad i = 1,2,\ldots,n . \qquad (4.25)$$

In matrix notation, the model given by Equation (4.25) is

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon} ,$$

where

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix},$$

$$\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \text{and} \quad \vec{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

The least-squares criterion is

$$S(\vec{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \vec{\varepsilon}'\vec{\varepsilon} = (\vec{y} - X\vec{\beta})'(\vec{y} - X\vec{\beta}). \tag{4.26}$$

The objective of multiple linear regression is to find the vector of least squares estimators, $\hat{\vec{\beta}}$, that minimizes the sum of the squared error terms, $S(\vec{\beta})$.

The least squares estimator of $\vec{\beta}$ is [51]

$$\hat{\vec{\beta}} = (X'X)^{-1} X'\vec{y} \tag{4.27}$$

provided that the inverse matrix $(X'X)^{-1}$ exists.

The fitted regression model corresponding to the levels of the independent variables $\vec{x}' = [1, x_1, x_2, \ldots, x_k]$ is

$$\hat{y} = \vec{x}'\hat{\vec{\beta}} = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_j. \tag{4.28}$$

The vector of fitted values $\hat{y}_i$ corresponding to the observed values y_i is

$$\hat{\vec{y}} = X\hat{\vec{\beta}} = X(X'X)^{-1} X'\vec{y} = H\vec{y} \tag{4.29}$$

where the $n \times n$ matrix $H = X(X'X)^{-1} X'$ is called the hat matrix.

The difference between the observed value y_i and the corresponding fitted value

$\hat{y}_i$ is called the residual e_i. In matrix notation, the n residuals can be written as

$$\vec{e} = \vec{y} - \hat{\vec{y}}$$

or
$$\vec{e} = \vec{y} - X\hat{\beta} = \vec{y} - H\vec{y} = (I - H)\vec{y} \qquad (4.30)$$

where I is a identity matrix.

4.4 Data Selection

All the data can be downloaded from Yahoo ! Finance. Data are also available

from some standard data libraries, such as COMPUSTAT.

Three risk factors are selected. They are R_{MRP}, R_{MktRP} and R_{VvsGP}. Details are in

Chapter 3.

Three stocks are chosen to form a portfolio: 1. International Business Machines

Corp. (IBM), 2. Microsoft Corp. (MSFT), and 3. Texas Instruments Inc. (TXN). Ten-year

weekly prices for these stocks are collected. They are denoted as S_{IBM}, S_{MSFT} and S_{TXN}.

Here, the price is not an open or closed price. The price is the adjusted closed price

because adjusted price is the price after adjustment for dividends and more importantly,

stock split [48]. Figures (4.3–4.5) are the trend charts for these three stocks from Jan. 2,

1996 to Jan. 17, 2006.
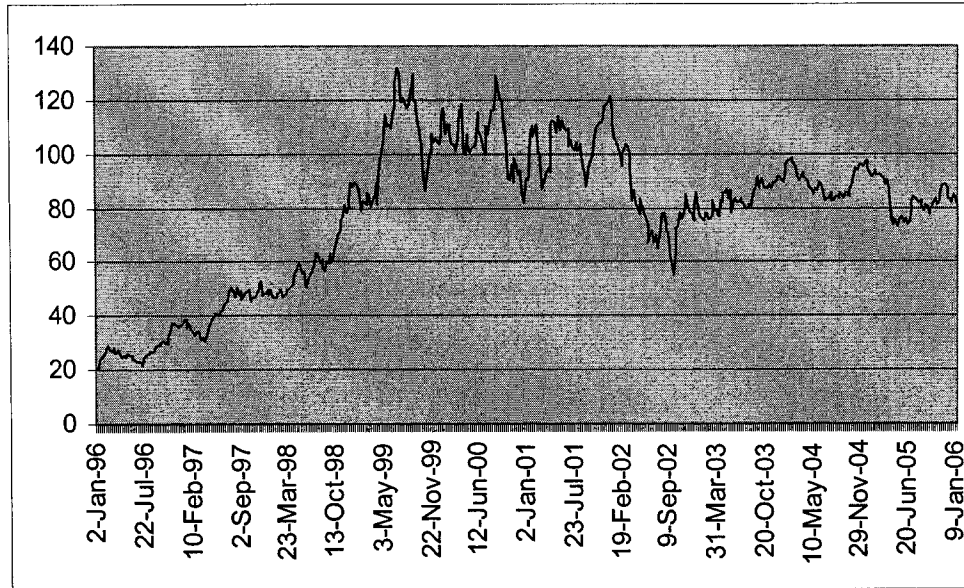
Figure 4.3 IBM (S_{IBM}) 96–06 weekly historical prices
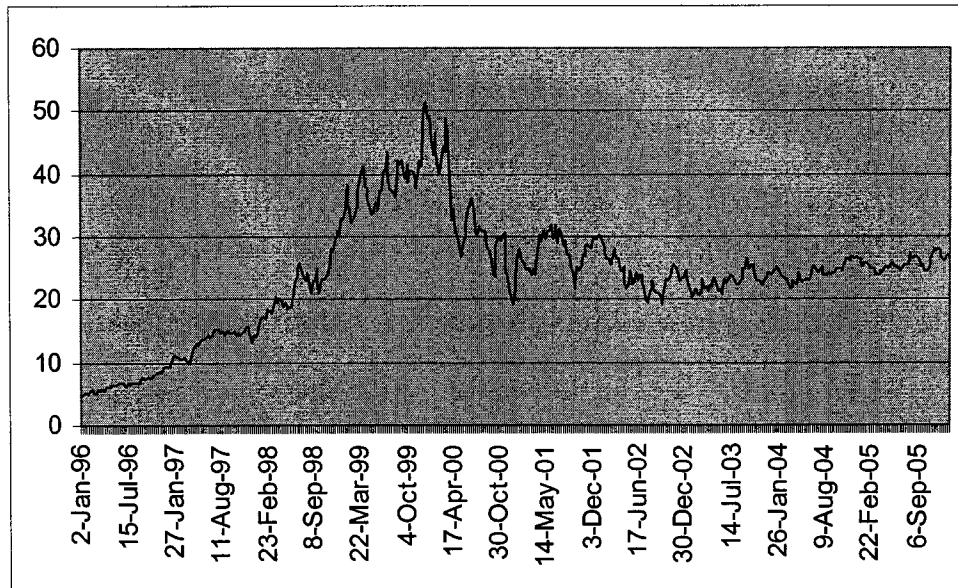
Figure 4.4 MSFT (S_{MSFT}) 96–06 weekly historical prices
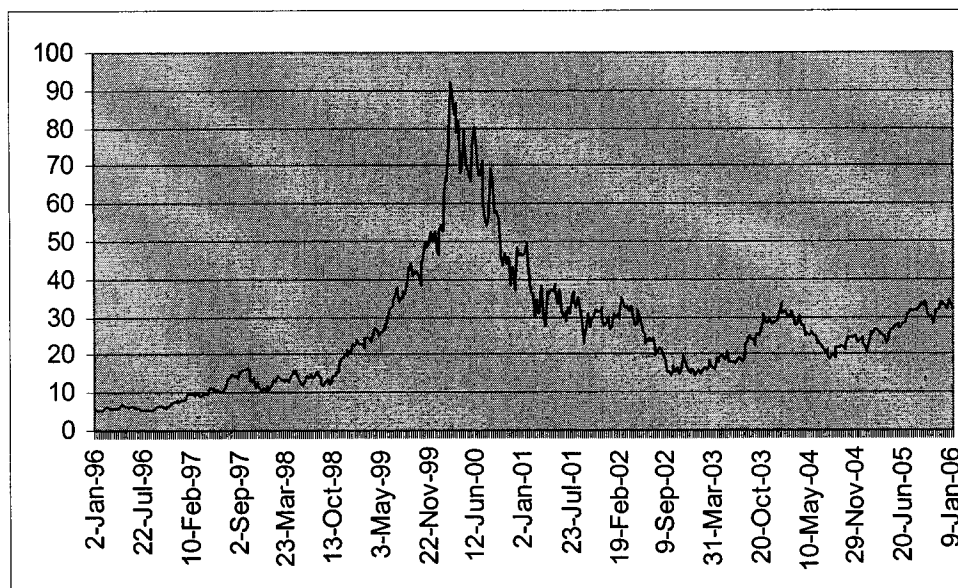
Figure 4.5 TXN (S_{TXN}) 96–06 weekly historical prices

4.5 Model Building

Regression analysis is one of the most widely used techniques for analyzing

multifactor data. The application of regression analysis in stock market is a statistical

technique used to forecast and to analyze the factors that influence the stock market. By

using the "multiple linear regression", studies have been performed in obtaining the best

regression model to do forecasting. The common types of "multiple linear regression" to

be studied here would be estimation of the model parameters, hypothesis testing and

confidence intervals.

Since building the risk management model needs heavy statistical computations,

SAS, a popular and powerful statistical software, is used as the tool to develop the model

in this research.

The following steps are needed to build the model:

1. Build linear models by least-squares method.

2. Build quadratic models by least-squares method.

    1) Build full models by least-squares method.

    2) Build reduced models by stepwise procedures: (1) forward selection, (2) backward elimination and (3) stepwise regression.

Before building our models, some tests should be run to investigate the relationships between each stock and each risk factor.

4.5.1 Checking for Relationships

A scatter plot reveals relationships between two variables. Such relationships manifest themselves by any non-random structure in the plot. Figures 4.6 - 4.14 display the relationships between each stock and each risk factor.
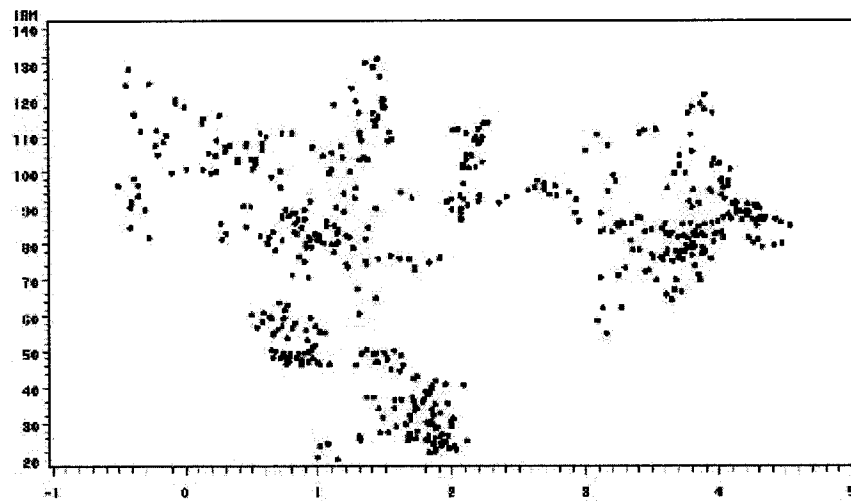
Figure 4.6 Scatter diagram for S_{IBM} vs. R_{MRP}
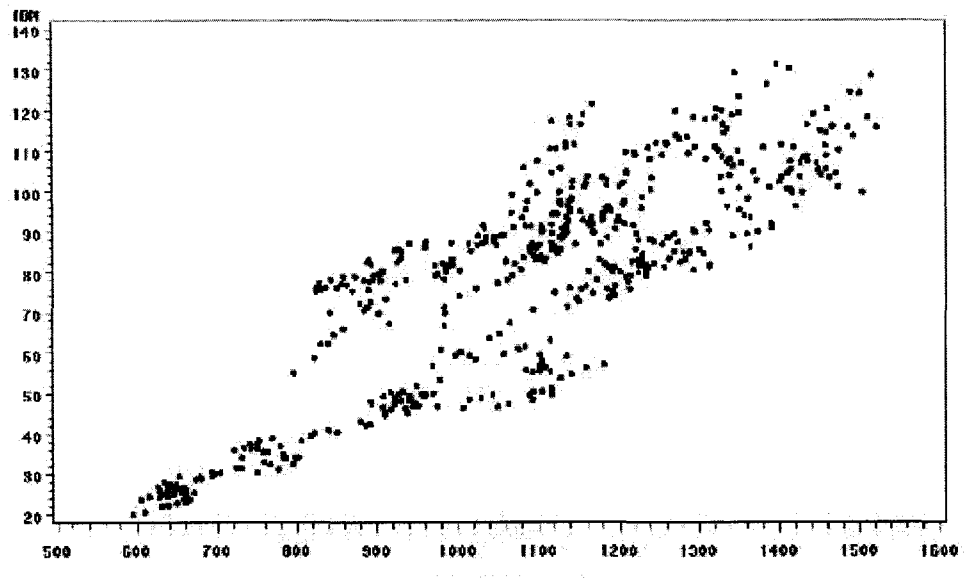
Figure 4.7 Scatter diagram for S_{IBM} vs. R_{MktRP}
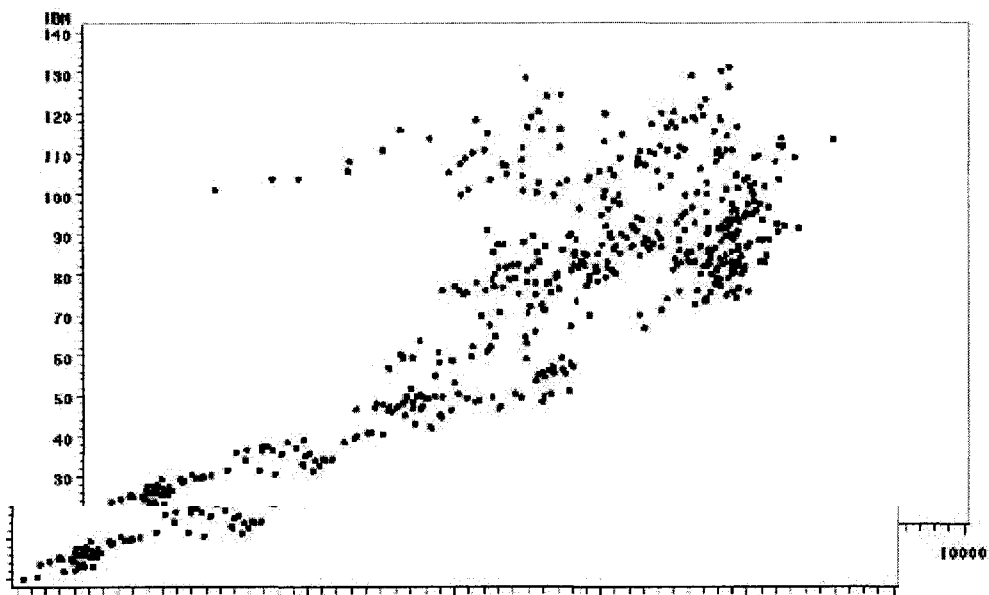
Figure 4.8 Scatter diagram for S_{IBM} vs. R_{VvsGP}
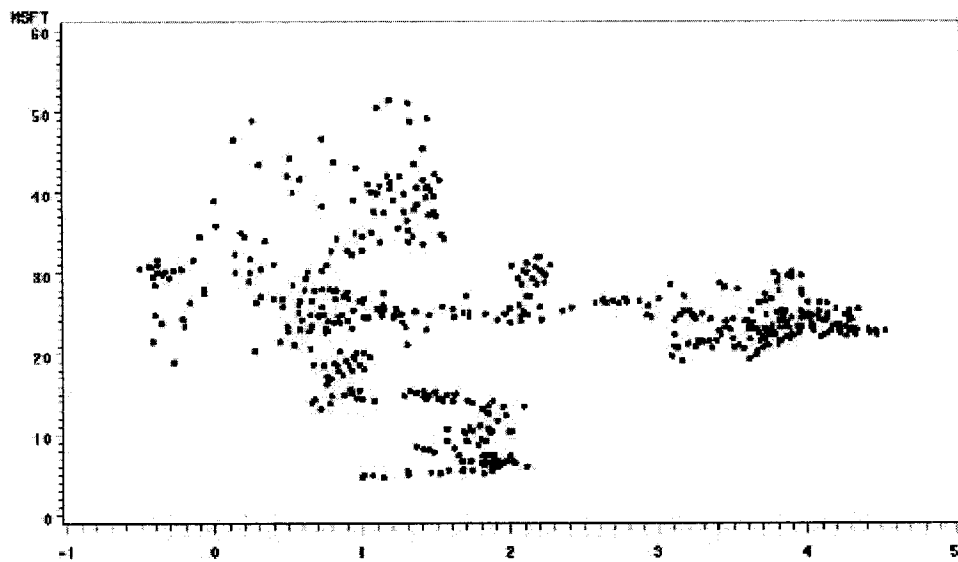
Figure 4.9 Scatter diagram for S_{MSFT} vs. R_{MRP}
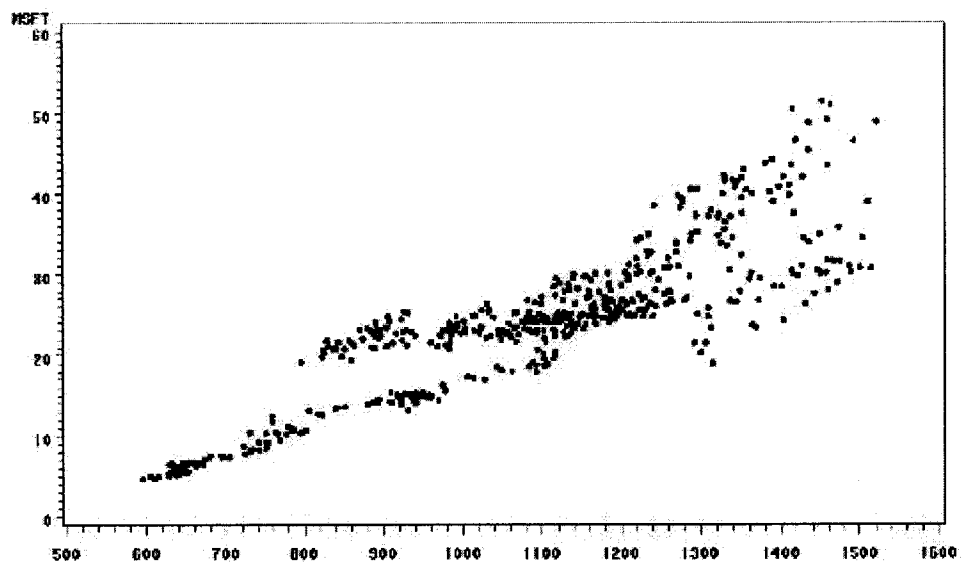
Figure 4.10 Scatter diagram for S_{MSFT} vs. R_{MktRP}
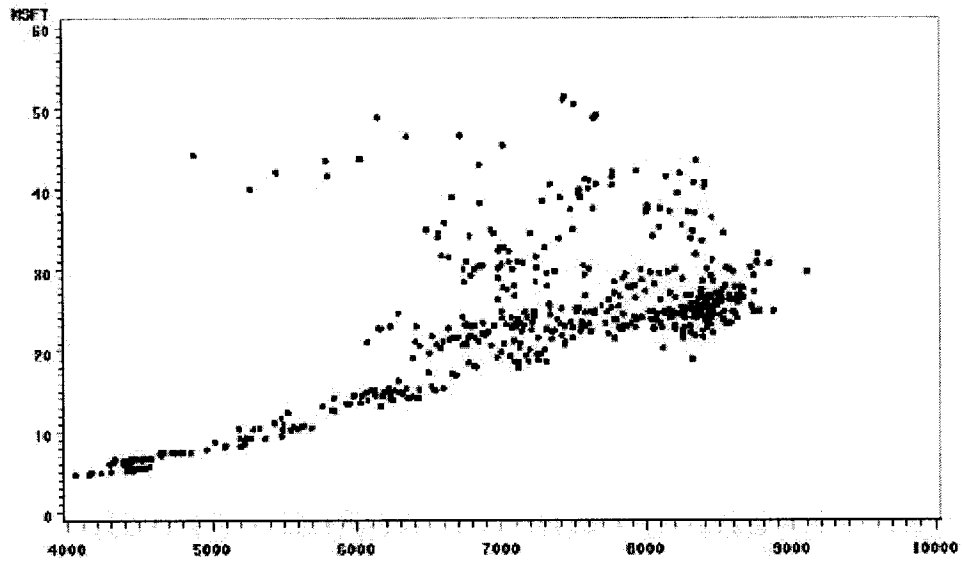
Figure 4.11 Scatter diagram for S_{MSFT} vs. R_{VvsGP}
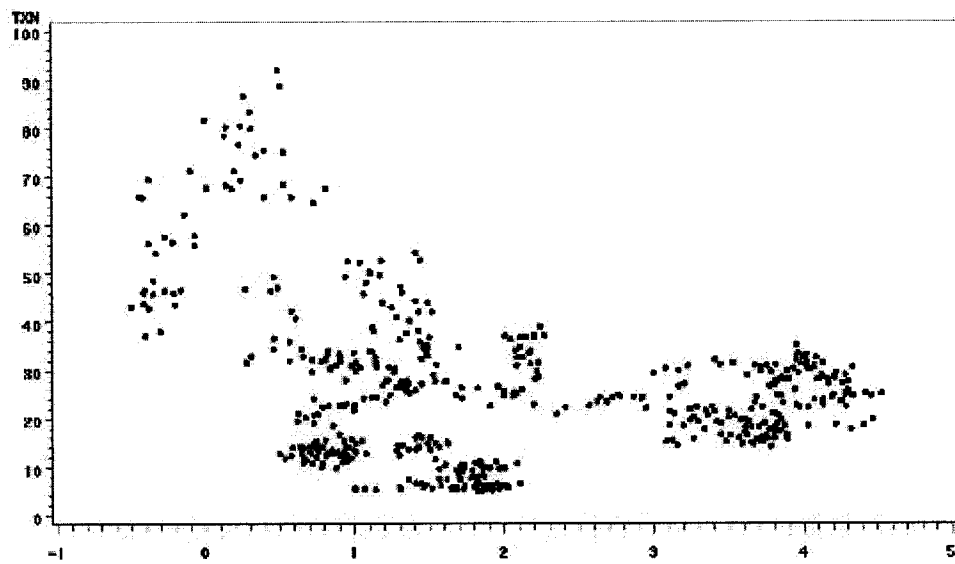
Figure 4.12 Scatter diagram for S_{TXN} vs. R_{MRP}
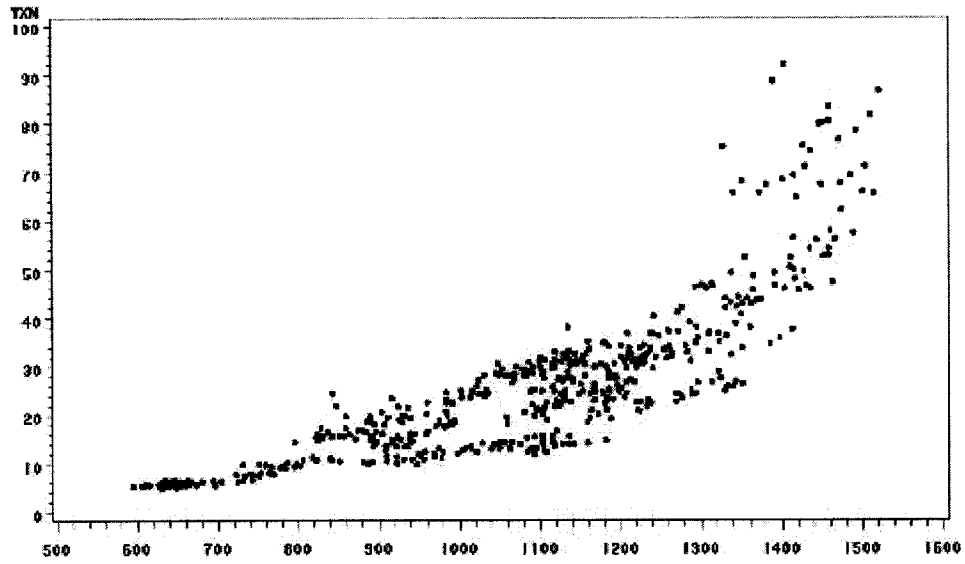
Figure 4.13 Scatter diagram for S_{TXN} vs. R_{MktRP}
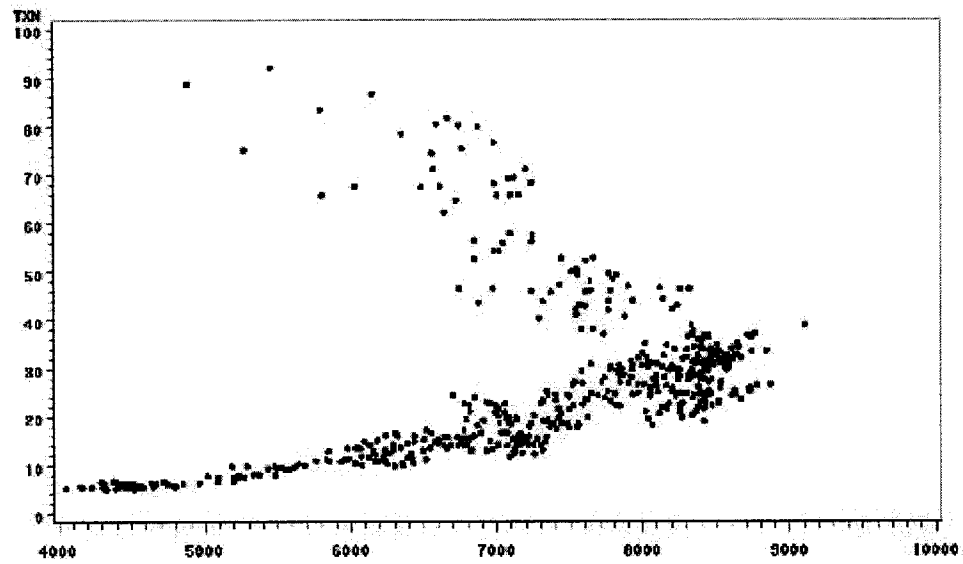
Figure 4.14 Scatter diagram for S_{TXN} vs. R_{VvsGP}

Table 4.1 summarizes our observation of trends (or correlation) from the above scatter plots.

Table 4.1 Summarization of the relationships between stocks and risk factors.

	R_{MRP}	R_{MktRP}	R_{VvsGP}
S_{IBM}	No clear tendency	correlated	correlated
S_{MSFT}	No clear tendency	correlated	correlated
S_{TXN}	Negative nonlinear correlated	correlated	correlated

4.5.2 Building Linear Models

The equation for three-factor linear regression models is

$$S_{stock} = \beta_0 + \beta_{MRP} \times R_{MRP} + \beta_{MktRP} \times R_{MktRP} + \beta_{VvsGP} \times R_{VvsGP} + \varepsilon \qquad (4.31)$$

where R_{MRP}, R_{MktRP}, and R_{VvsGP} represent the risk factor MRP, MktRP and VvsGP respectively, and β_{MRP}, β_{MktRP} and β_{VvsGP} are the stock's exposure to the risk factor MRP, MktRP and VvsGP respectively.

Ten-year weekly historical data for three stocks and three risk factors are used to fit the regression model in Equation (4.28) by least-squares method which is described in Section 4.3.2. The sample regression model corresponding to Equation (4.28) for each stock can be written as

$$S_{stock,i} = \beta_0 + \beta_{MRP} \times R_{MRP,i} + \beta_{MktRP} \times R_{MktRP,i} + \beta_{VvsGP} \times R_{VvsGP,i} + \varepsilon_i, \qquad (4.32)$$

where $i = 1,2,\ldots,524$ represents 524 observations for each stock and each risk factor.

To express the three-factor linear regression model, Equation (4.29), in matrix terms, we need to define the following matrices:

$$\vec{S}_{stock} = \begin{bmatrix} s_{stock,1} \\ s_{stock,2} \\ \vdots \\ s_{stock,524} \end{bmatrix} \quad (4.33a) \qquad R = \begin{bmatrix} 1 & r_{MRP,1} & r_{MktRP,1} & r_{VvsGP,1} \\ 1 & r_{MRP,2} & r_{MktRP,2} & r_{VvsGP,2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & r_{MRP,524} & r_{MktRP,524} & r_{VvsGP,524} \end{bmatrix} \quad (4.33b)$$

$$\vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_{MRP} \\ \beta_{MktRP} \\ \beta_{VvsGP} \end{bmatrix} \quad (4.33c) \qquad \vec{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{524} \end{bmatrix} \quad (4.33d)$$

Figure 4.15 presents a portion of the computer output from SAS PROC REG for S_{IBM} in our portfolio.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	337983	112661	1310.62	<.0001
Error	520	44699	85.95978		
Corrected Total	523	382682			

Root MSE	9.27145	R-Square	0.8832	
Dependent Mean	78.03670	Adj R-Sq	0.8825	
Coeff Var	11.88088			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	77.97910	0.40503	192.52	<.0001	0
MRP	1	9.38579	0.46865	20.03	<.0001	2.40032
MktRP	1	0.12977	0.00374	34.69	<.0001	4.18659
VvsGP	1	-0.00161	0.00065719	-2.44	0.0149	3.93103

Figure 4.15 SAS PROC REG output for S_{IBM} in linear model

The F value of 1310.62 is used to test the null hypothesis $H_0 : \beta_{MRP} = \beta_{MktRP} = \beta_{VvsGP} = 0$.

The associated p value of < 0.001 leads to the rejection of the null hypothesis and indicated that at least one of the coefficients is not zero. Since all three p values

associated with three regression coefficients are less than 0.05, all three risk factors contribute significantly to the model. R-Square = 0.8832 means that a major portion (> 88.32%) of the variation in S_{IBM} is explained by variation in the independent variables of the model. The statistic Adj R-Sq = 0.8825 (an alternative to R-Square) is adjusted for the number of parameters in the model. The maximum VIF = 4.187 < 10 implies that there is no serious problems with multicollinearity in this model. The fitted linear model for S_{IBM} is

$$\hat{S}_{IBM} = 77.98 + 9.386 R_{MRP} + 0.130 R_{MktRP} - 0.0016 R_{VvsGP}.$$ (4.34)

A very effective way to investigate the adequacy of the fitted regression model and to check the underlying assumptions is to draw residual plots. Figure 4.16 is the plot of residuals versus predicted IBM price S_{IBM} and Figure 4.17 is the plot of studentized residual (R-student) versus predicted S_{IBM}. The figures show no specific error patterns exhibited by the residuals of the linear model, but it shows that larger residuals are associated with larger values of the response variable. Figure 4.18 shows that most of the points are close to a straight line but a few extreme points slightly deviate from the line. It may be inferred that the distribution is close to normal.
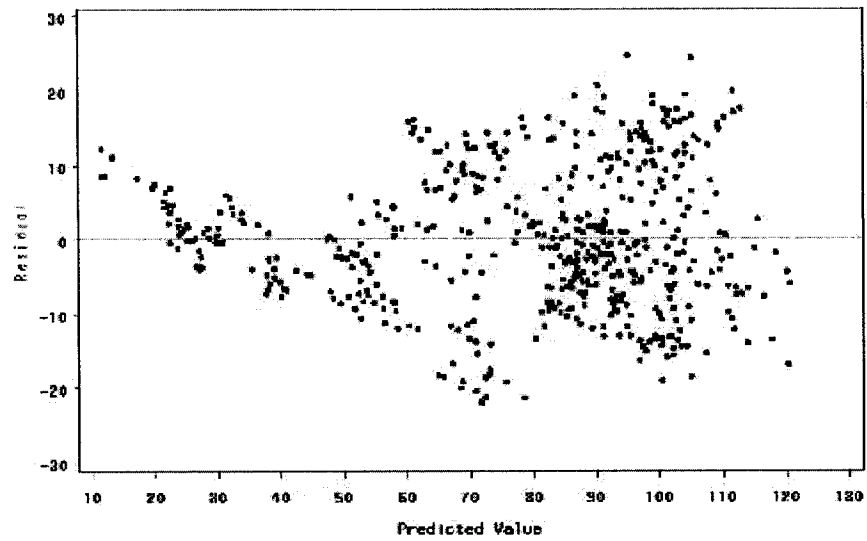
Figure 4.16 Plot of residuals vs. predicted values S_{IBM} in linear model
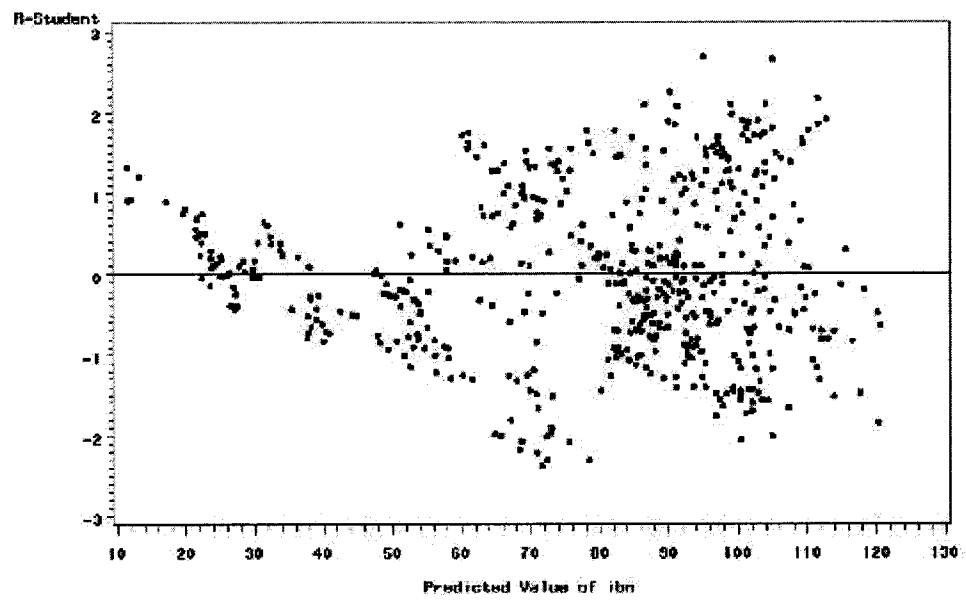
Figure 4.17 Plot of R-student values vs. predicted S_{IBM} in linear model
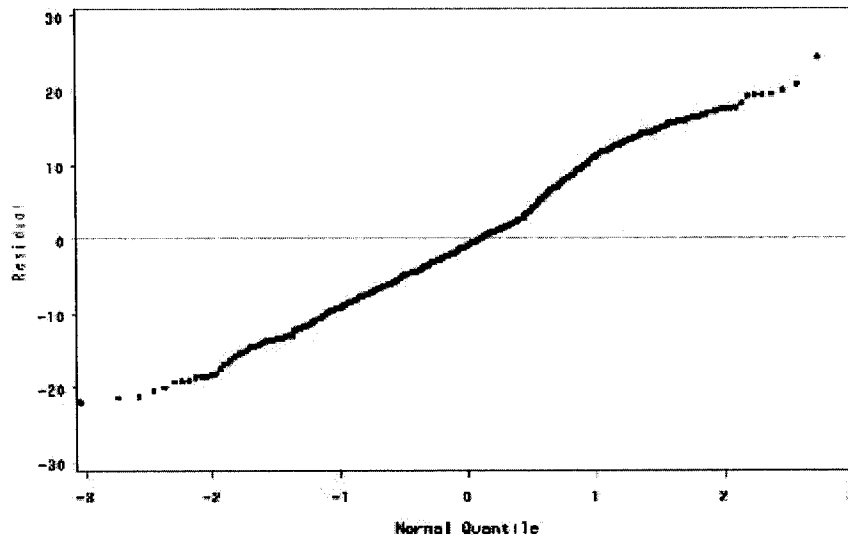
Figure 4.18 Normal probability plot for S_{IBM} in linear model

Figure 4.19 presents a portion of the computer output from SAS PROC REG for

the S_{MSFT} in our portfolio.

Analysis of Variance

Source		DF	Sum of Squares	Mean Square	F Value	Pr > F
Model		3	37683	12561	933.99	<.0001
Error		520	6993.24113	13.44854		
Corrected	Total	523	44676			

Root MSE		3.66723	R-Square	0.8435
Dependent	Mean	23.77229	Adj R-Sq	0.8426
Coeff	Var	15.42647		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	23.75662	0.16021	148.29	<.0001	0
MRP	1	2.63248	0.18537	14.20	<.0001	2.40032
MktRP	1	0.05079	0.00148	34.32	<.0001	4.18659
VvsGP	1	-0.00223	0.00025994	-8.57	<.0001	3.93103

Figure 4.19 SAS PROC REG output for S_{MSFT} in linear model

The F value of 933.99 is used to test the null hypothesis $H_0 : \beta_{MRP} = \beta_{MktRP} = \beta_{VvsGP} = 0$.

The associated p value of < 0.0001 leads to the rejection of the null hypothesis and indicated that at least one of the coefficients is not zero. Since all three p values associated with the three regression coefficients are less than 0.0001, all three risk factors contribute significantly to the model. R-Square = 0.8435 means that a major portion ($>$ 84.35%) of the variation of S_{MSFT} is explained by variation in the independent variables of the model. The statistic Adj R-Sq = 0.8426 (an alternative to R-Square) is adjusted for the number of parameters in the model. The maximum VIF = 4.187 < 10 implies that there is no serious problems with multicollinearity in this model. The fitted linear model for S_{MSFT} is

$$\hat{S}_{MSFT} = 23.76 + 2.632 R_{MRP} + 0.0508 R_{MktRP} - 0.00223 R_{VvsGP}. \tag{4.35}$$

Figure 4.20 is the plot of residuals versus predicted MSFT price and Figure 4.21 is the plot of the studentized residuals versus predicted MSFT price. They both show an outward–opening funnel pattern implying that the variance is an increasing function of S_{MSFT}. Figure 4.22 shows that most of the points are close to a straight line but a few extreme points slightly deviate from the line. It may be inferred that the distribution is close to normal.
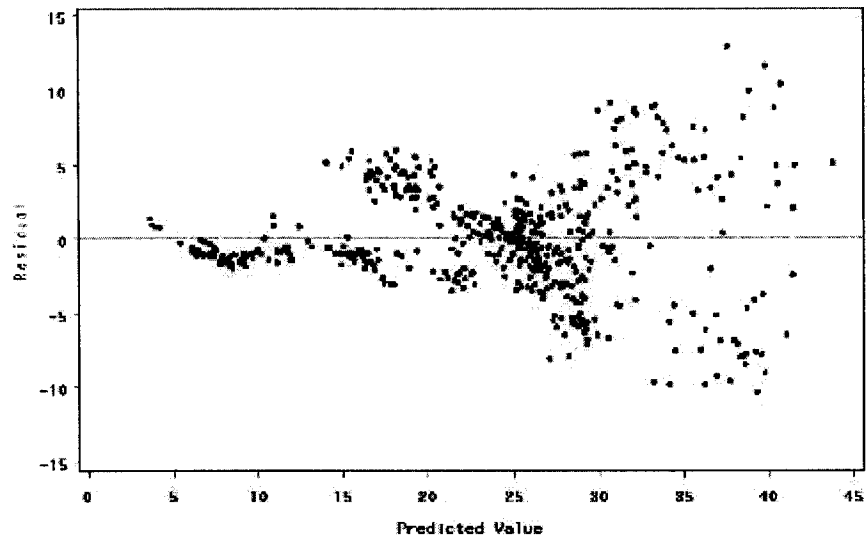
Figure 4.20 Plot of residuals vs. predicted S_{MSFT} in linear model
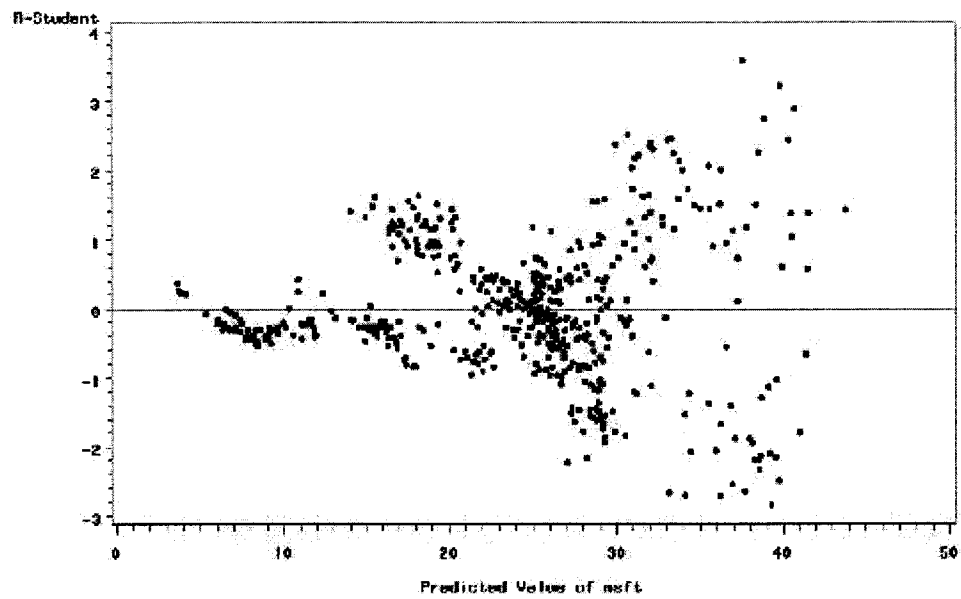
Figure 4.21 Plot of R-student values vs. predicted S_{MSFT} in linear model
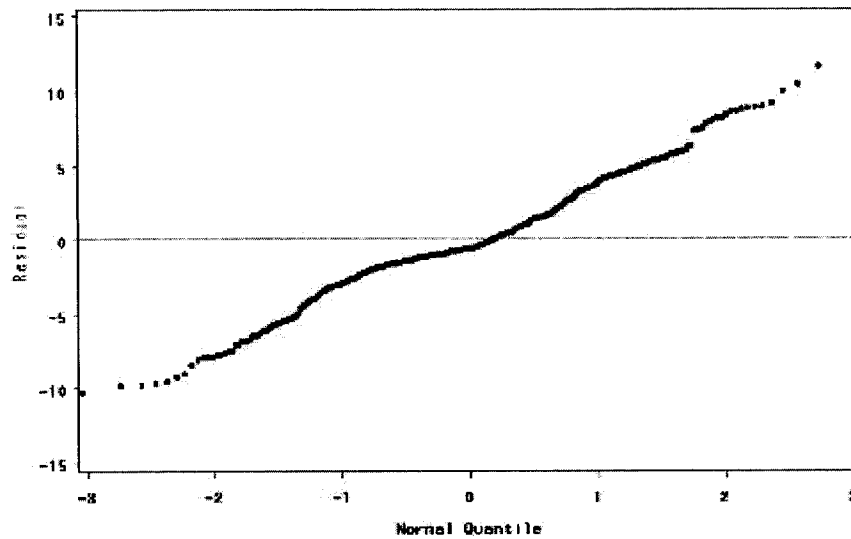
Figure 4.22 Normal probability plot for S_{MSFT} in linear model

Figure 4.23 presents a portion of the computer output from SAS PROC REG for

S_{TXN} in our portfolio.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	112747	37582	791.92	<.0001
Error	520	24678	47.45734		
Corrected Total	523	137425			

Root MSE	6.88893	R-Square	0.8204
Dependent Mean	26.09403	Adj R-Sq	0.8194
Coeff Var	26.40040		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	26.06792	0.30095	86.62	<.0001	0
MRP	1	4.72193	0.34822	13.56	<.0001	2.40032
MktRP	1	0.10550	0.00278	37.95	<.0001	4.18659
VvsGP	1	-0.00913	0.00048831	-18.70	<.0001	3.93103

Figure 4.23 SAS PROC REG output for S_{TXN} in linear model

The F value of 791.92 is used to test the null hypothesis $H_0 : \beta_{MRP} = \beta_{MktRP} = \beta_{VvsGP} = 0$.

The associated p value of < 0.001 leads to the rejection of the null hypothesis and

indicated that at least one of the coefficients is not zero. Since all three p values associated with three regression coefficients are less than 0.0001, all three risk factors contribute significantly to the model. R-Square = 0.8204 means that a major portion (> 82.04%) of the variation of S_{TXN} is explained by variation in the independent variables in the model. The statistic Adj R-Sq = 0.8426 is adjusted for the number of parameters in the model. The maximum VIF = 4.187 < 10 implies that there is no serious problems with multicollinearity in this model. The fitted linear model for S_{TXN} is

$$\hat{S}_{TXN} = 26.07 + 4.722R_{MRP} + 0.1055R_{MktRP} - 0.00913R_{VvsGP} \qquad (4.36)$$

Figure 4.24 is the plot of residuals versus predicted S_{TXN} and Figure 4.25 is the plot of R-student versus predicted S_{TXN}. They both also show an outward–opening funnel pattern implying that the variance is an increasing function of S_{TXN}. Figure 4.26 shows that most of the points are close to a straight line but a few extreme points slightly deviate from the line. It may be inferred that the distribution is close to normal.
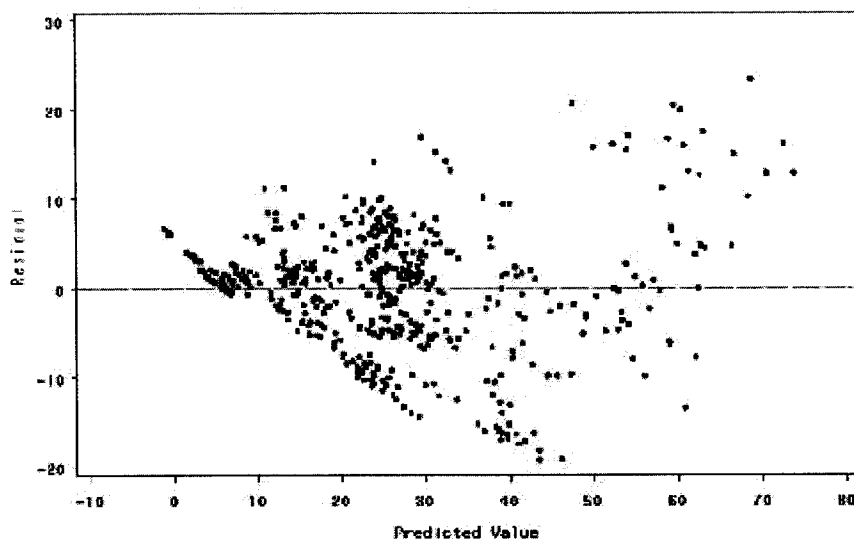
Figure 4.24 Plot of residuals vs. predicted S_{TXN} in linear model
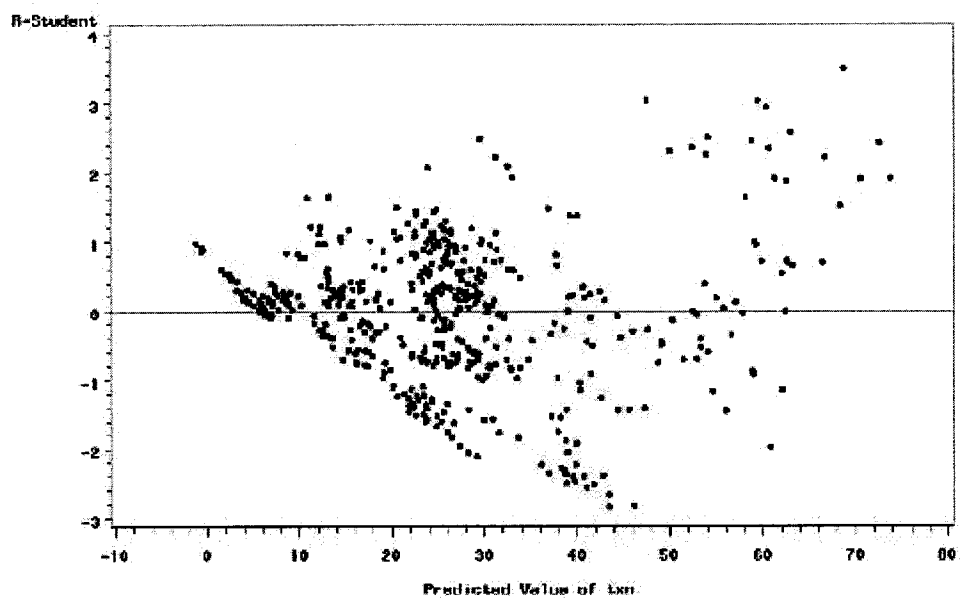
Figure 4.25 Plot of R-student values vs. predicted S_{TXN} in linear model
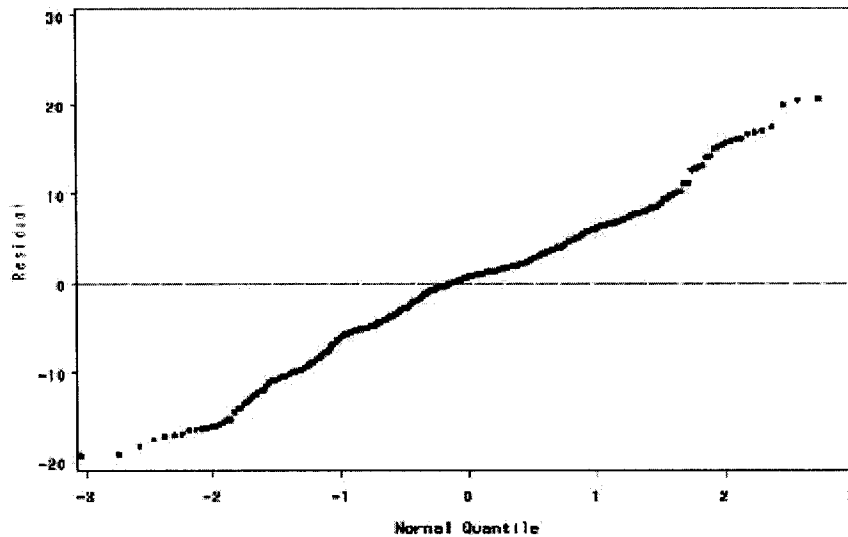
Figure 4.26 Normal probability plot for S_{TXN} in linear model

4.5.3 Building Quadratic Models

4.5.3.1 Full model with non-centered regressors

The fitted three-factor quadratic regression models is

$$S_{stock} = \beta_0 + \beta_1 \times R_{MRP} + \beta_2 \times R_{MktRP} + \beta_3 \times R_{VvsGP} + \beta_{11} \times R_{MRP}^2$$

$$+ \beta_{22} \times R_{MktRP}^2 + \beta_{33} \times R_{VvsGP}^2 + \beta_{12} \times R_{MRP} \times R_{MktRP} \tag{4.37}$$

$$+ \beta_{13} \times R_{MRP} \times R_{MktRP} + \beta_{23} \times R_{MktRP} \times R_{VvsGP} + \varepsilon.$$

Figures 4.27-4.29 are the SAS outputs for the three stocks using the full model, Equation (4.35). The outputs show that for all three models the VIFs are very large, from several hundreds to several thousands, which means that those models have serious multicollinearity problems.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	345166	38352	525.45	<.0001
Error	514	37516	72.98847		
Corrected Total	523	382682			

Root MSE	8.54333	R-Square	0.9020	
Dependent Mean	78.03670	Adj R-Sq	0.9002	
Coeff Var	10.94784			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-71.71400	13.61525	-5.27	<.0001	0
MRP	1	43.56375	5.64680	7.71	<.0001	410.41678
MktRP	1	0.15550	0.06105	2.55	0.0112	1313.07762
VvsGP	1	-0.01607	0.00910	-1.77	0.0779	886.98318
x11	1	-1.68134	0.55078	-3.05	0.0024	81.70546
x22	1	-0.00005132	0.00003797	-1.35	0.1771	2268.22844
x33	1	-6.04057E-7	6.102116E-7	-0.99	0.3227	718.66617
x12	1	-0.02570	0.00802	-3.21	0.0014	927.31986
x13	1	0.00036518	0.00117	0.31	0.7543	1177.42576
x23	1	0.00002029	0.00000737	2.75	0.0061	2326.52905

Figure 4.27 SAS PROC REG output for S_{IBM} in full model with non-centered regressors

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	40571	4507.93507	564.53	<.0001
Error	514	4104.42744	7.98527		
Corrected Total	523	44676			

Root MSE	2.82582	R-Square	0.9081	
Dependent Mean	23.77229	Adj R-Sq	0.9065	
Coeff Var	11.88704			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-67.19204	4.50343	-14.92	<.0001	0
MRP	1	16.41506	1.86776	8.79	<.0001	410.41678
MktRP	1	0.03640	0.02019	1.80	0.0720	1313.07762
VvsGP	1	0.00991	0.00301	3.29	0.0011	886.98318
x11	1	-0.66141	0.18218	-3.63	0.0003	81.70546
x22	1	0.00002644	0.00001256	2.11	0.0357	2268.22844
x33	1	-1.04906E-7	2.018357E-7	-0.52	0.6035	718.66617
x12	1	0.00978	0.00265	3.69	0.0003	927.31986
x13	1	-0.00271	0.00038578	-7.03	<.0001	1177.42576
x23	1	-0.00000696	0.00000244	-2.86	0.0045	2326.52905

Figure 4.28 SAS PROC REG output for S_{MSFT} in full model with non-centered regressors

```
                        Analysis of Variance

                                Sum of        Mean
        Source           DF     Squares      Square     F Value    Pr > F

        Model             9      125212       13912      585.56    <.0001
        Error           514       12212    23.75942
        Corrected Total 523      137425


                Root MSE          4.87436   R-Square     0.9111
                Dependent Mean   26.09403   Adj R-Sq     0.9096
                Coeff Var        18.68000


                        Parameter Estimates

                     Parameter    Standard                          Variance
        Variable  DF  Estimate       Error   t Value   Pr > |t|    Inflation

        Intercept  1   29.25293     7.76813      3.77    0.0002           0
        MRP        1    2.85524     3.22176      0.89    0.3759    410.41678
        MktRP      1   -0.08890     0.03483     -2.55    0.0110   1313.07762
        VvsGP      1    0.00073257  0.00519      0.14    0.8878    886.98318
        x11        1    1.39059     0.31425      4.43    <.0001     81.70546
        x22        1    0.00023193  0.00002167  10.71    <.0001   2268.22844
        x33        1    0.00000469  3.481539E-7  13.46    <.0001    718.66617
        x12        1    0.01535     0.00457      3.36    0.0009    927.31986
        x13        1   -0.00329     0.00066544  -4.94    <.0001   1177.42576
        x23        1   -0.00005539  0.00000420 -13.18    <.0001   2326.52905
```

Figure 4.29 SAS PROC REG output for S_{TXN} in full model with non-centered regressors

4.5.3.2 Full model with centered regressors

The fitting of quadratic polynomials of several independent variables with means not equal to zero can create difficult multicollinearity problems, as shown in Section 4.5.3.1. Specifically, the polynomials will be highly correlated due to the mean of the primary independent variable. With large numbers, this problem is very serious, and if proper protections are not put in place, can cause wrong results. One possible solution is to "center" the independent variables, i.e., to subtract the mean, and then to compute the polynomials. The formula for the centered-regressor quadratic regression model is

$$S_{stock} = \beta_0 + \beta_1 \times R'_{MRP} + \beta_2 \times R'_{MktRP} + \beta_3 \times R'_{VvsGP} + \beta_{11} \times R'^2_{MRP}$$

$$+ \beta_{22} \times R'^2_{MktRP} + \beta_{33} \times R'^2_{VvsGP} + \beta_{12} \times R'_{MRP} \times R'_{MktRP} \tag{4.38}$$

$$+ \beta_{13} \times R'_{MRP} \times R'_{MktRP} + \beta_{23} \times R'_{MktRP} \times R'_{VvsGP} + \varepsilon,$$

where $R'_{MRP} = R_{MRP} - \mu_{MRP}$, $R'_{MktRP} = R_{MktRP} - \mu_{MktRP}$ and $R'_{VvsGP} = R_{VvsGP} - \mu_{VvsGP}$.

Figures 4.30-4.32 are the SAS outputs for the three stocks in the full regressor-centered model. The outputs show that for all three models the VIFs are improved very much but they are still larger than 10, which still indicates that those models have multicollinearity problems. To solve this problem, we seek variable selection using all possible regressions and stepwise regression methods.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	345166	38352	525.45	<.0001
Error	514	37516	72.98847		
Corrected Total	523	382682			

Root MSE	8.54333	R-Square	0.9020
Dependent Mean	78.03670	Adj R-Sq	0.9002
Coeff Var	10.94784		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	77.82871	0.91638	84.93	<.0001	0
MRP	1	11.60955	0.70271	16.52	<.0001	6.35592
MktRP	1	0.13897	0.00781	17.79	<.0001	21.48959
VvsGP	1	-0.00207	0.00139	-1.49	0.1375	20.69011
x11	1	-1.68134	0.55078	-3.05	0.0024	6.21987
x22	1	-0.00005132	0.00003797	-1.35	0.1771	35.62985
x33	1	-6.04057E-7	6.102116E-7	-0.99	0.3227	11.23565
x12	1	-0.02570	0.00802	-3.21	0.0014	26.54771
x13	1	0.00036518	0.00117	0.31	0.7543	10.86933
x23	1	0.00002029	0.00000737	2.75	0.0061	47.33267

Figure 4.30 SAS PROC REG output for S_{IBM} in full model with centered regressors

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	40571	4507.93507	564.53	<.0001
Error	514	4104.42744	7.98527		
Corrected Total	523	44676			

Root MSE	2.82582	R-Square	0.9081	
Dependent Mean	23.77229	Adj R-Sq	0.9065	
Coeff Var	11.88704			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	27.23243	0.30310	89.85	<.0001	0
MRP	1	4.78464	0.23243	20.59	<.0001	6.35592
MktRP	1	0.06317	0.00258	24.45	<.0001	21.48959
VvsGP	1	-0.00459	0.00045953	-9.98	<.0001	20.69011
x11	1	-0.66141	0.18218	-3.63	0.0003	6.21987
x22	1	0.00002644	0.00001256	2.11	0.0357	35.62985
x33	1	-1.04906E-7	2.018357E-7	-0.52	0.6035	11.23565
x12	1	0.00978	0.00265	3.69	0.0003	26.54771
x13	1	-0.00271	0.00038578	-7.03	<.0001	10.86933
x23	1	-0.00000696	0.00000244	-2.86	0.0045	47.33267

Figure 4.31 SAS PROC REG output for S_{MSFT} in full model with centered regressors

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	125212	13912	585.56	<.0001
Error	514	12212	23.75942		
Corrected Total	523	137425			

Root MSE	4.87436	R-Square	0.9111	
Dependent Mean	26.09403	Adj R-Sq	0.9096	
Coeff Var	18.68000			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	18.41585	0.52284	35.22	<.0001	0
MRP	1	1.36356	0.40093	3.40	0.0007	6.35592
MktRP	1	0.04498	0.00446	10.09	<.0001	21.48959
VvsGP	1	0.00168	0.00079266	2.12	0.0343	20.69011
x11	1	1.39059	0.31425	4.43	<.0001	6.21987
x22	1	0.00023193	0.00002167	10.71	<.0001	35.62985
x33	1	0.00000469	3.481539E-7	13.46	<.0001	11.23565
x12	1	0.01535	0.00457	3.36	0.0009	26.54771
x13	1	-0.00329	0.00066544	-4.94	<.0001	10.86933
x23	1	-0.00005539	0.00000420	-13.18	<.0001	47.33267

Figure 4.32 SAS PROC REG output for S_{TXN} in full model with centered regressors

4.5.3.3 Best model with centered regressors

The following procedure is used for selecting the best model for our stock portfolio:

1. Select candidate models from stepwise, forward, backward and all possible regressions based on R-Square, Adj R-sq, MSE and c(p).

2. For each candidate model, run PROC REG again to check on the variation inflation factor (VIF).

3. Select the best model that has the least VIF value.

4. Use the model to check for normality of residuals and outliers by normal probability plot, R-student plot and residual plot.

The following figures show the final best model for each stock.

Figure 4.33 presents a portion of the computer output from SAS PROC REG for the IBM stock in the reduced quadratic model. The F value of 1163.81 is used to test the null hypothesis $H_0 : \beta_{MRP} = \beta_{MktRP} = \beta_{MRP^2} = \beta_{MRP,VvsGP} = 0$. The associated p value of < 0.001 leads to the rejection of the null hypothesis and indicated that at least one of the coefficients is not zero. Since the p values associated with the four regression coefficients are less than 0.0001, all factors in the model are significant. In addition to the risk factors R_{MRP} and R_{MktRP}, the quadratic term of R_{MktRP} and the cross-product term of R_{MRP} and R_{VvsGP} contribute significantly to the model. R-Square = 0.8997 which is better than the linear model's R-Square (0.8832). The statistic Adj R-Sq = 0.8989 is also higher than the linear model's R-Square (0.8825). The maximum VIF = 1.88 < 10 implies that there is no serious problems with multicollinearity in this model.

```
                          Analysis of Variance

                                   Sum of          Mean
          Source          DF       Squares         Square     F Value     Pr > F

          Model            4        344297         86074       1163.81    <.0001
          Error          519         38385         73.95923
          Corrected Total 523       382682


                    Root MSE            8.59996     R-Square    0.8997
                    Dependent Mean     78.03670     Adj R-Sq    0.8989
                    Coeff Var          11.02040


                            Parameter Estimates

                          Parameter       Standard                          Variance
          Variable   DF    Estimate          Error    t Value    Pr > |t|   Inflation

          Intercept   1    76.11562        0.54773     138.97     <.0001           0
          MRP         1    10.86804        0.38464      28.26     <.0001     1.87926
          MktRP       1     0.12880        0.00203      63.35     <.0001     1.43691
          x22         1     0.00006074     0.00000785    7.74     <.0001     1.50208
          x13         1    -0.00268        0.00039797   -6.73     <.0001     1.24892
```

Figure 4.33 SAS PROC REG output for S_{IBM} in best model with centered regressors

The fitted best model for S_{IBM} is

$$\hat{S}_{IBM} = 76.116 + 10.868 R_{MRP} + 0.1288 R_{MktRP} + 0.0000006074 R^2{}_{MktRP} - 0.00268 R_{MRP} R_{VvsGP}.$$

$$(4.39)$$

Figure 4.34 is the plot of R-Student residuals versus predicted IBM price. The pattern shows no specific error pattern exhibited by the residuals of the linear model, but it shows that larger residuals are associated with larger values of the response variable. Figure 4.35 shows that most of the points are close to a straight line but a few extreme points slightly deviate from the line. It may be inferred that the distribution is close to normal.
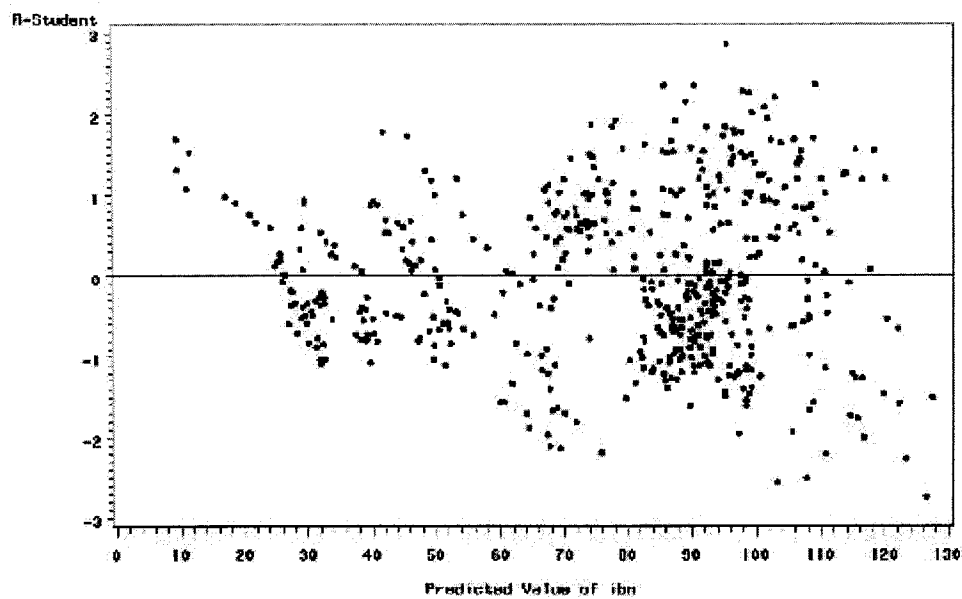
Figure 4.34 Plot of R-student residuals vs. predicted S_{IBM} in best model with centered regressors
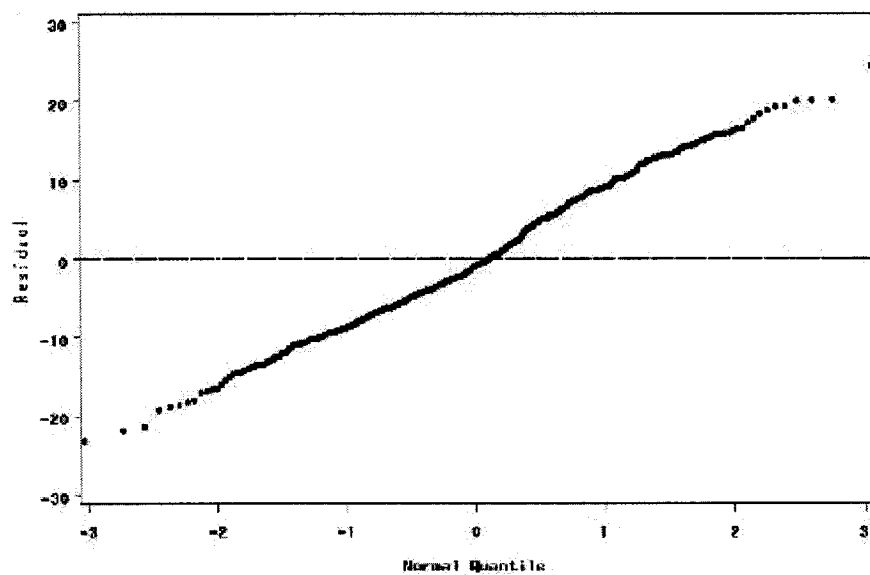
Figure 4.35 Normal probability plot for S_{IBM} in best model with centered regressors

Figure 4.36 presents a portion of the computer output from SAS PROC REG for

the MSFT stock in the case of the best reduced model.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	40434	6739.02188	821.38	<.0001
Error	517	4241.71179	8.20447		
Corrected Total	523	44676			

Root MSE	2.86434	R-Square	0.9051	
Dependent Mean	23.77229	Adj R-Sq	0.9040	
Coeff Var	12.04909			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	27.52256	0.28756	95.71	<.0001	0
MRP	1	4.99701	0.19564	25.54	<.0001	4.38267
MktRP	1	0.06698	0.00151	44.47	<.0001	7.10990
VvsGP	1	-0.00531	0.00031943	-16.62	<.0001	9.73044
x11	1	-1.15058	0.09166	-12.55	<.0001	1.53252
x22	1	-0.00002010	0.00000359	-5.60	<.0001	2.83389
x13	1	-0.00173	0.00015527	-11.16	<.0001	1.71374

Figure 4.36 SAS PROC REG output for S_{MSFT} in best model with centered regressors

The F value of 821.38 is used to test the null hypothesis

$$H_0 : \beta_{MRP} = \beta_{MktRP} = \beta_{VvsGP} = \beta_{MRP^2} = \beta_{MktRP^2} = \beta_{MRP,VvsGP} = 0.$$

The associated p value of < 0.0001 leads to the rejection of the null hypothesis and

indicated that at least one of the coefficients is not zero. Since all the p values associated

with six regression coefficients are less than 0.0001, all three risk factors contribute

significantly to the model. Also, the quadratic terms of R_{MRP} and R_{MktRP} and the cross-

product term of R_{MRP} and R_{VvsGP} contribute significantly to the model. R-Square = 0.905

is better than the linear model's (0.8435). The statistic Adj R-Sq = 0.904 is also larger

than the linear model's (0.8426). The maximum VIF = 9.73 < 10 implies that there are no serious problems with multicollinearity in this model.

The fitted linear model for S_{MSFT} is

$$\hat{S}_{MSFT} = 27.523 + 4.997 R_{MRP} + 0.067 R_{MktRP} - 0.00531 R_{VvsGP} - 1.151 R_{MRP}^2$$

$$- 0.00002 R_{MktRP}^2 - 0.00173 R_{MRP} R_{VvsGP} \qquad (4.40)$$

Figure 4.37 is the plot of R-Student residuals versus predicted MSFT price. The figure shows no specific error pattern exhibited by the residuals of the linear model. Figure 4.38 shows that most of the points are close to a straight line, but a few extreme points slightly deviate from the line. It may be inferred that the distribution is close to normal.
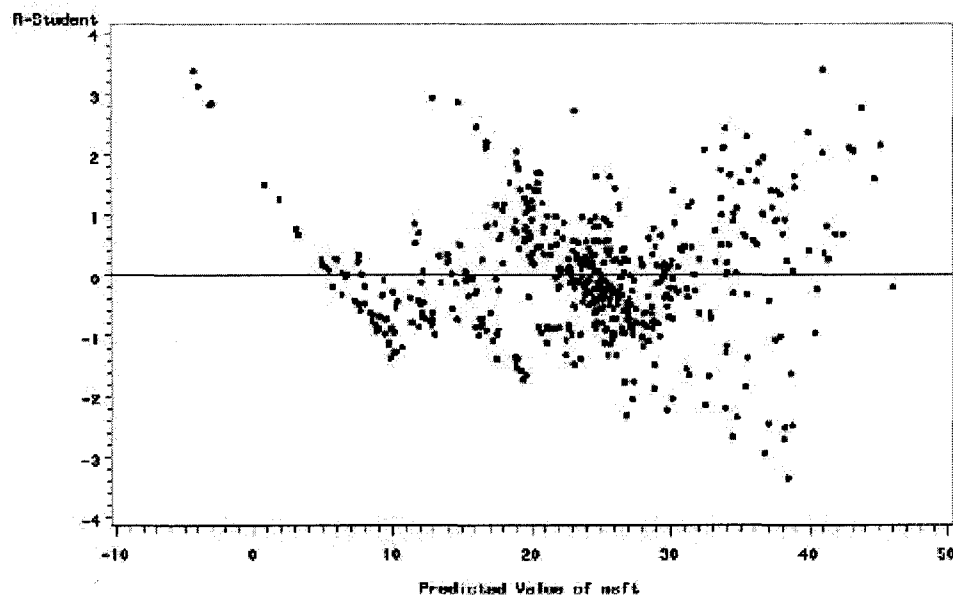
Figure 4.37 Plot of R-student residuals vs. predicted S_{MSFT} in best model with centered regressors
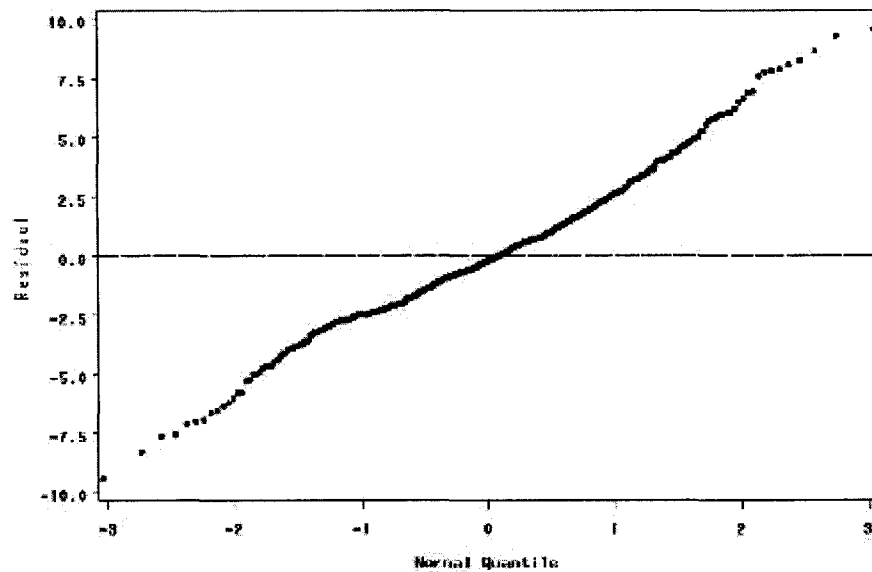
Figure 4.38 Normal probability plot for S_{MSFT} in best model with centered regressors

Figure 4.39 presents a portion of the computer output from SAS PROC REG for the TXN stock in the case of the best reduced model.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	124372	24874	987.16	<.0001
Error	518	13053	25.19804		
Corrected Total	523	137425			

Root MSE	5.01976	R-Square	0.9050	
Dependent Mean	26.09403	Adj R-Sq	0.9041	
Coeff Var	19.23722			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	18.76481	0.39686	47.28	<.0001	0
MktRP	1	0.03249	0.00228	14.28	<.0001	5.28577
VvsGP	1	0.00397	0.00043183	9.20	<.0001	5.78995
x22	1	0.00020977	0.00000910	23.06	<.0001	5.92484
x33	1	0.00000378	2.70189E-7	13.98	<.0001	6.38058
x23	1	-0.00004664	0.00000189	-24.70	<.0001	9.00264

Figure 4.39 SAS PROC REG output for S_{TXN} in best model with centered regressors

The F value of 987.16 is used to test the null hypothesis

$$H_0 : \beta_{MktRP} = \beta_{VvsGP} = \beta_{MktRP^2} = \beta_{VvsGP^2} = \beta_{MktRP,VvsGP} = 0.$$

The associated p value of < 0.0001 leads to the rejection of the null hypothesis and indicated that at least one of the coefficients is not zero. Since all the p values associated with five regression coefficients are less than 0.0001, the R_{MktRP} and R_{VvsGP}, the quadratic terms of R_{MktRP} and R_{VvsGP} and the cross-product term of R_{MktRP} and R_{VvsGP} contribute significantly to the model. R-Square $= 0.9050$ is a large improvement on the linear model's 0.8204. The statistic Adj R-Sq $= 0.9041$ is also much better than 0.8426 of the linear model. The maximum VIF $= 9.00264 < 10$ implies that there is no serious problems with multicollinearity in this model.

The fitted linear model for S_{TXN} is

$$\hat{S}_{TXN} = 18.765 + 0.0325 R_{MktRP} + 0.00397 R_{VvsGP} + 0.00021 R_{MktRP}^2$$

$$+ 0.00000378 R_{VvsGP}^2 - 0.004664 R_{MktRP} R_{VvsGP} \tag{4.41}$$

Figure 4.40 is the plot of R-student residuals versus predicted TXN price. It shows an outward –opening funnel pattern implying that the variance is an increasing function of S_{TXN}. Figure 4.41 shows that extreme points deviate slightly from a straight line which may infer that the distribution has a heavy tail, but not far from normal.
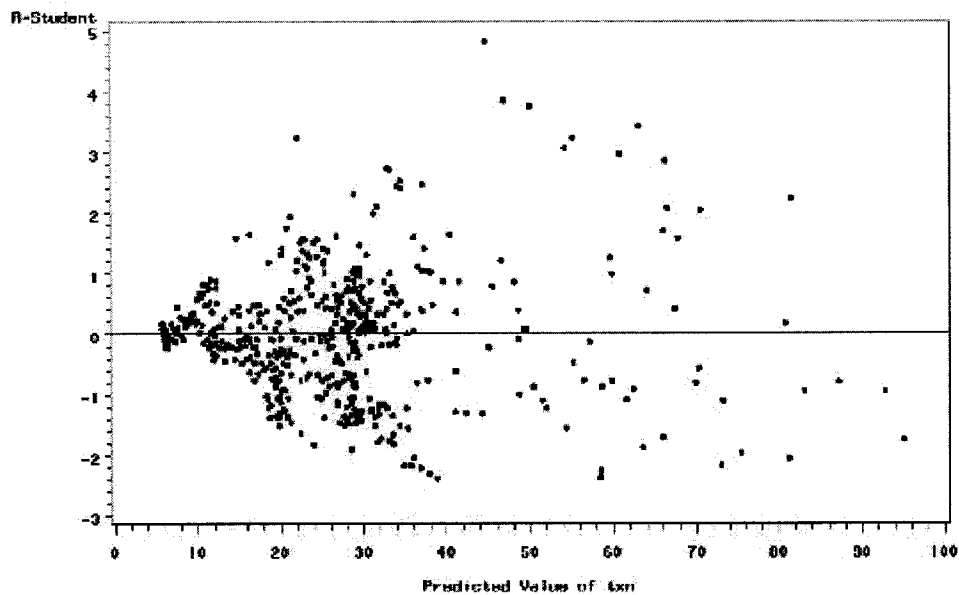
Figure 4.40 Plot of R-Student residuals vs. predicted S_{TXN} in best model with centered regressors
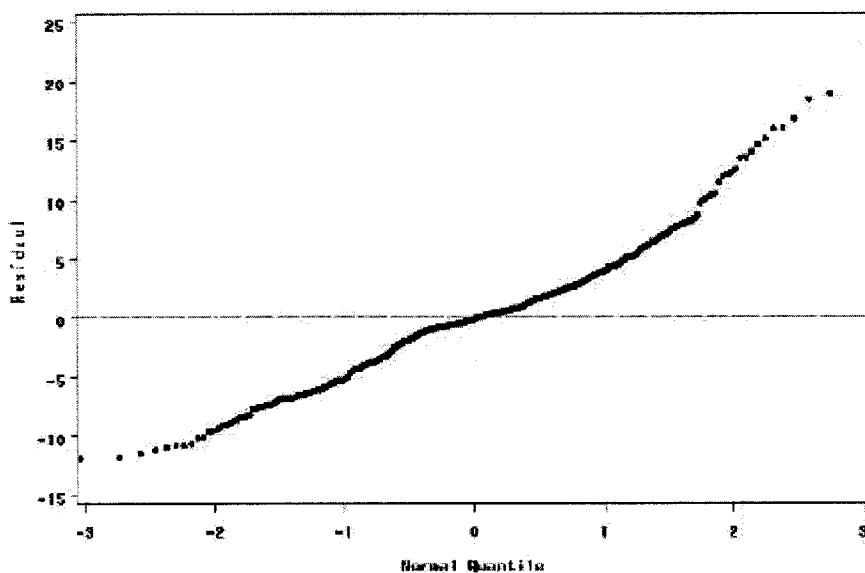
Figure 4.41 Normal probability plot for S_{TXN} in best model with centered regressors

4.6 Results and Discussions

1. Regression analysis can be used to quantitatively measure the impact of multiple independent variables (risk factors) on a dependent variable (stock).

2. Quadratic models are better than linear models in our stock portfolio.

3. Statistical methods, such as all possible regression and stepwise regression (forward, backward, and stepwise) can be used to determine the set of risk factors that most effectively explain the stock value. The best model is selected according to criterions, such as R-square, Adj Sq, MSE, c(p) and VIF.

4. Graphical analysis is used to check on model adequacy.

5. The models developed in this chapter will be used in the VaR transformation procedure which is the content of the next chapter.

CHAPTER 5

SIMULATIONS IN TRANSFORMATION

PROCEDURE

5.1 Introduction

There are six sections of this chapter. The first section is an introduction – a review of a transformation procedure in VaR measure. The second section is a discussion about the Monte Carlo method. The third section is a discussion about statistical multivariate distributions. The fourth section is simulation of four kinds of multivariate distributions using the Monte Carlo method. The fifth section is Monte Carlo transformation based on the three-factor models in Chapter 4. The last section is the results and discussions for this chapter.

As mentioned in Chapter 4, risk has two components: exposure and uncertainty. By specifying a portfolio mapping function θ in $^1P = \theta(^1\vec{R})$, a mapping procedure describes exposure. By characterizing the joint distribution for $^1\vec{R}$, an inference procedure describes uncertainty. A transformation procedure combines exposure and uncertainty to describe the distribution of ^1P, according to which we can summarize the value of any VaR metric. Hence, the transformation procedure describes risk.

The position of transformation procedure is highlighted in Figure 5.1.

148
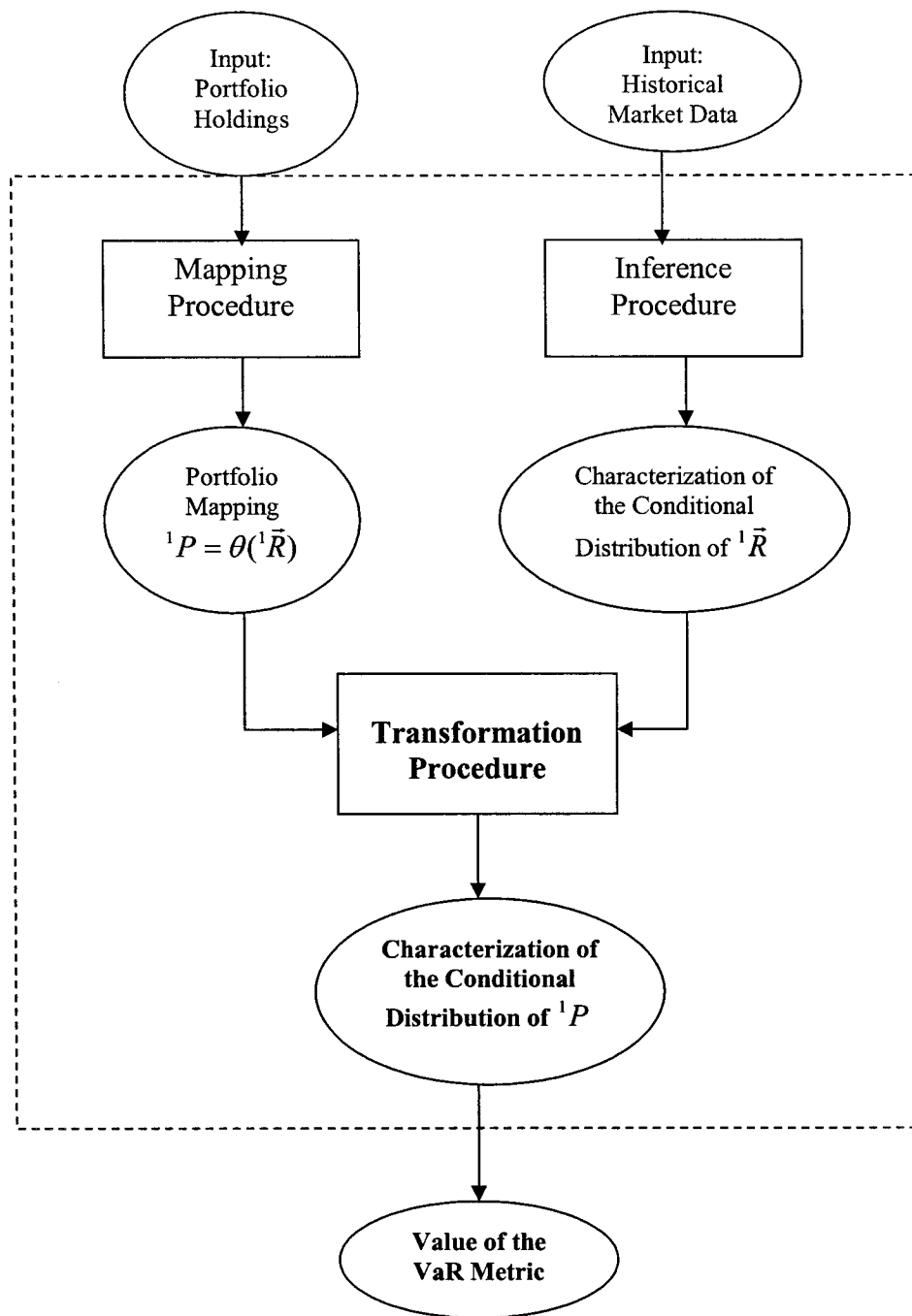
Figure 5.1 A reproduction of Figure 2.1 [21]

The chart in Figure 5.1 shows that a transformation procedure combines the outputs from the mapping and inference procedures and uses them to characterize the distribution of ^1P, conditional on information available at time 0. Based on that characterization, and perhaps the portfolio's current value ^0p, the transformation procedure determines the value of the desired VaR metric. The result is the VaR measurement. The characterization of the conditional distribution of ^1P may be a standard deviation, probability density function (pdf), characteristic function, or some other representation. If the characterization is sufficiently general to support any reasonable VaR metric, the transformation is complete. Otherwise, it is incomplete.

There are three types of transformations in VaR measures: linear transformations, quadratic transformations, and Monte Carlo transformations. Linear and quadratic transformations apply to linear and quadratic portfolios, respectively. Monte Carlo transformations apply to all portfolios.

5.1.1 Linear Transformations

Linear transformations are applicable to portfolios whose portfolio mapping function is a linear polynomial. Such portfolios include portfolios of equities, physical commodities, or futures. The market value of such portfolios depends linearly upon applicable key factors. Consider a portfolio $(^0p,^1P)$ with a linear portfolio mapping

$$^1P = \theta(^1\vec{R}) = \vec{b}'^1\vec{R} + a \tag{5.1}$$

where $\vec{b}$ is a vector and a is a scalar.

The mean and standard deviation of ^1P are given by

$$^0E(^1P) = \vec{b}' \times \vec{\mu}_{\bar{R}} + a, \tag{5.2}$$

and

$$^0Std(^1P) = \sqrt{^0Var(^1P)} = \sqrt{\vec{b}'^{1|0}\Sigma\vec{b}}, \tag{5.3}$$

where $^{1|0}\Sigma$ is the covariance matrix for $\vec{R}$ at time 1 conditional on the information at time 0. On its own, a standard deviation is not sufficient to determine a quantile, so it's an incomplete transformation. To complete the transformation, additional assumptions have to be made. Usually the standard assumptions are to specify a value for $^0E(^1P)$ and assume that ^1P is conditionally normal.

$$^1P \sim N(^0E(^1P),(^0Std(^1P))^2) \tag{5.4}$$

$$^1P =\,^0Std(^1P)Z+^0E(^1P), \quad Z \sim N(0,1) \tag{5.5}$$

The standard deviation of loss is

$$^0Std(^1L)=\,^0Std(^0p-^1P)=\,^0Std(^1P). \tag{5.6}$$

Based on Equation (5.5), it follows that any quantile of an $N(\mu,\sigma^2)$ distribution occurs a distance from its mean μ that is a fixed multiple of σ. Therefore, a q-quantile–of–loss VaR metric is calculated as

$$^{1|0}\Phi^{-1}_{^1_L}(q)=\,^0p-^{1|0}\Phi^{-1}_{^1_P}(1-q)$$

$$\begin{aligned}&=\,^0p-\left[^0E(^1P)+\Phi^{-1}_Z(1-q)\times^0Std(^1P)\right]\\&=\,^0p-^0E(^1P)+\Phi^{-1}_Z(q)\times^0Std(^1P).\end{aligned} \tag{5.7}$$

If the VaR horizon is short, e.g. a day or a week, it is reasonable to assume $^0E(^1P)=^0p$. In this case, Equation (5.7) becomes

$$^{1|0}\Phi^{-1}_{^1_L}(q) = \Phi^{-1}_Z(q)^0Std(^1P). \tag{5.8}$$

Figure 5.2 is an example of $q = 0.9$.
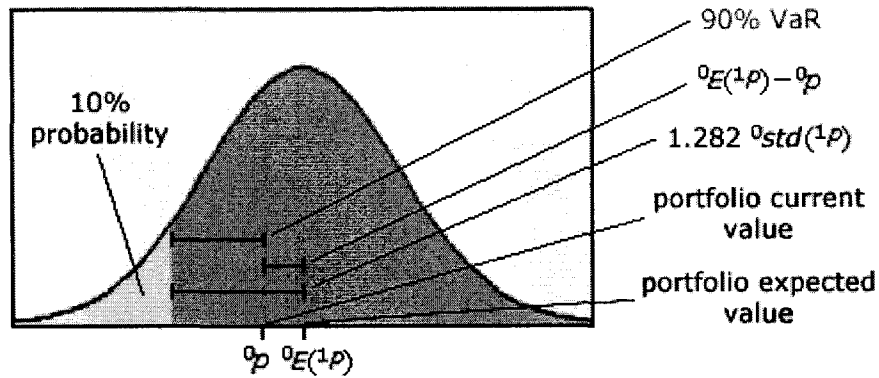
Figure 5.2 A graphical derivation of Equation (5.7) when $q = 0.9$. [46]

The 90% loss occurs at a portfolio value 1.282 standard deviations below the portfolio's expected value (the mean of the distribution). However, loss is calculated relative to the portfolio's current value as opposed to its expected value, which is why Equation (5.7) includes the $[{}^0E({}^1P)-{}^0p]$ term.

5.1.2 Quadratic Transformations

Quadratic transformations are applicable to quadratic portfolios. Consider a portfolio $({}^0p,{}^1P)$ with quadratic portfolio mapping

$$ {}^1P = \theta({}^1\vec{R}) = {}^1\vec{R}C{}^1\vec{R} + \vec{b}'\vec{R} + a , \tag{5.9} $$

where ${}^1\vec{R} \sim N_n({}^{1|0}\vec{\mu},{}^{1|0}\Sigma)$, C is a symmetric n x n matrix, $\vec{b}$ is a vector and a is a scalar. ${}^{1|0}\Sigma$ is assumed positive definite. The standard deviation of loss is

$$ {}^0Std({}^1L) = {}^0Std({}^1P) = \sqrt{{}^0E({}^1P^2)-{}^0E({}^1P)^2} . \tag{5.10} $$

Quantile-based VaR metrics are difficult to calculate. Various solutions have been proposed, such as approximate solution using Johnson's curves [52][53], Cornish-Fisher expansion [54][55], and Fast Fourier Transform [56]. The quadratic transformations described in these papers differ in various respects, but they all employ the mathematics of quadratic polynomials of joint-normal random vectors. Their collective solution can reasonably be called the quadratic transformation. However, all these methods are based on the assumption that the key factors have a joint-normal distribution. In the real world, data are not normally distributed. They may be skewed and heavy-tailed. Since we are using four different multivariate distributions to simulate the key factors, the quadratic transformation is not appropriate for our research. An obvious solution to our situation is to apply the Monte Carlo method.

5.1.3 Monte Carlo Transformations

There are two steps in Monte Carlo transformations of VaR measures: (1) randomly generate realizations $^1\vec{r}^{[k]}$ based on the joint distribution of $^1\vec{R}$ characterized in the inference procedure; and (2) construct a histogram of realizations $^1p^{[k]} = \theta(^1\vec{r}^{[k]})$ of ^1P.

Historical transformations are identical to Monte Carlo transformations except for one difference. Both employ the Monte Carlo method to construct a histogram of realizations $^1p^{[k]} = \theta(^1\vec{r}^{[k]})$ of ^1P. The difference lies in how they construct realizations $^1\vec{r}^{[k]}$ for $^1\vec{R}$. Monte Carlo transformations randomly generate them based upon a characterization of the distribution of $^1\vec{R}$. Historical transformations employ realizations $^1\vec{r}^{[k]}$ constructed from historical market data for $^1\vec{R}$.

Monte Carlo transformations are flexible. We can model either the portfolio mapping function θ or the joint distribution that is assumed for $^1\vec{R}$ or both. The joint distribution is not limited to joint-normal.

One of the major purposes of this research is to study how changes in the distribution of risk factors will affect the distributions of stock prices assuming that the stock is exposed to those three risk factors. In other words, instead of assuming joint-normal distribution for risk factors and normal distribution for portfolio value, other distributions may be chosen according to the historical data of the risk factors. In this case, the distribution of the portfolio value ^1P could be different from normal distribution. Following the two steps mentioned above, Monte Carlo transformations make it possible to model different joint distributions for $^1\vec{R}$ and to investigate the impact of the different distributions of $^1\vec{R}$ on the distribution of ^1P by comparing them with results under the normal assumption. This study provides more complete and valuable information for risk analysis and portfolio optimization.

5.2 Monte Carlo Method

5.2.1 Overview

The Monte Carlo (MC) method was originally practiced under more generic names such as "statistical sampling". It includes any technique of statistical sampling used to approximate solutions to quantitative problems. The real use of the Monte Carlo method as a research tool stems from work on the atomic bomb during the Second World War. This work involved a direct simulation of the probabilistic problems concerned with random neutron diffusion in fissile material. The generally accepted birth date of the

Monte Carlo method is 1949, when the article entitled "The Monte Carlo method" by Metropolis and Ulam appeared [57]. The name "Monte Carlo" comes from the city in the Monaco principality because of a roulette, a simple random number generator. (Monte Carlo is a city of Monaco, where the primary attractions are casinos containing games of chance, such as roulette wheels, dice, and slot machines, all of which exhibit random behavior.)

The Monte Carlo method is a numerical method of solving a mathematical problem by the simulation of random variables (random sampling) [57]. It provides approximate solutions to a variety of mathematical problems by performing statistical sampling experiments on a computer. This method applies to probabilistic problems as well as deterministic problems. Its efficiency relative to other numerical methods increases when the dimension of the problem increases [21].

Monte Carlo simulation produces results using a stochastic model, rather than an actual experiment with a real system under study. It is regarded as mathematical experimentation and best fit to modern computers [58].

There are two distinctive features of the Monte Carlo Method [57]: (1) The simple structure of the computation algorithm. As a rule, a program is written to carry out one random trial. This trial is repeated N times, each trial being independent of the rest, and then the results of all trials are averaged. Therefore, the Monte Carlo method is also called the method of statistical trials. (2) The error of calculations is proportional to $\sqrt{D/N}$, where D is some constant, and N is the number of trials. Hence, to decrease the error, it is necessary to increase N.

Because of the repetition of algorithms and the large number of calculations involved, Monte Carlo is a method suited to calculations using a computer. The Monte Carlo method enables simulation of any process whose development is influenced by random factors. For many mathematical problems involving no chance, the method also enables us to artificially construct a probabilistic model (or several such models), making possible the solution of the problems. Thus, the Monte Carlo method is a universal numerical method for solving mathematical problems. It is used in everything from economics to nuclear physics to regulating the flow of traffic. Of course the way it is applied varies widely from field to field. But, strictly speaking, to call something a "Monte Carlo" experiment, all you need to do is use random numbers to examine some problem. For example, the Monte Carlo method is useful for modeling phenomena with significant uncertainty in inputs, such as the calculation of risk in business.

5.2.2 Mathematical Background

5.2.2.1 Prerequisites

For a family of independent and identically distributed (iid) random variables $X_1, X_2, \ldots$, define the partial-sum process $\{S_n\}$ by

$$S_n \equiv X_1 + X_2 + \cdots + X_n, \quad n = 1, 2, \ldots, \tag{5.11}$$

with $S_0 = 0$. Associated with S_n is the sample mean defined by $\overline{X}_n \equiv S_n/n$. The iid assumption leads to such classic limit theorems as the strong law of large numbers and the central limit theorem. In a statistical context, laws of large numbers imply that the average of a random sample from a large population is likely to be close to the mean of the whole population [58].

Proposition 1. (Strong Law of Large Numbers) [57] [59] For a family of iid

random variables $X_1, X_2, \ldots$, suppose that the mean $\mu = E(X_i)$ exists. Then

$$P\left(\lim_{n \to \infty} \overline{X}_n = \lim_{n \to \infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu\right) = 1.$$ (5.12)

The strong law of large numbers ensures that the sample mean $\overline{X}_n$ converges to

the population mean μ almost surely as $n \to \infty$. Therefore, the sample mean is used as

an estimate of the population mean for large samples.

For a family of iid random variables $X_1, X_2, \ldots$, suppose that $E(X_i^2) < \infty$. Then

for S_n defined by Equation (5.11) and $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, we have

$$E(S_n) = n\mu, \qquad Var(S_n) = n\sigma^2.$$ (5.13)

Proposition 2. (Central Limit Theorem) [58][59] For a family of iid random

variables $X_1, X_2, \ldots$ with finite mean μ and finite variance $\sigma^2 > 0$, define

$$Z_n = \frac{X_1 + X_2 + \cdots X_n - n\mu}{\sigma\sqrt{n}}, \qquad n = 1, 2, \ldots.$$ (5.14)

Then,

$$\lim_{n \to \infty} P\{Z_n \leq x\} = \Phi(x), \qquad x \in \Re,$$ (5.15)

where $\Phi(x)$ is the cumulative distribution function (CDF) of the standard normal

distribution given by

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt, \qquad x \in \Re.$$ (5.16)

The central limit theorem shows that the distribution of Z_n converges towards the

standard normal distribution $N(0,1)$ as n approaches ∞. If we draw a random sample of

size n $(x_{1,}x_2,\ldots,x_n)$ from an infinite population with mean μ and variance σ^2, then the probability distribution of the partial sum S_n can be approximated by the normal distribution $N(n\mu, N\sigma^2)$ for sufficiently large n. This property makes the normal distribution special in stochastic models.

5.2.2.2 The general scheme of the
Monte Carlo method

For any random variable X and a real-valued function $h(x)$, the composition $h(X)$ is again a random variable under a regularity condition. Also, if $X_1, X_2, \ldots$ are iid, then so are the random variables $h(X_1), h(X_2), \ldots$. Hence, for a family of IID random variables $X_1, X_2, \ldots$, the strong law of large numbers ensures that [59]

$$P\left(\lim_{n\to\infty} \frac{h(X_1) + h(X_2) + \cdots + h(X_n)}{n} = E[h(X)]\right) = 1, \tag{5.17}$$

provided that the expectation $E[h(X)]$ exists.

Suppose that we need to calculate some unknown quantity I. Let us try to find a function $h(X)$ with $E[h(X)] = I$. Assume that the variance of $h(X)$ is $Var[h(X)] = \sigma^2$. Consider N independent random variables $h(X_1), h(X_2), \cdots, h(X_N)$. If $X_1, X_2, \ldots$ are iid and N is sufficiently large, then it follows from the central limit theorem that the distribution of the sum $\rho_N = h(X_1) + h(X_2) + \ldots + h(X_N)$ will be approximately normal, with $E(\rho_N) = N \times I$ and $Var(\rho_N) = \sigma^2 N$. I can be estimated based on Equation (5.17) for sufficiently large N. Hence, I is approximated by the sample mean,

$$I \approx \overline{Z}_N = \frac{h(X_1) + h(X_2) + \cdots + h(X_N)}{N}. \tag{5.18}$$

The confidence interval of the value I can be calculated based on the central limit theorem. Consider the probability

$$P\{|\overline{Z}_n - I| \leq \varepsilon\} = \alpha, \qquad \varepsilon > 0,$$

For a given confidence coefficient α. By the central limit theorem, $\overline{Z}_n$ is approximately normally distributed with mean $I = E[h(X)]$ and variance σ^2/n, where $\sigma^2 = Var[h(X)]$. It follows that

$$P\left\{-\frac{\varepsilon}{\sigma/\sqrt{n}} \leq \frac{\overline{Z}_n - I}{\sigma/\sqrt{n}} \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right\} = \alpha$$

and the normalized random variable $\sqrt{n}(\overline{Z}_n - I)/\sigma$ is approximated by the standard normal variable $N(0,1)$. Denoting the $100(1 - \alpha)$-percentile of the standard normal distribution by x_α, we have $\sqrt{n}\varepsilon/\sigma = x_{\alpha/2}$. It follows that

$$\varepsilon = x_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \qquad (5.19)$$

and the confidence interval for the estimate I with the confidence α is given by

$$\left[\overline{Z}_n - x_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{Z}_n + x_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

When the variance σ^2 is unknown, σ^2 is approximated by the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(h(X_i) - \overline{Z}_n)^2.$$

Equation (5.19) can be seen in another way. Suppose that the confidence coefficient α as well as the confidence interval $\varepsilon > 0$ is given. Then, the problem is how many samples are enough to guarantee the confidence level. The answer is $n = \left(\frac{\sigma x_{\alpha/2}}{\varepsilon}\right)^2$.

The idea behind Monte Carlo method is very easy to understand. Each expected value can be estimated by sampling. The expected value is a weighted sum of function values, where all weights are equal and the sampling points are chosen at random. Using statistics, one can make statements about the accuracy of the result. The uncertainty in the result decreases proportional to the square root of the number of points, independent of the complexity of the original problem.

5.2.3 Generation of Random Numbers

A random number is a realization of a random variable. A random vector is a realization of a multivariate random variable. Random numbers (or vectors) are called iid, if they are realizations of iid (multivariate) random variables. For iid random variables $X_1, X_2, \ldots$, the sample mean $\overline{X}_n$ is an estimator of the population mean $\mu = E(X_1)$. For realizations $x_1, x_2, \ldots$ of those random variables, the realization of $\overline{X}_n$, i.e. $(x_1 + x_2 + \cdots x_n)/n$, is called an estimate of the mean. In statistics, an estimate is a realization of an estimator. In the case of Monte Carlo simulation, the terms "random numbers" and "random variables" are used interchangeably.

Suppose that a random variable X is defined by

$$X = f(Z_1, Z_2, \ldots, Z_n),$$

for some random variables $Z_1, Z_2, \ldots, Z_n$ and a function $f(z_1, z_2, \ldots, z_n)$. Let $(z_1^i, z_2^i, \ldots, z_n^i)$ be the ith random vector of the multivariate random variable $(Z_1, Z_2, \ldots, Z_n)$. Realizations of X are generated by

$$x^i = f(z_1^i, z_2^i, \ldots, z_n^i), \qquad i = 1, 2, \ldots$$

If the random vectors $(z_1^i, z_2^i, ..., z_n^i)$ are iid, the value x^i are iid random numbers of X. According to the strong law of large number,

$$E(X) = \mu = \lim_{n \to \infty} \frac{x^1 + x^2 + \cdots + x^n}{n} \ .$$

A random number generated from the standard uniform distribution $U(0, 1)$ is called a uniform random number.

Interestingly, the Monte Carlo method does not require truly random numbers to be useful. Much of the most useful techniques use deterministic, pseudo-random sequences, making it easy to test and re-run simulations. The only quality usually necessary to make good simulations is for the pseudo-random sequence to appear "random enough" in a certain sense. What this means depends on the application, but typically they should pass a series of statistical tests. Testing that the numbers are uniformly distributed or follow another desired distribution when a large enough number of elements of the sequence are considered is one of the simplest and most common ones.

Computer-generated numbers aren't really random, since computers are deterministic. But, given a number to start with-generally called a random number seed-a number of mathematical operations can be performed on the seed so as to generate unrelated (pseudorandom) numbers. The output of random number generators is tested with rigorous statistical tests to ensure that the numbers are random in relation to one another. One should be aware of this: If you use a random number seed more than once, you will get identical random numbers every time. Thus, for multiple trials, different random number seeds must be used. Commercial programs, like Mathematica, pull a

random number seed from somewhere within the system, perhaps the time on the clock, so the seed is unlikely to be the same for two different experiments.

5.2.3.1 Linear congruential method

The most widely used pseudorandom number generators are linear congruential generators (LCGs). A sequence of integers are generated by the recursive relation

$$x_n = ax_{n-1} + c \pmod{m}, \qquad n = 1, 2, \ldots, \tag{5.20}$$

$$u_n = x_n / m, \tag{5.21}$$

where a, c, and m are preselected constants. There is another preselected constant, x_1, called seed. It is the first number in the LCG output stream. The modular notation "mod" indicates that x_n is the remainder after dividing the quantity $ax_{n-1} + c$ by m.

The sequence $\{x_n\}$ is called pseudo random because they are not random. It is well known that $x_{n+l} = x_n$ for some $l \le m$, so that the same value appears periodically.

Numerical Recipes in C advocates a generator of this form with: a = 1664525, c = 1013904223, m = 2^{32} [66]. The following LCG is used in various operating systems and software packages, such as IMSL and MATLAB [14]:

$$x_n = 16807 x_{n-1} \bmod(2^{31} - 1).$$

5.2.3.2 Multiple-recursive generators

Multiple-Recursive Generators (MCGs) are generalizations of LCGs that have high periods but still can be implemented with single precision floating-point arithmetic.

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_k x_{n-k} \pmod{m}, \tag{5.22}$$

and $$u_n = x_n / m, \tag{5.23}$$

for some positive integer n, positive integers $a_1, a_2, \ldots, a_k$, and seed values $x_0, x_1, \ldots, x_{k-1}$.

Through judicious selection of the parameters, an MCG can have a period as high as $m^n - 1$ [14].

5.2.3.3 Inversive generators

Inversive generators are nonlinear generators which use modular inversion to generate pseudorandom numbers.

An integer $\bar{b}$, $0 \le \bar{b} < m$, is the inverse of an integer b mod m if $b\bar{b} = 1 \bmod m$ for $b \neq 0$ and $\bar{b} = 0$ for b = 0.

Inverse congruential generator (ICG) is defined as

$$x_n = a\bar{x}_{n-1} + c \pmod{m},$$

$$u_n = x_n / m.$$

Explicit inverse congruential generator (EICG) is defined as

$$x_n = \overline{a(n + n_0) + c} \pmod{m},$$

$$u_n = x_n / m.$$

5.2.4 Variance Reduction Methods

From Equation (5.19) we can see that the accuracy of Monte Carlo simulation is given by $C / \sqrt{N}$, where $C = x_{\alpha/2}\sigma$ is the constant of the proportionality. To improve the accuracy of Monte Carlo estimate, we can reduce the variance σ. Any methods to reduce the variance of an estimate are called variance reduction methods.

5.2.4.1 Importance sampling

Importance sampling (IS) is a variance reduction technique that can be used in the Monte Carlo method. The idea behind IS is that certain values of the input random

variables in a simulation have more impact on the parameter being estimated than others. If these "important" values are emphasized by sampling more frequently, then the estimator variance can be reduced.

5.2.4.2 Antithetic variates

In this technique, for some unknown quantity I, we need to find two unbiased estimators Y' and Y'' which have a strong negative correlation. In this case, $\frac{1}{2}(Y' + Y'')$ will be an unbiased estimator of I with variance [60]

$$Var\left[\frac{1}{2}(Y' + Y'')\right] = \frac{1}{4}Var(Y') + \frac{1}{4}Var(Y'') + \frac{1}{2}Cov(Y', Y''). \tag{5.24}$$

The variance can be reduced dramatically if the covariance term is strongly negative.

5.2.4.3 Control variates [58]

Suppose that two estimates X and Y are obtained by the same simulation experiment, and the mean μ_Y of Y is known while that of X is unknown. For some constant α, let

$$Z = X + \alpha\{Y - \mu_Y\}. \tag{5.25}$$

Since $E(Z) = E(X)$, Z is an unbiased estimator for X. The variance of Z is given by

$$Var(Z) = Var(X) + \alpha^2 Var(Y) + 2\alpha Cov(X, Y).$$

When $\alpha = -Cov(X, Y)/Var(Y)$, the variance Var(Z) is minimized and equal to

$$Var(Z) = Var(X) - \frac{(Cov(X, Y))^2}{Var(Y)}.$$

The random variable Y is called a control variate for the estimation of $E(X)$.

5.2.4.4 Stratified sampling [60]

In stratified sampling, the region to be sampled is split into intervals or subregions, and each interval is sampled with a different sample size. The contributions are not added at the level of individually sampled points, but partial sums are added with appropriate weights. The idea of this technique is similar to the idea of importance sampling where more samples are taken in the parts of the region that are more important. The effect of reducing the variance is achieved by gathering more samples in more important subregions.

5.2.5 MC Simulation in Risk Management

Simulation is any analytical method meant to imitate a real-life system, especially when other analyses are too mathematically complex or too difficult to reproduce. In risk management, Monte Carlo simulation methods are used to estimate VaRs and other financial risk measures.

In professional risk assessments, risk combines the probability of a negative event occurring with how harmful that event would be. Financial risk is often defined as the unexpected variability or volatility of returns, and thus includes both potential worse than expected as well as better than expected returns. There are two major areas, market risk and credit risk. Different models are used for different areas. This research focuses on measuring market risk.

The major application of Monte Carlo simulation to risk management is measuring the risk in a portfolio of assets, rather than computing the price of individual securities. Simulation is used in estimating the profit and loss distribution of a portfolio and thus in computing risk measures that summarize this distribution. Particular attention

is given to the problem of estimating the probability of large losses, which requires simulation of rare but significant events.

5.2.5.1 Loss probabilities and value-at-risk

In order to manage market risk, we must measure market risk, especially the risk of large losses. Any method for measuring market risk must address two questions in particular [59]:

1. What statistical model accurately yet conveniently describes the movements in the individual sources of risk and co-movements of multiple sources of risk affecting a portfolio?

2. How does the value of a portfolio change in response to changes in the underlying sources of risk?

The first question asks for the joint distribution of changes in risk factors – the exchange rates, interest rates, stock indices, and bond indices to which a portfolio may be exposed. The second asks for a mapping from risk factors to portfolio values. Once the two questions are answered, the distribution of portfolio profit and loss is determined, and any risk measure that summarizes this distribution can be calculated.

The first question raises statistical issues and the second question raises computational issues. Addressing these two questions inevitably involves balancing the complexity required by the first with the tractability required by the second.

The purpose of this research is to answer these two questions.

The following notation will be used in our discussion:

$\vec{S}$ = vector of m market prices and rates;

Δt = risk-measurement horizon or time interval;

$\Delta \vec{S}$ = change in $\vec{S}$ over the interval Δt;

$V(\vec{S},t)$ = portfolio value at time t and market prices $\vec{S}$;

L = loss over the interval Δt

$$= -\Delta V = V(\vec{S},t) - V(\vec{S} + \Delta\vec{S}, t + \Delta t);$$

$F_L(x) = P(L < x)$, the cumulative distribution of L.

The number m of relevant risk factors could be very large, potentially reaching the hundreds or thousands. In bank supervision the interval Δt is usually quite short, with regulatory agencies requiring measurement over a two-week horizon.

The portfolio's Value-at-Risk is a percentile of its loss distribution over a fixed horizon Δt. For example, the 99% VaR is a point x_p satisfying

$$1 - F_L(x_p) \equiv P(L > x_p) = p$$

with $p = 0.01$. A quantile provides a simple way of summarizing information about the tail of a distribution, and this particular value is often interpreted as a reasonable worst-case loss level. As mentioned in Chapter 1, VaR might more accurately be called a measure of capital adequacy than simply a measure of risk since it is used primarily to determine if a bank has sufficient capital to sustain losses from its trading activities.

The significance of VaR lies in its focus on the tail of the loss distribution. It emphasizes a probabilistic view of risk. And through this probabilistic view, it draws attention to the co-movements of market risk factors in a portfolio-based approach to risk. The more fundamental issues of measuring the tail of the loss distribution, particularly at large loss, is to find $P(L>x)$ for large thresholds x. Once these loss probabilities are

determined, it is a comparatively simple matter to summarize them using VaR or some other measures.

5.2.5.2 Calculating VaR by Monte Carlo Simulation

The idea behind Monte Carlo methods is to simulate repeatedly from the random processes governing the prices or returns of the financial instruments we are interested in. If we are interested in estimating a VaR, each simulation would give use a possible value for our portfolio at the end of the VaR horizon. If we take enough of these simulations, the simulated distribution of portfolio values will converge to the portfolio's unknown 'true' distribution, and we can use the simulated distribution of end-period portfolio values to infer the VaR [19].

The simulation is based on the following assumptions: Only the net loss over the horizon Δt is considered. The composition of the portfolio remains fixed, though the value of its components may change in response to the market movements $\Delta \vec{S}$ and the passage of time Δt. The algorithm may be described as follows:

For each of n independent replications

- generate a vector of market moves $\Delta \vec{S}$;

- revalue the portfolio and compute loss $V(\vec{S}, t) - V(\vec{S} + \Delta \vec{S}, t + \Delta t)$.

Estimate $P(L > x)$ using

$$\frac{1}{n} \sum_{i=1}^{n} 1\{L_i > x\}$$

Where L_i is the loss on the ith replication.

If $\Delta \vec{S}$ is drawn directly from historical data (historical distribution), it is called historical

simulation. However, if $\Delta \vec{S}$ is drawn from a theoretical distribution, it is called Monte

Carlo simulation.

Quantile Estimation: Let $\hat{F}_{L,n}$ denote the empirical distribution of portfolio losses

based on n simulated replications,

$$\hat{F}_{L,n} = \frac{1}{n}\sum_{i=1}^{n} 1\{L_i \leq x\}.$$

A simple estimate of the VaR at probability p (e.g., p = 0.01) is the empirical quantile

$$\hat{x}_p = \hat{F}_{L,n}^{-1}(1-p),$$

With the inverse of the piecewise constant function $\hat{F}_{L,n}$ defined as

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}.$$

Applying piecewise linear interpolation to $\hat{F}_{L,n}$ before taking the inverse generally

produces more accurate quantile estimates [61].

5.2.5.3 Modeling heavy tails

As mentioned before, the normal distribution has shortcomings when used to

model changes in market prices. In fact, the distribution of observed price changes in all

markets displays a higher peak and heavier tails than can be captured with a normal

distribution. This is especially true over short time horizons. High peaks and heavy tails

are characteristic of a market with small price changes in most periods accompanied by

occasional very large price changes [59].

The qualitative property of having a high peak and heavy tails is often measured through kurtosis. The kurtosis of a random variable X with a mean μ is defined as

$$\frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2},$$

assuming X has a finite fourth moment. The kurtosis of a normally distributed random variable is 3. Sometimes, distributions are compared on the basis of excess kurtosis, the difference between the kurtosis and 3.

Kurtosis normalizes the fourth central moment of a distribution by the square of its variance. If two distributions have the same variance, the one with higher kurtosis will usually have a higher peak and heavier tails. Such a distribution is called leptokurtosis.

Kurtosis provides some information about the tails of the distribution, but it is far from a complete measure of the heaviness of the tails. An extension of the normal distribution that provides genuinely heavy tails is the Student t distribution with density

$$f_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \qquad -\infty < x < \infty, \qquad (5.26)$$

where $\Gamma(\cdot)$ is the gamma function. The degree of freedom ν controls the heaviness of the tails. If X has this t_ν density, then

$$P(X > x) \sim \text{constant} \times x^{-\nu}$$

as $x \to \infty$, and ν determines the number of finite moments of $|X|$ such that $E(X^r) < \infty$ for all $r < \nu$ and $E(X^r) = \infty$ for all $r > \nu$, for some $\nu \in (0,\infty)$.

A t_ν random variable can be represented as a ratio $Z\big/\sqrt{Y/\nu}$ in which Z has the standard normal distribution and Y has a chi-square distribution χ_ν^2 and is independent of Z.

To measuring the risk in a portfolio, a multivariate t distribution with heavy-tailed marginals is needed to model the changes in market prices. A t density in $\Re^m$ is given by [59]

$$f_{\nu,\Sigma}(\vec{x}) = \frac{\Gamma((\nu+m)/2)}{(\nu\pi)^{m/2}\Gamma(\nu/2)\,|\Sigma|^{1/2}}\left(1+\frac{1}{\nu}\vec{x}'\Sigma^{-1}\vec{x}\right)^{-(\nu+m)/2}, \qquad \vec{x}\in\Re^m. \tag{5.27}$$

Here, Σ is a symmetric, positive definite matrix, $|\Sigma|$ is its determinant. If $\nu > 2$, the distribution has covariance matrix $\nu\Sigma/(\nu-2)$. In the limit as $\nu\to\infty$, Equation (5.27) becomes the density of the multivariate normal distribution $N(0,\Sigma)$.

If $(X_1,X_2,\dots,X_m)$ have Equation (5.27) as their joint density, then

$$(X_1,X_2,\dots,X_m) =_d \frac{(\xi_1,\xi_2,\dots,\xi_m)}{\sqrt{Y/\nu}}, \tag{5.28}$$

where $=_d$ denotes equality in distribution, $\vec{\xi}=(\xi_1,\xi_2,\dots,\xi_m)$ has the distribution $N(0,\Sigma)$, and Y has distribution χ_ν^2 independent of $\vec{\xi}$. A multivariate t random vector is therefore a multivariate normal vector with a randomly scaled covariance matrix. According to Equation (5.27), the vector $\vec{X}$ with density $f_{\nu,\Sigma}$ can be represented as

$$\vec{X} = \frac{A\vec{Z}}{\sqrt{Y/\nu}} = A\widetilde{\vec{X}}, \tag{5.29}$$

where $AA' = \Sigma$, $\vec{Z} \sim N(0,I)$, and $\widetilde{X}$ is a multivariate t random vector with density $f_{\nu,I}$.

The components of $\widetilde{X}$ are uncorrelated (their correlation matrix is the identity), but not independent. Dependence is introduced by the shared denominator.

The t-copula: Let F_ν denote the CDF of the univariate t_ν distribution. Let the vector $\vec{X}$ have the representation in Equation (5.28) with Σ having all diagonal entries equal to 1. This implies that $X_i \sim t_\nu$ and then that $F_\nu(X_i)$ is uniformly distributed on the unit interval. Just as in the inverse transformation method, applying an inverse distribution $F_{\nu_i}^{-1}$ gives $F_{\nu_i}^{-1}(F_\nu(X_i))$ -- the t distribution with ν_i degrees of freedom. Applying such a transformation to each coordinate produces a vector

$$(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_m) = (F_{\nu_1}^{-1}(F_\nu(X_1)), F_{\nu_2}^{-1}(F_\nu(X_2)), \ldots, F_{\nu_m}^{-1}(F_\nu(X_m))), \tag{5.30}$$

the components of which have t distributions with arbitrary parameters $\nu_1, \nu_2, \ldots, \nu_m$.

Since a copula is defined as a function that joins univariate distribution functions to form multivariate distribution functions, Equation (5.30) is called a t-copula.

5.3 Multivariate Distributions

5.3.1 Introduction

Distributions are one means of expressing uncertainty in a problem.

5.3.1.1 Parameters

A parameter is a constant that characterizes a probability distribution, such as a standard deviation, mean vector, or covariance matrix.

Definition 1: Let X be a random variable. If X is discrete, its expectation is

$$E(X) = \sum_x x\phi(x), \tag{5.31}$$

where $\phi(x)$ is the probability function (PF) of X. If X is continuous,

$$E(X) = \int_{-\infty}^{\infty} x\phi(x)dx,$$

(5.32)

where $\phi(x)$ is the probability density function (pdf) of X.

Definition 2: Suppose X is a random variable and f is a function from $\Re$ to $\Re$. The $f(X)$ is a new random variable whose probability distribution can be inferred from that of X. If X is discrete, the expectation of $f(X)$ is

$$E(f(X)) = \sum_x f(x)\phi(x).$$

(5.33)

. If X is continuous, the expectation of $f(X)$ is

$$E[f(X)] = \int_{-\infty}^{\infty} f(x)\phi(x)dx.$$

(5.34)

Definition 3: If $\mu = E(X)$ is the mean or expected value of the random variable X, the variance of X is

$$Var(X) = \sigma^2 = E[(X - \mu)^2],$$

(5.35)

and the standard deviation of X is

$$Std(X) = \sigma = \sqrt{Var(X)}.$$

(5.36)

In probability theory and statistics, the variance of a random variable is a measure of its statistical dispersion, indicating how far from the expected value its values typically are.

Definition 4: The skewness of a random variable X is defined as

$$Skew(X) = \eta_1 = \frac{E[(X - \mu)^3]}{\sigma^3}.$$

(5.37)

In probability theory and statistics, skewness is a measure of the asymmetry of the probability distribution of a random variable. Roughly speaking, a distribution has positive skew (right-skewed) if the higher tail is longer and negative skew (left-skewed) if the lower tail is longer. Figure 5.3 illustrates two cases of skewness. Both pdf.s have the same expectation and variance. The one on the left is positively skewed. The one on the right is negatively skewed [46].
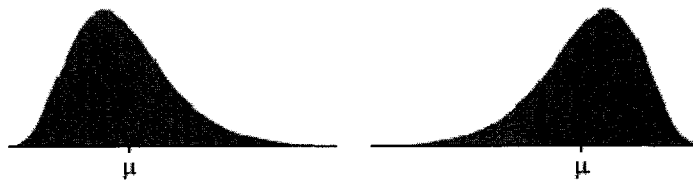
Figure 5.3 The two graphs illustrate the notion of skewness.

For a sample of N values the sample skewness is [62]

$$\hat{\eta}_1 = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^{3/2}}, \tag{5.38}$$

where $\bar{x}$ is the sample mean.

Definition 4: The kurtosis of a random variable X is defined as

$$Kurt(X) = \eta_2 = \frac{E[(X - \mu)^4]}{\sigma^4}. \tag{5.39}$$

In probability theory and statistics, kurtosis is a measure of the "peakedness" of the probability distribution of a random variable. Higher kurtosis means more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations. Figure 5.4 illustrates two cases of skewness. The pdf on the right has higher

kurtosis than the pdf on the left. It is more peaked at the center, and it has fatter tails. The pdf on the left is platykurtic. The one on the right is leptokurtic.

Figure 5.4 The two graphs illustrate the notion of kurtosis [46].

A normal random variable has a kurtosis of 3 irrespective of its mean or standard deviation. If a random variable's kurtosis is greater than 3, it is said to be leptokurtic. If its kurtosis is less than 3, it is said to be platykurtic. Leptokurtosis is associated with pdf.s that are simultaneously "peaked" and have "fat tails." Platykurtosis is associated with PDFs that are simultaneously less peaked and have thinner tails.

Definition 5: The excess kurtosis of a random variable X is defined as

$$Kurt(X) = \eta_2 = \frac{E[(X-\mu)^4]}{\sigma^4} - 3.$$ (5.40)

The "minus 3" at the end of this formula is often explained as a correction to make the kurtosis of the normal distribution equal to zero. A high kurtosis distribution has a sharper "peak" and fatter "tails", while a low kurtosis distribution has a more rounded peak with wider "shoulders".

According to Definition 5, distributions with zero kurtosis are called mesokurtic, such as normal distributions. A distribution with positive kurtosis is called leptokurtic. A distribution with negative kurtosis is called platykurtic.

Definition 6: If X is a random variable with CDF Φ, a q-quantile of X is any value x such that $P(X \le x) = q$.

In most VaR applications, all q-quantiles exist and are unique for $q \in (0,1)$ [21]. In such cases, a q-quantile is a parameter equals to the inverse CDF evaluated at q and is denoted as $x = \Phi^{-1}(q)$.

Definition 7: For any positive integer k, the kth moment of a random variable X is defined as

$$\mu_k' = E(X^k).$$ (5.43)

Its kth central moment is defined as

$$\mu_k = E[(X - \mu)^k].$$ (5.44)

Definition 8: If X_i and X_j are components of a random vector, their (k,l) joint moment is defined as

$$E[X_i^k X_j^l].$$ (5.45)

Their (k, l) joint central moment is defined as

$$E[(X_i - \mu_i)^k (X_j - \mu_j)^l].$$ (5.46)

Definition 9: If X_i and X_j are components of a random vector, their (1,1) joint central moment is called covariance:

$$\Sigma_{i,j} = Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)].$$ (5.47)

If $i = j$, $\Sigma_{i,i} = E[(X_i - \mu_i)^2] = \sigma_i^2$ is X_i's variance.

Covariance is the measure of how much two variables vary together. That is to say, the covariance becomes more positive with each pair of values which differ from their mean in the same direction, and becomes more negative with each pair of values which differ from their mean in opposite directions. In this way, the more often they

differ in the same direction, the more positive the covariance, and the more often they differ in opposite directions, the more negative the covariance.

Definition 10: If X_i and X_j are components of a random vector, their correlation is defined as

$$Corr(X_i, X_j) = \rho_{i,j} = \frac{Cov(X_i, X_j)}{\sigma_i \sigma_j}. \tag{5.48}$$

Correlation indicates the strength and direction of a linear relationship between two random variables.

5.3.1.2 Distributions

1. Uniform distribution: A random variable X defined over the interval (a, b) and having a constant density is said to be uniformly distributed over (a, b), denoted as $U(a,b)$. Figure 5.5 illustrates a uniform distribution. Its pdf is

$$\phi(x) = \frac{1}{b-a} \qquad a < x < b,$$

$$= 0 \qquad \text{otherwise.}$$
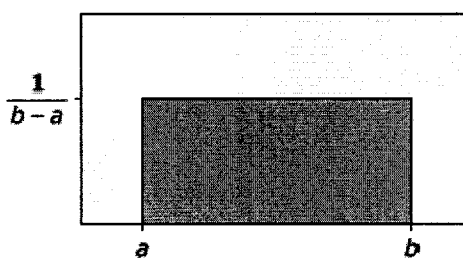
Figure 5.5 The pdf of a uniform distribution [46]

A $U(a,b)$ random variable's properties are listed in Table 5.1.

Table 5.1 Properties of $U(a,b)$

CDF	$\Phi(x) = \dfrac{x-a}{b-a} \qquad a < x < b$
Inverse CDF	$\Phi^{-1}(q) = a + (b-a)q \quad 0 < q < 1$
Mean	$\mu = \dfrac{a+b}{2}$
Standard deviation	$\sigma = \dfrac{b-a}{2\sqrt{3}}$
Skewness	$\eta_1 = 0$
Kurtosis / Excess Kurtosis	$\eta_2 = \dfrac{9}{5} \ / \ \eta_2 = -\dfrac{6}{5}$

2. Normal distribution: A normal distribution function is specified by two parameters: a mean μ and variance σ^2, denoted as $N(\mu,\sigma^2)$. Figure 5.6 shows an example of normal distribution. A $N(\mu,\sigma^2)$ random variable's properties are listed in Table 5.2. Its pdf is

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \qquad (5.49)$$

Its Cumulative Distribution Function (CDF) is

$$\Phi(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du \qquad (5.50)$$

μ
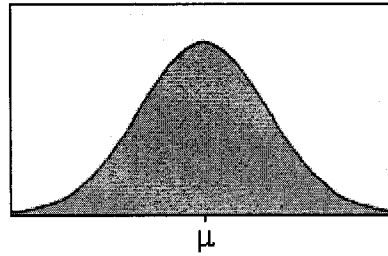
Figure 5.6 The pdf of a normal distribution [46]

Table 5.2 Properties of $N(\mu, \sigma^2)$

Mean	μ
Standard deviation	σ
Skewness	$\eta_1 = 0$
Kurtosis / Excess Kurtosis	$\eta_2 = 3$ / $\eta_2 = 0$

A linear polynomial of a normal random variable is also normal. If

$X \sim N(\mu, \sigma^2)$, then $bX + a \sim N(b\mu + a, (b\sigma)^2)$ for any scalars a, b. This means that any

$N(\mu, \sigma^2)$ random variable X can be expressed as a linear polynomial of some $N(0,1)$

random variable Z:

$$X = \sigma Z + \mu.$$

3. Chi-Squared Distribution: If X_i are k independent, normally distributed random

variables with means μ_i and variances σ_i^2, then the random variable

$$Z = \sum_{i=1}^{k} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$$

is distributed according to the chi-square distribution. This is usually written as

$\chi^2(v,\delta^2)$. If $\delta^2 = 0$, the distribution is centrally chi-squared, denoted as χ_v^2. Figure 5.7

shows an example of chi-square distribution. Its pdf is

$$\phi(x) = \frac{x^{(v-2)/2}\exp(-x/2)}{2^{v/2}\Gamma(v/2)} \qquad x > 0, \qquad (5.51)$$

$$= 0 \qquad\qquad\qquad \text{otherwise.}$$
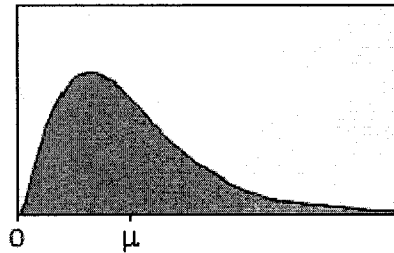
Figure 5.7 The pdf of a chi-squared distribution [46]

A $\chi^2(v,\delta^2)$ random variable's properties are listed in Table 5.3[14]:

Table 5.3 Properties of $\chi^2(v,\delta^2)$

Mean	$\mu = v + \delta^2$
Standard deviation	$\sigma = \sqrt{2(v + 2\delta^2)}$
Skewness	$\eta_1 = 2^{3/2}(v + 3\delta^2)\big/(v + 2\delta^2)^{3/2}$
Kurtosis / Excess Kurtosis	$\eta_2 = 3 + 12(v + 4\delta^2)\big/(v + 2\delta^2)^2$ / $\eta_2 = 12(v + 4\delta^2)\big/(v + 2\delta^2)^2$

4. Student t-Distribution: Suppose $X_1, X_2, \ldots, X_n$ are independent random variables that are normally distributed with expected value μ and variance σ^2. The sample mean is given by

$$\overline{X} = (X_1 + X_2 + \cdots + X_n)/n,$$

and the sample variance is given by

$$S_n^2 = \sum_{i=1}^{n}(X_i - \overline{X}_n)\Big/(n-1).$$

The quantity $T = \dfrac{\overline{X} - \mu}{S_n/\sqrt{n}}$ has the following probability density function

$$\phi_v(x) = \frac{\Gamma((v+1)/2)}{\sqrt{v\pi}\,\Gamma(v/2)}\left(1 + \frac{x^2}{v}\right)^{-(v+1)/2}, \tag{5.52}$$

with v equal to $n-1$. The distribution of t is now called the t-distribution, denoted as t_v.

In probability and statistics, the t-distribution or Student's t-distribution is a probability distribution that arises in the problem of estimating the mean of a normally distributed population when the sample size is small. It is the basis of the popular Student's t-tests for the statistical significance of the difference between two sample means, and for confidence intervals for the difference between two population means.

A t_v random variable's properties are listed in Table 5.4 [14].

Table 5.4 Properties of t_v

mean	$\mu = 0$
Standard deviation	$\sigma = \sqrt{\dfrac{v}{v-2}}$ for $v > 2$
skewness	$\eta_1 = 0$ for $v > 3$
Kurtosis / Excess Kurtosis	$\eta_2 = \dfrac{3v-6}{v-4}$ / $\eta_2 = \dfrac{6}{v-4}$ for $v > 4$

5.3.2 Multivariate Normal Distribution (MVN) [21][33][63]

A random vector $\vec{X} = (X_1, X_2, \ldots, X_n)'$ follows a multivariate normal distribution

with mean $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)'$ and positive definite covariance matrix $\Sigma = [\sigma_{ij}]$ if its

probability density function (pdf) is

$$\phi(\vec{x} \mid \vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{n/2} \mid \Sigma \mid^{1/2}} \exp\left[-\frac{1}{2}(\vec{x} - \vec{\mu})'\Sigma^{-1}(\vec{x} - \vec{\mu}) \right], \qquad (5.53)$$

where $\mid \Sigma \mid$ is the determinant of Σ. The multivariate normal can be written in the

following notation: $\vec{X} \sim N_n(\vec{\mu}, \Sigma)$.

If $\vec{X} \sim N_n(\vec{\mu}, \Sigma)$, B is a constant $m \times n$ matrix and $\vec{A}$ is an m-dimensional

constant vector, then:

$$B\vec{X} + \vec{A} \sim N_m(B\vec{\mu} + \vec{A}, B\Sigma B'). \qquad (5.54)$$

This generalizes the analogous one-dimensional property of univariate normal

distributions.

If $\vec{Z} \sim N_n(\vec{0}, I_n)$, the above $\vec{X} \sim N_n(\vec{\mu}, \Sigma)$ can be expressed as

$$\vec{X} = \Sigma^{1/2}\vec{Z} + \vec{\mu}.$$ (5.55)

5.3.2.1 Bivariate normal distribution

Bivariate normal distribution is a multivariate normal distribution when n = 2.

Here, $\vec{X} = (X_1, X_2)'$, $\vec{\mu} = (\mu_1, \mu_2)'$, and $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. The inverse matrix of Σ is

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}.$$ (5.56)

The probability density function of $\vec{X}$ can be written as [71][72]

$$\phi(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{q}{2(1-\rho^2)}\right], \quad -\infty < x_1, x_2 < \infty$$ (5.57)

where,

$$q = \frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)}{\sigma_2^2},$$ (5.58)

and $\rho = Corr(X_1, X_2) = \sigma_1\sigma_2/\sigma_{12}$ is the correlation coefficient between X_1 and X_2.

The marginal probability functions are

$$f(x_1) = \int_{-\infty}^{\infty}\phi(x_1, x_2)dx_2 = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left[-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right],$$ (5.59)

and

$$f(x_2) = \int_{-\infty}^{\infty}\phi(x_1, x_2)dx_1 = \frac{1}{\sigma_2\sqrt{2\pi}}\exp\left[-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right].$$ (5.60)

5.3.3 Multivariate Skew-Normal Distribution [64][65][66]

5.3.3.1 Univariate skew-normal distribution

Suppose X is a random variable having the probability density function:

$$f(x) = 2\phi(x)\Phi(\alpha x) \tag{5.61}$$

where α is a fixed arbitrary number, $\phi(x)$ is the standard normal density function, and $\Phi(\alpha x)$ is its cumulative distribution function evaluated at point αx. $\phi(x)$ and $\Phi(\alpha x)$ are respectively given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \tag{5.62}$$

and

$$\Phi(\alpha x) = \int_{-\infty}^{\alpha x} \phi(t)dt. \tag{5.63}$$

The density function $f(x)$ has some interesting properties:

1. If $\alpha = 0$, then $f(x)$ becomes the standard normal density function $\phi(x)$.

2. As the absolute value of α increases, the skewness of the distribution increases.

3. If $\alpha \to \infty$, then $f(x)$ converges to the "half-normal" density or "folded" normal distribution.

4. If the sign of α changes, then $f(x)$ is reflected on the opposite side of the vertical axis. If $\alpha > 0$, $f(x)$ is positively skewed; if $\alpha < 0$, $f(x)$ is negatively skewed.

Figures 5.8 and 5.9 [67] illustrate two cases of skew-normal distribution.
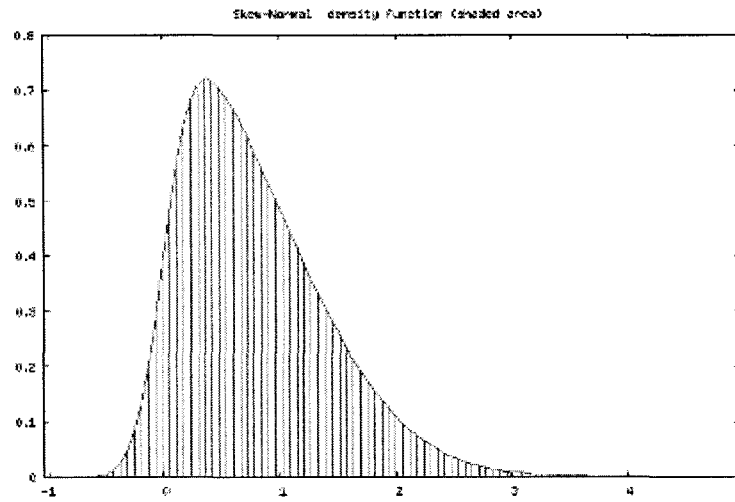
Figure 5.8 Skew-normal density with $\xi = 0$; $\omega = 1$; $\alpha = 5$
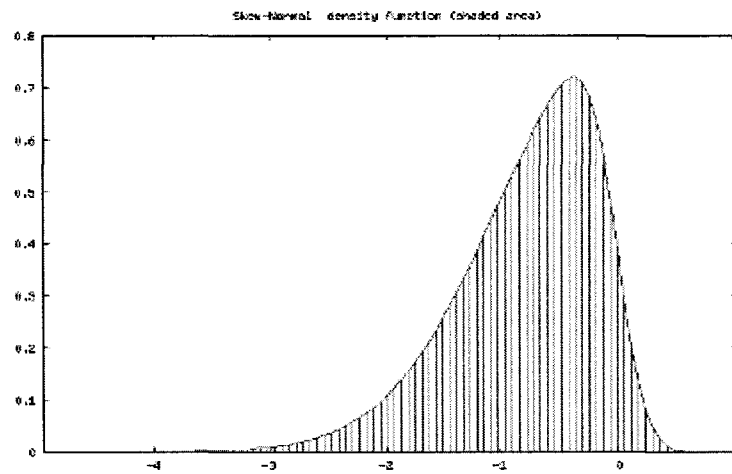
Figure 5.9 Skew-normal density with $\xi = 0$; $\omega = 1$; $\alpha = -5$

The random variable X with density function $f(x)$ in Equation (5.61) can be used to construct the general skew-normal distribution. Consider the linear transformation

$$Y = \xi + \omega X ,$$ (5.64)

The random variable Y has a skew-normal distribution with parameters (ξ, ω, α) and is denoted as

$$Y \sim SN(\xi, \omega^2, \alpha) ,$$ (5.65)

where $\xi, \omega,$ and α are called the location, the scale and the shape parameters. If $\alpha = 0$, $SN(\xi, \omega^2, 0)$ becomes $N(\mu, \sigma^2)$ with $\mu = \xi$ and $\sigma = \omega$.

Given that $\delta = \alpha / \sqrt{1 + \alpha^2}$, the expectation of X is

$$E(X) = \sqrt{\frac{2}{\pi}} \delta ,$$ (5.66)

and the variance of X is

$$Var(E) = 1 - \frac{2}{\pi} \delta^2$$ (5.67)

A $SN(\xi, \omega^2, \alpha)$ random variable Y's properties are listed in Table 5.5.

Table 5.5 Properties of $SN(\xi, \omega^2, \alpha)$

Mean	$\mu = \xi + \omega \left(\dfrac{2}{\pi}\right)^{1/2} \delta$
Standard deviation	$\sigma = \sqrt{\omega^2 \left(1 - \dfrac{2}{\pi} \delta^2\right)}$
Skewness	$\eta_1 = \dfrac{4 - \pi}{2} \times \dfrac{\left(\sqrt{\dfrac{2}{\pi}} \delta\right)^3}{\left(1 - \dfrac{2}{\pi} \delta^2\right)^{3/2}}$
Kurtosis / Excess Kurtosis	$\eta_2 = 2(\pi - 3) \dfrac{\left(\sqrt{\dfrac{2}{\pi}} \delta\right)^4}{\left(1 - \dfrac{2}{\pi} \delta^2\right)^2}$

5.3.3.2 Multivariate skew-normal distribution [66]

Equation (5.61) for univariate skew-normal distribution can be extended to the multivariate case. A k-dimensional random vector $\vec{Z}$ has a multivariate skew-normal distribution if its probability density function is given by

$$f(\vec{z}) = 2\phi_k(\vec{z}; \Omega_Z)\Phi(\vec{\alpha}'\vec{z}) \qquad (\vec{z} \in \Re^k), \qquad (5.68)$$

where $\phi_k(\vec{z}; \Omega_Z)$ is the k-dimensional normal density with zero mean and correlation Ω_Z, $\Phi(\cdot)$ is the $N(0, 1)$ distribution function and $\vec{\alpha}$ is a k-dimensional vector.

If $\vec{\alpha} = \vec{0}$, $f(\vec{z})$ becomes $\phi_k(\vec{z}; \Omega_Z)$ which is the density function for $N_k(\vec{0}, \Omega_Z)$.

$\vec{\alpha}$ is called the shape parameter. The mean vector and the covariance matrix of $\vec{Z}$ are

$$\vec{\mu}_Z = E(\vec{Z}) = (2\pi)^{1/2}\delta, \tag{5.69}$$

and

$$\Sigma_Z = Var(\vec{Z}) = \Omega_Z - \vec{\mu}_Z \vec{\mu}_Z', \tag{5.70}$$

where $\vec{\delta} = \dfrac{1}{(1+\vec{\alpha}'\Omega_Z\vec{\alpha})^{1/2}}\Omega_Z\vec{\alpha}$. $\vec{Z}$ is denoted as $SN_k(\Omega_Z,\vec{\alpha})$.

Equation (5.68) can be extended to allow location and scale parameters. Assume $\vec{Y}$ is a k-dimensional random vector that satisfies the following transformation:

$$\vec{Y} = \vec{\xi} + \omega\vec{Z} \tag{5.71}$$

where $\vec{\xi} = (\xi_1,\xi_2,\ldots,\xi_k)'$ and $\omega = diag(\omega_1,\omega_2,\ldots,\omega_k)$ are location and scale parameters respectively. Here, the diagonal components of ω are positive. The density function of $\vec{Y}$ is

$$f(\vec{y}) = 2\phi_k(\vec{y}-\vec{\xi};\Omega)\Phi(\vec{\alpha}'\omega^{-1}(\vec{y}-\vec{\xi})) \tag{5.72}$$

where $\Omega = \omega\Omega_Z\omega$ is a covariance matrix . $\vec{Y}$ has a multivariate skew-normal distribution and is denoted as $\vec{Y} \sim SN_k(\vec{\xi},\Omega,\vec{\alpha})$.

5.3.3.3 Bivariate skew-normal distribution

Set $k = 2$ in Equation (5.68), the density function of $\vec{Z} = (Z_1,Z_2)'$ is

$$f_2(z_1,z_2) = 2\phi_2(z_1,z_2;\omega)\Phi(\alpha_1 z_1 + \alpha_2 z_2), \tag{5.73}$$

where ω is the off-diagonal element of Ω_Z.

5.3.4 Multivariate t-Distribution [59][68][69]

The random vector $\vec{X} = (X_1, X_2, \ldots, X_n)'$ follows a multivariate t-distribution with mean $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)'$ and positive definite covariance matrix C if its probability density (pdf) function is

$$f_{v,\Sigma}(\vec{x}) = \frac{\Gamma[(v+n)/2]}{(v\pi)^{n/2} \Gamma(v/2) |\Sigma|^{1/2}} \exp\left[1 + \frac{1}{v}(\vec{x} - \vec{\mu})'\Sigma^{-1}(\vec{x} - \vec{\mu})\right]^{-(v+n)/2}, \qquad (5.74)$$

where $\Sigma = (v - 2)C/v$. The multivariate student-t distribution can be written in the following notation: $\vec{X} \sim t_n(v, \Sigma)$.

The multivariate t random vector $\vec{X}$ can be generated by

$$\vec{x} = \vec{\mu} + \sqrt{\frac{v}{s}} \vec{z},$$

where $\vec{z}$ is generated from $N_n(\vec{0}, \Sigma)$ and s is generated from χ_v^2.

5.3.5 Multivariate Skew t-Distribution [70][71]

The family of skew t-distributions is an extension of the Student's t family, via the introduction of a shape parameter which regulates skewness; when shape = 0, the skew t-distribution reduces to the usual t-distribution. The skew t-distribution is related to the skew-normal distribution by the following transformation:

$$\vec{Y} = \vec{\xi} + V^{-1/2}\vec{Z}, \qquad (5.75)$$

where $\vec{Z}$ has the multivariate skew-normal distribution, $SN_n(\vec{0}, \Omega, \vec{\alpha})$, and $V \sim \chi_v^2/v$, independent of $\vec{Z}$. According to Equation (5.72), the density function of $\vec{Z}$ is

$$f(\vec{z}) = 2\phi_n(\vec{z}; \Omega)\Phi(\vec{\alpha}'\omega^{-1}\vec{z}). \qquad (5.76)$$

The density function of $\vec{Y}$ is

$$f(\vec{y}) = 2t_n(\vec{y};v)T_1\left\{\vec{\alpha}'\omega^{-1}(\vec{y}-\vec{\xi})\left(\frac{v+n}{Q_y+v}\right)^{1/2};v+n\right\} \qquad (5.77)$$

where $\omega = diag(\omega_1,\omega_2,\ldots,\omega_n)$ is defined in Section 5.3.3.2,

$$Q_y = (\vec{y}-\vec{\xi})'\Omega^{-1}(\vec{y}-\vec{\xi}),$$

$$t_n(\vec{y};v) = \frac{1}{|\Omega|^{1/2}}g_n(Q_y;v) = \frac{\Gamma\{(v+n)/2\}}{|\Omega|^{1/2}(\pi v)^{n/2}\Gamma(v/2)}\left(1+\frac{Q_y}{v}\right)^{-(v+n)/2},$$

is the density function of n-dimensional t-variate with v degrees of freedom and $T_1(x;v+n)$ represents the scalar t-distribution function with $t+n$ degrees of freedom. The distribution in Equation (5.63) is called multivariate skew t-distribution and is denoted as

$$\vec{Y} \sim St_n(\vec{\xi},\Omega,\vec{\alpha},v), \qquad (5.78)$$

and

$$E(\vec{Y}) = \omega\mu, \qquad \text{for } v > 2,$$

$$Var(\vec{Y}) = \frac{v}{v-2}\Omega - \omega\vec{\mu}\vec{\mu}'\omega,$$

where $\vec{\mu} = \vec{\delta}\left(\frac{v}{\pi}\right)^{1/2}\frac{\Gamma\left\{\frac{1}{2}(v-1)\right\}}{\Gamma\left(\frac{1}{2}v\right)}$ if $v > 1$.

5.3.5.1 Univariate skew t-distribution [70]

Suppose Z is a skew-normal random variable denoted by $Z \sim SN(0,\omega^2,\alpha)$, random variable Y has a skew t-distribution through the following transformation:

$$Y = \xi + V^{-1/2}Z \qquad (5.79)$$

where $V \sim \chi_v^2 / v$ and Y is denoted as $Y \sim St(\xi, \omega^2, \alpha, v)$.

A $St(\xi, \omega^2, \alpha, v)$ random variable Y's properties are listed in Table 5.6.

Table 5.6 Properties of $St(\xi, \omega^2, \alpha, v)$

Mean	$\mu = \left\{ \delta \left(\dfrac{v}{\pi}\right)^{1/2} \dfrac{\Gamma\left[\frac{1}{2}(v-1)\right]}{\Gamma\left(\frac{1}{2}v\right)} \right\} \omega,$ for $v > 1$
Standard Deviation	$\sigma = \sqrt{\omega^2 \left[\dfrac{v}{v-2} - \left(\dfrac{\mu}{\omega}\right)^2 \right]},$ for $v > 1$
Skewness	$\eta_1 = \dfrac{\mu}{\omega} \left[\dfrac{v(3-\delta^2)}{v-3} - \dfrac{3v}{v-2} + 2\left(\dfrac{\mu}{\omega}\right)^2 \right] \left[\dfrac{v}{v-2} - \left(\dfrac{\mu}{\omega}\right)^2 \right]^{-3/2}$ for $v > 3$
Kurtosis/Excess Kurtosis	$\eta_2 = \left[\dfrac{3v^2}{(v-2)(v-4)} - \dfrac{4\left(\frac{\mu}{\omega}\right)^2 v(3-\delta^2)}{v-3} + \dfrac{6\left(\frac{\mu}{\omega}\right)^2 v}{v-2} - 3\left(\dfrac{\mu}{\omega}\right)^4 \right] \left[\dfrac{v}{v-2} - \left(\dfrac{\mu}{\omega}\right)^2 \right]^{-2} - 3$

5.4 Simulation of Multivariate Distributions

5.4.1 General Sampling Methods

5.4.1.1 Inverse method

Let F(x) be the distribution function of a random variable X defined on $\mathfrak{R}$, and suppose the inverse function $F^{-1}(x)$ exists.

$$X = F^{-1}(U), \qquad U \sim U(0,1).$$

Given a sequence of iid uniform random numbers $\{u_n\}$ and the inverse function $F^{-1}(x)$, we can generate a sequence of iid random numbers generated from X according to the equation

$$x_n = F^{-1}(u_n), \qquad n = 1,2,\ldots \qquad (5.80)$$

This method to generate the random numbers is called the inverse transformation method and is useful for the case where the inverse function $F^{-1}(x)$ is known in closed form.

Since the standard normal distribution function $\Phi(x)$ cannot be represented in closed form, an analytical form of its inverse function is not known. Hence, when applying the inverse transformation method to the normal random numbers, we need either to employ a numerical technique to calculate Equation (5.80), or to develop an approximation of the inverse function $\Phi^{-1}(x)$ [72].

5.4.1.2 Conditional distribution approach [73]

Suppose that $\vec{X} = (X_1,\ldots,X_k)'$ is k-dimensional random vector which can be expressed in a multivariate distribution function $f(\vec{x})$. The vector $\vec{x} = (x_1,\ldots,x_k)$ can be generated by the conditional distribution approach as follows:

1. Generate $X_1 = x_1$ from the marginal distribution of X_1.

2. Generate $X_2 = x_2$ from the conditional distribution of X_2 given $X_1 = x_1$.

3. Generate $X_3 = x_3$ from the conditional distribution of X_2 given $X_1 = x_1$ and $X_2 = x_2$.

$$\vdots$$

k. Generate $X_k = x_k$ from the conditional distribution of X_k given $X_1 = x_1$, $X_2 = x_2$, ..., and $X_{k-1} = x_{k-1}$.

The advantage of this approach is that it reduces the problem of generating a k-dimensional random vector into a series of k univariate generation problem, which make it possible to use techniques available for the univariate case.

5.4.1.3 Transform method [73]

Distributions with intractable inverse functions can be generated with a transformation other than $F^{-1}(U)$. For example, the t distribution with ν degree of freedom can be generated by

$$t_\nu = X_1 / (X_2 / n)^{1/2},$$

where X_1 is standard normal, X_2 is χ_n^2, and X_1 is independent of X_2. This method can be used to construct univariate and multivariate distributions.

5.4.1.4 Acceptance-rejection (AR) method [58]

For generating samples from an arbitrary probability distribution function $f(x)$ by using an instrumental distribution $g(x)$ under the only restriction that $f(x) < cg(x)$ where $c > 1$ is an appropriate bound on $f(x)/g(x)$. One uses the following approach.

1. Generate x from $g(x)$.

2. Generate U from $U(0,1)$.

3. Check whether or not $U < f(x) / cg(x)$.

   If this holds, accept x as a realization of $f(x)$;

   if not, reject the value of x and go to step 1.

The AR method is used in cases where $f(x)$ is difficult to sample from directly. This difficulty can be avoided by sampling from a distribution function $g(x)$, which is easy to

sample from. This method can be used to construct univariate and multivariate distributions.

5.4.2 Simulation of Multivariate Normal Distributions

5.4.2.1 Generating univariate normals

There are several algorithms available for generating univariate normal distributions:

1. Box-Muller Method [59][73] – The following steps can be used to generate pairs of independent standard normally distributed random numbers $\vec{Z} = (Z_1, Z_2) \sim N_2(\vec{0}, I_2)$, given a source of independent uniformly distributed random numbers $U_1, U_2 \sim U(0,1)$.

    a. Generate $U_1, U_2 \sim U(0,1)$;

    b. $Z_1 = \sqrt{-2\log(U_1)}\cos(2\pi U_2)$ and $Z_2 = \sqrt{-2\log(U_1)}\sin(2\pi U_2)$

    c. Repeat a and b to generate enough number of normal variates.

2. The Inverse Method (see section 5.4.1.1)[59][73] To compute $\Phi^{-1}(U)$, we can use Newton's method, $x_{n+1} = x_n - \dfrac{\Phi(x_n) - U}{\phi(x_n)}$, to find $\Phi(x^*) = U$.

5.4.2.2 Simulation results for univariate normals

The inverse method is used to generate univariate normal distribution. All the simulation processes in this chapter are written in C language, and Microsoft Excel is used to do graphical analysis.

First, standard normal distribution is generated by using different sample sizes, N = 100, N = 500, N = 5000, and N = 10000. Figure 5.10 shows the simulation results: as

the sample size increases, the distribution looks more like a bell-shaped normal distribution probability function.
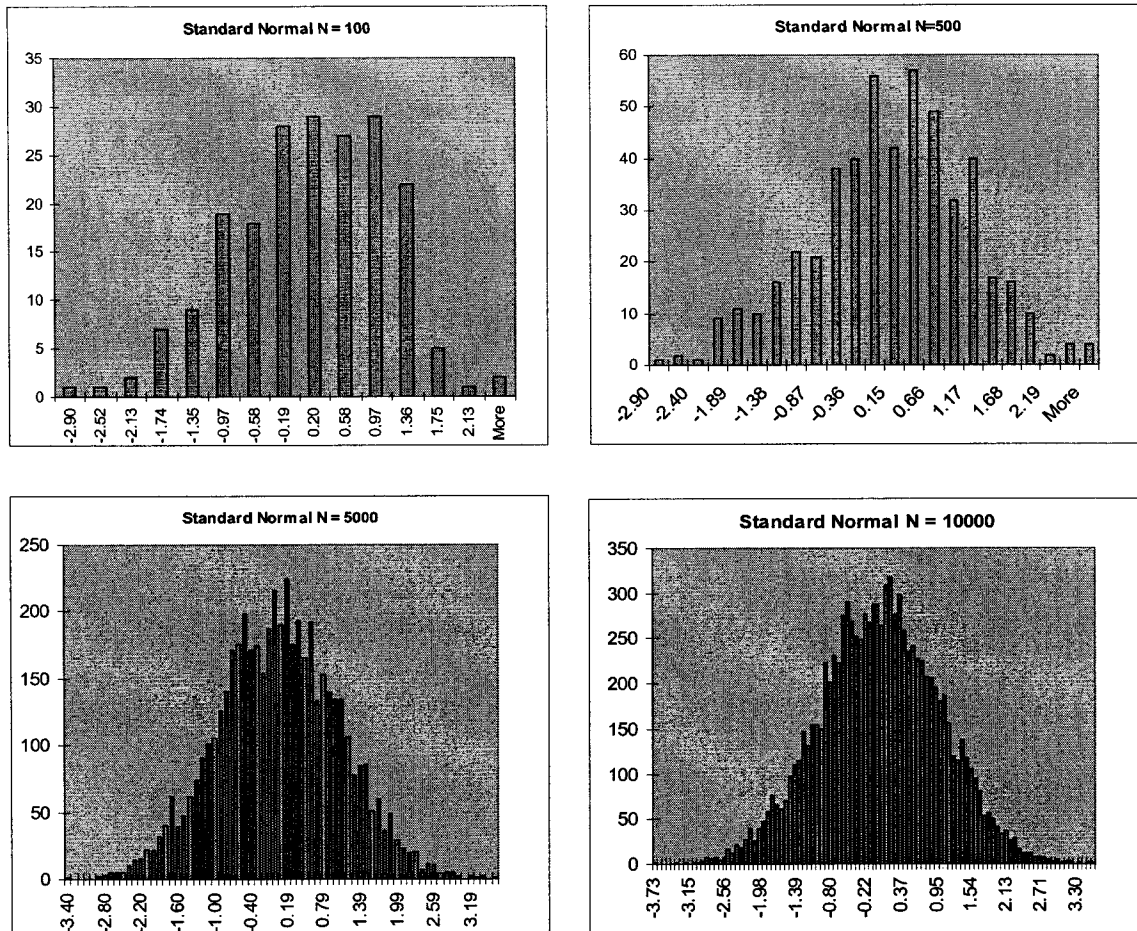
Figure 5.10 Histograms of a standard normal distribution with different sample size

Second, normal distributions with different means and variances are generated by using the same sample size, $N = 10000$. Figure 5.11 (a) shows the simulation results with the same variance: the cumulative distribution functions for $N(-2,4)$, $N(0,4)$, and $N(2,4)$. Figure 5.11 (b) shows the simulation results with the same mean: the cumulative distribution functions for $N(0,0.2)$, $N(0,0.5)$, $N(0,1)$ and $N(0,5)$.
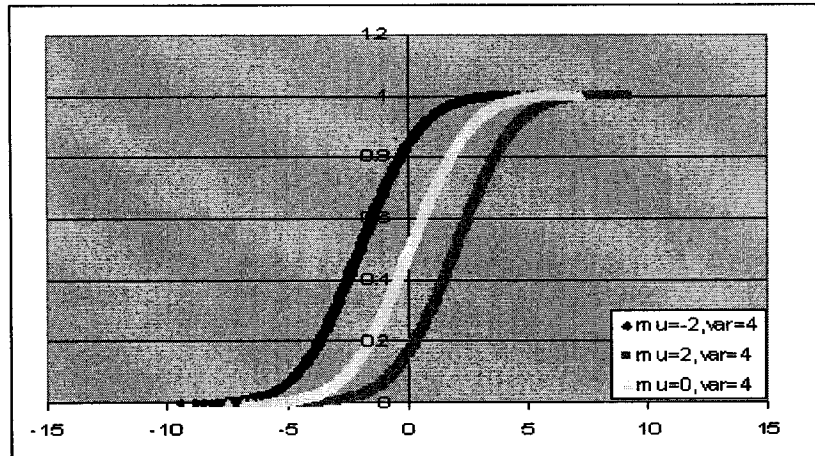
Figure 5.11 (a) Simulation of CDF of normal distributions with same variance but different means.
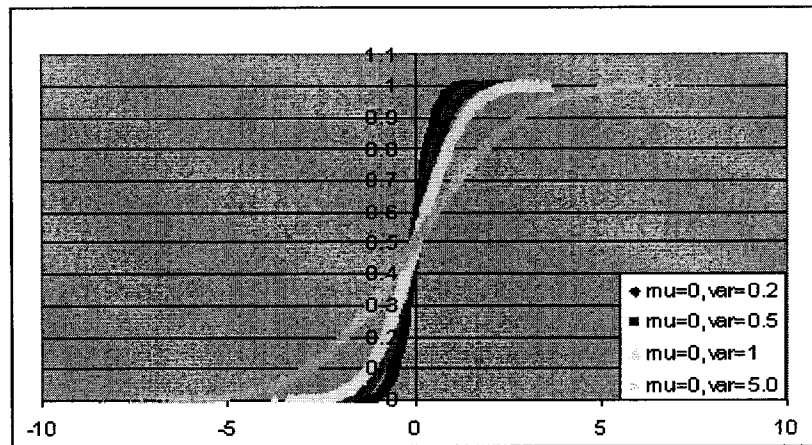
Figure 5.11 (b) Simulation of CDF of normal distributions with same mean but different variances.

5.4.2.3 Generating multivariate normal
         distributions

A widely used method for drawing a random vector X from the n-dimensional

multivariate normal distribution with mean vector $\bar{\mu}$ and covariance matrix Σ (required to

be symmetric and positive definite) works as follows:

1. Compute the Cholesky decomposition (matrix square root) of Σ, that is, find the unique lower triangular matrix C such that $CC' = \Sigma$.

2. Let $\vec{Z} = (z_1, z_2, \ldots, z_n)$ be a vector whose components are n independent standard normal variates.

3. Let $\vec{X}$ be $\vec{\mu} + C\vec{Z}$.

Let $\vec{X} = (X_1, X_2, \ldots, X_n)'$ follow an n-variate normal distribution with a mean vector $\vec{\mu} = (\mu_1, \mu_2, \cdots, \mu_n)'$, $\mu_i = E(X_i)$, and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}, \quad \sigma_{ij} = Cov[X_i, X_j].$$

For some non-singular matrix C, the decomposition is not unique. Here we use the Cholesky Decomposition method.

Proposition 3 (Cholesky Decomposition): Given the covariance matrix $\Sigma = (\sigma_{ij})$, the matrix $C = (c_{ij})$ can be defined as follows:

$$c_{11} = \sqrt{\sigma_{11}},$$

$$c_{i1} = \frac{\sigma_{i1}}{\sqrt{\sigma_{11}}}, \qquad i = 1, 2, \ldots, n,$$

$$c_{jj} = \sqrt{\sigma_{jj} - \sum_{k=1}^{j-1} c_{jk}^2}, \qquad j = 2, 3, \ldots, n,$$

$$c_{ij} = \frac{1}{c_{jj}}\left(\sigma_{ij} - \sum_{k=1}^{j-1} c_{ik} c_{jk}\right), \qquad j < i, \quad i = 2, 3, \ldots, n-1.$$

Thus, the matrix C is a lower triangular.

The next result suggests how to obtain random vectors distributed as $N_n(\vec{\mu}, \Sigma)$ from iid normal random numbers.

Proposition 4: Suppose that $\vec{Z} \sim N_n(\vec{0}, I)$ and that a positive definite covariance matrix Σ is decomposed as $CC' = \Sigma$. Then, we have

$$\vec{Y} = C\vec{Z} + \vec{\mu} \quad \sim \quad N_n(\vec{\mu}, \Sigma).$$

Conversely, for $\vec{Y} \sim N_n(\vec{\mu}, \Sigma)$, we have

$$\vec{Z} = C^{-1}(\vec{Y} - \vec{\mu}) \quad \sim \quad N_n(\vec{0}, I).$$

Based on Proposition 4, iid random vectors distributed as $N_n(\vec{\mu}, \Sigma)$ can be generated as follows:

1. Generate iid normal random numbers $z_1, z_2, \ldots, z_n \sim N(0,1)$, and define

$$\vec{z} = (z_1, z_2, \ldots, z_n)'.$$

2. Define the random vector $\vec{y} = (y_1, y_2, \ldots, y_n)'$ by

$$\vec{y} = C\vec{z} + \vec{\mu}.$$

3. Repeat step 1 and 2 until enough random vectors are generated.

5.4.2.4 Simulation results for MVN

According to the simulation method discussed in Section 5.4.2.3, C program is written for generating multivariate normal distributions. Figures 5.12 – 5.13 are some examples on bivariate and trivariate normal distributions. Excel is used to do graphical analysis. There are multiple purposes for this demonstration:

First, verify if the method is working; second, verify if the program (the method implemented in C) is correct; picture the results directly to help us better understand the multivariate normal distribution. For example, in Figure 5.12 (a), since there is no

correlation between $X1$ and $X2$, and they both have the same variance, the scatter plot shows a round shape which is just what we expect from the bivariate normal distribution function. In Figue 5.12 (b), other things stay the same, only the correlation change from 0 to 0.5. The scatter plot becomes an elliptical shape and rotates $45°$ from the x-axes. If the correlation change to -0.5, we can expect that the scatter plot is still an elliptical shape but rotate $135°$ to reflect the negative correlated fact between $X1$ and $X2$. This rotation is shown in Figure 5.12 (b). If the correlation coefficient keeps the same, but the variances of $X1$ and $X2$ are changed to values larger than 1, the direction of the scatter plot doesn't change, but the area is larger which demonstrates that the larger the standard deviation, the more dispersed the data is. Compare (a) and (b) in Figure 5.12, it is not difficult to get this conclusion.

(a)   (b)
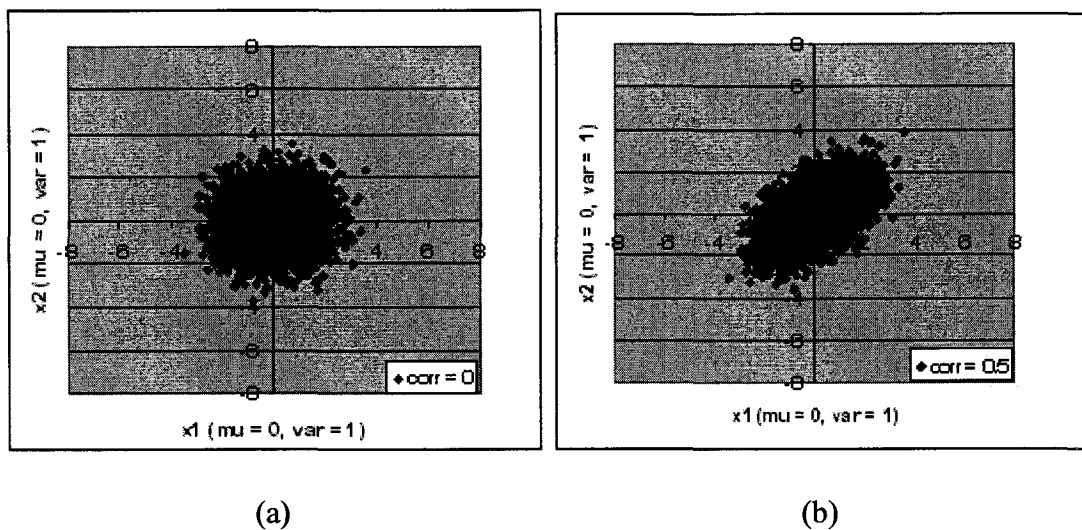
Figure 5.12 (a) Scatter plots of bivariate normal distributions with same sample size $N =$ 5000: (a) $\rho_{12} = 0$ and (b) $\rho_{12} = 0.5$.

(a)   (b)
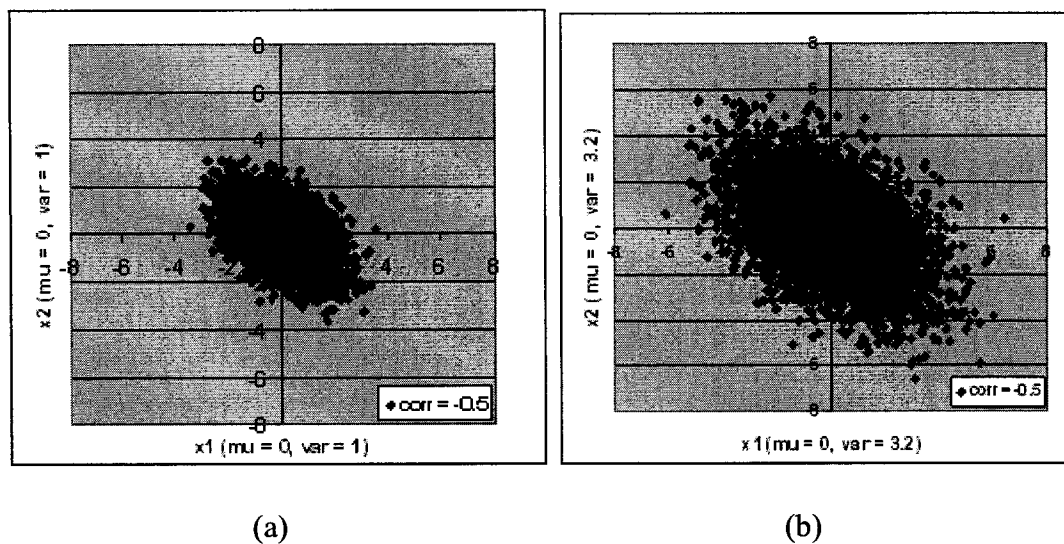
Figure 5.12 (b) Scatter plots of bivariate normal distributions with same sample size $N =$ 5000: (a) $\rho_{12} = -0.5$ and (b) $\rho_{12} = 0.5$.

Figure 5.13 shows the simulation results for a trivariate normal distribution with

variance-covariance matrix $\Sigma_X = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$ and mean $\vec{\mu}_X = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. Figure 5.13 (a) is the

scatter plot of $X1$ vs. $X2$; figure 5.13 (b) is the scatter plot of $X1$ vs. $X3$; figure 5.13 (c) is

the scatter plot of $X2$ vs. $X3$; and figure 5.13 (d) is the overlay of $X1$ vs CDF($X1$), $X2$ vs.

CDF($X2$) and $X3$ vs. CDF($X3$).

(a)

(b)

(c)

(d)
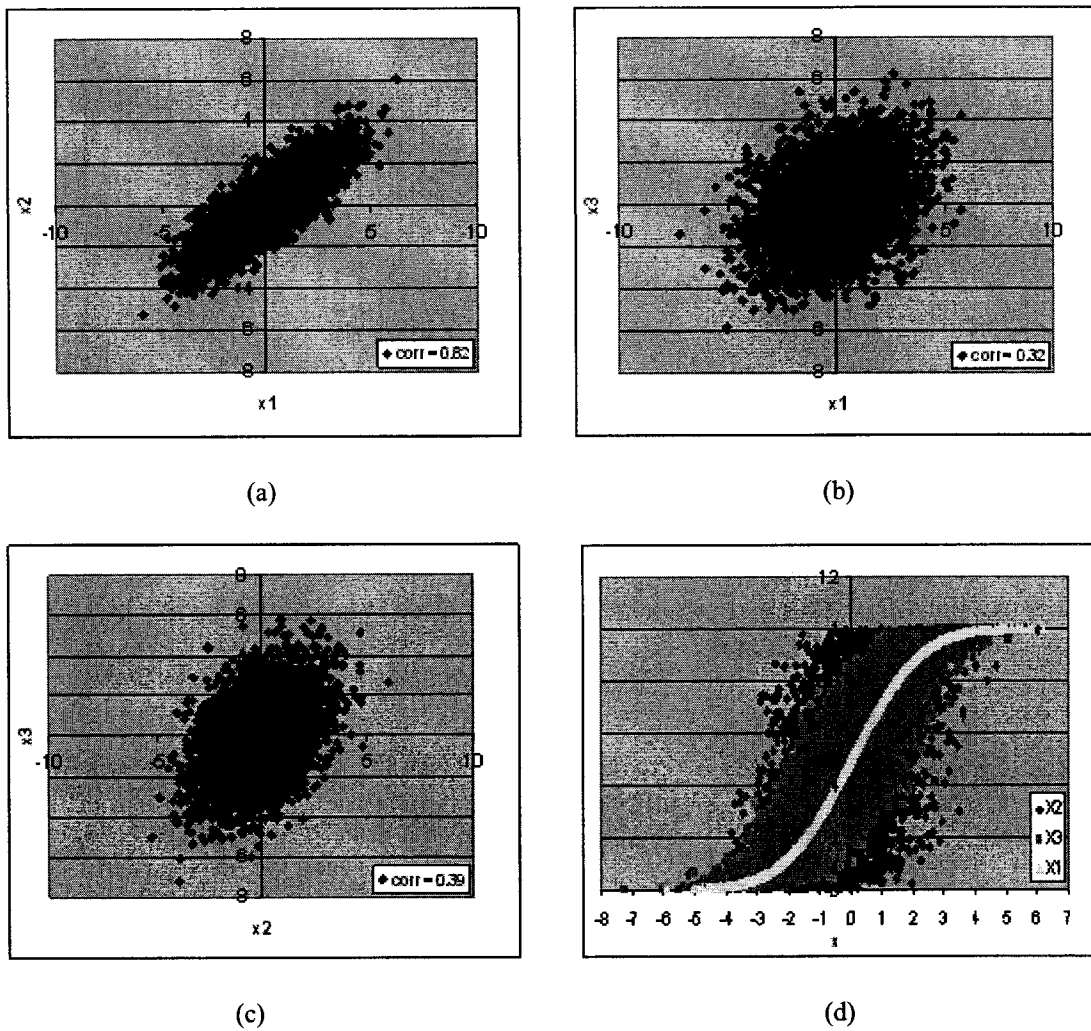
Figure 5.13 Scatter plots of trivariate normal distribution with sample size $N = 5000$

The covariance matrix calculated from sampled data is $\hat{\Sigma}_X = \begin{pmatrix} 2.98 & 2 & 0.963 \\ 2.003 & 2.013 & 0.962 \\ 0.963 & 0.962 & 2.975 \end{pmatrix}$ and

the correlation matrix is $\hat{\rho} = \begin{pmatrix} 1 & 0.818 & 0.323 \\ 0.818 & 1 & 0.393 \\ 0.323 & 0.393 & 1 \end{pmatrix}$.

5.4.3 Simulation of Multivariate
      Skew-Normal Distributions

Skew-normal distribution is a family of distributions including the normal, but with an extra parameter to regulate skewness. Azzalini and Vale [64] investigate its properties. According to the properties, we can find ways to generate random numbers from skew-normal distribution.

5.4.3.1 Generating univariate
      skew-normal distributions

Azzalini and Vale [64] listed the following properties for the univariate skew-normal distribution.

Proposition 1. If Y and W are independent $N(0,1)$ variates, and Z is set equal to Y conditionally on $\alpha Y > W$, for some real α, then $Z \sim SN(0,1,\alpha)$.

For generating random numbers of $Z \sim SN(0,1,\alpha)$, Azzalini and Vale [64] suggest a more efficient way to avoid rejection of samples:

If $(\alpha Y > W)$, then $Z = Y$.

Otherwise, $Z = -Y$.

Proposition 2. If (X, Y) is a bivariate normal random vector with standardized marginals and correlation δ, then the conditional distribution of Y given $X > 0$ is $Y \sim SN(0,1,\alpha(\delta))$.

The correlation coefficient δ is related to the shape parameter α by the functions

$$\delta(\alpha) = \frac{\alpha}{(1+\alpha^2)^{1/2}}, \quad \alpha(\delta) = \frac{\delta}{(1-\delta^2)^{1/2}}. \tag{5.81}$$

The simple way to generate random numbers Y for $Y \sim SN(0,1,\alpha(\delta))$ according to proposition 2 is,

If ($X > 0$), then $Z = Y$.

Otherwise, $Z = -Y$.

Proposition 3. *If* Y_0 *and* Y_1 *are independent N(0, 1) variables and* $\delta \in (-1,1)$, *then*

$$Z = \delta \mid Y_0 \mid +(1-\delta^2)^{1/2} Y_1 \qquad (5.82)$$

is $Z \sim SN(0,1,\alpha(\delta))$.

We can also generate skew-normal random numbers according to proposition 3.

5.4.3.2 Simulation results for univariate
skew-normal distributions

Figure 5.14 shows the simulation results with different shape parameters based on a standard normal distribution. Figure 5.14 (a) shows that when the shape parameter α is zero, the distribution becomes a standard normal distribution. Figure 5.14 also shows that the larger the shape parameter is, the more skewed the distribution is. When the shape parameter changes from a positive number to a negative numbers, the distribution changes from positive skewed to negative skewed.
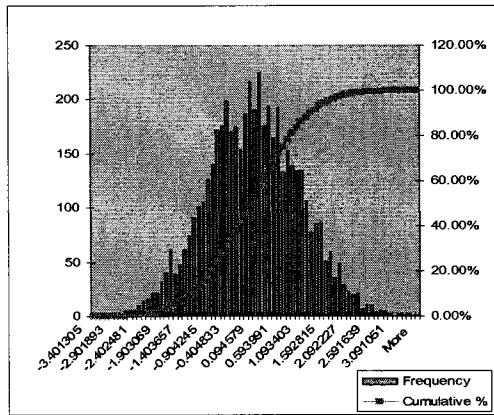
(a) $\alpha = 0$
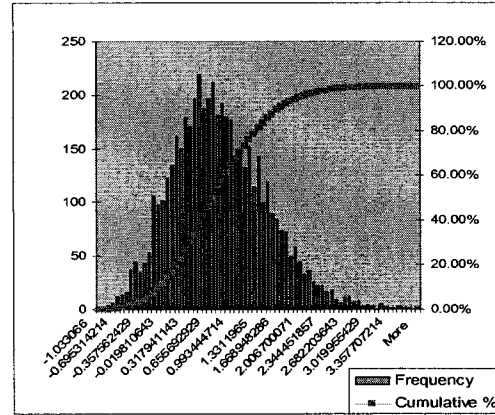
(b) $\alpha = 2$
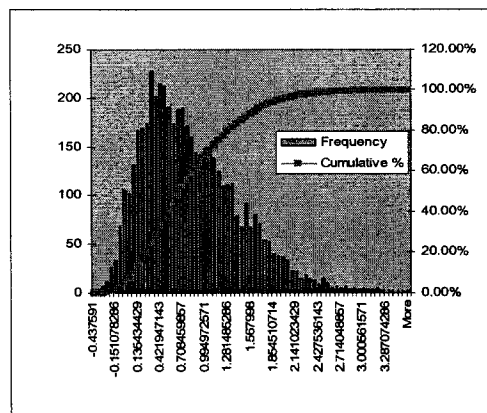
(c) $\alpha = 5$
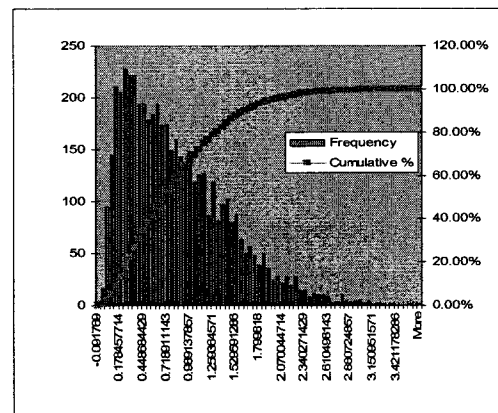
(d) $\alpha = 20$
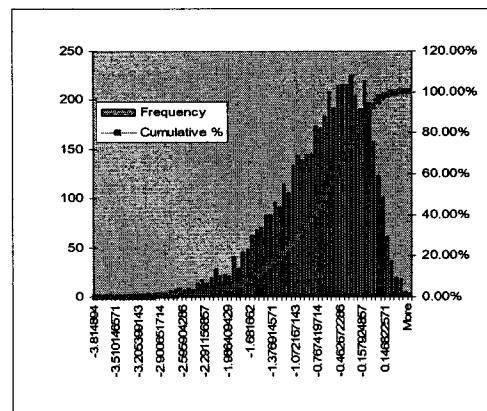
(e) $\alpha = -5$
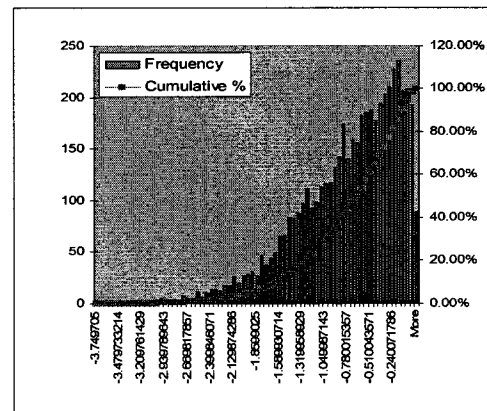
(f) $\alpha = -50$

Figure 5.14 Histograms of the $SN(0,1,\alpha)$ for (a) $\alpha = 0$, (b) $\alpha = 2$, (c) $\alpha = 5$, (d) $\alpha = 20$, (e) $\alpha = -5$ and (f) $\alpha = -50$.

5.4.3.3 Generating multivariate
skew-normal distributions

There are two methods for generation of multivariate skew-normal distributions [64].

1. Transformation method. Suppose $\vec{Y} = (Y_1, Y_2, ..., Y_k)'$ is a k-dimensional normal random vector with standardized marginals, independent of $Y_0 \sim N(0,1)$; thus

$$\begin{pmatrix} Y_0 \\ \vec{Y} \end{pmatrix} \sim N_{k+1} \left\{ 0, \begin{pmatrix} 1 & 0 \\ 0 & \Psi \end{pmatrix} \right\}, \tag{5.83}$$

with Ψ is a $k \times k$ correlation matrix. If $(\delta_1, \delta_2, ..., \delta_k)$ are in the range (-1, 1), define

$$Z_j = \delta_j \mid Y_0 \mid + (1 - \delta_j^2)^{1/2} Y_j \quad \text{for} \quad j = 1, 2, ..., k, \tag{5.84}$$

so that $Z_j \sim SN(0, 1, \lambda(\delta_j))$, according to Equation (5.82) in Proposition 3. $\vec{Z}$ is denoted as $SN_k(\vec{\lambda}, \Psi)$.

2. Conditioning method. Suppose $\vec{Y} = (Y_1, Y_2, ..., Y_k)'$ is a k-dimensional normal random vector and Y_0 is a scalar random veriable, such that

$$\begin{pmatrix} Y_0 \\ \vec{Y} \end{pmatrix} \sim N_{k+1} \left\{ 0, \begin{pmatrix} 1 & \delta' \\ \delta' & \Omega \end{pmatrix} \right\},$$

let $\vec{Z}$ be defined as

$$\vec{Z} \leftarrow \vec{Y}; \quad Y_0 > 0,$$

$$\vec{Z} \leftarrow -\vec{Y}; \quad \text{Otherwise.}$$

Hence $\vec{Z} \sim SN_k(\vec{\lambda}(\delta), \Psi\}$, where $\lambda_j(\delta_j) = \dfrac{\delta_j}{(1 - \delta_j)^{1/2}}$ and the elements of Ψ are from

those of Ω by solving the following equation:

$$corr(Y_i, Y_j) = \omega_{ij} = \delta_i \delta_j + \psi_{ij}(1 - \delta_i^2)(1 - \delta_j^2)^{1/2}.$$

5.4.3.4 Simulation results

Figures 5.15 and 5.16 display the simulation results with different shape parameters $\vec{\alpha} = (\alpha_1, \alpha_2)$ based on a bivariate standard normal distribution, and sample size $N = 5000$.

(a) $\rho_{1,2} = 0$, $\vec{\alpha} = (0,0)'$ (b) $\rho_{1,2} = 0$, $\vec{\alpha} = (0.707,0.707)'$

(c) $\rho_{1,2} = 0$, $\vec{\alpha} = (1,5)'$ (e) $\rho_{1,2} = 0$, $\vec{\alpha} = (40,-40)'$
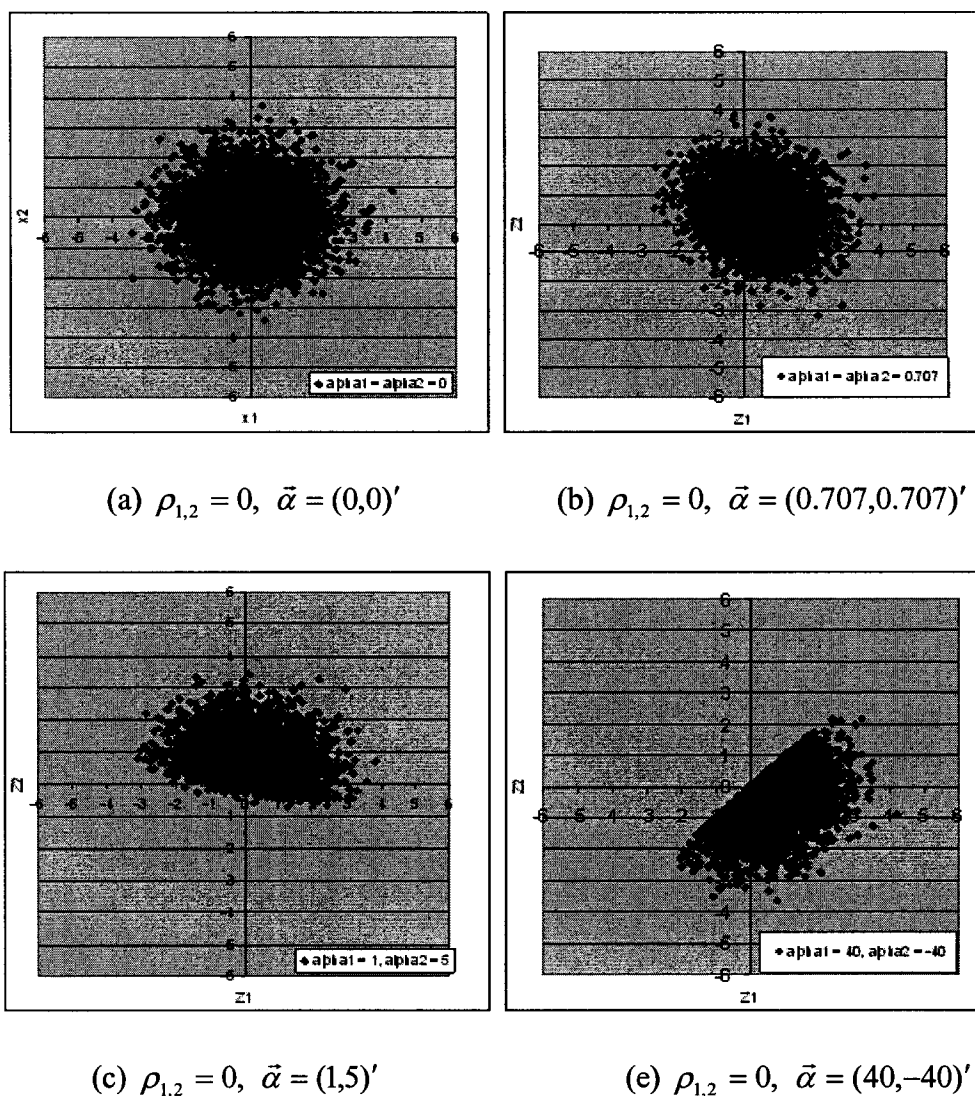
Figure 5.15 Scatter plots of bivariate skew-normal distributions with various shape parameters, zero correlation between $X1$ and $X2$.

(a) $\rho_{1,2} = 0.4,\ \vec{\alpha} = (0,0)'$ (b) $\rho_{1,2} = 0.4,\ \vec{\alpha} = (0.45,0.45)'$

(c) $\rho_{1,2} = 0.4,\ \vec{\alpha} = (1.584,-1.188)'$ (d) $\rho_{1,2} = 0.4,\ \vec{\alpha}(-1.323,6.086)'$
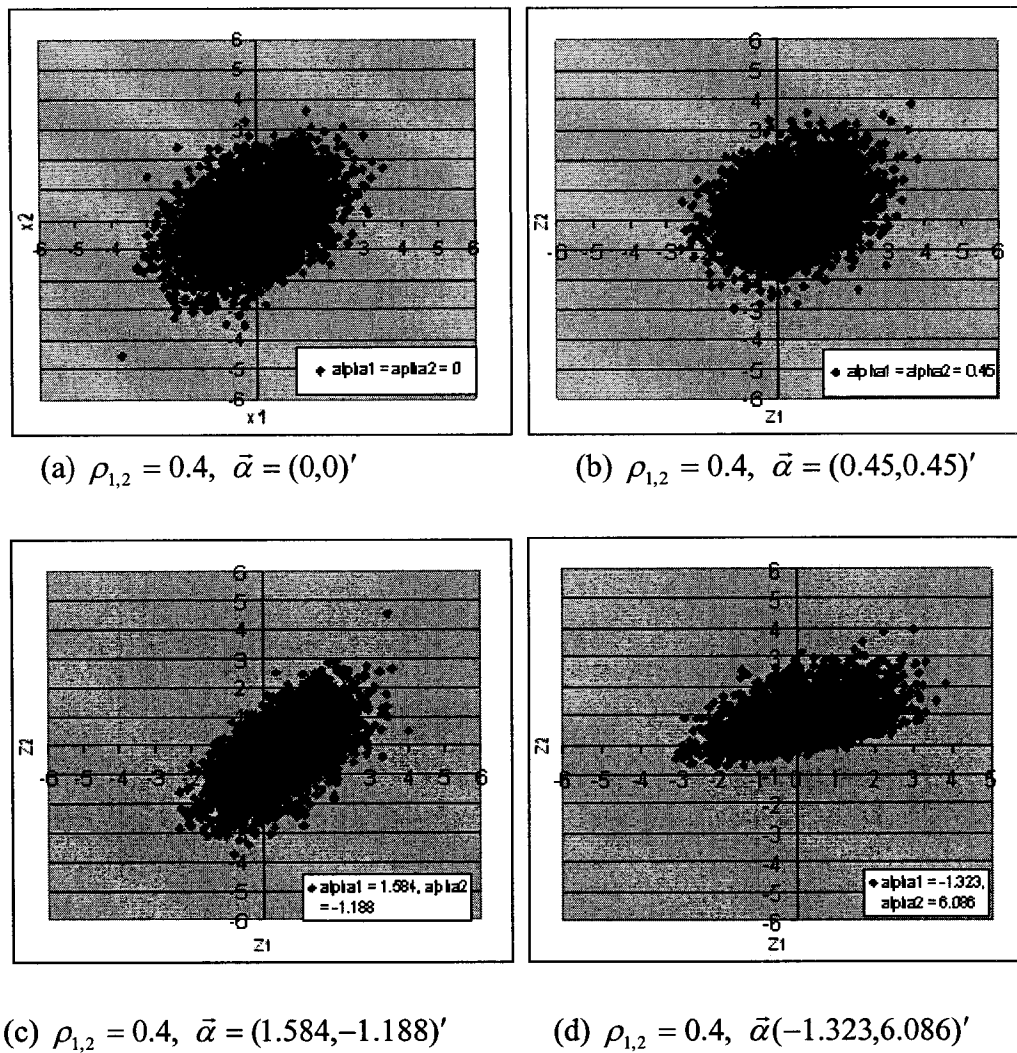
Figure 5.16 Scatter plots of bivariate skew-normal distributions with various shape parameters, and a correlation coefficient between $X1$ and $X2\ \rho_{1,2} = 0.4$.

As Figures 5.15 (a) and 5.16 (a) show, if the shape parameter $\vec{\alpha}$ is zero, a bivariate skew-normal distribution becomes bivariate normal distribution. Generally, the patterns of the scatter plots change depending on the signs and magnitude of the shape parameters.

5.4.4 Simulation of Multivariate
t-Distributions

5.4.4.1 Generating univariate t

Let Z have a normal distribution with mean 0 and variance 1. Let V have a chi-square distribution with v degrees of freedom. Further suppose that Z and V are independent. Then the ratio

$$\frac{Z}{\sqrt{V/v}}$$

has a t-distribution with v degrees of freedom.

5.4.4.2 Simulation results for univariate
t-distribution

Figure 5.17 shows the simulation results with different degrees of freedom based on a standard normal distribution. The sample size for this simulation is $N = 1000$. The degrees of freedom determine the shape of the t-distribution. The smaller the degrees of freedom are, the higher the peak is and the fatter the tail is. The larger the degrees of freedom are, the closer the distribution is to the standard normal distribution. For $v = \infty$, the t-distribution converges to the normal distribution.

(a) $v = 5$

(b) $v = 10$

(c) $v = 30$ (d) $v = \infty$
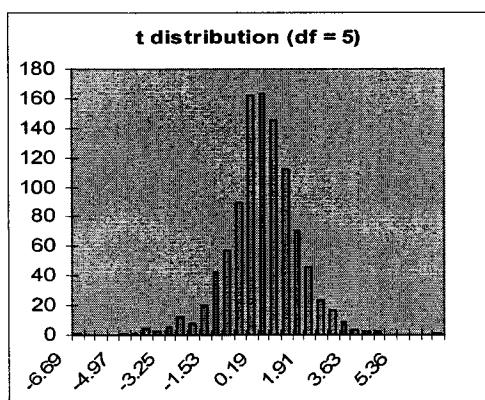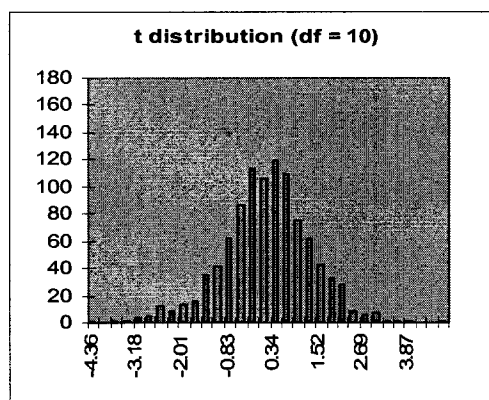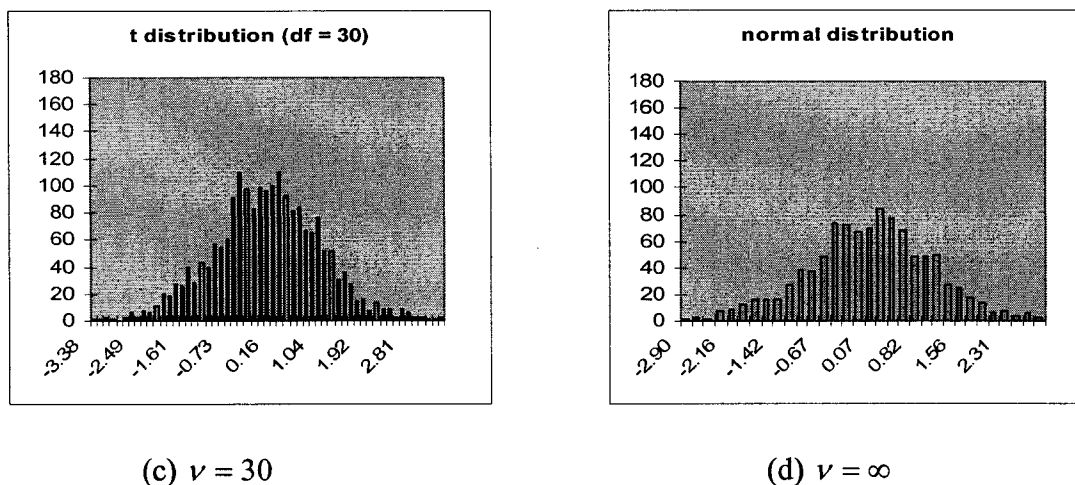
Figure 5.17 Histograms of various t-distribution with different degrees of freedom v.

5.4.4.3 Generating multivariate t

According to Equation (5.27), the vector $\vec{X}$ with density $f_{v,\Sigma}$ can be represented as

$$\vec{X} = \frac{A\vec{Z}}{\sqrt{Y/v}}, \tag{5.85}$$

where $AA' = \Sigma$, $\vec{Z} \sim N(0,I)$.

5.4.4.4 Simulation results for multivariate t

Figure 5.18 shows the simulation results with degrees of freedom 10 based on bivariate and trivariate standard normal distributions and no correlations between the normal distribution variables.
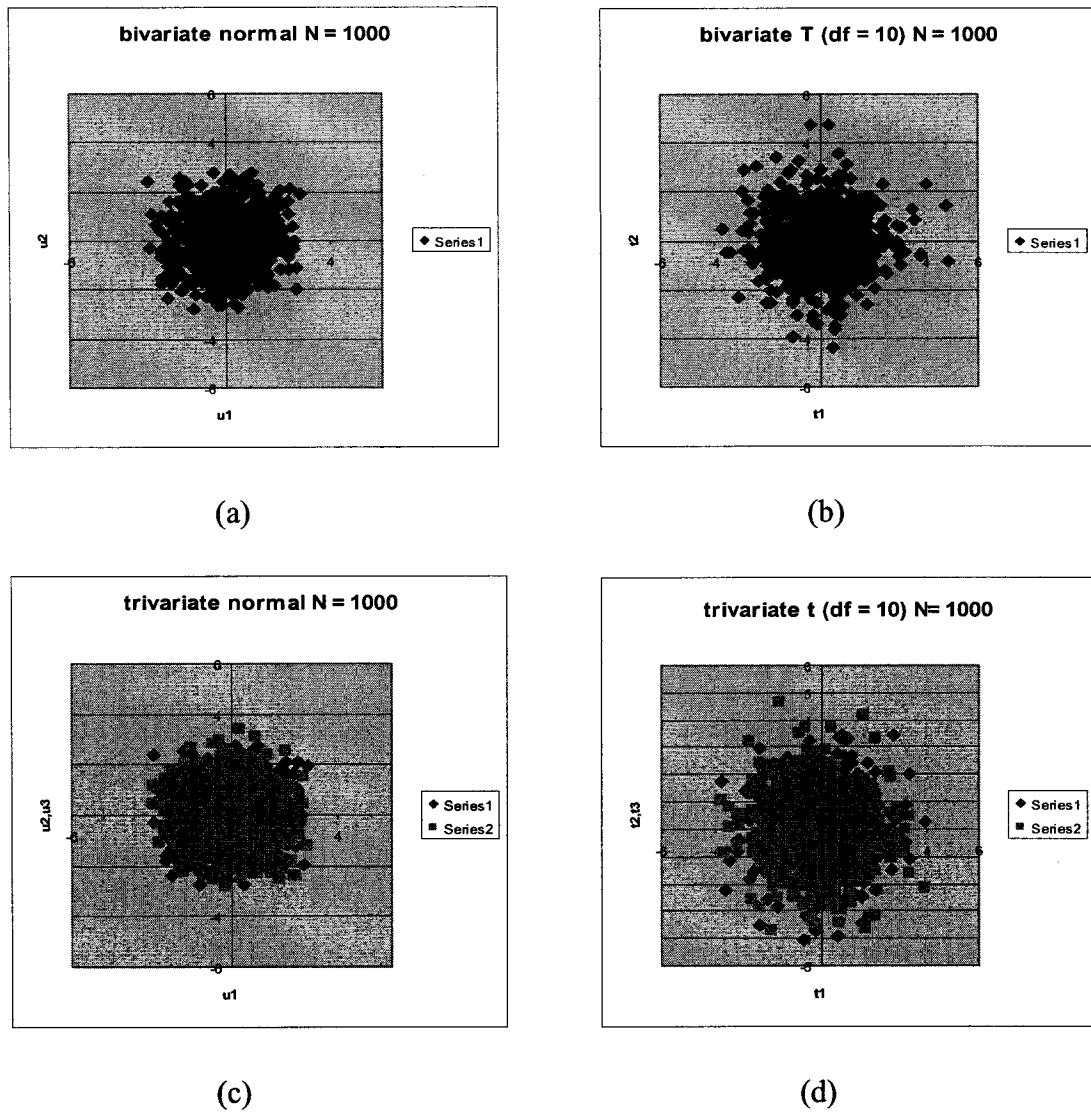
Figure 5.18 (a) Scatter plot of a bivariate normal distribution with zero correlations between the random variables. (b) Scatter plot of a bivariate t-distribution with $v = 10$. (c) The overlay of scatter plots of a trivariate normal distribution. (d) The overlay of scatter plots of a trivariate t-distribution.

Figure 5.19 shows the simulation results with degrees of freedom 10 based on a bivariate standard normal distribution with a correlation coefficient is 0.6. The sample size for this simulation is $N = 1000$. From those figures we can see that compared with normal distributions, the points in t-distributions are more spread out.

(a)   (b)
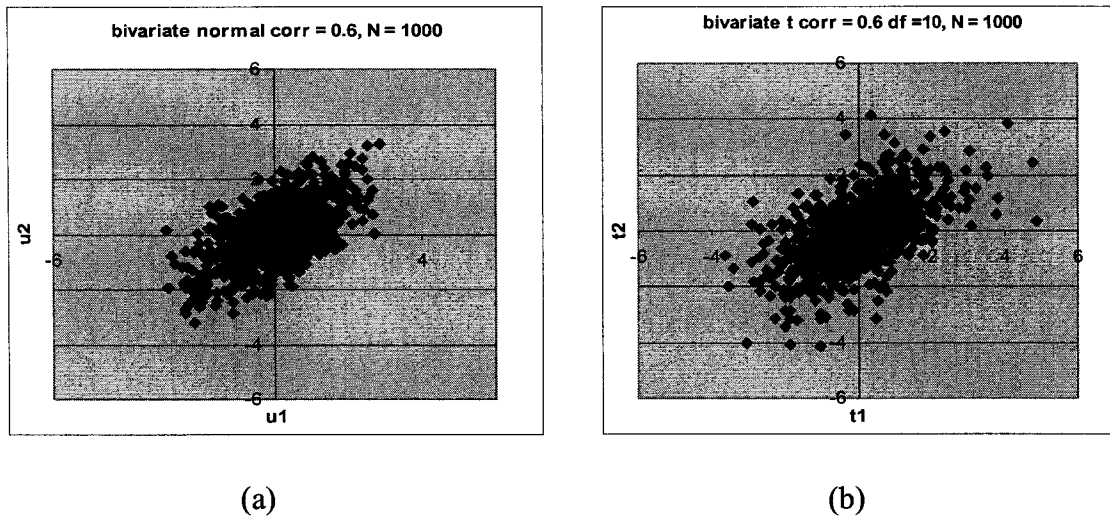
Figure 5.19 (a) Scatter plot of bivariate normal distribution, $\rho_{12} = 0.6$. (b) Scatter plot of bivariate t-distribution with degree of freedom $v = 10$.

Figure 5.20 shows the simulation results with degree of freedom $v = 10$ based on

a trivariate standard normal distribution and a correlation matrix $\Sigma = \begin{pmatrix} 1 & 0.6 & 0.6 \\ 0.6 & 1 & 0.3 \\ 0.6 & 0.3 & 1 \end{pmatrix}$.

The estimated correlation matrix from the 1000 samples is

$$\hat{\Sigma} = \begin{pmatrix} 1.0087 & 0.6079 & 0.5697 \\ 0.6079 & 1.0101 & 0.2912 \\ 0.5697 & 0.2912 & 0.9618 \end{pmatrix}.$$

trivariate normal N=1000

u2,u3

u1

◆ u2 vs u1
■ u3 vs u1

(a)

trivariate normal N = 1000

u1,u3

u2

◆ u1 vs u2
■ u3 vs u2

(b)

trivariate T df = 10 N = 1000

t2

t1

◆ t2 vs t1
■ t3 vs t2

(c)

trivariate t df = 10 N = 1000

t1,t3

t2

◆ t1 vs t2
■ t3 vs t1

(d)
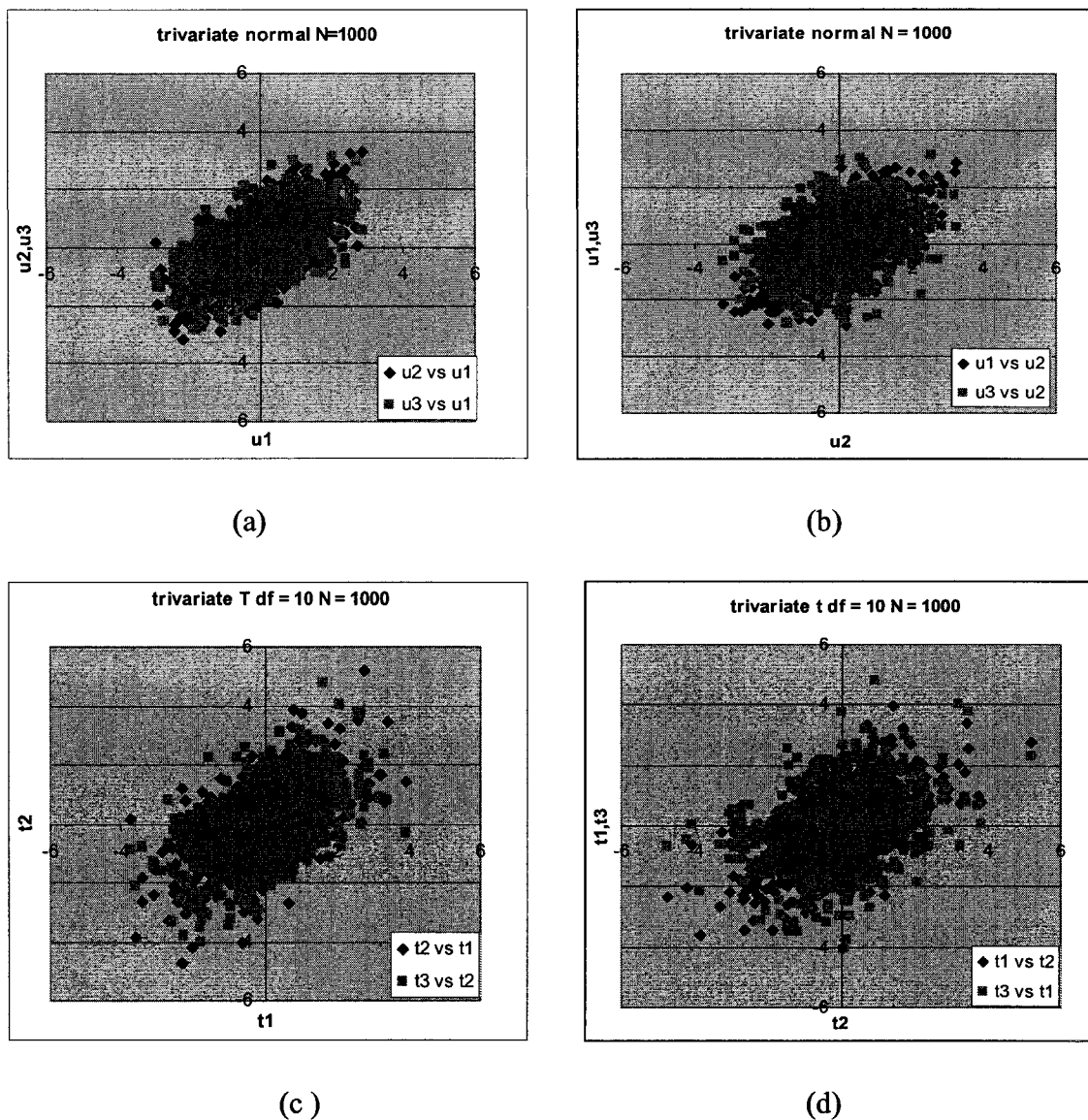
Figure 5.20 (a) and (b) present an overlay of scatter plots of a trivariate normal distribution, $\rho_{12} = \rho_{13} = 0.6$ and $\rho_{23} = 0.3$. (c) and (d) present an overlay of scatter plots of a trivariate t-distribution.

5.4.5 Simulation of Multivariate Skew
t-Distributions

5.4.5.1 Generating univariate skew t

To generate random numbers for univariate skew t-distribution, first, generate a univariate skew normal random number Z following the method discussed in Section

5.4.3.1, then generate a chi-squared random number V with degrees of freedom v, and

let $w = z / \sqrt{V / v}$, the function w has the skew t-distribution with v degrees of freedom,

denoted as $ST(0,1,\alpha,v)$. If location and scale parameters ξ and ω are included,

$Y = \xi + \omega Z$ and is denoted as $ST(\xi,\omega,\alpha,v)$.

5.4.5.2 Simluation results

Figure 5.21 displays four different distributions together to illustrate the changes and differences between them.

(a) normal

(b) skew-normal $\alpha = 3$

(c) t $v = 10$

(d) skew t $\alpha = 3$, $v = 10$
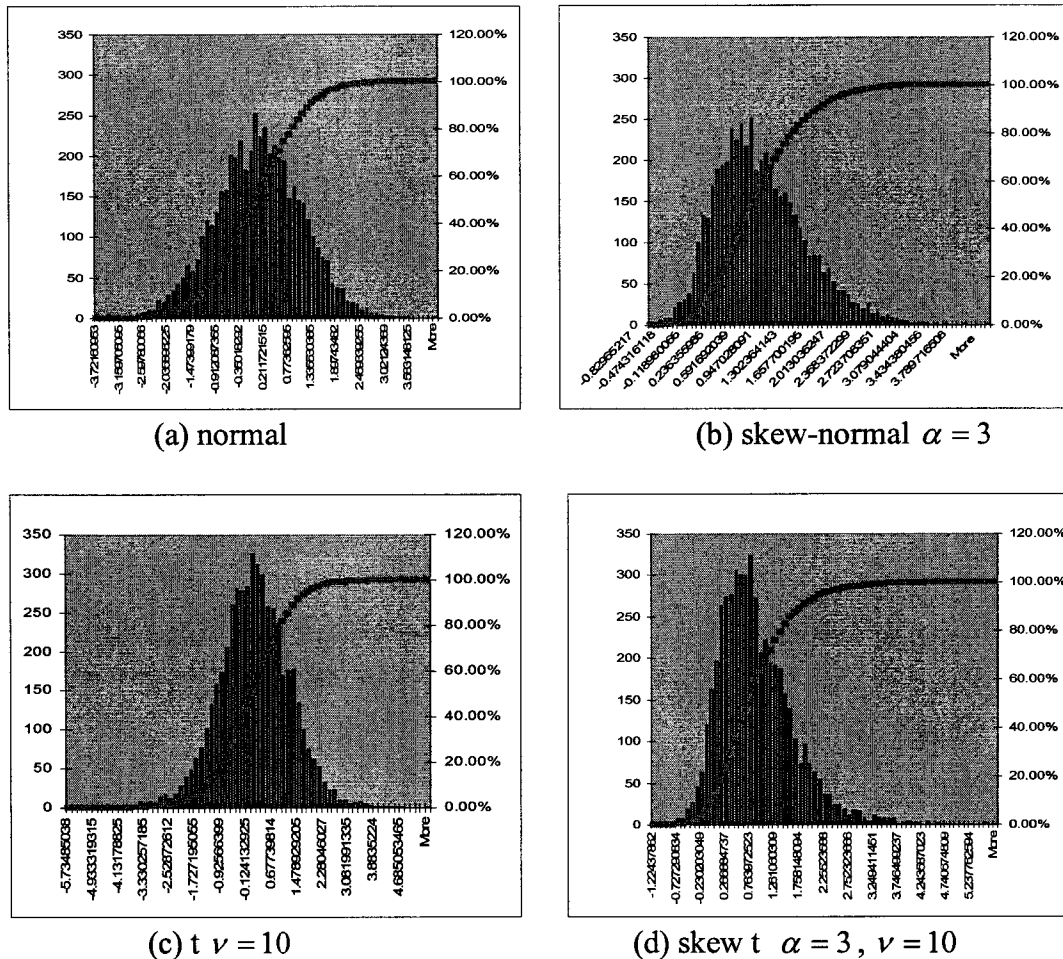
Figure 5.21 (a) Histogram of a normal distribution; (b) Histogram of a skew-normal distribution with $\alpha = 3$ based on (a); (c) Histogram of a t-distribution with $v = 10$ based on (a); (d) Histogram of skew t-distribution with $\alpha = 3$ and $v = 10$ based on (a).

Here the sample size is $N = 5000$. The shape parameter α is 3 for skewness and the degrees of freedom v is 10 to construct t related distributions.

5.4.5.3 Generating multivariate skew t

The method for generating univariate skew t-distribution random numbers can be extended to generate random vector from multivariate skew t-distribution. First, generate a k-dimensional multivariate skew normal vector $\vec{Z}$, then generate a chi-squared random number V with degrees of freedom v, and let $\vec{W} = \vec{Z} / \sqrt{V/v}$, $\vec{W}$ has the multivariate skew t-distribution with v degrees of freedom, denoted as $ST_k(0,I,\vec{\alpha},v)$. Take into account of location and scale parameters, $Y = \vec{\xi} + \vec{\omega}\vec{Z}$ is denoted as $ST_k(\vec{\xi},\Omega,\vec{\alpha},v)$. Detail can be found in Section 5.3.5.

5.4.5.4 Simulation results

Figure 5.22 shows the simulation results of four different bivariate distributions based on the following correlation matrix and parameters:

$$\Omega = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}, \quad \vec{\delta} = \begin{pmatrix} 0.6 \\ -0.3 \end{pmatrix}, \quad v = 5. \text{ The shape parameter calculated from } \vec{\delta} \text{ is}$$

$$\vec{\alpha} = (1.6,-1.2)'.$$

(a) bivariate normal   (b) bivariate skew-normal

(c) bivariate t (d) bivariate skew t
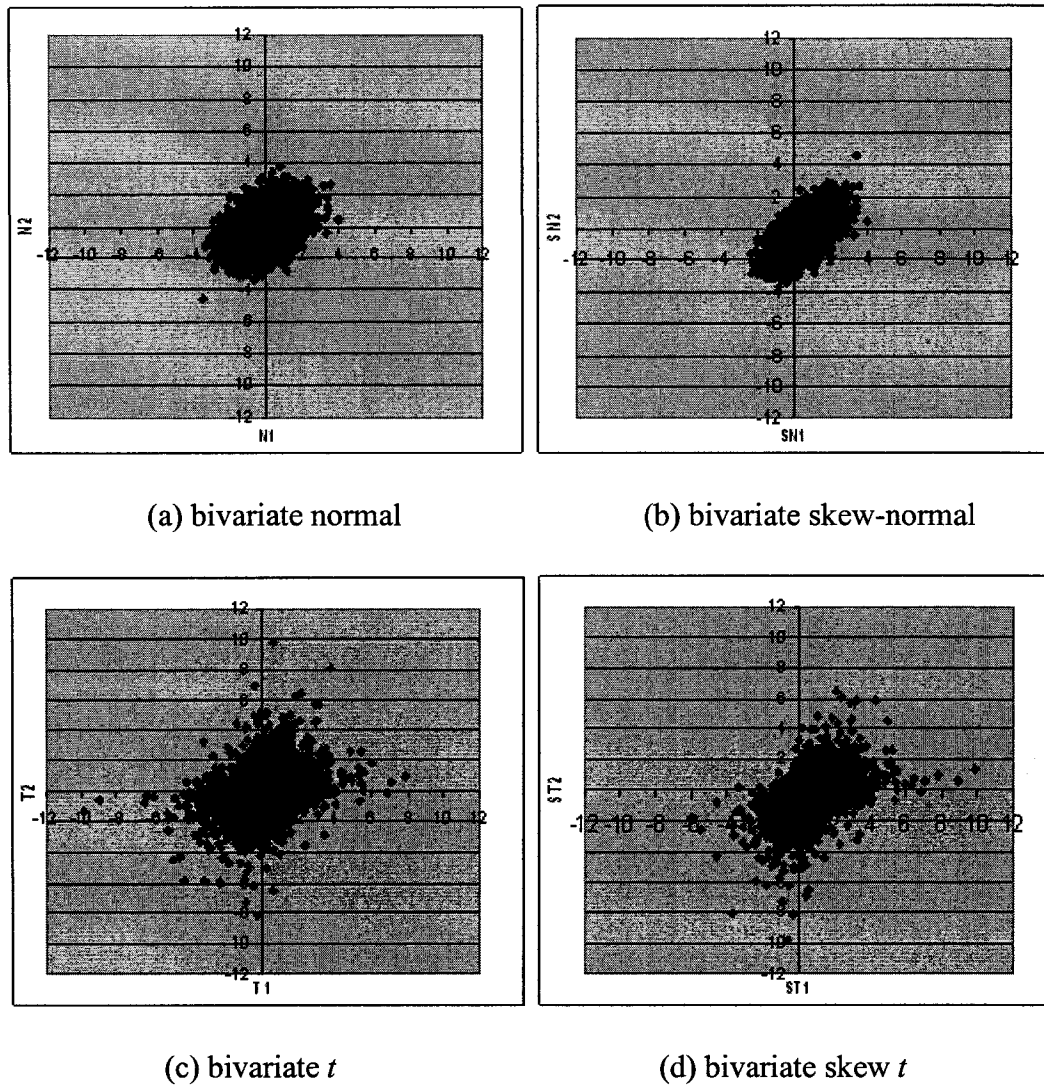
Figure 5.22 Four different bivariate distributions

Figure 5.23 shows the simulation results of four different trivariate distributions

based on the following correlation matrix and parameters: $\Omega = \begin{pmatrix} 1 & 0.6 & 0.3 \\ 0.6 & 1 & 0.5 \\ 0.3 & 0.5 & 1 \end{pmatrix}$,

$\vec{\delta} = \begin{pmatrix} 0.6 \\ 0.5 \\ 0.4 \end{pmatrix}$, $v = 8$.

(a) trivariate normal

(b) trivariate skew-normal

(c) trivariate t

(d) trivariate skew t
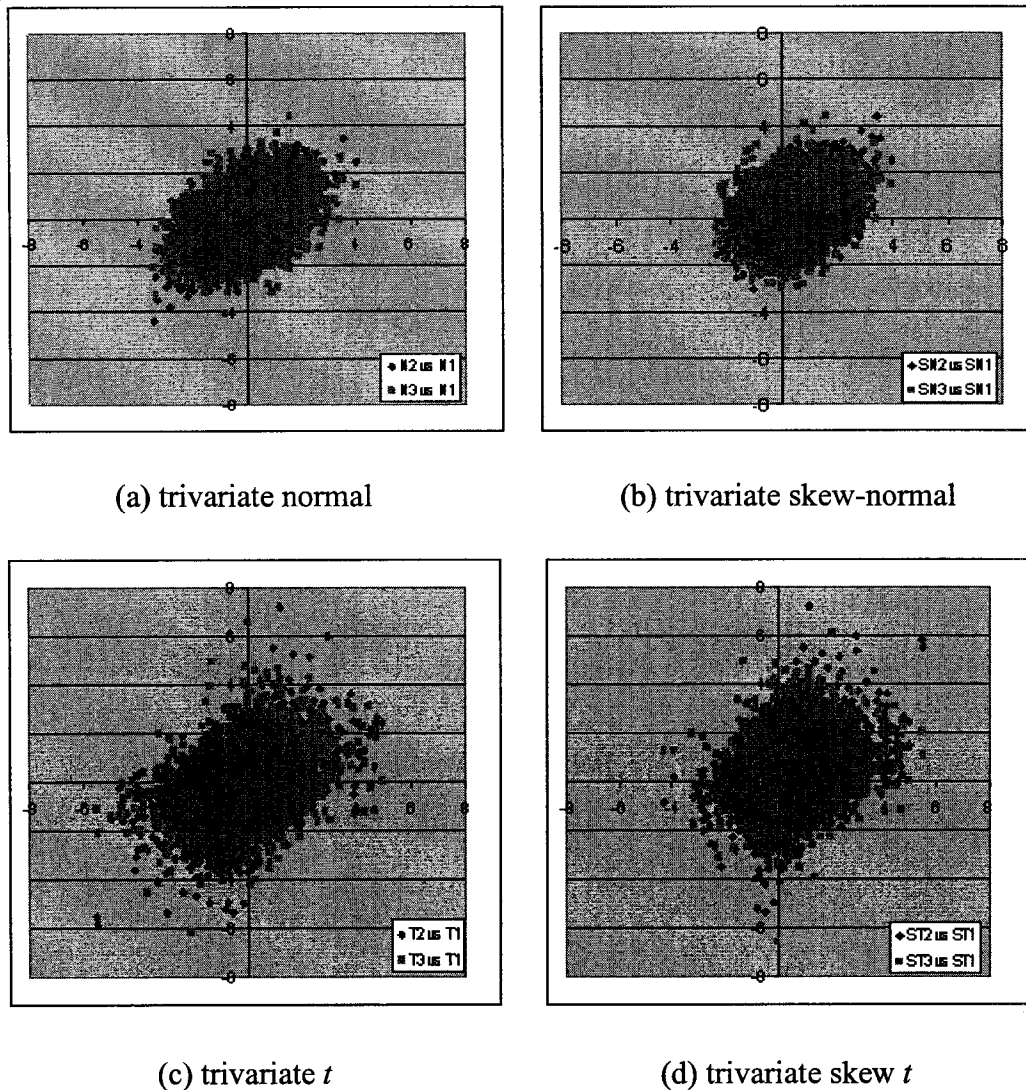
Figure 5.23 Four different trivariate distributions

5.5 Simulations of the Three-Factor Models
in VaR Transformation Procedure

The three risk factors are random variables to be simulated. Four different kinds

of multivariate distributions are considered for the three risk factors: normal, t, skew-

normal and skew t. These four scenarios are analyzed and the results are compared. For

each scenario, linear models and quadratic models developed in chapter 4 will be used in

the simulation. Values at Risk (VaRs) are calculated for each stock and the whole

portfolio at different confidence levels. In the simulation, the variance-covariance matrix

is calculated from 54 weekly historical data of the three risk factors in year 2005. The

idea is to use the most recent data to reflect the real market.

<u>5.5.1 Simulation of Three-Factor</u>
<u>Linear Models</u>

The fitted linear models for IBM, MSFT and TXN are, respectively,

$$\hat{S}_{IBM} = 77.98 + 9.386 R_{MRP} + 0.130 R_{MktRP} - 0.0016 R_{VvsGP}, \tag{5.86}$$

$$\hat{S}_{MSFT} = 23.76 + 2.632 R_{MRP} + 0.0508 R_{MktRP} - 0.00223 R_{VvsGP} \tag{5.87}$$

and

$$\hat{S}_{TXN} = 26.07 + 4.722 R_{MRP} + 0.1055 R_{MktRP} - 0.00913 R_{VvsGP} \tag{5.88}$$

<u>5.5.1.1 Multivariate normal distribution</u>

The one-year variance-covariance matrix of the three risk factors is

$$\Sigma = \begin{pmatrix} 0.3226 & -14.4215 & 10.117 \\ -14.4215 & 1193.473 & 3039.45 \\ 10.112 & 3039.45 & 27267.48 \end{pmatrix}.$$

The sampling size is $N = 5000$ for each risk factor. In this simulation, the three risk

factors are assumed to have a multivariate normal distribution with variance-covariance

matrix Σ. The simulation results are shown in the Figure 5.24.

The simulation results are shown in the Figure 5.24: (a), (b) and (c) are the

distributions of R_{MRP}, R_{MktRP}, and R_{VvsGP} generated from the variance-covariance matrix

Σ based on a multivariate normal distribution assumption. (d), (f), (g) are the distributions

of S_{IBM}, S_{MSFT}, and S_{TXN} calculated according to the linear models in Equations (5.86),

(5.87) and (5.88).

(a) R_{MRP}

(b) R_{MktRP}

(c) R_{VvsGP}

(d) S_{IBM}

(e) S_{MSFT}

(f) S_{TXN}
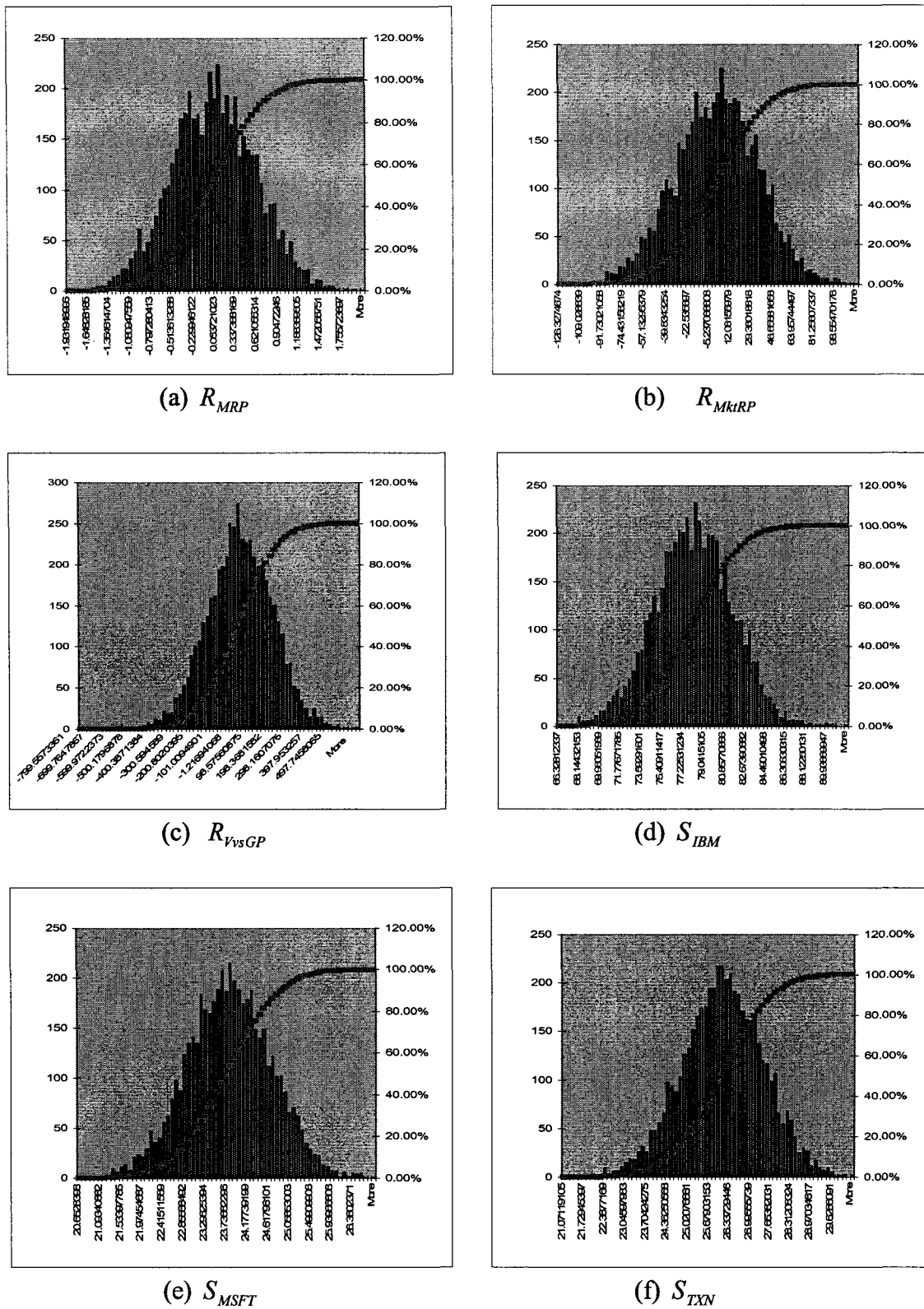
Figure 5.24 The distributions (from simulation) of the three risk factors and three stocks

Figure 5.24 (d), (f), and (g) show that S_{IBM}, S_{MSFT}, and S_{TXN} are all normally distributed. This is as expected since R_{MRP}, R_{MktRP}, and R_{VvsGP} are multivariate normal and the models apply linear transformations from the three risk factors to each of the three stocks.

Table 5.7 displays the input and estimated parameters for the risk factors $\vec{R} = (R_{MRP}, R_{MktRP}, R_{VvsGP})$. As a comparison, Table 5.8 lists the means of S_{IBM}, S_{MSFT}, and S_{TXN} calculated from simulation, denoted as $\hat{\mu}$, and the means of S_{IBM}, S_{MSFT}, and S_{TXN} calculated from 2005 one-year weekly historical data, denoted as μ.

Table 5.7   Risk factor information in MVN

	Input parameters	Estimated results from simulation
Variance-covariance matrix	$\Sigma = \begin{pmatrix} 0.3226 & -14.4215 & 10.117 \\ -14.4215 & 1193.473 & 3039.45 \\ 10.112 & 3039.45 & 27267.48 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.3205 & -14.1716 & 10.472 \\ -14.1716 & 1175,447 & 3004.544 \\ 10.472 & 3004.544 & 27021.58 \end{pmatrix}$
Mean	$\vec{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\hat{\vec{\mu}} = \begin{pmatrix} 0.0058 \\ -0.7938 \\ -4.8734 \end{pmatrix}$
Skewness $(\alpha = 0)$	$\eta_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\hat{\eta}_1 = \begin{pmatrix} -0.0231 \\ -0.0677 \\ -0.0435 \end{pmatrix}$
Kurtosis $(v = 0)$	$\eta_2 = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$	$\hat{\eta}_2 = \begin{pmatrix} 2.888 \\ 2.926 \\ 3.043 \end{pmatrix}$

Table 5.8 Means of three stocks from simulation ($\hat{\mu}$) and from historical data (μ)

	S_{IBM}	S_{MSFT}	S_{TXN}
$\hat{\mu}$ (N, linear)	77.98	23.74	26.06
μ (1 year)	83.26	25.76	28.9

Table 5.9 lists all the VaRs in terms of profit/loss (P/L) at different confidence levels for each stock individually and for the portfolio with three stocks.

Table 5.9 VaR P/L per share (multivariate normal distribution, linear model)

VaR (P/L) at $(1-\alpha)\%$		75%	90%	95%	97.5%	99%
$\hat{\mu}_N = \begin{pmatrix} 77.98 \\ 23.74 \\ 26.06 \end{pmatrix}$	ΔS_{IBM}	-2.34	-4.44	-5.7	-6.79	-8.18
	ΔS_{MSFT}	-0.60	-1.14	-1.47	-1.75	-2.10
	ΔS_{TXN}	-0.87	-1.66	-2.13	-2.53	-3.05
	$\Delta S_{Portfolio}$	-1.20 (-1.27)	-2.20 (-2.41)	-3.00 (-3.10)	-3.50 (-3.67)	-4.20 (-4.44)
$\mu_{1y} = \begin{pmatrix} 83.26 \\ 25.76 \\ 28.9 \end{pmatrix}$	ΔS_{IBM}	-7.62	-9.72	-10.98	-12.07	-13.46
	ΔS_{MSFT}	-2.62	-3.16	-3.49	-3.77	-4.12
	ΔS_{TXN}	-3.62	-4.5	-4.97	-5.37	-5.89
	$\Delta S_{Portfolio}$	-4.55 (-4.62)	-5.50 (-5.79)	-6.40 (-6.48)	-6.96 (-7.07)	-7.65 (-7.82)

For simplicity, we assumed that the weight of each stock in the portfolio is the same. So the portfolio value $P = nS_{IBM} + nS_{MSFT} + nS_{TXN}$. Here n is the number of shares for each stock. The price per share in the portfolio is calculated according to

$P = nS_{IBM} + nS_{MSFT} + nS_{TXN}$:

$$S_{Portfolio} = P/3n = \frac{1}{3}(S_{IBM} + S_{MSFT} + S_{TXN}) \ . \tag{5.89}$$

and

$$\Delta S_{Portfolio} = \frac{1}{3}(\Delta S_{IBM} + \Delta S_{MSFT} + \Delta S_{TXN}). \tag{5.90}$$

Table 5.9 compares VaRs of the portfolio and the VaRs of individual stock at certain confidence levels. The VaRs here are the profit or loss per share. The VaRs of each stock are calculated by constructing histograms based on Monte Carlo simulation results. The VaRs of the portfolio are calculated by constructing P/L histograms according to Equation (5.90). The calculations results shown in Table 5.9 are based on the sample mean $\hat{\mu}_N$ and the mean of one-year historical data, denoted as μ_{1y}. Diversification works as this: For example, at 90% confidence level and based on $\hat{\mu}_N$, S_{IBM} will lose no more than \$4.44 per share, S_{MSFT} will lose no more than \$1.14 per share, and S_{TXN} will lose no more than \$1.66 per share. Totally, if add them together and divide by 3, the portfolio maximum loss per share is supposed to be \$2.41 at 90% confidence level. From simulation, for a portfolio of the three stocks, the maximum loss per share is \$2.20 at 90% confidence level.

Figure 5.25 graphically describes in another way the relationships between the probabilities α and the VaRs of the four cases. It is seen from Figure 5.25 that there is 5% probability that the portfolio will lose more than \$3.00 per share.
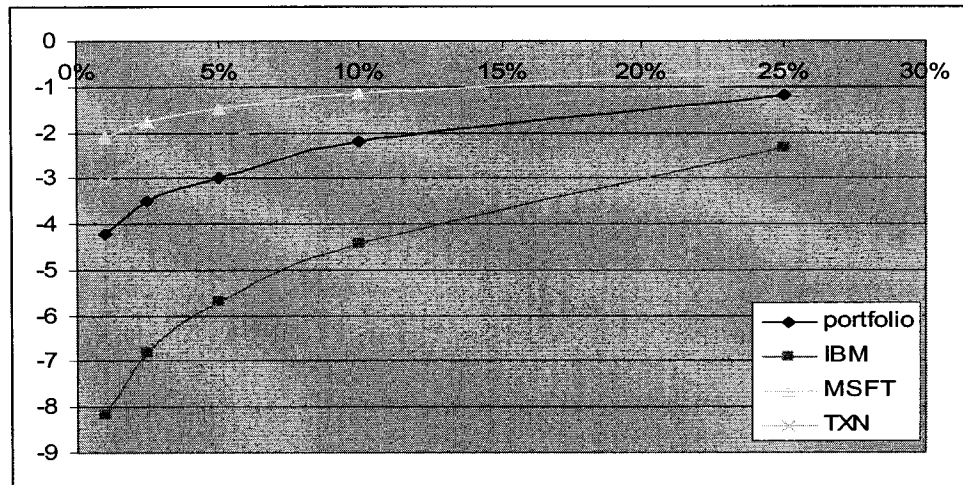
Figure 5.25 Probability α vs. P/L referred to $\hat{\mu}_N$ for multivariate normal, linear model

5.5.1.2 Multivariate t-distribution

From now on, the sample mean $\hat{\mu}_N$ will be used as a reference to compare VaRs

from different distributions. In this simulation, the three risk factors are assumed to be

multivariate t-distributions based on the multivariate normal distribution with variance-

covariance matrix Σ in Section 5.5.1.1. The new parameter added in this simulation is

the degree of freedom v. The distribution of the three risk factors changes from $N_3(\vec{0},\Sigma)$

to $t_3(\vec{0},\Sigma,v)$. The mean vector is zero since the centered risk factors are used in deriving

the three-factor model.

The simulation results are shown in Figure 5.26: (a), (b) and (c) are the

distributions of R_{MRP}, R_{MktRP}, and R_{VvsGP} generated from $t_3(\vec{0},\Sigma,5)$. (d), (f), (g) are the

distributions of S_{IBM}, S_{MSFT}, and S_{TXN} calculated according to the linear models in

Equations (5.86), (5.87) and (5.88).

(a) R_{MRP}

(b) R_{MktRP}

(c) R_{VvsGP}

(d) S_{IBM}

(e) S_{MSFT}

(f) S_{TXN}
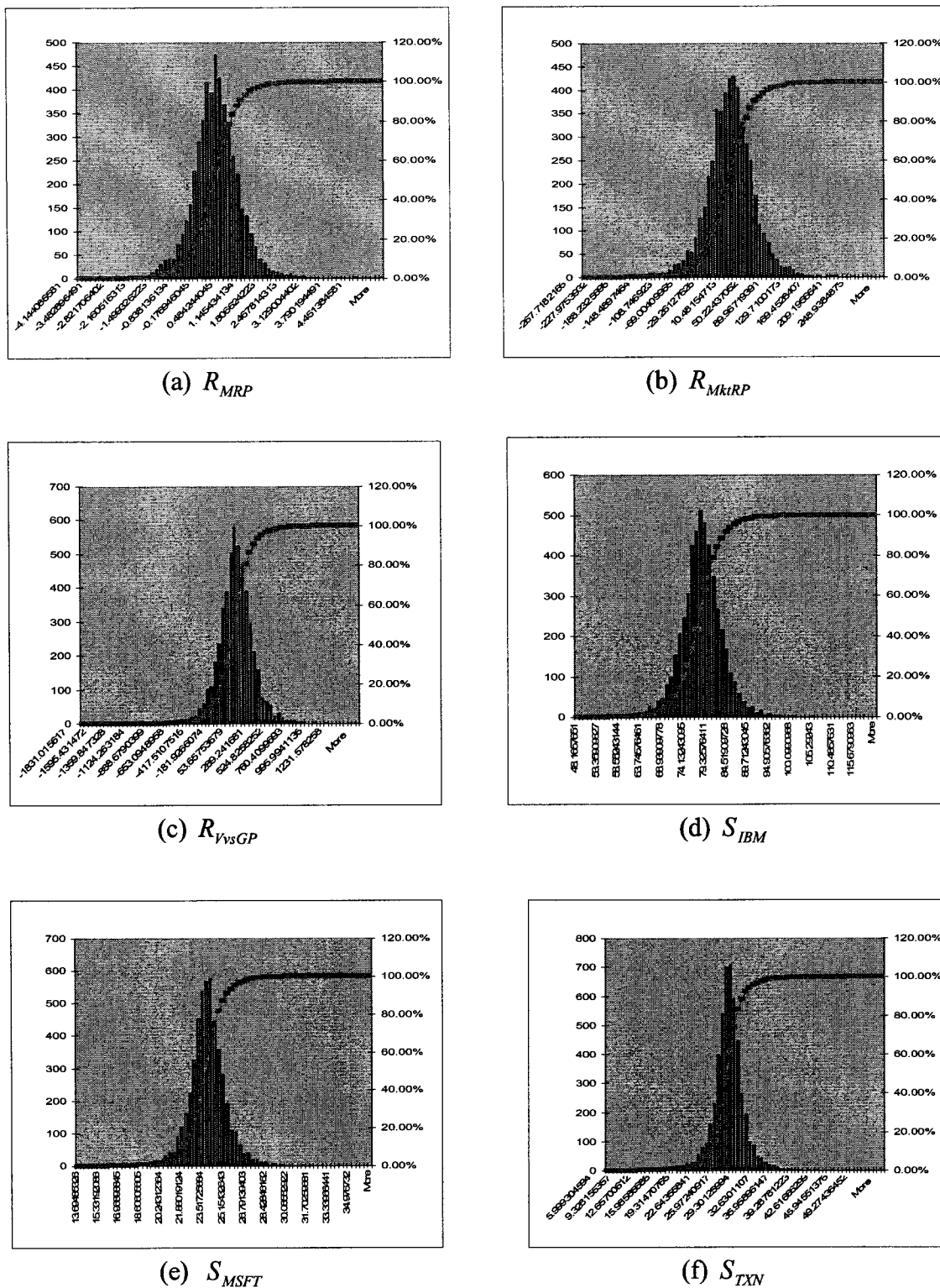
Figure 5.26 The distributions of the three risk factors and three stocks when $\vec{R}$ is a multivariate t-distribution (MVT)

From the Figure 5.26 (d), (f), and (g) we can see that S_{IBM}, S_{MSFT}, and S_{TXN} still follow t-distributions. This is because R_{MRP}, R_{MktRP}, and R_{VvsGP} are a multivariate t-distribution and the transformation is linear from $\vec{R} = (R_{MRP}, R_{MktRP}, R_{VvsGP})$ to S_{IBM}, S_{MSFT} and S_{TXN}, respectively.

Table 5.10 displays the input and estimated parameters for the risk factors $\vec{R} = (R_{MRP}, R_{MktRP}, R_{VvsGP})$. As a comparison, Table 5.11 lists the means of S_{IBM}, S_{MSFT}, and S_{TXN} calculated from the t-distribution samples, denoted as $\vec{\hat{\mu}}_t$, and the means estimated in multivariate normal (MVN) case $\vec{\hat{\mu}}_N$.

Table 5.10 Risk factor information in multivariate t-distribution

	Input parameters	Estimated results from samples
Variance-covariance matrix	$\Sigma = \begin{pmatrix} 0.3226 & -14.4215 & 10.117 \\ -14.4215 & 1193.473 & 3039.45 \\ 10.112 & 3039.45 & 27267.48 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.3205 & -14.1716 & 10.472 \\ -14.1716 & 1175,447 & 3004.544 \\ 10.472 & 3004.544 & 27021.58 \end{pmatrix}$
Mean	$\vec{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\hat{\vec{\mu}} = \begin{pmatrix} 0.0049 \\ -0.8883 \\ -5.1755 \end{pmatrix}$
Skewness $(\alpha = 0)$	$\eta_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\hat{\eta}_1 = \begin{pmatrix} 0.0790 \\ -0.1396 \\ -0.0358 \end{pmatrix}$
Kurtosis $(v = 5)$	$\eta_2 = \begin{pmatrix} 6 \\ 6 \\ 6 \end{pmatrix}$	$\hat{\eta}_2 = \begin{pmatrix} 5.593 \\ 6.140 \\ 6.845 \end{pmatrix}$

Table 5.11 Means of three stocks in two cases

	S_{IBM}	S_{MSFT}	S_{TXN}
$\bar{\hat{\mu}}_t$ (t, linear)	77.92	23.74	26.04
$\bar{\hat{\mu}}_N$ (N, linear)	77.98	23.74	26.06

Table 5.12 compares VaRs of the portfolio and the VaRs of individual stock at certain confidence levels. The VaRs here are the profit or loss per share. The VaRs of each stock are calculated by constructing histograms based on Monte Carlo simulation results on MVT of the risk factors and the linear factor models. The VaRs of the portfolio are calculated by constructing P/L histograms according to Equation (5.90). The calculations results shown in Table 5.12 are based on the sample mean $\hat{\mu}_t$, and $\hat{\mu}_N$ of MVN. Diversification effects are shown by comparing the two numbers listed for $\Delta S_{Portfolio}$: the number in the parenthesis is the sum of ΔS_{IBM}, ΔS_{MSFT}, and ΔS_{TXN}. The number without the parenthesis is the value of $\Delta S_{Portfolio}$ by constructing P/L histogram using Monte Carlo simulation results of S_{IBM}, S_{MSFT} and S_{TXN}.

Figure 5.27 graphically describes in another way the relationships between probabilities α and the VaRs of the four cases. For example, there is 5% possibility that the portfolio will lose more than \$4.70 per share. Compare with Figure 5.25, the possibility to lose \$4.70 per share is almost zero. If the real data has t distribution, the analysis based on normal distribution assumption could underestimate the risk (real loss).

Table 5.12 VaR P/L per share (multivariate t-distribution, linear model)

| VaR (P/L) at $(1-\alpha)\%$		75%	90%	95%	97.5%	99%
$\hat{\mu}_t = \begin{pmatrix} 77.92 \\ 23.74 \\ 26.04 \end{pmatrix}$	ΔS_{IBM}	-3.60	-6.83	-8.76	-10.44	-12.57
	ΔS_{MSFT}	-1.07	-2.03	-2.60	-3.10	-3.73
	ΔS_{TXN}	-1.93	-3.67	-4.71	-5.61	-6.76
	$\Delta S_{Portfolio}$	-1.50 (-2.20)	-3.27 (-4.17)	-4.67 (-5.36)	-6.17 (6.38)	-9.09 (-7.69)
$\hat{\mu}_N = \begin{pmatrix} 77.98 \\ 23.74 \\ 26.06 \end{pmatrix}$	ΔS_{IBM}	-3.66	-6.89	-8.82	-10.50	-12.63
	ΔS_{MSFT}	-1.07	-2.03	-2.60	-3.10	-3.73
	ΔS_{TXN}	-1.95	-3.69	-4.73	-5.63	-6.78
	$\Delta S_{Portfolio}$	-1.60 (-2.23)	-3.50 (-4.20)	-4.70 (-5.38)	-6.30 (-6.41)	-9.11 (-7.71)
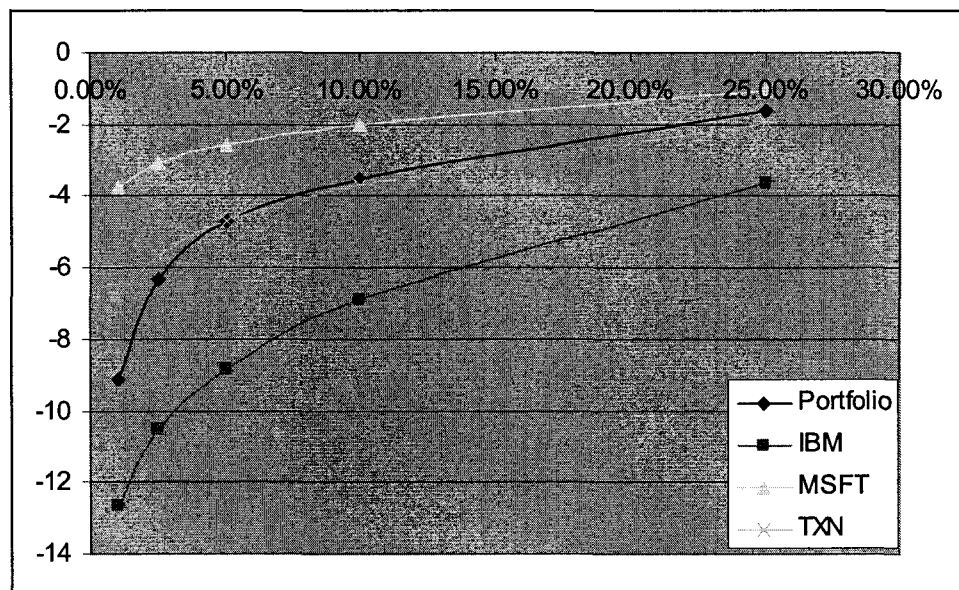
Figure 5.27 Probability α vs. P/L referred to $\hat{\mu}_N$ for a multivariate t, linear model

5.5.1.3 Multivariate skew-normal distribution

In this simulation, the three risk factors are assumed to have a multivariate skew-normal (MVSN) distribution based on the multivariate normal distribution with variance-covariance matrix Σ in Section 5.5.1.1. The new parameter added in this simulation is the shape parameter $\vec{\alpha}$. The distribution of the three risk factors changes from $N_3(\vec{0}, \Sigma)$ to $t_3(\vec{0}, \Sigma, \vec{\alpha})$, where $\vec{\alpha} = (2,2,2)'$. As discussed before, the larger the shape parameter, the more skewed the distribution becomes. The mean vector is zero since the centered risk factors are used in deriving the three-factor model.

The simulation results are shown in Figure 5.28: (a), (b) and (c) are the distributions of R_{MRP}, R_{MktRP}, and R_{VvsGP} generated from $t_3(\vec{0}, \Sigma, \vec{\alpha})$. Here $\vec{\alpha} = (2,2,2)'$. (d), (f), (g) are the distributions of S_{IBM}, S_{MSFT}, and S_{TXN} calculated according to the linear models in Equations (5.30), (5.31) and (5.32). From Figure 5.28 (d), (f), and (g) we can see that S_{IBM}, S_{MSFT}, and S_{TXN} are still skew-normal distributions, which further confirms that linear transformation doesn't change the type of the distribution.

Table 5.13 displays the input and estimated parameters for the risk factors $\vec{R} = (R_{MRP}, R_{MktRP}, R_{VvsGP})$. As a comparison, Table 5.14 lists the means of S_{IBM}, S_{MSFT}, and S_{TXN} calculated from the skew-normal distribution samples, denoted as $\hat{\mu}_{sn}$, and the means estimated in MVN case $\hat{\mu}_N$.

Figure 5.28 The distributions of the three risk factors and three stocks when $\vec{R}$ is multivariate t-distribution (MVT)

Table 5.13   Risk factor information in multivariate skew-normal distribution (MVSN)

	Input parameters	Estimated results from samples
Variance-covariance matrix	$\Sigma = \begin{pmatrix} 0.3226 & -14.4215 & 10.117 \\ -14.4215 & 1193.473 & 3039.45 \\ 10.112 & 3039.45 & 27267.48 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.3205 & -14.1716 & 10.472 \\ -14.1716 & 1175,447 & 3004.544 \\ 10.472 & 3004.544 & 27021.58 \end{pmatrix}$
Mean	$\vec{\mu} = \begin{pmatrix} 0.4053 \\ 24.6537 \\ 117.8404 \end{pmatrix}$	$\hat{\vec{\mu}} = \begin{pmatrix} 0.4025 \\ 23.9712 \\ 114.1011 \end{pmatrix}$
Skewness $\vec{\alpha} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$	$\eta_1 = \begin{pmatrix} 0.453741 \\ 0.453741 \\ 0.453741 \end{pmatrix}$	$\hat{\eta}_1 = \begin{pmatrix} 0.508583 \\ 0.464333 \\ 0.462669 \end{pmatrix}$
Kurtosis	$\eta_2 = \begin{pmatrix} 3.304975 \\ 3.304975 \\ 3.304975 \end{pmatrix}$	$\hat{\eta}_2 = \begin{pmatrix} 3.362526 \\ 3.322886 \\ 3.313236 \end{pmatrix}$

Table 5.14 Means of three stocks in two cases

	S_{IBM}	S_{MSFT}	S_{TXN}
$\hat{\mu}_{sn}$ (sn, linear)	84.68	25.78	29.46
$\hat{\mu}_N$ (N, linear)	77.98	23.74	26.06

Table 5.15 compares VaRs of the portfolio and the VaRs of individual stock at certain confidence levels. The VaRs of each stock are calculated by constructing histograms based on Monte Carlo simulation results on MVSN of the risk factors and the linear factor models. The VaRs of the portfolio are calculated by constructing P/L histograms according to Equation (5.34). The calculations results shown in Table 5.15 are based on the sample mean $\hat{\mu}_{sn}$, and $\hat{\mu}_N$ of MVN. Diversification effects are shown

by comparing the two numbers listed for $\Delta S_{Portfolio}$: the number in the parenthesis is the

sum of ΔS_{IBM} , ΔS_{MSFT} , and ΔS_{TXN} . The number without the parenthesis is the value of

$\Delta S_{Portfolio}$ by constructing P/L histogram using Monte Carlo simulation results of S_{IBM} ,

S_{MSFT} and S_{TXN} .

Table 5.15 VaR P/L per share (MVSN distribution, linear model)

VaR (P/L) at $(1-\alpha)\%$		75%	90%	95%	97.5%	99%
$\hat{\mu}_{sn} = \begin{pmatrix} 84.68 \\ 25.78 \\ 29.46 \end{pmatrix}$	ΔS_{IBM}	-3.61	-6.86	-8.80	-10.48	-12.62
	ΔS_{MSFT}	-1.08	-2.05	-2.63	-3.13	-3.77
	ΔS_{TXN}	-1.79	-3.40	-4.36	-5.19	-6.25
	$\Delta S_{Portfolio}$	-2.40 (-2.16)	-3.5 (-4.10)	-4.0 (-5.26)	-4.4 (-6.27)	-4.87 (-7.55)
$\hat{\mu}_{N} = \begin{pmatrix} 77.98 \\ 23.74 \\ 26.06 \end{pmatrix}$	ΔS_{IBM}	3.09	-0.16	-2.10	-3.78	-5.92
	ΔS_{MSFT}	0.96	-0.01	-0.59	-1.09	-1.73
	ΔS_{TXN}	1.61	0.00	-0.96	-1.79	-2.85
	$\Delta S_{Portfolio}$	1.70 (1.89)	0.70 (-0.06)	0.00 (-1.22)	-0.35 (-2.22)	-0.80 (-3.50)

Figure 5.29 graphically describes the relationships between probabilities α and

the VaRs of the four cases in another way. For example, in Figure 5.29 there is 5%

possibility that the portfolio will lose more than $4.70 per share. Here for the same 5%

possibility the portfolio will lose $0.0 per share. Different distributions give totally

different VaR information.
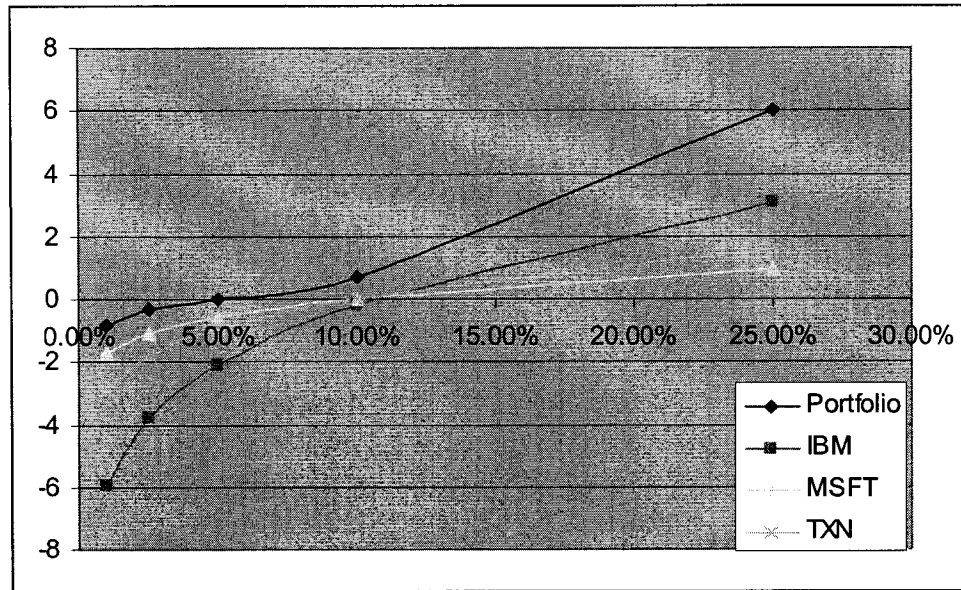
Figure 5.29 Probability α vs. P/L referred to $\hat{\mu}_N$ for multivariate SN, linear model
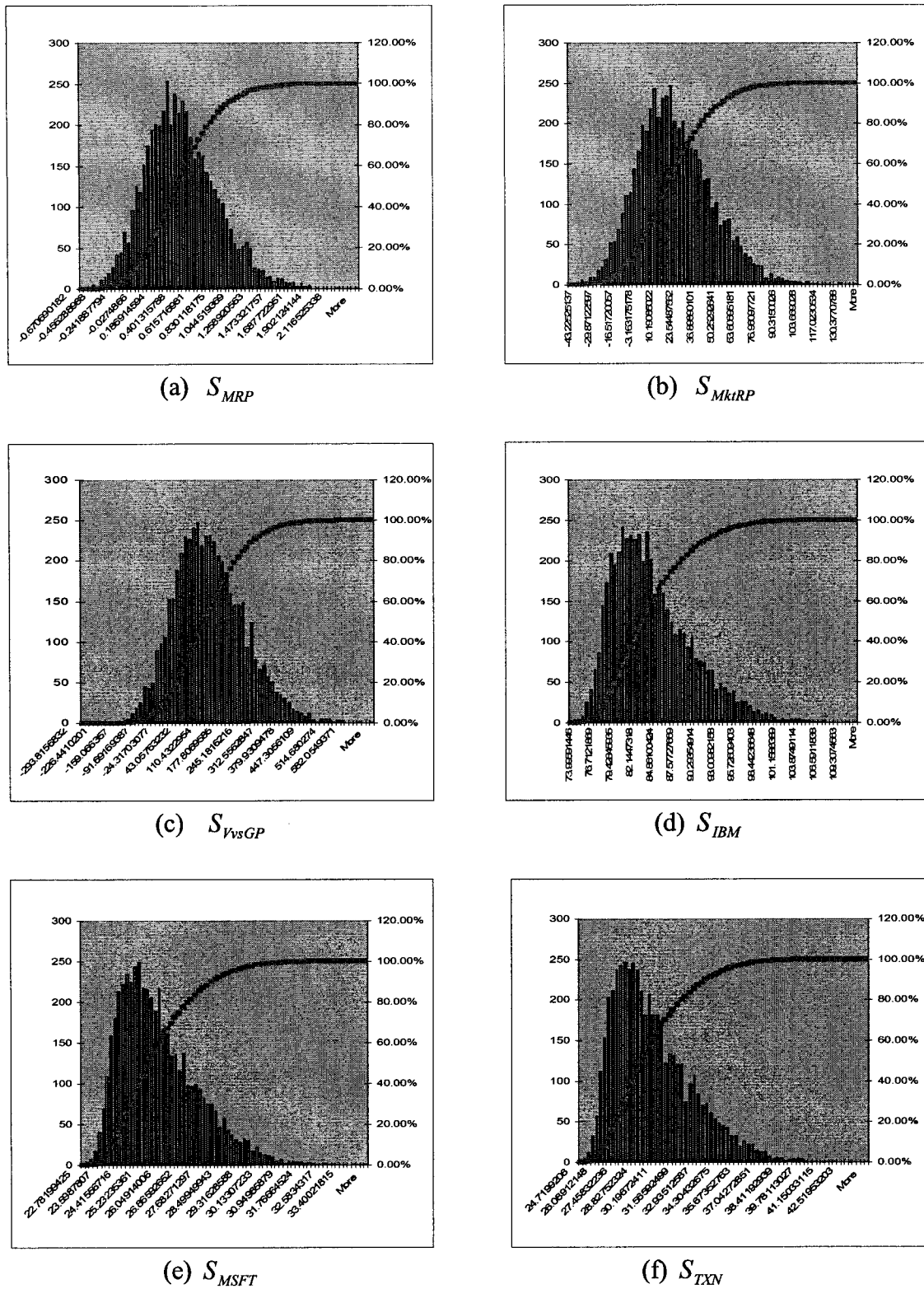
5.5.1.4 Multivariate skew t-distribution

In this simulation, the three risk factors are assumed to be multivariate skew-t (MVST) distribution based on the multivariate normal distribution with variance-covariance matrix Σ in Section 5.5.1.1. The new parameter added in this simulation is the shape parameter $\vec{\alpha}$ and the degrees of freedom v. The distribution of the three risk factors changes from $N_3(\vec{0},\Sigma)$ to $ST(\vec{0},\Sigma,\vec{\alpha},v)$, where $\vec{\alpha} = (2,2,2)'$ and $v = 2$. As discussed before, the larger the shape parameter, the more skewed the distribution becomes. The smaller the degrees of freedom, the higher peaked and heavier tailed the distribution is. The mean vector is zero since the centered risk factors are used in deriving the three-factor model.

The simulation results are shown in Figure 5.30: (a), (b) and (c) are the distributions of R_{MRP}, R_{MktRP}, and R_{VvsGP} generated from $SN_3(\vec{0},\Sigma,\vec{\alpha},v)$. Here

$\vec{\alpha} = (2,2,2)'$ and $v = 2$. (d), (f), (g) are the distributions of S_{IBM}, S_{MSFT}, and S_{TXN} calculated according to the linear models in Equations (5.30), (5.31) and (5.32). From the Figure 5.30 (d), (f), and (g) we can see that S_{IBM}, S_{MSFT}, and S_{TXN} are still skew t-distributions since linear transformation doesn't change the type of the distribution.

Table 5.16 displays the input and estimated parameters for the risk factors $\vec{R} = (R_{MRP}, R_{MktRP}, R_{VvsGP})$. As a comparison, Table 5.17 lists the means of S_{IBM}, S_{MSFT}, and S_{TXN} calculated from the skew t-distribution samples, denoted as $\hat{\mu}_{st}$, and the means estimated in MVN case $\hat{\mu}_N$.

Table 5.18 compares VaRs of the portfolio and the VaRs of individual stock at certain confidence levels. The VaRs of each stock are calculated by constructing histograms based on Monte Carlo simulation results on MVST of the risk factors and the linear factor models. The VaRs of the portfolio are calculated by constructing P/L histograms according to Equation (5.34). The calculations results shown in Table 5.18 are based on the sample mean $\hat{\mu}_{st}$, and $\hat{\mu}_N$ of MVN. Diversification effects are shown by comparing the two numbers listed for $\Delta S_{Portfolio}$: the number in the parenthesis is the sum of ΔS_{IBM}, ΔS_{MSFT}, and ΔS_{TXN}. The number without the parenthesis is the value of $\Delta S_{Portfolio}$ by constructing P/L histogram using Monte Carlo simulation results of S_{IBM}, S_{MSFT} and S_{TXN}.

(a) R_{MRP}

(b) R_{MktRP}

(c) R_{VvsGP}

(d) S_{IBM}

(e) S_{MSFT}

(f) S_{TXN}
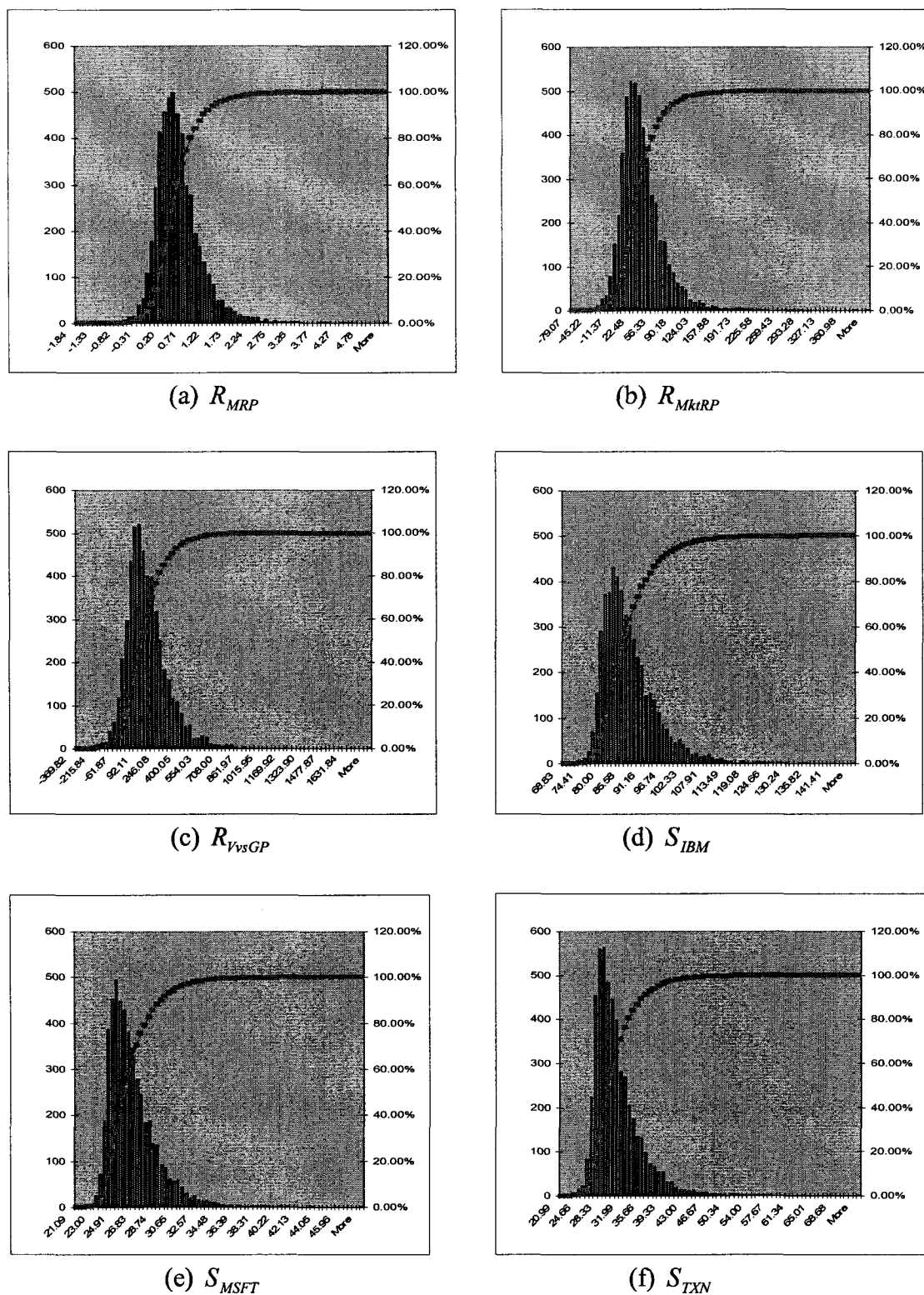
Figure 5.30 The distributions of the three risk factors and three stocks when $\vec{R}$ is multivariate skew t-distribution (MVST)

Table 5.16 Risk factors information for MVST

	input parameters	Estimated results from samples
Variance-covariance matrix	$\Sigma = \begin{pmatrix} 0.3226 & -14.4215 & 10.117 \\ -14.4215 & 1193.473 & 3039.45 \\ 10.112 & 3039.45 & 27267.48 \end{pmatrix}$	$\hat{\Sigma} = \begin{pmatrix} 0.3205 & -14.1716 & 10.472 \\ -14.1716 & 1175,447 & 3004.544 \\ 10.472 & 3004.544 & 27021.58 \end{pmatrix}$
mean	$\bar{\mu}_{\bar{R}} = \begin{pmatrix} 0.482137 \\ 29.32324 \\ 140.1613 \end{pmatrix}$	$\hat{\bar{\mu}}_{\bar{R}} = \begin{pmatrix} 0.472764 \\ 29.2074 \\ 134.1451 \end{pmatrix}$
Skewness $\alpha = 2$	$\eta_1 = \begin{pmatrix} 1.790046 \\ 1.790046 \\ 1.790046 \end{pmatrix}$	$\hat{\eta}_1 = \begin{pmatrix} 1.668538 \\ 1.783777 \\ 1.273095 \end{pmatrix}$
Kurtosis $v = 5$	$\eta_2 = \begin{pmatrix} 13.52727 \\ 13.52727 \\ 13.52727 \end{pmatrix}$	$\hat{\eta}_2 = \begin{pmatrix} 12.71914 \\ 14.59714 \\ 10.77981 \end{pmatrix}$

Table 5.17 Means of three stocks in two cases

	S_{IBM}	S_{MSFT}	S_{TXN}
$\hat{\bar{\mu}}_{st}$ (st, linear)	85.99	26.19	30.16
$\hat{\bar{\mu}}_{N}$ (N, linear)	77.98	23.74	26.06

Table 5.18 VaR P/L per share (MVST distribution, linear model)

VaR (P/L) at $(1-\alpha)\%$		75%	90%	95%	97.5%	99%
$\hat{\mu}_{st} = \begin{pmatrix} 85.99 \\ 26.19 \\ 30.16 \end{pmatrix}$	ΔS_{IBM}	-4.94	-9.38	-12.03	-14.34	-17.27
	ΔS_{MSFT}	-1.53	-2.90	-3.73	-4.44	-5.35
	ΔS_{TXN}	-2.72	-5.17	-6.63	-7.90	-9.51
	$\Delta S_{Portfolio}$	-3.00 (-3.06)	-4.27 (-5.82)	-5.0 (-7.46)	-5.5 (-8.89)	-6.00 (-10.71)
$\hat{\mu}_{N} = \begin{pmatrix} 77.98 \\ 23.74 \\ 26.06 \end{pmatrix}$	ΔS_{IBM}	3.07	-1.37	-4.02	-6.33	-9.26
	ΔS_{MSFT}	0.92	-0.45	-1.28	-1.99	-2.90
	ΔS_{TXN}	1.39	-1.07	-2.53	-3.80	-5.41
	$\Delta S_{Portfolio}$	1.80 (1.79)	0.70 (-0.96)	0.00 (-2.61)	-0.70 (-4.04)	-1.40 (-5.86)

Figure 5.31 graphically describes the relationships between probabilities α and the VaRs of the four cases. For MVSN, the portfolio diversification effect is very obvious. The VaRs of $\Delta S_{Portfolio}$ are much smaller than those of any individual stocks.

Figure 5.32 compares the VaRs of the portfolio in four different distributions: MVN, MVT, MVSN and MVST. Since every elements of the shape parameter $\vec{\alpha}$ is 2 > 0, the MVSN and MVST are both positively skewed, the left-side tail in these two cases is short, the VaRs are smaller compared with MVN case. MVT has fat and long tail, the VaRs of MVT is much larger than MVN, just as it should be. If the shape parameter becomes negative, it is reasonable to infer that the VaRs for negative MVSN or MVST will become larger.
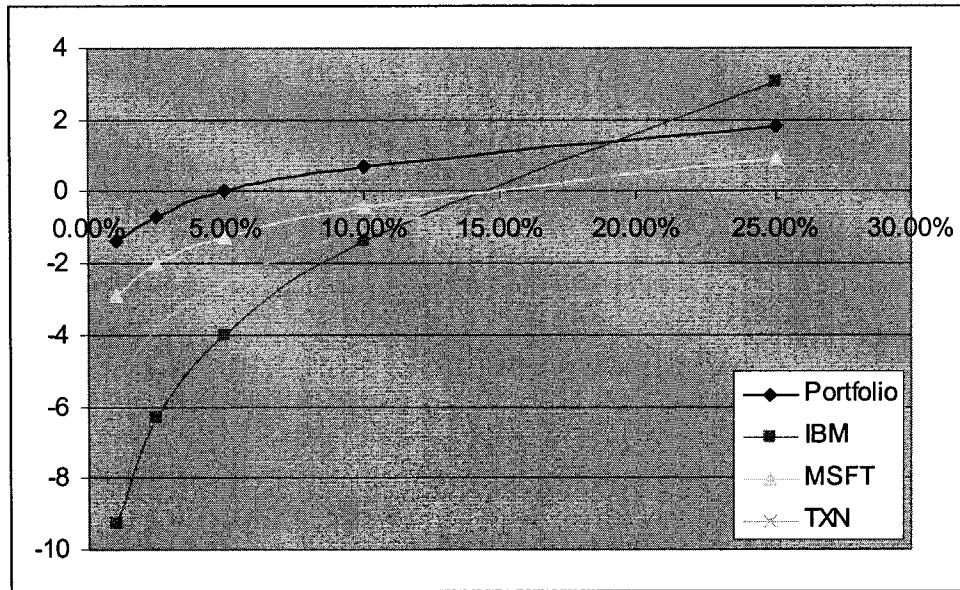
Figure 5.31 Probability α vs. P/L referred to $\hat{\mu}_N$ for MVST, linear model
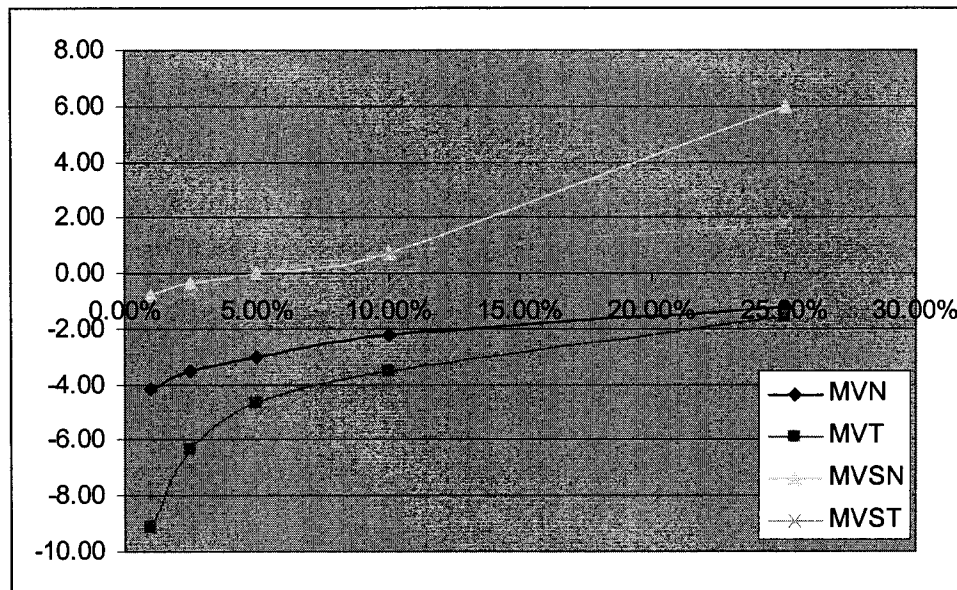
Figure 5.32 Probability α vs. P/L for $\Delta S_{Portfolio}$ comparison of four distributions for linear model

5.5.2 Simulation on Three-Factor
     Quadratic Models

The simulation techniques are the same as in Section 5.5.1, and the method of

analysis is also the same as in analyzing the linear models. In this section, the simulation

results will be displayed and only short descriptions will be given.

The fitted quadratic model for IBM is

$$\hat{S}_{IBM} = 76.116 + 10.868R_{MRP} + 0.1288R_{MktRP} + 0.0000006074R^2{}_{MktRP} - 0.00268R_{MRP}R_{VvsGP} \ .$$

$$(5.34)$$

The fitted quadratic model for MSFT stock is

$$\hat{S}_{MSFT} = 27.523 + 4.997R_{MRP} + 0.067R_{MktRP} - 0.00531R_{VvsGP} - 1.151R^2_{MRP}$$

$$- 0.00002R^2_{MktRP} - 0.00173R_{MRP}R_{VvsGP} \qquad (5.35)$$

The fitted quadratic model for TXN stock is

$$\hat{S}_{TXN} = 18.765 + 0.0325R_{MktRP} + 0.00397R_{VvsGP} + 0.00021R^2_{MktRP}$$

$$+ 0.00000378R^2_{VvsGP} - 0.004664R_{MktRP}R_{VvsGP} \qquad (5.36)$$

5.5.2.1 Multivariate normal distribution

The simulation results are shown in Figure 5.33: (a), (b) and (c) are the

distributions of R_{MRP}, R_{MktRP}, and R_{VvsGP} generated from $N_3(\vec{0}, \Sigma)$. (d), (f), (g) are the

distributions of S_{IBM}, S_{MSFT}, and S_{TXN} calculated according to the quadratic models in

Equations (5.34) - (5.36).

(a) R_{MRP}

(b) R_{MktRP}

(c) R_{VvsGP}

(d) S_{IBM}

(e) S_{MSFT}

(f) S_{TXN}
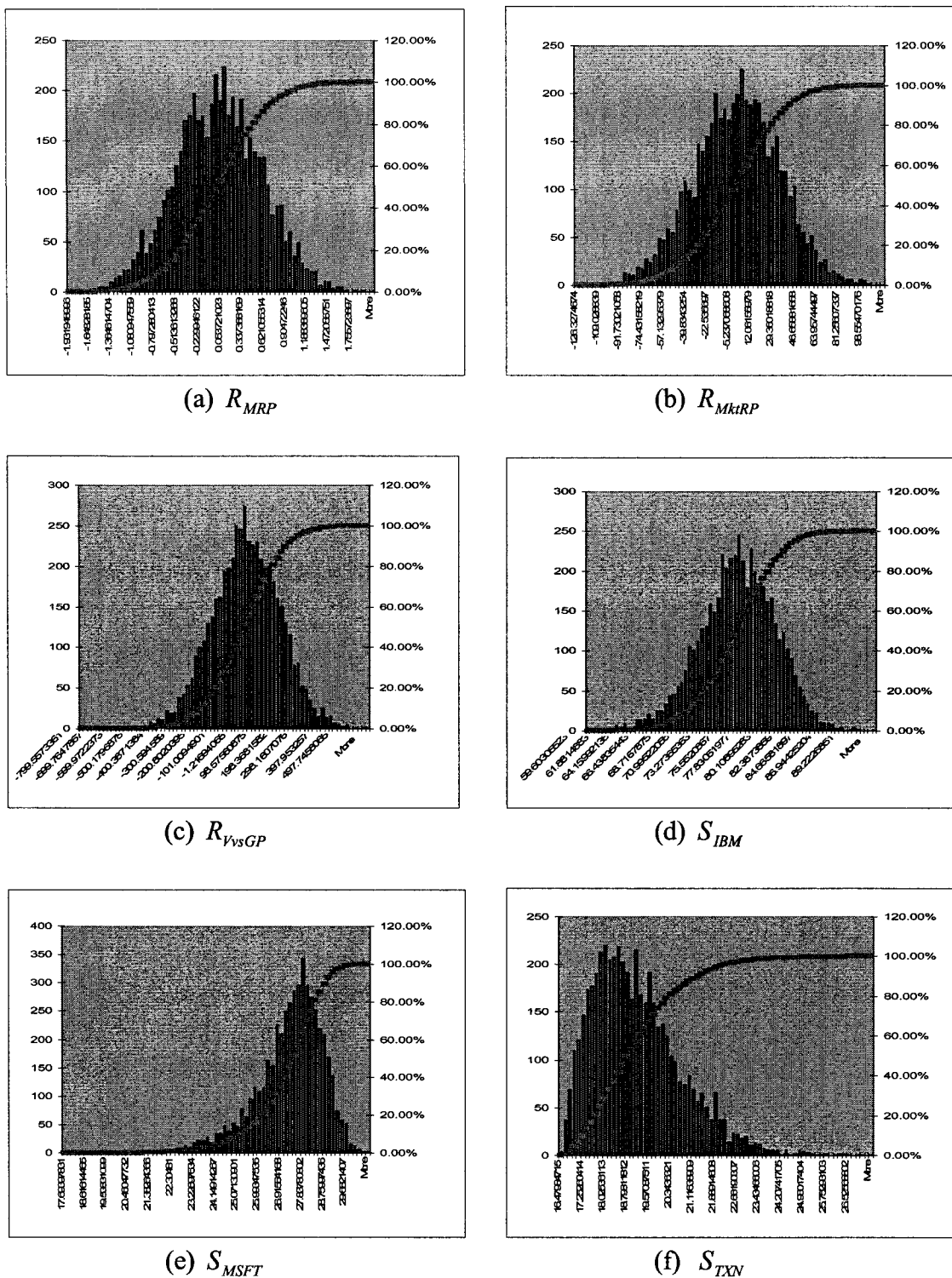
Figure 5.33 The distributions of the three risk factors and three stocks when $\vec{R}$ is multivariate normal distribution (MVN) in quadratic models

From Figure 5.33 we can see the generated distributions for S_{MSFT} in (e) and S_{TXN}

in (f) are not normal any more. In fact, they are skewed in opposite directions, even

though the distribution of R_{MRP}, R_{MktRP}, and R_{VvsGP} is MVN. Since the linear

transformation terms of the quadratic models don't change the dependent variable's

distribution. The change of distribution is caused by the squared terms and cross-product

terms. Table 5.19 summarizes the new finding.

Table 5.19 Analysis of the causes of skewness for S_{IBM}, S_{MSFT} and S_{TXN}

	$R^2{}_{MktRP}$	$R_{MRP}R_{VvsGP}$		Skewness effect is canceled out.
S_{IBM}	+	-		
	$R^2{}_{MRP}$	$R^2{}_{MktRP}$	$R_{MRP}R_{VvsGP}$	Negatively skewed
S_{MSFT}	-	-	-	
	$R^2{}_{VvsGP}$	$R^2{}_{MktRP}$	$R_{MktRP}R_{VvsGP}$	Positively skewed
S_{TXN}	+	+	-	

The risk factors information is the same as in MVN linear models. As a

comparison, Table 5.20 lists the means of S_{IBM}, S_{MSFT}, and S_{TXN} calculated from the

simulation of MVN quadratic model, denoted as $\hat{\mu}_{N^2}$, and the means estimated in MVN

linear model $\hat{\mu}_N$, The means of S_{IBM} and S_{MSFT} don't change much from the means in

linear models. However, the mean of S_{TXN} changes a lot to a smaller number.

Table 5.20 Means of S_{IBM}, S_{MSFT} and S_{TXN} in two cases

	S_{IBM}	S_{MSFT}	S_{TXN}
$\hat{\mu}_{N^2}$ (N, quadratic)	76.12	27.17	19.07
$\hat{\mu}_N$ (N, linear)	77.98	23.74	26.06

Table 5.21 compares VaRs of the portfolio and the VaRs of individual stock at certain confidence levels. The VaRs of each stock are calculated by constructing histograms based on Monte Carlo simulation results on MVN of the risk factors and the quadratic factor models. The VaRs of the portfolio are calculated by constructing P/L histograms according to Equation (5.34). The calculations results shown in the table 5.21 are based on the sample mean $\hat{\mu}_{st}$, and $\hat{\mu}_N$ of MVN diversification effects are shown by comparing the two numbers listed for $\Delta S_{Portfolio}$: the number in the parenthesis is the sum of ΔS_{IBM}, ΔS_{MSFT}, and ΔS_{TXN}. The number without the parenthesis is the value of $\Delta S_{Portfolio}$ by constructing P/L histogram using Monte Carlo simulation results of S_{IBM}, S_{MSFT} and S_{TXN}.

Figure 5.34 graphically describes the relationships between probabilities α and the VaRs of the four cases. The VaRs of $S_{Portfolio}$ is smaller than those of S_{IBM} and S_{TXN}, which shows the benefit of diversification.

Table 5.21 VaR P/L per share (MVN distribution, quadratic model)

VaR (P/L) at $(1-\alpha)\%$		75%	90%	95%	97.5%	99%
$\hat{\mu}_{N^2} = \begin{pmatrix} 76.12 \\ 27.17 \\ 19.07 \end{pmatrix}$	ΔS_{IBM}	-2.85	-5.42	-6.95	-8.28	-9.97
	ΔS_{MSFT}	-1.00	-1.90	-2.43	-2.90	-3.49
	ΔS_{TXN}	-1.07	-2.03	-2.60	-3.10	-3.73
	$\Delta S_{Portfolio}$	-1.29 (-1.64)	-2.62 (-3.12)	-3.51 (-3.99)	-4.33 (-4.76)	-5.29 (-5.73)
$\hat{\mu}_N = \begin{pmatrix} 77.98 \\ 23.74 \\ 26.06 \end{pmatrix}$	ΔS_{IBM}	-4.71	-7.28	-8.81	-10.14	-11.83
	ΔS_{MSFT}	2.43	1.53	1.00	0.53	-0.06
	ΔS_{TXN}	-8.06	-9.02	-9.59	-10.09	-10.72
	$\Delta S_{Portfolio}$	-3.10 (-3.45)	-4.43 (-4.93)	-5.32 (-5.80)	-6.20 (-6.57)	-7.10 (-7.54)
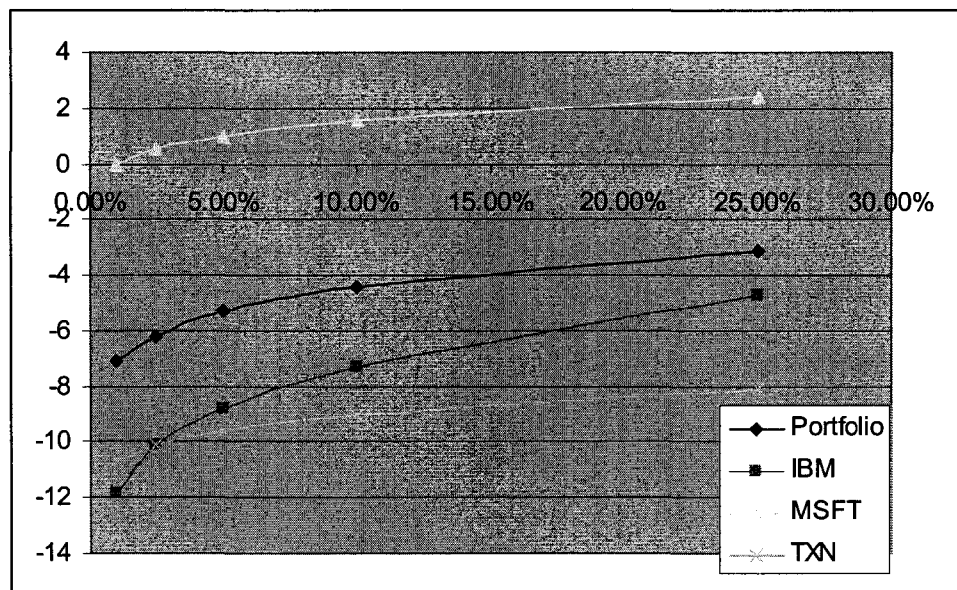
Figure 5.34 Probability α vs. P/L referred to $\hat{\mu}_N$ for MVN, quadratic model

5.5.2.2 Multivariate t-distribution

The simulation results are shown in Figure 5.35: (a), (b) and (c) are the distributions of R_{MRP}, R_{MktRP}, and R_{VvsGP} generated from $t_3(\bar{0}, \Sigma, v)$, where $v = 5$. (d), (f), (g) are the distributions of S_{IBM}, S_{MSFT}, and S_{TXN} calculated according to the quadratic models in Equations (5.34) - (5.36).

As we expect, the generated distributions for S_{MSFT} (in Figure 5.35(e)) and S_{TXN} (in Figure 5.35(f)) are skewed in opposite directions, even though the distribution of R_{MRP}, R_{MktRP}, and R_{VvsGP} is MVT. The reason is the same as analyzed in Section 5.5.2.1 MVN case.

The risk factors information is the same as in MVT linear models. As a comparison, Table 5.22 lists the means of S_{IBM}, S_{MSFT}, and S_{TXN} calculated from the simulation of MVT quadratic model, denoted as $\hat{\mu}_{t^2}$ and the means estimated in MVN linear model $\hat{\mu}_N$. The means of S_{IBM} doesn't change much from the means in MVN linear models. However, the mean of S_{MSFT} moves to a higher value and the mean of S_{TXN} changes to a smaller number, as noted in MVN quadratic model.

Table 5.22 Means of S_{IBM}, S_{MSFT}, and S_{TXN} in two cases

	S_{IBM}	S_{MSFT}	S_{TXN}
$\hat{\mu}_{t^2}$ (t, quadratic)	76.14	26.91	19.32
$\hat{\mu}_N$ (N, linear)	77.98	23.74	26.06
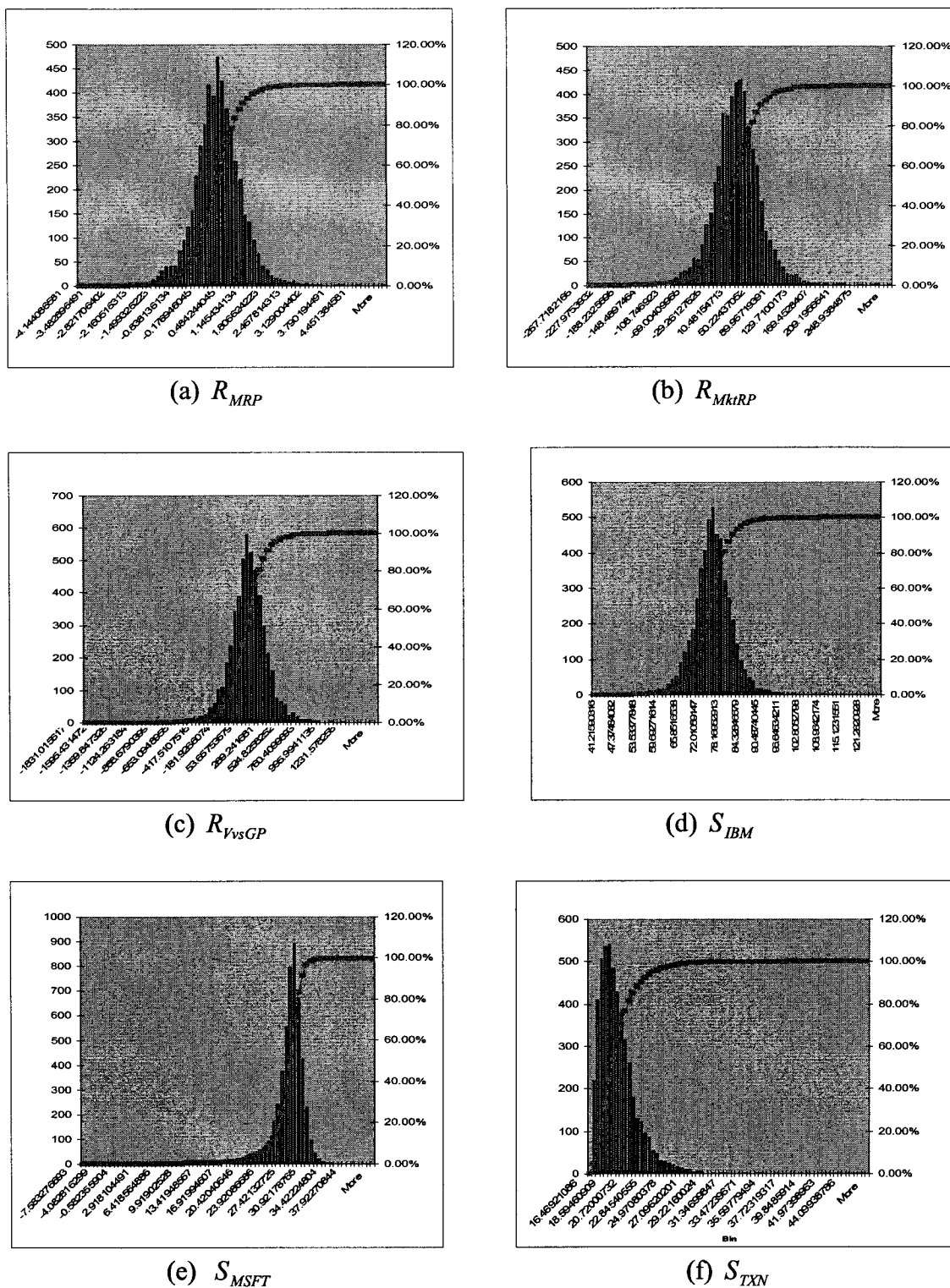
Figure 5.35 The distributions of the three risk factors and three stocks when $\vec{R}$ is multivariate t-distribution (MVT) in quadratic models

Table 5.23 compares VaRs of the portfolio and the VaRs of individual stock at certain confidence levels. The VaRs of each stock are calculated by constructing histograms based on Monte Carlo simulation results on MVT of the risk factors and the quadratic factor models. The VaRs of the portfolio are calculated by constructing P/L histograms according to Equation (5.34). The calculations results shown in Table 5.23 are based on the sample mean $\hat{\mu}_{t^2}$ and $\hat{\mu}_N$ of MVN in linear model. Diversification effects are shown by comparing the two numbers listed for $\Delta S_{Portfolio}$: the number in the parenthesis is the sum of ΔS_{IBM}, ΔS_{MSFT}, and ΔS_{TXN}. The number without the parenthesis is the value of $\Delta S_{Portfolio}$ by constructing P/L histogram using Monte Carlo simulation results of S_{IBM}, S_{MSFT} and S_{TXN}.

Table 5.23 VaR P/L per share (MVT distribution, quadratic model)

VaR (P/L) at $(1-\alpha)\%$		75%	90%	95%	97.5%	99%
$\hat{\mu}_{t^2} = \begin{pmatrix} 76.14 \\ 26.91 \\ 19.32 \end{pmatrix}$	ΔS_{IBM}	-4.17	-7.93	-10.17	-12.12	-14.60
	ΔS_{MSFT}	-1.84	-3.49	-4.48	-5.34	-6.43
	ΔS_{TXN}	-1.93	-3.67	-4.71	-5.61	-6.76
	$\Delta S_{Portfolio}$	-1.57 (-2.65)	-3.27 (-5.03)	-4.67 (-6.45)	-6.33 (-7.69)	-8.67 (-9.26)
$\hat{\mu}_N = \begin{pmatrix} 77.98 \\ 23.74 \\ 26.06 \end{pmatrix}$	ΔS_{IBM}	-6.01	-9.77	-12.01	-13.96	-16.44
	ΔS_{MSFT}	1.33	-0.32	-1.31	-2.17	-3.26
	ΔS_{TXN}	-8.67	-10.21	-11.45	-12.35	-13.50
	$\Delta S_{Portfolio}$	-3.41 (-4.45)	-5.46 (-6.77)	-6.83 (-8.26)	-8.21 (-9.48)	-10.30 (-11.07)

Figure 5.36 graphically describes the relationships between probabilities α and the VaRs of the four cases. The VaRs of $S_{Portfolio}$ is smaller than those of S_{IBM} and S_{TXN}, which shows the benefit of diversification.
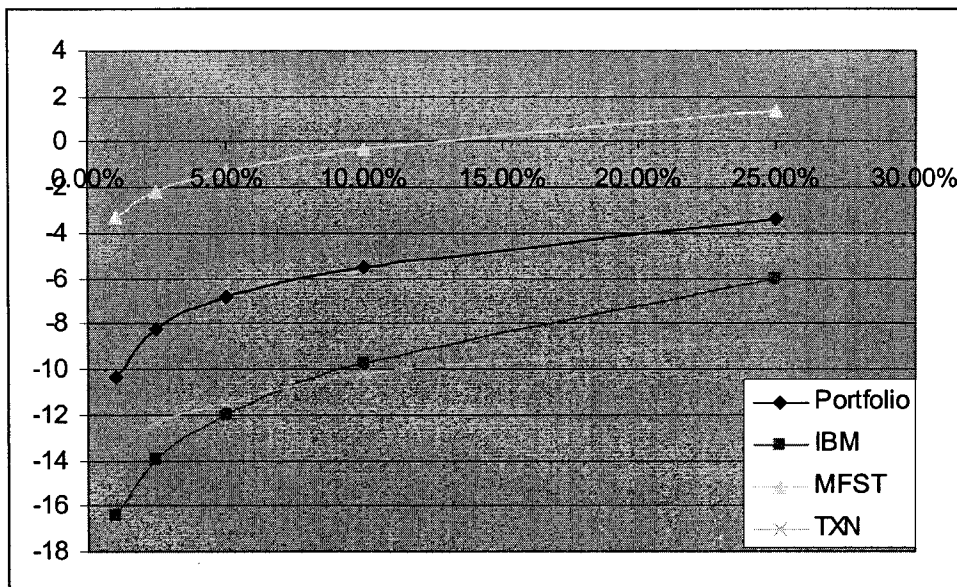
Figure 5.36 Probability α vs. P/L referred to $\hat{\mu}_N$ for MVT, quadratic model

5.5.2.3 Multivariate skew-normal distribution

The simulation results are shown in Figure 5.37: (a), (b) and (c) are the distributions of R_{MRP}, R_{MktRP}, and R_{VvsGP} generated from $SN_3(\vec{0}, \Sigma, \vec{\alpha})$, where $\vec{\alpha} = (2,2,2)'$. (d), (f), (g) are the distributions of S_{IBM}, S_{MSFT}, and S_{TXN} calculated according to the quadratic models in Equations (5.34) - (5.36).

(a) R_{MRP}

(b) R_{MktRP}

(c) R_{VvsGP}

(d) S_{IBM}

(e) S_{MSFT}

(f) S_{TXN}
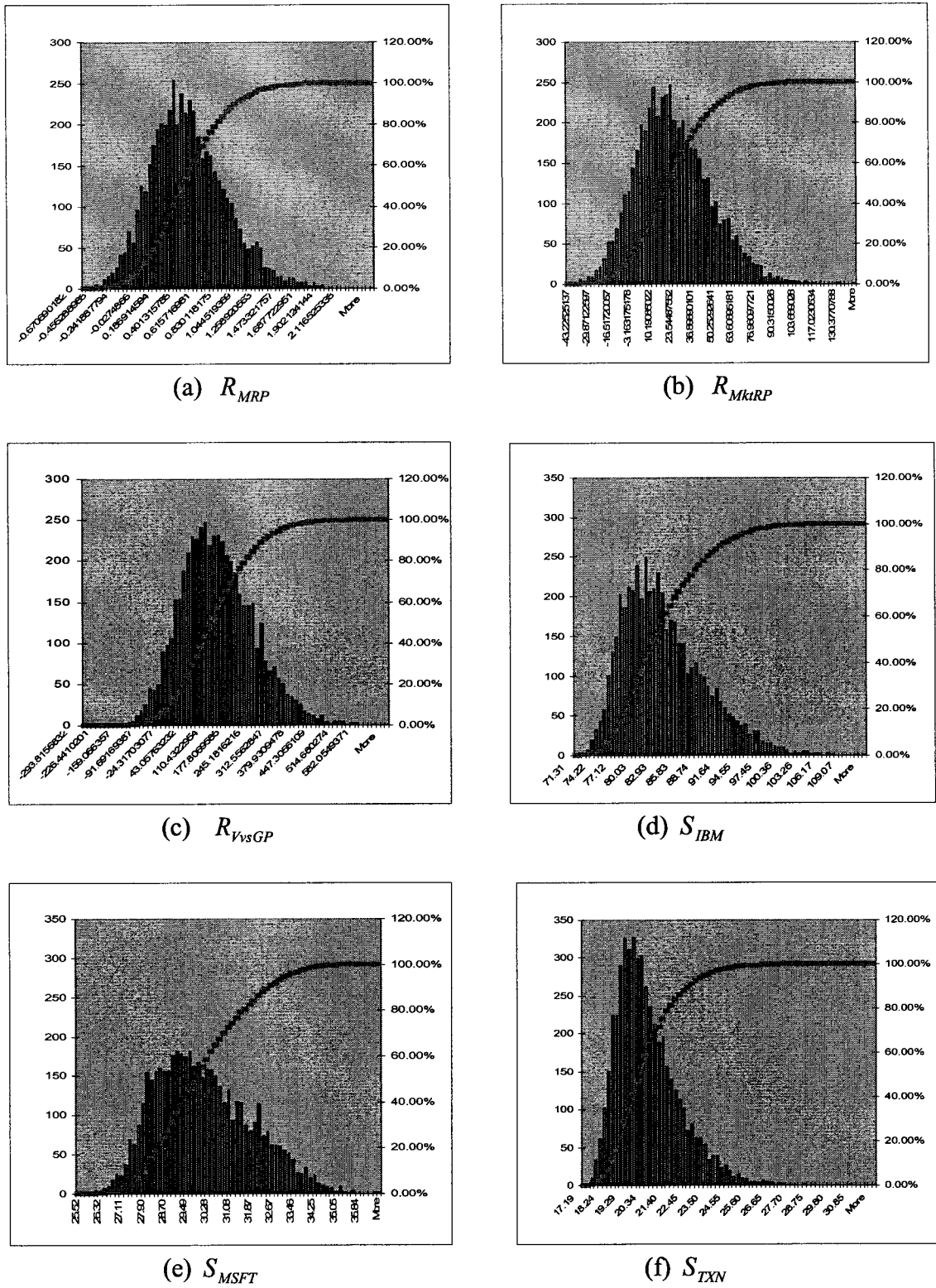
Figure 5.37 The distributions of the three risk factors and three stocks when $\vec{R}$ is multivariate skew-normal distribution (MVSN) in quadratic models

Comparing the simulation results of MVSN in linear models, it is interesting to observe in Figure 5.37 that the skewness of S_{IBM} doesn't change, the distribution of S_{MSFT} becomes less skewed, and the distribution of S_{TXN} becomes more skewed. These phenomena can also be explained by the signs of the squared-terms and cross-product terms in quadratic models.

The risk factors information is the same as in MVSN linear models. As a comparison, Table 5.24 lists the means of S_{IBM}, S_{MSFT}, and S_{TXN} calculated from the simulation of MVT quadratic model, denoted as $\hat{\mu}_{sn^2}$, and the means estimated in MVN linear model $\hat{\mu}_N$. The means of S_{IBM} and S_{MSFT} move up from the means in MVN linear models. The mean of S_{TXN} is still a smaller number than the mean in MVN linear model.

Table 5.24 Means of three stocks in two cases

	S_{IBM}	S_{MSFT}	S_{TXN}
$\hat{\mu}_{sn^2}$ (sn, quadratic)	83.45	30.07	20.34
$\hat{\mu}_N$ (N, linear)	77.98	23.74	26.06

Table 5.25 compares VaRs of the portfolio and the VaRs of individual stock at certain confidence levels. The VaRs of each stock are calculated by constructing histograms based on Monte Carlo simulation results on MVSN of the risk factors and the quadratic factor models. The VaRs of the portfolio are calculated by constructing P/L histograms according to Equation (5.34). The calculations results shown in Table 5.25 are based on the sample mean $\hat{\mu}_{sn^2}$, and $\hat{\mu}_N$ of MVN in linear model. Diversification effects are shown by comparing the two numbers listed for $\Delta S_{Portfolio}$: the number in the

parenthesis is the sum of ΔS_{IBM}, ΔS_{MSFT}, and ΔS_{TXN}. The number without the parenthesis

is the value of $\Delta S_{Portfolio}$ by constructing P/L histogram using Monte Carlo simulation

results of S_{IBM}, S_{MSFT} and S_{TXN}.

Table 5.25 VaR P/L per share (MVSN distribution, quadratic model)

VaR (P/L) at $(1-\alpha)\%$		75%	90%	95%	97.5%	99%
$\hat{\mu}_{sn^2} = \begin{pmatrix} 83.45 \\ 30.07 \\ 20.34 \end{pmatrix}$	ΔS_{IBM}	-3.90	-7.41	-9.52	-11.34	-13.65
	ΔS_{MSFT}	-1.20	-2.28	-2.92	-3.48	-4.20
	ΔS_{TXN}	-1.09	-2.07	-2.65	-3.16	-3.80
	$\Delta S_{Portfolio}$	-2.23 (-2.06)	-3.20 (-3.92)	-3.93 (-5.03)	-4.33 (-5.99)	-4.77 (-7.22)
$\hat{\mu}_N = \begin{pmatrix} 77.98 \\ 23.74 \\ 26.06 \end{pmatrix}$	ΔS_{IBM}	1.57	-1.94	-4.05	-5.87	-8.18
	ΔS_{MSFT}	5.13	4.05	3.41	2.85	2.13
	ΔS_{TXN}	-6.81	-7.79	-8.37	-8.88	-9.52
	$\Delta S_{Portfolio}$	3.80 (-0.04)	-1.20 (-1.89)	-1.80 (-3.00)	-2.37 (-3.97)	-2.70 (-5.19)

Figure 5.38 graphically describes the relationships between probabilities α and

the VaRs of the four cases. The VaRs of $S_{Portfolio}$ is smaller than those of S_{IBM} and S_{TXN},

which shows the benefit of diversification.
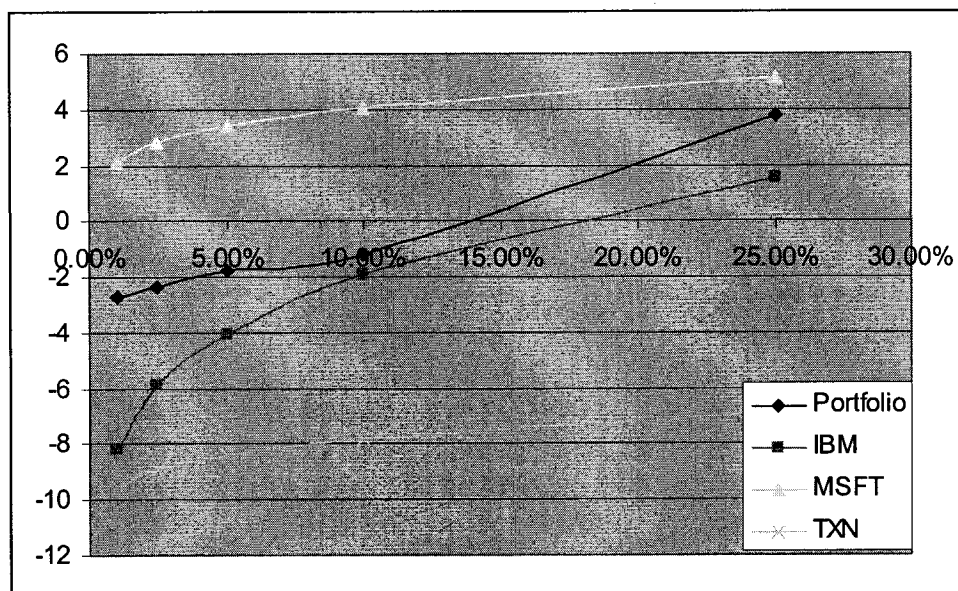
Figure 5.38 Probability α vs. P/L referred to $\hat{\mu}_N$ for MVSN, quadratic model

5.5.2.4 Multivariate skew t-distribution

The simulation results are shown in Figure 5.39: (a), (b) and (c) are the distributions of R_{MRP}, R_{MktRP}, and R_{VvsGP} generated from $ST_3(\vec{0}, \Sigma, \vec{\alpha}, \nu)$, where $\vec{\alpha} = (2,2,2)'$ and $\nu = 5$. Figure 5.39 (d), (f), (g) are the distributions of S_{IBM}, S_{MSFT}, and S_{TXN} calculated according to the quadratic models in Equations (5.34) ~ (5.36).

Similar to the case in MVSN, the distribution of S_{IBM} doesn't change. However, the distribution of S_{MSFT} becomes less skewed and less peaked and the distribution of S_{TXN} becomes more skewed and more peaked. These phenomena can also be explained by the signs of the squared-terms and cross-product terms in quadratic models.
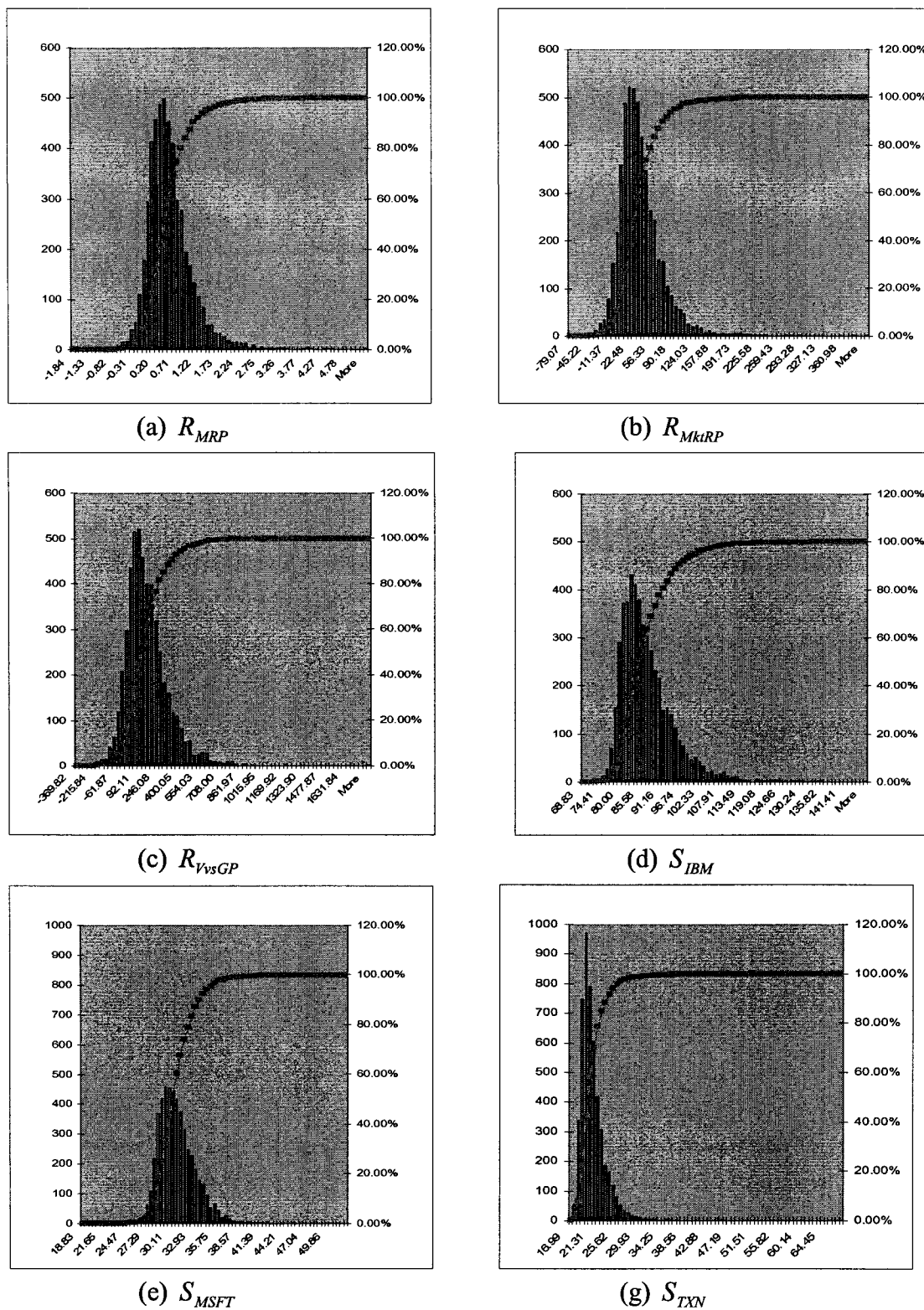
Figure 5.39 The distributions of the three risk factors and three stocks when $\bar{R}$ is multivariate skew t-distribution (MVST) in quadratic models

The risk factors information is the same as in MVST linear models. As a comparison, Table 5.26 lists the means of S_{IBM}, S_{MSFT}, and S_{TXN} calculated from the simulation of MVST quadratic model, denoted as $\hat{\mu}_{st^2}$, and the means estimated in MVN linear model $\hat{\mu}_N$. Similar to the case in MVSN, the means of S_{IBM} and S_{MSFT} move up from the means in MVN linear models. The mean of S_{TXN} is still a smaller number than the mean in MVN linear model.

Table 5.26 Means of three stocks in two cases

	S_{IBM}	S_{MSFT}	S_{TXN}
$\hat{\mu}$ (st, quadratic)	84.86	30.36	20.83
$\hat{\mu}$ (N, linear)	77.98	23.74	26.06

Table 5.27 compares VaRs of the portfolio and the VaRs of individual stock at certain confidence levels. The VaRs of each stock are calculated by constructing histograms based on Monte Carlo simulation results on MVST of the risk factors and the quadratic factor models. The VaRs of the portfolio are calculated by constructing P/L histograms according to Equation (5.34). The calculations results shown in Table 5.27 are based on the sample mean $\hat{\mu}_{st^2}$ and $\hat{\mu}_N$ of MVN in linear model. Diversification effects are shown by comparing the two numbers listed for $\Delta S_{Portfolio}$: the number in the parenthesis is the sum of ΔS_{IBM}, ΔS_{MSFT}, and ΔS_{TXN}. The number without the parenthesis is the value of $\Delta S_{Portfolio}$ by constructing P/L histogram using Monte Carlo simulation results of S_{IBM}, S_{MSFT} and S_{TXN}.

Table 5.27 VaR P/L per share (MVST distribution, quadratic model)

VaR (P/L) at $(1-\alpha)\%$		75%	90%	95%	97.5%	99%
$\hat{\mu}_{st^2} = \begin{pmatrix} 84.86 \\ 30.36 \\ 20.83 \end{pmatrix}$	ΔS_{IBM}	-5.33	-10.13	-13.00	-15.49	-18.65
	ΔS_{MSFT}	-1.62	-3.09	-3.96	-4.72	-5.68
	ΔS_{TXN}	-1.80	-3.42	-4.38	-5.22	-6.29
	$\Delta S_{Portfolio}$	-3.00 (-2.92)	-4.00 (-5.55)	-4.83 (-7.11)	-5.33 (-8.48)	-5.67 (-10.21)
$\hat{\mu}_{N} = \begin{pmatrix} 77.98 \\ 23.74 \\ 26.06 \end{pmatrix}$	ΔS_{IBM}	1.75	-3.05	-5.92	-8.41	-11.57
	ΔS_{MSFT}	5.00	3.53	2.66	1.90	0.94
	ΔS_{TXN}	-7.03	-8.86	-9.61	-10.45	-11.52
	$\Delta S_{Portfolio}$	-0.20 (-0.09)	-1.30 (-2.79)	-1.95 (-4.29)	-2.50 (-5.65)	-3.00 (-7.38)
	$\Delta S_{Portfolio}$	-3.33 (-1.31)	-4.67 (-6.43)	-5.33 (-8.00)	-6.00 (-9.36)	-6.67 (-11.06)

Figure 5.40 graphically describes the relationships between probabilities α and the VaRs of the four cases. The VaRs of $S_{Portfolio}$ is smaller than those of S_{IBM} and S_{TXN} when α is less than 17%, which shows the benefit of diversification.

Figure 5.41 compares the VaRs of the portfolio in four different distributions for quadratic models: MVN, MVT, MVSN and MVST. Since the shaper parameter is positive, even after taking into account the negative impact of squared-terms and cross-product terms, S_{IBM}, S_{MSFT} and S_{TXN} are still positively skewed in MVSN and MVST case. Therefore, the VaRs of the portfolio in MVSN and MVST cases are smaller than those in MVN. The VaRs of MVT is larger than MVN because of the left heavy tail.
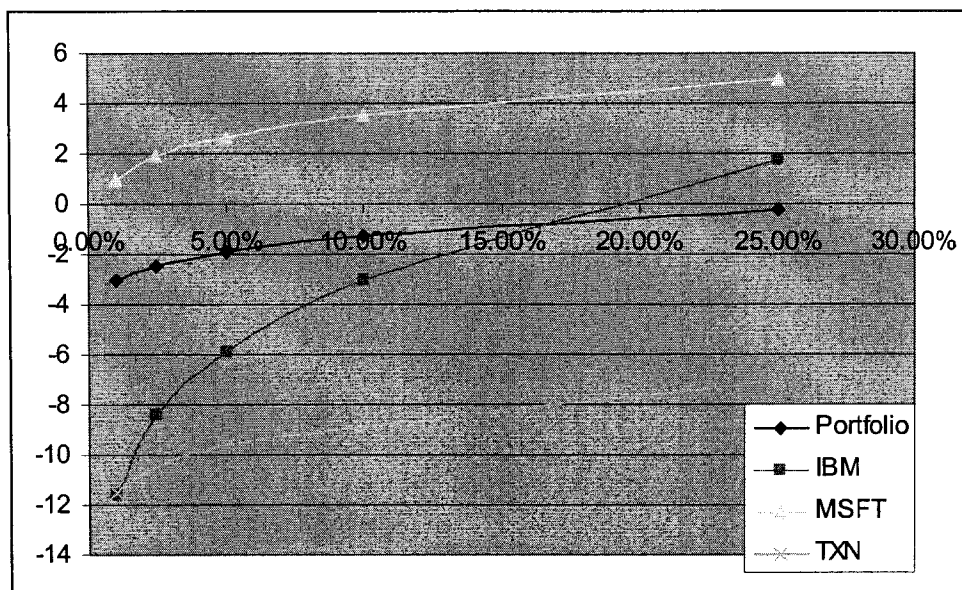
Figure 5.40 Probability α vs. P/L referred to $\hat{\mu}_N$ for MVST, quadratic model
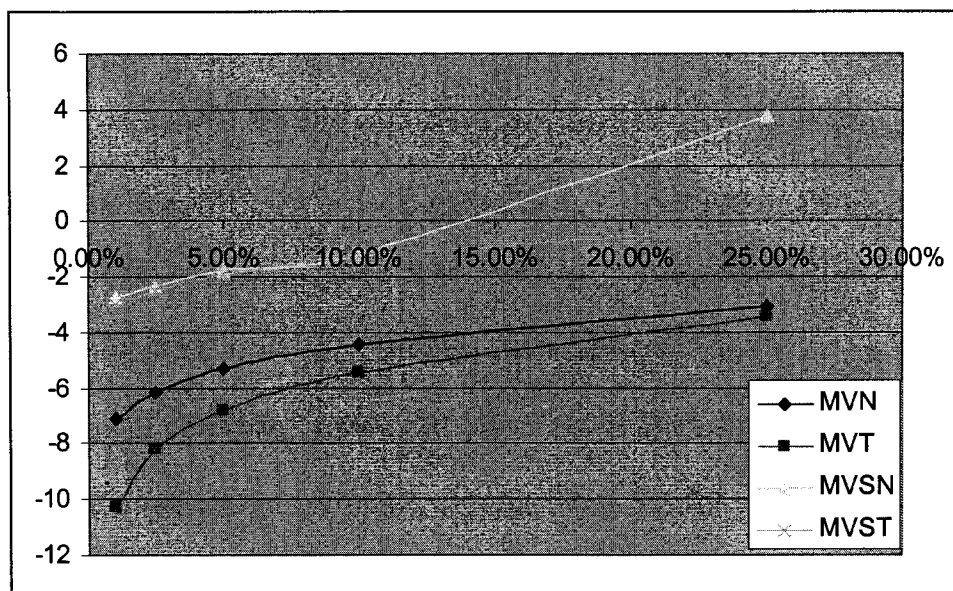
Figure 5.41 Comparison of portfolio VaRs in four distributions for quadratic model

5.6 Discussion

In this chapter, extensive study and simulation has been done on the linear and quadratic three-factor models with four most important multivariate distributions used to describe financial risk factors. One interesting finding is that linear models keep the distribution unchanged. However, the quadratic models may change the distribution depending on the signs of the coefficients of squared terms and cross-product terms. The VaRs of a portfolio with several stocks could be smaller than the VaRs of some individual stock. This difference is the result of diversification. Normal distributions may underestimate the risk and are not good for extreme case simulation since most of the financial data are peaked, skewed, and heavy-tailed. However, skewness makes the situation more complicated, one may overestimate or underestimate the risk. The most important thing is to find out what distribution is appropriate for the financial data being researched.

CHAPTER 6

SUMMARY AND FUTURE STUDY

6.1 Summary and Contributions

Value at Risk (VaR) is a very important risk measure in risk management. Recent years have witnessed an explosive growth and extension of application of VaR in various research areas. Since market risk is the central risk faced by financial institutions, this dissertation focuses on quantifying market risk of an equity portfolio by using VaR methodology.

Since the distribution of the portfolio depends on the possibilities of the risk factors' movements, which can be described as multivariate distributions, it is very interesting to know how the distribution of the risk factors affects the distribution of the portfolio. For this reason, four different multivariate distributions are used to model the possible movements of the three risk factors.

In this research, three risk factors are identified by literature review. Chapter 4 gives a detailed discussion on this. Regression analysis techniques are used to find first-order and second-order models which relate an asset's value to the three common risk factors. All the risk factors are treated as random variables and simulated by Monte Carlo method according to four possible distributions: multivariate normal distribution, multivariate t-distribution, multivariate skew-normal distribution and multivariate skew t-

255

distribution. Through simulation, the change of the assets' value is forecasted as a distribution that is determined by the factor models and the distribution of the risk factors. Based on the asset's distribution, VaRs are calculated for individual assets and for the whole portfolio. This modeling approach can be used for any VaR measurement in any area as long as the relationship between an asset and risk factors can be found and the joint distribution of risk factors can be characterized.

The main contributions of this research are:

1. Extensively simulated a total of four multivariate distributions of the risk factors $\vec{R}$. These four distributions are most often used distributions to describe financial market data. Most of the studies in the literature use normal distributions for simplicity. Few studies use t-distribution, the skew-normal, or skew t. In this dissertation, all four distributions are studied, and the simulation results are compared.

2. Developed linear (first-order) model and quadratic (second-order) model to describe the relationship of an individual asset to the three common risk factors by using regression analysis techniques.

3. Characterized the distribution of portfolio value changes by using both the first-order model and the second-order model $S_i = f(\vec{R})$ to approximate the portfolio function $P = \sum_{i=1}^{3} \omega_i S_i$ according to different multivariate distributions used to model changes in the risk factors $\vec{R}$.

The simulation results in this dissertation shows:

1. Linear transformation doesn not change the characteristics of distribution. In other words, if the three risk factors are multivariate normal distributed, then the distribution of the portfolio is still normal.

2. Quadratic transformation may change the characteristics of the distribution depending on the signs of the coefficients of its second-order terms.

3. The VaR measues of a portfolio with several stocks could be smaller than those of the individual stocks. This is due to the effect of diversification.

4. Normal distributions may underestimate risk and are not good for most practical applications since most of the financial data are peaked, skewed and heavy-tailed.

5. Skewness makes the situation more complicated. One may overestimate or underestimate the risk based on the distribution.

These results underline the importance of determining the proper distribution for the financial data under consideration in order to predict risk accurately.

6.2 Future Study

This research explored a way through simulation to integrate computer science, mathematics, statistics, and financial risk management to calculate VaRs, the market risk of a stock portfolio. This approach helps to design simulations in quantifying market risks in various scenarios according to the distributions of the real data, and to improve the understanding of the VaR results. However, there is still much room for improvement. Future studies may involve the following:

1. Instead of using historical data to construct the variance-covariance matrix of risk factors as input to characterize the future change of the risk factors, ARIMA and state-space models can be used to forecast risk factors. Chapter 3 gives a detailed discussion. In this manner, the joint distribution of the risk factors in the future is determined by time series forecasting techniques, instead of generating from the historical variance-covariance matrix.

2. Construct a more diversified portfolio by choosing stocks from different industry in order to reduce further the VaRs of a portfolio.

3. Similar to constructing efficient frontier (EF) in modern portfolio theory (MPT), an EF (as described in Section 4.2.1) can also be constructed for a portfolio in terms of the value of the portfolio and the standard deviation of the portfolio by varying the weights of each asset in the portfolio.

4. The present simulation technique can be improved by using copula (as described in Section 5.2.5.3), which makes it possible to simulate multivariate t-distribution with different degrees of freedom for each element of the random vector. This will be also useful in the simulation of multivariate t related distribution.

5. Based on VaR measurement, another important financial risk measure-expected tail loss (ETL) or expected shortfall (ES)-can be calculated as $ES = E[L \,|\, L > VaR]$. Here, L means loss. There is a short description about ES in Chapter 1.

APPENDIX

SOURCE CODES

259

```
/*
        Monte Carlo Simulation on Multivariate Distribution
        Four multivariate distributions are simulated by using Monte Carlo simulation
        method. They are

            1. Multivariate Normal Distribution
            2. Multivariate t-Distribution
            3. Multivariate Skew-Normal Distribution
            4. Multivariate Skew t-Distribution

        "nrutil.h" and "nrutil.c" from <<Numerical Recipes in C>> are used for
        vector and matrix implementation.
*/


#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include "nrutil.h"
#include "nrutil.c"
#define TRUE 1
#define FALSE 0
#define PI 3.141592653589793

struct t_bc {double a, b, h;} bc;

double func_standard_normal(float x);
void trapz_int(double *ss, double a, double b, int n);
void dvec_uniform( int n, int *seed, double r[] );
void func(double x, double *y_derivative);
void gen_normal_vector(double **z, int dimension, int size);
void gen_skewnormal_vector(double **z, double **u, int dimension, int size);
void choldc(float **a, int n, float p[]);
void LU_decom(double **L, int s);
int MatrixInversion1(double** A, int n, double** AInverse);
void shape_vector(double **Omega_inv, double *Delta, int n);
void gen_chi_sqr_ran(double *v, int size, int df);
double AChiSq(double p, int n);
double ChiSq(double x, int n);
void gen_uni_chi_sqr(double *v, double *r, int size, int df);
void gen_uni_normal_vector(double *z, double *r, int size);
void LU_decom_4T(double **L, int s);

void main()
{
        int  i, j, k, size, dimension, n, s, df, yes;
        int seed = 123456789, seed_in;
```

```
double mu, var, **z, **u, **L, *sn, **SN, **v, *vv, *tt, *st, **ST, x, **r, **zz,
*u1, *u0, *Z, **T, *t;
float delta, alpha;

/*    Univariate T Distribution   */

printf(" Please enter how many skew t random variables in your simulation.\n");
scanf("%d", &dimension);
printf("\n");

n = 2 * dimension + 1;

printf(" Please enter how many pseudorandom numbers you would like to
generate for each t random variable.\n");
scanf("%d", &size);
printf("\n");

r = (double **) dmatrix(1, n, 1, size);

for (i = 1; i <= n; i++)
{
   seed_in = seed;
   dvec_uniform(size, &seed, r[i]);
}

if (dimension == 1)
{
        zz = (double **) dmatrix(1, dimension + 1, 1, size);
        for (i = 1; i <= dimension + 1; i++)
                gen_uni_normal_vector(zz[i], r[i], size);


        // alpha = delta * zz[1] + sqrt(1 - delta ** 2) * zz[2]
        printf(" Please enter the shape parameter for the SN.\n");
        scanf("%f", &alpha);

        delta = alpha / sqrt(1 + alpha * alpha);
        u0 = (double *) dvector(1, size);
        u1 = (double *) dvector(1, size);

        // Normal with correlation delta
        for (i = 1; i <= size; i++)
                u0[i] = zz[1][i];

        for (i = 1; i <= size; i++)
                u1[i] = delta * zz[1][i] + sqrt(1 - delta * delta) * zz[2][i];
```

```
// Skew Normal
Z = (double *) dvector(1, size);

for (i = 1; i <= size; i++)
{
        if (u0[i] >= 0.0)
                Z[i] = u1[i];
        else
                Z[i] = (-1.0) * u1[i];
}

printf(" Please enter the degree of freedom for the chi-square random
variable.\n");
scanf("%d", &df);

vv = (double *) dvector(1, size);

gen_uni_chi_sqr(vv, r[3], size, df);   // chi-squared random numbers

st = (double *) dvector(1, size);
t  = (double *) dvector(1, size);

/* T distribution */
for (i = 1; i <= size; i++)
        t[i] = u1[i] / sqrt(vv[i] / df);

/* Skew T distribution */
 for (i = 1; i <= size; i++)
        st[i] = Z[i] / sqrt(vv[i] / df);

free_dvector(vv, 1, size);
free_dvector(st, 1, size);
free_dvector(Z, 1, size);
free_dvector(u0, 1, size);
free_dvector(u1, 1, size);
free_dvector(t, 1, size);
free_dmatrix(zz, 1, dimension+1, 1, size);
}
else
{
        z = (double **) dmatrix(1, dimension + 1, 1, size);
        for (i = 1; i <= dimension + 1; i++)
                gen_uni_normal_vector(z[i], r[i], size);
```

```
L = (double **)dmatrix(1, dimension + 1, 1, dimension + 1);

LU_decom_4T(L, dimension + 1);

//Multivariate Standard Normal Distribution with above variance-
covariance matrix

u = (double **) dmatrix(1, dimension + 1, 1, size);

for (i = 1; i <= dimension + 1; i++)
        for (j = 1; j <= size; j++)
                u[i][j] = 0.0;

// Normal with var-covar matrix
for (j = 1; j <= size; j++)
{
        for (i = 1; i <= dimension + 1; i++)
        {
                for (k = 1; k <= dimension + 1; k++)
                        u[i][j] += L[i][k] * z[k][j];
        }
}

//Skew Normal
zz = (double **) dmatrix(1, dimension, 1, size);

for (i = 1; i <= size; i++)
{
        if (u[1][i] >= 0.0)
        {
                for (j = 1; j <= dimension; j++)
                        zz[j][i] = u[j+1][i];
        }
        else
        {
                for (j = 1; j <= dimension; j++)
                        zz[j][i] = (-1.0) * u[j+1][i];
        }
}

printf(" Please enter the degree of freedom for the chi-square random
variables.\n");
scanf("%d", &df);

v = (double **) dmatrix(1, dimension, 1, size);
```

```
        for (i = 1; i <= dimension; i++)
                gen_uni_chi_sqr(v[i], r[i+dimension+1], size, df);

        // T
        T = (double **) dmatrix(1, dimension, 1, size);

        for (i = 1; i <= dimension; i++)
        {
                for (j = 1; j <= size; j++)
                        T[i][j] = u[i+1][j] / sqrt(v[i][j] / df);
        }

        //Skew T
        ST = (double **) dmatrix(1, dimension, 1, size);

        for (i = 1; i <= dimension; i++)
        {

                for (j = 1; j <= size; j++)
                        ST[i][j] = zz[i][j] / sqrt(v[i][j] / df;
        }
        free_dmatrix(zz, 1, dimension, 1, size);
        free_dmatrix(v, 1, dimension, 1, size);
        free_dmatrix(u, 1, dimension + 1, 1, size);
        free_dmatrix(ST, 1, dimension, 1, size);
        free_dmatrix(T, 1, dimension, 1, size);
        free_dmatrix(r, 1, dimension * 2 + 1, 1, size);

    }
}

void gen_normal_vector(double **z, int dimension, int size)
{
        /* r - random number, z - indep. std normal, u - std normal with corr matrix */
        double **r, y_derivative;
        int i, j, seed_in, n;
        int seed = 123456789;
        double s, epsilon, error, x, xb, y;

        r = (double **) dmatrix(1, dimension, 1, size);

        for (i = 1; i <= dimension; i++)
        {
                seed_in = seed;
                dvec_uniform(size, &seed, r[i]);
        }
```

```
epsilon = 0.00001;

for (i = 1; i <= dimension; i++)
{
        for (j = 1; j <= size; j++)
        {

                x = 0.0;       /* initial guess for x */
                xb = x;
                n = 0;
                error = 1.0e10;
                while (error > epsilon)
                {
                        n = n + 1;
                        func(x, &y_derivative);
                        /* lower_limit = -10, # of intervals = n, upper_limit = x */
                        trapz_int(&s, -10, x, 800);
                        y = s - r[i][j];
                        x = x - y / y_derivative;       /* finds new x  */

                        error = fabs(x - xb);
                        xb = x;
                        if (n > 50)
                        {
                                x = x + 1.0;
                                n = 0;
                        }

                }
                /*   Normal Distribution with expection 0 and variance 1  */
                z[i][j] = x;
        }
}
free_dmatrix(r, 1, dimension, 1, size);
}

void gen_uni_normal_vector(double *z, double *r, int size)
{
        double y_derivative;    /* r - random number, z - indep. std normal */
        int i, j, n;
        double s, epsilon, error, x, xb, y;

        epsilon = 0.00001;

        for (i = 1; i <= size; i++)
        {
```

```
x = 0.0;        /* initial guess for x */
xb = x;
n = 0;
error = 1.0e10;
while (error > epsilon)
{
        n = n + 1;
        func(x, &y_derivative);
        /* lower_limit = -10, # of intervals = n, upper_limit = x */
        trapz_int(&s, -10, x, 800);
        y = s - r[i];
        x = x - y / y_derivative;        /* finds new x */
        error = fabs(x - xb);
        xb = x;

        if (n > 50)
        {
                x = x + 1.0;
                n = 0;
        }
}
        /*   Normal Distribution with expection 0 and variance 1   */
        z[i] = x;
}
}


void gen_skewnormal_vector(double **z, double **u, int dimension, int size)
{
        int i,j;
        double alpha, delta;
        alpha = 10.0;
        delta = alpha / sqrt (1 + alpha * alpha);
        for (i = 1; i <= dimension; i++)
        {
                for (j = 1; j <= size; j++)
                {
                        u[i][j] = delta * z[i][j] + sqrt(1 - delta * delta) * z[i+1][j];

                        if (z[i][j] >= 0.0)
                                u[i][j] = z[i][j];
                        else
                                u[i][j] = (-1.0) * z[i][j];
                }

        }
}
```

```
void trapz_int(double *ss, double a, double b, int n)   /* Trapezoidal rule */
{
        int i;
        double sum, w, x;
        sum = 0;

        bc.a = a;
        bc.b = b;
        bc.h = (bc.b - bc.a) / n;

        for (i = 0; i <= n; i++)
        {
                x = bc.a + i * bc.h;
                w = 2;
                if ( i == 0 || i == n)
                        w = 1;
                sum = sum + w * func_standard_normal (x);
        }
        *ss = sum * bc.h / 2;
        return;
}

double func_standard_normal(float x)
{
        double func_v;

        /* f(x) is a pdf for standard normal distribution.*/
        func_v = exp((-0.5) * pow( x, 2)) /(sqrt(2 * PI));
        return (func_v);
}

/*
   This routine implements the recursion

     seed = 16807 * seed mod ( 2**31 - 1 )
     unif = seed / ( 2**31 - 1 )
*/
void dvec_uniform( int n, int *seed, double r[] )
{
        int i;
        int k;

        for ( i = 1; i <= n; i++ )
        {
                k = *seed / 127773;
```

```
            *seed = 16807 * ( *seed - k * 127773 ) - k * 2836;

            if ( *seed < 0 )
                    *seed = *seed + 2147483647;

            r[i] = ( double ) ( *seed ) * 4.656612875E-10;
        }
}

/* This function computes y and y_derivative */
void func(double x, double *y_derivative)
{
    *y_derivative = exp((-0.5) * pow( x, 2)) /(sqrt(2 * 3.1415926));
}

/* Given a positive-definite symmetric matrix a[1..n][1..n], this routine constructs
   its Cholesky decomposition, A = L * L(T). On input, only the upper triangle of a
   need be given; it is not modified. The Cholesky factor L is returned in the lower
   triangle of a, except for its diagonal elements which are returned in p[1..n] */

void choldc(float **a, int n, float p[])
{
        int i, j, k;
        float sum;

        for (i = 1; i <= n; i++)
        {
                for (j = 1; j <= n; j++)
                {
                        for (sum = a[i][j], k = i - 1; k >= 1; k--)
                                sum -= a[i][k] * a[j][k];
                        if (i == j)
                        {
                                if (sum <= 0.0)
                                        nrerror ("cholesky decomposition failed");
                                p[i] = sqrt(sum);
                        }
                        else
                                a[j][i] = sum / p[i];
                }
        }
}

void LU_decom(double **L, int s)
{
```

```
float **a, *p;
int i, j, t;
double *delta, **omega, **omega_inv, del, alp, E_X, Var_X, SKEW;

a = (float **)matrix(1, s, 1, s);
p = (float *)vector(1,s);

printf("Please enter the variance-covariance matrix:\n");

for (i = 1; i <= (s - 1) + 1; i++)
{
        for (j = 1; j <= (s - 1) + 1; j++)
                scanf(" %f", &a[i][j]);
}

if (s == 2)    // Univariate
{
        del = a[1][2]; // del is the correlation coeffient
        alp = del / sqrt(1 - del * del);

        printf(" For univariate SN, delta = %6.4f, the shape parameter alpha =
        %6.4f.\n", del, alp);

        E_X = sqrt(2 / PI) * del;
        Var_X = 1 - 2 * del * del / PI;
        SKEW = ((4 - PI) / 2) * pow (E_X, 3) / pow (Var_X, 1.5);
}
else if (s > 2) // multivariate
{
        delta = (double *)dvector(1, s - 1);
        omega = (double **)dmatrix(1, s - 1, 1, s - 1);

        for (j = 1; j <= (s - 1); j++)
        {
                delta[j] = a[1][j+1];
        }

        for (i = 1; i <= (s - 1); i++)
        {
          for (j = 1; j <= (s - 1); j++)
          {
                omega[i][j] = a[i+1][j+1];
          }
        }
        omega_inv = (double **)dmatrix(1, s - 1, 1, s - 1);
        t = MatrixInversion1(omega, s - 1, omega_inv);
```

```
                shape_vector(omega_inv, delta, s - 1);
        }

        choldc(a, s, p);

        for (i = 1; i <= (s - 1) + 1; i++)
        {
                for (j = 1; j <= (s - 1) + 1; j++)
                        L[i][j] = a[i];
        }

        for (i = 1; i <= s; i++)
                L[i][i] = p[i];

        free_matrix(a, 1, s, 1, s);
        free_vector(p, 1, s);

        if (s > 2)
        {
          free_dmatrix(omega, 1, s - 1, 1, s - 1);
          free_dmatrix(omega_inv, 1, s - 1, 1, s - 1);
        }
}

void LU_decom_4T(double **L, int s)
{
        float **a, *p;
        int i, j, t;

        a = (float **)matrix(1, s, 1, s);
        p = (float *)vector(1,s);

        printf("Please enter the variance-covariance matrix:\n");

        for (i = 1; i <= (s - 1) + 1; i++)
        {
                for (j = 1; j <= (s - 1) + 1; j++)
                        scanf(" %f", &a[i][j]);
        }

        choldc(a, s, p);

        for (i = 1; i <= (s - 1) + 1; i++)
        {
                for (j = 1; j <= (s - 1) + 1; j++)
                        L[i][j] = a[i][j];
```

```
        }

        for (i = 1; i <= s; i++)
                L[i][i] = p[i];

        free_matrix(a, 1, s, 1, s);
        free_vector(p, 1, s);
}

//Matrix Inversion Routine
int MatrixInversion1(double** A, int n, double** AInverse)
{
        // A = input matrix (n x n)
        // n = dimension of A
        // AInverse = inverted matrix (n x n)
        // This function inverts a matrix based on the Gauss Jordan method.
        // The function returns 1 on success, 0 on failure.
        int i, j, iPass, imx, icol, irow;
        double det, temp, pivot, factor;
        //float* ac = (float*)calloc(n*n, sizeof(float));
        double **ac;

        ac = (double **)dmatrix(1, n, 1, n);

        det = 1;

        for (i = 1; i <= n; i++)
        {
                for (j = 1; j <= n; j++)
                {
                        AInverse[i][j] = 0;
                        ac[i][j] = A[i][j];
                }
                AInverse[i][i] = 1;
        }

        // The current pivot row is iPass.
        // For each pass, first find the maximum element in the pivot column.
        for (iPass = 1; iPass <= n; iPass++)
        {
                imx = iPass;
                for (irow = iPass; irow <= n; irow++)
                {
                        if (fabs(A[irow][iPass]) > fabs(A[imx][iPass])) imx = irow;
                }
```

```
                    // Interchange the elements of row iPass and row imx in both A and
AInverse.
        if (imx != iPass)
        {
                for (icol = 1; icol <= n; icol++)
                {
                        temp = AInverse[iPass][icol];
                        AInverse[iPass][icol] = AInverse[imx][icol];
                        AInverse[imx][icol] = temp;

                        if (icol >= iPass)
                        {
                                temp = A[iPass][icol];
                                A[iPass][icol] = A[imx][icol];
                                A[imx][icol] = temp;
                        }
                }
        }


        // The current pivot is now A[iPass][iPass].
        // The determinant is the product of the pivot elements.
        pivot = A[iPass][iPass];
        det = det * pivot;
        if (det == 0)
        {
                free_dmatrix(ac, 1, n, 1, n);
                return 0;
        }

        for (icol = 1; icol <= n; icol++)
        {
                // Normalize the pivot row by dividing by the pivot element.
                AInverse[iPass][icol] = AInverse[iPass][icol] / pivot;
                if (icol >= iPass) A[iPass][icol] = A[iPass][icol] / pivot;
        }

        for (irow = 1; irow <= n; irow++)
        // Add a multiple of the pivot row to each row.  The multiple factor
        // is chosen so that the element of A on the pivot column is 0.
        {
                if (irow != iPass) factor = A[irow][iPass];
                for (icol = 1; icol <= n; icol++)
                {
                        if (irow != iPass)
                        {
                                AInverse[irow][icol] -= factor * AInverse[iPass][icol];
```

```
                                A[irow][icol] -= factor * A[iPass][icol];
                              }
                        }
                  }
            }

            free_dmatrix(ac, 1, n, 1, n);
            return 1;
}

void shape_vector(double **Omega_inv, double *Delta, int n)
{
            double *alpha, *temp1_v, *temp2_v, temp;
            int i, j;

            alpha = (double *)dvector(1, n);
            temp1_v = (double *)dvector(1, n);
            temp2_v = (double *)dvector(1, n);

            for (i = 1; i <= n; i++)
            {
                  temp1_v[i] = 0.0;
                  temp2_v[i] = 0.0;
            }
            temp = 0.0;
            for (i = 1; i <= n; i++)
            {
                  for (j = 1; j <= n; j++)
                        temp1_v[i] += Omega_inv[i][j] * Delta[j];
            }

            for (j = 1; j <= n; j++)
            {
                  for (i = 1; i <= n; i++)
                        temp2_v[j] += Delta[i] * Omega_inv[i][j] ;
            }

            for (i = 1; i <= n; i++)
            {
                  temp += temp2_v[i] * Delta[i];
            }

            for (i = 1; i <= n; i++)
            {
                  alpha[i] = temp1_v[i] / sqrt(1 - temp);
            }
```

```c
        free_dvector(alpha, 1, n);
        free_dvector(temp1_v, 1, n);
        free_dvector(temp2_v, 1, n);
}

void gen_chi_sqr_ran(double *v, int size, int df)
{
        double *r;      /* r - random number */
        int i, seed_in;
        int seed = 123456789;

        for (i = 1; i <= size; i++)
                v[i] = AChiSq(1 - r[i], df);
}

void gen_uni_chi_sqr(double *v, double *r, int size, int df)
{
        int i;

        for (i = 1; i <= size; i++)
                v[i] = AChiSq(1 - r[i], df);
}

double ChiSq(double x, int n)
{
        double p, t;
        int k, a;

        p = exp(-0.5*x);

        if((n%2)==1)
                p = p * sqrt(2 * x / PI) ;
        k = n;

        while(k >= 2)
        {
                p = p * x / k;
                k = k-2;
        }

        t = p;
        a = n;

        while(t > 1e-15 * p)
        {
```

```
                    a = a + 2;
                    t = t * x /a;
                    p = p+t;
            }
        return 1-p;
}

double AChiSq(double p, int n)
{
            double v, dv, x;

            v = 0.5;
            dv = 0.5;
            x = 0.0;

            while(dv > 1e-10)
            {
                    x = 1/v - 1;
                    dv = dv / 2;
                    if(ChiSq(x,n) > p)
                            v = v - dv;
                    else
                            v = v + dv;
            }
        return x;
}
```

REFERENCES

1. Jorion, P. 2001. *Value at Risk: the new benchmark for managing financial risk.* McGraw-Hill, New York.

2. News: Analysis & Commentary: *Today, Orange County....* Business Week: December 19, 1994

3. Skora, R. K. 1998. Review of The Practice of Risk Management – Implementing processes for managing firmwide market by Goldman Sachs and SBC Warburg Dillon Road, Euromoney Books, London.

4. Gallati, R. 2003. *Risk Management and Capital Adequacy.* McGraw-Hill, Inc.

5. Dowd. K. 2005. *Measuring Market Risk,* second edition. John Wiley & Sons, Chichester, England.

6. Best, P. W. 1998. *Implementing Value at Risk,* John Wiley & Sons, Chichester, England.

7. Alexander, C. 1998. *Risk Management And Analysis Volume 1 Measuring And Modeling Financial Risk* (ed.), John Wiley & Sons, Chichester, England.

8. Markowitz, H. M. 1952. Portfolio Selection, *Journal of Finance,* 7 (1): 77-91.

9. US Department of the Treasury, 2003. Supervisory Guidance on Operational Risk: Advanced Measurement Approaches for Regulatory Capital.

10. Holton, G. A. 1997. Subjective Value-at-Risk. *Financial Engineering News.* 1(1), 1, 8-9, 11.

11. Glasserman, P. 1999. *Value at Risk (A),* Columbia Business School.

12. J.P Morgan / Reuters, 1996. RiskMetrics™ -- Technical Document.

13. Alexander, C. 2001. *Market Models,* John Wiley & Sons, Chichester, England.

14. Wiener, Z. 1997. Introduction to VaR (Value-at-Risk). *Risk Management and Regulation in Banking.*

15. Gerhard, S. 1997. Three Cheers. *Risk*. Vol. **10,** no. **5**: 67-69.

16. Wilson, C. T. 1998. *"Value at Risk", Risk Management and Analysis Volume 1: Measuring and Modeling Financial Risk*. John Wiley & Sons, Chichester, England.

17. Danielson, J., de Vries, C.G., Jorgensen, B. N. 1998. The Value of Value at Risk: Statistical, Financial, and Regulatory Considerations. FRBNY Economic Review.

18. Li, S., Kasthuri, R. 2004. Sector Exposure Models for Risk Management of Security Portfolio, *working paper*, found from www.google.com.

19. Madhaven, A., Yang, J. 2002. Practical Risk Analysis for Portfolio Managers and Traders, *working paper*, ITG Inc., 380 Madison Avenue, New York, NY 10017.

20. Fama, E. F., French, K. R. 1993. Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* **33**: 3-56.

21. Holton, G. A. 2003. *Value-at-Risk: Theory and Practice*, Academic Press.

22. http://www.riskglossary.com/link/var_measure.htm

23. Christoffersen, P. F. 2003. *Elements of Financial Risk Management*, Academic Press.

24. Cornish, E. A., Fisher, R.A. 1937. Moments and Cumulants in the Specification of Distributions. *Review of the International Statistical Institute*, **5**: 307-320.

25. Pindyck, R. S. and Rubinfeld, D. L. 1981. *Econometric Models & Economic Forecasts*, second edition. McGraw-Hill Book Company.

26. Franses, P. H., 2002. *Time Series Models for Business and Economic Forecasting*. Cambridge University Press.

27. NIST/SEMATECH e-Handbook of Statistical Methods, 2006, http://www.itl.nist.gov/div898/handbook/.

28. Engle, R.F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, **50**: 987-1007.

29. Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**:. 307-327.

30. Terasvirta, T. 1994. Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*. **89**: 208-218.

31. Van Dijk, D., Terasvirta, T., Davis, Franses, P.H. 2002. Smooth Transition Autoregressive Models – A Survey of Recent Developments. *Econometric Reviews.* **21**(1): 1-47.

32. Hamilton, J.D. 1989. A New Approach to the Econometric Analysis of Nonstationary Time Series and Business Cycles, *Econometrica*, **57**: 357-384.

33. Tsay, R S. 2002.*Analysis of Financial Time Series.* John Wiley & Sons, New York.

34. Kim, C.J., Nelson, C.R. 1999. *State-Space Models with Regime Switching.* The MIT Press.

35. http://documents.wolfram.com/applications/timeseries/

36. Yakir, B. 2002. Notes: Advance Statistical Models.

37. Brockwell, Peter J. 2002. *Introduction to Time Series and Forecasting.* New York: Springer.

38. Risk-Factor Regression Analysis, http://www.csulb.edu/~pammerma/fin450/ .

39. SAS Institute Inc., 1996. *Forecasting Examples for Business and Economics Using the SAS System.* Cary, NC, USA.

40. *You Changed America's Heart* – A 50th anniversary tribute to the participants in the Framingham Heart Study. http://www.nhlbi.nih.gov/about/framingham/fhsbro.pdf.

41. Yang, J., Madhaven, A. 2002. Practical Risk Analysis for Portfolio Managers and Traders. ITG Inc..

42. Portfolio Analytics Group, 2003. Selecting Risk Factors. Financial Engineering, ITG Inc.

43. http://www.finflowholdings.com/ffh/multifactor_model.shtml

44. Gallati, R. 2003. *Risk Management and Capital Adequacy.* McGraw-Hill.

45. http://www.moneychimp.com/articles/risk/riskintro.htm

46. http://www.riskglossary.com/link/measuring_market_risk.htm

47. Farrell, J.L.1983. *Guide to Portfolio Management.* McGraw-Hill.

48. Li, S. Kasthuri, R. 2004. Sector Exposure Models for Risk Management of Security Portfolio, *working paper*.

49. Fama, E.F., French, K.R. 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*. **33**(1): 3–56.

50 http://faculty.washington.edu/ezivot/econ512/econ512factormodels.pdf

51. Montgomery, D.C. Peck, E.A. Vining, G.G. 2004. *Introduction to Linear Regression Analysis*. John Wiley & Sons, Inc..

52. Zangari, P. 1996. Market risk Methodology, *RiskMetrics – Technical Document*, 4th ed., New York: Morgan Guaranty: 105 – 148.

53. Johnson, N. L. 1949. Systems of Frequency Curves Generated by Methods of Translation, *Biometrika*, **36**: 149 – 176.

54. Fallon, W. 1996. Calculating Value-at-Risk, *working paper*.

55. Pichler, S., Selitsch, K. 2000. *A Comparison of Analytic VaR Methodologies for Portfolios that Include Options, Model Risk: Concepts, Calibration and Pricing*, Rajna Gibson (editor), London: Risk Books.

56. Cárdenas, J., Fruchard, E., Picron, J., Reyes, C., Walters, K., Yang W. 1999. Monte Carlo within a Day, *Risk*, **12**(2): 55-59.

57. Sobol', I. M.. 1994. *A Primer for the Monte Carlo Method*, CRC Press.

58. Kijima, M. 2003. *Stochastic Processes with Applications to Finance*, Chapman & Hall/CRC, New York.

59. Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering*, Springer Inc., New York.

60. Rubinstein, R. Y., 1981. *Simulation and the Monte Carlo Method*, John Wiley & Sons, New York.

61. Avramidis, A.N., Wilson, J.R. 1998. Correlation-Induction Techniques for Estimating Quantiles in Simulation Experiments, *Operations Research*, **46**:574-591.

62. http://en.wikipedia.org/wiki/Skewness

63. Hogg, R. V., McKean, J. W., Craig, A. T., 2005. *Introduction to Mathematical Statistics, Pearson Prentice Hall*, New Jersey.

64. Azzalini, A., Valle, A.D., 1996. The Multivariate Skew-Normal Distribution, *Biometrika*, **83**(4): 715-726.

65. Azzalini, A., 1985. A Class of Distributions which Includes the Normal Ones. *Scandinavian Journal of Statistics, Theory and Applications*, **12**:171-178.

66. Azzalini, A., 1999. Statistical Application of the Multivariate Skew-Normal Distribution, *Journal of Royal Statistical Society* B, **61**: 579-602.

67. http://azzalini.stat.unipd.it/SN/plot-SN1.html

68. http://www.nag.co.uk/numeric/FL/manual/pdf/G05/g05lxf.pdf

69. Kan, R., Zhou, G., 2006. Modeling Non-Normality Using Multivariate t: Implication for Asset Pricing. *Working paper*.

70. Azzalini, A., 2003. Distribution Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t-Distribution, *Journal of Royal Statistical Society* B, **65**: 367-389.

71. Kim, H-M., 2004. Quadratic Forms of Skew t-Distribution with an Application to Spacial and Time Series Analysis. *Working paper*.

72. Moro, B. 1995. The full Monte. *Risk*, Feb: 57-58.

73. Johnson, M. E. 1987. *Multivariate Statistical Simulation*, John Wiley & Sons, New York.