Risk Factors in Greek Companies: An Empirical Analysis

Theodore Priniotakis

Thesis Submitted for the Degree of PhD in Finance

CITY UIVERSITY LONDON

CASS BUSINESS SCHOOL

MARCH 2004

Table of contents

Chapter 2: Descriptive Analysis of Credit Risk Factors

Chapter 3: Empirical Results on credit risk factors

Chapter 4: Measuring the Market Risk of Equities: A Review of Alternative Approaches and Some Empirical Tests on the Their Assumptions.

List of tables

List of illustrations

Acknowledgements

I would like to thank the many people who have helped me to complete this thesis. First of all I would like to express my debt to Professor Elias Dinenis who has been a helpful and understanding mentor throughout my studies. I would also like to thank Professor Kate Phylaktis and Dr Alistair Milne for the useful comments they made on earlier drafts of this thesis. I would also like to thank Professor Giovanni Urga for the support that he has provided as director of the PhD programme. Finally, I would like to express my deep gratitude to my family for the financial and moral support they have given me throughout my studies.

Declaration

I hereby grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

The purpose of this thesis is to develop a framework for measuring and modelling the market and credit risks of Greek companies. The thesis is motivated by the tremendous transformation of both the regulatory requirements as well as the techniques for measuring risks and by the fact that stakeholders in companies have become aware in the last few years of the need for understanding the risks to which companies are exposed.

This thesis is divided into two parts. In the first part of the thesis we deal with the first major category of risks related to a company, that is credit risks. The credit risks are estimated using a sample of companies that were unable to service their bank loans to a specific bank. The issue we deal with in the first part is to identify a set of risk factors that will help us identify companies which may exhibit a high degree of credit risk as it is measured by their inability to fulfil their obligation to a bank.

The empirical results of the thesis shed some light on the behaviour of companies in Greece and highlight the factors that may help us to quantify and predict the credit risk of a company.

In the second part we deal with issues relating to the measurement of market risk of Greek quoted companies. The approach we take in this part is to look at the issues involved both at the level of measuring the risk of individual equities as well as measuring the risk in a portfolio context and to apply the various techniques that have been proposed to the study of the market risks of Greek companies. Such an analysis will be helpful to investors, portfolio managers and regulators since, we hope, it will produce a clearer picture of market risks.

The two parts are complementary, for whereas in the second part we deal large companies, since only the larger companies are quoted on the Athens Stock Exchange, in the first part we deal primarily with small and medium-sized enterprises which represent the majority of corporate entities in Greece.

Introduction

Overview of the Thesis

The thesis is organized into five chapters, of which the first three deal with aspects of credit risk and the last two with issues of market risk.

Chapter 1: Review of the literature on measuring credit risk

This chapter, which is the first of Part One deals with two issues. First it sets out the objectives of this part of the thesis and the research design. This is placed in the appropriate context by reviewing the literature on credit risk measurement which is growing exponentially. Starting with the review of the literature on credit risk measurement, we identify various strands under which the various approaches could be classified. We have grouped the various approaches into the portfolio approach and the statistical approach which deals with the risks of an individual company. These two broad approaches are of course linked and the development of the more recent portfolio approach compliments earlier efforts but it has also been influenced by them. In a portfolio context, we have examined the four main approaches, CrediMetrics, CreditPortfolio View, Credit Risk + and the KMV model. All these models require estimates of the probability of default and of the probability of migration. The probability of default can be estimated directly using one of the statistical approaches such as the logistic approach, discriminant analysis or classification trees. The probability of migration is also derived from historical migration matrices of the credit rating companies. We critically assess the assumptions of these models and their performance. Looking more closely at the statistical approach, which is more akin to the objective of this thesis, we review a number of approaches such as the linear probability model, linear and quadratic discriminant analysis, logistic regression, the Tobit model, survival analysis and classification trees. We pay particular evidence to the empirical studies of discriminant analysis and logistic regression since these are the techniques we shall employ in this thesis.

In the second part of Chapter 1 we explain the research strategy and the design of our test. We explain how using a unique set of data on companies that experienced financial distress manifested by their inability to service their debt to a bank, the identification of risk factors will be conducted and the estimation of credit risk will be realized. These are not bankrupt firms in a legal sense since the declaration of bankruptcy is a judicial decision that may take many years. Our definition is more akin to the economic definition of bankruptcy rather than the legal one. The fundamental hypothesis is whether, two years prior to being made bankrupt a company can be identified on the basis of a set of company characteristics, primarily financial ratios.

Chapter 2 Descriptive Analysis of Credit Risk Factors

In Chapter 2 we define the characteristics of a company that may give rise to credit risk and examine the statistical properties of these characteristics. In the first part of Chapter 2, we define the set of risk factors that may potentially predict which companies are likely to be experiencing financial problems in the following two years. The set of risk factors was selected on the basis of theoretical model such as Merton's model of risky debt, from variables that have been used by credit rating agencies or variable which have been found significant in other studies. We concentrate exclusively on financial characteristics and leave out qualitative factors such as the quality management, factors on which we have no data. In the second part of Chapter 2 we examine the statistical properties of a number of financial ratios of both financially distressed and healthy companies. We find that the average values of the financial ratios are significantly different in the two groups. This is supported both from the univariate as well as from the multivariate test. The median values are also significantly different but the variances of financial ratios in the two groups are not significantly different, showing that the distribution of ratios in the two groups differ in terms of the location parameters only. We have also tested for two assumptions which are crucial for the statistical analysis of Chapter 3, namely the assumption of normally distributed ratios and the assumption of equal covariance matrices. We reject both assumptions. We found that most of the ratios exhibit significant skewness and kurtosis coefficients. We also reject the hypothesis of equal covariance matrices, which in conjunction with the univariate acceptance of equal variances means that the differences in the covariance matrices must be due to differences in the correlation structure of the financial ratios. An important issue in the use of financial ratios is the stability of the ratios. We have found that the majority of the means of the ratios in our sample have remained relatively stable.

Chapter 3: Empirical Results on credit risk factors

Chapter 3 deals with the empirical identification and quantification of the impact of risk factors as well as with the predictive ability of these factors. We use three classification methods, linear discriminant analysis, quadratic discriminant analysis and logistic regression. In this chapter we start with a comprehensive review of the classification methods and the assumptions on which they are based. Linear classification rules are derived under the assumption of normality and equal covariance matrices in the two samples. Quadratic classification rules are derived when the assumption of equal covariance matrices is relaxed.

The empirical analysis based on the discriminant method is starts with the application of the linear discriminant analysis. There are two general approaches to identifying discriminating variables in the linear discriminant analysis based on whether the objective of the research is to classify or to predict group membership. Unfortunately the two methods tend to generate different sets of risk factors and thus make the selection of the factors difficult. We have therefore used a criterion that incorporates both requirements as our modelling strategy. The results from the linear discriminant analysis support the view that there is a set of risk factors in the form of financial characteristics of the company which predict which companies will experience financial distress within two years. These risk factors include profitability, leverage, solvency and liquidity ratios.

Next we relax the assumption of equal covariance matrices, which is rejected by the data as the statistical analysis of Chapter 2 shows, and classify companies in the two groups on the basis of quadratic classification functions. The empirical results do not show an improvement, over the simpler linear discriminant analysis.

The third approach we have employed to estimate the credit risk factors of a company is the logistic regression approach. This approach models the posterior probability directly and does not rely on either the assumption of normality or the assumption of equal covariance matrices in the two groups. The empirical data suggest that there is a set of risk factors that can be used to identify companies at least two tears before they were downgraded. These factors tend to be stable across all the years of the samples producing stable probabilities of being downgraded. Four variables in particular, the return on assets, working capital over assets, interest payments over sales and debt over total assets are capable of predicting correctly on average 70 percent of the companies that were eventually downgraded.

Chapter 4: Measuring the Market Risk for Equities: A Review of Alternative Approaches and Some Empirical Tests of their Assumptions.

Chapter 4 is a comprehensive attempt to deal with the issues surrounding the measurement of the market risk of individual equities. In this chapter we present in a unified statistical framework the various approaches to measuring the market risk of equities such as moment – based and quantile –based measures. We analyse the advantages and disadvantages of risk measures such as standard deviation, lower partial moment, Value-at-Risk and expected shortfall from a number of perspectives such as portfolio and time aggregation, consistency with the implicit distributional assumptions on equity returns and consistency with the theoretical properties of risk measures such as coherence.

We pay particular attention to the assumptions on which these approaches are based, their properties and their performance by reviewing the large body of empirical literature that has been using these approaches.

Particular emphasis is based on the parametric approach because Value at Risk (VaR) is based on both distributional assumptions and estimates of volatility and its empirical performance depends therefore on the validity of the assumption of the probability distribution (e.g. normally distributed returns) as well as on the estimate of the volatility.

The original approach to Value-at-Risk as introduced in RiskMetrics was based on normally distributed returns and a particular model of volatility, the exponentially weighted model. Given the inconsistency of the assumption of normally distributed returns with the actual data due primarily to the existence of fatter tails in the empirical density functions than the normal one, a number of alternative distributions have been suggested. These alternative distributions include the class of stable distributions which exhibit heavy tails, as well as the Student-T distribution and its generalisations, the generalised beta distribution, the inverse Gaussian and other which are reviewed in the chapter. An alternative way of introducing kurtosis is through the class of ARCH and GARCH volatility models which are tremendously useful since not only introduce fat tails, but they also allow a far more flexible way of modelling and forecasting volatility. The empirical performance of the class of GARCH models is also reviewed.

The assumptions on which risk models are based are tested using a sample of 132 stock returns from companies quoted on the Athens Stock Exchange. We test for normality which is rejected for all the companies. We also test for serial correlation in returns and heteroskedasticity which all the returns exhibit. These findings cast doubt on the simple rules of risk aggregation over time.

We have also fitted to the data two alternative probability distributions to normal namely the generalised skewed and the family of stable distributions. The generalised skewed distribution was the dominant distribution. Finally we have estimated the lower tails of the distribution using extreme value theory. We have found that all the companies possess either semi-or heavy tails.

Chapter 5: Evaluation of the Forecasting Performance of Alternative Value-at-Risk Models.

The purpose of this chapter is to evaluate the forecasting performance of the various approaches to estimating VaR. The approach we adopt is to evaluate the out-of–sample performance of the various models as the primary purpose of a VaR model is to forecast future losses.

The chapter starts with the development of a framework for the evaluation of forecast performance. Kupiec (1995) suggested a statistic for evaluating the performance of a model based on the proportion of prediction failure. Evaluation based on the proportion of correct forecasts ignores conditioning or time variation in the data which may introduce serial correlation in the prediction failure. Christoffersen (1998) developed a test that takes this aspect into account when a model is evaluated and this is the test we have used in this chapter.

Next we review the data we have used to compare the performance of the alternative models. More specifically we explore the statistical properties of returns of the Athens Stock Exchange General Price Index. We test for normality and compare three distributions in terms of quantile prediction, namely the normal, the Laplace and the Student-Distribution. Estimation of the density function is useful on its own right but it also gives us insights as to the right density for the error term when we model the dynamics of the returns.

In Section 3 we estimate the tails of the distribution using the Pareto model paying particular emphasis to the specification of the correct threshold for the characterisation of the extreme observations.

In Section 4 we tackle the issue of heteroskedasticity in returns by modelling the volatility dynamics through a GARCH model. The implications of applying extreme value theory to the heteroskedasticity - corrected series are further explored by calculating the Value-at-Risk for the standardised residuals.

In Section 5 we present the specification of the alternative models that we compare. The approach we have adopted is to estimate each model over a certain sample and then to forecast the next day VaR producing a total of 2000 forecasts. In total we compare eight different models which are differentiated in terms of their distributional assumptions and in terms of their conditioning methods. The first group of models assumes that the parameters remain fixed within the sample on which estimation is conditioned. In the second group of models we assume that volatility varies over time and is not fixed within the sample.

Finally in Section 6 we present the results of the comparison using the metrics we have developed. The results show an overwhelming support for the extreme value approach and the models which show allow for time-varying volatility. The standard normal RiskMetrics models, as well as models with Cornish Fisher corrections, are not found to be adequate equity risk measurement models.

Contribution of the Thesis

This thesis makes a contribution to the literature on the measurement of credit and market risk in a number of different areas. These contributions are both practical and methodological.

1. This study is one of only a handful of studies that examine the issue of credit risk assessment and identification of risk factors in Greece and this thesis makes a practical contribution to our understanding of what determines the credit risk of Greek companies. To the best of our knowledge there have been only a small number of studies using Greek data. These studies include the study by Gloubos and Grammatikos we have already reviewed, the study by Michalopoulos which is based on expert systems and the study by Dimitras and Zoumbounides that is based on neural networks. Our study, thanks to the availability of a better data set allows to pursue a more systematic study of the characteristic of companies and to examine a wider set of issues such as the stability of the risk factors and the discriminant and logistic equations across number of years.
2. The second contribution is methodological and relates to both the way we have defined and measured credit risk as well as the way we have estimated the discriminant and logistic equations. Rather than measuring the probability of default we have found a stable set of financial ratios that predict which companies will be downgraded not just those companies that will go bankrupt. This is a more meaningful measure of credit risk, since consistent with the international experience the number of companies that actually go bankrupt is very small. For example, Altman (1968), Deakin (1972), Altman, Haldeman and Narayan (1977), Dambolena and Khoury (1980), Hamer (1983) and Hillegeist, Keating, Cram and Lundstedt, (2002) rely on samples of bankrupt firms numbering only 33, 32, 53, 23, 44 and 30 companies respectively.

   This thesis also makes a contribution in terms of the statistical methodology we have employed. In most previous studies on bankruptcy prediction the sample of bankrupt firms is drawn from a number of years not just from one year. Then this sample is matched by a sample of surviving firms matched by year and normally size. However such a procedure is not optimal from a statistical point of view.

   Our approach of estimating discriminant functions and logistic equations for every year instead of pulling all the data as in previous studies, allows us to test the stability of the estimated coefficient over time and to test whether we can get stable relationships which can be used in the future. Our methodology is better in the sense that whereas all the other studies have used bankruptcy data over a different period ignoring therefore cyclical effects, we have used data on downgraded firms from a single year. Our approach is therefore more informative in the sense of testing the model to different macro conditions at different phases of the business cycle. We do not have to resort to more complicated econometric procedures such as hazard models to avoid econometric issues that the traditional approaches entail. .
3. The third contribution is the empirical results themselves which show that both models whether they are based on discriminant analysis or logistic regression predict reasonably well which of the companies may be downgraded. In our research we have adopted a

stricter criterion, because we are predicting the downgrading of a company over a two –year horizon rather than over a one year horizon. The available empirical evidence shows that the predictive power deteriorates as the credit horizon becomes longer. Also estimating the probability of company being downgraded is more difficult than estimating the probability of a company going bankrupt because in the latter case the companies tend to more homogenous.

4. The fourth contribution of this research is to show that the ratios identified by Altman and other researchers which were reviewed in early parts of this chapter are indeed important in the Greek context as well. That means that international comparisons and comparable credit rating is possible.
5. The fifth contribution of this thesis is that the financial ratios we have used remain reasonably stable in their values over the sample period making the model much more reliable and easy to use. The cut-off points for instance are stable.
6. Credit risk is the probability that a company may not be able to repay its debt. This can happen without the company going bankrupt. Our model is therefore closer to the spirit of measuring credit risk rather than the bankruptcy studies.

There are a number of implications from our study both for financial institutions as well as for policy makers. For financial institution we have provided a reliable tool to assess the probability of a company becoming financially distressed and consequently to calculate the potential risk of a borrower and adjust its loan pricing policy. Our results are particular useful for financial institutions in view of the Basle II provisions that allow banks to use their own internal rating systems. For policy makers, our results give early warning signs of the performance of he corporate sector and the thus afford the opportunity for policy measures to be taken.
With regard to our understanding of the market risk of Greek equities we believe that this thesis also makes a number of contributions.

1. The study analyses in a systematic way the issues surrounding the implementation of risk measurement methodologies in a specific market environment. The thesis provides a systematic attempt to understand the statistical properties of individual company returns. First it estimates alternative probability distribution of returns for individual stock returns. The question of whether normality is an appropriate assumption is central to financial theory and practice and we have shown that the overwhelming majority of stock returns are not normally distributed. Secondly it estimates the tail characteristics for stock returns. This latter effort represents the first systematic attempt to model the tail-characteristics of Greek companies. Earlier studies, [see e.g. Dinenis and Priniotakis (2002), (2003)] have examined risk modelling issues for indices only. The estimation of the risk characteristics for individual stocks may lead to different portfolio construction than the one predicted from standard mean variance analysis. Thirdly we have estimated volatility models for individual

companies. These results help us decide not only which the appropriate volatility model is but also whether risk aggregation techniques are legitimate.

2. The second major contribution, beyond increasing our understanding of an under-researched market, is the evaluation of the performance of alternative models for the forecast of value-at-risk. The empirical evidence available to the Greek risk manager and investor comes from European and American stock markets, which may not be appropriate when risk assessment in the Greek market is the objective. By analysing the performance of these models in the specific market environment of the Greek stock market we believe that we have provided results that should be of help to risk managers and investors in deciding which method to employ.

Chapter 1 - Review of the literature on credit risk factors identification and measurement

1.1 Introduction

The financial stability of a company is of great concern to its stakeholders (employees, investors, lenders, government) and this importance is reflected both in the vast literature on measuring credit risk as well as in the recent proposals on assessment of credit risk (BIS 2003) for financial institutions. The credit risk of a borrower is defined as the potential that a borrower will fail to meet its obligations in accordance with agreed terms. Most of the literature on credit risk measurement has traditionally concentrated on the credit risk of corporate bonds, analyzing risk factors which are important for investors. However for banks, and other financial intermediaries, whose financial stability affect the entire system, loans are the largest and most obvious source of credit risk. In this chapter we concentrate on a review of the literature on the measurement of credit risk for loans and try to identify consistent risk factors that may explain the inability of companies to service their debts. The purpose of Chapters 2 and 3 is to describe and analyse the statistical properties of company characteristics which may help us identify which companies are more likely to become financially distressed.

This chapter is organised as follows: First we review the various approaches to credit risk measurement in the framework we have developed above. For each approach we examine the assumptions on which it is based and how credit risk is assessed, that is whether it is the probability of migration or the probability of default. The advantages and disadvantages of each method are critically examined and their empirical relevance assessed. We start first with the empirical approach to determining the probability of migration and default and then review three of the latest approaches to modelling credit risk, CreditMetrics developed by JP Morgan, CreditPortfolio View developed by KPMG and CreditPortfolio+ developed by McKinsey. We then present the theoretical model of estimating default probabilities developed by Merton (1974) and its most famous variant the KMV model. Then we examine a number of statistical approaches to modelling credit risk. These statistical approaches are based either on techniques of predicting whether companies will go bankrupt or not, or on estimating the probability of default. Models in the first category include the univariate approach, multivariate discriminant analysis, classification trees and neural networks. These models deal with classification of companies into two groups and do not provide an estimate of credit risk in any of the forms we discussed above, or the probabilities of migration and default. However credit risk indices can be calculated which can map the risk index to the probability of default. In the second category we have the Logit and Probit models, the linear probability model, the survival and hazard models. These models provide a direct estimate of the probability of default. The second part of this chapter explains the research design and our approach to identifying credit risk factors.

1.2 Fundamental Concepts

Quantification of credit risk is not unique. Some models concentrate on estimating the probability of default. Other models try to estimate the probability of upgrading or downgrading of the credit quality of a borrower and the impact on the value of the loan or the portfolio of loans. A third category of models estimates the probability distribution of losses. All these approaches are of course compatible as the credit risk of an individual loan or bond (see e.g. Ong 2000 or Dinenis 2002b for a fuller review of these points) has three components:

- The probability of default. The probability that the obligor or counterparty will default on its contractual obligations to repay its debt.
- The recovery rate. The extent to which the face value of an obligation can be recovered once the obligor has defaulted.
- Credit migration. Short of a default, the extent to which the credit quality of the obligor has defaulted.

In addition in a portfolio context, there are two other elements, namely default and credit quality correlation and risk contribution and credit concentration.

- Default and credit quality correlation. The degree to which the default or credit quality of one obligor is related to the default or credit quality of another.
- Risk contribution and credit concentration. The extent to which an individual instrument or the presence of an obligor in the portfolio contributes to the totality of risk in the overall portfolio.

All the risk elements can be expressed by the expected loss, the unexpected loss, the maximum potential loss and economic capital on a loan, which are defined as follows:

- Expected loss. The expected loss E(L) shows the amount of money the lender should expect to lose on average each year,

$$E(L) = (1 - \delta)pF_T$$

where

p is the probability of default

F_T is the face value of the loan at maturity.

δ is the recovery rate.

- Unexpected loss. The unexpected loss on the other hand is given by

$$UL = F_T \times (1 - \delta)\sqrt{p \times (1 - p)}$$

which is the standard deviation of the loss distribution.

- Maximum possible loss. The Maximum Possible Loss (MPL) is a confidence level such that there is only a small probability (p) that losses could be worse than MPL. The required value would depend on the probability of default. For example for a single A - rated bank the probability of default p is around 0.1% and the MPL is the level of the loss that is so bad that there is a one-in –one thousand chance of the loss being greater than the MPL.
- Economic Capital. The economic capital (EC) the amount of capital that the lender has to hold so as to meet the Maximum Possible Loss. It is defined as

  $$EC = MPL - E(L)$$

- The Value at Risk (VaR). The VaR of an individual bond, loan or of a portfolio of bonds and loans is closely related to the concept of economic capital. VaR is another way of expressing the unexpected loss of a portfolio and can be expressed as a multiple of standard deviations from the expected value. Unlike the previous definition of unexpected loss, VaR takes into account both the upside and downside move in the credit quality of an obligor. Thus VaR departs from the previous approaches by looking at the distribution of values rather than at the distribution of losses.

The value at risk of a position is the difference between the expected value of the position and the value corresponding to some confidence interval say 99% or 95% (minimum possible value).

VaR = E (V) - MPV

All the above concepts are explained in the Figures 1.1 and 1.2.

Figure 1.1: Credit Risk Concepts

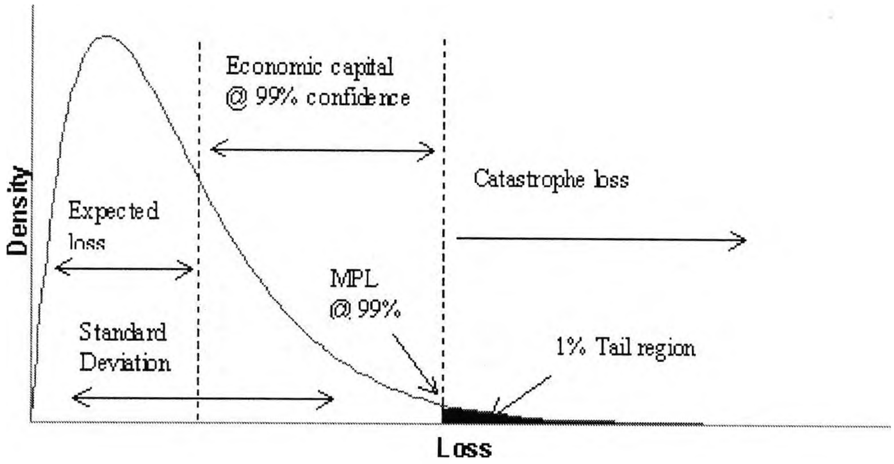
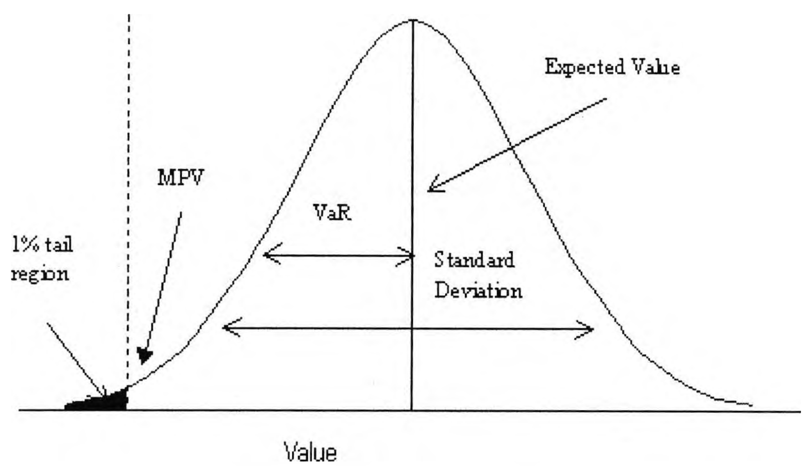


Figure 1.2: Value at Risk Definition



A third approach to measuring credit risk is by constructing a credit scoring index. This has been the main approach adopted by banks over a number of years. The index is constructed either on the basis of quantitative and qualitative information on companies or on the basis of subjective judgments by credit officers. The index takes the form of a numerical scale, with companies which have a high score deemed as good in terms of credit quality and companies with a low score deemed as of low credit quality [see Tamari (1984), Altman and Saunders (1998) or Dinenis (2003b)] for examples of such indices.

The measurement of credit risk is equated with the migration of a certain debt or loan to another credit category. In markets where loans or corporate bonds are traded this migration is reflected in increased spreads. However most of the empirical studies in the accounting and finance literature are invariably concentrate on modeling the probability of default rather than the probability of migration from one category to another. In this case credit risk is determined by the probability of default. Thus a company that has a very low or even zero probability of default but a high probability of being downgraded will have its credit risk significantly underestimated. Yet with the recent changes in the regulatory capital (BIS 2000) it is important for a bank to make the distinction between the probability of default and the probability of being downgraded.

The various approaches or models that have been employed have been classified in a rather confusing manner as modern, traditional, expert systems, statistical and structural, reduced form, marked to market or default mode. In this review we examine the various approaches and classify them in terms of the technique they have employed. The two general groups are those, which deal with credit risk at the level of the portfolio, and those that deal with credit risk at the level of the firm. Although some of the portfolio approaches have developed their own method of dealing with individual risk measurement, the two groups are compatible. Since the purpose of this part of the thesis is the identification of individual risk factors, we pay particular attention to the measurement of risk at firm level. There are two general approaches to measuring the credit risk of an individual company, the statistical approach and the structural model (see e.g. Moody's 2000). The statistical approach is the oldest one of the two approaches and is reviewed extensively in Jones (1987) and more recently in Dimitras, Zanakis and Zopunidis (1996), Saunders (1999) and Caoutte, Altman and Narayanan (1998). The structural approach is based on Merton's model of default debt (1974) and has been the basis for the portfolio approaches to credit risk, measurement such as KMV, and CreditMetrics. The latter approach is not as popular because it requires data that are not readily available.

1.3 Empirical Method

The most widely used way of estimating default probabilities is from historical data on defaults of companies that have issued corporate bonds. Credit rating companies such Standard & Poor, Moody's Investor Services of Fitch collect data on defaults and publish regular migration tables presenting the average probability of a company in a given credit category being upgraded, downgraded of defaulting. Table 1 shows an example of a one-year migration table prepared by Standard and Poor and which is based on the migration behaviour of bonds for a period of 60 years. It is common to use migration data from bonds to estimate the migration behaviour of loans but the two asset classes may not behave in exactly the same way.

Table 1.1: Example of a one-year transition matrix.

Initial Rating	**Rating at year-end**							
	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	90.81%	8.33%	0.68%	0.06%	0.12%	0.00%	0.00%	0.00%
AA	0.70%	90.65%	7.79%	0.64%	0.06%	0.14%	0.02%	0.00%
A	0.09%	2.27%	91.05%	5.52%	0.74%	0.26%	0.01%	0.06%
BBB	0.02%	0.33%	5.95%	86.93%	5.30%	1.17%	0.12%	0.18%
BB	0.03%	0.14%	0.67%	7.73%	80.53%	8.84%	1.00%	1.06%
B	0.00%	0.11%	0.24%	0.43%	6.48%	83.46%	4.07%	5.20%
CCC	0.22%	0.00%	0.22%	1.30%	2.38%	11.24%	64.86%	19.79%

Source:[Standard & Poor (1997)]

Each row corresponds to an initial rating and each column corresponds to a year-end rating. To find the probability that a bond rated A today will be rated BBB one year from today, we read across the A row until we get to the BBB column and find the probability is 5.52%. Each row sums to 100% because a bond has to end the year in one of the eight column categories.5 Notice that the largest probabilities are on the diagonal, indicating that most ratings do not change in the course of a year. Notice also that some transitions probabilities are 0 (at least to several decimal places). This indicates, for example, that the chance of a AAA-rated bond defaulting within one year is negligible.

Table 1.2: Standard and Poor cumulative default probabilities

Initial Rating	**Years**							
Term	1	2	3	4	5	7	10	15
AAA	0.00%	0.00	0.07	0.15	0.24	0.66	1.40	1.40
AA	0.00%	0.02	0.12	0.25	0.43	0.89	1.29	1.48
A	0.06%	0.16	0.27	0.44	0.67	1.12	2.17	3.00
BBB	0.18%	0.44	0.72	1.27	1.78	2.99	4.34	4.70
BB	1.06%	3.48	6.12	8.68	10.97	14.46	17.73	19.91
B	5.20%	11.00	15.95	19.40	21.88	25.14	29.02	30.65
CCC	19.79%	26.92	31.63	35.97	40.15	42.64	45.10	45.10

In Table 2 the one year –horizon is extended to show the probability of default over a longer horizon. Although the probability of default for a AAA company is practically zero over one period, it becomes 1.40% over a ten year period. Similar increases are apparent for the other categories. Similarly for a B rated company, the default probability increases from 5.20% over one year to 19.40 over a period of 4 years and to 30.65% over a period of fifteen years.

A similar empirical method was employed by Moody's Investor Services (1997) covering the period 1920 to 1996 and using the credit histories of more than 14,000 US and non-US corporate debt issuers. The 77-year time frame decidedly allows comparison of rating change patterns over a variety of business, interest rate and economic cycles.

Table 1.3: Moody's one year migration matrix average values 1970-2001

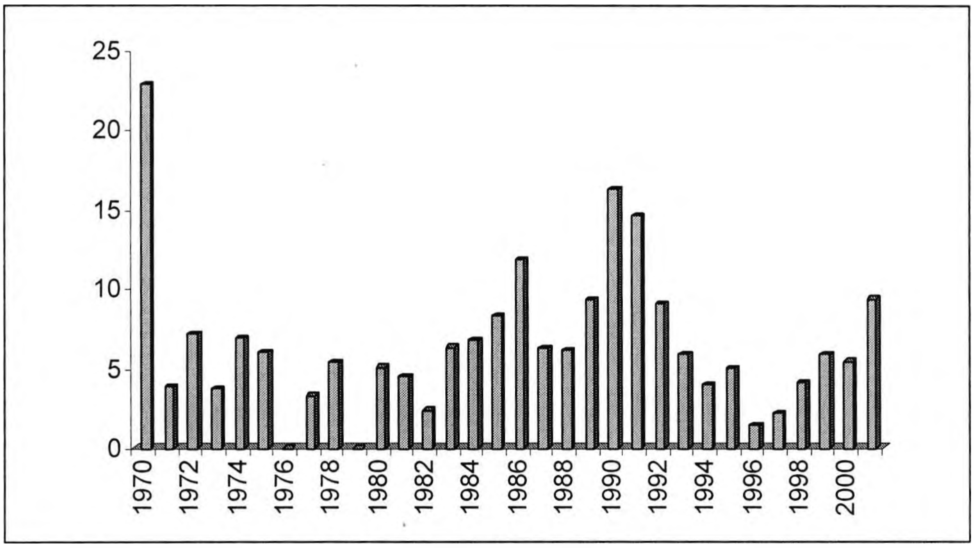
Initial Rating	**Rating at year-end**							
	Aaa	Aa	A	Baa	Ba	B	Caa-C	Default
Aaa	89.09	7.15	0.79	0.00	0.02	0.00	0.00	0.00
Aa	1.17	88.00	7.44	0.27	0.08	0.01	0.00	0.02
A	0.05	2.41	89.01	4.68	0.49	0.12	0.01	0.01
Baa	0.05	0.25	5.20	84.55	4.51	0.69	0.09	0.15
Ba	0.02	0.04	0.47	5.17	79.35	6.23	0.42	1.19
B	0.01	0.02	0.13	0.38	6.24	77.82	2.40	6.34
Caa-C	0.00	0.00	0.00	0.57	1.47	3.81	62.90	23.69

Source: Moody's Investors Service (2002): Default & Recovery Rates of Corporate Bond Issuers, February.

One of the more important criticisms of the empirical approach employed by the rating agencies to determine default and transition probabilities is the apparently static nature of the resulting average historical probabilities.

In reality, actual transition and default probabilities are very dynamic and can vary quite substantially over the years, depending on general economic conditions and business cycles. As an example of how default rates fluctuate over time, we have plotted the annual default rates for Moody's B –rated companies in Figure 3. The figure of the mean default value of 6.3 percent can be quite misleading when the standard deviation of 4.65 is taken into account.

Figure 1.3: Annual Default Rates for a Moody's B rated company.



This issue is particularly critical if the analysis horizon is rather long. The empirical relevance of this criticism can be seen from the following table which shows the average default probabilities and its standard deviation by rating category for a portfolio of obligors rated by Moodys during the period 1970 to 1995.

Table 1.4: Statistics of default rates

	One-Year Default Rate	
Credit Rating	Average (%)	Standard Deviation (%)
Aaa	0.00	0.0
Aa	0.03	0.1
A	0.01	0.0
Baa	0.13	0.3
Ba	1.42	1.3
B	7.62	5.1

Source: Carty and Lieberman (1996)

Comparing the above results to those of Table 1.3 and Graph 1.£ we see that not only the default probabilities are not fixed over time but that the standard deviation also changes. We see that the standard deviation of default rates was 4.65 for a B company for the period 1970-2001 but 5.1 for the period 1970-1995.

A survey of the internal rating systems of 18 major bank holding companies suggested that as much as 60 percent of their collective loan portfolios may be below the equivalent investment grade [Treacy and Carey (1998)] and the default rates on low –quality credits (including junk bonds) are highly sensitive to the state of the business cycle. Moreover there is empirical evidence that rating transitions in general may depend on the state of the economy [see Nickell, Perraudin and Varotto (1998) and Wilson (1997a, 1997b)]. This evidence suggests that the probability of downgrades and defaults may be significantly greater in a cyclical downturn than in an upturn and thus the historical average probabilities should be adjusted. A number of models have been developed that link the probability of default to macroeconomic factors.

A related problem is the differences between the transition and default probabilities produced by the main risk rating companies. Looking across the diagonals of Tables 1 and Table 3 and comparing the probability of a company remaining in the same credit category we see that this probability is always lower for the bonds in the Moody's sample than for the firms in the Standard & Poor's sample.

1.4 CreditMetrics

CreditMetrics is the first of the portfolio approaches to credit risk measurement. It is reviewed first because it is a complete approach to the measurement of all the elements of credit risk. CreditMetrics rather than concentrating on estimating default probabilities, attempts to measure the risk resulting from ratings transition. The key characteristics of the approach are (a) it takes into account both upside and downside moves in credit quality; (b) it requires the distribution of estimates of future loan values in calculating a capital requirement on a loan.

Credit Metrics analysis is based on a ratings transition matrix giving the probabilities of rating changes over a period of for example., one year. Table 1 is an example of such a matrix. The credit matrix approach calculates the values of a loan or bond at the end of a specified period and then calculates the risk associated with the bond as the standard deviation of the future bond values.

Let $v_i, \ldots, v_m$ represent the values of a bond at the end of a year and let $p_1, \ldots, p_m$ be the associated probabilities. Table 5 shows the values and the associated probabilities for a BBB bond.

Table 1.5: Values of bonds for different credit ratings

Year-end Rating	Probability Of State	New Bond Value
AAA	0.02%	$ 109.37
AA	0.33%	$ 109.19
A	5.95%	$ 108.66
BBB	86.93%	$ 107.55
BB	5.30%	$ 102.02
B	1.17%	$ 98.10
CCC	0.12%	$ 83.64
Default	0.18%	$ 51.13

The value of the bond at each state is calculated from

$$v_i = \sum_{k=1}^{T} \frac{C}{(1 + r_k + s_i)^k} \qquad i = 1,\ldots,m$$

where

C = the coupon of the bond

r_k = the k-maturity spot rate

s_i = the credit spread associated with a bond in credit rating state i.

The following example again taken from the CreditMetrics document shows how this valuation is done. For the BBB bond of our example we use the values from the Creditmetrics document. Suppose the bond has a face value of $100 and pays an annual coupon of 6%. Using the discount factors from the Table 6 for the BBB bond, we calculate the value of the bond at the end of a year if it remained rated BBB using

$$v_{BBB} = 6 + \frac{6}{1.041} + \frac{6}{(1.0467)^2} + \frac{6}{(1.0525)^3} + \frac{106}{(1.0563)^4} = 107.55$$

If it is upgraded to A, it will be worth $108.66.

$$v_A = 6 + \frac{6}{1.0372} + \frac{6}{(1.0432)^2} + \frac{6}{(1.0493)^3} + \frac{106}{(1.0532)^4} = 108.66$$

If it is downgraded to CCC, it will be worth .$83.64

$$v_{CCC} = 6 + \frac{6}{1.1505} + \frac{6}{(1.1502)^2} + \frac{6}{(1.1403)^3} + \frac{106}{(1.1352)^4} = 83.64$$

Table 1.6 : Forward rates by credit category

Category	Year 1	Year 2	Year 3	Year 4
AAA	3.60	4.17	4.73	5.12
AA	3.65	4.22	4.78	5.17
A	3.72	4.32	4.93	5.32
BBB	4.10	4.67	5.25	5.63
BB	5.55	6.02	6.78	7.27
B	6.05	7.02	8.03	8.52
CCC	15.05	15.02	14.03	13.52

The results from the valuation are shown in Table 7. According to the table, in case of default the value will be $51.13; This latter is the amount bondholders can expect to recover in case of bankruptcy — roughly fifty cents on the dollar. In practice, not all bonds with the same credit rating will have the same value (even if they have the same coupon and maturity). The values in the Table 7 may be thought of as average values within each category. The CreditMetrics document explains how to include information about the standard deviation of bond values within each category; however, for simplicity, we will ignore this issue and pretend that the possible bond values are exactly as given in the table.

We can now use the table of possible bond values to measure risk. Two measures of risk are discussed in the CreditMetrics document: standard deviation and a percentile or value-at-risk measure. Although CreditMetrics considers risk at a portfolio level, the framework can be sued to calculate the risk of a single asset. We consider the same BBB bond whose possible values one year from today are given in Table 7. The expected value and the standard deviation are given by

$$E(v) = \bar{v} = \sum_{i=1}^{m} v_i p_i \quad i = 1,\ldots,m$$

and

$$Var(v) = \sigma_u = \sum_{i=1}^{m} (v_i - \overline{v})^2 p_i \quad i = 1,...,m$$

Table 1.7 shows the calculations for the expected value and the standard deviation of the bond which are 107.09 and 2.99 respectively. From this table, we can also immediately determine the second measure of risk proposed by CreditMetrics a percentile or value-at-risk measure. Suppose we want to find a dollar amount such that the probability that the bond will be worth that amount or less is at most 5%. From the table we find that there is a 0.18% chance that the bond will be worth 51.13; there is a 0.30% (=0.18+0.12) chance that the bond will be worth 83.64 or less; there is a 1.47% (=0.18+0.12+1.17) chance that the bond will be worth 98.10 or less; and there is a 6.77% that the bond will be worth 102.02 or less. So, we can say that the probability the bond will be worth 98.10 or less is below the 5% limit, but we cannot say that about 102.02. To hit 5% exactly, we may interpolate to find a value that correspond s to 5% equal to 100.71[1].

Table 1.7: Calculation of values and standard deviation.

Standard deviation calculation for bond initially rated BBB						
	Probability	**New**	**Probability**	**Difference**	**Probability**	
Year-end	**Of**	**Bond**	**Weighted**	**of value**	**Weighted difference**	
Rating	**State**	**Value**	**Value**	**From mean**	**Squared**	
AAA	0.02%	109.37	0.02	2.28	0.0010	
AA	0.33%	109.19	0.36	2.10	0.0146	
A	5.95%	108.66	6.47	1.57	0.1471	
BBB	86.93%	107.55	93.49	0.46	0.1856	
BB	5.30%	102.02	5.41	(5.07)	1.3612	
B	1.17%	98.10	1.15	(8.99)	0.9452	
CCC	0.12%	83.64	0.10	(23.45)	0.6598	
Default	0.18%	51.13	0.09	(55.96)	5.6363	
	Mean=	107.09	107.09		Variance=	8.9508
				Standard deviation =		2.99

To summarize, the expected value of the bond is $107.09 and its standard deviation is $2.99. We estimated that there is a 5% chance that the bond will be worth $100.71 or less. Notice that the difference 107.09 –100.71 is 6.38 and 6 .38/2.99 = 2.1 .This indicates that 100.71 is about 2.1 standard deviations below the mean. The fact that there is approximately a 5% probability that the value will be more than 2.1 standard deviations below the mean reflects the skew in the distribution of value;

We can compare the risk measures provided by risk metrics with the default mode model as a further check of its robustness. The expected loss in the default mode is given by $E(L) = (1-\delta)pF_T$

Using the following values derived from our example.

$p = 0.0018$

$F_T = 100{,}000{,}000$

$\delta = 0.5113$

the expected loss is

$$E(L) = (1-\delta)pF_T = 0.0018 \times 0.4887 \times 100{,}000{,}000 = 87{,}966$$

The unexpected loss on the other hand is given by

$$UL = F_T \times (1-\delta)\sqrt{p \times (1-p)} = 100{,}000{,}000 \times 0.4887 \times \sqrt{0.0018 \times 0.9982}$$
$$= 2{,}071{,}511$$

The unexpected loss in CreditMetric is

$$UL_{creditmetrics} = \sigma F_T = 0.0299 \times \times 100{,}000{,}000 = 2{,}990{,}000$$

The difference in the two values of the unexpected loss is due to the fact that CreditMetrics is a mark-to-market approach and allows an upside as well a downside to a loan's value. The default mode approach on the other hand fixes the maximum upside to the face value of the bond which in the example we have used is $100 million.

The economic capital under the Default Mode approach is more closely related to book value accounting concepts than to the market value accounting concepts used in the MTM approach.

Although Creditmetrics represented a significant breakthrough in term of managing credit risks it suffers from a number of problems.

1. The calculation of transition matrices based on averaging one-year transitions over a past data period assumes that transition probabilities follows a stable Markov process, which means that the probability of a loan moving to any particular state during this period is independent of any outcome in the past period. There is evidence however that rating transition are autocorrelated over time. For example a loan that was downgraded in the previous period has a higher probability (compared to another loan that was not downgraded) of being downgraded in the current period. [Nickell, Perraudin and Varotto (2001)]. This suggests that a second or higher order Markov process may better describe rating transition over time.
2. A second serious weaknesses of the Credit Metrics approach is that it ignores the fact that migration and default probabilities vary with the cycle. We have already presented evidence in Section 1, that the probability of default varies considerably. One way to deal with cyclical factors in the CreditMetrics framework is to dived the sample period into recession years and non recession years and calculate two separate historic transition matrices (a recession matrix and non-recession matrix) to yield two separate VAR calculations. [See Saunders (2000)].
3. A third problem is the portfolio of bonds used in calculating the transition matrix. Altman and Kishore (1997) found a noticeable impact of bond "aging" on the probabilities calculated in the transition matrix. Indeed a material difference is noted depending on whether the bond sample used to calculate transitions is based on new bonds or on all bonds outstanding in a rating class at a particular moment in time.
4. A final problem with Creditmetrics is using bond transition matrices to value loans. There are many differences between the two instruments such as collateral covenants that make the two instruments quite different. Using bond transition matrices may therefore result in biased valuations. The internal models developed by banks based on loan transition matrices may offer more reliable inputs for the calculation of the VaR measures of loan risk.

---

1 The value is calculated as follows $100.71 = (5-1.47) \times \frac{102.02-98.10}{6.77-1.47} + 98.10$

1.5 CreditPortfolioView

McKinsey proposed this model and it is a disrcetised multi-period econometric model that measures only default risk. Default probabilities depend on macroeconomic variables and it therefore allows default probabilities to vary over the cycle contrary to the CreditMetrics approach. Directly modelling the relationship between transition probabilities and macro factors is a second way of dealing with cyclical factors and effects. [See Saunders (2000)].

The starting point in this approach is also the transition matrix which is reproduced here again for convenience.

Initial Rating	**Rating at year-end**							
	AAA	AA	A	BBB	BB	B	CCC	Default
AAA	90.81%	8.33%	0.68%	0.06%	0.12%	0.00%	0.00%	0.00%
AA	0.70%	90.65%	7.79%	0.64%	0.06%	0.14%	0.02%	0.00%
A	0.09%	2.27%	91.05%	5.52%	0.74%	0.26%	0.01%	0.06%
BBB	0.02%	0.33%	5.95%	86.93%	5.30%	1.17%	0.12%	0.18%
BB	0.03%	0.14%	0.67%	7.73%	80.53%	8.84%	1.00%	1.06%
B	0.00%	0.11%	0.24%	0.43%	6.48%	83.46%	4.07%	5.20%
CCC	0.22%	0.00%	0.22%	1.30%	2.38%	11.24%	64.86%	19.79%

If we look this time to a CCC-rated borrower, the estimated probability that it will default over the next year is 19.79%. It is reasonable to assume that this probability will change over the economic cycle and to be higher in recessions than in expansions. Because the probabilities in each row of the transition matrix must sum to 1, an increase in the probability of default must be compensated for by a decrease in other probabilities-for example those involving upgrades of initially CCC-rated debt. Let P_{jt} denote the default probability of bond in credit category j, at time t. The CreditPortfolioView approach assumes that this probability will vary at time t along with a set of macro factors indexed by variable Q.

The default probabilities are assumed to be generated by a logit function that relates them to a "country speculative-grade specific index" which is itself related to country

$$P_{jt} = \frac{1}{1+\exp[-Q_{jt}]} \qquad (1.4.1)$$

where

$$Q_{jt} = \alpha_j + \boldsymbol{\beta}'\mathbf{X}_t + u_{jt} \qquad (1.4.2)$$

$$u_{jt} \sim N(0, \sigma_j)$$

The macro-variables are specified for each country and, when enough data are available, the vector $\boldsymbol{\beta}$ can be consistently calibrated. Moreover in the implementation proposed by McKinsey, each of the independent variables is assumed to follow an autoregressive model of order 2 (AR(2)), such that the process $\mathbf{X}_t$ has some memory. It can be written as

$$\mathbf{X}_t = \mathbf{a} + \mathbf{B}\mathbf{X}_{t-1} + \boldsymbol{\Gamma}\mathbf{X}_{t-2} + \boldsymbol{\varepsilon}_t$$

Substituting the above equation into equation (1.4.1) yields [2]

$$Q_{jt} = \alpha_j + \boldsymbol{\beta}'[\mathbf{a} + \mathbf{B}\mathbf{X}_{t-1} + \boldsymbol{\Gamma}\mathbf{X}_{t-2} + \boldsymbol{\varepsilon}_t] + u_{jt} \qquad (1.4.3)$$

[2] In empirical applications the coefficient matrices $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are normally assumed to be diagonal.

Then substituting (1.4.3) into equation (1.4.1) the probability of a grade j bond/loan moving to default category during the next year will be determined by:

$$P_{jt} = \frac{1}{1+\exp\{-\alpha_j - \beta'[\mathbf{a} + \mathbf{B}\mathbf{X}_{t-1} + \mathbf{\Gamma}\mathbf{X}_{t-2} + \varepsilon_t] + u_{jt}\}} \tag{1.4.5}$$

Equation (1.4.5) means that the determinants of the transition probability is a function of lagged macro variables, a general economic shock factor u_{jt} and shock factors or innovations for each of the macro variables ε_t. Because $\{\mathbf{X}_{t-1}, \mathbf{X}_{t-2}\}$ are predetermined, the key variables driving P_{jt} will be the innovations or shocks u_{jt} and ε_t. Using structured Monte Carlo simulation approach, values for u_{jt} and ε_t can be generated for periods in the future that occur with the same probability as that observed from history. We can use the simulated u_{jt} and ε_t along with the fitted macro model to simulate scenario values for P_{jt} in periods t, t+1, t+2, ...t+n and on into the future.

Suppose that based on current macro conditions, the simulated value of P_{jt}, labeled P^S_{jt} is 25% as opposed to the value of 19.7% in the historical unconditional transition matrix. Because the unconditional transition value of 19.7% is less than the value estimated conditional on the macro economic state we are likely to underestimate the VAR of loans and a loan portfolio – especially in a low quality end.
The ratio of the two probabilities is given by

$$m_t = \frac{P^S_{jt}}{P_{jt}} = \frac{25}{19} = 1.2$$

Based on the simulated macro model, the probability of a CCC – rated borrower's defaulting over the next year is 20 percent higher than the average (unconditional) historical transition relationship implies. Similar adjustments to the historical default probabilities can be calculated for longer periods. For example, suppose , based on simulated innovations and macro-factor relationships, the simulation predicts P^S_{jt+1} to be 30%. The relevant ratio for next year m_{t+1} will be

$$m_{t+1} = \frac{P^S_{jt}}{P_{jt}} = \frac{30}{19} = 1.5$$

In this example using the unconditional transition matrix will underestimate the risk of default on a CCC-grade borrower in this period.
Following the above procedure, different transition matrices are generated for each year into the future (t, t+1, t+2, ...t+n) reflecting the simulated effect of the macro economic shocks on transition probabilities. This approach could be used along with CreditMetrics to calculate a cyclically sensitive VAR for one year, two years, ... n years.
VAR estimates over longer horizons can also be produced. The transition matrix over the next two years for example can be calculated the two migration matrices

$$M_{t,t+1} = M_t \times M_{t+1}$$

The final column of this new matrix will give the simulated (cumulative) probabilities of default on differently rated loans over the next two years.
Wilson (1997a, b) notes that if long run average levels of the macro-variables are used as initial conditions for simulating the AR(2) processes, then the cumulative default probability of the revised conditional transition matrix would be equal to the original unconditional transition matrix as given by

Standard & Poors or Moody's. This remark shows that the impact of the credit effects is well encompassed by the technology.

One of the limitations of the model is the need of reliable (and frequent) data on the countries and for the industrial sectors within this country. Also, what is criticized is the ad-hoc rule to adjust the transition matrix even though it is a powerful suggestion to be able to relate the classical ratings' approach to the ongoing economic reality. Crouhy et al. (2000) state that there is no proof that this method performs better than the Bayesian alternative that would be based on the internal expertise and appreciation of the credit department of the institution. One of the answers to this criticism is that for regularity purposes, we require a standardized approach. Even though this calibration and simulation techniques can be in part subjective, they rely at least on a formal methodology with defined adjustment procedures.

Another limitation of the model is the ad hoc procedure to adjust the migration matrix.

1.6 CreditRisk+

The CreditRisk+ model is founded on the actuarial principles of mortality tables rather than transition matrices and seeks to estimate the probability of default as a process that is independent of any macro-factors or specific characteristics of the borrower. While CreditMetrics seeks to estimate the full VAR of a loan or loan portfolio by viewing rating upgrades and downgrades and the associated effects of spreads in the discount rate as part of the VAR exposure of a loan, CreditRisk+ views spread risk as part of the market risk rather than credit risk. As a result in any period only two states of the world are considered – default and non-default- and the focus is on measuring expected and unexpected losses rather than expected and unexpected changes in value (or VAR) as under CreditMetrics. Thus CreditMetrics is a mark-to-market (MTM) model; CreditRisk+ is a default model (DM) model.

The second major difference is that in CreditMetrics, the default probability in any year is discrete (as are the upgrade/downgrade probabilities). In CreditRisk+, default is modeled as a continuous variable with a probability distribution. The probability distribution of defaults is assumed to follow a Poisson distribution with parameter λ. The probability function is given by

$$\Pr(n\ defaults) = \frac{\lambda^n e^{-\lambda}}{n!}$$

The expected mean default rate is λ and standard deviation $\sqrt{\lambda}$. Now suppose that based on historical data we are able to derive an estimate of λ from $\bar{n}$ = average number of defaults per year $\bar{n} = \sum p$

The empirical distribution of defaults is then given by

$$\Pr(n\ defaults) = \frac{\bar{n}^n e^{-\bar{n}}}{n!}$$

In order to calculate the expected and unexpected loss of a loan or a portfolio of loans, we also need to have estimates of the severity of the losses themselves. In CreditRisk+, the fact that severity rates are uncertain is tackled not by specifying a probability distribution for the severity rates but by creating bands for loan exposures, e.g. bands of $20,000. Thus we can produce a distribution of losses for each exposure band. Summing these losses across exposure bands produces a distribution of losses for the portfolio of loans.

An example of how to create bands is given in Cruhy et al (2000) which is reproduced here. The bands are based on the size of the Loss Given Default which in this approach is called exposure. In terms of our notation bands are created according to the size of $LGD_i = (1-\delta)F_T$. Table 8 shows the data for this example

Table 1.8: Methodology explanation

LGD_i	Exposure in $	Number of units in category
150	1.5	2
460	4.6	5
435	4.35	5
370	3.7	4
190	1.9	2
480	4.8	2

For each bands we have

$$l_i = v_i \lambda_i \quad \text{or} \quad \lambda_i = \frac{l_i}{v_i}$$

Table 9 below provides an illustration of the results of those calculations. To derive the distribution of losses, for the entire portfolio, we proceed in two steps

First the probability generating function for each band is derived. Each band is viewed as a portfolio of exposures itself. The probability generating function for any band, say band j, is by definition

$$G_i(z) = \sum_{n=0}^{\infty} P(loss = nLGD)z^n = \sum_{n=0}^{\infty} P(\text{n defaults})z^{v_i n}$$

Table 1.9: Calculations

Band j	Number of obligors	l_i	λ_i
1	30	1.5	1.5
2	40	8	4
3	50	6	2
4	70	25.2	6.3
5	100	35	7
6	60	14.4	2.4
7	50	38.5	5.5
8	40	19.2	2.4
9	40	25.2	2.8
10	20	4	0.4

where the losses are expressed in the unit LGD of exposure. Since we have assumed that the number of defaults follows a Poisson distribution the probability generating function for band j is given by

$$G_i(z) = \exp(-\lambda_i + \lambda_i z^{v_i n})$$

To derive the distribution of losses for the entire portfolio we derive the probability generating function for the entire portfolio. Since we have assumed that each band is a portfolio of exposures independent from the other bands, the probability generatinf function for the entire portfolio is simply the product of the probability generating functions for each band

$$G(z) = \prod_{i=1}^{m} \exp(-\lambda_i + \lambda_i z^{v_i}) = \exp\left\{ -\sum_{i=1}^{m} \lambda_i + \sum_{i=1}^{m} \lambda_i z^{v_i} \right\}$$

The loss distribution for the entire portfolio is derived from the probability generating function as follows

$$P(\text{loss of nLGD}) = \frac{1}{n!} \frac{d^n G(z)}{dz^n} \Bigg|_{z=0} \quad \text{for } n = 1,2,\ldots..,$$

These probabilities can be expressed in closed form and depend only on 2 sets of parameters: λ_i and v_i.

CreditRisk+ has the advantage that is relatively easy to implement. First, it allows for closed-form expressions to be derived for the probability of portfolio loan losses and this makes it very attractive from the computational point of view. Second, marginal risk contribution by obligor can be easily computed. Third, CreditRisk+ focuses only on default, requiring relatively few inputs to estimates. For each instrument only the probability of default and the exposure are required.

Its principal limitation is that credit risk has no relationship with market risk. This is a property shared by both CreditMetrics and KMV (which will be reviewd later). In addition CreditRisk+ ignores completely migration risk. The exposure of each obligor is fixed and is not sensitive to future changes in the credit quality of the issuer, or to the variability of future interest rates. Even in its more general form, where the probability of default depends upon several stochastic background factors, the credit exposures are taken to be constant and not related to changes in these factors.

Finally, like CreditMetrics and KMV , CreditRisk+ does not deal with nonlinear products such as, e.g., options and foreign currency swaps.

Crouhy et al. (2000) compare the CreditMetrics and CreditRsk+ model with the BIS (1988) standardised approach on a benchmark portfolio of 1800 bonds across 13 currencies, various maturities, and the entire range of credit standings. The main result is that the models produce Credit VaR values within a relative range of 1.5. This is not entirely surprising since all the models are driven by the default risk.

Gordy (2000) compares CreditMetrics and CreditRsk+ . He shows that under restrictive assumptions each model can mathematically be mapped into the other, despite what seemed ex ante fundamental differences

1.7 Structural Approach - Merton's Model of Default debt

The prevalent theoretical model on which the structural approach is based and which at the same time provides a theoretical explanation of how firm characteristics affect the probability of default is Merton's (1974) model of risky-debt valuation. This is actually a two-state distinguishing firms between bankrupt and non-bankrupt and therefore its use may be limited as companies do not move directly from a state of financial health to a state of bankruptcy. However the model can be extended to cover those intermediate steps and it is indeed extended in CreditMetrics to include migration possibilities.

Let V(t) denote the date t value of a firm's assets. The firm is assumed to have a very simple capital structure. In addition to shareholders' equity, it has issued a single zero-coupon bond that promises to pay an amount F at date T >t. Also let T-t be the time until this debt matures. The balance sheet of this prototype firm is given by

Assets	Liabilities
V(t)	Equity E(t)
	Debt D(t)

Where D(t) is the value of the risky corporate bond, i.e. the discounted value of F using a discount factor that includes the default premium. From the balance sheet, the value of equity in period T, the maturity period will be

$$E_T = V_T - F$$

The company will default on its debt if $V_T < F$ in which case the equity value of the company will be zero. From the viewpoint of the shareholder, and the company's lender the probability of default can be calculated as $\text{Prob}(V_T \leq F)$. Credit risk exists as long as the probability of the value of the assets falling below the face value of the debt at debt maturity is positive, i.e.

$$P(V_T < F) > 0$$

In order to calculate the probability of default we need a model that describes the evolution of the value of the firm's assets over time. We make the standard assumption that the value of the firm V(t) follows a Geometric Brownian motion which expressed in terms of the instantaneous rate of return on the assets of the firm dV_t / V_t given by

$$\frac{dV_t}{V_t} = \mu dt + \sigma dW_t$$

where μ and σ are the mean and the standard deviation of the instantaneous return $\frac{dV}{V}$ and W_t is a standard Brownian motion.

Using standard stochastic calculus results, the value of the assets of the firm in period T will be given by

$$V_T = V_t \exp\left((\mu - \tfrac{1}{2}\sigma^2)(T-t) + \sigma\sqrt{T-t}Z_T\right)$$

where $Z_T \sim N(0,1)$ and the term $\sqrt{T-t}Z_T = W_T - W_t$ is normally distributed with a zero mean and a variance equal to (T-t). Note that the value of the firm at maturity time T is log-normally distributed with expected value at time T given by

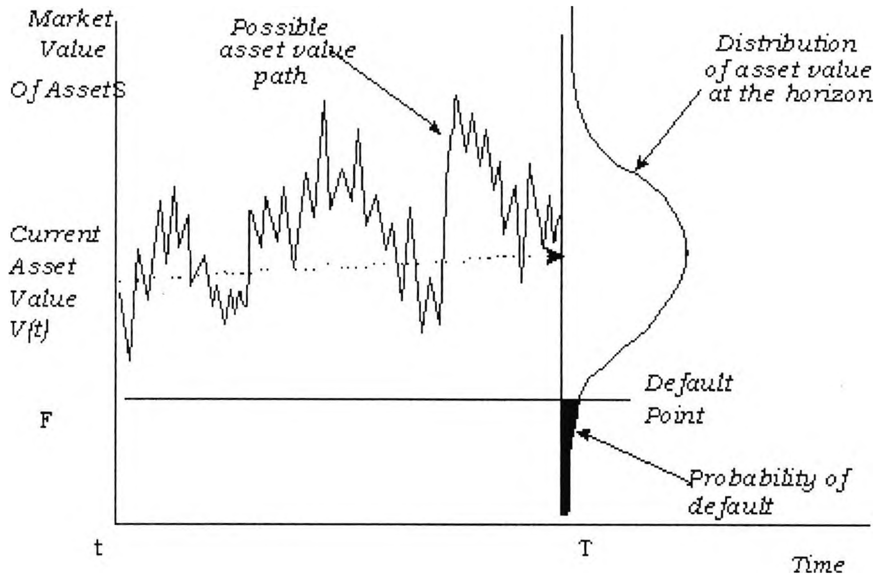
$$E(V_T) = V_t \exp(\mu(T-t))$$

Figure 1.4 shows the distribution of the value of the assets at time T, the maturity of the zero-coupon corporate debt, and the probability of default, the latter being representd by the shaded area on the left-hand side of the default point, F.

The default condition $V_T \leq F$ implies $V_t \exp\left((\mu - \tfrac{1}{2}\sigma^2)(T-t) + \sigma\sqrt{T-t}Z_T\right) \leq F$ which after dividing both sides of the equation by the term V_t can be written as

$$\exp\left((\mu - \tfrac{1}{2}\sigma^2)(T-t) + \sigma\sqrt{T-t}Z_T\right) \leq \frac{F}{V_t}$$

Figure 1.4: Default Probability in the Merton Model



Finally taking logs the default condition can be written as

$$(\mu - \tfrac{1}{2}\sigma^2)(T-t) + \sigma\sqrt{T-t}Z_T \le \ln\frac{F}{V_t}$$

This last inequality can be expressed in terms of the standardised normal random variable Z_T as follows:

$$Z_T \le \frac{\ln\frac{F}{V_t} - (\mu - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \quad \text{or } Z_T \le -\frac{\ln\frac{V_t}{F} + (\mu - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \quad \text{or } Z_T \le -d_2$$

Where $d_2 = \dfrac{\ln\frac{V_t}{F} + (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$ So that

$Prob(Default) = Prob(V_T \le F) = Prob(Z_T \le -d_2) = N(-d_2)$

Default is therefore triggered when $Z_T \le -d_2$

The quantity

$$d_2 = \frac{\ln\frac{V_t}{F} + (\mu - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

is also known as distance to default.

The probability of default depends therefore on

- the debt/asset ratio, F/V
- the expected rate of return on the assets of the company (μ).
- the volatility of the company assets (σ).
- The maturity of debt (T-t).

The effect of three of the factors is unambiguous. A higher F/V ratio increases the probability of default and so does an increase in volatility. A higher expected return on the assets of the company will also increase the probability of default as a higher expected return in equilibrium will lead to higher risk. The effect of maturity depends on whether F/V>1 or whether F/V<1. If F>V, the firm is technically insolvent. To avoid bankruptcy, it will need to have increasing earnings. As maturity increases, there is more time for the increase in earnings to occur and for the risk to be reduced If F<V and the loan has only a short time to go before maturity, it is unlikely that the loan will default. .

1.8 The KMV Model

The most widely used variant of the Merton model is the KMV model (Kealhofer 1997). The model utilises the Merton framework and treats equity as a European call option on the assets of the company with an exercise price equal to debt F. The maturity of the call option is the same as the maturity of the corporate bond issued. From standard Black-Scholes option pricing theory, the value of the equity is given by

$E = E(V_t, F, (T-t), \sigma, r)$ [3]

The volatility of equity is linked to the volatility of the assets (Hull (2000))

$\sigma_E = h(\sigma)$

Once we have values for the current value of equity and estimates for the volatility of equity from historical data, the two equations can be solved for the two unobservable variables V_t, σ using an iterative non-linear algorithm. Once we have estimates for the unobservable variables, then the probability of default can be calculated using the Merton formula. KMV use these estimates to calculate the expected default frequency (EDF) for any given borrower. The idea is shown in Figure 1.5. The EDF is the shaded area below F.

If it is assumed that future asset values are normally distributed around the firm's current asset value, we can measure today's distance from default at a one-year horizon as

$$DFD = \frac{V_t - F}{\sigma}$$

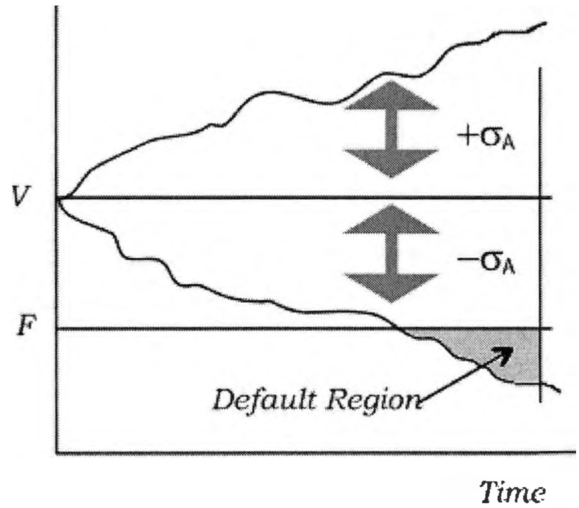
Despite the adoption of normality assumption by CrediMetrics this is a questionable assumption and KMV rather than using theoretical EDFs generates empirical EDFs as follows.

$$\text{Empirical } EDF = \frac{\textit{Number of firms that defaulted within a year with asset values of } 2\sigma \textit{ from } F \textit{ at the beginning of the year}}{\textit{Total population of firms with asset values of } 2\sigma \textit{ from } F \textit{ at the beginning of the year}}$$

Figure 1.5. Calculation of the Theoretical EDF in the KMV model.

---

[3] The variable r is the risk –free rate. When the equity of the firm is valued using the Black-Scholes model, the expected return on assets is replaced by the risk-free rate. In the risk-neutral world we define $d_2 = \dfrac{\ln \frac{V_t}{F} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$ and $N(-d_2)$ is the risk-neutral probability of default



The structural approach as it is applied by KMV has a number of advantages, notably the fact that it utilizes stock market data rather than historic accounting data and it reflects market changes immediately. It is also consistent with the modern theory of finance and the variables used are theoretically justified. The statistical approach lacks such an underpinning.

Against that one of course should count the fact that it is primarily a model for public quoted companies. It is also very difficult to be implemented without the assumption of normality. It is also considered static in the sense that over the time horizon the company is not expected to change its capital structure. Finally it assumes a very simple type of debt when in practice companies use a variety of debt instruments. For all these reasons the objective of this study is to develop a credit measuring system based on the statistical approach.

Loffler (1999) proposes a comparative study of the sensitivity to input errors of the KMV model and CreditMetrics™. Two inputs that are shared by both seem particularly problematic: correlations and recovery rates. CreditMetrics™ appears to be more sensitive to recovery rates errors but possibly less to correlation errors.

KMV like the other models suffers from defects. Recovery rates are determined in an ad hoc way (beta distributions for most, with levels determined from US-based studies even for implementation of the models in Europe or in Asia, where bankruptcy laws differ substantially form US laws).

None of the models currently integrates some minimal form of market risks, at least via stochastic interest rates. None of the models can handle nonlinear instruments (from guarantees to swaps to credit derivatives and letters of credit or call features and OTC options). Some technical issues, such as discrete approximation of the continuous distribution, may lead to difficulties (and not only on small samples) for all of them.

However some of the models by looking at skewed and fat-tailed distributions present a significant improvement on risk management techniques based on the standard deviation.

1.9 Univariate Statistical Approach

In the statistical approach, the financial characteristics of a company are normally summarised through a set of financial ratios constructed from the company's balance sheet and profit and loss account. The use of ratios instead of actual values means that all pertinent financial variables are normalised for size, which eliminates potential distortions induced by the fact that company size varies by several orders of magnitude across firms in our sample[4]. The use of ratios also provides a deflator and avoids differences in the value of money across years and thus facilitates intertemporal comparisons.

Table 1.10: Fraction of Sample that is misclassified

	Ratio			
Years before failure	Cash flow/Total assets	Cash flow/Total debt	Net Income /Total debt	Sample size
1	0.10* (0.10)	0.13 (0.10)	0.15 (0.08)	158
2	0.20 (0.17)	0.21 (0.18)	0.20 (0.16)	153
3	0.24 (0.20)	0.23 (0.21)	0.22 (0.20)	150
4	0.28 (0.26)	0.24 (0.24)	0.26 (0.26)	128
5	0.28 (0.25)	0.22 (0.22)	0.32 (0.26)	117

*

Source: Beaver (1966, table A-4)

The statistical approach to credit risk measurement started with Patrick (1932) who employed ratio analysis to predict corporate failure and was extended by Beaver (1966) who developed a framework for the univariate analysis of bankruptcy.

Beaver used more than 30 financial ratios which he divided into six groups and tested their predictive power individually. In most cases the individual ratios were able to differentiate between bankrupt and non- bankrupt firms providing the first empirical evidence that financial ratios can predict the financial health or credit risk of a company. Beaver's study was criticized for its dependence on single ratios rather than looking at combinations of factors that may affect the financial state of a company and its probability of default.

Table 10 contains some of Beaver's results and suggests that single ratios can predict failure reasonably well. Despite the statistical shortcomings, as Beaver's results indicate, a univariate analysis allows the researcher to understand the nature of the data and provides valuable insights for the multivariate analysis.

Casey and Bartczack also applied univariate analysis of the predictive cash flow from operations (CFO) and related cash flow ratios. They found that although the ratios were good at identifying bankrupt companies (e.g., using CFO, one could correctly classify 92 percent of the bankrupt firms), they did poorly in classifying nonbankrupt firms, compared to a discriminant model composed of six popular accrual-based ratios. Furthermore when looking at overall accuracy (on the total of bankrupt and nonbankrupt firms), Casey and Bartczack found that the multivariate accrual model was superior.

1.10 Linear Multivariate Discriminant Analysis

[4] Size is measured by total assets

The next generation of models were based on multivariate discriminant analysis (MDA) with Altman (1968) and Altman (1983) employing MDA to discriminate between bankrupt and nonbankrupt companies. Most of the studies that employ MDA restrict themselves to two-group classification and this is the version we shall concentrate on in this review. In MDA the random vector r of dimension p is measured in two populations, bankrupt and non-bankrupt companies with mean vector $E_1(\mathbf{r}) = \boldsymbol{\mu}_1$, and covariance matrix $\boldsymbol{\Omega}$ in population 1 and mean vector $E_2(\mathbf{r}) = \boldsymbol{\mu}_2$, and covariance matrix $\boldsymbol{\Omega}$ in population 2. Let $\boldsymbol{\beta} = (\beta_1, ...\beta_p)'$ denote the vector of coefficients of a linear combination, and set $Z = \boldsymbol{\beta}'\mathbf{r}$. The mean and the variance of the random variable Z are $E(Z) = \mu_Z = \boldsymbol{\beta}'\boldsymbol{\mu}$ and $\mathrm{Var}(Z) = \sigma_Z^2 = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}$ respectively. Applying the same linear combination to the points $\mathbf{r}_1$ and $\mathbf{r}_2$ representing the values of the vector $\mathbf{r}$ in the two populations we have $Z_1 = \boldsymbol{\beta}'\mathbf{r}_1$ and $Z_2 = \boldsymbol{\beta}'\mathbf{r}_2$ with $E(Z_1) = \boldsymbol{\beta}'\boldsymbol{\mu}_1$ and $E(Z_2) = \boldsymbol{\beta}'\boldsymbol{\mu}_2$.

The vector of coefficients $\boldsymbol{\beta} = (\beta_1, ...\beta_p)'$ is selected so as to maximize the quantity

$$\frac{\mathrm{E}(Z_1) - \mathrm{E}(Z_2)}{\sigma_Z} = \frac{\boldsymbol{\beta}'\boldsymbol{\mu}_1 - \boldsymbol{\beta}'\boldsymbol{\mu}_2}{(\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta})^{1/2}}$$

The vector that maximises the distance is given by[5]

$$\boldsymbol{\beta} = \boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

The classification rule based on discriminant analysis is defined as follows: Classify a company in population 1 if

$$Z > \frac{1}{2}[E(Z_1) + E(Z_2)] \quad \text{or equivalently} \quad \boldsymbol{\beta}'\mathbf{r} > \frac{1}{2}(\boldsymbol{\beta}'\boldsymbol{\mu}_1 + \boldsymbol{\beta}'\boldsymbol{\mu}_2)$$

In empirical studies the population means are replaced by the sample means $\bar{\mathbf{r}}_1$ and $\bar{\mathbf{r}}_2$ so that the classification rule becomes $\boldsymbol{\beta}'\mathbf{r} > \frac{1}{2}(\boldsymbol{\beta}'\bar{\mathbf{r}}_1 + \boldsymbol{\beta}'\bar{\mathbf{r}}_2)$

The linear classification rule that maximizes the distance between the population means of the linear combination of random variables is optimal in the sense of minimizing the misclassification error only if the random vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ follow multivariate normal distributions with the same variance covariance matrix[6] $\mathbf{r}_1 \sim \mathrm{N}(\boldsymbol{\mu}_1, \boldsymbol{\Omega})$ $\mathbf{r}_2 \sim \mathrm{N}(\boldsymbol{\mu}_2, \boldsymbol{\Omega})$

---

[5] This vector is not unique. Any vector $\boldsymbol{\beta}^* = c\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ will also maximise the distance. In that sense the discriminant function $Z = \boldsymbol{\beta}'\mathbf{r}$ is uniquely defined.

[6] This can be seen very easily in the case of a single discriminating variable $r \sim N(\mu_1, \sigma^2)$ and $r \sim N(\mu_2, \sigma^2)$. In this case

$$\Pr(r < \gamma / r \in 1) = \Pr\left(\frac{r - \mu_1}{\sigma} < \frac{\gamma - \mu_1}{\sigma} / r \in 1\right) = \Pr\left(\zeta < \frac{\frac{1}{2}(\mu_1 + \mu_2) - \mu_1}{\sigma} / r \in 1\right) = \Pr\left(\zeta < -\frac{\mu_1 - \mu_2}{2\sigma} / r \in 1\right) = \Phi\left(-\frac{D}{2}\right)$$

Using similar arguments we can show that $\Pr(r > \gamma / r \in 2) = \Phi\left(-\frac{D}{2}\right)$. This shows that the higher the distance the lower the probability of misclassification.

From the discriminant function we can calculate a credit index which allows the derivation of the probability of default and the calculation of expected and unexpected losses from the values of the Z score. For example using the Z –score risk rating system developed by Altman et al. (1993), Altman and Saunders (1998) have assigned a bond equivalent rating to each of the loans/bonds evaluated using the following discriminant function

$$Z = 6.56X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4 + 3.25$$

where

X_1 = *working capital / total assets*
X_2 = *retained earnings / total assets*
X_3 = *EBIT/ total assets*
X_4 = *equity (book value) / total liabilities*

A constant term has been added to the equation to standardise the bond rating equivalent analysis, so that scores of zero indicate a D (default) rating and positive scores indicate ratings above D.

The rating equivalents of each score can be used to estimate the default probability using the migration tables and thus the expected loss over time. Using the actual loss experience for every year over a number of years the standard deviation can be calculated, which is a measure of the unexpected loss.

Table 1.11: Z scores and Credit rating

US equivalent rating	Average Z –score	Sample Size
AAA	8.15	8
AA+	7.60	-
AA	7.30	18
AA-	7.00	15
A+	6.85	24
A	6.65	42
A-	6.40	38
BBB+	6.25	38
BBB	5.85	59
BBB-	5.65	52
BB+	5.25	34
BB	4.95	25
BB-	4.75	65
B+	4.50	78
B	4.15	115
B-	3.75	95
CCC+	3.20	23
CCC	2.50	10
CCC-	1.75	6
D	0.00	14

Source: Altman and Saunders (1998)

Buy, Kaminski, Pinnamaneni and Shanbhogue (1998) used a similar technique to match financial ratios to the rating scales of the credit rating agencies. They produced a numerical scale corresponding to the alphabetic system and they then "explained" the numeric scale by a set of financial ratios.
They used data on 90 public electric utilities for the years 1994-96 which were rated by Standard and Poor. They regressed the ratings on each company for the years 95-97 on a set of financial ratios for the years 94-96. They estimated the following equation with an adjusted R^2 of 65.7%

$$Rating = 8.65 + 0.17X_1 - 1.08X_2 - 0.06X_3$$

where

$X_1 = total\ debt/\ total\ assets$

$X_2 = return\ on\ assets$

$X_3 = book\ value\ per\ share$

Table 1.12: Credit Rating Correspondence

Standard and Poor's	Moodys	Numerical Scale
AAA	Aaa	2
AA+	Aa1	4
AA	Aa2	5
AA-	Aa3	6
A+	A1	7
A	A2	8
A-	A3	9
BBB+	Baa1	10
BBB	Baa2	11
BBB-	Baa3	12
BB+	Ba1	13
BB	Ba2	14
BB-	Ba3	15
B+	B1	16
B	B2	17
B-	B3	18
CCC+	Caa	19
CCC	Ca	20
CCC-	C	21

Having explained the fundamental aspects of this approach we turn now to the empirical performance of this approach. Our interest is on three aspects. First we are interested to see what risk factors each study has identified. Second we would like to review the performance of the method. The proportion of correctly classified cases is invariably used to measure the performance of discriminant analysis. The third aspect we are interested in is, where possible, to present results testing the assumptions of the approach.

Altman (1968)

The first empirical study employing this technique was Altman (1968). Altman considered 22 ratios and found five of them to be significant in discriminating between bankrupt and non-bankrupt companies. Altman found the following rates as the most significant variables (Table 12).

- Working Capital / Total Assets
- Retained Earnings / Total Assets
- EBIT / Total Assets
- Market Value of Equity / Total Assets
- Sales / Total Assets

Altman, Haldeman and Narayanan (1977)

Altman developed three variants of the bankruptcy prediction models, the initial Z-score, the Zeta model and finally the quadratic ZETA model whose performance was tested and found it predicted bankrupt companies up to four years before bankruptcy occurred. In this paper they presented the Zeta model using discriminant analysis to a sample of 53 bankrupt firms and 58 non-bankrupt firms matched on the basis of industry and year data. The 1968 model was refined by incorporating prior probabilities and cost of misclassification and using quadratic as well as linear discriminant analysis. The discriminating variables included common financial ratios

Table 1.13 Altman's results

Variables	Parameter values[7]
Working Capital / Total Assets	-0.120
Retained Earnings / Total Assets	-0.014
EBIT / Total Assets	-0.033
Market Value of Equity / Total Assets	-0.006
Sales / Total Assets	-0.999

as well as trend, stability and stock market measures. When applied to holdout samples, a linear discriminant model correctly classified 93 percent of bankrupt firms and 90 percent of non-bankrupt firms one year before bankruptcy and 70 percent and 82 percent, respectively five years before bankruptcy. Even though the covariance matrices were unequal, a quadratic discriminant model did not improve classification results.

Lis (1972)

Altman's approach has been used by many researches in many countries. Lis(1972) was the first UK study and showed that the framework developed by Altman was applicable in other countries . The study develops a 4-variable discriminant function with ratios based on Altman (1968) using as is bankrupt sample 30 major quoted manufacturing construction and retailing failures between 1964 and 1972 and with his equal size continuing sample matched by industry, asset size and year. The model is given by

$$Z = 0.063X_1 + 0.092X_2 + 0.057X_3 + 0.0014X_4$$

[7] Altman does report the significance levels of individual variables

X_1 = working capital / total assets

X_2 = EBIT/total assets

X_3 = retained earnings (adjusted for scrip issues)/ total assets

X_4 = net worth / total debt

A selected cut-off of 0.037 misclassified one –failed company and 5 non-failures. The mean of the failed group was –0.003 and that of the matched sample was 0.063.

Taffler 1974

This study was the second study in the UK to use Altman's approach and the one that had the most significant impact in the area. The following ratios were found significant in separating bankrupt from non-bankrupt companies. 23 failing and 45 non-failing firms over the period 1968 to 1973 were employed to test the significance of the model.

X_1 = *EBIT/ total assets*

X_2 = *Total liabilities / net capital employed*

X_3 = *quick assets / total assets*

X_4 = *working capital / net worth*

X_5 = *stockturn*

The cut-off point was set taking into account prior probability odds of 1:10 (failed:solvent) for the population . the performane of the model was good with only 12 percent of the bankrupt firms being missclasified.

Taffler (1977)

In a subsequent study Taffler experimented with other ratios and found that the following ratios were significant discriminating factors.

X_1 = Profit before tax / average curent liabilities

X_2 = Current assets /total liabilities

X_3 = Current liabilities / total assets

X_4 = the no - credit interval

The sample of this study consisted of 46 manufacturing firms quoted on the London Stock Exchange and which had failed over the period 1968-1976. The solvent companies were matched on a 1:1 basis by industry and size but not by year, with the latest year available being used. The Z sore was used to classify 825 listed companies, of which 115 (14%) had a Z value of less than zero at the end of 1976. Nearly 43 percent of those companies had failed within six years, whereas another 29 percent were still at risk. These results, show that the predictive ability of the financial ratios employed is high.

Mason and Harris (1978)

This is one of the earliest studies developed for a particular sector, in this case the construction industry.

$$Z = 25.4 - 51.2X_1 + 87.8X_2 - 4.8X_3 - 14.5X_4 - 9.1X_5 - 4.5X_6$$

X_1 = *Profit before interest and tax / opening net assets*
X_2 = *Profit before interest and tax/ opening net capital employed*
X_3 = *debtors/ creditors*
X_4 = *Current liabilities / current assets*
$X_5 = log_{10}$*(day debtors)*
X_6 = *creditors trend measurement*

The performance of the model, judged on the basis of misclassification errors, was adequate with 63.7 of the failing companies being correctly classified.

El Hennaway and Morris (1983)

The authors used a longer period (1955-1974) to test the predictive power of the model and also introduced dummy variables for the various sectors from which the companies in the sample come from recognising the fact that different companies have different probabilities of default. The authors estimated discriminant function for the prediction of failure one year ahead, which is given below

$$Z = -6.17 + 11.43X_1 + 14.07X_2 + 0.55X_3 - 1.57X_4 + 0.98X_5$$

where

X_1 = *Operating profit before depreciation / total assets*
X_2 = *Long - term debt/ net capital employed*
X_3 = *Current assets / total assets*
X_4 = *Quarrying and construction industry dummy*
$X_5 = log_{10}$*(day debtors)*
X_6 = *distribution industry dummy*

and discriminant function for the prediction of failure 5 years before failure. They also found that industry membership was important.

Ko (1982)

Ko estimated a discriminant analysis model to,the Japanese problem firms. His sample included 41 pairs of bankrupt and non-bankrupt entities for the period 1960-1980. The proportion of correct classifications was about 82.9 percent. Ko found that five variables were significant and with the correct sign. Three of these ratios were the same as in the original study of Altman (1968).

The estimated discriminant function is given below.

$$Z = 0.868X_1 + 0.198X_2 - 0.048X_3 + 0.436X_3 + 0.436X_4 + 0.115X_5$$

where

$X_1 = EBIT / SALES$
X_2 = *Inventory turnover 2 years prior / inventory turnover 3 years prior*
X_3 = *standard error of net income (4 years)*
X_4 = *working capital / total debt*
X_5 = *market value equity / total debt*

The standardised form results in a zero cut-off score, i.e. any score greater than zero indicates a healthy situation, with probability of classification of bankruptcy less than 0.5 and probabilities greater than 0.5 for negative scores. A significant innovation in the set of discriminating variables was the inclusion of the standard deviation of net income based on the previous four years as a proxy for the impact of earnings volatility. The negative sign indicates that income volatility increases the probability of bankruptcy.

Altman and Lavalee (1981)

This is one of the first studies using Canadian data by Altman and his associates. They data cover the period 1968-1980 and bankrupt companies are matched by surviving companies from the same year and the same industry. The study was based on a sample of 54 publicly traded firms, half-failed and half-continuing entities. The companies had failed over the period 1970-1979. The companies were selected to have the same size as well. The estimated discriminant function is given below.

$$Z = -1.626 + 0.234X_1 - 0.531X_2 + 1.002X_3 + 0.972X_4 + 0.612X_5$$

where

$X_1 = Sales / Assets$
$X_2 = Total\ debt\ / total\ assets$
$X_3 = current\ assets / total\ liabilities$
$X_4 = net\ profits\ after\ tax\ / total\ debt$
$X_5 = rate\ of\ growth\ of\ equity - rate\ of\ asset\ growth$

The classification criterion was based on a zero cut-off score with positive scores indicating a non-failed classification and negative scores a failed assignment. The overall classification accuracy of the model on the original 54 firms was 83.3 % which is quite high.

Bilderbeek (1977)

Bilderbeek analysed a sample of 38 Dutch firms which went bankrupt from 1950-1974 and 59 ongoing Dutch companies. Bilderbeek analysed 20 ratios within a step-wise discriminant framework and arrived at a 5 variable model of the form

$$Z = 0.45 - 5.03X_1 - 1.57X_2 + 4.55X_3 + 0.17X_4 + 0.15X_5$$

where

$X_1 = Retained\ earnings\ / total\ assets$
$X_2 = Added\ value\ / total\ assets$
$X_3 = Account\ payable / Sales$
$X_4 = Sales / Total\ assets$
$X_5 = net\ profits\ after\ tax\ / equity$

Two of the five signs are positive and contrary to the expectations since for this model negative scores indicate a healthy situation and positive scores indicate a failure classification. His model was base don observations over 5 periods prior to failure and is not based on 1-year intervals. His results were only mildly impressive with accuracies ranging from 70-80% for 1 year prior and remaining stable over a 5 year period prior to failure.

Micha (1984)

Micha used data on French companies that went bankrupt and he found the following significant discriminant function. Again matched data (for size, industry and year) were used and the following discriminant analysis was estimated

$$100Z = -1.255X_1 + 2.003X_2 - 0.824X_3 + 5.221X_4 - 0.689X_5 - 1.164X_6 + 0.706X_7 + 1.408X_8 - 85.544X_9$$

where

$X_1 = \textit{Interest Charge/ Gross Operating Profit}$
$X_2 = \textit{Long term resources / capital employed}$
$X_3 = \textit{net cash flow / total financial debt}$
$X_4 = \textit{Gross Operating Profit / net sales}$
$X_5 = \textit{trade Debt / purchases, incl. VAT}$
$X_6 = \textit{Annual variation in value added}$

$$X_7 = \frac{\textit{Work in progress - customer's prepayment + trade accounts receivable}}{\textit{output, incl. VAT}}$$

$$X_8 = \frac{\textit{Fixed assets investment}}{\textit{value added}}$$

Izan (1984)

Izan's study covers the period 1960-1980 using Australian data. The standardised coefficients of the discriminant function are shown below

$$Z = 0.23X_1 + 0.53X_2 + 0.24X_3 - 0.25X_4 + 0.44X_5$$

where

$X_1 = \textit{EBIT / Tangible total assets}$
$X_2 = \textit{EBIT / Interest payments}$
$X_3 = \textit{current assets / current liabilities}$
$X_4 = \textit{funded debt (borrowings) / shareholder funds}$
$X_5 = \textit{rmarket value of equity /total liabilities}$

Izan found that the interest coverage and the ratio of the market value of equity to total liabilities had the most important relative contribution whereas the liquidity variable X3 contributed the least. This is consistent with many past studies.

Frydman Altman and Ko

The primary purpose of this study was to compare classification trees and discriminant analysis. The discriminant analysis models were constructed using a forward stepwise procedure. The first model included 10 variables while the second model included only the four most significant variables as provided by the stepwise method. The two discriminant functions were

$$Z_1 = -1.255X_1 + 2.003X_2 - 0.824X_3 + 5.221X_4 - 0.689X_5 - 1.164X_6 + 0.706X_7 + 1.408X_8 - 85.544X_9$$
$$Z_2 = -1.255X_1 + 2.003X_2 - 0.824X_3 + 5.221X_4 - 0.689X_5 - 1.164X_6 + 0.706X_7 + 1.408X_8 - 85.544X_9$$

where

$X_1 = \textit{Net income / Total assets}$
$X_2 = \textit{Current Assets / Current Liabilities}$
$X_3 = \textit{log(Total Assets)}$
$X_4 = \textit{Gross Operating Profit / net sales}$
$X_5 = \textit{trade Debt / purchases, incl. VAT}$
$X_6 = \textit{Annual variation in value added}$

$$X_7 = \frac{\textit{Work in progress - customer's prepayment + trade accounts receivable}}{\textit{output, incl. VAT}}$$

$$X_8 = \frac{\textit{Fixed assets investment}}{\textit{value added}}$$

In this study a number of new variables were introduced that did not appear in previous studies.

Gloubos and Grammatikos (1988)

Gloubos and Grammatikos, in the earliest study, which used Greek data, found significant risk factors that were able to differentiate between bankrupt and non-bankrupt companies one year prior to going bankrupt. The predictive ability of the model was good with the proportion of correct predictions around 70 percent. The following risk factors were found significant.

- Current Assets / Current Liabilities
- Net Working Capital / Total assets
- Total Debt / Total Assets
- Gross Income / Total Assets
- Gross Income / Current Liabilities.

The data used in the study covered the period 1970-1988 and the failing companies were companies that had applied for bankruptcy proceedings to start.

Ooghe and Verbaere (1992)

Ooghe and Verbaere using Belgian data for the period 1972-1990 ended up with the following significant factors as discriminating variables in a linear regression model.

- Overdue short-term priority debts / Short-term liabilities
- Accumulated profits / Total liabilities
- Gross earnings before interests and taxes / Total Assets
- Equity Capital / Total Liabilities
- Cash / Current Assets

The data used in the Belgian study consisted of a matching sample of 45 failing and 45 non-failing industrial firms.

Shirata (2001)

In a recent study Shirata, using more recent Japanese data, found the following ratios significant in a linear discriminant analysis model.

- Retained Earnings / Total Assets
- Accounts payable / sales
- Current Gross Capital / Gross Capital a year ago
- Interest and discount expenses / total borrowing and bond issues

Shirata used a matched sample of 62 failing and non-failing firms and reports small classification errors. The proportion of correct forecasts is around 78 percent a year prior to bankruptcy. The set of factors that were identified as significant is rather small, but Shirata reports exhaustive tests that indicate that other variables do not play a role is discriminating between the two groups.

Conclusion on empirical discriminant studies

In this section we have reviewed most of the international studies that have been published over the last 30 years. The emphasis was on presenting the empirical evidence in order to establish two things. First, the range of financial ratios that have been identified as significant by the empirical studies. The several international studies reviewed so far have found, with minor exceptions, the same list of factors as significant discriminating variables between bankrupt and non-bankrupt firms. In selecting the initial set of factors for the statistical analysis of Chapter 3 we can therefore rely on the available empirical evidence. The second issue that is clear from this review is that discriminant analysis seems to be able to differentiate between companies that went bankrupt and companies that survived. Whether this ex post performance analysis can guarantee ex ante predictive success is open to debate, because all the studies have based their methodology on matched pairs of companies spanning a very large period of time. This approach may be ignoring macroeconomic and other influences that may affect the probability of default over the business cycle and consequently may not capture the changing environment in which companies may be operating.

Problems with discriminant analysis

Despite their popularity, bankruptcy prediction models based on the Discriminant Analysis method, suffer from a number of problems as they are based on assumptions that are difficult to be satisfied by the data. (Zavgren, 1985). One of the most important problems relates to the assumption of normality for the financial ratios employed in the construction of the linear Discriminant function. This assumption is invalidated as soon as we use categorical variables as risk factors. Other problems related specific to the linear discriminant function, such as the assumption that the variance covariance matrices in the bankrupt and non-bankrupt groups are equal.

As a result of these problems alternative models that require less demanding models have been proposed. One line of extension is to relax the assumption of equal covariance matrices in the two populations. This is done in the following section. Another line of extension is to relax the assumption of normality. The logistic regression model (Logit Model) is the most popular of models, which do not make the assumption that the risk factors are normally distributed. In fact it can be shown that the DA method is a special case of the logistic regression model (Kennedy, 1991). Which of the two models performs better is still debatable. Ohlson (1980) claims that if deviations from normality are small, then the DA method produces Maximum Likelihood estimators that are asymptotically more efficient than the Logit Model.

1.11 Quadratic Discriminant Analysis

One of the major criticisms of the linear discriminant analysis is the assumption of equal covariance matrices in the two populations (Eisembeis 1977). There is little evidence that the dispersal matrices in the two populations are the same and if the assumption is rejected, but multivariate normality is retained, then the classification rule becomes (Huberty (1994) , Flury (1997))

$$\frac{1}{2}\mathbf{r}'(\boldsymbol{\Omega}_2^{-1}-\boldsymbol{\Omega}_1^{-1})\mathbf{r}+(\boldsymbol{\Omega}_2^{-1}\boldsymbol{\mu}_2-\boldsymbol{\Omega}_1^{-1}\boldsymbol{\mu}_1)\mathbf{r} > \frac{1}{2}\ln\frac{\boldsymbol{\Omega}_2}{\boldsymbol{\Omega}_1}+\frac{1}{2}[\boldsymbol{\mu}_2'\boldsymbol{\Omega}_2^{-1}\boldsymbol{\mu}_2-\boldsymbol{\mu}_1'\boldsymbol{\Omega}_1^{-1}\boldsymbol{\mu}_1)$$

which reduces to the linear discriminant function when

$$\boldsymbol{\Omega}_1=\boldsymbol{\Omega}_2=\boldsymbol{\Omega}$$

The quadratic version of MDA (QMDA) which is not based on the assumption of equal covariance matrices, is in principle more robust way of classifying firms, but implementation of the QMDA is not supported by most statistical packages that make its implementation difficult.

1.12 Linear Probability Model

The next three methods attempt to estimate the probability of default directly rather than classifying observations into groups of bankrupt and non-bankrupt companies. The linear probability model is a special case of ordinary least squares regression with a dichotomous (0-1) dependent variable. It represents the simplest attempt to calculate the probability of default by directly linking it to characteristics of a company. The model assumes that the random variable x describes the state of a company. If a company is downgraded then, $x = 1$ and if the company s not downgraded then $x = 0$. We believe that there is a vector $\mathbf{r}$ explaining the state of the company, so that

$$\Pr(x = 1) = F(\mathbf{r}, \boldsymbol{\beta})$$
$$\Pr(x = 0) = 1 - F(\mathbf{r}, \boldsymbol{\beta})$$

The linear probability model results if we make the assumption that

$$F(\mathbf{r}, \boldsymbol{\beta}) = \boldsymbol{\beta}'\mathbf{r}$$

Since $E(x / \mathbf{r}) = F(\mathbf{r}, \boldsymbol{\beta})$ the following regression model can be constructed

$$x = E(x / \mathbf{r}) + [x - E(x / \mathbf{r})] = \boldsymbol{\beta}'\mathbf{r} + \boldsymbol{\varepsilon}$$

Serious statistical problems however exist with the linear probability model. First of all the error term is heteroscedastisc with variance

$$Var\ (\boldsymbol{\varepsilon} / \mathbf{r}) = \boldsymbol{\beta}'\mathbf{r}(1 - \boldsymbol{\beta}'\mathbf{r})$$

The second problem is that there is no way to guarantee the predicted values of model will not lie outside the (0-1) interval. For this reason this approach has not been used very frequently in empirical studies [see e.g. Meyer and Pifer (1970) and Theodossiou (1991), Altman et. al. (1981)] both because of the statistical problems already mentioned but also because it was found to produce results not significantly different from Discriminant Analysis.

1.13 Logistic Regression

Consider a random variable x, which indicates membership in the population 1 or 2. The random variable takes the values 0 and 1 with probabilities p_1 and p_2. Since both x and $\mathbf{r}$ are random one could study their joint density function $f(x, \mathbf{r})$ defined as

$$f_{x,r}(j, \mathbf{r}) = \begin{cases} p_j f_j(\mathbf{r}) & \text{for } j = 1,2 \\ 0 & \text{otherwise} \end{cases}$$

Discriminant analysis looks at the conditional distribution of $\mathbf{r}$, given x. In discriminate analysis we assume that conditionally on $x = 1$, $\mathbf{r}$ follows a p-variate distribution with probability density function $f_1(\mathbf{r})$. Similarly, assume that conditionally on $x = 0$, $\mathbf{r}$ follows a p-variate distribution with probability density function $f_2(\mathbf{r})$. We also make the further assumption that the conditional distribution of $\mathbf{r}$ is multivariate normal. With all these assumptions we get the classification rules studied in an earlier section. If we assume equal covariances we get linear classification rules, if we assume unequal covariances we get quadratic classification rules.

However it is possible instead of studying the conditional distribution of $\mathbf{r}$ given $x = 1,2$ to study the conditional distribution of x given $\mathbf{r} = \mathbf{r}^*$. Denoting the marginal probability density function of $\mathbf{r}$ as

$$f_r(\mathbf{r}) = f_1(\mathbf{r})p_1 + f_2(\mathbf{r})p_2$$

The conditional probability of membership in group 1, given $\mathbf{r} = \mathbf{r}^*$ can be calculated by

$$\pi_{1r} = \Pr(x = 1 / \mathbf{r} = \mathbf{r}^*) = \frac{f_1(\mathbf{r}^*)p_1}{f_1(\mathbf{r}^*)p_1 + f_2(\mathbf{r}^*)p_2}$$

The π_{1r} is usually called the posterior probability and by construction $\pi_{1r} + \pi_{2r} = 1$.

In some cases it may be more reasonable to model the conditional distribution of x given $\mathbf{r} = \mathbf{r}^*$ directly, without making any distributional assumptions on $\mathbf{r}$. In this setup one can treat x as the dependent variable and $\mathbf{r}$ as a vector of p "independent" or regressor variables. The most popular statistical technique based on this approach is logistic regression. Logistic regression assumes that

$$\pi_{1r} = \Pr(x = 1 / \mathbf{r} = \mathbf{r}^*) = \frac{\exp[\alpha + \boldsymbol{\beta}'\mathbf{r}]}{1 + \exp[\alpha + \boldsymbol{\beta}'\mathbf{r}]}$$

In the logistic regression framework therefore, the probability of bankruptcy can be estimated directly on the basis of a vector of financial ratios.

Ohlson (1980)

Ohlson's was the first study that used logistic regression instead of linear discriminant analysis for the estimation of credit risk. Ohlson found the following variables significant in the logistic regression equation

- Size
- Total Liabilities / Total Assets (TL/TA)
- Working Capital / Total Assets (WC/TA)
- Current Liabilities / Total Assets (CL/TA)
- Net Income / Total Assets (NI/TA)
- Pre-tax income plus depreciation and amortisation / Total Liabilities (FFO/TL)
- INTWO is one if Net Income was negative for the last two years, zero otherwise
- OENEG is one if owners' equity is negative, zero otherwise
- CHIN is the scaled change in net income [8]

The coefficients of the estimated model are shown in table 1.14 below.

Zavgren (1985)

This is another well cited study that used US data in the spirit of Ohlson. Zavgren found the following risk factors as significant in the logistic regression.

- Inventory /Sales
- Receivables / Inventory
- Cash / Total Assets
- Quick Assets / Current Assets (Acid test)
- Total Income / Total Capital

- Debt / Total Capital
- Sales / Net Plant

Zavgren;s study was influential because it demonstrated that logistic regression performs better than discriminant analysis. Compared to the studies based on discriminant analysis this study shows that the discriminating factors is a larger set compared to factors identified earlier using z-score type functions.

Table 1.14: Ohlson's results

Variables	Parameter Estmates
Constant	-1.320
Size	-0.407*
TL/TA	6.030*
WC/TA	-1.430**
CL/TA	0.076
NI/TA	-2.370**
FFO/TL	-1.830*
INTWO	0.285
OENEG	-1.720*
CHIN	-0.521*

* significant at 1% or lower (two-sided test)

** significant at 5% or lower (two-sided test)

Westgaard and van der Wijst (2001):

Westgaard and van der Wijst (2001) estimated a logistic model using data on bankruptcies of Norwegian companies for the period 1995-1999. The authors use the 1996 accounting data to predict bankruptcies in 1998. The results are shown in the following table

Table 1.15: Logistic regression results

Variables	Parameter Estmates
Cash/debt	-1.7698
Size	-0.1204
Liquidity	-0.2775
Finance cover	-0.00891
Solidity	-0.5807
Age	-0.3232
Realserv	-0.9530
Hotresta	0.2455
Mnorway	0.3631
Nnorway	0.2470

Where

[8] $CHIN = (NI_t - NI_{t-1}) / (|NI_t| + |NI_{t-1}|)$

Cash/debt	Operating income plus depreciation (cash flow) over total debt
Size	The base 10 logarithm of the firm's total assets.
Liquidity	Liquidity measured as current assets divided by current debt
Finance cover	Financial coverage measured as net results before financial costs divided by financial costs
Solidity	Solidity is measured as equity divided by total capital
Age	Age o company in years
Realserv	Dummy variable for the real estate and services industry
Hotresta	Dummy variable for the hotel and restaurant industry
Mnorway	Dummy variable for geography: Mid Norway
Nnorway	Dummy variable for geography: Northern Norway

The authors have used dummy variables to account for the differences in the bankruptcy rates in the different industries and regions. The performance of the model was evaluated by using only half the 1996 sample and using the other half as a "hold out" sample. The number of companies in their sample was 35, 287 of which 954 went bankrupt.

Lin and Pierce (2001)

More recently Lin and Pierce (2001) found the following ratios as significant in a conditional probability analysis approach to modelling default probabilities in the UK.

- After-tax Profit / Total Assets
- Retained Earnings / Total Assets
- Change in Cash / Total liabilities
- Working Capital / Total Assets
- Working Capital / Operating Expenditure

Foreman (2003)

Foreman' study applied logistic regression to a particular industry. More specifically, Foreman estimated logistic regression based on 1999 balance sheet to predict the bankruptcy of telecommunication companies two years later. He used two specifications (Specification 1 and Specification 2). The following variables were found significant in the two specifications. The main difference is whether the market value is significant in predicting the state of a company as it would be predicted by the Merton model.

Specification 1	Specification 2
Earnings per share	Earnings per share
Employees	Employees
	FCC pages to sales
	Market-to-book
Retained earnings to assets	Retained earnings to assets
Total debt proportion	
Working capital to sales	
Constant	Constant

Lo (1985)

In this study Lo although the primary objective was the development of a test framework for the comparison of the performance of the Logit and Discriminat Analysis models, proposing a Hausman - type specification test to chose the appropriate model, .he also presented interesting results. Lo estimated the following model using a matched sample of 77 failing and non-failing firms between the years 1975 and 1983 inclusively.

The following variable were used

Variable	Definition
SIZE	Log(total assets/GNP price deflator)
CLTA	Current debt liabilities, divided by total assets
OLTA	Other debt liabilities divided by total assets
CATA	Current assets, divided by total assets
NITA	Net income divided by total assets
BANK	Bankruptcy index suggested by Lo (1984)

The estimated coefficients of the two alternative models are given in table 1.16.

Table 1.16: Lo's results

	Logit estimate (Std. Error)	DA estimate (Std. Error)
Constant	1.2140	
	(2.51)	
SIZE	-0.0441	-0.0418
	(0.29)	(0.28)
CLTA	0.1074	-2.3803
	(2.42)	(2.02)
OLTA	-3.3258	-2.6131
	(1.81)	(1.51)
CATA	-1.2321	-0.6597
	(1.98)	(1.71)
NITA	10.7971	5.1616
	(4.82)	(2.25)
BANK	4.1642	1.0686
	(2.27)	(0.62)

Lo computed a Hausman test statistic to choose between the two models and found that the assumption of normality could not be rejected. On the basis of this test he chose the Discriminant Analysis model, since as Efron (1975) demonstrated, under the assumption of normality, the Discriminant Analysis estimates are more efficient than the Logit Model.

1.14 Probit Model

In the probit model the probability of a firm being in the downgraded group is given by

$$\pi_{1r} = \Pr(x = 1 / \mathbf{r} = \mathbf{r}^*) = \int_{-\infty}^{\beta' r} f(t)dt = \Phi(\boldsymbol{\beta}'\mathbf{r})$$

where $\Phi(\boldsymbol{\beta}'\mathbf{r})$ is the cumulative function of the standard normal distribution. Although there practical reasons for choosing the logistic model over the probit, there are no theoretical reasons that can be invoked for justification. The logistic distribution is similar to the normal except in the tails, which are

considerable heavier (in fact it resembles the t distribution with seven degrees of freedom). For intermediate values of $\boldsymbol{\beta}'\mathbf{r}$ (say between –1.2 and +1.2),the two distributions tend to give similar probabilities[9]. The logistic distribution tends to give larger probabilities to $\Pr(x = 0)$ when $\boldsymbol{\beta}'\mathbf{r}$ is extremely small than the normal distribution. The two models will produce different predictions of course when the sample contains very few observations of one of the two possible values of the random variable x or if one of the independent variables exhibited very wide variation in the values it attained.
Recent empirical applications of the probit model include Espahbodi and Espahdodi (2003) where the model is compared to a logit and a model based on discriminant analysis.

1.15 Survival Analysis

Survival Analysis assumes that both failed and non-failed firms in a sample are from the same population, with non-failed firms considered as some kind of censored observations. The risk of failure is measured by the survival time that is calculated for each firm. Assuming that T is the time on which a firm will fail, the survivor function S(t) is the probability that the time T is greater than t. The probability F(t) for a firm to fail before t is

$$F(t) = 1 - S(t)$$

the hazard function h(t) is given by

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

and

$$h(t/x) = h_0 \exp(\mathbf{x}'\hat{\mathbf{a}})$$

where **x** is the variable vector (vector of characteristics) and $\hat{\mathbf{a}}$ the coefficient vector. Vector $\hat{\mathbf{a}}$ which is the parametric part of the function is estimated by a technique similar to that of maximum likelihood; $h_0(t)$ the non parametric part is calculated by setting $\mathbf{x} = \mathbf{0}$.

The survivor function $S(t/x)$ is then

$$S(t/x) = S_0 \exp(\mathbf{x}'\hat{\mathbf{a}})$$

where $S_0(t)$ is given by

$$S_0(t) = \exp\left(-\int_0^t h_0(u)du \right)$$

As $h_0(t)$ is distribution free survivor analysis, is a semi-parametric methodology and it partially skips the criticisms against parametric techniques. Furthermore the method appears to be more natural in dealing with the failure problem.
Survivor analysis was employed by Lane *et al.* (1986), in the explanation of bank failure in the US and by Luoma and Latinen (1991) for the prediction of business failures in Finland. Their sample consisted of 36 industrial and retailing failed Finish companies, matched by size and industry type to 36 non-failed firms. The goal of the method was to calculate the survival time starting from the end of accounting period, assuming that at this time the failure process starts. This time is not necessarily a natural start of

---

9 see Green (2000) p 815

the failure process. This model consisted of financial ratios as well as a measure of the size and it performed satisfactorily compared to discriminant analysis and logit regression.

The interpretation of the results according to the expected failure time, provides decision makers with important information about a firm. The survival analysis method, although a viable alternative to statistical approaches has not been often applied for the prediction of business failure.

1.16 Recursive Partitioning

The recursive-partitioning algorithm is a non-parametric classification technique. The method starts with the sample of firms their financial characteristics, the actual group classification, the prior probabilities and the misclassification costs. A binary classification tree is built where a rule is associated to any mode. These are usually univariate rules; that is a certain financial characteristic and a cut-off point that minimise the cost of the misclassification for the rest of the firms. The risk of misclassification in any node t, q(t) is given by

$$q(t) = (c_{21} + c_{12})\pi_1\pi_2 \frac{1}{\pi(t)} \frac{m_1(t)m_2(t)}{n_1 n_2}$$

where

$m_1(t), m_2(t)$: The number of firms in each group on node t

n_1, n_2 : The total number of firms in each group (failed and non-failed firms)

c_{12} : Cost of misclassifying a firm in group 1 while in group 2

c_{21} : Cost of misclassifying a firm in group 2 while in group 1

$\pi(t)$: Probability of classifying a firm on node t

p_1, p_2 : Prior probability of a firm to be a member of group 1 or group 2.

After the classification tree is constructed, the risk of the final nodes and the risk for the entire tree is calculated. For the classification of any new firm, the firm descends the tree and falls into a final node that identifies the group membership for the specific firm and the associated probability. Breiman et al (1984) provide an extensive description of the method, including theory of binary trees, splitting rules, etc.

Frydman et al. (1985) first employed RPA as an alternative method to study the failure problem. The purpose of this study was to introduce RPA for the prediction of business failure and to compare resulting classification trees to models derived by discriminant analysis.

A sample of 58 bankrupt industrial companies and of 142 non-bankrupt manufacturing and retail companies was selected at random from the period 1971-1981. The RPA classification trees and two discriminant analysis models were constructed and compared in the study. The classification trees from RPA(1) and RPA(2) are shown in Figures 6 and 7 respectively.

The RPA trees, RPA(1) and RPA(2) were constructed for different costs of misclassification.

Figure 1.6: The RPA(1) Classification Tree

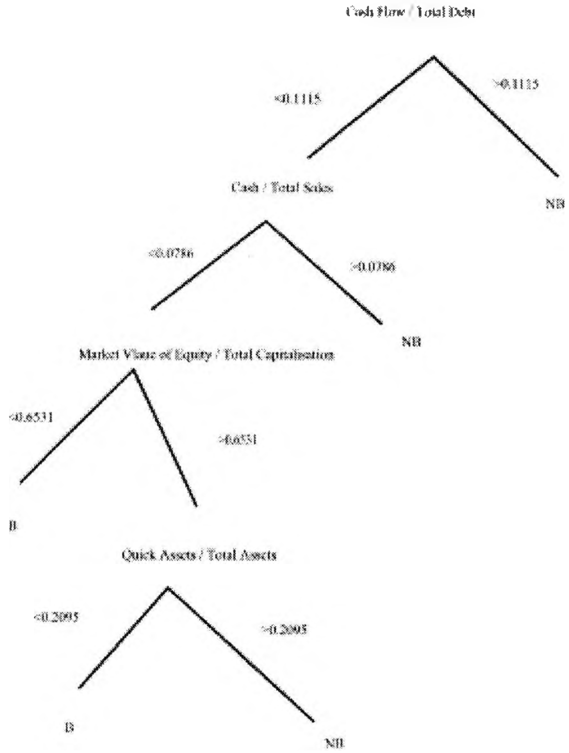
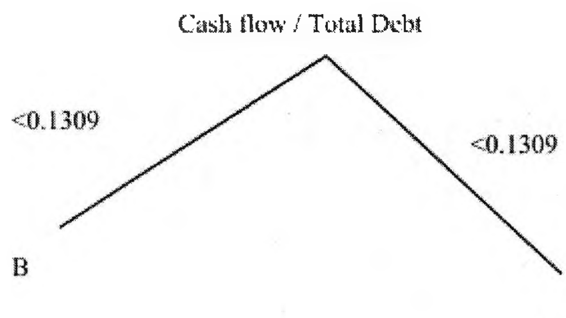


Figure 1.7 : The RPA(2) tree in Fydman, Altman and Kao (1985).



For RPA(1) it was assumed and for RPA(2) it was assumed . For both trees, the prior probabilities for bankrupt and non-bankrupt groups were set at $p_1 = 0.02, p_2 = 0.98$. From the various trees constructed according to the above parameters, the trees with the smallest cross validation costs were selected.

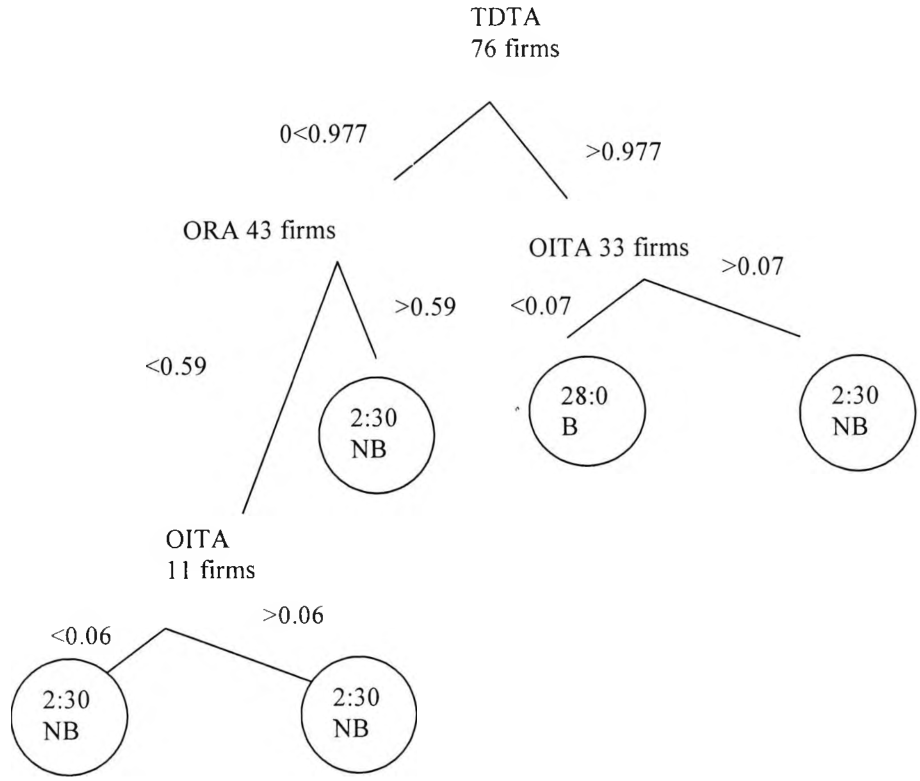
RPA has the characteristics of both the univariate and multivariate statistical methods. Classification on a node is made upon the rule of a single variable (although rules in the form of linear combination of financial characteristics can also be assigned to a node). On the other hand the method uses a sequence of nodes, i.e. a sequence of characteristics to classify a firm.

RPA is a forward selection method. It does not review previous classifications while it introduces new classification rules and this can result in a reappearing of the same variable to a later stage, with a different cut-off score. There is also a problem of overfitting, as continuation of partitioning process can result in a tree were each firm is classified by one terminal mode. To avoid such problems different trees of various degrees of complexity are derived and tested.

One of the main advantages of RPA is that the binary tree easily explains failure for a specific firm. This simplicity is eliminated if, instead of single variable rules, rules of linear combination of characteristics are used. In such cases, the resulting model can not easily explain failure. The estimation of the tree is also difficult. From a decision maker point of view, RPA just classifies firms into categories of risk. It does not permit comparisons between firms in the same category, making it very difficult to assess the relative performance of firms. Another weakness in RPA is that the technique does not provide an estimate of the probability of classification, although this problem can be overcome employing the technique suggested by Frydman et al.

The number of studies that have used RPA is rather small compared to the other methods. Michalopoulos et al. (1993) applied the method with good results to Greek data to predict bankrupt firms. Marais et al. (1984) applied RPA to commercial bank loan classification as did Srinivasan and Kim (1988). Another example of this approach is Laitinen and Kankaanpaa (1999) who estimated the following classification trees using Finish data.

Figure 1.8: Laitinen and Kankaanpaa (1999) Classification Tree



1.17 Other approaches to modelling credit risk

Recent developments in the credit risk area include the use of nonlinear statistical models such as neural networks. Instead of assuming a linear relationship between the risk factors and Z, neural networks allow more complex relationships to be present. The success of these models is however debatable and there is no consensus that they perform better than the traditional model based on either logistic regression or discriminant analysis.

1.18 Conclusions

In this chapter we have sought to accomplish two objectives. The first objective was to review the development of credit measurement techniques over the last 30 years. There has been a tremendous growth in the literature in the techniques that have been employed for the calculation of the fundamental elements of credit risk, such as the probability of default and expected loss at portfolio level. We have seen that the approaches to modelling credit risk have evolved since Beaver's original study and Altman's popularisation of quantitative techniques in many directions. Different and more complicated statistical models have been proposed to model the probability of default to microeconomic or macroeconomic factors, either for a single obligor or for a portfolio of obligors. The range of financial ratios used in the various studies has also grown although the review has revealed that there is a set of financial ratios that works well in all countries.

Most of the studies have employed one method for the assessment of credit risk and comparison of the various approaches over time and in different countries is not therefore easy. The Laitinen and Kankaanpaa (1999) comparison study provides some results on the performance of the various approaches for Finland. The results are interesting not least because they highlight that the best method differs according to the prediction horizon. The logit, for instance is the best method for the one year horizon, but it is relegated to fourth place for the 3 year horizon, with the survival method being a surprise top method for the two and three-year horizon. There is no question that much more research is needed in order to be able to draw definite conclusion on the performance of the various methods.

We have shown how to construct risk indices and to calculate the fundamental building blocks or credit risk management such as expected loss, unexpected loss and value at risk. However these portfolio approaches although important in terms of developing methodologies for aggregating credit risks, offer very little insight into the determination of the probability of default for a single loan. These approaches will not be employed in the rest of this thesis.

Table 1.17 : Methods ranked according to prediction accuracy

1 year		2 years		3 years	
I	Logit	I	Survival	I	Survival
II	NN	II	DA	II	NN
III	DA	III	Logit	III	RPA
IV	HIP	IV	NN	IV	Logit
V	RPA	V	RPA	V	DA
VI	Survival				

Source: Laitinen and Kankaanpaa (1999)

The second objective of this chapter is to position our thesis in this developing area both in terms of its research design as well as in terms of the contribution to the literature. This is achieved in the next two sections where the research design followed in this thesis is explained in view of the problems that most other studies have encountered and the nature of our data. The contribution of this thesis to the empirical literature on credit and market risk measurement is presented in the final section.

1.19 Research Design

One of the major problems in measuring credit risk and validating credit risk models is the lack of sufficiently large numbers of bankruptcies. In most cases a very small sample of companies that have gone bankrupt is employed. The declaration of a company as bankrupt may come a long time after the company is financially insolvent and that generates a problem in terms of relating financial characteristics to bankruptcy as what we want the data to predict is when a company becomes insolvent rather than when the company becomes legally bankrupt. The data, which we employ for this study, come from a sample of companies which have experienced problems in servicing their debt to a particular commercial bank in Greece for the period 1994-2001. We have followed the bank's own classification and methodology in classifying companies according to their ability to meet interest payments to the bank..

The bank classifies the companies in its loan portfolio in two broad categories depending on their interest payment experience. The two categories are described below.

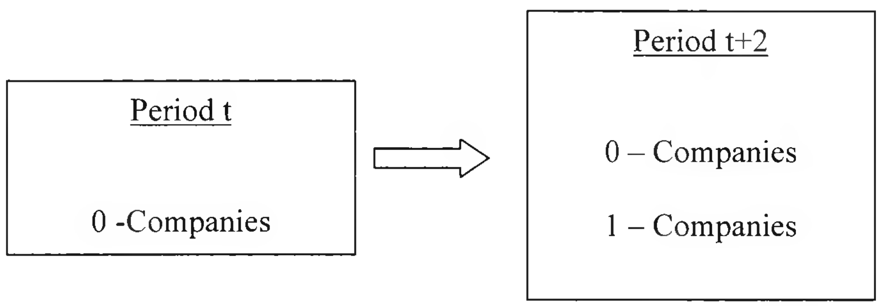
- Category 0 Companies: These are companies that fulfil all of the bank's credit criteria, have not delayed payments during the preceding year and have not had any major managerial or other non-financial problems. These companies are not expected to default over the duration of the loan.
- Category 1 Companies: These are borrowers which exhibit inability to service its debt to the bank. The degree of difficulties in servicing their debt in the preceding year will vary. The severity of the situation is measured by the delay in payments. A company is classified in category 1 once there is a delay of three months. The classification takes place after the credit event has happened and is not therefore based on any judgmental criteria. Some of the companies will resume normal payment at some time in the future whereas other companies will be led to bankruptcy. The key criterion for classification of a company as financially distressed is the delay in payment.

As we have already discussed the approach of this dissertation is to identify financial characteristics of a company that may predict the financial health of the company at some future date. There are three issues that arise from such a task. First we have to define the time horizon. There is no preferred time period over which such predictions are evaluated and the time horizon can be one year or four years. In our study we have opted for a time horizon of two years. The choice of the two-year credit event horizon was dictated by the size of the sample, since the number of companies that migrated from one category to another over a single year is rather small. The second reason we adopted the two-year credit event horizon was for the test to have some meaningful statistical predictive power. Given that we use financial ratios from published accounts of companies and given the delays in the publication of financial statements of between 6 and 9 months after the end of the financial year to which they refer, a one-year analysis would have amounted to forecasting where a company would have been in the past 6-9 months.

The second issue is to identify the set of financial characteristics that we want to use as explanatory variables. This is done in Chapter 2. The third issue is to find the combination of these factors so as to maximise predictive performance. This is done in Chapter 3 where we employ Multivariate Discriminant Analysis and Logistic Regression.

The testable hypothesis is then to identify factors which predict which of the companies will be downgraded in two years time. A graphic representation of the test design is given in the figure below.

Figure 1.9: Test Design



The sample of companies we are using in this study to test the predictive power of financial ratios comprises 492 companies in 1994, 546 companies in 1995, 474 companies in 1996, 403 companies in 1997, 375 companies in 1998 and 303 companies in 1999. The sample came from a population of about 14,000 companies and was selected on basis of available balance sheet and profit and loss data for the years under consideration. The table below shows the number of 0-category companies in the six base years (1994, 1995, 1996, 1997, 1998 and 1999) and their classification 2 years later in 1996, 1997, 1998, 1999, 2000 and 2001 respectively.

Table 1.18: The Sample of companies

	0-Category Companies in 1994 and their classification in 1996	0-Category Companies in 1995 and their classification in 1997	0-Category Companies in 1996 and their classification in 1997	0-Category Companies in 1997 and their classification in 1999	0-Category Companies in 1998 and their classification in 2000	0-Category Companies in 1999 and their classification in 2001
0	426	479	418	355	337	252
1	66	67	56	48	38	51
Total	492	546	474	403	375	303

For every year in our sample, say in 1994, we split the sample of 492 companies into two groups of 426 companies that remained in the 0 – category in 1996 and 66 companies that were downgraded in 1996. As we have already explained the idea behind this test is to see whether those 66 companies that were downgraded in 1996, exhibited different characteristics from those companies that were not eventually downgraded. The actual characteristics that will be employed to do this as well as the statistical techniques used will be investigated in chapters 2 and 3 of the thesis..

As the table makes clear most of the companies (about 88%) in the sample remain in the normal 0-category after two years. Given therefore the small number of companies that were actually downgraded we decided not to use a matched group of non- downgraded companies but instead to use the entire sample.

Unlike the previous studies on bankruptcy prediction our model is one that identifies companies whose credit quality is expected to deteriorate rather than companies that are expected to default. When comparing the results of this study with other studies on bankruptcy prediction, this should be taken into account since the characteristics of a company that is about to go bankrupt are different from those of a company that is merely downgraded. We would therefore expect a lower predictive power when compared with bankruptcy prediction studies.

1.20 Appendix

Table A1: Number of Firms failed (Hillegeist, Keating, Cram and Lundstedt, 2002)

Year	Number of Firms	Number of Bankruptcies	% Firms Failed
1979	2812	19	0.68%
1980	2816	21	0.75%
1981	2895	25	0.86%
1982	3015	17	0.56%
1983	3266	37	1.13%
1984	3333	49	1.47%
1985	3308	35	1.06%
1986	3473	32	0.92%
1987	3524	34	0.96%
1988	3374	34	1.01%
1989	3255	38	1.17%
1990	3211	27	0.84%
1991	3299	16	0.48%
1992	3423	23	0.67%
1993	3786	17	0.45%
1994	3987	19	0.48%
1995	4157	18	0.43%
1996	4554	24	0.53%
1997	4472	31	0.69%

Table A2: Most Common Financial Ratios

	WC/TA	*TD/TA*	*CA/CL*	*EBIT/TA*	*NI/TA*	*CF/TD*	*OA/CL*	*CF/S*	*RE/TA*	*S/TA*
Australia			2	1	1				1	
Canada		1	1							1
Finland		1	1	3	1	2	1	6		2
France	1					1		1		1
Greece	5	5	2	1	2				1	
Israel			1							
Italy										
Japan	1									
Sweden				1						
Netherlands										1
UK	5			1			5		1	
USA	4	8	5	5	7	6	3	1	3	2
Total	16	15	12	12	11	9	9	8	6	7

	S/TA	*GP/TA*	*NI/SE*	*CASH/TA*	*PBT/S*	*STP/TC*	*INV/S*	*QA/TA*	*TA/GNP*
Australia									
Canada	1								
Finland	2								
France	1								
Greece		6	1						
Israel									
Italy			1						
Japan				1					
Sweden							1		
Netherlands	1		1						
UK			3	2	5	5	2	1	2
USA	2			2			1	3	2
Total	7	6	6	5	5	5	4	4	4

Source: Dimitras, Zanakis and Zopounidis (1996)

Chapter 2 – Descriptive Statistics and Data Analysis

2.1 Introduction

The objective of this chapter is to define and describe the statistical properties of company characteristics, that is risk factors, that may help us predict whether a company will become financially distressed or not. The survey of empirical studies of the previous chapter has shown that most of the risk factors that have been used in discriminating bankrupt from non-bankrupt firms are financial ratios derived from the financial statements of a company. The number of financial ratios that have been used in the literature to describe the financial condition of a company is very large. In a review of existing practices, Chen and Shimeda (1981) identified more than a hundred ratios that had been employed in financial distress and bankruptcy studies. Such a vast array of possibilities is of course of little use in empirical analysis. We have therefore decided in this study, to narrow down the choice by looking at theoretical models to come up with a basic set of risk factors. Adding factors that have been found significant in other empirical studies then augments the theoretical set of risk factors.

The Merton (1974) model has identified some factors, such the leverage ratio, asset volatility and debt maturity that would explain the behaviour of companies but it is too simple to accommodate other factors that may impact on the behaviour of companies, such as liquidity and quality of management. Another source of information as to the performance of financial ratios is the published credit scoring models of rating companies.

We restrict the examination to two of the most important credit rating companies, namely Moody's and Standard and Poor's. There are other rating companies that used slightly different approaches but as these two companies are the dominant ones in the field, we felt that they provide a good indication of risk factors which are widely accepted at least implicitly as valid.

<u>Moody's Risk Factors (2000)*</u>

Moody's use the following set of seven factors in a model that combines discriminant analysis and logistic regression.

- Assets/Consumer Price Index
- Inventories/Cost of goods sold
- Liabilities/Assets
- Net Income Growth
- Net Income/Assets
- Quick Ratio
- Retained Earnings/Assets
- Sales Growth
- Cash/Assets
- Debt Service Coverage Ratio

Standard and Poor (2000) **

Standard and Poor report the following eight factors as important in terms of classifying issuers of corporate bonds.

- EBIT interest coverage
- EBITDA interest coverage
- Funds flow/total debt
- Free open cash flow/total debt
- Return on capital
- Operating income/sales
- Long-term debt/capital
- Total debt/capital

Both agencies employ ratios that are consistent with the basic Merton model but include ratios that capture aspects outside the model's framework such as short-term solvency, growth and efficiency ratios. The Chapter deals with two issues. First, we define characteristics of companies that should explain why companies migrate from one credit category to another and predict which of the companies may exhibit a high probability of migration. We deal with this issue in the first part of this chapter, Sections 2.2 and 2.3. The second issue we deal with in this chapter is to test statistically whether companies that were downgraded exhibited financial characteristics, two years before the migration, which were different to those companies that did not migrate. The question we examine is as follows: Does the distribution of financial ratios discriminate between companies that may be downgraded and companies that may remain in the same criteria category?

Our general approach in dealing with this second issue is by testing whether those of the 0-classified companies in a particular year (year t) that were downgraded within the following two years exhibited different characteristics from those companies that were not downgraded. The design of the test is stated in Chapter 1. We perform six statistical tests:

- Univariate Tests of equality of the means of the ratios in the two groups
- Multivariate tests of the equality of the vector of means in the two groups of companies
- Tests of equality of the medians in the two groups
- Univariate Tests of equality of variances of the financial ratios in the two groups of companies
- Multivariate Tests of equality of the covariance matrices in the two groups.
- Tests of normality of the distribution of financial ratios.

2.2 Company Characteristics and Financial Distress

In describing the state of company researchers normally rely on financial characteristics, as well as on characteristics such as size, quality of management and industry structure. It is possible of course that good management and monopolistic power in an industry would ultimately be reflected in financial performance, however, the timing issues may be important and therefore the effect of management

---

* Moody's Investor Service, Rating Methodology for Private Companies May 2000

** "Adjusted Key U.S. Industrial Financial Ratios" Standard and Poor's Credit Week, 20th September 2000.

change which is measured today may not be fully reflected within the time frame of our analysis. Thus one may want to investigate differences in companies in terms of non-financial characteristics.

The fundamental factors that may cause corporate failure have been studied widely and are classified by Argenti (1976) and Dambolena and Khouri (1980) as internal factors and external factors. The main internal factor is bad management manifested through , lack of responsiveness to change in technology, bad communication, misfeasance and fraud, insufficient consideration for cost factors, poor knowledge of financial matters and high leverage position. The main external factors are effect of labour unions where too high a wage settlement causing the firm to pay its employees in excess of their marginal product, government regulations which impede, in some instances, the functioning of the market system distorting in the process its signals to the corporate decision makers, and natural causes such as natural disasters, demographic changes, etc. All the above causes may or may not be reflected on the balance sheet in timely or accurate fashion for use by the research.

So in addition to the analysis of the information contained in the financial statements, non-financial factors exist that may modify the evaluation of the company. Some of these non-financial factors are:

- Foreign exposure
- Quality of management
- Ownership structure

Foreign exposure is important for exporting or importing companies, as part of the revenue or expenses may be influenced by events outside Greece. Foreign currency exposure is another possible source of risk, especially if the publication of information on hedging currency risk is not obligatory.

The quality of management and the depth of administrative structure are more difficult to evaluate. Positive aspects of management will be reflected in a steady growth pattern. Negative aspects would include a firm founded and headed by one person who is approaching retirement and has made no plan for succession. Equally negative is the firm that has had numerous changes of management and philosophy. On the other hand excessive stability is not always desirable. Characteristics of a good management team include depth, a clear line of succession if the chief officers are nearing retirement, and a diversity of age within the management team.

Ownership of the firm should also be considered as a factor. If one family or group of investors owns a controlling interest in a firm, they may be too conservative in reacting to changes in the market. Owners should also be judged in terms of whether they are strategic or financial. Often financial buyers invest for the short to intermediate term, hoping to sell their positions (or the entire company) at a profit. In such a case a company may not plan for longer-term growth and it may be in its interest to boost short-term growth at the expense of long term prospects.

A second issue that faces the investigator is to determine when a financial characteristic of a company deviates from the norm. In other words we need a norm in order to compare these characteristics to other companies or to the industry. For example, a financial characteristic of a company such as sales growth may appear attractive on its own for a company, but for an industry it may not be. There are many industry considerations that should be taken account of such as:

- Economic cyclicality
- Growth prospects

- Research and Development expenses
- Competitors
- Sources of supply
- Degree of regulation
- Labour relationships
- Accounting policies

An additional degree of complexity is added because these factors should be considered in global context. Companies in the tourist industry are a good example, although the elimination of trade barriers seems to have increased the competitive pressures on all sectors of an economy. The economic crisis in the Far East, the Russian bond crisis, Oil price volatility, all constitute factors that make the economic analysis of a company from the view point of domestic market seriously deficient.

2.3 Financial Ratios

The financial ratios employed in this study are grouped into 6 broad categories, which reflect aspects of the financial conditions of a company. The groups of ratios are:

1. Profitability ratios
2. Liquidity ratios
3. Leverage ratios
4. Solvency ratios
5. Activity ratios
6. Growth ratios and size

Some important variables such as free cash flows could not be used because of lack of data on capital expenditure. Any ratios that would be important require cash flow data and market value, which were not available in our data set, are therefore not employed in this study.

Profitability Ratios

Higher profitability will normally raise a firm's equity value. It also implies a longer way for revenues to fall or costs to rise before losses occur. There is a number of profitability ratios that can be defined depending on the profitability measure employed. We have employed two ratios to measure the impact of profitability, the Return on Equity (ROE)

$$ROE = \frac{\text{Operating Profit}}{\text{Share Capital}} = \frac{AV}{AD}$$

and return on total assets.

$$ROA = \frac{\text{EBIT}}{\text{Total Assets}} = \frac{AR + AS - AU + BA}{AB}$$

Return on equity is crucial to the shareholder. Anticipated year over year return on equity is a key driver of the share price, and one of management's major goals is maximizing the shareholders' value over the long term. Return on assets is independent of financing, but return on equity is strongly influenced by financing. Highly leveraged companies (those with a high debt ratio) can expect to have wider swings (more volatility) in their return on equity. Table 2.1 shows the average values of this ratio in each year for

Category 0 companies that were still in that category two years later (Group 0) and companies that had been downgraded in the following two years (Group 1). The average of the whole sample is also given.

Table 2.1: Return on Equity

Group	94	95	96	97	98	99
zero	0.471	0.520	0.436	0.431	0.442	0.344
One	0.198	0.194	0.056	0.130	0.245	0.243
Total	0.598	0.444	-0.071	0.494	0.402	0.218

As with other ratios, some variation is possible. The formula above gives return on all equity (for example, common and preferred shares). Some analysts extract the cost of servicing the preferred shares from the numerator (i.e. they subtract the preferred share dividends from net income), and make the denominator the common share equity. This gives a return on common shares only.

Another issue is whether profitability is best measured relative to equity, assets or some other variable such as sales. Each measure is more appropriate for different types of companies. We have tried therefore to capture the effect of profitability by using a variety of ratios

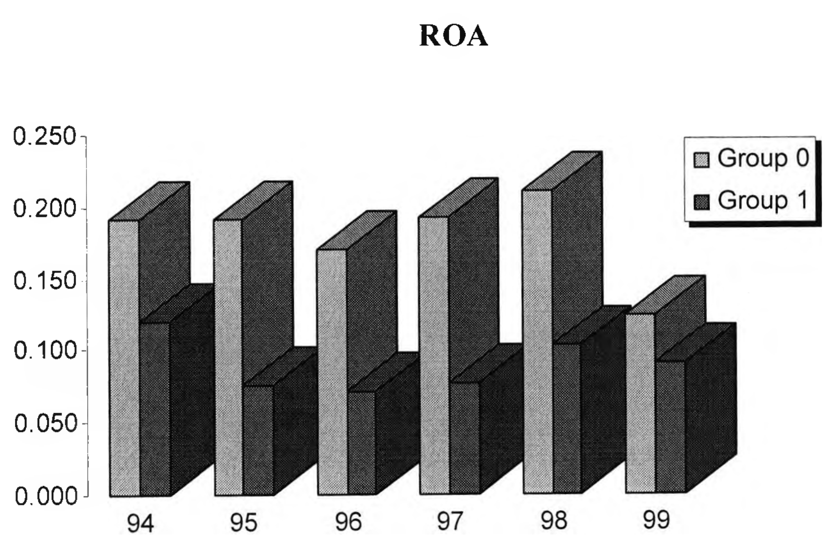
The ROA is a measure of what a company earns on all of its assets, whether they are financed by debt or shareholder equity. This ratio, again always expressed as a percent, is a powerful tool for comparing between similar companies. Table 2.2 and the Figure that follows show the average values of this ratio for the two groups for every year of our sample. As it was expected, the ROA of companies in Group 0 are consistently higher than those in Group 1, most of the time twice as high.

Table 2.2: ROA

Group	94	95	96	97	98	99
zero	0.192	0.192	0.171	0.193	0.211	0.124
One	0.120	0.075	0.071	0.077	0.103	0.090
Total	0.182	0.177	0.159	0.179	0.200	0.118

The following graph shows the Return on Assets for the two group of companies for each year.

Figure 2.1 ROA



An important feature of the table and the graph is the relative stability of the ROA especially for the group zero companies, that is the companies that were not downgraded.

Leverage Ratios

The second group of risk factors are leverage ratios. The higher the leverage the larger the cost of capital for the company and the smaller the ability of the company to accommodate adverse profitability shocks. We have used two leverage ratios. Total Debt to Total Assets (DR) defined as

$$DR = \frac{Total\ Debt}{Total\ Assets} = \frac{AG + AH}{AC + AG + AH}$$

and Total Debt over Paid Up Capital

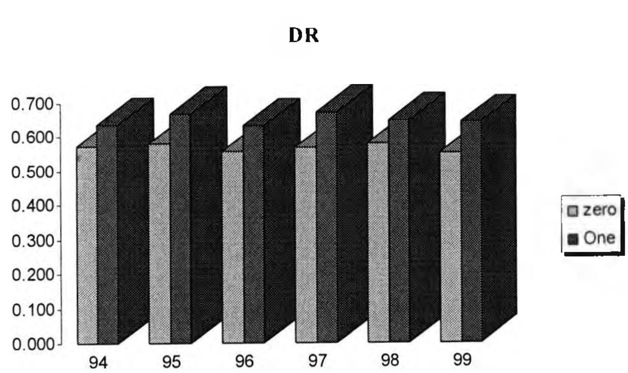
$$DC = \frac{Total\ Debt}{Paid\ up\ capital} = \frac{AG + AH}{AD}$$

Debt ratio is a critical number for both lenders and investors. Lenders have a strong focus on the security of a debt, i.e. its likelihood of repayment. To a lender, a high debt ratio increases the risk that the ongoing company will not service the debt, or that the debt cannot be recovered if the company fails. An investor would have the same concern if the debt ratio was very high: if the company can not meet its debt payment schedule, the lenders will push the firm into bankruptcy and there will be nothing left for the equity holders. However, investors often don't like a very low debt ratio, because their return is not "leveraged". Debt also increases the volatility of the return on equity generated by a company. This effect is called leverage, and is an important concept in understanding a company's financial performance. The average values of the DR ratio for each of the groups are shown in Table 2.3 and the accompanying graph. Again as it was expected, the Group 1 companies have higher leverage ratios that the companies that did not migrate (Group zero).

Table 2.3: DR

	zero	One	Total
94	0.574	0.635	0.582
95	0.582	0.669	0.592
96	0.559	0.633	0.568
97	0.570	0.672	0.582
98	0.583	0.648	0.589
99	0.553	0.643	0.568

Figure 2.2 DR



From the table and the graph we see that the ratios remain relatively stable throughout the period. Ratio stability is an important issue in the use of financial ratios for the prediction of financially distressed companies.

Solvency Ratios

Some kind of solvency ratio has been used in all the empirical studies we have reviewed in the previous studies as it is accepted that the larger the burden on a company from servicing its loan the larger its probability of default. There is not a single accepted way of measuring this risk, so we have calculated three ratios to be used in the analysis. The first one is Interest Payments over Cash Flow (IPF) defined as

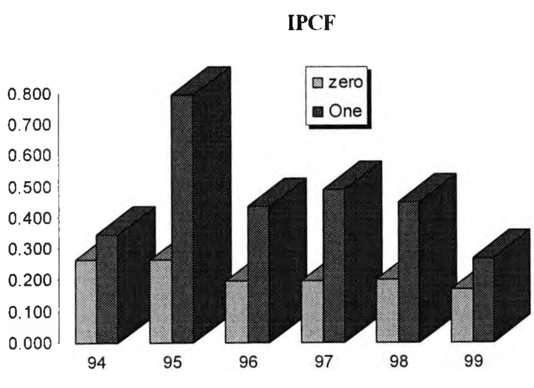
$$IPF = \frac{\text{Interest Payments}}{\text{Gross Profit} + \text{Depreciation}} = \frac{AT}{AR + BA}$$

This ratio is the inverse of times interest earned which is one of the most important criteria for setting the interest rate on debt. Times interest earned looks at earnings with interest and income taxes added back in. Interest payments are added back in because they are earnings that are available to pay interest. The theory for adding taxes is that if a company becomes unprofitable its tax bill goes to zero (since taxes are calculated as a percentage of net earnings or profits), and the earnings that had been going to pay taxes can now be used to service debt. The average values of this ratio are shown in Table 2.4 and the graph that follows.

Table 2.4: IPF

Group	94	95	96	97	98	99
zero	0.264	0.264	0.198	0.199	0.200	0.169
One	0.346	0.796	0.434	0.488	0.450	0.271
Total	0.275	0.329	0.226	0.234	0.226	0.186

Figure 2.3 IPCF



The proportion of cash flow devoted to interest payments has decreased steadily reflecting the significant reduction in interest rates during the period. This is particularly true for companies that were not downgraded eventually. There is no theoretical guidance as to what is an appropriate benchmark value. As a rule of thump companies for which EBIT is more than five times its interest are considered to be very good credit risk, and bonds of such companies have a high grade (rating). Applying this rule of thump in our definition, high grade companies should have IPCF values of less than 0.2, which is actually the case in our sample apart from 1994 and 1995.

The second ratio employed is Interest Payments over Sales (IPS) defined as

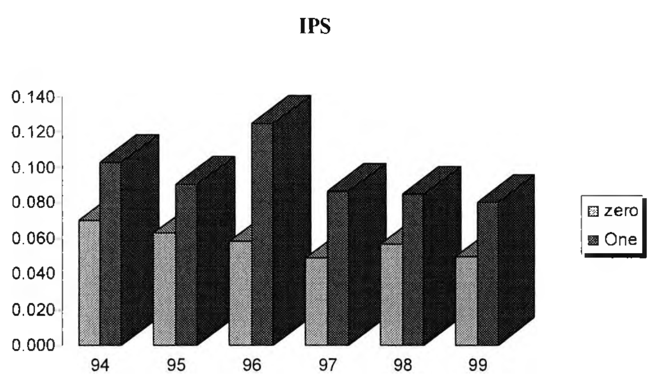
$$IPS = \frac{Interest\ Payments}{Sales} = \frac{AT}{AP}$$

Sales are less amenable to manipulation than either cash flow or EBITDA and therefore a more objective measure. However a high sales volume may not necessarily be translated into a better position to meet the debt servicing obligations of a company. The average values are shown in Table 2.5 and in the accompanying graph.

Table 2.5: IPS

	zero	One	Total
94	0.070	0.103	0.074
95	0.063	0.090	0.066
96	0.058	0.125	0.066
97	0.049	0.087	0.053
98	0.057	0.085	0.060
99	0.049	0.080	0.054

Figure 2.4 IPS



This ratio follows the same downward trend as IPCF reflecting the reduction in interest rates during the period.

The third ratio employed is Interest Payments over EBITDA[1], that is this time we calculate the interest coverage on EBITAD. This last variable is closely related to the interest coverage on Cash Flow, but it includes other items on top of AR and BA.

$$IPE = \frac{\text{Interest Payments}}{EBITDA} = \frac{AT}{AR + BA + AS - AU + AW - AX}$$

Liquidity Ratios

[1] *As lease financing has become more popular, it may be prudent to move from "times interest earned" to factoring in all fixed (non-avoidable) charges. Sometimes the before tax income required to fund a "sinking fund" is also included in the denominator. Financial statements such as the ones in the ICAP database from which the financial data are drawn don't have enough detail to complete the fixed charge coverage, because details on leasing payments are not included.*

Liquidity is considered an important variable for credit rating decisions. Sufficient current assets allow a company to meet its current liabilities and to reduce its requirements for working capital. We have used two ratios in this category. The Current Ratio defined as

$$CR = \frac{Current\ Assets}{Current\ Liabilities} = \frac{R+V+AA}{AH}$$

and Working Capital over Total Assets (WA)

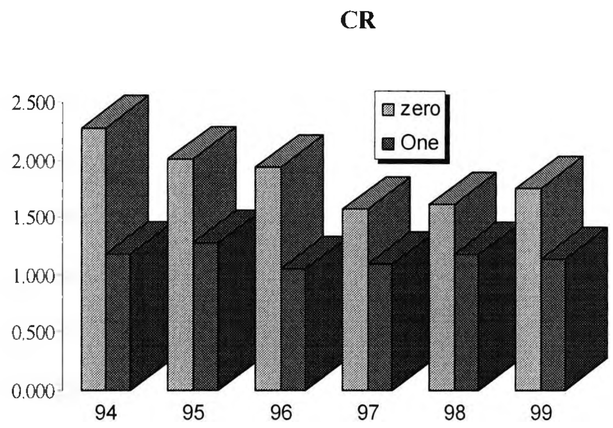
$$WA = \frac{V+R+AA-AH}{AB}$$

The current ratio for the two groups is shown in the graph that follows and in Table 2.6

Table 2.6: CR

	1994	1995	1996	1997	1998	1999
zero	2.275	2.003	1.931	1.576	1.614	1.753
One	1.176	1.278	1.049	1.099	1.179	1.142
Total	2.127	1.914	1.827	1.519	1.570	1.650

Figure 2.5: CR



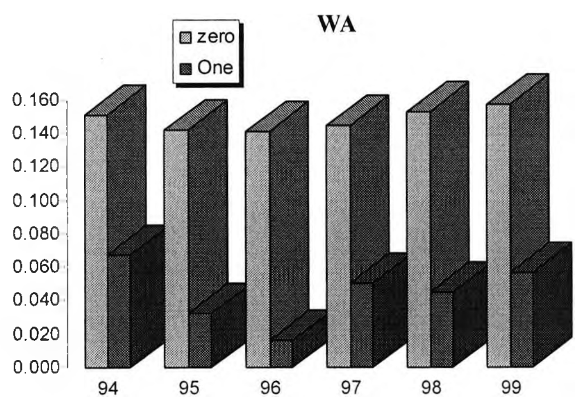
The current ratio is of critical importance to a short-term lender because the lender normally is secured by current assets only, i.e. the fixed assets are pledged against long-term debt and the short-term lender can't seize them if the firm goes bankrupt. A current ratio of less than one says that for every dollar of readily accessible ("liquid") assets, such as inventory and receivables, there is more than a dollar of debt owing to a creditor. This is viewed as a risk by a lender except in unusual circumstances. Short-term lenders will frequently have a covenant that requires the business owners to maintain or exceed a set current ratio or to have an absolute amount of working capital (current assets minus current liabilities). For large companies with many channels for raising funds a high current ratio is more a sign of incompetent cash management than prudence. That relationship does not exist however for smaller companies that cannot tap the direct lending market.

Table 2.6 and the accompanying graph show the working capital over assets ratio (WA) for the two groups and for each year. This ratio is remarkably stable for the companies that did not experience financial distress. For the companies that were eventually downgraded this ratio is much smaller and more volatile.

Table 2.7: WA

	zero	One	Total
94	0.151	0.067	0.139
95	0.142	0.033	0.128
96	0.141	0.016	0.127
97	0.145	0.050	0.133
98	0.153	0.045	0.142
99	0.157	0.057	0.140

Figure 2.6 WA



Activity and Efficiency Ratios

Activity and Efficiency ratios measure how intensively a company uses its assets. The more efficiently a company uses its assets the higher the return on the assets for given utilization costs and ceteris paribus the larger the ability of a company to withstand shocks. The use of these ratios in predicting default probabilities is problematic because they are not suitable for comparison across industries and it is not surprising that activity and efficiency ratios tend to be only weakly associated with the probability of default. The following ratios have been suggested in the literature as relevant proxies of efficiency and are used in this study.

- Sales over Total Assets

$$SOA = \frac{Sales}{Total\ Assets} = \frac{AP}{AB}$$

- Operating Profit over Sales

$$OS = \frac{Operating\ \ Profit}{Sales} = \frac{AR + AS - AU + BA}{AP}$$

- Net Income over Sales

$$NIS = \frac{Net\ Income}{Sales} = \frac{BB - BN}{AP}$$

- EBITDA over Sales

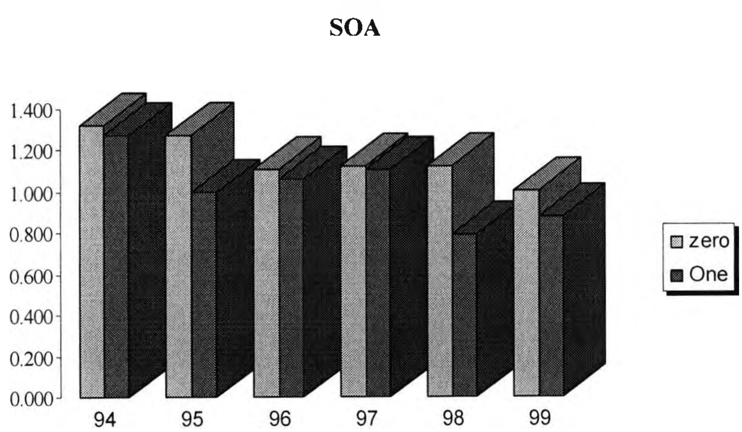
$$ES = \frac{EBITDA}{Sales} = \frac{AR + AS - AU + BA + AW - AX}{AP}$$

The first ratio total asset turnover measures, in theory, the extent to which management is using the assets of a company efficiently and can be used to compare two companies within an industry. This ratio could however be meaningless for comparing companies from different industries, since for example, a retail chain will have a huge turnover (low assets and high sales), whereas a highly capital intensive project will be the opposite. This ratio can also have problems within an industry that makes its use highly problematic. The accounting framework reflects the principle of prudence and assets have rarely their value increased to reflect inflation; in essence, an asset is depreciated from its original cost, with no reflection of inflation[2]. Hence, if one were to compare two companies, one with old assets and one with newer assets, there would be a depreciation and inflation impact that would distort this ratio. Total asset turnover ratios are therefore of limited value in inter-company company comparisons.

Table 2.8: SOA

	94	95	96	97	98	99
zero	1.318	1.273	1.106	1.124	1.124	1.004
One	1.276	0.997	1.059	1.108	0.792	0.884
Total	1.312	1.239	1.101	1.122	1.090	0.983

Figure 2.7 SOA



The other three ratios measure short term profitability on turnover and measure the efficiency with regard to turnover as opposed to assets. This would be a more appropriate ratio for comparing companies from industries with different capital requirements. The three ratios vary in the way the numerator is defined. OS uses as the numerator net operating income before income tax and interest (to show the yield of the business before financing and tax related charges that distort inter-company comparisons). NIS looks at an after-tax and interest number, and defines the numerator as "net income available to common shareholders". Finally, ES uses EBITDA as the numerator, which includes depreciation in the calculation, and is as close as we can get to a proxy for the cash flow of a company.

Growth Ratios and Size

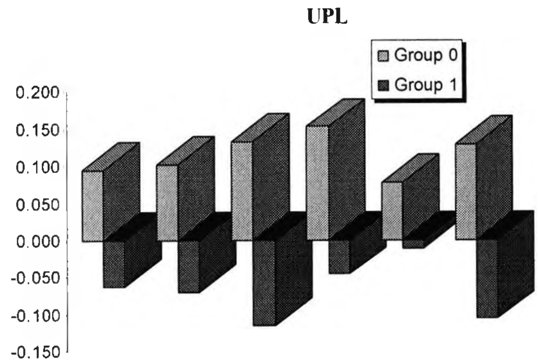
[2] *Revaluation of fixed assets does take place periodically*

The final set of ratios measure the financial health of a company and its potential for growth as well its potential to withstand adverse conditions. We use two related ratios here. The first one is undistributive profits & accumulated losses over paid up capital (UPL) defined as

$$UPL = \frac{\text{AF}}{\text{AD}}$$

The larger the proportion of undistributive profits & accumulated losses to paid up capital the larger the ability of the company to withstand adverse developments. Also, this can be interpreted as a growth rate since the larger the profitability of the company, the larger this ratio is.

Figure 2.8 UPL



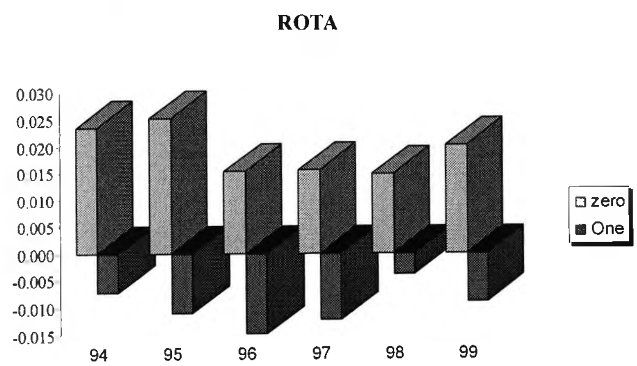
As the plot of the values of UPL shows for the two groups of companies, companies that experience financial distress two years later had a negative UPL today.

The second ratio is retained profits over total assets defined as

$$\text{RTA} = \frac{\text{BB} - \text{BN} - \text{BF}}{\text{AB}}$$

this is one of the most popular ratios in discriminant analysis as the survey of literature in Chapter 1 indicated.

Figure 2.9: ROTA



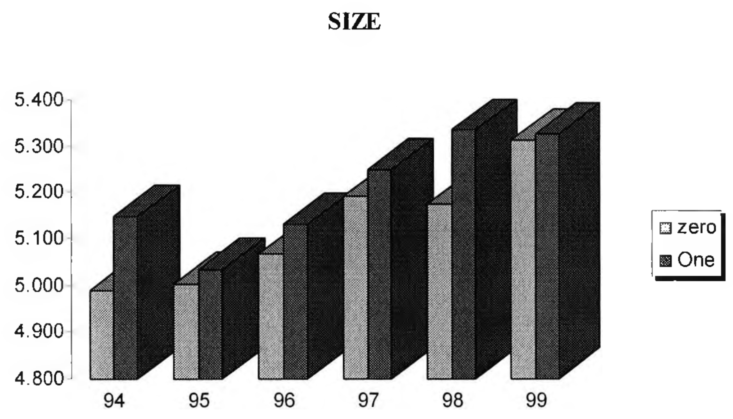
Size

The last ratio we employed was a proxy for the size of the company. There is no single, universally accepted proxy and we have therefore used the natural logarithm of total assets as a proxy.

$$\text{SZ} = \log(AB)$$

Are small companies more likely to be downgraded than larger companies? In a number of empirical studies, this hypothesis has been tested and found that size does matter.

Figure 2.10: SIZE



However, our sample estimates do not support the view that smaller companies are more likely to be downgraded than larger ones as the analysis of the following chapter shows. From the graph we see that the downgraded companies on average had more assets than those that were not migrated.

This completes the description of the financial ratios employed in this study and their broad movement over time. A summary of the ratios together with and explanation of the terms in terms of the format of the Balance Sheet and Profit and Loss Account are given in Tables 2.9, 2.10 and 2.11. The notation adopted is follows the one used by Dinenis (2003a).

Table 2.9: profit and loss account

PROFIT AND LOSS ACCOUNT ITEM	SYMBOL
Turnover	AP
Cost of Goods sold	AQ
Gross profit	**AR**
Commissions and other operating expenses	AS
Financial expenses	AT
Other Operating Expenses	AU
Operating Profit	**AV**
Non operating income	AW
Non operating expenses	AX
Total Depreciation	**AZ**
Depreciation included in the cost of goods sold.	BA
Depreciation not included in the cost of goods sold.	AY
Profit before tax	**BB**
Tax	BN
After –tax profit	
Dividends	BF
Retained profits	RT

Table 2.10: Balance sheet

BALANCE SHEET ITEM	SYMBOL
Net fixed assets	**I**
Land	J
Buildings	K
Other fixed assets	BH
Good will	L
Long –term Credit	P
Investment in Subsidiaries	Q
Stocks	**R**
Finished goods	S
Semi-processed	T
Raw materials	U
Non fixed assets	**V**
Receivables	W
Financial investments	Y
Other claims	Z
Cash and bank balances	AA
Accumulated depreciation	**O**
Depreciation of Machinery and equipment	BI
Depreciation of buildings	BO
Other depreciation	BP
Total assets	**AB**
Share capital	**AC**
Paid-up capital	AD
Reserves	AE
Undistributed Profits – accumulated losses.	AF
Medium and long-term liabilities – provisions	**AG**
Medium and long-term liabilities	BK
Provisions	BL
Short-term liabilities	**AH**
Bank Loans- Interest on Long term loans	AI
Suppliers – Creditors – Accounts payables	AJ
Payable dividends	AK
Share accounts	AL
Other liabilities	AM
Total liabilities	**AN**

Table 2.11: Summary Definitions of Financial Ratios employed in the study by category

Profitability Ratios		
ROE	Return on equity	AV/AD
ROA	EBIT / total assets	(AR+AS-AU+BA)/AB
Solvency Ratios		
IPCF	interest payments /cash flow	AT/(AR+BA)
IPS	Interest payments/ sales	AT/AP
IPE	Interest payments / EBITDA	AT / (AR+BA+AS-AU+AW-AX)
Leverage Ratios		
DC	total debt / paid up capital	(AH+AG)/AD
DR	Total debt/ (total debt & total equity)	(AG+AH)/(AC+AG+AH)
Liquidity Ratios		
CR	current ratio	(R+V+AA)/AH
WA	Working capital/ total assets	(V+R+AA-AH)/AB
Activity Ratios		
SOA	Sales/ total assets	AP/AB
OS	Operating profit / sales	(AR+AS-AU+BA)/AP
NIS	Net income / sales	(BB-BN)/AP
ES	EBITDA/ Sales	(AR+AS-AU+BA+AW-AX)/AP
Growth Ratios		
UPL	(undistributive profits & accumulated losses)/ paid up capital	AF/AD
RTA	Retained profits/ total assets	(BB-BN-BF)/AB
SZ	Size	Log(AD)

2.4 Testing the Equality of Means – Univariate Tests

The first test we perform is to test whether the mean values of the financial ratios of the companies that migrated were different (statistically) from the mean values of the financial ratios of the companies that did not migrate. Let $\bar{r}_{j1}$ be the mean of ratio j at time t for a company that remained k in t+2 and $\bar{r}_{j2}$ be the mean of ratio j at time t for a company that migrated in period t+2. The univariate test is

$$Ho: \bar{r}_{j1} = \bar{r}_{j2}$$

$$H_1: \bar{r}_{j1} \neq \bar{r}_{j2}$$

and where the means have been calculated as follows

$$\bar{r}_{j1} = \frac{1}{n_1}\sum_{i=1}^{n_1} r_{j1i} \text{ and } \bar{r}_{j2} = \frac{1}{n_2}\sum_{i=1}^{n_2} r_{j2i}$$

the hypothesis of equal means was tested using the t-test given by

$$t_j = \frac{\bar{r}_{j1} - \bar{r}_{j2}}{s_j}$$

where s_p is the pooled standard deviation of the two samples.

The arithmetic means for each financial variable for the years 94, 95, 96, 97, 98 and 99 and for each of the two categories is shown in the Data Appendix to this chapter. Table 2.12 shows the values of the T-test for testing the hypothesis that the mean values of the ratios between the two groups of companies are the same.

Table 2.12: Value of the t-statistics for testing the equality of the means in the two groups of companies

Ratio	94-96	95-97	96-98	97-99	98-00	99-01
(Undistributive profits & accumulated losses)/ paid up capital	1.663	2.056	2.369	2.699	0.574	1.430
Total debt / paid up capital	0.174	-0.609	-0.851	-1.391	-1.581	-1.081
Current ratio	2.270	2.233	-0.319	4.830	1.859	1.539
Return on equity	2.520	3.027	2.954	4.013	1.117	1.564
Interest payments /cash flow	-0.715	-1.395	-3.507	-2.999	-5.083	0.245
Operating profit / sales	1.192	1.052	0.733	1.012	-0.350	-0.232
EBIT / total assets	1.982	6.712	6.527	4.709	2.218	2.155
Net income / sales	1.836	1.320	1.243	2.320	-0.008	2.757
EBITDA/ Sales	1.196	1.183	0.670	1.506	-0.427	-0.234
Retained profits/ total assets	3.480	5.103	4.457	4.071	2.094	2.519
Sales/ total assets	0.187	3.335	0.366	0.090	2.982	2.268
Working capital/ total assets	3.700	4.548	4.721	3.736	3.348	2.918
Interest payments/ sales	-1.940	-2.727	-3.153	-3.547	-1.929	-3.439
Interest payments/ EBITAD	-1.835	-3.295	-1.967	-1.673	1.350	1.661
Total debt/ (total debt & total equity)	-2.211	-3.612	-2.683	-3.556	-1.853	-1.963
Size	-1.863	-0.353	-0.920	-0.814	-1.407	-1.346

From the tables in the Data Appendix, that give the values of the means for the two categories, one can observe significant differences of the average values between the two groups for the overwhelming majority of the discriminator variables and this is reflected in the values of the t –statistics in the table which statistically significant for more that 2 years for most of the 16 ratios. Most of the profitability and leverage ratios are significant but the liquidity and activity ratios less so. Size is not significant as a factor indicating large and small companies have the same probability of being downgraded. One can safely conclude that on the whole, the companies that were downgraded had different average values for their ratios.

2.5 Testing the statistical differences in the vector of means

The univariate test of the equality of means overwhelmingly supports the hypothesis that the average value of most financial ratios differed between migrating and non-migrating companies. However the univariate test suffers from the problem of ignoring interrelationships between the various ratios and consequently its power is lower. In order to overcome this problem we shall therefore employ a series of tests that test the hypothesis that the vector of means in the two groups differ.

The population means of the two samples will be denoted by $\bar{\boldsymbol{\mu}}_1$ and $\bar{\boldsymbol{\mu}}_2$. To test whether the two samples come from the same population we have to test the hypothesis $Ho: \bar{\boldsymbol{\mu}}_1 = \bar{\boldsymbol{\mu}}_2$ against the alternative $H_1: \bar{\boldsymbol{\mu}}_1 \neq \bar{\boldsymbol{\mu}}_2$ Where, $\bar{\boldsymbol{\mu}}_i = (\bar{\mu}_{i1}, \bar{\mu}_{i2}, ..., \bar{\mu}_{ip})'$ for $i = 1,2$, denotes the mean vector of population i.

Let $\mathbf{r}_i = (r_{i1}, r_{i2}, ..., r_{ip})'$ be the vector of p ratios in group i and let $\bar{r}_i = (\bar{r}_{i1}, \bar{r}_{i2}, ..., \bar{r}_{ip})'$ for $\mathrm{i} = 0,1$, denote the sample mean vector of group i, where the mean of each ratio has been calculated over all the companies in the particular sample.

The null and alternative hypothesis for testing the sample means becomes,

$$\mathrm{Ho}: \bar{\mathbf{r}}_1 = \bar{\mathbf{r}}_0 \quad \text{and} \quad \mathrm{H}_1: \bar{\mathbf{r}}_1 \neq \bar{\mathbf{r}}_0$$

There are a number of test statistics that can be employed to test the above hypothesis. It is intuitively appealing to use a generalisation of the univariate t-test for the and such a test is Hotelling's T^2 test defined as

$$\mathrm{T}^2 = \frac{\mathrm{n}_1 \mathrm{n}_0}{\mathrm{n}_1 + \mathrm{n}_0} (\bar{\mathbf{r}}_1 - \bar{\mathbf{r}}_0)' \hat{\Omega}_{\mathbf{p}}^{-1} (\bar{\mathbf{r}}_1 - \bar{\mathbf{r}}_0)$$

where $\hat{\boldsymbol{\Sigma}}_{\mathrm{p}}$ is the sample variance covariance matrix. An alternative statistic is Wilks' lambda, defined as [3]

[3] Other test statistics can be defined in terms of Hotelling's T^2 statistic the most common of which are: Hotelling's trace $\mathrm{H} = \frac{1}{\mathrm{n}_1 - \mathrm{n}_0 - 2} \mathrm{T}^2$:Pillai-Bartlett trace, V test. $V = \frac{H}{1+H}$ and Roy's greatest characteristic root $R = \frac{H}{1+H}$. In the two-group case all the test-statistics produce the same result although in cases where the

$$\Lambda = \frac{1}{1+H} = \frac{n_1 - n_0 - 2}{n_1 - n_0 - 2 + T^2}$$

The null hypothesis can be tested using any of the above criteria. The statistical significance of any of the tests can be evaluated using two simple transformations based either on a chi-square distribution or an F-distribution. For Wilk's lambda we define the test statistic based on the chi-square distribution for g groups as[4]

$$\Lambda_c = -\ln\Lambda[n_1 + n_0 - 2 - \frac{g}{2}] \sim \chi^2_2$$

Hence, the selected 16 financial ratios is an appropriate set of discriminator variables, since the null hypothesis is rejected for each year respectively apart from 99/01 when the hypothesis could not be rejected.

Table 2.13 – Values of Test Statistic

	n_1	n_0	Wilk's lambda	Chi-square	Degrees of freedom	Critical value
94-96	426	66	0.895	52.889	16	41.34
95-97	479	67	0.829	99.648	16	41.34
96-98	418	56	0.853	73.801	16	41.34
97-99	355	48	0.847	65.337	16	41.34
98-00	337	38	0.855	57.108	16	41.34
99-01	252	51	0.859	45.553	16	41.34

2.6 Testing the Equality of Medians

Although the use of means provides us with information about the average value of financial ratios for the two groups of companies it is more informative to use the median as a measure of location for the distribution of financial ratios. The use of median underpins most of the work of credit rating companies and as the following table shows there is a high correlation between the median value of a financial ratio and the credit rating of a company.

comparison involves more than two groups their performance differ. Olson (1976) found the Pillai-Bartlett trace V to be the most robust of the four tests and is sometimes preferred for this reason.

[4] Wilk's lambda test based on the F-distribution again for g groups is defined as:

$\Lambda_F = \frac{1-\Lambda}{\Lambda}\frac{n_1 + n_0 - g - 1}{g} \sim F^g_{n_1+n_0-p-1}$ Alternatively we can use a test statistic based on Hotelling's T^2 that follows the F- distribution is defined as $H_F = \frac{n_1 + n_0 - p - 1}{2(n_1 + n_0 - p)} T^2 \sim F^p_{n_1+n_0-p-1}$

Both Λ_F and H_F follow the F-distribution with p and $n_1 + n_0 - p - 1$ degrees of freedom. Using Hotellings test statistic, we reject the null hypothesis $Ho: \bar{\mu}_1 - \bar{\mu}_0 = 0$ at level of confidence a if: $H_F \geq F^p_{n_1+n_0-p-1}(\alpha)$ where $F^p_{n_1+n_0-p-1}(a)$ is the value of the f-distribution, above, which only 5% of the cases lie.

Table 2.14 Three-year (1997-1999) medians for various credit categories

	AAA	AA	A	BBB	BB	B	CCC
EBIT interest coverage.	17.5	10.8	6.8	3.9	2.3	1.0	0.2
EBITDA interest coverage	21.8	14.6	9.6	6.1	3.8	2.0	1.4
Funds flow/total debt	105.8	55.8	46.1	30.5	19.2	9.4	5.8
Free operating Cash flow/total debt (%)	55.4	24.6	15.6	6.6	1.9	(4.5)	(14.0)
Return on capital (%)	28.2	22.9	19.9	14.0	11.7	7.2	0.5
Operating Income/sales (%)	29.2	21.3	18.3	15.3	15.4	11.2	13.6
Long-term debt/capital (%)	15.2	26.4	32.5	41.0	55.8	70.7	80.3
Total debt/capital (%)	26.9	35.6	40.1	47.7	61.3	74.6	89.4
Number of Companies	10	34	150	234	276	240	23

Source: "Adjusted Key US Industrial Financial Ratios " Standard & Poor's Credit Week, September 20, 2000, pp. 39-44.

For instance, 50 percent of the companies that are classified as CCC had a debt-capital ratio of more than 89.4 whereas 50 % of those companies rated as AAA had a debt-capital ratio of less than 26.9 percent. In a univariate framework it will be unlikely for a company with a debt-capital ratio of less than 26.9 percent to be a CCC. The predictive power of the median can therefore be stronger as it does not depend on outliers, which may affect the mean value.

The test for the equality of medians is the Wilkoxon test also known as the rank sum test and one variant as the Mann-Whitney test. This is a non-parametric test and compares two unpaired groups.

Let $x_1, x_2, ..., x_m$ and $y_1, y_2, ..., y_n$ be two independent random samples of sizes m and n with m<n for the continuous populations X and Y. We wish to test the hypothesis

$$H_0: \quad median(x) = median(y)$$
$$H_a: \quad median(x) \neq median(y)$$

The test procedure is as follows. Arrange all m + n observations in ascending order of magnitude and assign ranks to them. If two or more observations are tied (identical) then we use the mean of the ranks that would have been assigned if the observations were differed. Let W(m) be the sum of the ranks in the smaller sample, and define W(n) as the sum of ranks in the other sample. Then

$$W(n) = \frac{(m+n)(m+n+1)}{2} - W(m)$$

If the sample medians do not differ, we will expect the sum of the ranks to be nearly equal for both samples after adjusting for the difference in sample size. Consequently, if the sums of the ranks differ greatly, we will conclude that the medians are not equal. The distribution of W(m) and W(n) can be derived but are complicated. However when both m and n are moderately large, normally greater than 8, the distribution of W(m) can be approximated by the normal distribution with mean

$E(W(m)) = \frac{m(m+n+1)}{2}$ and variance $Var(m) = \frac{mn(m+n+1)}{12}$

and therefore the quantity

$$z = \frac{W(m) - E(W(m))}{\sqrt{Var(m)}}$$

follows the standard normal distribution and can be employed as the test statistic. The null hypothesis of equal medians can be rejected if the value of the test statistic is above the critical value at some level of confidence.

The Wilkoxon test can be formulated in terms of the sum of the ranks in the larger sample as follows

$E(W(n)) = \frac{n(m+n+1)}{2}$ and variance $Var(n) = \frac{mn(m+n+1)}{12}$

and therefore the quantity

$z = \frac{W(n) - E(W(n))}{\sqrt{Var(n)}}$ follows the standard normal distribution and can be employed as the test statistic.

The null hypothesis of equal medians can be rejected if the value of the test statistic is above the critical value at some level of confidence. The table below shows the value of the Wilkoxon normal approximation.

As in the case of the test for the means, the test is significant for most of the financial ratios and for most of the years. The values of the medians are given in the Data Appendix.

Table 2.15: Values of the test statistic z for each year

Ratio	94	95	96	97	98	99
(undistributive profits & accumulated losses)/ paid up capital	-2.492	-3.727	-3.465	-3.613	-1.199	-3.613
Total debt / paid up capital	-1.349	-1.444	-1.359	-1.753	-1.796	-1.753
Current ratio	-3.184	-4.377	-4.453	-3.636	-3.494	-3.636
Return on equity	-3.900	-5.175	-4.234	-3.295	-1.988	-3.295
Interest payments /cash flow	-1.679	-3.577	-5.373	-6.107	-4.323	-6.107
Operating profit / sales	-0.788	-2.131	-0.366	-1.189	-0.340	-1.189
EBIT / total assets	-3.458	-5.363	-5.031	-4.672	-4.165	-4.672
Net income / sales	-3.833	-6.011	-4.349	-4.482	-2.639	-4.482
EBITDA/ Sales	-0.992	-2.190	-0.387	-1.467	-0.666	-1.467
Retained profits/ total assets	-4.042	-5.135	-4.274	-3.831	-2.777	-3.831
Sales/ total assets	-1.320	-2.573	-2.273	-1.387	-3.217	-1.387
Working capital/ total assets	-3.310	-4.446	-4.555	-3.318	-3.518	-3.318
Interest payments/ sales	-2.890	-3.909	-5.190	-4.730	-3.652	-4.730
Interest payments/ EBITAD	-4.326	-5.208	-6.134	-5.174	-2.854	-5.174
Total debt/ (total debt & total equity)	-2.111	-3.348	-2.530	-3.326	-1.839	-3.326
Size	-1.509	-0.340	-0.951	-0.841	-1.127	-0.841

2.7 Testing the equality of variances

The next statistical attribute of the financial ratios that we are testing is the equality of the variances in the two groups of financial ratios. Before conducting the t- test, one should normally evaluate the variances of the two distributions to see if they differ. There are two reasons for this:

(a) It is important to know if a ratio's values vary more in one group of companies than in another as this has important implications for the risk of particular group of companies

(b) The calculations for the two- sample t- test depend on whether the sample variances are the same or different.

The formal test takes the form of equality of variances in g groups

$H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_g$

$H_a : \sigma_i \neq \sigma_j$ for at least one pair (i,j).

The most common test for equality of variances across g samples is Bartlett's test which is very sensitive to departures from normality. If the samples come from non-normal distributions, then Bartlett's test may simply be testing for non-normality. However, If there is strong evidence that data employed do in fact come from a normal, or nearly normal, distribution, then Bartlett's test has better performance.

The Levene Test is an alternative to the Bartlett test that is less sensitive to departures from normality. The test can be defined as follows: Given a variable Y with sample of size n divided into g subgroups, where n_i is the sample size of the ith subgroup, the Levene test statistic is defined as:

$$W = \frac{n-g}{g-1} \frac{\sum_{i=1}^{g} n_i(\overline{z}_i - \overline{z})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i} n_i(\overline{z}_{ij} - \overline{z}_i)^2}$$

where $\overline{z}_{ij}$ can have one of the following three definitions:

1. $\overline{z}_{ij} = \left| Y_{ij} - \overline{Y}_i \right|$

   where $\overline{Y}_i$ is the mean of the ith subgroup.

2. $\overline{z}_{ij} = \left| Y_{ij} - \tilde{Y}_i \right|$

   where $\tilde{Y}_i$ is the median of the ith subgroup.

3. $\overline{z}_{ij} = \left| Y_{ij} - \tilde{\tilde{Y}}_i \right|$

   where $\tilde{\tilde{Y}}_i$ is the 10% trimmed mean[5]of the ith subgroup.

$\overline{z}_i$ are the group means of the Z_{ij} and $\overline{z}$ is the overall mean of the Z_{ij}.

The three choices for defining Z_{ij} determine the robustness and power of Levene's test. By robustness, we mean the ability of the test to not falsely detect unequal variances when the underlying data are not normally distributed and the variables are in fact equal. By power, we mean the ability of the test to detect unequal variances when the variances are in fact unequal. Levene's original paper only proposed using the mean. Brown and Forsythe (1974) performed Monte Carlo studies that indicated that using the trimmed mean performed best when the underlying data followed a Cauchy distribution (i.e., heavy-tailed) and the median performed best when the underlying data followed a χ_4^2 (i.e., skewed) distribution. Using the mean provided the best power for symmetric, moderate-tailed, distributions.

Although the optimal choice depends on the underlying distribution, the definition based on the median is recommended as the choice that provides good robustness against many types of non-normal data while

[5] Trimmed Mean - similar to the mid-mean except different percentile values are used. A common choice is to trim 5% of the points in both the lower and upper tails, i.e., calculate the mean for data between the 5th and 95th percentiles.

retaining good power. If you have knowledge of the underlying distribution of the data, this may indicate using one of the other choices.

The Levene test rejects the hypothesis that the variances are equal if

$W > F(\alpha, g-1, n-g)$

where $F(\alpha, g-1, n-g)$ is the upper critical value of the F-distribution with g - 1 and n- g degrees of freedom at a significance level of α . The empirical results from the test are shown in the following table.

Table 2.16– Values of the test Statistic – Levene Test

Ratio	94	95	96	97	98	99
(undistributive profits & accumulated losses)/ paid up capital	0.60	2.47	1.03	1.14	0.180	0.536
Total debt / paid up capital	1.19	0.01	0.52	2.97	1.846	2.178
Current ratio	1.46	1.44	1.92	4.74	3.216	1.924
Return on equity	1.01	2.01	3.56	2.62	0.234	3.503
Interest payments /cash flow	20.94	20.70	11.95	17.92	21.403	0.364
Return on fixed assets	0.40	1.04	2.88	1.04	0.702	1.788
Operating profit / sales	0.21	25.13	3.08	0.23	0.057	0.074
EBIT / total assets	0.61	3.83	5.08	2.32	3.415	4.799
Net income / sales	0.06	40.58	15.81	0.04	0.180	9.139
EBITDA/ Sales	0.21	31.71	1.20	0.65	0.156	0.223
Retained profits/ total assets	0.00	0.52	0.16	0.54	0.763	0.002
Sales/ total assets	0.33	3.99	4.41	3.80	4.619	1.516
Working capital/ total assets	0.93	0.02	0.02	2.26	0.137	0.059
Interest payments/ sales	1.58	3.00	3.39	2.60	0.237	6.826
Interest payments/ EBITAD	3.23	0.04	0.30	4.11	16.771	19.909
Total debt/ (total debt & total equity)	0.05	0.49	0.92	1.51	0.605	0.012
Size	7.24	0.63	1.61	2.45	1.902	0.120
n	492	546	479	403	374	327
g – 1	1	1	1	1	1	1
n – 2	492	546	474	402	374	327
Critical Value @ 5%	3.860	3.859	3.861	3.865	3.866	3.875
Critical Value @ 1%	6.687	6.681	6.688	6.698	6.704	6.726

Starting with the 1994 sample, for most of the ratios the value of the test is below the critical value and we cannot reject the hypothesis that the variances in the two groups are the same. In fact only 4 ratios appear to have a different sample variance. However for subsequent years, the number of ratios for which the assumption of equality cannot be accepted rises to 12 out of 28 in 1995 and then remains the same at 7 for the two remaining years of 1996 and 1997. When we use the 1% significance level, the number of ratios for which we can reject the assumption of equal variances is very small. The conclusion from this test is that the results of the test statistics on the means and in particular the t-test should be accepted as valid.

2.8 Testing the Homogeneity of Covariance Matrices

The final statistical attribute that we are testing is the equality of the variance-covariance matrices in the two groups. Suppose we have n observations on p variables which come from g groups with n_i

observations in group i and $n = \sum_{i=1}^{g} n_i$. Let Ω_i be the variance covariance matrix in group i and the pooled variance – covariance matrix

$$\Omega_p = \frac{1}{(n-g)} \sum_{i=1}^{g} (n_i - 1)\Omega_i$$

The log-likelihood ratio statistic for testing the hypothesis of equality of covariance matrices $\Omega_1 = \Omega_2 = \cdots \Omega_g$ is given by

$$\text{LRRS} = n \ln\left|\Omega_p\right| - \sum_{i=1}^{g} n_i \ln\left|\Omega_i\right|$$

and for large samples (large value of n) it follows under the null hypothesis of equal variance covariance matrices, a chi-square distribution with (g-1)p(p+1)/2 degrees of freedom.

For small samples a modified version of the log-likelihood ratio is employed, known a Box's M statistic which is defined as

$$M = (n-g) \ln\left|\Omega_p\right| - \sum_{i=1}^{g} (n_i - 1) \ln\left|\Omega_i\right|$$

In order to test the null hypothesis the following transformation is normally used

$$G_1 = cM$$

where

$$c = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \left(\sum_{i=1}^{g} \frac{1}{n_i - 1} - \frac{1}{n-g} \right)$$

which is distributed as a chi-square with degrees of freedom given by

$$df_1 = \frac{(g-1)p(p+1)}{2}$$

The test statistic can also be expressed in the form of an F-test by adopting a second transformation

$$G_2 = \frac{c - df_1 / df_2}{df_1} M$$

which is distributed as an F with degrees of freedom df_1 and df_2, and where

$$df_2 = \frac{df_1 + 2}{\left|\tau - (1-c)^2\right|}$$

and

$$\tau = \frac{(p-1)(p+2)}{6(g+1)} \left(\sum_{i=1}^{g} \frac{1}{(n_i - 1)^2} - \frac{1}{(n-g)^2} \right)$$

The empirical results, the values of M and the test statistic together with the critical value of the F statistic are shown in the table below.

Table 2.17: Empirical results

Period	M	G_2	df_1	df_2	Critical Value	significance
94-96	1083.1	7.253	136	41661	1.12	0.000
95-97	5995.9	40.224	136	42494	1.12	0.000
96-98	2312.5	15.205	136	29384	1.12	0.000
97-99	1050.2	6.579	136	21470	1.12	0.000
98-00	957.9	5.916	136	13112	1.12	0.000
99-01	1240.3	8.038	136	25577	1.12	0.000

The probability value of this F should be greater than .05 to demonstrate that the assumption of homoscedasticity is upheld (lower than the critical value). However the results indicate that the assumption of equal variance-covariance matrices is not supported for any of the years. We reject therefore the hypothesis that the two variance-covariance matrices may be assumed drawn from a common population. The two groups may be thought of as coming from populations with varying in location and shape distribution parameters.

Heteroscedasticity in a multivariate may arise from group differences in covariance as well as the variance of individual variables. We have already seen, using the Levene test, that individual variances are on the whole not overwhelmingly different in the two groups. In fact in any one period the number of unequal variances never exceeds more than a quarter of the total number of the variances. Differences in covariances can arise because two variables correlate differently in different populations. For example the ROE may correlate differently with the solvency ratio in the group of downgraded companies than in the group of non-downgraded companies.

Table 2.18: Value of determinants (Natural log of determinant)

	Group 1	Group 0	Pooled
94-96	-47.862	-38.163	-37.239
95-97	-32.427	-31.820	-27.163
96-98	-46.729	-31.924	-28.750
97-99	-58.452	-40.883	-40.323
98-00	-52.994	-40.857	-39.488
99-01	-49.722	-48.951	-44.958

The value of the determinant can be used to asses the approximate correlation structure of the group variance-covariance structure. The larger a group's determinant is, the more nearly uncorrelated the set of predictors are within the group. If two or more variables correlate perfectly, the resulting correlation matrix is singular and will have a determinant of 0. The more the group determinants vary among themselves, the more the group correlation structure will differ.

Although no formal test is available to test the equality of determinants, it is obvious from the table above that the determinants in the two groups differ substantially. We may therefore conclude that the differences in the covariance matrices are primarily due to the different correlation structure in the two groups.

2.9 Testing for Normality

The assumption of a normal distribution for the various financial ratios employed in empirical studies of discriminant analysis is the justification for that particular approach. Testing for normality is therefore imperative in order to see if the assumption is supported by the data. Deviations from normality are manifested in the non-symmetry of the empirical distribution function and the presence of kurtosis. In the next two tables we show the values of the coefficient of skewness and kurtosis in order to see if they different from the coefficients of the normal distribution. .

Skewness is the lack of symmetry in a distribution. It is calculated as the third central moment divided by the standard deviation raised to the power of three.

$$sk = \frac{C_3}{\sigma^3} \quad \text{where } C_3 = E(X-\mu)^3$$

If $-2 < \dfrac{\text{sk}}{\text{standard error of skewness}} < +2$ then the data are normally distributed otherwise we reject the hypothesis of symmetry.

Most of the ratios show significant deviations from symmetry with the coefficients of skewness for the overwhelming majority of ratios being statistically different from zero. For only three ratios we can accept the hypothesis of symmetry in every single year (WA, DR and SIZE) whereas ROTA has a symmetric distribution in four years.

Next we consider the second sample statistic that shows departures from normality and this is the coefficient of kurtosis or peakedness of a distribution which is defined as

$$ku = \frac{C_4}{\sigma^4} - 3 \quad \text{where } C_4 = E(X-\mu)^4$$

Table 2.19: Skewness Coefficients

	94	95	96	97	98	99
UPL	3.17	6.44	5.84	8.39	6.92	10.12
DC	6.00	10.67	4.91	2.24	4.01	2.72
CR	15.07	15.17	19.73	6.95	4.50	9.34
ROE	5.92	5.03	4.36	4.54	4.62	5.92
IPF	3.06	16.16	-2.36	8.28	6.12	-9.03
OS	0.96	-23.20	-0.72	-1.78	-14.95	0.51*
ROA	7.22	5.07	8.99	13.88	3.89	0.66*
NIS	-2.21	-21.45	-17.43	-13.99	-8.83	-4.15
ES	0.90	-22.56	-3.07	1.41	-14.64	1.08
ROTA	-0.65	0.68	0.09*	0.37*	-0.60*	0.18*
SOA	3.85	3.62	2.48	1.80	1.40	2.38
WA	0.16*	-0.41*	-0.29*	-0.12*	0.20*	-0.21*
IPS	3.04	5.94	4.13	2.38	11.69	3.11
IPE	7.02	-17.38	-21.21	4.32	-10.97	-14.56
DR	-0.32*	-0.50*	-0.35*	-0.37*	-0.47*	-0.32*
SIZE	-0.09*	-0.13*	-0.23*	0.00*	-0.07*	0.08*
N	492	546	474	403	375	303

A common rule-of-thumb test for normality is to estimate the coefficient of kurtosis, then divide it by its standard error. Kurtosis should be within the +2 to -2 range when the data are normally distributed (a few authors use +3 to -3). Negative kurtosis indicates too many cases in the tails of the distribution. Positive kurtosis indicates too few cases in the tails. Note that in SPSS the kurtosis is centred on 0 rather than the value of 3 which is the value of the coefficient for the normal distribution and this is reflected in our definition above.

Looking at the coefficient of kurtosis, we get a similar picture with most of the ratios exhibiting significant kurtosis. Only three ratios have kurtosis which is not statistically different from zero, WA, DR and SIZE.

Table 2.20 : Kurtosis –Whole sample

	94	95	96	97	98	99
UPL	32.30	84.20	57.54	104.18	74.84	134.51
DC	54.07	160.98	39.72	5.87	29.64	8.62
CR	231.12	266.84	411.73	59.21	24.29	108.87
ROE	50.23	40.18	29.78	27.47	26.93	47.33
IPF	63.21	287.74	36.77	92.83	52.09	130.77
OS	7.78	540.68	27.24	23.55	267.42	3.15
ROA	87.21	33.90	120.98	234.83	19.73	2.68
NIS	11.79	478.30	342.71	253.12	140.90	32.60
ES	6.29	519.23	44.73	10.04	260.52	2.47
ROTA	10.36	4.15	7.79	4.71	11.72	5.71
SOA	25.49	25.89	14.39	6.55	3.23	10.35
WA	0.40*	2.20	1.29	0.96*	1.24	0.88*
IPS	13.38	54.73	23.66	9.65	186.55	14.95
IPE	69.95	377.83	457.26	34.60	215.83	234.56
DR	-0.55*	-0.39*	-0.54*	-0.43*	-0.41*	-0.33*
SIZE	0.30*	0.34*	0.45*	0.59*	0.44*	0.61*
N	492	546	474	403	375	303

A formal test of normality was also conducted using the Kolmogorov-Smirnov (KS) test. The KS test compares the empirical cumulative distribution function (cdf) to the cdf of the normal distribution. The empirical cdf is computed as an estimate of the theoretical distribution and is simply defined as

$$F_n(r) = \text{No of observations} < r$$

The test statistic is calculated as

$$D_n = \underset{r}{Max}\left|F_n(r) - F(r)\right|$$

where the maximum is taken over r. If the D statistic is significant, then the hypothesis that the respective distribution is normal should be rejected. The probability values that are reported can be based on those tabulated by Massey (1951); those probability values are valid when the mean and standard deviation of the normal distribution are known a-priori and not estimated from the data.

When the cumulative distribution function parameters are not known, in this case the mean and the variance of the distribution, they have to be estimated from the data. The distribution of D using estimated parameters has been derived by Lilliefors (1967) and it is this values that SPSS uses to test normality and the ones that we are reporting.

We have tested for univariate normality for the whole sample as well as for the two groups of companies. The results for each variable and each year are shown in the Appendix. Again the overwhelming evidence from this test is that the assumption of normality is not supported by the data. However the results f or the three variables (WA, DR and SIZE) are confirmed using this more robust test.

2.10 Conclusions

The results of this chapter can be summarised in terms of the six tests we performed as follows. The location parameters of the distribution, i.e. the means and the medians of the two groups of ratios are significantly different in the two groups. This is confirmed by both the univariate and multivariate tests. Companies, which are downgraded, are characterised by higher leverage and solvency ratios and lower profitability and growth ratios. Size is not an important variable in predicting whether a company will be downgraded. The variances are not on the whole different, indicating distributions with the same shape parameters but different location parameters. The correlation structure of the ratios however, is different in the two groups. Finally, the tests of normality have shown that the distribution of most financial ratios significantly different from the Gaussian normal distribution. This finding should be taken into account in the application of techniques that are based on the assumption of normality and equal covariance matrices in the two groups. The impact of the assumptions is of course an empirical question which will be assessed in the next chapter.

Turning now to the stability of the financial ratios over time, we see that some important ratios such the leverage ratios and liquidity ratios such as working capital over total assets have remained stable. This is important if the purpose of the analysis is to predict financial distress. It is also noticeable that the financially healthy companies had more stable ratios than companies that became financially distressed. The results would therefore indicate that the statistical characteristics of the financial ratios of the two groups of companies were different two years before they were downgraded. As we have already discussed, statistical inference based on univariate tests may be contradictory and misleading. What we are testing in the next chapter is the significance of linear combinations of those characteristics.

2.11 Data Appendix

Table A1: Values Of the Means 94-96

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.091	-0.062
Total debt / paid up capital	4.200	4.104
Current ratio	2.310	1.176
Return on equity	0.470	0.198
Interest payments /cash flow	0.269	0.346
Operating profit / sales	0.192	0.134
EBIT / total assets	0.257	0.120
Net income / sales	0.054	-0.021
EBITDA/ Sales	0.191	0.133
Retained profits/ total assets	0.024	-0.007
Sales/ total assets	1.312	1.276
working capital/ total assets	0.151	0.067
Interest payments/ sales	0.076	0.103
Interest payments/ EBITAD	0.570	0.837
Total debt/ (total debt & total equity)	0.573	0.635
Size	4.988	5.147

Table A2: Mean Values 95-97

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.154	-0.036
Total debt / paid up capital	4.616	5.699
Current ratio	1.923	2.360
Return on equity	0.450	0.153
Interest payments /cash flow	0.207	0.422
Operating profit / sales	0.137	0.112
EBIT / total assets	0.177	0.073
Net income / sales	-0.001	-0.155
EBITDA/ Sales	0.135	0.113
Retained profits/ total assets	0.014	-0.013
Sales/ total assets	1.102	1.043
working capital/ total assets	0.139	0.020
Interest payments/ sales	0.064	0.122
Interest payments/ EBITAD	-0.210	0.818
Total debt/ (total debt & total equity)	0.563	0.639
Size	5.053	5.116

Table A3: Mean Values 96-98

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.154	-0.036
Total debt / paid up capital	4.616	5.699
Current ratio	1.923	2.360
Return on equity	0.450	0.153
Interest payments /cash flow	0.207	0.422
Operating profit / sales	0.137	0.112
EBIT / total assets	0.177	0.073
Net income / sales	-0.001	-0.155
EBITDA/ Sales	0.135	0.113
Retained profits/ total assets	0.014	-0.013
Sales/ total assets	1.102	1.043
working capital/ total assets	0.139	0.020
Interest payments/ sales	0.064	0.122
Interest payments/ EBITAD	-0.210	0.818
Total debt/ (total debt & total equity)	0.563	0.639
Size	5.053	5.116

Table A4: Mean Values 97-99

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.146	-0.046
Total debt / paid up capital	3.990	5.222
Current ratio	1.571	1.099
Return on equity	0.427	0.130
Interest payments /cash flow	0.203	0.488
Operating profit / sales	0.138	0.124
EBIT / total assets	0.192	0.077
Net income / sales	0.011	-0.028
EBITDA/ Sales	0.144	0.122
Retained profits/ total assets	0.015	-0.012
Sales/ total assets	1.121	1.108
working capital/ total assets	0.143	0.050
Interest payments/ sales	0.052	0.087
Interest payments/ EBITAD	0.461	0.759
Total debt/ (total debt & total equity)	0.571	0.672
Size	5.189	5.248

Table A5 Mean Values 98-00

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.081877	0.027401
Total debt / paid up capital	4.23012	5.677777
Current ratio	1.608952	1.181371
Return on equity	0.449329	0.270893
Interest payments /cash flow	0.199207	0.441866
Operating profit / sales	0.131971	0.148785
EBIT / total assets	0.219047	0.104959
Net income / sales	-0.00015	0.000351
EBITDA/ Sales	0.132548	0.153245
Retained profits/ total assets	0.014944	-0.00309
Sales/ total assets	1.120559	0.803113
working capital/ total assets	0.151156	0.04876
Interest payments/ sales	0.0565	0.083469
Interest payments/ EBITAD	0.435287	0.091693
Total debt/ (total debt & total equity)	0.584133	0.650915
Size	5.170886	5.31415

Table A6 Mean Values 99-01

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.101949	-0.119
Total debt / paid up capital	3.322427	4.084223
Current ratio	1.746211	1.150794
Return on equity	0.333314	0.104319
Interest payments /cash flow	0.518238	0.291693
Operating profit / sales	0.147327	0.151571
EBIT / total assets	0.12296	0.092675
Net income / sales	0.029536	-0.01506
EBITDA/ Sales	0.149912	0.154346
Retained profits/ total assets	0.019222	-0.0043
Sales/ total assets	1.000982	0.771127
working capital/ total assets	0.150766	0.052965
Interest payments/ sales	0.04905	0.078182
Interest payments/ EBITAD	0.309811	-0.44326
Total debt/ (total debt & total equity)	0.557004	0.627514
Size	5.311664	5.452448

Table B1 Median Values 94-96

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.000	-0.008
Total debt / paid up capital	2.248	3.419
Current ratio	1.250	1.103
Return on equity	0.176	0.046
Interest payments /cash flow	0.229	0.284
Operating profit / sales	0.130	0.109
EBIT / total assets	0.147	0.112
Net income / sales	0.021	0.005
EBITDA/ Sales	0.128	0.106
Retained profits/ total assets	0.013	0.001
Sales/ total assets	1.109	1.007
Working capital/ total assets	0.137	0.068
Interest payments/ sales	0.057	0.076
Interest payments/ EBITAD	0.481	0.764
Total debt/ (total debt & total equity)	0.587	0.649
Size	5.025	5.033

Table B2 Median Values 95-97

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.000	-0.040
total debt / paid up capital	2.228	3.490
current ratio	1.263	1.057
Return on equity	0.187	0.046
interest payments /cash flow	0.199	0.303
Operating profit / sales	0.121	0.095
EBIT / total assets	0.131	0.089
Net income / sales	0.022	0.004
EBITDA/ Sales	0.120	0.090
Retained profits/ total assets	0.012	0.004
Sales/ total assets	1.090	0.934
working capital/ total assets	0.136	0.044
Interest payments/ sales	0.048	0.085
Interest payments/ EBITAD	0.422	0.753
Total debt/ (total debt & total equity)	0.588	0.672
Size	5.076	5.049

Table B3 Median Values 96-98

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.000	-0.024
Total debt / paid up capital	2.366	3.275
Current ratio	1.250	1.083
Return on equity	0.163	0.025
Interest payments /cash flow	0.181	0.350
Operating profit / sales	0.117	0.115
EBIT / total assets	0.128	0.083
Net income / sales	0.019	0.001
EBITDA/ Sales	0.118	0.115
Retained profits/ total assets	0.009	0.001
Sales/ total assets	0.991	0.675
Working capital/ total assets	0.139	0.048
Interest payments/ sales	0.044	0.083
Interest payments/ EBITAD	0.344	0.749
Total debt/ (total debt & total equity)	0.583	0.673
Size	5.079	5.149

Table B4 Median Values 97-99

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.000	0.000
total debt / paid up capital	2.659	3.583
current ratio	1.245	1.079
Return on equity	0.182	0.031
interest payments /cash flow	0.157	0.334
Operating profit / sales	0.121	0.094
EBIT / total assets	0.125	0.072
Net income / sales	0.020	0.004
EBITDA/ Sales	0.118	0.090
Retained profits/ total assets	0.008	0.001
Sales/ total assets	1.034	0.885
working capital/ total assets	0.138	0.051
Interest payments/ sales	0.039	0.069
Interest payments/ EBITAD	0.330	0.622
Total debt/ (total debt & total equity)	0.587	0.708
Size	5.179	5.241

TABLE B5 Median Values 98-00

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.000	0.000
total debt / paid up capital	2.565	4.052
current ratio	1.239	1.053
Return on equity	0.191	0.061
interest payments /cash flow	0.157	0.312
Operating profit / sales	0.128	0.141
EBIT / total assets	0.138	0.071
Net income / sales	0.019	0.007
EBITDA/ Sales	0.122	0.122
Retained profits/ total assets	0.010	0.003
Sales/ total assets	1.015	0.781
working capital/ total assets	0.137	0.043
Interest payments/ sales	0.046	0.086
Interest payments/ EBITAD	0.356	0.553
Total debt/ (total debt & total equity)	0.603	0.672
Size	5.213	5.301

TABLE B6 Median Values 99-01

	0	1
(undistributive profits & accumulated losses)/ paid up capital	0.001	0
total debt / paid up capital	2.118	2.011
current ratio	1.307	1.064
Return on equity	0.123	0.051
interest payments /cash flow	0.155	0.221
EBIT / total assets	0.117	0.091
Net income / sales	0.021	0.007
EBITDA/ Sales	0.132	0.138
Retained profits/ total assets	0.010	0.001
Sales/ total assets	0.893	0.704
working capital/ total assets	0.145	0.054
Interest payments/ sales	0.042	0.060
Interest payments/ EBITAD	0.326	0.403
Total debt/ (total debt & total equity)	0.559	0.667
Size	5.303	5.361

Table C1: Kolmogorov-Smirnov Test of Normality 1994[6]

Ratios	*Group*	*KS Statistic*	*Degrees of freedom*	*Significance*
UPL	0	0.280	426.000	0.000
	1	0.231	66.000	0.000
DC	0	0.264	426.000	0.000
	1	0.141	66.000	0.002
CR	0	0.418	426.000	0.000
	1	0.127	66.000	0.010
ROE	0	0.254	426.000	0.000
	1	0.279	66.000	0.000
IPF	0	0.165	426.000	0.000
	1	0.284	66.000	0.000
OS	0	0.126	426.000	0.000
	1	0.121	66.000	0.018
ROA	0	0.204	426.000	0.000
	1	0.170	66.000	0.000
NIS	0	0.244	426.000	0.000
	1	0.248	66.000	0.000
ES	0	0.135	426.000	0.000
	1	0.118	66.000	0.024
ROTA	0	0.185	426.000	0.000
	1	0.181	66.000	0.000
SOA	0	0.145	426.000	0.000
	1	0.267	66.000	0.000
WA	0	0.056	426.000	0.003
	1	0.049	66.000	0.200
IPS	0	0.145	426.000	0.000
	1	0.180	66.000	0.000
IPE	0	0.254	426.000	0.000
	1	0.257	66.000	0.000
DR	0	0.063	426.000	0.000
	1	0.105	66.000	0.071
SIZE	0	0.058	426.000	0.002
	1	0.114	66.000	0.033

[6] Significance has been calculated using Lilliefors Significance Correction

Table C2: Kolmogorov-Smirnov Test of Normality 1995

Ratios	*Group*	*KS Statistic*	*Degrees of freedom*	*Significance*
UPL	0	0.299	479.000	0.000
	1	0.313	67.000	0.000
DC	0	0.312	479.000	0.000
	1	0.225	67.000	0.000
CR	0	0.380	479.000	0.000
	1	0.327	67.000	0.000
ROE	0	0.252	479.000	0.000
	1	0.260	67.000	0.000
IPF	0	0.353	479.000	0.000
	1	0.427	67.000	0.000
OS	0	0.112	479.000	0.000
	1	0.491	67.000	0.000
ROA	0	0.241	479.000	0.000
	1	0.088	67.000	0.200
NIS	0	0.273	479.000	0.000
	1	0.492	67.000	0.000
ES	0	0.144	479.000	0.000
	1	0.484	67.000	0.000
ROTA	0	0.164	479.000	0.000
	1	0.205	67.000	0.000
SOA	0	0.137	479.000	0.000
	1	0.115	67.000	0.029
WA	0	0.050	479.000	0.006
	1	0.108	67.000	0.051
IPS	0	0.230	479.000	0.000
	1	0.202	67.000	0.000
IPE	0	0.389	479.000	0.000
	1	0.330	67.000	0.000
DR	0	0.061	479.000	0.000
	1	0.114	67.000	0.030
SIZE	0	0.051	479.000	0.005
	1	0.091	67.000	0.200

Table C3 Kolmogorov-Smirnov Tests of Normality 1996

Ratios	*Group*	*KS Statistic*	*Degrees of freedom*	*Significance*
UPL	0	0.293	418.000	0.000
	1	0.170	56.000	0.000
DC	0	0.220	418.000	0.000
	1	0.266	56.000	0.000
CR	0	0.392	418.000	0.000
	1	0.099	56.000	0.200
ROE	0	0.227	418.000	0.000
	1	0.168	56.000	0.000
IPF	0	0.214	418.000	0.000
	1	0.163	56.000	0.001
OS	0	0.149	418.000	0.000
	1	0.245	56.000	0.000
ROA	0	0.233	418.000	0.000
	1	0.077	56.000	0.200
NIS	0	0.299	418.000	0.000
	1	0.397	56.000	0.000
ES	0	0.180	418.000	0.000
	1	0.252	56.000	0.000
ROTA	0	0.175	418.000	0.000
	1	0.227	56.000	0.000
SOA	0	0.086	418.000	0.000
	1	0.207	56.000	0.000
WA	0	0.060	418.000	0.001
	1	0.085	56.000	0.200
IPS	0	0.203	418.000	0.000
	1	0.185	56.000	0.000
IPE	0	0.473	418.000	0.000
	1	0.273	56.000	0.000
DR	0	0.046	418.000	0.035
	1	0.103	56.000	0.200
SIZE	0	0.057	418.000	0.003
	1	0.075	56.000	0.200

Table C4: Kolmogorov-Smirnov Tests of Normality 1997

Ratios	*Group*	*KS Statistic*	*Degrees of freedom*	*Significance*
UPL	0	0.308	355.000	0.000
	1	0.310	48.000	0.000
DC	0	0.189	355.000	0.000
	1	0.238	48.000	0.000
CR	0	0.264	355.000	0.000
	1	0.197	48.000	0.000
ROE	0	0.248	355.000	0.000
	1	0.186	48.000	0.000
IPF	0	0.205	355.000	0.000
	1	0.277	48.000	0.000
OS	0	0.154	355.000	0.000
	1	0.155	48.000	0.005
ROA	0	0.309	355.000	0.000
	1	0.080	48.000	0.200
NIS	0	0.342	355.000	0.000
	1	0.292	48.000	0.000
ES	0	0.169	355.000	0.000
	1	0.143	48.000	0.016
ROTA	0	0.189	355.000	0.000
	1	0.228	48.000	0.000
SOA	0	0.088	355.000	0.000
	1	0.193	48.000	0.000
WA	0	0.043	355.000	0.179
	1	0.158	48.000	0.004
IPS	0	0.127	355.000	0.000
	1	0.138	48.000	0.023
IPE	0	0.263	355.000	0.000
	1	0.294	48.000	0.000
DR	0	0.048	355.000	0.044
	1	0.106	48.000	0.200
SIZE	0	0.042	355.000	0.200
	1	0.092	48.000	0.200

Table C5 Kolmogorov-Smirnov Test of Normality 1998

Ratios	Group	KS Statistic	Degrees of freedom	Significance
UPL	0	0.275	337.000	0.000
	1	0.217	38.000	0.000
DC	0	0.216	337.000	0.000
	1	0.241	38.000	0.000
CR	0	0.285	337.000	0.000
	1	0.290	38.000	0.000
ROE	0	0.262	337.000	0.000
	1	0.255	38.000	0.000
IPF	0	0.190	337.000	0.000
	1	0.264	38.000	0.000
OS	0	0.321	337.000	0.000
	1	0.162	38.000	0.013
ROA	0	0.234	337.000	0.000
	1	0.231	38.000	0.000
NIS	0	0.299	337.000	0.000
	1	0.188	38.000	0.002
ES	0	0.320	337.000	0.000
	1	0.177	38.000	0.004
ROTA	0	0.213	337.000	0.000
	1	0.208	38.000	0.000
SOA	0	0.095	337.000	0.000
	1	0.129	38.000	0.114
WA	0	0.066	337.000	0.001
	1	0.145	38.000	0.043
IPS	0	0.253	337.000	0.000
	1	0.096	38.000	0.200
IPE	0	0.259	337.000	0.000
	1	0.466	38.000	0.000
DR	0	0.063	337.000	0.003
	1	0.139	38.000	0.063
SIZE	0	0.052	337.000	0.026
	1	0.097	38.000	0.200

Table C6 Kolmogorov-Smirnov (KS) Test of Normality 1999

Ratios	*Group*	*KS Statistic*	*Degrees of freedom*	*Significance*
UPL	0	0.336	252.000	0.000
	1	0.240	51.000	0.000
DC	0	0.214	252.000	0.000
	1	0.229	51.000	0.000
CR	0	0.323	252.000	0.000
	1	0.113	51.000	0.115
ROE	0	0.261	252.000	0.000
	1	0.285	51.000	0.000
IPF	0	0.282	252.000	0.000
	1	0.098	51.000	0.200
OS	0	0.096	252.000	0.000
	1	0.141	51.000	0.013
ROA	0	0.090	252.000	0.000
	1	0.112	51.000	0.156
NIS	0	0.196	252.000	0.000
	1	0.291	51.000	0.000
ES	0	0.104	252.000	0.000
	1	0.160	51.000	0.002
ROTA	0	0.166	252.000	0.000
	1	0.200	51.000	0.000
SOA	0	0.124	252.000	0.000
	1	0.241	51.000	0.000
WA	0	0.057	252.000	0.049
	1	0.053	51.000	0.200
IPS	0	0.128	252.000	0.000
	1	0.212	51.000	0.000
IPE	0	0.337	252.000	0.000
	1	0.458	51.000	0.000
DR	0	0.044	252.000	0.200
	1	0.113	51.000	0.099
SIZE	0	0.058	252.000	0.037
	1	0.088	51.000	0.200

Chapter 3 – Empirical Results on Credit Risk Factors

3.1 Introduction

In the previous chapter we have found that the financial ratios of downgraded companies have different statistical properties from the financial ratios of the companies that were not downgraded. One could argue that on the basis of their values one could use financial ratios to differentiate between companies that were likely to be downgraded and companies, which were not likely to be downgraded. However, comparing and classifying companies on the basis of univariate analysis may lead to contradicting classifications. An alternative way is needed that will take into account all the characteristics of a company not just one at a time, and form an index which will assign weights to its components. The objective of this chapter is to identify the factors that may explain the migration behaviour of the companies in our sample.

To make our objectives clear we use some notation. Consider a company whose state is characterized by its set of financial attributes in the form of p financial ratios $\{r_1, r_2, \ldots, r_p\}$. The company can be in one of two groups. In the univariate analysis of the last chapter, we found that the values of the elements of the set $\{r_1, r_2, \ldots, r_p\}$ were drawn from different populations for companies in the two different groups. In this chapter we address two questions:

(a) Which of the elements in the set $\{r_1, r_2, \ldots, r_p\}$ are better at differentiating companies between the two groups and how do we choose the weights of the set of the "best" differentiating characteristics?

(b) Which is the best method to predict membership of the two groups on the basis of an observed set of financial characteristics? This latter question can be reformulated in terms of providing a probability of membership of a company in a particular group on the basis of an observed set of characteristics.

The statistical technique that we shall employ in this chapter to answer the first question is linear discriminant analysis. The derivation of probabilities of group membership will be tackled using logistic regression.

This chapter is divided in two main parts. In the first part (Sections 3.1 to 3.4) we introduce some basic probability concepts underpinning the theory of classification and define a general classification criterion. We then derive classification criteria for the case when the vector $\{r_1, r_2, \ldots, r_p\}$ is multivariate normal and the covariance matrices in the two groups are equal. The classification rule is the well-known linear discriminant analysis. A quadratic rule is next derived relaxing the assumption of equal covariance matrices. We consider next the case of classification rules not based on the normality assumption or equal covariance assumption in the form of the logistic regression approach.

In the second part (Sections 3.5 and 3.9) we present methods of estimation and inference in the case of discriminant analysis and logistic regression as well as the empirical result of this chapter. We have devoted separate sections to estimation and statistical inference for discriminant analysis and logistic regression in order to make clear the statistical and methodological underpinnings of our empirical results.

We conclude the chapter with an evaluation of our results in Section 3.10.

3.2 Review of Two-Group Classification Theory

Suppose that the financial characteristics of a company are summarized by the p-vector of financial ratios $\mathbf{r} = [r_1, r_2, \ldots, r_p]$ and assume that sample space of $\mathbf{r}$ is partitioned into two subspaces R_0 and R_1. Now define a binary function of $\mathbf{r}$ $\xi = \xi^*(\mathbf{r})$ such that

$$\xi = \xi^*(\mathbf{r}) = \begin{cases} 0 & \text{if } \mathbf{r} \in R_0 \\ 1 & \text{if } \mathbf{r} \in R_1 \end{cases}$$

with prior probabilities $p_0 = \Pr(\xi = 0)$ and $p_1 = \Pr(\xi = 1)$ with $p_1 + p_0 = 1$. The joint probability distribution of $(\xi, \mathbf{r})$ is a mixture with marginal conditional density functions of $\mathbf{r}$ defined as

$f_0(\mathbf{r})$ is the density of $\mathbf{r}$ if $\mathbf{r}$ belongs to group 0

$f_1(\mathbf{r})$ is the density of $\mathbf{r}$ if $\mathbf{r}$ belongs to group 1

The unconditional marginal distribution of $\mathbf{r}$ is given by

$f_r(\mathbf{r}) = p_0 f_0(\mathbf{r}) + p_1 f_1(\mathbf{r})$

The conditional probability of an observation being from group j given that $\mathbf{r} = \mathbf{r}^*$ is given by

$$\pi_{jr} = \Pr(\xi = j / \mathbf{r} = \mathbf{r}^*) = \frac{p_j f_j(\mathbf{r}^*)}{p_0 f_0(\mathbf{r}^*) + p_1 f_1(\mathbf{r}^*)} = \frac{p_j f_j(\mathbf{r}^*)}{f_r(\mathbf{r}^*)} \quad \text{for } j = 0,1$$

The probabilities π_{jr} are called the posterior probabilities of membership in the jth group. The posterior probabilities use the information obtained by measuring $\mathbf{r}$ and are therefore "informed guesses" as opposed to the "uninformed guesses" expressed by the prior probabilities.

Probabilities of Misclassification.

Given the definitions of probability, it follows that the (conditional) probability of classifying an object to group 1 when it in fact is from group 0 equals

$$\begin{aligned} q_{10} &= \Pr(\xi^* = 1 / \xi = 0) \\ &= \Pr(r \in R_1 / \xi = 0) \\ &= \int_{R_1} f_0(\mathbf{r}) d\mathbf{r} \end{aligned}$$

where a single integral is used to represent multiple integration over p dimensions, and the (conditional) probability of classifying an object to group 0 when it in fact is from group 1 equals

$$\begin{aligned} q_{01} &= \Pr(\xi^* = 0 / \xi = 1) \\ &= \Pr(r \in R_0 / \xi = 1) \\ &= \int_{R_0} f_1(\mathbf{r}) d\mathbf{r} \end{aligned}$$

Similarly, we define the conditional probabilities of correct classification q_{00} and q_{11} as

$$q_{00} = \Pr(\xi^* = 0 / \xi = 0)$$
$$= \Pr(\mathbf{r} \in R_0 / \xi = 0)$$
$$= \int_{R_0} f_0(\mathbf{r}) d\mathbf{r}$$

and

$$q_{11} = \Pr(\xi^* = 1 / \xi = 1)$$
$$= \Pr(r \in R_1 / \xi = 1)$$
$$= \int_{R_1} f_1(\mathbf{r}) d\mathbf{r}$$

It follows that the overall probabilities of correctly and incorrectly classifying companies are given by:

Probability of company 1 correctly classified as 1	$p_1 q_{11}$
Probability of company 0 misclassified as 1	$p_0 q_{10}$
Probability of company 0 correctly classified as 0	$p_0 q_{00}$
Probability of company 1 misclassified as 0	$p_1 q_{01}$

Classification Errors

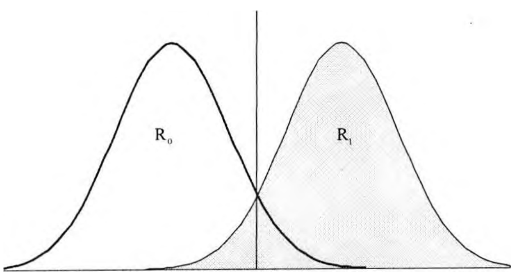
The error rate of a classification rule ξ^* is defined as $\theta(\xi^*) = \Pr(\xi^* \neq \xi)$. For a given classification rule with classification regions R_1 and R_0 the overall probability of misclassification or error rate of the classification rule can be calculated as

$$\theta(\xi^*) = \Pr(\xi = 1 / \xi^* \neq 1) + \Pr(\xi = 0 / \xi^* \neq 0) = q_{10} p_0 + q_{01} p_1$$
$$= 1 - \Pr(\xi = 1 / \xi^* = 1) + \Pr(\xi = 0 / \xi^* = 0)$$
$$= 1 - p_1 q_{11} - p_0 q_{00}$$

Given the classification regions R_1 and R_0 the error rate $\theta(\xi^*)$ can be computed from the above equation provided prior probabilities are known and provided that the probability density function is also specified.

A graphical exposition of the classification errors is shown in Figure 3.2.1.

Figure 3.2.1: Classification regions



The classification error is the area under the two curves which is common to the two distributions. In some cases it may be possible to assign a cost to the classification of an observation. To consider the cost of misclassification denote by g_{10} the cost of classifying a company from group 0 as group 1 and by g_{01} the cost of classifying a company from group 1 as group 0.

The expected cost conditional on the observation coming from population group j is denoted by $\vartheta_j(\xi^*)$.

Specifically $\vartheta_1(\xi^*) = g_{01}q_{01}$ and $\vartheta_0(\xi^*) = g_{10}q_{10}$. The unconditional expected cost is

$$\vartheta(\xi^*) = p_1\vartheta_1(\xi^*) + p_0\vartheta_0(\xi^*)$$

and the total expected cost of misclassification in the presence of costs is therefore given by

$$\vartheta(\xi^*) = g_{01}p_1q_{01} + g_{10}p_0q_{10}$$

Optimal Classification Rules

There are many optimal classification rules [see for example Seber (1986)] depending on the criterion function that is optimized. Classification rules can be obtained by

(a) Minimizing the total probability of misclassification

(b) Minimizing the expected cost of misclassification $\vartheta(\xi^*)$.

(c) Maximizing the likelihood function

(d) Maximizing the posterior probability

In what follows we shall derive the optimal rules by minimizing the expected cost of misclassification as cases (a) and (c) will be shown to be just special cases of this criterion.

The regions R_1 and R_0 that minimize $\vartheta(\xi^*)$ are given by

$$R_0 = \left\{ \mathbf{r} \in R; \frac{f_0(\mathbf{r})}{f_1(\mathbf{r})} \geq \frac{g_{01}}{g_{10}} \frac{p_1}{p_0} \right\}$$

$$R_1 = \left\{ \mathbf{r} \in R; \frac{f_0(\mathbf{r})}{f_1(\mathbf{r})} < \frac{g_{01}}{g_{10}} \frac{p_1}{p_0} \right\}$$

The proof is straightforward because by using the fact that $q_{11} + q_{01} = 1$ (since $R_1 \cup R_0 = R$) we obtain

$$\vartheta(\xi^*) = g_{01}p_1q_{01} + g_{10}p_0q_{10} = g_{01}p_1 + g_{10}p_0q_{10} - g_{01}p_1q_{11}$$

$$= g_{01}p_1 + \int_{R_1}[g_{10}f_0(\mathbf{r})p_0 - p_{01}p_1f_1(\mathbf{r})]d\mathbf{r}$$

Since $g_{ij}, p_i, q_{ij} \geq 0$ the quantity $\vartheta(\xi^*)$ will be minimal if $g_{10}f_0(\mathbf{r})p_0 - g_{01}p_1f_1(\mathbf{r}) \leq 0$ for $\mathbf{r} \in R_1$. This of course implies $g_{10}f_0(\mathbf{r})p_0 \leq g_{01}p_1f_1(\mathbf{r})$ or $\frac{g_{10}}{g_{01}} \frac{p_0}{p_1} \leq \frac{f_1(\mathbf{r})}{f_0(\mathbf{r})}$

Note that the regions only depend on the three ratios

- Density ratios $\frac{f_0(\mathbf{r})}{f_1(\mathbf{r})}$
- Cost ratios $\frac{g_{10}}{g_{01}}$

- Prior probability ratios $\frac{p_1}{p_0}$

Depending on the values of these ratios we have a number of special cases (Dinenis (2002b)):

1. Equal prior probabilities $\frac{p_1}{p_0} = 1$. In this case the classification rule becomes

$$R_0 : \frac{f_0(\mathbf{r})}{f_1(\mathbf{r})} \geq \frac{g_{01}}{g_{10}},\ R_1 : \frac{f_0(\mathbf{r})}{f_1(\mathbf{r})} < \frac{g_{01}}{g_{10}}$$

2. Equal classification costs $\frac{g_{01}}{g_{10}} = 1$. In this case the classification rule becomes

$$R_1 : \frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} > \frac{p_0}{p_1}$$

   In this case the classification rule can be expressed in terms of the posterior probabilities since $R_1 : \frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} \geq \frac{p_0}{p_1}$ is equivalent to $R_1 : f_1(\mathbf{r})p_1 > f_0(\mathbf{r})p_0$ or $R_1 : \frac{f_1(\mathbf{r})p_1}{f(\mathbf{r})} > \frac{f_0(\mathbf{r})p_0}{f(\mathbf{r})}$ or equivalently $\pi_{1r} > \pi_{0r}$ that is we compare the value of the posterior probabilities for the two categories and the category with the highest posterior probability can be identified to classify an observation.

3. The third special case is when we have both equal classification costs and equal prior probabilities

$$\frac{g_{01}}{g_{10}} = \frac{p_1}{p_0} = 1$$

   in which case the classification rules are the simple maximum likelihood rule

$$R_1 : f_1(\mathbf{r}) > f_0(\mathbf{r})$$

4. The likelihood rule is also appropriate in the case where classification costs and prior probabilities are not necessarily the same but the condition

$$\frac{g_{10}}{g_{01}} = \frac{p_1}{p_0}$$

   holds. That is the ratio of classification costs is the same as the ratio of the prior probabilities. In this case we have

$$R_1 : f_1(\mathbf{r})p_1 > f_0(\mathbf{r})p_0$$

According to the maximization of the posterior probability criterion, we would assign an observation to the group with the larger posterior probability, that is assign $\mathbf{r}$ to group 1 if $\pi_{1r} > \pi_{0r}$ which is the same as

$$\frac{p_1 f_1(\mathbf{r}^*)}{f_r(\mathbf{r}^*)} > \frac{p_0 f_0(\mathbf{r}^*)}{f_r(\mathbf{r}^*)} \quad \text{or} \quad \frac{f_1(\mathbf{r}^*)}{f_0(\mathbf{r}^*)} > \frac{p_0}{p_1}$$

So the maximisation of the posterior probability criterion leads to the same classification rules.

3.3 Normal Discriminant Analysis.

To make any of the classification rules derived so far operational we need to specify the conditional density functions. The above classification rules can be applied for any type of probability density function. Simpler classification rules are derived if we assume that $f_1(\mathbf{r})$, $f_0(\mathbf{r})$ are multivariate normal densities with mean vector $\boldsymbol{\mu}_1$, and covariance matrix $\boldsymbol{\Omega}_1$ for the first group and, mean vector $\boldsymbol{\mu}_0$, and covariance matrix $\boldsymbol{\Omega}_0$ for the second group.

$$f_1(r) = (2\pi)^{-\frac{p}{2}}|\boldsymbol{\Omega}_1|^{-\frac{1}{2}} e^{-\frac{1}{2}(r-\boldsymbol{\mu}_1)'\boldsymbol{\Omega}_1^{-1}(r-\boldsymbol{\mu}_1)}$$

and

$$f_0(r) = (2\pi)^{-\frac{p}{2}}|\boldsymbol{\Omega}_0|^{-\frac{1}{2}} e^{-\frac{1}{2}(r-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}_0^{-1}(r-\boldsymbol{\mu}_0)}$$

and the classification rule becomes

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = \ln\left\{\frac{\dfrac{1}{\sqrt{2\pi}|\boldsymbol{\Omega}_1|^{\frac{1}{2}}}\exp\left[-\dfrac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_1)'\boldsymbol{\Omega}_1^{-1}(\mathbf{r}-\boldsymbol{\mu}_1)\right]}{\dfrac{1}{\sqrt{2\pi}|\boldsymbol{\Omega}_0|^{\frac{1}{2}}}\exp\left[-\dfrac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}_0^{-1}(\mathbf{r}-\boldsymbol{\mu}_0)\right]}\right\} \geq \ln\left[\frac{p_0}{p_1}\right] + \ln\left[\frac{g_{10}}{g_{01}}\right]$$

This classification rule can be actually expressed either as a linear or as a quadratic function of the observation vector $\mathbf{r}$ and these two possibilities are explored in the next two sections.

3.3.1 Linear Classification Rules

Linear classification rules are derived if we assume that both groups have the same covariance matrix $\Omega_1 = \Omega_0 = \Omega$ in which case the densities $f_1(\mathbf{r})$ and $f_0(\mathbf{r})$ are defined as follows:

$$f_1(\mathbf{r}) = \frac{1}{\sqrt{2\pi}|\Omega|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_1)'\Omega^{-1}(\mathbf{r}-\boldsymbol{\mu}_1)\right]$$

and

$$f_0(\mathbf{r}) = \frac{1}{\sqrt{2\pi}|\Omega|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_0)'\Omega^{-1}(\mathbf{r}-\boldsymbol{\mu}_0)\right]$$

According to the classification criterion we assign the observation to group 1 if:

$$\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} \geq \frac{p_0}{p_1}\frac{g_{10}}{g_{01}}$$

Or equivalently,

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} \geq \ln\left[\frac{p_0}{p_1}\right] + \ln\left[\frac{g_{10}}{g_{01}}\right]$$

The above equation can be further simplified by substituting $f_1(\mathbf{r})$ and $f_0(\mathbf{r})$ as:

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = \ln\left\{\frac{\frac{1}{\sqrt{2\pi}|\boldsymbol{\Omega}|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_1)'\boldsymbol{\Omega}^{-1}(\mathbf{r}-\boldsymbol{\mu}_1)\right]}{\frac{1}{\sqrt{2\pi}|\boldsymbol{\Omega}|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}(\mathbf{r}-\boldsymbol{\mu}_0)\right]}\right\}$$

$$= \ln\left\{\exp\left[-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_1)'\boldsymbol{\Omega}^{-1}(\mathbf{r}-\boldsymbol{\mu}_1)+\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_2)\boldsymbol{\Omega}^{-1}(\mathbf{r}-\boldsymbol{\mu}_0)\right]\right\}$$

$$= -\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_1)'\boldsymbol{\Omega}^{-1}(\mathbf{r}-\boldsymbol{\mu}_0)+\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}(\mathbf{r}-\boldsymbol{\mu}_0)$$

$$= -\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}\mathbf{r}+\frac{1}{2}\mathbf{r}'\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)-\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_0)$$

$$= (\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}\mathbf{r}-\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_0)$$

denoting

$$\alpha = -\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_0+\boldsymbol{\mu}_1)$$

and $\boldsymbol{\beta} = \boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)$

we can write the ratio of the density functions as:

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = \alpha + \boldsymbol{\beta}'\mathbf{r}$$

and the optimal classification rule becomes:

$$\xi^*_{opt} = \begin{cases} 1 & \text{if } \alpha+\boldsymbol{\beta}'\mathbf{r} \geq \ln\left[\frac{p_0}{p_1}\right]+\ln\left[\frac{g_{10}}{g_{01}}\right] \\ 0 & \text{otherwise} \end{cases}$$

in the special case when

$$\frac{p_1}{p_0} = \frac{g_{01}}{g_{10}} = 1$$

the rule becomes

$$\alpha + \boldsymbol{\beta}'\mathbf{r} \geq 0$$

or since $\alpha = -\frac{1}{2}\boldsymbol{\beta}'(\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1)$

$$\boldsymbol{\beta}'\mathbf{r} \geq \frac{1}{2}(\boldsymbol{\beta}'\boldsymbol{\mu}_0 + \boldsymbol{\beta}'\boldsymbol{\mu}_1)$$

That is if the linear combination is above or equal to the midpoint we classify in group 1 and if it is below the mid point we classify in group 0.

Fisher Discriminant Analysis

Fisher actually arrived at the linear classification rule using an entirely different argument. Fisher's idea was to transform the multivariate observations $\mathbf{r}$ into univariate observations z so that the z's derived from populations 1 and 0 were separated as much as possible. Fisher suggested taking linear combinations of $\mathbf{r}$ to create the univariate variables $z = \boldsymbol{\beta}'\mathbf{r}$. Fisher's approach does not assume that the populations are normal, but it does implicitly assume that the population covariances are equal because a common covariance matrix is used.

$$\mathrm{d} = \frac{(\bar{z}_1 - \bar{z}_0))^2}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}} = \frac{(\boldsymbol{\beta}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0))^2}{\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}}$$

The $p \times 1$ nonzero vector, that maximises the distance is given by

$$\boldsymbol{\beta} = \gamma\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

and is not unique since it depends on the value of the constant $\gamma \neq 0$. The maximum value attained is $\mathrm{d}_{\max} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = \Delta_p^2$. That is for two populations, the maximum relative separation that can be obtained by considering linear combinations of the multivariate observations is equal to the distance Δ_p^2. The vector of weights $\boldsymbol{\beta}$ that maximises the distance is the same vector that minimises the classification error rate defined earlier.

Using Fisher's discriminant function z and its mean values for the two groups as $\bar{z}_1 = \boldsymbol{\beta}'\boldsymbol{\mu}_1$, and $\bar{z}_0 = \mathbf{w}'\boldsymbol{\mu}_0$ we can write the ratio of the density functions equivalently as

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = z - \frac{1}{2}(\bar{z}_1 + \bar{z}_0)$$

and the classification rule becomes:

Classify observation $\mathbf{r}$ into group 1 if

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = z - \frac{1}{2}(\bar{z}_1 + \bar{z}_0) \geq \ln\left[\frac{p_0}{p_1}\right] + \ln\left[\frac{g_{10}}{g_{01}}\right]$$

or equivalently

$$z \geq \frac{1}{2}(\bar{z}_1 + \bar{z}_0) + \ln\left[\frac{p_0}{p_1}\right] + \ln\left[\frac{g_{10}}{g_{01}}\right]$$

That is, the classification of a vector of observations is based on the value of a linear combination of the elements of vector $\mathbf{r}$ with weights given by $\boldsymbol{\beta}' = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)'\Omega^{-1}$.

The function $L(\mathbf{r}) = \alpha + \boldsymbol{\beta}'\mathbf{r}$ is sometimes called Fisher's linear discriminant function because it is a linear function of the vector of discriminating variables $\mathbf{r}$. The special cases examined above can be restated in terms of the values of the function $z = \boldsymbol{\beta}'\mathbf{r}$. When the classification costs are the same $\frac{g_{10}}{g_{01}} = 1$ the classification rule becomes $z \geq \frac{1}{2}(\bar{z}_1 + \bar{z}_0) + \ln\left[\frac{p_0}{p_1}\right]$.

When both the classification costs and the prior probabilities are the same, the classification rule becomes

$$z \geq \frac{1}{2}(\bar{z}_1 + \bar{z}_0)$$

This is the z – score function rule introduced by Altman (1968). Following from the theoretical exposition we have presented so far it is clear that this simple rule is valid only under restrictive assumptions, violation of which will also invalidate any inference based on it.

3.3.2 Quadratic Classification Rule

We now examine classification schemes under normality but without the assumption of equal covariance matrices. The classification criterion for classifying a company on the basis of the set of values $\mathbf{r}$ for any arbitrary density function was defined before as.

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} \geq \ln\left[\frac{p_0}{p_1}\right] + \ln\left[\frac{g_{10}}{g_{01}}\right]$$

However even under multivariate normality when $\Omega_1 \neq \Omega_0$ the ratio of densities cannot be simplified as before. In this case we have

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = \ln\left\{\frac{\frac{1}{\sqrt{2\pi}|\boldsymbol{\Omega}_1|^{\frac{1}{2}}}e^{-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_1)'\boldsymbol{\Omega}_1^{-1}(\mathbf{r}-\boldsymbol{\mu}_1)}}{\frac{1}{\sqrt{2\pi}|\boldsymbol{\Omega}_0|^{\frac{1}{2}}}e^{-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}_0^{-1}(\mathbf{r}-\boldsymbol{\mu}_0)}}\right\}$$

or

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = -\frac{1}{2}\ln|\Omega_1| + \frac{1}{2}\ln|\Omega_0| - \frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_1)'\Omega_1^{-1}(\mathbf{r}-\boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{r}-\boldsymbol{\mu}_0)\Omega_0^{-1}(\mathbf{r}-\boldsymbol{\mu}_0)$$

After collecting terms the above expression simplifies to

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = \mathbf{r}'\mathbf{A}\mathbf{r} + \mathbf{b}'\mathbf{r} - k$$

where

$$\mathbf{A} = -\frac{1}{2}\left(\Omega_1^{-1} - \Omega_0^{-1}\right)$$

$$\mathbf{b} = \Omega_1^{-1}\boldsymbol{\mu}_1 - \Omega_0^{-1}\boldsymbol{\mu}_0$$

$$k = \frac{1}{2}\ln|\Omega_1| - \frac{1}{2}\ln|\Omega_0| + \frac{1}{2}\boldsymbol{\mu}_1'\Omega_1^{-1}\boldsymbol{\mu}_1 - \frac{1}{2}\boldsymbol{\mu}_0'\Omega_0^{-1}\boldsymbol{\mu}_0$$

In this case the ratio of the density functions is a quadratic function of the observation vector. The classification rule for classifying in population 1 now becomes

$$\ln\frac{f_1(\mathbf{r})}{f_0(\mathbf{r})} = \mathbf{r}'\mathbf{A}\mathbf{r} + \mathbf{b}'\mathbf{r} - k \geq \ln\left[\frac{p_0}{p_1}\right] + \ln\left[\frac{g_{10}}{g_{01}}\right]$$

or

$$Q(\mathbf{r}) = \mathbf{r}'\mathbf{A}\mathbf{r} + \mathbf{b}'\mathbf{r} - c \geq 0$$

where

$$c = k + \ln\left[\frac{p_0}{p_1}\right] + \ln\left[\frac{g_{10}}{g_{01}}\right] \geq 0$$

$Q(\mathbf{r})$ is usually referred as the quadratic discriminant function. The boundary of the classification regions consists of all $\mathbf{r} \in R$ such that $Q(\mathbf{r}) = 0$.

The Linear discriminant function is a special case when $\Omega_1 = \Omega_0$ in which case $\mathbf{A} = 0$

$\mathbf{b} = \boldsymbol{\beta}$ and $\alpha = -c$

The posterior probability in the case of a quadratic classification rule is given by

$$\pi_{1r} = \Pr(\xi = 1/\mathbf{r} = \mathbf{r}^*) = \frac{e^{q_1(\mathbf{r}^*)}}{e^{q_1(\mathbf{r}^*)} + e^{q_0(\mathbf{r}^*)}}$$

with

$$q_i(r) = \ln[p_i f_i(r)] + \frac{p}{2}\ln(2\pi)$$

The quadratic classifier's performance can be degraded when the number of discriminating variables is large compared to the sample size due to the instability of sample estimates. In particular, the sample covariance estimate becomes highly variable and may even be singular. One way to deal with the instability of covariance estimate is to employ the linear classifier. By replacing each class covariance estimate with their average, leading to the linear classifier, the number of parameters is reduced and thus the variances of their estimates become smaller. Even though each class covariance matrix may differ substantially, studies have shown that the decrease in variances of the parameter estimates accomplished by using the linear classifier often leads to better classification performance than the quadratic classifier for small training sample size.

Although a linear classifier often performs better than a quadratic classifier for small training set size, the choice between linear and quadratic classifiers is rather restrictive.

Several methods have been proposed where the sample covariance estimate is replaced by partially pooled covariance matrices of various forms. In this formulation, some degree of regularization is applied to reduce the number of parameters to be estimated, thus improving classification performance with small training set size.

Therefore, regularization techniques can also be viewed as choosing an intermediate classifier between the linear and quadratic classifiers.

3.4 Logistic Discrimination

The second method of identifying credit risk factors is logistic regression. The logistic regression model has a close relationship with Discriminant Analysis and is derived from it in a straightforward way. Remember that for general conditional functions the probability of a vector **r** belonging to group 1 is also defined by the posterior probability

$$\pi_{1r} = \frac{f_1(\mathbf{r})p_1}{f_1(\mathbf{r})p_1 + f_0(\mathbf{r})p_0} = \frac{\dfrac{f_1(\mathbf{r})p_1}{f_0(\mathbf{r})p_0}}{1+\dfrac{f_1(\mathbf{r})p_1}{f_0(\mathbf{r})p_0}}$$

with

$\pi_{0r} = 1 - \pi_{1r}$

which implies that $\ln\dfrac{\pi_{1r}}{1-\pi_{1r}} = \ln\dfrac{f_1(\mathbf{r})p_1}{f_0(\mathbf{r})p_0}$

When the conditional distribution of **r** given ξ, is multivariate normal with the same covariance matrix in both groups then we can write

$$\ln\frac{f_1(r)p_1}{f_0(r)p_0} = \frac{1}{2}\left[(\mathbf{r}-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}(\mathbf{r}-\boldsymbol{\mu}_0) - (\mathbf{r}-\boldsymbol{\mu}_1)'\boldsymbol{\Omega}^{-1}(\mathbf{r}-\boldsymbol{\mu}_1)\right] + \ln\frac{p_1}{p_0}$$

$$= \frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1) + \ln\frac{p_1}{p_0} + (\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}\mathbf{r}$$

$$= \ln\frac{p_1}{p_0} - \frac{1}{2}\boldsymbol{\beta}'(\boldsymbol{\mu}_1+\boldsymbol{\mu}_0) + \boldsymbol{\beta}'\mathbf{r}$$

$$= \alpha_L + \boldsymbol{\beta}'\mathbf{r}$$

where $\alpha_L = \ln\dfrac{p_1}{p_0} - \dfrac{1}{2}\boldsymbol{\beta}'(\boldsymbol{\mu}_1+\boldsymbol{\mu}_0)$.

From the last expression we have

$$\frac{f_1(r)p_1}{f_0(r)p_0} = \exp(\alpha_L + \boldsymbol{\beta}'\mathbf{r})$$

and the posterior probability takes the form

$$\pi_{1r} = \frac{f_1(\mathbf{r})p_1}{f_1(\mathbf{r})p_1 + f_0(\mathbf{r})p_0} = \frac{\dfrac{f_1(\mathbf{r})p_1}{f_0(\mathbf{r})p_0}}{1 + \dfrac{f_1(\mathbf{r})p_1}{f_0(\mathbf{r})p_0}} = \frac{\exp(\alpha_L + \boldsymbol{\beta}'\mathbf{r})}{1 + \exp(\alpha_L + \boldsymbol{\beta}'\mathbf{r})}$$

Setting $z = \alpha_L + \boldsymbol{\beta}'\mathbf{r}$ we obtain the posterior probability in terms of the linear function z

$$\pi_{1r} = \frac{e^Z}{1 + e^Z}$$

the posterior probability has the functional form of the logistic function.

The above way of estimating Bayes rule is known as the density estimation approach and requires knowledge of the conditional density functions in the two groups as well as knowledge of the prior probabilities. In order to use this criterion for classification, the conditional probabilities have to be estimated first, together with the prior probabilities and then use the posterior probability formula to classify observations. This approach is also known as the Gaussian maximum likelihood discriminant rules or discriminant analysis, which was derived in the previous section starting from a slightly different standpoint;

An alternative approach to estimating the Bayes rules is through what is called the *direct function estimation approach*. In this case the posterior probabilities are estimated directly based on function estimation methodology such as regression. The main techniques that follow this approach are

- Logistic regression;
- Neural networks;
- Classification trees;
- Projection pursuit;
- Nearest neighbor classifiers

Under this approach we have to model directly the posterior probability as a function of a set of risk factors, using an appropriate function say $F(\boldsymbol{\beta}, \mathbf{r})$ such that

$$\Pr(\xi = 1) = F(\boldsymbol{\beta}, \mathbf{r})$$
$$P(\xi = 0) = 1 - F(\boldsymbol{\beta}, \mathbf{r})$$

For the function $F(\boldsymbol{\beta}, \mathbf{r})$ to be a proper probability function, it has to be bounded between 1 and 0, and this can be achieved by using functional forms that guarantee that

$$\lim_{\boldsymbol{\beta}'\mathbf{r} \to \infty} \Pr(\xi = 1) = 1$$
$$\lim_{\boldsymbol{\beta}'\mathbf{r} \to -\infty} \Pr(\xi = 1) = 0$$

The logistic function

$$F(\boldsymbol{\beta}, \mathbf{r}) = \Lambda(\boldsymbol{\beta}'\mathbf{r}) = \frac{\exp(c + \boldsymbol{\beta}'\mathbf{r})}{1 + \exp(c + \boldsymbol{\beta}'\mathbf{r})}$$

satisfies this requirement[1] and consequently the probability of an observation belonging to group 1 is given by

$$\pi_{1r} = \Pr(\xi = 1 / \mathbf{r} = \mathbf{r}) = \frac{\exp(c + \boldsymbol{\beta}'\mathbf{r})}{1 + \exp(c + \boldsymbol{\beta}'\mathbf{r})}$$

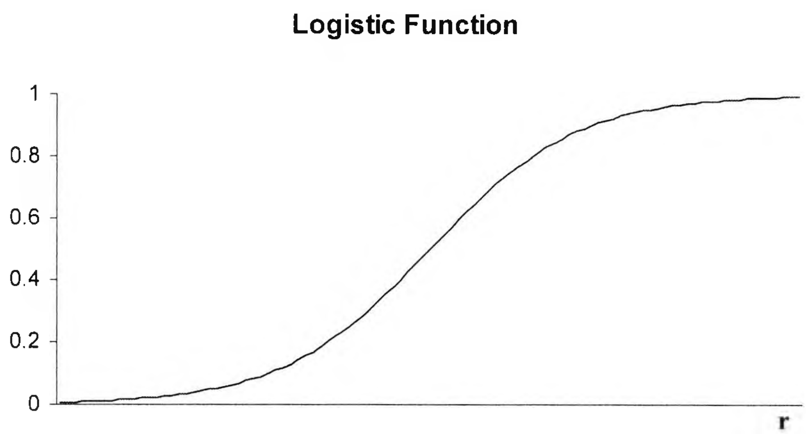
where β is the vector of parameters that determine the impact of the exogenous variables in this instance the financial ratios.

Since

$$\lim_{\boldsymbol{\beta}'\mathbf{r} \to \infty} \pi_{1r} = 1$$

$$\lim_{\boldsymbol{\beta}'\mathbf{r} \to -\infty} \pi_{1r} = 0$$

Figure 3.4.1: Logistic Function



The functional form of the function ensures that the probability will lie between 0 and 1. The functional form of the logistic probability function is shown in Figure 3. 4.1

The probability defined above can be equivalently restated as

$$g(\mathbf{r}) = \log \frac{\pi_{1r}}{1 - \pi_{1r}} = c + \boldsymbol{\beta}'\mathbf{r}$$

The function $g(\mathbf{r}) = \ln \frac{\pi_{1r}}{1 - \pi_{1r}}$ is called the logit of π and it maps the unit interval onto the real line since $g(\mathbf{r}) = c + \boldsymbol{\beta}'\mathbf{r}$.

Logistic regression does not rely on the distributional assumptions of discriminant analysis namely the multivariate normality of the risk factors and the equality of covariance matrices and it is therefore a more robust method [Lo (1986)]. Neither of these assumptions is usually supported by the data and this is the case in the empirical analysis conducted so far. As discussed in previous sections, violation of the multivariate normality assumption affects the significance of tests and the classification rates. However in cases where mild violations of multivariate normality are occurred, the much more computationally demanding logistic regression may not be warranted. Indeed as Efron (1975) has shown, when the two

---

1 Another function that satisfies this requirement is the normal distribution $F(\boldsymbol{\beta}, \mathbf{r}) = \Phi(\boldsymbol{\beta}'\mathbf{r})$ which gives rise to the Probit model

assumptions are satisfied the discriminant analysis estimates of the coefficient vector β, are more efficient than the estimates of the logistic regression .

Coefficient Interpretation

Consider the univariate logistic model

$g(r) = \beta_0 + \beta_1 r$

the regression coefficient is given by $\beta_1 = g(r+1) - g(r)$ and represents the change in the logit for a change of one unit in the independent variable r. However proper interpretation of the coefficients in a logistic regression model depends on being able to place meaning on the difference between two logits, and that in turn depends on the definition and meaning of a one unit change in the independent variable. For a dichotomous independent variable where the independent variable takes the values of r = 0 and r =1, we have $g(0) = \ln \frac{\pi(0)}{1-\pi(0)}$ and $g(1) = \ln \frac{\pi(1)}{1-\pi(1)}$

Define

$$\Psi = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}$$

then

$\ln \Psi = g(1) - g(0)$

From the definition of the odds ratio we can show that

$\Psi = e^{\beta_1}$

and therefore

$\ln \Psi = \ln e^{\beta_1} = \beta_1$

For example if $\Psi = 2$ then companies with a value of r =1, have twice as large a probability of being downgraded.

With continuous independent variables a change of one unit in the independent variable may not make sense especially if we are using ratios as in the current study and we may want to calculate the impact of a change of m units on the odds ratio. The log odds for a change of m units is obtained from the logit difference

$g(r+m) - g(r) = m\beta_1$

and the associated odds ratio is obtained by exponentiating this logit difference

$\Psi(m) = \Psi(r+m, r) = \exp(m\beta_1)$

3.5 Estimation and inference in LDA

In deriving the classification rules we have ignored so far the fact that the density function $f_i(\mathbf{r}/\xi_i)$ depends on the unknown parameters $\{\boldsymbol{\mu}_i, \boldsymbol{\Omega}_i\}$. For samples $\{\mathbf{r}_{11}, \mathbf{r}_{12}, \ldots\ldots, \mathbf{r}_{1n_1}\}$ and $\{\mathbf{r}_{01}, \mathbf{r}_{02}, \ldots\ldots, \mathbf{r}_{0n_0}\}$ from group 1 and group 0 respectively, we shall examine in this section ways of deriving estimators for the vector of parameters $\{\boldsymbol{\mu}_i, \Omega_i\}$ and consequently $\boldsymbol{\beta}$ and α, and how to test statistical hypotheses concerning the elements of the vector $\boldsymbol{\beta}$.

The joint likelihood function of $\xi, \mathbf{r}$ is given by

$$f(\xi,\mathbf{r}) = \pi(\xi/r)f(r) = \left[p_1^{n_1}\prod_{i=1}^{n_1} f_1(\mathbf{r}^i)\right]\left[p_0^{n_0}\prod_{i=n_1+1}^{n_1+n_0} f_0(\mathbf{r}^i)\right]$$

with the log likelihood function being written as

$$L(\xi,\mathbf{r}) = (2\pi)^{-\frac{p}{2}} - \frac{n}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\sum_{i=1}^{n_0}(\mathbf{r}_i - \boldsymbol{\mu}_0)'\boldsymbol{\Omega}^{-1}(\mathbf{r}_i - \boldsymbol{\mu}_0)$$

$$-\frac{1}{2}\sum_{i=1}^{n_1}(\mathbf{r}_i - \boldsymbol{\mu}_1)'\boldsymbol{\Omega}^{-1}(\mathbf{r}_i - \boldsymbol{\mu}_1) + n_0 \ln p_0 + n_1 \ln p_1$$

The maximum likelihood estimators are given by

$$\hat{\boldsymbol{\mu}}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0}\mathbf{r}_i$$

$$\hat{\boldsymbol{\mu}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\mathbf{r}_i$$

$$\hat{\boldsymbol{\Omega}}_{MLE} = \frac{1}{n}\sum_{i=1}^{n_0}(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_0)(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_0)' + \frac{1}{n}\sum_{i=1}^{n_1}(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_1)(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_1)'$$

$$\hat{p}_0 = n_0 / n$$
$$\hat{p}_1 = n_1 / n$$

and the parameters of the discriminant function are obtained using the invariance property of the maximum likelihood estimators as

$$\hat{a}_D = -\frac{1}{2}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_0)'\hat{\boldsymbol{\Omega}}^{-1}(\hat{\boldsymbol{\mu}}_0 + \hat{\boldsymbol{\mu}}_1) + \ln \hat{p}_1 / \hat{p}_0$$

$$\hat{\boldsymbol{\beta}}_D = \hat{\boldsymbol{\Omega}}_p^{-1}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_0)$$

Asymptotic standard errors based on Richard (1975) and Lo (1986) can be calculated from the formula

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_{DA} - \boldsymbol{\beta}\right) \sim N(0, \mathbf{V}_{DA})$$

where

$$\mathbf{V}_{DA} = \delta\boldsymbol{\Omega}^{-1} + 2(\boldsymbol{\mu}' \otimes \mathbf{I}_k)(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}) \times \mathbf{R}'\mathbf{Q}(\boldsymbol{\Omega} \otimes \boldsymbol{\Omega})\mathbf{Q}'\mathbf{R}(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1})(\boldsymbol{\mu} \otimes \mathbf{I}_k)$$

where

$$\delta = \frac{1}{\pi_0 \pi_1}$$

R is the $\frac{1}{k}k(k+1)\times k^2$ selection matrix such that $\mathbf{R}'\boldsymbol{\sigma} = \text{vec}(\boldsymbol{\Omega})$

$\boldsymbol{\sigma} = \{\sigma_1^2, \sigma_{12}, \cdots, \sigma_m^2\}$ is the vector of the elements of the matrix $\boldsymbol{\Omega}$

$\mathbf{Q}'$ and is the Moore-Penrose inverse of R, so that $\mathbf{R}'\mathbf{Q} = \mathbf{I}$

Alternatively, rather than using MLE the parameters can be estimated from the sample by using sample means and the sample variances. The following estimates of the means and variances are efficient

$$\hat{\boldsymbol{\mu}}_0 = \bar{\mathbf{r}}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0}\mathbf{r}_{0i}$$

$$\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{r}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\mathbf{r}_{1i}$$

and

$$\hat{\boldsymbol{\Omega}}_p = \frac{(n_1 - 1)\hat{\boldsymbol{\Omega}}_1 + (n_0 - 1)\hat{\boldsymbol{\Omega}}_0}{n_1 + n_0 - 2}$$

with

$$\hat{\boldsymbol{\Omega}}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0}(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_0)(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_0)'$$

$$\hat{\boldsymbol{\Omega}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_1)(\mathbf{r}_i - \hat{\boldsymbol{\mu}}_1)'$$

Note that the pooled estimator $\hat{\boldsymbol{\Omega}}_p$ is an unbiased estimator whereas $\hat{\boldsymbol{\Omega}}_{MLE}$ is not.

If we derive the optimal rule based on Fisher's approach we need to assume a particular value for the constant γ. Choosing the constant $\gamma = 1$, the vector $\boldsymbol{\beta}$ is estimated using sample estimates of the mean and the covariance as follows $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Omega}}_p^{-1}(\bar{\mathbf{r}}_1 - \bar{\mathbf{r}}_0)$.

3.5.1 Tests of overall significance

The estimated discriminant function is given by

$L(r) = \hat{\alpha} + \hat{\boldsymbol{\beta}}'\mathbf{r}$

and statistical inference on the coefficients of the discriminant function parameters can be based either on a classical framework in direct analogy to the linear regression framework, or in a predictive framework [Huberty (1994)] depending on whether the emphasis is on group separation or prediction. Statistical inference deals with tests concerning the vector of coefficients $\boldsymbol{\beta}$. There are two types of tests we shall employ. The first type tests the overall significance of the discriminant function and the second type tests hypotheses on individual parameters of the vector $\boldsymbol{\beta}$. As in linear regression analysis a test of the overall significance of the discriminant function i.e. a test of the hypothesis $H_0 : \boldsymbol{\beta} = 0$ leads to the derivation of summary measures of significance analogous to R^2. A test of the hypothesis $H_0 : \boldsymbol{\beta} = 0$ could be conducted using the estimated variance – covariance matrix of the regression coefficient $\mathbf{V}_{DA}$ which is a rather difficult task since it is not produced by any of the statistical packages that estimate discriminant functions. However since $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Omega}}_p^{-1}(\hat{\boldsymbol{\mu}}_0 - \hat{\boldsymbol{\mu}}_1)$ and since $\hat{\boldsymbol{\Omega}}_p^{-1}$ is assumed to be nonnegative, it follows that $H_0 : \boldsymbol{\beta} = 0$ is equivalent to $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_0$. For this hypothesis we can apply the test statistic of Chapter 2 such as Hotelling's T^2 a Likelihood Ratio Test or Wilk's lambda. Hotelling's test is defined as

$$T^2 = \frac{n_1 n_0}{n_1 + n_0}(\bar{\mathbf{r}}_1 - \bar{\mathbf{r}}_0)'\hat{\boldsymbol{\Omega}}_p^{-1}(\bar{\mathbf{r}}_1 - \bar{\mathbf{r}}_0) = \frac{n_1 n_0}{n_1 + n_0}\Delta_p^2$$

The distribution of this ratio can be derived by simple algebraic transformations. The transformation

$$H = \frac{n_1 + n_0 - p + 1}{p(n_1 + n_0)}T^2 = \frac{n_1 + n_0 - p + 1}{p(n_1 + n_0)}\frac{n_1 n_0}{n_1 + n_0}\Delta_p^2$$

has an F distribution with degrees of freedom p and $n_1 + n_0 - p + 1$ under the null hypothesis assuming independence of observation vectors, normality and equal covariances. If H_0 is rejected we can conclude that the separation between the two populations is significant.

An alternative test based on a chi-square distribution can be derived using a slightly more interesting justification which may utilize the maximum likelihood framework. The Likelihood Ratio test to test the hypothesis $H_0 : \boldsymbol{\beta} = 0$ can be written as [Flury (1995)].

$$LRS = -2\ln(LR) = (n_1 + n_0)\ln[1 + \frac{1}{n_1 + n_0 - 2}T^2]$$

Under the null hypothesis $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0$ LRS is distributed as chi-square with p degrees of freedom. Notice that when $n_1 + n_0$ is large, then $\ln[1 + \frac{1}{n_1 + n_0 - 2}T^2] \approx \frac{1}{n_1 + n_0}T^2$ [2] and

$LRS \approx T^2$ so that we may use the T^2-statistic and compare it directly to the $(1-\alpha)$ - quintile of the chi-square distribution with p degrees of freedom.

The third test of the overall significance of the discriminant function is a generalisation of the univariate Wilks' statistic and is given by

$$\Lambda = \frac{|\mathbf{SSCP}_w|}{|\mathbf{SSCP}_t|}$$

The term $|\mathbf{SSCP}_w|$ denotes the determinant of the within-groups SSCP matrix and the term $|\mathbf{SSCP}_t|$ denotes the determinant of the total-sample SSCP matrix. This can be approximated as a chi-square statistic using the transformation

$$\chi^2 = (n_1 + n_0 - 1 - \frac{p+2}{2})\ln\Lambda$$

The χ^2 statistic is distributed as a chi-square distribution with p degrees of freedom. The statistical significance of the discriminant function can also be assessed by transforming Wilks' Λ into an exact F ratio using

2 This based on the result $\ln(1+x) \approx x$ for small $|x|$

$$F = \frac{1-\Lambda}{\Lambda}\left(\frac{n_1 + n_0 - p - 1}{p}\right)$$

Which follows the F distribution with p and $(n_1 + n_0 - p - 1)$ degrees of freedom under the null hypothesis of $\boldsymbol{\beta} = 0$.

As in the linear regression framework where the value of the overall test of significance is also an indication of the association between the dependent variable and the regression in an analogous manner, the values of the overall test statistics, be it T^2 or Wilk's lambda can be used to produce an overall significance index. The following statistic has been suggested

$$\eta^2 = \frac{T^2}{T^2 + n_1 + n_0 + p - 1}$$

$$\eta^2 = 1 - \Lambda F$$

Tatsuoka (1998) suggests the following degrees of freedom adjustment in direct analogy to the adjustment to the R^2

$$\eta_{adj}^2 = \eta^2 - \frac{p^2 + q^2}{3(n_1 + n_0)}(1 - \eta^2)$$

3.5.2 Statistical inference based on individual coefficients

Now suppose that a T^2 or its variants described above has rejected the hypothesis that $\boldsymbol{\beta} = 0$. However not all potential variables that measure the characteristics of a company may be included in the discriminant function because they may not contribute to the discrimination between the two categories of companies. This is equivalent to the variable selection problem in regression analysis. We may then proceed to a finer analysis and asses, which of the variables are needed for discriminating and which would be omitted without loss of information. Such a specification test is important because although as the number of discriminating variables increases, classification prediction may not increase in analogy with the r-squared criterion in multiple regression analysis

The question may be asked whether a subset $\mathbf{r}_q$ of the p variables in $\mathbf{r} = \{\mathbf{r}_{p-q}, \mathbf{r}_q\}$ will discriminate just as well.

$$L(r) = \hat{\alpha} + \hat{\boldsymbol{\beta}}'_{p-q}\mathbf{r}_{p-k} + \hat{\boldsymbol{\beta}}'_q\mathbf{r}_q$$

There is no theory predicting which of the p variables should discriminate between the two populations. The criteria for selecting the subset of the p variables entering the function will be statistical. A test can be constructed as follows. Define the distance functions for the p and the q variables

$$\hat{\Delta}_p^2 = (\bar{\mathbf{r}}_1^p - \bar{\mathbf{r}}_0^p)'\boldsymbol{\Omega}_p^{-1}(\bar{\mathbf{r}}_1^p - \bar{\mathbf{r}}_0^p)$$

$$\hat{\Delta}_q^2 = (\bar{\mathbf{r}}_1^q - \bar{\mathbf{r}}_0^q)'\boldsymbol{\Omega}_q^{-1}(\bar{\mathbf{r}}_1^q - \bar{\mathbf{r}}_0^q)$$

The hypothesis $H_0 : \boldsymbol{\beta}_q = \mathbf{0}$ is equivalent to $H_0 : \hat{\Delta}_p^2 = \hat{\Delta}_q^2$ and a test statistic is given by

$$F = \frac{(\hat{\Delta}_p^2 - \hat{\Delta}_q^2)(n_1 + n_0 - p - 1)}{(m + \hat{\Delta}_q^2)(p-1)}$$

where

$$m = \frac{(n_1 + n_0)(n_1 + n_0 - 2)}{n_1 n_0}$$

The test statistic follows the F distribution $F \sim F^{p-q}_{n_1+n_2-p-1}$ and the hypothesis $\boldsymbol{\beta}_q = 0$ is rejected if $F > F^{p-q}_{n_1+n_0-p-1}(\alpha)$ where $F^{p-q}_{n_1+n_0-p-1}(\alpha)$ is the critical value of the $F^{p-q}_{n_1+n_0-p-1}$ at level α.

The most common test for the significance of the individual variables in the discriminant function is Wilks' lambda statistic defined already in Chapter 2 as $\Lambda = \frac{1}{1+H}$ and which can be converted into a an F ratio $F = \frac{1-\Lambda}{\Lambda}(n_1 + n_0 - 2) \sim F^1_{n_1+n_0-2}$ which follows the F distribution with 1 and $(n_1 + n_0 - 2)$ degrees of freedom under the null hypothesis of $w_i = 0$. Since $w_i = 0$ is equivalent to $\mu_{i1} = \mu_{i2}$ the F test is equivalent to the univariate t-test. In fact $F = t^2$. This is the test used by SPSS for the selection of variables in the discriminant function and the one we have employed. A mispecification test to test the statistical significance of individual variables that were excluded can be performed by employing a Hausman –type test (see e.g. Flury (1997)) which involves the regression of the excluded variable on the other variables and the indicator variable which takes the values of 1 and 0. That is we estimate the model

$$E[r_j / x, \mathbf{r}) = \sum_{i=1, i \neq j}^{p} a_i r_i + \gamma x$$

and test the hypothesis that $H_0 : \gamma = 0$ using the t-test

$$t_i = \frac{\hat{\gamma}}{se(\hat{\gamma})}$$

If the null hypothesis $H_0 : \gamma = 0$ is rejected then the variable needs to be included in the discriminant function.

In performing the above tests we should not forget that they are all based on the assumption that the covariance matrix is the same in both groups and that violation of the assumption may undermine the validity of the tests.

3.5.3 Statistical inference based on classification performance

There are many ways to assess the performance of classification rules estimated from a particular sample. The methods can be split into approaches that make explicit assumptions about the distribution of financial ratios and those which are distribution free. We present here four methods. The first two methods depend explicitly on the normality assumption. The last two are distribution free.

3.5.3.1. Misclassification Error Function Approach

The most obvious way is to use the misclassification error of misclassification cost function we defined in section 3.1 and which was used to derive the optimal classification rules. Explicit formulas for the misclassification errors can be derived under the assumption of normality using the linear discriminant function

$$L(\mathbf{r}) = \alpha + \boldsymbol{\beta}'\mathbf{r}$$

The conditional expectation of the discriminant function when $\xi = i$ is given by

$$E[L(\mathbf{r}) / \mathbf{r} \in R_i] = L(\boldsymbol{\mu}_i) = \boldsymbol{\beta}'\boldsymbol{\mu}_i - \frac{1}{2}\boldsymbol{\beta}'(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = \frac{1}{2}(-1)^{i+1}\Delta_p^2$$

where

$$\Delta_p^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \boldsymbol{\beta}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}$$

The variance of the discriminant function is given by

$$V[L(\mathbf{r})] = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta} = \Delta_p^2$$

For sample estimates $\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2 \hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\mu}}_2 \hat{\boldsymbol{\Omega}}$ we derive estimates of $\hat{\boldsymbol{\beta}}, \hat{\Delta}_p^2$ and of the misclassification probabilities from

$$\hat{q}_{10} = P[L(r) \le \ln c / r \in R_0] = P\left(\frac{L(r) - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p} \le \frac{\ln c - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p}\right)$$

$$= P\left(Z \le \frac{\ln c - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p}\right) = \Phi\left(\frac{\ln c - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p}\right)$$

and

$$\hat{q}_{01} = P[L(\mathbf{r}) > \ln c / \mathbf{r} \in R_1] = P\left(\frac{L(\mathbf{r}) - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p} > \frac{\ln c - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p}\right)$$

$$= P\left(Z > \frac{\ln c - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p}\right) = \Phi\left(\frac{-\ln c - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p}\right)$$

where Φ is the CDF of the standardized normal distribution and $\ln c = \ln\left[\frac{p_2}{p_1}\right] + \ln\left[\frac{g_{12}}{g_{21}}\right]$

The total misclassification error is then given by

$$\hat{\theta} = p_1\Phi\left(\frac{\ln c - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p}\right) + p_2\Phi\left(\frac{-\ln c - \frac{1}{2}\hat{\Delta}_p^2}{\hat{\Delta}_p}\right)$$

The hit rate estimates are determined by

$$\hat{q}_{00} = 1 - \hat{q}_{10}$$

and

$$\hat{q}_{11} = 1 - \hat{q}_{01}$$

The overall success probability can be estimated from

$$S = \hat{p}_1\hat{q}_{11} + \hat{p}_0\hat{q}_{00}$$

On the basis of the above criterion we select the classification rule with the minimum misclassification error or the maximum hit rate. However it has been shown that the plug-in estimates of the true success rate $p_1q_{11} + p_0q_{00}$ are biased and tend to overestimate the true success rate. The following correction has been proposed to the estimate of Δ_p^2 (Anderson (1984))

$$\tilde{\Delta}_p^2 = \frac{n_1 + n_0 - p - 3}{n_1 + n_0 - 2}\hat{\Delta}_p^2 - \frac{p(n_1 + n_0)}{n_1 n_0}$$

The above hit rate estimates depend on the normality assumption as well as on the assumption of equal covariances. Under these assumptions variable selection depends on various functions of Δ_p^2. The vector $\boldsymbol{\beta}$ that maximizes the likelihood function and η^2, will also maximize the success rate $S = p_1q_{11} + p_0q_{00}$. However the two criteria in practice will produce different results because of sampling errors and because of deviations from the normality and equal covariance matrices assumption.

3.5.3.2 Posterior Probability Approach

A second predictive framework method of selecting and evaluating discriminating variables is via the posterior probabilities.

$$\hat{\pi}_{1r} = \frac{\hat{p}_1 f_1(\mathbf{r}/\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Omega}})}{\hat{p}_1 f_1(\mathbf{r}/\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Omega}}) + \hat{p}_0 f_0(\mathbf{r}/\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Omega}})}$$

This method is usually termed the Maximum Posterior Probability [Huberty (1994)] and depends again on the assumptions of normality and equal covariance matrices. According to this criterion we select the risk factors that maximize the posterior probability. This is the function that SPSS uses to estimate classification errors.

3.5.3.3 Apparent Error Rate

The third criterion, which is a distribution-free classification performance measure, is the Apparent Error Rate (APER) [Johnson and Wichern (2002)] which gives the proportion of observations misclassified when the classification rule is applied to the sample data. This is a very popular measure because of its simplicity and is defined by introducing the binary random variable

$$e_i = \begin{cases} 0 & \text{if } \xi_i = \xi_i^* \\ 1 & \text{if } \xi_i \neq \xi_i^* \end{cases}$$

The plug-in error (i.e. estimated error) is given by

$$\hat{\theta}_{APER} = \frac{1}{n_1 + n_0} \sum_{i=1}^{n} e_i$$

the above formula does not differentiate between the two misclassification cases but it is possible to evaluate the performance with regard to one category.

The APER can be calculated using the 2 x 2 classification table below which tallies correct and incorrect estimates.

	Predicted Group		
Actual Group	Group 0	Group 1	Totals
Group 0	n_{00}	n_{10}	n_0
Group 1	n_{01}	n_{11}	n_1
Totals	$n_{00} + n_{01}$	$n_{10} + n_{11}$	$n_1 + n_0$

Let n_0 denote the number of cases that truly belong to Group 0 (companies that have not migrated) and n_1 denote the number of cases that truly belong to Group 1 (companies that have migrated).

Then,

n_{00} = Number of cases that belong to Group 0 and are assigned to Group 0 (i.e. correctly classified)

n_{10} = Number of cases that belong to Group 0 but are assigned to Group 1 (i.e. incorrectly classified)

n_{01} = Number of cases that belong to Group 1 but are assigned to Group 0 (i.e. incorrectly classified)

n_{11} = Number of cases that belong to Group 1 and are assigned to Group 1 (i.e. correctly classified)

$n_{11} + n_{01}$ = Total number of cases assigned to Group 1

$n_{10} + n_{00}$ = Total number of cases assigned to Group 0

$n_{11} + n_{00}$ = Total number of cases correctly classified

The columns are the two predicted values of the dependent, while the rows are the two observed (actual) values of the dependent. In a perfect model, all cases will be on the diagonal and the overall percent correct will be 100%. As in the case of MDA we are looking at the three rates that measure overall performance $\frac{n_{11}+n_{00}}{n_1+n_0}$ the performance in relation to the companies that were not downgraded $\frac{n_{11}}{n_1}$ and

the performance in relation to the companies that were downgraded $\frac{n_{00}}{n_0}$.

3.5.3.4 Leave-one-out error rate

The final variable selection and classification performance measure is the leave-one-out error rate which is obtained as follows. Omit the ith observation from the sample and calculate the classification rule $\hat{x}^*_{-i}$ using the remaining n_1+n_0-1 observations. Apply the classification rule $\hat{x}^*_{-i}$ to omitted observation and check if it would be classified correctly. This process is repeated n_1+n_0 times, once for each observation. Define $e_{i,-i}=0$ if group membership of the ith observation was predicted correctly by $\hat{x}^*_{-i}$ and $e_{i,-i}=1$ otherwise. The leave-one-out error rate (L-O-O) is given by

$$\hat{\theta}_{L-O-O}=\frac{1}{n_1+n_0}\sum_{i=1}^{n}e_{i,-i}$$

The methods of selecting a classification rule are used at the same time to assessing the rule. This is similar to the use of forecasting errors to use select the appropriate regression model. This framework is not as flexible to work and especially to test the statistical significance of individual variables

3.5.3.5 Evaluation of Classification Performance

Significant separation, as expressed by the statistical significance of the discriminating parameters in the discriminant function, does not necessarily imply good classification. The efficacy of the classification procedures can be evaluated independently of any test of separation. On the other hand, if the separation is not significant, the search for a useful classification rule will be fruitless. [Johnson and Wichern (2002)]

The central question in evaluating predictive discriminant analysis is whether the classification rule does better than a rule allocating cases randomly to either of the two groups. A simple benchmark that has been suggested in the literature is the maximum chance criterion (MCC) which is defined as

$$MCC=\frac{n_0}{n_1+n_0}$$

This is the proportion of companies that are expected to be classified as non-migrating by chance. For samples that differ slightly, the value of benchmark will be about 0.50 and this may be an acceptable benchmark. However in cases where the two samples differ substantially, for example when the sample of non migrating companies is as large as eighty percent of the total sample, then classifying all companies as "non-migrating" will result in the benchmark being satisfied. Have such a classification scheme will have disastrous effects for a bank.

A second criterion the proportional Chance Criterion (PCC) on the other hand, assumes that the proportion of correct forecasts due entirely to chance is given by

$$PCC=\left(\frac{n_0}{n}\right)^2+\left(\frac{n_1}{n}\right)^2$$

For tests of equal size we can statistically test the difference between the success rate of a classification benchmark using a t-test

$$t = \frac{\frac{n_{11} + n_{00}}{n} - 0.5}{\sqrt{0.5(1-0.5)}}$$ with n-2 degrees of freedom .

For unequal sizes we can use Press's Q statistic defined as

$$Q = \frac{[n - 2(n_{11} + n_{00})]^2}{n-1}$$

which follows a chi-square distribution with one degree of freedom.

3.6 Estimation and inference in Logistic Regression

The estimation and inference procedures for the regression analysis are more akin to the classical regression framework than to the predictive framework of discriminant analysis.

The dependent variable, say ξ, in the logistic regression is binary and take the value of 1 if the company has migrated or the value of 0 if the company has not migrated.

The probability that a company has migrated $\Pr(\xi = 1)$ is given by

$$\pi_{1r} = \Pr(\xi = 1/\mathbf{r}) = \frac{\exp(\alpha_L + \boldsymbol{\beta}'\mathbf{r})}{1 + \exp(\alpha_L + \boldsymbol{\beta}'\mathbf{r})}$$

with associated logit function

$$g(\mathbf{r}) = \ln\frac{\pi_{1r}}{1-\pi_{1r}} = \alpha_L + \boldsymbol{\beta}'\mathbf{r}$$

The value of the dependent variable (i.e. the log of odds) is not available but if we consider each observation as a Bernoulli trial, we have for the ith observation

$$P(\xi_i) = \pi^{\xi_i}(1-\pi)^{1-\xi_i} \text{ for } \xi_i = 0,1$$

Assuming that the n observations of the sample are independent, then

$$P(\xi_1, \xi_2, \ldots, \xi_n) = P(\xi_1)P(\xi_2)\cdots P(\xi_n)$$

and the likelihood function is given by

$$L = \prod_{i=1}^{n} \pi_i^{\xi_i}(1-\pi_i)^{1-\xi_i} = \prod_{i=1}^{n}\left(\frac{e^{\alpha_L + \beta'\mathbf{r}_i}}{1 + e^{\alpha_L + \beta'\mathbf{r}_i}}\right)^{\xi_i}\left(\frac{1}{1 + e^{\alpha_L + \beta'\mathbf{r}_i}}\right)^{1-\xi_i}$$

where $\mathbf{r}_i$ is the vector of ratios for company i. the log of the likelihood function is given by

$$\ln L = \sum_{i=1}^{n} \xi_i \ln\frac{e^{\alpha_L + \beta'\mathbf{r}_i}}{1 + e^{\alpha_L + \beta'\mathbf{r}_i}} + \sum_{i=1}^{n}(1-\xi_i)\ln\left(\frac{1}{1 + e^{\alpha_L + \beta'\mathbf{r}_i}}\right)$$

the vector of coefficients β and the intercept can be estimated by maximising the above function. The first order conditions are

$$\frac{\partial \ln L}{\partial \alpha_L} = \sum_{i=1}^{n}(\xi_i - \pi_i) = 0 \quad \frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}(\xi_i - \pi_i)\mathbf{r}_i = \mathbf{0}$$

The above equations are highly nonlinear in the vector of parameters and can be solved using the Newton-Raphson algorithm. Once we have obtained the maximum likelihood estimates $\hat{\alpha}_L, \hat{\boldsymbol{\beta}}$, we can obtain the estimated probability of migration of company i with vector of and expected frequencies from

$$\hat{\pi}_i = \frac{\exp(\hat{\alpha}_L + \hat{\boldsymbol{\beta}}'\mathbf{r}_i)}{1 + \exp(\hat{\alpha}_L + \hat{\boldsymbol{\beta}}'\mathbf{r}_i)}$$

and $\hat{\xi}_i = \hat{\pi}_i$

For given maximum likelihood estimates $\hat{\alpha}_L, \hat{\boldsymbol{\beta}}$ we can evaluate the log-likelihood function at its maximum. Writing the maximum as a function of the estimated probabilities we have.

$$\ln L(\hat{\pi}_1, \hat{\pi}_0) = \sum_{i=1}^{n} \xi_i \ln \hat{\pi}_1 + \sum_{i=1}^{n} (1 - \xi_i) \ln(1 - \hat{\pi}_1)$$

The value of the likelihood function $\ln L(\hat{\pi}_1, \hat{\pi}_0)$ can be compared to the value of the log-likelihood function of different models and assess the significance of individual of group of parameters.

The joint distribution of the parameter estimates is approximately normal with mean and variance given by

$$\begin{pmatrix} \hat{\alpha}_L \\ \hat{\boldsymbol{\beta}} \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_L \\ \boldsymbol{\beta} \end{pmatrix}, \boldsymbol{\Omega} \right)$$

$$\hat{\boldsymbol{\Omega}} = \left[\mathbf{I}(\hat{\alpha}_L, \hat{\boldsymbol{\beta}}) \right]^{-1}$$

where $\mathbf{I}(\hat{\alpha}_L, \hat{\boldsymbol{\beta}})$ is the information matrix defined given by

$$\mathbf{I}(\hat{\alpha}_L, \hat{\boldsymbol{\beta}}) = \begin{bmatrix} \frac{\partial^2 \ln LF}{\partial \alpha_L^2} & \frac{\partial^2 \ln LF}{\partial \alpha_L \partial \boldsymbol{\beta}} \\ \frac{\partial^2 \ln LF}{\partial \alpha_L \partial \boldsymbol{\beta}} & \frac{\partial^2 \ln LF}{\partial \boldsymbol{\beta}^2} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \pi_i (1 - \pi_i) & \sum_{i=1}^{n} \xi_i \pi_i (1 - \pi_i) \\ \sum_{i=1}^{n} \xi_i \pi_i (1 - \pi_i) & \sum_{i=1}^{n} \xi_i^2 \pi_i (1 - \pi_i) \end{bmatrix}$$

3.6.1 Testing the Significance of the Model.

Once we have fit a particular multivariate logistic regression model, we need to assess the statistical significance of the overall model as well as the significance of individual coefficients. The overall performance of the model is evaluated by comparing the value of the likelihood function for the unconstrained and the constrained model (i.e. the model that contains only the constant term). That is the null hypothesis $H_0 : \beta_j = 0 \quad j = 1,2,\ldots\ldots, p$ can be tested using the likelihood ratio test[3]

$$LR = 2[\ln L_p - \ln L_0]$$

Where L_0 and L_p denote the value of the likelihood function for models containing only the intercept, and the model containing the intercept plus the p covariates, respectively. The statistic LR is distributed approximately as a chi-square distribution with p degrees of freedom for large samples. How close the

3 Both $\ln L_0$ and $\ln L_p$ are negative numbers with $|\ln L_0| > |\ln L_p|$.

distribution of the likelihood ratio test is to the chi-square distribution depends on the size of the sample. Kent (1983) shows for instance that there is still substantial upward bias even for large samples. The likelihood ratio test is an overall test of significance that does not assume that every independent variable is individually significant. The null hypothesis is that none of the independent variables are linearly related to the log odds of the dependent.

An alternative statistic is also commonly used to test the overall significance of the model. It is based on the deviance residuals and is defined as

$$d_i(\xi_i,\hat{\pi}_i) = (-1)^{1-\xi_i}\sqrt{2\xi_i \ln\left(\frac{\xi_i}{\hat{\pi}_i}\right) + 2(1-\xi_i)\ln\left(\frac{1-\xi_i}{1-\hat{\pi}_i}\right)}$$

where $\hat{\pi}_i$ is the estimated probability for case i.

When $\xi_i = 1$

$$d_i(1,\pi_i) = \sqrt{-2\ln(\pi_i)}$$

When

$\xi_i = 0$

$$d_i(1,\pi_i) = -\sqrt{-2\ln(1-\pi_i)}$$

Large values for the deviance residuals indicate that the model does not fit the case well. For large samples the deviance residuals is approximately normally distributed and therefore we can base our empirical evaluation of the fit on the normal distribution tables.

A summary statistic based on the deviance residuals can be calculated by taking the sum of the square of the deviance residuals and is shown below

$$DV = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n}\left[2\xi_i \ln\left(\frac{\xi_i}{\hat{\pi}_i}\right) + 2(1-\xi_i)\ln\left(\frac{1-\xi_i}{1-\hat{\pi}_i}\right)\right] = -2\sum_{i=1}^{n}\left[\xi_i \ln\left(\frac{\hat{\pi}_i}{\xi_i}\right) + (1-\xi_i)\ln\left(\frac{1-\hat{\pi}_i}{1-\xi_i}\right)\right]$$

The above statistic is called the deviance of a model and plays the same role for the logistic regression as the residual sum of squares plays in linear regression. It approximately follows the chi-square distribution with n-p-1 degrees of freedom.

Tests for the statistical significance of a single variable are often performed using the Wald Statistic defined as the ratio

$$\text{wald} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

For large n if the hypothesis of redundancy of the jth variable holds true, the Wald ratio will be approximately standard normal. As a result, under the null hypothesis that $\beta_j = 0$ the ratio

$$\text{wald}^2 = \left(\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}\right)^2$$

follows a chi-square distribution with one degree of freedom, and this is the statistic used in SPSS. Unfortunately as Hauck and Donner (1977) note, the Wald statistic has a very undesirable property. When the absolute value of the regression coefficient becomes large, the estimated standard error is too large. This produces a Wald statistic that is too small, resulting in a failure to reject the null hypothesis when it should be rejected. In this situation it is safer to use the Likelihood Ration test $LR = 2[\ln L_p - \ln L_{p-1}]$ to test the significance of the pth variable.

3.6.1 Goodness of fit statistics

Goodness of fit statistics describe how effective the model we have estimated is in describing the outcome variable. There are three different statistics that used to this purpose in logistic regression.
The first one, is the Hosmer and Lemeshow's Goodness of Fit Test, and tests the null hypothesis that the data were generated by the model fitted by the researcher. The test divides subjects into deciles based on predicted probabilities, then computes a chi-square from observed and expected frequencies. Then a probability (p) value is computed from the chi-square distribution with 8 degrees of freedom to test the fit of the logistic model. If the Hosmer and Lemeshow Goodness-of-Fit test statistic is .05 or less, we reject the null hypothesis that there is no difference between the observed and model-predicted values of the dependent. If the H-L goodness-of-fit test statistic is greater than .05, as we want, we fail to reject the null hypothesis that there is no difference, implying that the model's estimates fit the data at an acceptable level. This does not mean that the model necessarily explains much of the variance in the dependent, only that however much or little it does explain is significant. As with other tests, as the sample size gets larger, the H-L test's power to detect differences from the null hypothesis improves
The second set of goodness –of- fit statistics are based on pseudo R^2 measures. There is no widely accepted logistic regression analogue to OLS regression's R^2. This is because an R^2 measure seeks to make a statement about the "percent of variance explained," but the variance of a dichotomous or categorical dependent variable depends on the frequency distribution of that variable. For a dichotomous dependent variable, for instance, variance is at a maximum for a 50-50 split and the more lopsided the split, the lower the variance. This means that R-squared measures for logistic regressions with differing marginal distributions of their respective dependent variables cannot be compared directly, and comparison of logistic R-squared measures with R^2 from OLS regression is also problematic.
Nonetheless, a number of logistic R-squared measures have been proposed the most common of which is the Cox and Snell's R-Square which is an attempt to imitate the interpretation of multiple R-Square based on the likelihood, but its maximum can be (and usually is) less than 1.0, making it difficult to interpret.
This statistic which is part of SPSS output is defined as

$$R^2_{CS} = 1 - \exp\left(\frac{1}{n}(2\ln L_0 - 2\ln L_p)\right)$$

The maximum value of this quantity is $1 - \exp\left(\frac{2}{n}\ln L_0\right)$ attainable if the model is perfect.

The Nagelkerke's R-Square, [Nagelkerke (1991)] is a modification of the Cox and Snell coefficient to assure that it can vary from 0 to 1. This is achieved by dividing the Cox and Snell's R^2 by its maximum.

$$R^2_{NG} = \frac{R^2_{CS}}{1 - \exp\left(\frac{2}{n}\ln L_0\right)}$$

Therefore Nagelkerke's R-Square will normally be higher than the Cox and Snell measure. It is also part of SPSS output.
Another popular measure of fit is McFadden's R-square defines as

$$R_M^2 = 1 - \frac{\ln L_p}{\ln L_0}$$

This measure varies between 0 and 1 with the value of zero being attained when all the coefficients apart from the constant term are zero, in, which case

$\ln L_p = \ln L_0$ and $R_M^2 = 1 - 1 = 0$

The value of 1 is attained when the model is perfect and the likelihood function is therefore 1. In this case

$\ln L_p = \ln(1) = 0$ and $R_M^2 = 1 - 0 = 1$.

The third statistic that is used to asses the performance of the model is the proportion of correctly predicted classifications as in the case of the discriminant analysis. Where such a criterion is valid in the logistic regression context is subject to debate [see Green (2000) p 835 for a discussion]. There are two issues here. The first one is that we use a criterion, which does not play any role in variable selection, that is, it is not based on the likelihood function. The only case again when the maximum likelihood parameter estimates would minimise some classification error function would be if the independent variables were jointly multivariate normal. Such a link does not exist of course when ad hoc error functions are used, such as the proportion of wrong predictions. In general one would expect a weak link, between a model that is chosen on the basis of maximising the likelihood function of the sample and the ad hoc prediction error. Hosmer and Lemeshow (1989) highlight the point using the following example in the case of a single independent variable with conditional probabilities $r/\xi = 0 \sim N(0,1)$ and $r/\xi = 1 \sim N(\mu,1)$ for the two populations.

In this model the slope coefficient for the logistic regression is the same as in the linear discriminant function and is given by $\beta_1 = \mu$ whereas the intercept is given by $\alpha_L = \ln(\pi_1/\pi_0) - \frac{1}{2}\mu^2$. The probability of misclassification in this case is given by

$$\hat{\theta} = \pi_1 N\left(\frac{\ln(\pi_0/\pi_1) - \frac{1}{2}\beta_1^2}{\beta_1}\right) + \pi_0 N\left(\frac{\ln(\pi_1/\pi_0) - \frac{1}{2}\beta_1^2}{\beta_1}\right)$$

which is a special case of the formula we derived in Section 3.5 for a single discriminating variable. "The expected error term is a function of the magnitude of the slope, not necessarily of the fit of the model. Accurate or inaccurate classification does not address our criteria of goodness of fit; that the distance between observed and expected values be unsystematic, and within the variation of the model." [Hosmer and Lemeshow (1989) p 147].

The second issue is the cut-off point used for classification of an observation in one of the two groups. If we use one of the ad hoc classification procedures based as before on the hit ratios

$\frac{n_{00}}{n_0}$ $\frac{n_{11}}{n_1}$ and $\frac{n_{11} + n_{00}}{n_1 + n_0}$

then we need to specify such a cut-off value. The usual cut-off point is 0.50, i.e. an observation is assigned to group 1 if the posterior probability associated with this observation exceeds 0.5. If the sample is unbalanced, that is one of the two groups has many more observations than the other, as it is the case here with our sample, then a cut-off point of 0.5 may never classify an observation in group 1. To consider a specific case, let us look at our 1997 sample. For that year our total sample of 404 companies is divided into 356 companies for group 0 (88.1%) and 48 companies for group 1 (11.9%). The average predicted probability in the sample will be about 0.12. It will require an extreme configuration of the independent variables, requiring them to take values not observed in the data, for the probability to exceed 0.5. If this is the criterion, then the model will always fail to classify observations in group 1. The obvious solution is to reduce the cut-off point so as to increase the number of correct classifications in group 1. However in doing so, we reduce the number of correct classifications in group 0. The problems are highlighted using the following example based on the estimation of a univariate logistic regression for 1997 with UPL as the only independent variable. The estimated logistic regression is given by

$$\hat{\pi}_{1r} = \frac{\exp(-1.977 - 0.871r)}{1 + \exp(-1.977 - 0.871r)}$$

with

$$\text{wald}(\hat{\alpha}_L) = 162.562 \quad \text{wald}(\hat{\beta}) = 4.252$$

$$2\ln L_0 = -294.557 \quad 2\ln L_p = -289.096 \text{ and } 2(lnL_p - lnL_0) = 5.462$$

$$R^2_{CS} = 0.013 \quad R^2_{NG} = 0.026$$

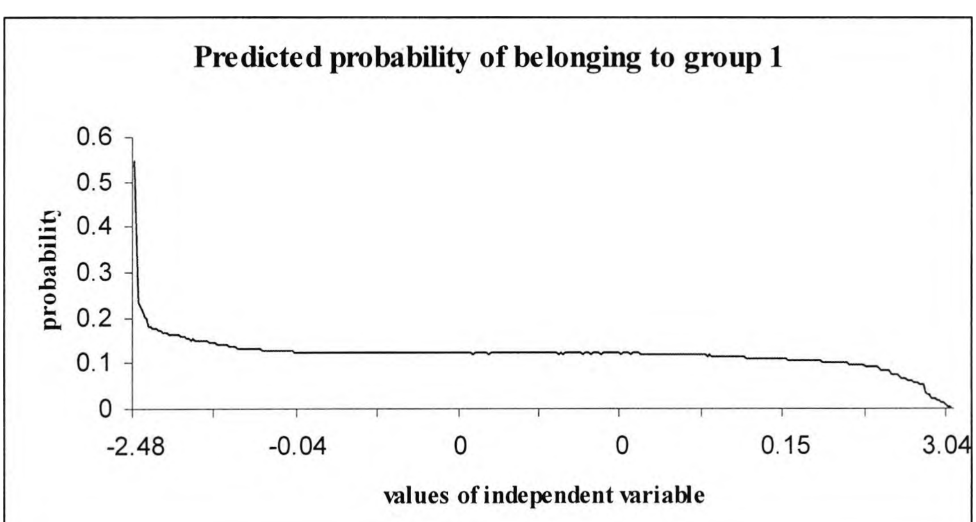
The estimated parameters are significant and the inclusion of the independent variable contributes to the explanation of the behaviour of dependent variable. When we assume a cut-off point of 0.15 the classification statistics are given bellow

$$\frac{n_{00}}{n_0} = 93.3\% \quad \frac{n_{11}}{n_1} = 20.8\% \quad \frac{n_{11} + n_{00}}{n_1 + n_0} = 84.7\%$$

If we reduce now the cut-off point to the average probability of belonging to group 1, that is to 0.12 we get the following classification results

$$\frac{n_{00}}{n_0} = 37.1\% \quad \frac{n_{11}}{n_1} = 85.4\% \quad \frac{n_{11} + n_{00}}{n_1 + n_0} = 42.8\%$$

The latter results represent a complete reversal of the classification performance of this rule. The following graph shows the predicted probabilities and why it is impossible to get sensible results using the 0.5 cut-off point. It is seen from the graph that only one, out of 48 group 1 observations (or one out of 404 total observations) exceeded the level of 0.5. The level of -2.58 required is an extreme value in terms of the sample means and standard deviations as they are reported in Chapter 2.



In the empirical analysis of the logistic approach we shall therefore adopt as the cut-off point the sample average probability of belonging to group 1 rather the 0.5 level.

3. 7 Empirical Results – Discriminant Analysis

The purpose of this section is to utilise the framework we developed in the previous sections and identify the characteristics of a firm that explain or predict group membership. The identification of risk factors and the quantification of their impact are done through the estimation of discriminant functions for every year of our sample. As we have already discussed in Section 3.5 there are two general approaches to the identification of discriminating factors in discriminant analysis depending on whether the main interest is to explain or predict group membership.

Explanation of group membership is done by using what Huberty (1994) (see also Dinenis (2002b)) calls descriptive discriminant analysis (DDA). In this approach, the specification and estimation of the discriminant functions is based on classical statistical criteria such as Wilk's lambda, or a distance measure and selects the risk factors that maximise the separation of the sample means of the two groups.

If prediction of group membership is on the other hand the main objective of risk factor identification, then the specification of the discriminant functions and the selection of the risk factors will be based on the classification performance of the discriminant function as measured by one of the error functions we have examined so far. This approach is known as Predictive Discriminant Analysis (PDA).

Unfortunately for the reasons we have already discussed in Section 3.5 the two approaches may not produce the same list of risk factors even if the theoretical conditions such as multivariate normality and equal covariance matrices are met because the empirical counterparts of the theoretical error measures are biased.

In this section we have pursued the estimation of discriminant functions using both approaches. First we estimate discriminant functions within the DDA framework and then we repeat our specification search within the PDA framework. The whole empirical analysis of this section has been performed using the software package SPSS version 10.

3.7.1 Descriptive Discriminant Analysis

The DDA framework is similar to that of regression analysis, in the sense that we need a well defined variable selection and specification search strategy in order to find the best equation that fits the data. Our variable selection starts with the estimation of a univariate discriminant function for all the risk factors for all the years. Univariate discriminant is useful for a number of reasons. First, we can test the significance of the factors prior to inclusion into the set of risk factors to be considered by the researcher. Variables which are not significant could be excluded from consideration. As such a univariate analysis is the same as the univariate test of the sample means that was conducted in Chapter 2 which showed that all the variables were significant at least in some years we have included all the variables in the set to be considered as candidates for inclusion in the discriminant function. Second, univariate analysis may still reduce he number of variables that should be considered as candidates for inclusion in the discriminant function by looking at the explanatory power of individual factors. Factors with very low discriminating power could be excluded from consideration in the specification search. Thirdly, since we have more that one ratio representing a particular characteristic such as liquidity, or profitability, univariate analysis may help us choose the best one for inclusion in the discriminant equation.

There are well known dangers of course in relying too much on a univariate approach, as it is well known that the inclusion or exclusion of other variables in the equation affects the significance of a variable. The discriminating performance of each single-variable equation is measured by the square of the canonical correlation coefficient η^2 which is defined as $\eta^2 = 1 - \Lambda$. In order to make the comparison between variables and across the years easier we have multiplied its value by 100. The results from the univariate approach are shown in Table 3.7.1.

The performance of the individual variables varies considerably. The profitability ratios (ROA and ROE) show consistently good performance. Of the solvency ratios (IPS, IPF, and IPE) the first two have strong discriminating power for most of the years with IPS slightly outperforming IPF. Of the leverage ratios DR is consistently high whereas DC has low discriminating power. Of the liquidity variables, WA is consistently high whereas CR is not. Of the growth ratios, RTA is high but the other two and especially size does not play any role in discriminating between the two groups. Finally, the activity ratios are very inconsistent with NIS having high values occasionally.

Table 3.7.1: Univariate Discriminant Criterion $100 \times \eta^2$

	94-96	95-97	96-98	97-99	98-00	99-01
ROA	1.246	2.464	1.750	0.897	1.526	2.581
ROE	0.826	1.119	2.068	1.402	0.407	0.208
IPS	2.373	1.106	6.921	7.008	1.040	4.785
IPF	0.469	1.710	6.577	7.797	6.650	1.764
IPE	1.086	0.733	0.116	1.383	0.498	0.703
DR	0.936	1.684	1.202	2.647	0.863	2.798
DC	0.003	0.047	0.234	0.802	0.520	1.376
CR	0.154	0.179	0.199	1.092	0.926	1.121
WA	2.303	3.290	4.815	2.885	3.290	4.220
UPL	0.691	0.702	1.740	0.773	0.291	1.134
RTA	2.283	4.401	3.156	3.993	1.212	5.460
SIZE	0.944	0.032	0.167	0.122	0.647	0.006
SOA	0.018	0.854	0.036	0.005	2.474	0.497
NIS	2.322	2.240	2.376	0.538	0.213	5.050
OS	0.110	1.489	0.362	0.131	0.039	0.316
ES	0.112	1.867	0.215	0.317	0.063	0.158

The next step is to devise a method of selecting the best subset of the risk factors. A widely used variable selection procedure which is employed by SPSS and other packages to estimate discriminant functions based on the separation of groups principle is the stepwise procedure which comes in three versions, forward, backward or combined, [see Sharma(1996) for a description].

In forward selection we choose from a univariate discriminant analysis the variable with the lowest Wilk's Lambda value or highest distance value, or highest canonical correlation. The second variable that is entered is the one that creates, together with the first variable the lowest Wilk's ratio. The procedure continues until no other variable increases the statistical criterion used. On the basis of the univariate analysis we can for example start a forward stepwise selection process by choosing the variable with the highest η^2 and then inserting variables that increase the criterion function. In this case of course one should use the adjusted statistics η^2.

In the backward selection, we start with all the variables in the discriminant function and remove one variable at each step. The variable that is removed is a redundant variable in the sense that the statistical criterion used does not change. The procedure continuous until there are no redundant variables left in the equation.

The combined method, sometimes called Stepwise Selection, combines both approaches. This time we start with no variables in the equation, and a variable is added or deleted at each step depending on its impact on the statistical criterion. A variable already in the discriminant function is removed if it does not significantly lower the discriminant power as measured by the statistical criterion. If no variable is removed at a given step then the variable that significantly adds the most discriminating power, as measured by the statistical criterion is added to the discriminating function. The procedure stops when at a given step no variable is added or removed from the discriminating function.

Each of the three procedures gives the same discriminant function if the variables are not correlated among themselves. However, the results could be very different if there is substantial amount of multicollinearity in the data.

SPP version 10 which was used for the empirical estimation of discriminant functions, uses the combined method only, and this variant of stepwise discriminant analysis was employed with the following criteria:

(a) Wilk's Λ was used as the selection criterion. That is, at each step either a variable is added or deleted from the discriminant function according to the value of Wilk's Λ.
(b) A tolerance level of 0.001 was used
(c) Priors were assumed to be equal
(d) The maximum significance of the F distribution for a variable to enter was 0.05 and the minimum significance for a variable to be removed was 0.10.

The results from the stepwise approach are shown in Table 3.7.2. The results of the stepwise approach are disappointing. The major problem is that for three years (1995, 1997 and 1999) out of the six years of our sample, the coefficients in the discriminant function have the wrong sign. They imply that companies with high debt equity ratios should be classified in group 0 (non migrating companies) and companies with high profitability should be classified in group 1 (financially distressed companies). Although the coefficients in the discriminant function are not uniquely defined, their sign must conform to the theoretically expected sign.

Focusing now on the discriminant functions for the three years that have the correct signs we find that nearly in every year we have a different set of factors with an overwhelming presence of solvency and leverage factors. For two years we have both IPF and IPS as risk factors at the expense of liquidity or profitability ratios. The fact that the set of risk factors changes from year to year creates problems if the objective of the researcher is to produce relationships that can be used to predict future membership.

Table 3.7.2: Estimates of Discriminant Functions-Stepwise Approach

	94-96	95-97	96-98	97-99	98-00	99-01
ROA		1.284	-1.519			
ROE		-0.206	1.651	1.637	-0.631	
IPS	10.208		6.794	11.792		-8.859
IPF					0.346	
IPE		-1.767		2.118	-0.457	
DR						-0.097
DC						
CR	-2.592	1.645	-2.374		0.286	2.041
WA						
UPL						11.653
RTA	0.737	5.958				
SIZE			0.324		0.582	
SOA		0.878				
NIS	-4.526					
OS		-0.672				
(Constant)	-3.440	0.661	-0.634	-2.245	-4.400	0.358
Wilks Λ	0.942	0.886	0.854	0.880	0.873	0.888
χ^2(df)	29.10	65.308	73.899	50.919	50.237	35.535
Df	4	7	5	3	5	4
BOX M	52.981	4705.53	225.66	134.421	514.855	69.940
F-Stattistic	5.178	161.590	14.544	21.876	32.624	6.300
DOF1	10	28	15	6	15	10
DOF2	57698	46813	36060	38190	16079	35562
η^2	0.240	0.337	0.382	0.346	0.356	0.335
η^2_{adj}	0.231	0.317	0.371	0.341	0.341	0.323
$\hat{\beta}'\hat{\mu}_0$	-0.0972	0.134	-0.151	-0.135	0.128	0.159
$\hat{\beta}'\hat{\mu}_1$	0.627	-0.956	1.126	1.000	-1.132	-0.787
$(\hat{\beta}'\hat{\mu}_1 + \hat{\beta}'\hat{\mu}_0)/2$	0.2649	-0.411	0.4875	0.4325	-0.502	-0.314
n_{11}/n_1	51.5	74.3	60.7	64.6	57.9	27.4
n_{00}/n_0	68.8	64.2	81.6	80.0	76.9	72.6
$(n_{11}+n_{00})/n$	66.5	73.1	79.1	78.2	74.9	56.6

Furthermore looking at the predictive power of the model using the leave-one-out (L-O-O) criterion we see that the performance is not good. The hit rate for the downgraded companies is low and in one particular year, 1999, only 27% of the companies that were eventually downgraded were predicted correctly. All in all, the stepwise procedure has failed to produce an economically meaningful and statistically credible model.

The results we have obtained are not entirely unexpected as many authors have highlighted the problems associated with the stepwise procedure. These problems are summarised aptly by Judd and McClelland (1989) who state that the main problems are:

1. Stepwise methods will not necessarily produce the best model if there are redundant variables.

2. All-possible-subset methods produce the best model for each possible number of terms, but larger models need not necessarily be subsets of smaller ones, causing serious conceptual problems about the underlying logic of the investigation.
3. Models identified by stepwise methods have an inflated risk of capitalising on chance features of the data. They frequently fail when applied to new datasets. They are rarely tested in this way.
4. Since the interpretation of coefficients in a model depends on the other terms included, "it seems unwise to let an automatic algorithm determine the questions we do and do not ask about our data".

They conclude their discussion stating: "It is our experience and strong belief that better models and a better understanding of one's data result from focused data analysis, guided by substantive theory." (p 204)

Henderson and Velleman (1981) also criticize the mechanical aspects of stepwise selection procedures by stating "The data analyst knows more than the computer and failure to use that knowledge produces inadequate data analysis."

In view of the unsatisfactory empirical results we have obtained and the problems associated with the stepwise procedure, we have adopted a second modelling approach that utilises the results of the univariate performance in a more structured way. This second approach we have adopted within the descriptive discriminant framework could be described as the best subset approach. That is, for every year we have selected the best variable from each category as it is measured by the univariate performance. This would be the best approach if the variables were completely uncorrelated.

More specifically, the strategy adopted is to use one or sometimes two variables from each category of risk factors (profitability, solvency, leverage, liquidity, activity, and growth) for inclusion in the discriminant function. The inclusion of variables was based exclusively on the statistical significance of the risk factor ignoring classification performance.

Under normality the two criteria will produce the same results but given the fact that not all our variables are normally distributed this will not be true.

Our approach consisted of the following steps.

(a) Include from each risk category the variable with the highest univariate performance. Given that we have six categories of risk factors we start with six factors. Size we treat separately, because although we have included it in the growth category is not highly correlated with the other "growth" variables. Using the 1994 sample as an example we have selected the following factors initially:
(b) Retain from each category the variable that remained significant after the estimation of the six-variable discriminant function.
(c) For each retained variable of a risk category, rotate all the other variables of the other risk categories to test the sensitivity of the estimated discriminant functions.

This procedure is not optimal in any sense since sequential testing of this nature is optimal only under restrictive assumptions. However when compared to the stepwise it makes sure that the variables in the equation are economically meaningful. Given also our requirement that this should be used to forecast membership of new companies, significance across the 6 years was also a criterion.

The best subset approach gives results which are superior to the stepwise approach. In every year the selected variables have the correct sign and thus they are meaningful as risk factors. The performance of the model as it is measured by η^2_{adj} is comparable to that of the stepwise model. The most important finding, however is that the cut-off points, which determine the classification regions, are reasonably stable over the years. Looking at the risk factors that have been selected, the return on assets (ROA) and the Debt –to-Equity ratio (DR) are present in every year. The liquidity ratio WA (Working capital over total assets) is also present in 5 of the 6 years. The solvency risk factors (IPS and IPF) each appear in three years. These two variables are interchangeable. IPF improves the statistical criterion whereas IPS the classification performance.

Table 3.7.3: Estimates of Discriminant Functions –Best subset approach

	94-96	95-97	96-98	97-99	98-00	99-01
ROA	-1.020	-2.563	-1.602	-0.506	-1.280	-6.222
WA	-1.237		-1.945	-1.788	-2.387	-2.080
IPS	5.015		8.092			10.810
IPF		0.341		2.350	2.554	
SIZE	0.868					
DR	2.450	2.511	1.542	1.618	0.774	2.004
RTA			-3.808		-1.036	
UPL		-0.358		-0.110		
NIS	-3.713					
(Constant)	-5.738	-1.116	-0.863	-1.48	-0.424	
Wilks Λ	0.934	0.941	0.874	0.888	0.900	0.892
χ^2(df)	34.713	32.858	63.007	47.376	39.149	32.219
Df	6	4	5	5	5	4
BOX M	141.2	502.3	202.1	282.9	121.9	47.1
F-Stattistic	6.497	49.11	13.023	18.125	7.723	4.573
DOF1	21	10	15	15	15	10
DOF2	47849	58786	36060	26349	16079	35562
η^2	0.257	0.243	0.354	0.335	0.317	0.329
η^2_{adj}	0.248	0.220	0.342	0.329	0.301	0.316
$\hat{\beta}'\hat{\mu}_0$	-0.104	-0.093	-0.138	-0.130	-0.112	-0.156
$\hat{\beta}'\hat{\mu}_1$	0.674	0.667	1.033	0.64	0.992	0.771
$(\hat{\beta}'\hat{\mu}_1 + \hat{\beta}'\hat{\mu}_0)/2$	0.285	0.287	0.4475	0.255	0.44	0.3075
n_{11}/n_1	56.1	65.7	67.9	66.7	63.2	60.8
n_{00}/n_0	67.6	66.6	78.9	77.5	76.6	68.
$(n_{11}+n_{00})/n$	66.1	66.5	77.6	76.2	75.2	67.0

Size is significant only in one year and with a positive sign implying that larger companies were more likely to become financially distressed over the following two years. Growth risk factors, such as accumulated profits or losses (UPL) and retained earnings were significant in three years and NIS was the only activity ratio to be significant in 1994.

The model has been subjected to a number of specification tests in order to test its robustness using the statistics of Section 3.5. In all cases we use the likelihood ratio statistic to test whether the excluded variables should be in the equation. The first test is to see whether any of the activity ratios should be in the equation. Only in 1994 we get a significant impact of the activity variables and we test whether any of the four variables should have been included in the discriminant equation. The results are shown in Table 3.7.4

Table 3.7.4 – LR tests for the inclusion of activity variables (values of chi square with 1 degree of freedom)

	1994	1995	1996	1997	1998	1999
SOA	0.78	1.23	1.45	1.56	2.01	1.78
NIS	3.92	0.99	2.70	2.90	3.3	2.64
OS	1.24	2.29	1.59	1.80	0.97	1.13
ES	1.9	0.78	0.98	1.35	0.45	1.46

The second test was to test whether any of the growth variables should be included in the equation in any of the years. The growth variables are included in four of the six samples we have used and the test is whether they should be included in he remaining two samples. The results are shown in Table 3.7.5. The final test concerns the impact of size on the classification of companies. Previous studies have shown that size could be an important factor. In our case though, it is only significant in one year.

Table 3.7.5 – LR tests for the inclusion of growth variables

	1994	1995	1996	1997	1998	1999
UPL	1.89	4.23	2.43	3.98	1.29	2.05
RTA	2.224	2.90	3.10	4.02	1.03	1.95

The lack of significance is not surprising, since in Chapter 2 it was found that the sample means of this characteristic were not statistically different in the two groups. The results of the test statistic are shown in Table 3.7.6 which formally confirm size is not a discriminating factor.

Table 3.7.6 – LR tests for the inclusion of size variables

	1994	1995	1996	1997	1998	1999
SIZE	3.85	2.74	1.56	2.01	1.07	1.91

3.7.2 Predictive Discriminant Analysis

The descriptive discriminant analysis has produced a set of factors that explain group membership in every year. We now turn to the estimation of predictive discriminant functions and to the variable selection problem using the classification performance as a criterion. Before we proceed we have to resolve a number of issues, the most of important of which is the classification criterion itself (See Dinenis (2003a) for a discussion of these issues). The criterion could be either the total proportion of correct forecasts or the proportion of correct forecasts in one of the two groups. The choice of the classification criterion is important because using the proportion of correct classifications of financially distressed companies n_{11}/n_1 instead of the proportion of total correct predictions $(n_{11}+n_{00})/n$ will produce different results. We give some examples of these later.

The second issue is that classification requires the use of a cut-off point. Such a cut-off point depends on the classification costs and the prior probabilities. In SPSS the allocation of observations to either of the two groups is based on the value of the calculated posterior probability which we defined earlier as

$$\pi_{jr} = \Pr(\xi = j / \mathbf{r} = \mathbf{r}^*) = \frac{p_j f_j(\mathbf{r}^*)}{p_0 f_0(\mathbf{r}^*) + p_1 f_1(\mathbf{r}^*)} = \frac{p_j f_j(\mathbf{r}^*)}{f_r(\mathbf{r}^*)}$$

The use of the posterior probability requires values for the two conditional density functions as well as for the two prior probabilities and classification costs. The effect of prior probabilities is to change the probability of classifying an observation into a particular group by making it harder to meet the condition for inclusion. The same effect has the incorporation of misclassification costs. High misclassification costs act as an additional hurdle to the classification of observation to a particular group. The specification of prior probabilities and classification costs is not a trivial problem as it requires some quantification of the loss that would be incurred if the wrong decision is taken. One way to estimate the prior probabilities is to use the relative sample sizes. This would have produced the following priors for each year

Table 3.7.7 Prior Probabilities

Year	Prior probability For Group 1	Prior Probability for Group 0
1994-96	0.134	0.866
1995-97	0.123	0.877
1996-98	0.118	0.882
1997-99	0.119	0.881
1998-00	0.101	0.899
1999-01	0.168	0.832

The issue of the classification costs is trickier. How much worse is a bank when it classifies a suspect company as healthy relative to classifying a healthy company as suspect? In this study we have used a ratio of 5:1 that is about 5 times more costly to misclassify a failing company relative to misclassifying a healthy one. The product is then about 0.5 and this is the prior we have used to classify the observations into the two groups.

Having discussed the two important issues of the classification criterion and the prior probabilities and costs, we can now explore the variable selection strategies available in predictive discriminant.

Our starting point therefore, as in the case of descriptive discriminant analysis is to perform a univariate classification analysis of all the variables. The results of the univariate classification are shown in Tables 3.7.8A and 3.7.8B (the tables have been split to accommodate all the data).

Table 3.7.8 A L-O-O hit rates

		94	95	96	97	98	99
ROA	n_{00}/n_0	34.3	50.5	51.3	44.1	42.8	55.3
	n_{11}/n_1	84.8	76.1	81	81.3	76.9	67.6
	$(n_{11}+n_{00})/n$	41.0	53.6	54.9	48.5	46.3	56.9
ROE	n_{00}/n_0	31.8	35.3	35.7	37.4	32.4	**36.5**
	n_{11}/n_1	83.2	82.1	81	72.9	74.4	**75.7**
	$(n_{11}+n_{00})/n$	38.7	41.0	41.2	41.6	36.7	**41.5**
DR	n_{00}/n_0	54.2	53.9	53.2	57.3	52.8	54.9
	n_{11}/n_1	53	62.7	62.1	66.7	66.7	59.5
	$(n_{11}+n_{00})/n$	54.0	55.0	54.3	58.4	54.2	55.5
DC	n_{00}/n_0	29.8	71.8	72.2	73.5	74.8	**76.6**
	n_{11}/n_1	60.6	34.3	32.1	35.4	36.8	**35.3**
	$(n_{11}+n_{00})/n$	33.9	67.2	67.5	69.0	70.9	**69.6**
IPS	n_{00}/n_0	73.1	72.9	84.9	78.1	73.7	72.9
	n_{11}/n_1	42.4	53.7	44.8	50	53.8	45.9
	$(n_{11}+n_{00})/n$	69.0	70.5	80.1	74.8	71.6	69.5
IPF	n_{00}/n_0	66.4	91.4	78.7	86.5	82.6	68.6
	n_{11}/n_1	43.9	19.4	53.4	47.9	46.2	45.9
	$(n_{11}+n_{00})/n$	63.4	82.6	75.6	81.9	78.8	65.7
IPE	n_{00}/n_0	75.4	71.0	47.1	76.6	62.0	91.3
	n_{11}/n_1	57.6	59.9	83.9	52.1	15.8	11.8
	$(n_{11}+n_{00})/n$	73.0	69.6	51.5	73.7	57.3	77.9
WA	n_{00}/n_0	55.6	60.8	66	59.6	60.5	61.6
	n_{11}/n_1	60.6	62.7	58.6	62.5	64.1	64.9
	$(n_{11}+n_{00})/n$	56.3	61.0	65.1	59.9	60.9	62.0
CR	n_{00}/n_0	22.4	42.1	30.7	42.1	35.1	36.9
	n_{11}/n_1	93.9	83.3	86.2	83.3	84.6	75.7
	$(n_{11}+n_{00})/n$	32.0	47.2	37.4	47.0	40.2	41.8

Table 3.7.8 B L-O-O hit rates

		94	95	96	97	98	99
SIZE	n_{00}/n_0	53.5	49.1	51.3	53.1	54	56.5
	n_{11}/n_1	45.4	53.7	58.6	52.1	53	48.6
	$(n_{11}+n_{00})/n$	52.4	49.7	52.2	53.0	53.9	55.5
UPL	n_{00}/n_0	33.2	34	28.1	32.3	27.1	72.2
	n_{11}/n_1	72.7	71.6	79.3	89.6	74.4	43.2
	$(n_{11}+n_{00})/n$	38.5	38.6	34.3	39.1	32.0	68.5
RTA	n_{00}/n_0	57.7	61.6	73.8	72.2	57.2	56.9
	n_{11}/n_1	66.7	64.2	46.6	50	64.1	75.7
	$(n_{11}+n_{00})/n$	58.9	61.9	70.5	69.6	57.9	59.3
NIS	n_{00}/n_0	54.7	98.7	90.5	87.6	66.4	71
	n_{11}/n_1	72.7	4.5	20.7	33.3	56.4	51.4
	$(n_{11}+n_{00})/n$	57.1	87.1	82.1	81.1	65.4	68.5
SOA	n_{00}/n_0	38.6	48.9	44	44.7	54.6	51.4
	n_{11}/n_1	65.2	68.7	69	43.8	74.4	73
	$(n_{11}+n_{00})/n$	42.2	51.3	47.0	44.6	56.6	54.1
OS	n_{00}/n_0	45.5	99.6	47.4	45.4	55.8	48.0
	n_{11}/n_1	59.1	6.0	57.1	60.4	52.6	58.8
	$(n_{11}+n_{00})/n$	47.7	88.1	48.5	47.1	55.5	49.8
ES	n_{00}/n_0	43.2	99.8	47.1	41.7	59.9	44.0
	n_{11}/n_1	59.1	6.0	53.6	64.6	47.4	62.7
	$(n_{11}+n_{00})/n$	45.3	88.3	47.9	44.4	58.7	47.2

The most striking feature of the results which are also shown graphically in Figures 3.7.1 and 3.7.2 is that some of the ratios have a very high hit rate for group 1 companies n_{11}/n_1 and some ratios have a very high hit rate for group 0 companies n_{00}/n_0 but not for both. The most successful variable predicting group 1 membership is the current Ratio (CR) with an average hit rate of 84.5 percent, followed by both profitability ratios ROA and ROE with an average hit rate of 78 percent.

The next three variables (UPL, SOA, and WA) represent growth activity and liquidity ratios. Leverage (DR and DC) and in particular solvency ratios (IPS and IPF) on the other hand are not very good at predicting financially distressed companies.
However, leverage (DR and DC) and particularly solvency ratios (IPS and IPF) are the best ratios to predict which of the companies will not become financially distressed. With the exception of NIS, it seems that profitability, liquidity and growth ratios are good predictors of the companies that will become financially distressed and leverage and solvency ratios of those that will not become financially distressed.

Figure 3.7.1: Ranking of Factors in terms of Group 0 average L-O-O hit rate (n_{00} / n_0).

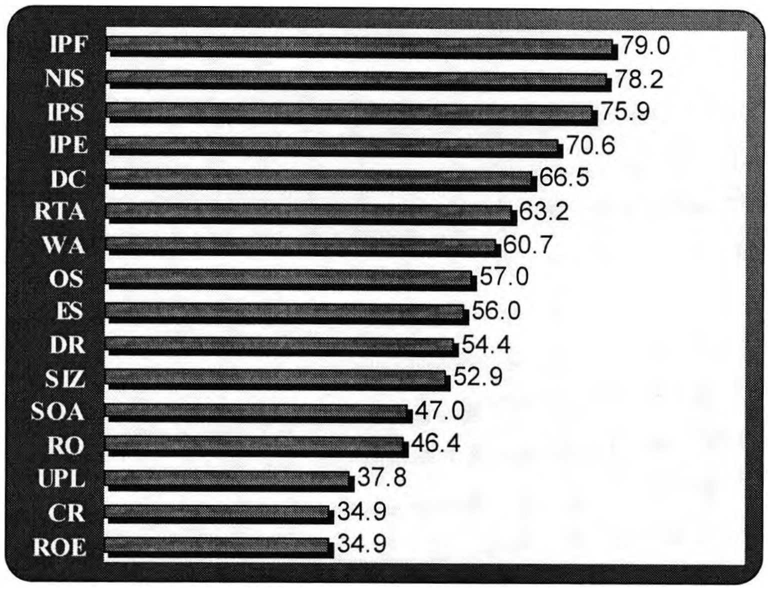
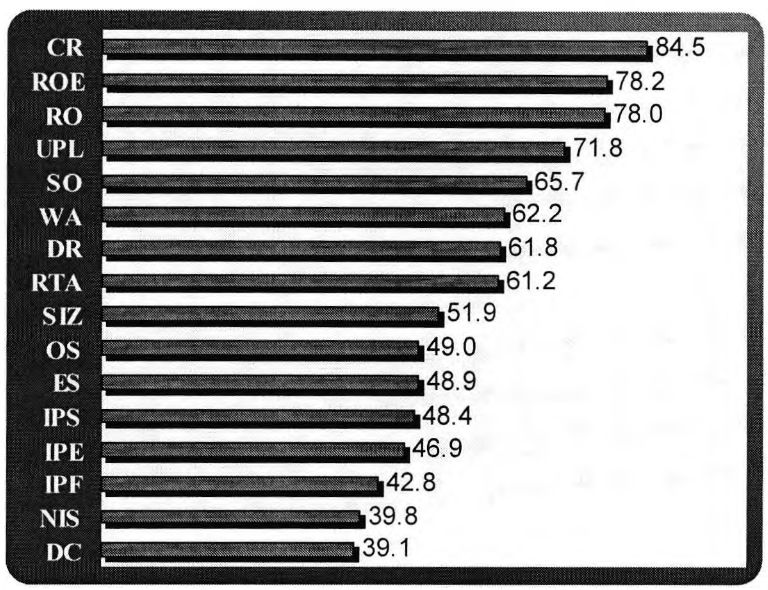


Figure 3.7.2: Ranking of factors in terms of group 1 average L-O-O hit rate n_{11} / n_1



There is no stepwise methodology based in classification criteria available in any of the major statistical packages. A popular forward stepwise procedure suggested by Smith (1984) involves the following steps:

1. Record p univariate PDAs and record the L-O-O hit rate of interest (total group or for a particular group) for each predictor; a best subset of size 1 consists of one variable that yields the highest of the hit rates.
2. Consider p-1 bivariate PDAs using $V1$ and $V2$, $V1$ and $V3, \ldots,$ $V1$ and Vp and record the L-O-O hit rate of interest for each pair; a best subset of size 2 (given that $V1$ is included) consists of the pair (containing $V1$) say $V1$ and $V2$ that yields the highest of the p-1 hit rates.

3. Conduct *p-2* trivariate PDAs using *V1* and *V2* and *V3, V1* and V2 and *V4,... V1* and *V2* and *Vp* and record the L-O-O hit rate for each triple; a best subset of six (given that *V1* and *V2* are included) consists of a triple (containing (*V1*and V2), say *V1* and *V2* and *V3* that yields the highest of the *p-2* hit rates.
4. Continue this through the subset of size *p-1*.

A total of *p(p+1)/2* analyses would need to be conducted, which in our case with p = 16 means 136 analyses. It is possible of course that the there may be more than one subsets at each stage which have the same hit rate. In that case one has to consider all the cases stemming from both sets.

The two main problems associated with a programme stepwise analysis also afflict the forward selection analysis we have adopted here, namely: (1) the best subset of a given size may not emerge, and (2) only one "good" subset of each size is suggested.

We have an additional problem in our case because of the nature of the univariate classification results. Suppose that we want to implement the forward selection procedure in a particular year and choose the variable with the highest group 1 hit rate. This will be R13 with a hit rate in excess of 80%. The inclusion of additional variables will never increase this hit rate and we are forced to conclude that the discriminant function will only contain one term. However such a discriminant function will only predict 35 percent of the of the healthy companies. The forward stepwise method cannot therefore be applied in the form it was suggested above.

A backward approach could also be implemented by starting with all the variables and eliminating those whose omission does not affect the classification performance. In this context the question still arises as to whether we should consider all 16 variables for inclusion in the discriminant function or a subset of those variables and how to arrive at such a subset. Such a process of factor reduction is useful because of the presence of multicollinearity in the data that makes the parameter estimates unreliable. One way to reduce the potential number of variables is to consider only those variables that are statistically significant on the basis of Wilk's lambda statistic. Since the testing of the equality of group means in Chapter 2 has shown that most of the financial ratios means are statistically different in the two groups for most of the years we have decided to consider all the variables as potential candidates in the discriminant function. A second way to reduce the number of factors for consideration is to look at the univariate performance and include the variables with the best classification performance from each category. However as in the case of the forward stepwise we have the problem of how to choose the "best" factors.

Our modelling approach within the predictive discriminant analysis framework is a pragmatic one. Rather than relying exclusively on classification performance, we have employed a combination of statistical and classification criteria. As Huberty (1994) suggests that the selection of variables should be based on both separation and classification performance if the objective of the study is to get a set of coefficients that could be used in subsequent samples.

In view of the problems with forward procedure we have instead adopted a backward procedure. Our starting point in this backward approach is then to start in every year with all the statistically significant variables and then rotate the various ratios and use classification criteria to select the best one. That means that we start more or less with the same set of variables as in the best subset DDA but the selection of variables and the finer specification analysis is now based on classification criteria. We are particularly interested in the correct prediction of companies that will be financially distressed, but since the factors that correctly predict those companies are not very good predictors of the healthy companies we have used a modified overall performance measure that gives a higher weight to the group 1 hit rate. The measure we have used is given by

$$0.5 \times \frac{n_{00}}{n_0} + 0.5 \times \frac{n_{11}}{n_1}$$

instead of the usual overall L-O-O hit rate given by

$$\frac{n_0}{n} \times \frac{n_{00}}{n_0} + \frac{n_1}{n} \times \frac{n_{11}}{n_1} = 0.88 \times \frac{n_{00}}{n_0} + 0.12 \times \frac{n_{11}}{n_1}$$

The advantages of this more flexible approach rests on the following. First, it is possible that there is a subset of variables that predict better than the whole. This is true if the additional variables do not contribute to the inter-group differences. This is an important difference between separation and classification. Secondly the positive bias in classification errors such as APER increases as p increases. Thirdly, where a set of variables could be deleted in order to enhance the predictive accuracy, this should be done so as to have a more robust classification tool. Fourthly a small number of variables would reduce the problem regarding the stability of the coefficients in the discriminant function. Finally a great source of instability will be the correlation between the independent variables. It may sensible then to exclude variables that have been highly correlated with other variables.

In summary we have established the following criteria for selection of variables in the discriminant function.

(a) The variables in the equation should meet both the separation and prediction criteria
(b) The variables should exhibit a low degree of collinearity
(c) The variables should be stable across all the years, that is the variables should have a permanent discriminating influence

The results of our modified backward selection procedure are shown in Table 3.7.9. The first major observation is that the set of factors that were selected on the basis of classification performance is very close to the best subset approach in DDA. Indeed in 1996, we arrived at the same model. In general the PDA approach produced for nearly all the years a more parsimonious model, which makes of course the model more robust. Now looking at the individual years we see that the 1994 model is the same as the DDA model but without the UPL variable. This is one of the cases were a variable should be in the equation according to a statistical criterion but its inclusion deteriorates the predictive performance of the model.

Having compared the two approaches and analyzed the forecasting performance of the model, we now turn to the analysis of the estimated equations. The signs of the coefficients are consistent with prior expectations. As it would be expected, increased profitability (ROA) and increased liquidity (WA) diminish the probability of financial distress. In contrast, the positive sign indicates that increased interest payments will increase the probability of financial distress. The table above (3.7.9) shows that the discriminant function coefficients are individually significant as indicated by the Wilks test in every single year and that the discriminant function was significant overall.

Table 3.7.9 Estimates of Predictive Discriminant Functions

	94-96	95-97	96-98	97-99	98-00	99-01
ROA	-2.018	-2.845	-1.602	-0.593	-1.301	-6.989
WA	-1.894		-1.945	-1.219	-2.469	
IPS	8.342	5.034	8.092	16.923		12.572
IPF					2.589	
DR	1.615	2.801	1.542	2.205	0.742	2.795
RTA			-3.808			
UPL				-0.074		
(Constant)	-0.930	-1.488	-0.863	-1.909	-0.410	-1.448
Wilks Λ	0.946	0.948	0.874	0.893	0.900	0.904
χ^2(df)	25.522	29.086	63.007	44.927	39.111	30.350
Df	4	3	5	5	4	3
BOX M	70.106	70.540	202.054	177.355	99.645	46.008
F-Stattistic	6.859	11.560	13.023	11.363	9.569	7.495
DOF1	10	6	15	15	10	6
DOF2	57697	75649	36060	26349	18107	46057
η^2	0.226	0.228	0.354	0.327	0.316	0.310
η^2_{adj}	0.217	0.204	0.342	0.321	0.300	0.297
$\hat{\beta}'\hat{\mu}_0$	-0.097	-0.087	-0.138	-0.127	-0.112	-0.146
$\hat{\beta}'\hat{\mu}_1$	0.587	0.626	1.033	0.937	0.990	0.724
$(\hat{\beta}'\hat{\mu}_1 + \hat{\beta}'\hat{\mu}_0)/2$	0.245	0.269	0.447	0.405	0.439	0.289
n_{11}/n_1	60.6	74.6	67.9	68.8	65.8	62.7
n_{00}/n_0	68.3	62.2	78.9	75.5	76.9	68.7
$(n_{11}+n_{00})/n$	67.3	63.7	77.6	74.7	75.7	67.7

Table 3.7.10 Wilk's test for individual significance

	94	95	96	97	98	99
ROA	6.185	13.743	8.409	3.630	5.781	7.976
IPS	11.912	6.084	35.095	30.221	3.921	15.128
IPF	2.307	9.462	33.228	33.910	26.571	5.405
DR	4.630	9.316	5.741	10.901	3.249	8.664
WA	11.549	18.509	23.876	11.913	12.691	13.262
UPL	3.409	3.846	8.356	3.125	1.090	3.452
RTA	11.446	25.047	15.380	16.680	4.576	17.385

The validity of the linear discriminant analysis is based on the equality of the covariance matrices in the two groups. We have already seen in Chapter 2 that we could not reject the hypothesis that the covariance matrices are different in the two groups. We repeat the same tests here. From the vales of the F –statistic in Table 3.7.9 we cannot again reject the hypothesis that the covariance matrices are different. The rejection of this hypothesis casts some doubts on the validity of our procedure, but in view of the stable results we have obtained we should not be unduly concerned about the failure of this property.

The amount of variation between the two groups explained by the discriminating variables is given by the adjusted for degrees of freedom canonical correlation coefficient η^2_{adj}. This is satisfactory in terms of the size that one expects on the strength of association from studies of this nature and it is consistent with the values obtained in other empirical studies.

The relative importance of a discriminator variable is assessed using the standardised coefficients which are defined as $\hat{\beta}^*_i = \frac{\hat{\beta}_i}{\hat{\sigma}_{ii}}$ where $\hat{\sigma}_{ii}$ is the (i,i) element of the pooled covariance matrix $\hat{\Omega}_p$.

The higher the standardized coefficient of a variable relative to the other variables, the more important its contribution is to the discrimination of the two groups. The standardized coefficients for the discriminant functions for each of the years in our sample are given in the table below.

Table 3.7.11: Standardized Canonical Function Coefficients

	94-96	95-97	96-98	97-99	98-00	99-01
ROA	-0.438	-0.685	-0.386	-0.235	-0.342	-0.536
WA	-0.352		-0.351	-0.217	-0.438	
IPS	0.606	0.430	0.640	0.757		0.648
IPCF					0.732	
DR	0.350	0.613	0.335	0.443	0.158	0.554
UPL				-0.054		
RTA			-0.205			

From the table it appears that the most important discriminating variable in all six samples is IPS and ROA, followed by DR and WA whereas the two variables RTA and UPL seem relatively less important.

The loading of a variable is calculated as $l_i = \sum_{j=1}^{p} \rho_{ij} \hat{\beta}^*_j$ where ρ_{ij} is the pooled correlation between variable i and variable j, measures the contribution of each variable to the formation of the discriminant function.

Table 3.7.12: Structure Matrix

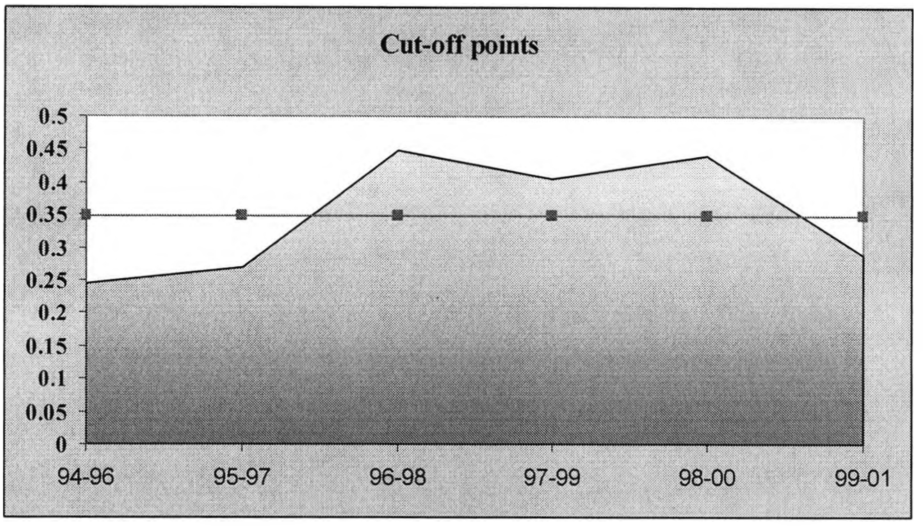
	94-96	95-97	96-98	97-99	98-00	99-01
IPS	0.673	0.451	720	0.795		0.686
WA	-0.663		-0.593	-0.499	-0.553	
ROA	-0.485	-0.677	-0.352	-0.275	-0.373	-0.498
RTA			-0.476			
IPF					0.800	
DR	0.420	0.558	0.291	0.477	0.280	0.520
UPL				-0.256		

The loadings of all seven variables used are high which means that the discriminant score can be interpreted as a measure of the financial health of a company and that all variables contribute highly towards the formation of the discriminant function.

The actual coefficients of the discriminant function used to calculate the cut-off points for classification are given in the appendix whereas the average values together with the cut-off point for each year are also reported in the table.

Apart from the sample for 1994 the cut-off points tend to be stable which is an important feature of our findings. The cut-off value indicates that companies with a value above that will be classified as migrating companies and those with a value less than that as non - migrating.

Figure 3.7.3 : Cut-off points



Turning now to the predictive power of the model we reproduce for ease of comparison the classification performance of the PDA and DDA in terms of all the four measures we have used.

Table 3.7.13 – Classification Performance

	94-96	95-97	96-98	97-99	98-00	99-01
DDA						
n_{11}/n_1	56.1	65.7	67.9	66.7	63.2	60.8
n_{00}/n_0	67.6	66.6	78.9	77.5	76.6	68.
$(n_{11}+n_{00})/n$	66.1	66.5	77.6	76.2	75.2	67.0
$0.5\times\frac{n_{00}}{n_0}+0.5\times\frac{n_{11}}{n_1}$	61.85	66.15	73.4	72.1	69.9	64.4
PDA						
n_{11}/n_1	60.6	74.6	67.9	68.8	65.8	62.7
n_{00}/n_0	68.3	62.2	78.9	75.5	76.9	68.7
$(n_{11}+n_{00})/n$	67.3	63.7	77.6	74.7	75.7	67.7
$0.5\times\frac{n_{00}}{n_0}+0.5\times\frac{n_{11}}{n_1}$	64.45	68.4	73.4	72.15	71.35	65.7

The predictive power of the model is satisfactory. The model consistently classifies more than the random 50 percent benchmark based on the proportional chance criterion for total and group hit rate where the proportion chance criterion for the total hit rate was defined as

$$PCC(T) = 0.5 \times \frac{n_0}{n} + 0.5 \times \frac{n_1}{n} = 0.5$$

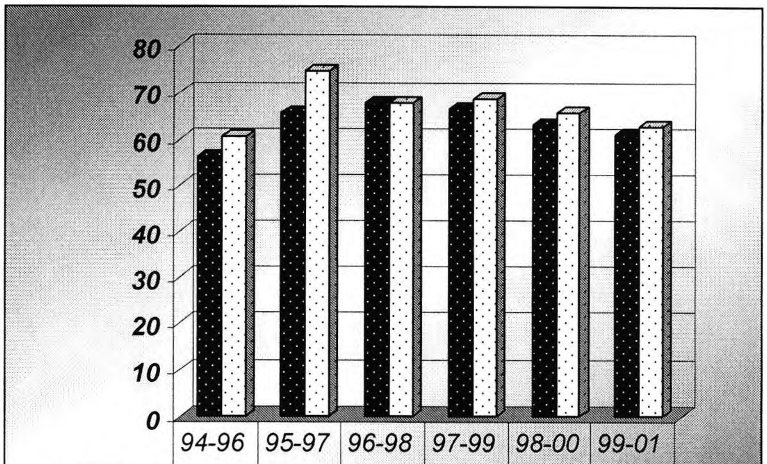
The predictive performance of the model as shown in table 3.7.14 depends on the criterion we have used. In the DDA approach the implicit emphasis was on the overall performance, whereas in the PDA approach we gave a higher weight to the correct classification to group 1, the downgraded companies. The performance according to the two criteria is given in Figures 3.7.4 and 3.7.5. Although the differences are not great for the total they are significant in terms of the hit rate for group 1. Under the DDA approach, the hit rate for group 1 is only 56 percent. In conclusion, the PDA is our preferred approach because it maximises the hit rate for group 1 without sacrificing too much the overall success rate.

We can test whether the classification performance of our model is significantly different from that of the proportional chance criterion. The difference in the two quantities can be tested using the standard normal z score whose values are shown in table 3.7.14. The performance is statistically different for every single year.

Table 3.7.14 : Values of z statistic

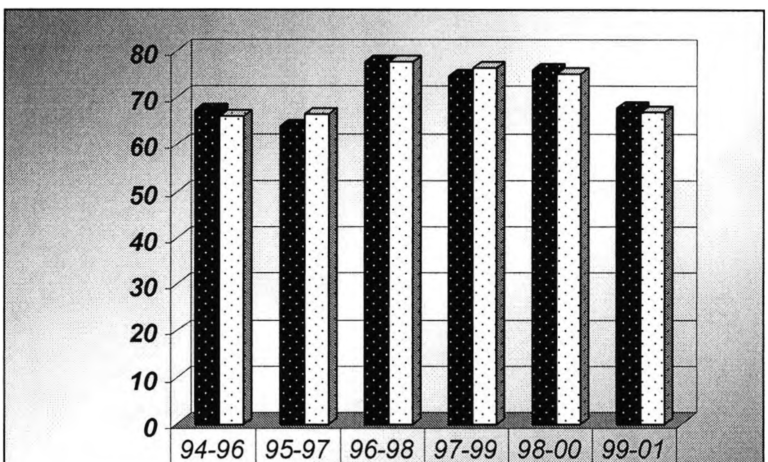
	94	95	96	97	98	99
DDA						
$(n_{11} + n_{00})/n$	7.14	7.71	12.02	10.52	9.76	5.92
$0.5 \times \frac{n_{00}}{n_0} + 0.5 \times \frac{n_{11}}{n_1}$	5.26	7.55	10.19	8.87	7.71	5.01
PDA						
$(n_{11} + n_{00})/n$	7.67	6.40	12.02	9.92	9.95	6.16
$0.5 \times \frac{n_{00}}{n_0} + 0.5 \times \frac{n_{11}}{n_1}$	6.41	8.60	10.19	8.89	8.27	5.47

Figure 3.7.4 Hit rates for group 1 by the two approaches



	94-96	95-97	96-98	97-99	98-00	99-01
■ DDA	56.1	65.7	67.9	66.7	63.2	60.8
□ PDA	60.6	74.6	67.9	68.8	65.8	62.7

Figure 3.7.5: Overall Hit rates by both approaches



	94-96	95-97	96-98	97-99	98-00	99-01
■ PDA	67.3	63.7	77.6	74.7	75.7	67.7
□ DDA	66.1	66.5	77.6	76.2	75.2	67

3.7.3 Conclusion from the empirical discriminant analysis

In this section we have pursued a systematic specification search for the set of risk factors that would explain and predict group membership. We have found that the classification criterion produced discriminant functions which were invariably nested within the DDA equation. The following set of risk factors were found to have a systematic effect on all years of our sample and constitute the set of risk factors we have derived of the basis of linear discriminant analysis.

1. Profitability ratios (ROA)
2. Leverage (DR)
3. Solvency ratios (IPF and IPS)

4. Liquidity ratios (WA)
5. Growth ratios (RTA and UPL)
6. Size (SIZE)
7. Activity ratios (NIS)

We have found that activity and growth ratios do not exert a systematic effect on the separation of companies. This does not exclude the possibility that these variables may be significant in particular years when some exogenous factors, e.g. the stage of the business cycle may exert some influence.

Compared to other studies which use discriminant analysis, our results are comparable in terms of the type of ratios we have found as significant (see Chapter 1 for a review of the relevant studies). An important feature of the discriminant functions we have estimated is the relative stability of the cut-off points that allow us to use the functions for classification of observations from new samples.

3.8 Results from Quadratic Discriminant Analysis

The results from the tests of hypothesis of equal covariance matrices both for the various ratios included in linear discriminant function indicated that the hypothesis of unequal variances cannot be rejected. In this situation the optimal classification rule should be based on a quadratic classification function rather than the linear one.

The classification according to this criterion was done using the estimated posterior probabilities from the equation

$$\hat{\pi}_{1\mathbf{r}} = \Pr(\xi = 1 / \mathbf{r} = \mathbf{r}^*) = \frac{e^{\hat{q}_1(\mathbf{r}^*)}}{e^{\hat{q}_1(\mathbf{r}^*)} + e^{\hat{q}_2(\mathbf{r}^*)}}$$

with

$$\hat{q}_i(r) = \ln[\hat{\pi}_i \hat{f}_i(\mathbf{r} / \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\Omega}}_i)] + \frac{p}{2}\ln(2\pi)$$

The same prior probability estimates were used, $\hat{\pi}_i = 0.5$ and the same estimates of the mean vectors $\hat{\boldsymbol{\mu}}_i$ and sample covariances $\hat{\boldsymbol{\Omega}}_i$ as in the linear discriminant analysis.

We have not pursued a separate specification search for the quadratic discriminant function, as this is computationally very demanding and no search properties have been established in the literature. Instead we have used the same variables that we employed in the linear discriminant function i.e. the variables (ROA, DR, IPF, IPS, WA, RTA, UPL, SIZE, and NIS). This allows direct comparison of the linear and the quadratic function approaches. The results are shown in the following table.

Table 3.8.1: Classificaion results from Quadratic Discriminant Function

	94-96	95-97	96-98	97-99	98-00	99-01
n_{11} / n_1	57.6	62.7	63.8	60.4	66.7	62.2
n_{00} / n_0	63.3	73.1	77.5	78.9	74.3	67.1
$(n_{11} + n_{00})/(n_1 + n_0)$	62.6	71.8	75.9	76.7	73.5	66.4

The classification performance is not markedly different from that of the linear discriminant analysis. Although the quadratic discriminant approach does not rely on the assumption of equal covariance matrices it still suffers from a number of problems discussed in Section 3.3. It still relies on the assumption of normally distributed ratios in the two samples if this is not a valid assumption, the results will of course affected. We have already discussed the performance of quadratic discriminant function in previous sections. Given the computational burden that the quadratic approach involves, the gains are not considerably higher.

3.9 Empirical Results from Logistic Regression

In this section we present the results from the estimation of the logistic regression model for the six time periods. We start the section with an explanation of the modelling strategies we have employed and in particular the variable selection strategy.

3.9.1 Modelling Strategies in Logistic Regression

As in the case of linear regression and discriminant analysis there is no standard approach to the variable selection problem in logistic regression. Hosmer and Lemeshow (1989) (see also Dinenis (2003b) for application of the procedure) suggest the following four steps in estimating a logistic model.

1. First, a univariate analysis of all the variables is undertaken. The univariate analysis is useful in determining if a variable should be included in the model . The significance of the variable can be based on the Wald statistic or on the likelihood ratio test defined in section 3.6. Another way to decide on the significance on the variable is to use the t statistic for the equality of means in the two samples. The discriminant function coefficient in the case of single normally distributed discriminating function is given by

   $$\frac{(\bar{r}_1 - \bar{r}_0)}{\hat{\sigma}_p^2} = \frac{t}{\hat{\sigma}_p}\sqrt{\frac{1}{n_1} + \frac{1}{n_0}}$$

   The use of univariate analysis has been criticised as ignoring the co-influence of other variables and the case where several variables may have individually a weak effect but taken together they may exercise a significant influence. Despite this criticism, univariate analysis does produce a lot of information on which a more complex analysis can be subsequently based. For instance it reduces the dimensionality of the problem considerably, because it reduces the number of variables for consideration in the multivariate analysis.
2. Secondly, a multivariate model should be estimated either by including all the variables an investigator considers as relevant, or using the smaller set of variables based on the results of the univariate analysis. If the latter approach is adopted, we consider for inclusion the variables from the univariate analysis that have met some criterion. The criterion is normally the statistical significance of the regression coefficient at some confidence level, normally at 95 percent. Once the set of variables on which the multivariate analysis will be based has been decided, a stepwise approach is normally employed. In SPPS the logistic regression programme there contains two variants the forward stepwise the backward stepwise approach.
3. Thirdly, following the fit of the multivariate model, the importance of each variable in the model should be verified using the statistical tests explained in section 3.6.
4. Fourthly, once a model has been estimated that contains the relevant variables, a closer inspection of the variables should take place and consider whether the model is correctly

specified, whether is a need to include interaction among the variables, whether there are influential data points that exert an undue influence on the model parameters.

3.9.2 Univariate Approach

The full results from the univariate analysis are presented in Tables B1-B6 in Appendix B. For each variable we present the estimated regression coefficients as well as various goodness-of-fit statistics such as the LR test, the R-squared, the Hosmer and Lemeshow test and classification performance tests.
The overwhelming majority of the coefficients are significant in all years. The most significant variables are the profitability leverage and solvency ratios with liquidity and growth ratios being inconsistent in their significance across the sample periods.

3.9.3 Stepwise Approaches

In the forward stepwise method we start with a model that initially contains only the constant term. At each step the variable with the smallest significance level for the score statistic, provided it is less than the chosen cut-off value (by default 0.05), is entered into the model. All variables in the forward stepwise that have been entered are then examined to see if they meet removal criteria. Variables are added according to some criterion. One such criterion is the value of the likelihood function and the likelihood ratio test. An alternative test is to use the value of the Wald Statistic.
In the backward stepwise approach one starts with all the variables in the model and drops the variables whose elimination does not reduce the value of the likelihood function.
The stepwise analysis is widely used but as it has been pointed out in the relevant discussion in discriminant analysis, is full of problems as this a mechanical procedure that may produce a model that may not make sense. It is possible for instance to include significant variables that have the wrong sign.
Having completed the univariate analysis we now move to the multivariate modelling stage. For the selection of the best set of variables for each period we use the two approaches described above, that is the forward and the backward stepwise methods. One of the uses of the univariate approach is the reduction in the number of variables, the investigator has to consider for consideration. Given the results of the univariate analysis one could justifiably exclude variables like OS which are not significant. However given the small number of variables under consideration we have decided to use all the variables in the forward and backward stepwise variable selection strategy. The results from the forward selection are shown in the Table 3.9.2. whereas the results from the backward selection procedure are shown in Table 3.9.3.
Starting with the Forward Method, we look first at the economic interpretation of the estimated equations and then we make an appraisal of their statistical performance. All the coefficients have the right sign except the coefficients for the variable SOA in 1996 and 1997. All the variables are individually significant (even SOA in the two years in which it had the wrong sign), with the values of the Wald statistics significant at 95 percent confidence level.

Table 3.9.2: Empirical Result of Logistic Regression- Forward Stepwise Method

	94	95	96	97	98	99
CONSTANT	-1.878 (78.023)	-2.725 (28.65)	-3.550 (31.88)	-5.055 (35.12)	-2.130 (19.754)	-1.535 (44.80)
IPS	2.863 (3.53)		7.549 (13.18)	18.978 (22.56)		
WA	-1.715 (4.805)		-2.353 (5.20)		-2.140 (3.405)	-2.211 (6.00)
RTA	-4.027 (4.14)					-14.719 (12.70)
ROA		-8.919 (20.67)	-12.385 (21.76)	-14.937 (23.15)	-3.459 (3.51)	
DR		2.795 (15.14)	2.335 (6.31)	4.305 (13.23)	3.920 (11.563)	
SOA			0.463 (5.34)	0.610 (4.65)	-1.892 (12.278)	
NIS		-1.225 (1.093)				
DC						0.068 (4.556)
IPF			1.563 (7.44)			
$2 \ln L_0$	-387.89	-406.54	-344.323	-294.30	-246.00	-274.644
$2 \ln L_p$	-367.07	-342.18	-254.621	-220.82	-203.97	-245.078
$LR = 2[\ln L_p - \ln L_0]$	20.82	64.36	89.702	73.48	42.03	29.565
R^2_{CS}	0.041	0.11	0.172	0.167	0.106	0.093
R^2_{CS}	0.076	0.21	0.334	0.322	0.220	0.156
$HL(\chi^2_8)$	11.060	11.88	3.415	3.910	10.388	7.467
Significance level	0.198	0.16	0.906	0.865	0.239	0.487
n_{00} / n_0	52.3	66.4	76.3	73.0	75.1	46.4
n_{11} / n_1	63.6	80.6	76.8	81.3	71.1	76.5
$(n_{00} + n_{11}) / n$	53.9	68.1	76.4	73.9	74.7	51.5

Table 3.9.3: Empirical Result of Logistic Regression- Backward Stepwise Method

	94	95	96	97	98	99
Constant	-5.911 (13.973)	-1.681 (5.58)	-3.104 (12.39)	-4.271 (22.02)	-5.236 (5.33)	-2.417 (5.634)
SIZE	0.569 (4.81)				0.589 (2.73)	
ROA	-4.255 (8.23)	-7.647 (15.61)	-13.535 (23.37)	-15.937 (20.41)	-3.423 (2.92)	-6.237 (4.32)
IPS	3.952 (6.86)		9.532 (15.22)	28.700 (19.34)		
DR	2.392 (10.12)	2.549 (9.35)	2.661 (6.85)	3.393 (7.28)	3.523 (6.64)	2.681 (6.15)
IPE		0.332 (3.54)	0.254 (4.86)			
CR		-0.526 (2.637)	-0.846 (3.95)			-0.492 (2.00)
SOA		-0.391 (2.71)	0.492 (5.91)	0.593 (4.07)	-2.019 (13.82)	
ES		-1.002 (2.83)		-5.700 (3.88)		4.169 (5.09)
NIS				4.367 (1.93)		
DC					0.072 (3.89)	
RTA						-12.249 (7.30)
WA					-2.241 (3.92)	
IPF			1.405 (6.20)			
$2\ln L_0$	-387.89	-406.54	-344.323	-294.30	-246.00	-274.644
$2\ln L_p$	-355.12	-329.24	-245.270	-215.77	-199.17	-238.381
$LR = 2[\ln L_p - \ln L_0]$	32.77	77.296	99.052	78.532	46.833	36.263
R^2_{CS}	0.064	0.132	0.189	0.177	0.117	0.113
R^2_{CS}	0.118	0.251	0.365	0.342	0.244	0.189
$HL(\chi^2_8)$	20.32	15.827	8.697	4.192	12.14	8.937
Significance level	0.009	0.045	0.368	0.839	0.145	0.348
n_{00}/n_0	53.5	68.5	75.8	72.1	74.8	46.8
n_{11}/n_1	72.7	77.6	82.1	79.2	71.1	80.4
$(n_{00}+n_{11})/n$	56.1	69.6	76.6	73.0	74.4	52.5

The statistical performance of the model as a whole can be evaluated by testing the hypothesis $H_0 : \beta_j = 0 \quad j = 1,2,.....,p$ using the likelihood ratio test. The likelihood ratio test indicates that collectively the parameter estimates are different from zero and accept the existence of a model. The Hosmer and Lemeshow Test is significant for every single year. The pseudo R-squared coefficients are small, but we have explained they cannot be interpreted in the same spirit as the corresponding quantities in the linear regression framework.

Looking now at the results of the backward stepwise method, we see that all the parameter estimates are significant and have the right sign except three variables. The first one is SOA which has the wrong sign for two years as in the case of the forward selection. The other two variables are NIS and ES both activity ratios, with the first one having the wrong sign in 1997 and the second one having the wrong sign in 1999. Note that ES has the correct sign in the other years in which it enters equations.

The likelihood ratio test rejects the hypothesis $H_0 : \beta_j = 0 \quad j = 1,2,.....,p$ and we accept the existence of a model, that is a valid relationship between the dependent variable and the set of independent variables.

The Hosmer and Lemeshow Test is significant for every single year whereas the pseudo R-squared coefficients are higher in every year than the corresponding values for the forward method.

Since parsimony is a desirable property of an econometric model, we see that the main difference in the estimated models using the two approaches is that the model produced using the backward stepwise method always contains more variables in each equation. A formal comparison of the two approaches could be performed using the pseudo R-squared adjusted for degrees of freedom or using some information criteria like the

Akaike information criteria and its variants. The Akaike Information Criterion, (AIC) defined as

$AIC = 2\ln L_p - 2p$.

Where L_p is the value of the likelihood function and p is the number of variables in the model.

Alternatively, we can use the Schwartz criterion, which is a modified version of AIC and is defined as $BIC = 2\ln L_p - p\ln(n)$. Both criteria adjust their values to penalize for the extra variables required to produce a given level of performance with the Schwartz criterion being much stricter. Table 3.9.4 shows the results of the comparison of the two models.

Table 3.9.4 : Information Criteria for Forward and Backward Methods.

	1994	1995	1996	1997	1998	1999
Forward						
Akaike	-370.07	-345.18	-260.621	-224.82	-207.97	-248.078
Schwartz	-385.665	-361.088	-291.588	-244.816	-227.678	-262.219
Backward						
Akaike	-359.12	-335.24	-252.27	-221.77	-205.17	-243.381
Schwartz	-379.914	-367.056	-288.398	-251.764	-234.732	-266.95

The two criteria produce conflicting results. In 1997 for example the Akaike criterion would claim that the backward method is best whereas the Schwartz criterion would say that the forward method is best. This is to be expected because the penalty of additional variables is larger in the latter model.

The comparison of the models should of course be based not just on statistical criteria but on classification performance as well. Although both the discriminant analysis and logistic regression procedures in SPSS estimate the posterior probability, they use a different procedure. In discriminant analysis we estimate the conditional probabilities first and then modify them with the introduction of classification costs and prior probabilities. In logistic regression, though, the posterior probability is estimated directly, and classification takes place on the basis of a predetermined cut-off point.

We can either adjust the constant term in the equation to incorporate costs forcing the constant term to behave as if the product of prior probabilities and classification costs are the same in both groups and then assume a cut-off point of 50 percent or we can leave use the estimated equation and assume a cut-off point equal to the sample size. We have adopted the latter since it is computationally more efficient.

Looking at the statistical results, we see that the probability of migration depends on profitability, interest payments and leverage. Size is occasionally important and as in the case of discriminant analysis, larger companies in the sample have a larger probability of default. The two variables that have been present in the estimated equation for all periods are the profitability and leverage ratios. Looking at the final criterion, the percentage of correct forecasts, the overall percentage of correct forecasts is satisfactory.

3.9.4 Stable Subset Approach

The automated approaches of the two stepwise variants have produced models that although statistically adequate for most of the years are different in term of the variables they use in the equations. In this section we shall try to synthesise the two approaches and come up with a model that is economically rational and meets the statistical and classification performance criteria. One way to do that is to follow one of the non-nested models hypothesis testing procedure. In this section however we have opted for a more pragmatic approach by using a multivariate modelling procedure which can be described as the "stable subset" variable selection procedure. With this procedure a number of models containing one, two, three, and so on, variables are examined which are considered the "best" according to some specified criteria. One criterion to determine the "best subset" model that would be appropriate in our case would be the statistical significance of the variables in every sub-period we have examined.

Table 3.9.5: Summary results from stepwise regression

	Forward	Backward	Number of Wrong signs
ROA	3	6	
ROE	0	0	
IPS	3	3	
IPE	0	2	
IPF	1	1	
DR	3	6	
DC	1	1	
WA	4	1	
CR	0	3	
RTA	2	1	
SIZE	0	2	
UPL	0	0	
SOA	3	4	4
ES	0	3	1
NIS	1	1	1
OS	0	0	

However, given the multicollinearity between the financial ratios, we have followed the same strategy as in the case of the MDA. That is we selected one or more financial ratios from each category and estimated the resulting logistic equation. We selected the model that produced the highest value for the likelihood function. The selection of the initial variables was best on both the univariate logistic regression analysis results as well as on the results of the backward and the forward selection. We would expect variables like DR and ROA which were present in most of the equations to also be significant in the "best subset" approach. Table 3.9.5 summarizes the results from the two approaches which helps us identify the most significant variables.

On the basis of the results from the forward and backward stepwise procedure were summarized in the above table we have selected the following ratios for inclusion in our "best subset" logistic regression equations.

Profitability Ratios	ROA
Solvency Ratios	IPS
Leverage	DR
Liquidity	CR and WA
Growth	RTA and SIZE

The results from the estimation of the logistic regression equations are shown in Table 3.9.6 . Starting again with the economic interpretation of the estimated coefficient parameters we see that this time all the parameters have the expected sign, with profitability (ROA) and liquidity (WA) reducing the probability of being downgraded and interest payments (IPS) and debt (DR) increasing the probability of being downgraded.

Table 3.9.6: Empirical Result of Logistic Regression- Stable Subset Method

	94	95	96	97	98	99
ROA	-3.671 (5.89)	-9.069 (24.18)	-11.527 (22.99)	-13.644 (20.90)	-5.277 (6.87)	-6.749 (7.25)
DR	2.090 (7.15)	2.282 (9.91)	3.146 (13.53)	5.006 (19.98)	1.570 (2.705)	2.106 (5.21)
IPS	3.472 (4.95)	1.811 (2.25)	7.714 (16.94)	15.198 (20.38)	2.610 (3.44)	7.561 (7.45)
WA	-1.010 (1.30)	-1.073 (1.945)	-2.188 (4.76)		-2.845 (6.37)	-1.918 (3.82)
SIZE	0.565 (4.77)					
Constant	-5.640 (12.62)	-4.153 (19.07)	-3.170 (29.72)	-4.696 (33.71)	-2.335 (12.50)	-2.667 (7.29)
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-353.8	-346.9	-267.4	-225.2	-218.2	-241.4
$LR = 2[\ln L_p - \ln L_0]$	34.1	59.7	76.9	69.1	27.8	33.3
R^2_{CS}	0.07	0.11	0.15	0.16	0.07	0.11
R^2_N	0.12	0.20	0.29	0.30	0.15	0.17
$HL(\chi^2_8)$	12.9	3.7	5.1	10.2	21.0	7.7
Significance level	0.11	0.88	0.75	0.25	0.01	0.46
n_{00}/n_0	54.9	67.0	72.7	71.0	69.4	49.2
n_{11}/n_1	69.7	76.0	78.6	75.0	68.4	74.5
$(n_{00}+n_{11})/n$	56.9	68.1	73.4	71.5	69.3	53.5

Three of the variables (ROA, DR and IPS) are significant in every year, whereas WA is not significant in 1997. Size again is significant in just one year as before. No profitability variables were found significant and no growth variables could be included in the equation despite exhaustive testing. In the case of the interest payments variables we tested the performance of the model using IPF instead of IPS but in all case the model performed better using the latter variable.

The statistical performance of the model as measured by the likelihood ratio test is good and we are able to reject the hypothesis $H_0 : \beta_j = 0 \quad j = 1,2,....., p$ and accept the existence of a model. The Hosmer and Lemeshow Test is significant for every single year. The pseudo R-squared coefficients are small, but we have explained they cannot be interpreted in the same spirit as the corresponding quantities in the linear regression framework.

Looking at the final criterion, the cut off point on which classification took place was 0.12 which corresponds to the relative sample sizes. The classification results as we have already discussed are very sensitive to the cut off point and this is something that we should have in mind when we assesses the results. Now looking at the classification performance, the overall percentage of correct forecasts is very high for the group 1 companies, that is companies that were going to be downgraded. The lowest hit rate was 68.4 percent in 1998 and the highest at 78.6 percent in 1996. However the high hit rate for group one companies is reflected in the low hit rate for group zero companies, that is companies that were not downgraded. In two particular years, in 1994 and in 1999 the hit rate is as low as 54.9 percent and 49.2 percent respectively.

The results are comparable to other studies using the same prediction period. Latinen and Kankaapaa (1999), for instance, using Finnish data, obtained a classification accuracy of 71.1 % two years prior to failure which is lower than the ones we achieved here. If the logistic model has homoscedasticity (not a logistic regression assumption), the percent correct will be approximately the same for both groups of companies. Here it is not, with the model having much more difficulty predicting companies that were not downgraded. While the overall percent correctly predicted seems good, consider that blindly estimating the most frequent category (non-downgraded companies) for all cases would yield a percent correct of 88% so the model may appear that it does not contribute to the overall performance. However as we have explained in the case of the MDA this is not a valid benchmark against which to compare the performance of the model since the benchmark would assume that no companies would be downgraded . This is of course a costly assumption to make. The difficulty in predicting the first and the last year of the sample mirrors the results of the discriminant analysis. The last year of the sample was contaminated by the phenomenal rise in the stick market valuation of many companies, which changed the behaviour of companies completely. The model does not contain any market values and is therefore unable to capture this effect.

3.9.5 Diagnostics

In the last section we concentrated on the log likelihood chi-square and pseudo R-square measures for the model. The log likelihood chi-square is an omnibus test if our model as a whole is significant. The pseudo R-square measures the portion of variation explained by the model. These are the measures a researcher normally looks at first a model has been estimated. We have also looked at another popular measure of goodness of fit the Hosmer and Lemeshow's goodness of fit test. The idea behind the Hosmer and Lemeshow's goodness of fit test is that the predicted frequency and observed frequency should match closely, the more they match, the better the fit. The Hosmer-Lemeshow goodness of fit statistic is computed as the Pearson chi-square from the contingency table of observed frequencies and expected frequencies.

However, when we build a logistic regression model, we assume that the logit of the outcome variable is a linear combination of the independent variables. This involves two aspects as we are dealing with the two sides of our logistic regression equation. First, the link function of the outcome variable on the left hand side of the equation, that is the logit function (in logistic regression) is the right function to use. Secondly, on the right hand side of the equation, we have included all the relevant variables and have not included any variables that should not be in the model and the logit function is a linear combination of them.

It is possible that the logit function as the link function is not the right choice or the relationship between the logit of outcome variable and the independent variables is not linear. In either case, we have a specification error. The misspecification of the link function is usually not too severe compared with using other alternative link function choices such as probit (based on the normal distribution). In practice, the researcher is more concerned with whether the model has all the relevant predictors and if the linear combination of them is sufficient.

So, in addition to the statistical performance of the model using the conventional tests, in order to have confidence in the estimated model we have performed a number of diagnostic tests to determine how robust our estimates are to data. The theoretical foundation for the use of diagnostics in logistic regression is due to Pregibon (1981) who adapted the diagnostic test for linear models to the case where the independent variable is a binary one. In ordinary least regression, we have several types of residuals and influence measures that help us understand how each observation behaves in the model, such as if the observation is too far away from the rest of the observations, or if the observation has too much leverage on the regression line. Similar techniques have been developed in logistic regression.

The diagnostic tests are designed to test whether the regression assumptions are valid, to check for adequate fit, to check for outliers, and to identify influential points. All of the diagnostic tests are based on the estimated probabilities, $\hat{\pi}_i$, the residuals $\hat{\varepsilon}_i$ and the leverage of observations.

Logistic regression residuals are defined as the difference between the classification variable and the estimated probability

$$\hat{\varepsilon}_i = \xi_i - \hat{\pi}_i$$

and the Pearson residual (standardised residual in SPSS) which are defined as

$$\zeta_i = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}} = \frac{\xi_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1-\hat{\pi}_i)}}$$

For large samples, the distribution of the Pearson residual is approximately normal with a mean of zero and variance of 1. Both Pearson and Deviance residuals are useful in identifying observations that are not explained well by the model. A large number of such cases is strong evidence of misspecification of the model. A summary statistic, the Pearson Chi-square statistic defined as

$$X2 = \sum_{i=1}^{n} \zeta_i^2$$

can be used to provide an overall statistic for all the cases. This statistic follows the chi-square distribution with n-p-1 degrees of freedom and the assumption that the fitted model is correct in all aspects. Small values of the statistic indicate a good fit, whereas large values indicate a poor fit. Plotting the standardised residuals against the fitted probabilities allows us to test whether the variance of the residuals is constant.

Pearson residual and its standardized version is one type of residuals. Pearson residual is defined to be the standardized difference between the observed frequency and the predicted frequency. It measures the relative deviations between the observed and fitted values. Deviance residual is another type of residual. It measures the disagreement between the maxima of the observed and the fitted log likelihood functions. Since the logistic regression uses the maximal likelihood principle, the goal in logistic regression is to minimize the sum of deviance residuals. Therefore, this residual is parallel to the raw residual in ordinary least square regression, where the goal is to minimize the sum of squared residuals.

Another statistic, sometimes called hat diagonal since technically it is the diagonal of the hat matrix, measures the leverage of an observation. It is sometimes called Pregibon leverage. The leverage of an observation can be used to check for outliers, that is unusual observations. The leverage of an observation is the diagonal element of the "hat" matrix in direct analogy with linear regression. The "hat" matrix in logistic regression is given by [Pregibon (1981)]

$$\mathbf{L} = \mathbf{V}^{1/2}\mathbf{R}(\mathbf{R'VR})^{-1}\mathbf{R'V}^{1/2}$$

where $\mathbf{R}$ denotes the $n \times (p+1)$ matrix containing the values of all observations for the m independent factors and the constant term and $\mathbf{V}$ is a $n \times n$ diagonal matrix with general element $v_i = \hat{\pi}_i(1-\hat{\pi}_i)$.

The leverage for the ith observation is defined as

$$l_i = \hat{\pi}_i(1-\hat{\pi}_i)(1,\mathbf{r}_i')(\mathbf{R'VR})^{-1}(1,\mathbf{r}_i')'$$

The leverage is used to detect observations that have a large impact on the predicted values. Unlike linear regression, the leverage values in logistic regression depend on both the dependent variable scores and the design matrix. Leverage values are bounded by 0 and 1. Their average value is p/n where p is the number of estimated parameters in the model including the constant and n is the sample size. For cases with predicted probabilities less than 0.1 or greater than 0.9, the leverage values may be small even when the cases are influential. Large values of the leverage are considered to be values that are more than 2 or 3 times its average value.

These three statistics, Pearson residual, deviance residual and Pregibon leverage are considered to be the three basic building blocks for regression diagnostics and it useful to inspect them first. A good way of looking at them is to graph them against either the predicted probabilities or simply case numbers. Figure 3.9.1 shows the Pearson residuals plotted against predicted probabilities for all the years. The range of the band remains relatively constant indicating that the variance is not unstable. However for most years the residual bands have a negative slope which means that the variables and the logit may be related in a nonlinear fashion. Also there a number of observation which at low probability levels are misclassified. Figure 3.9.2 shows the leverage of the equation for every single year. The leverage varies between 0 an1 with the larger values indicating the presence of outliers. In each year there are a handful of outliers which are manifested in values of the leverage in excess of 0.15. The largest outlier with a value of 0.9 is found in 1998.

We have seen so far how to detect potential problems with model building. We will focus now on detecting potential observations that have significant impact on our model. There are several reasons why we need to detect influential observations. First, there are possibly data entry errors. Secondly influential observations may be of interest by themselves for us to study. After all, influential data points may badly skew our regression estimation. The last diagnostic test is therefore designed to identify influential observations. Influential observations are those which have a large impact on the value of the regression coefficients and the statistics use to test the overall significance of the model. An influential point is one whose removal will have a substantial change on coefficient estimates. Influential points pull the regression line in their direction and distort the fit. Influential points do not necessarily produce large residuals, i.e. they do not necessarily produce large outliers. Because they change the slope of the regression they may avoid being outliers. The effect deleting a particular observation has on the value of the estimated coefficients and the overall summary measures of fit such as Pearson's chi-square statistic and deviance is also very important.

For example, the change in the second coefficient in a logistic regression equation when case j is deleted is given by

$$\Delta f(\hat{\beta}_{2(-j)}) = \hat{\beta}_2 - \hat{\beta}_{2(-j)}$$

where $\hat{\beta}_2$ is the value of the coefficient when all cases are included and $\hat{\beta}_{2(-j)}$ is the value of the coefficient when the jth case is excluded. Large values for statistic identify observations that should be examined. The $\Delta f(\hat{\beta}_{2(-j)})$ values calculated by the logistic regression program are an approximation to the true value but they provide a good indication of the sensitivity of the estimated coefficients to particular observations.

An overall measure of the impact of influential observations on the equation has been provided by Pregibon (1981) and is analogous to the measure proposed by Cook (1977, 1979) for linear regression. Cook's distance for the logistic equation is given by

$$C_i = \frac{\zeta_i^2 h_i}{1 - h_i}$$

Using similar approximations it can be shown that the decrease in the value of the Pearson chi-square statistic due to deletion of observation j is given by

$$\Delta X2 = X - X_{(-j)} = \frac{\zeta_i^2}{1 - h_i}$$

A similar quantity may be obtained for the change in the deviance

$$\Delta DV = DV - DV_{(-j)} = \frac{d_i^2}{1 - h_i}$$

The squared root of this quantity is called studentized residuals in SPSS. The change in the deviance and the Pearson chi-square statistic – ΔDV and $\Delta X2$ are diagnostics for detecting ill-fitted observations; in other words, observations that contribute heavily to the disagreement between the data and the predicted values of the fitted model.

These diagnostic statistics are very useful as they allow us to identify those covariate patterns that are poorly fit (large values of the statistic) and those that have a great deal of influence on the values of the estimated parameters (large values of the statistic).

Looking at the value of the Cook statistic for instance, we see again that we have a small number of influential variables that correspond to the outliers in the data. These outliers were in the group of downgraded companies and their removal would have eliminated valuable information about their behaviour.

The studentized residuals on the other hand show that despite the existence of some influential observations, the measures of overall performance are not affected.

Figure 3.9.1 Pearson Residuals

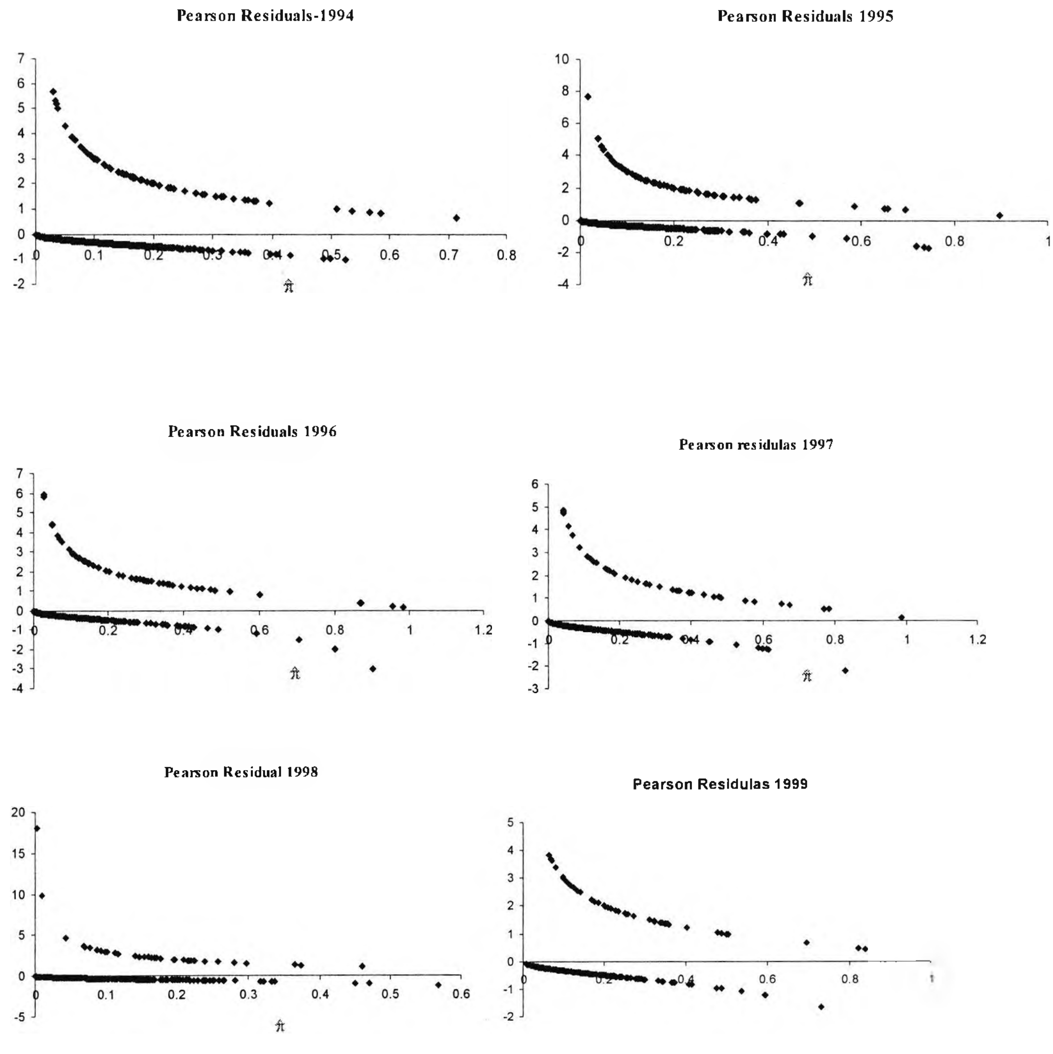


Figure 3.9.2 – Leverage Values

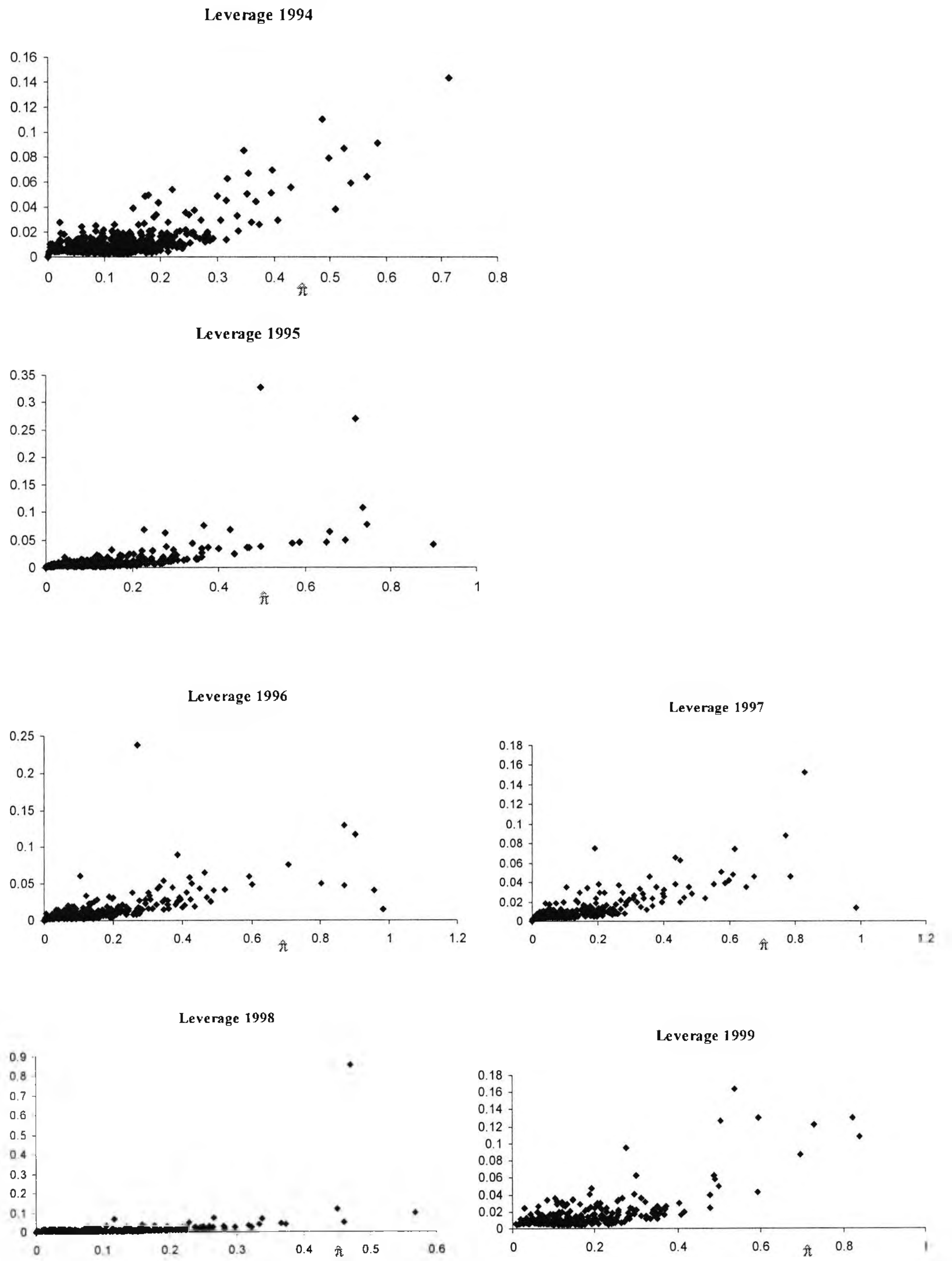


Figure 3.9.3 Cook's statistic

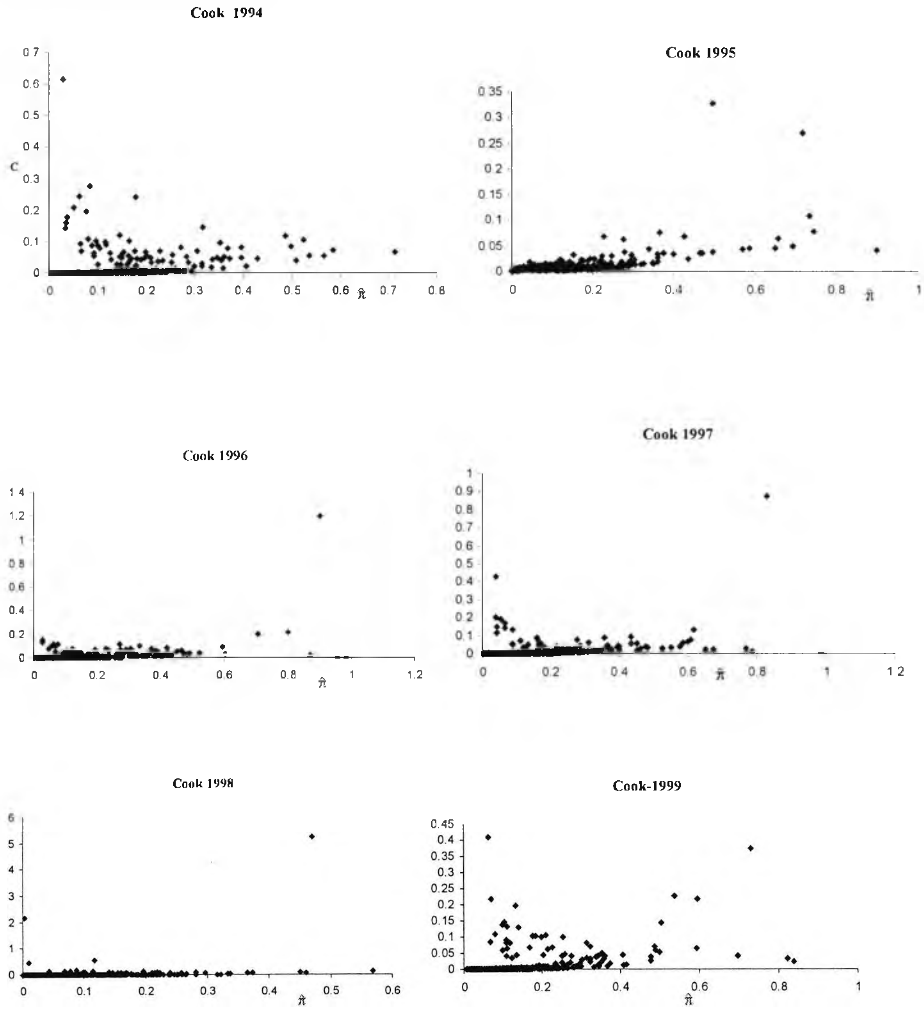
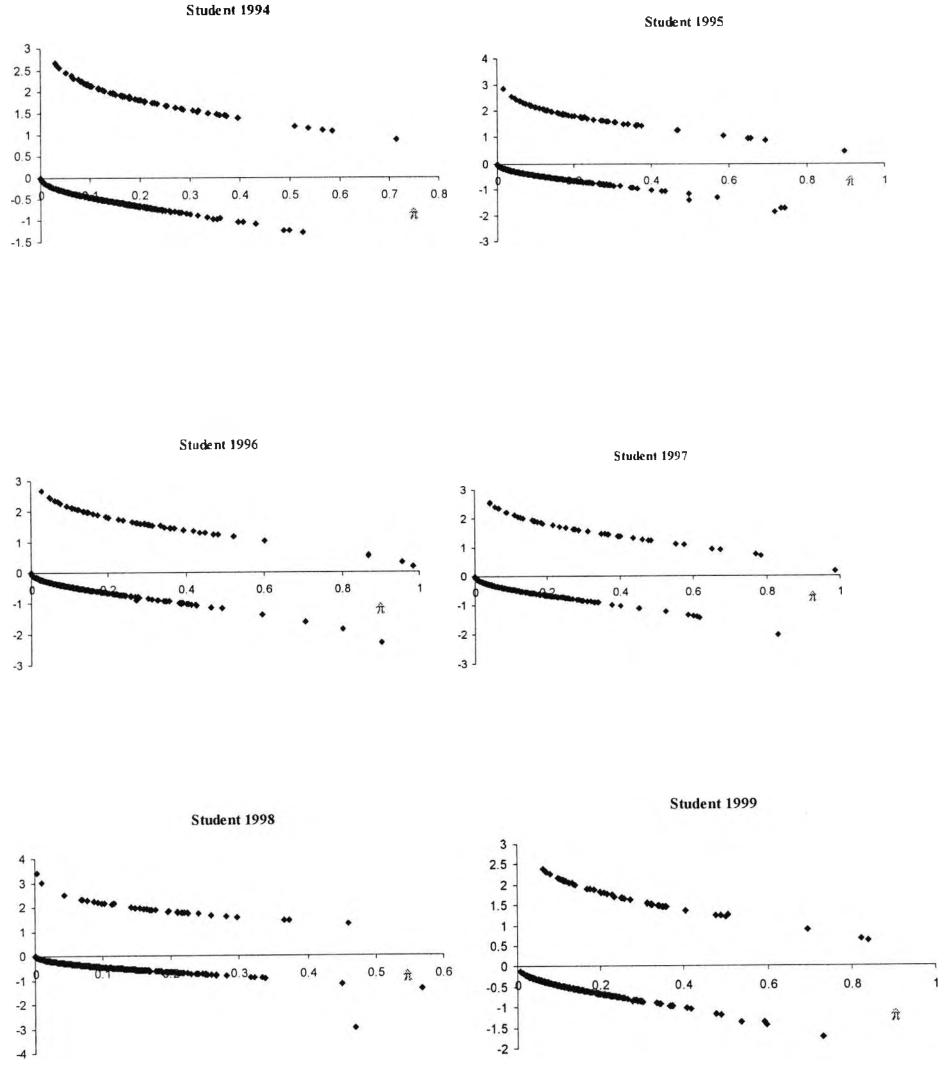


Figure 3.9.4 Studentized Residuals



3.10 Conclusions from the Empirical Study

The purpose of the study undertaken in this chapter was the identification of factors affecting the credit risk of a company. We have carried out this task using three alternative methods: Linear discriminant analysis, quadratic discriminant analysis and logistic regression.

First we estimated linear discriminant functions for six years, 94, 95, 96, 97, 98 and 99 and we found that there is a group of variables that can differentiate between downgraded and non downgraded companies in a statistically significant way throughout the period. The classification results based on in-sample estimates show that in all cases at least 60 percent of the companies that were downgraded could be predicted to do so.

The second technique we used was quadratic discriminant function, which relaxes the assumption of equal covariance matrices in the two groups of companies, although it still relies on the assumption of normally distributed returns. The results from the quadratic discriminant analysis do not improve the results of the linear case. In most years the results were in fact inferior.

The third approach to classification of observation into one of the two groups of companies in our sample, i.e. migrating and non migrating companies, the logistic regression approach, provides an alternative to Discriminant Analysis when the distributional assumption underpinning DA are not met.

In the case of logistic regression, we have found variables that were significant throughout the whole period. The most important variables were those measuring profitability, leverage, solvency liquidity. Growth variables were occasionally significant whereas activity ratios were insignificant in the final equation. Size is not an important differentiating factor between the two group of companies, and in the only case when it came in as significant, it implied that size contributes to the probability of being downgraded.

The classification of results on the basis of logistic regression is comparable to that of Discriminant analysis in terms of the overall performance but it is significantly better when the hit rate for group 1 companies is the criterion. All in all, in terms of classification performance by both methods, the hit rates are satisfactory when one considers the non-homogeneity of the sample.

Although the two techniques have produced slightly different results so far, it would be correct to say that there has been a set of factors that appear to be significant in both approaches. Although a specification test has been proposed by Lo (1986) to choose between the two techniques we have not conducted a formal test due to the computational requirements.

The entire statistical analysis was based on the use of financial ratios only, disregarding other qualitative factors that may exert an influence on the performance of a company. The results indicate that there is enough information in the financial statements of a company to help the identification of companies that may face problems. On the other hand it is also clear that the inclusion of qualitative variables would possibly produce stronger results.

3.11 Appendix – Univariate logistic regressions

Table A1 - Profitability Ratios

	94	95	96	97	98	99
R16						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-375.364	-366.886	-314.969	-268.57	-233.986	-257.28
$LR(\chi_1^2)$	12.536	39.414	29.331	25.73	12.014	17.32
R(N)	0.046	0.133	0.115	0.125	0.066	0.0955
HL	8.456	9.422	10.826	11.263	17.561	14.441
n_{00}/n_0	65.7	60.5	59.1	63.2	67.4	65.9
n_{11}/n_1	50.0	70.1	71.9	74.6	68.4	71.3
$(n_{00}+n_{11})/(n_0+n_1)$	63.6	61.7	60.6	66.8	67.5	67.85
$\hat{\alpha}_L$	-1.277*	-0.960*	-1.067*	-1.832	-1.489	-1.669
β	-4.033*	-8.776*	-8.769*	-6.732	-5.050	-5.981
R12						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-380.723	-395.239	-327.342	-286.38	-233.278	-266.45
$LR(\chi_1^2)$	7.177	11.061	16.958	7.92	12.722	8.15
R(N)	0.027	0.038	0.067	0.072	0.07	0.06
HL	18.587	23.010	10.108	11.213	4.652	6.78
n_{00}/n_0	25.6	36.3	43.8	56.7	76.0	55.63
n_{11}/n_1	86.4	82.1	75.4	74.6	55.3	69.8
$(n_{00}+n_{11})/(n_0+n_1)$	33.7	41.9	47.6	60.1	73.9	60.1
$\hat{\alpha}_L$	-1.682*	-1.718*	-1.714*	-1.912	-1.830	-1.711
β	-0.614*	-0.809*	-1.331*	-2.34	-3.542	-3.023

TableA2- Solvency Ratios

	94	95	96	97	98	99
R24						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-385.254	-401.823	-322.70	-276.96	-227.84	-260.45
$LR(\chi_1^2)$	2.646	4.477	21.592	17.34	18.16	14.15
R_N^2	0.010	0.016	0.085	0.073	0.013	0.043
$HL(\chi_8^2)$	9.182	21.536	6.558	17.34	19.831	18.86
n_{00}/n_0	94.4	63.3	78.0	85.6	96.7	91.31
n_{11}/n_1	12.1	55.2	54.4	49.12	13.2	31.26
$(n_{00}+n_{11})/n$	83.4	62.3	75.2	76.21	88.3	82.25
$\hat{\alpha}_L$	-2.005*	-2.153*	-2.497*	-3.011	-2.328	-2.669
β	1.563**	2.538*	6.312*	4.213	2.222	3.417
R13						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-383.44	-400.794	-337.58	-285.97	-241.98	-269.49
$LR(\chi_1^2)$	4.46	5.506	6.72	8.33	4.02	5.11
R_N^2	0.028	0.019	0.017	0.017	0.012	0.018
$HL(\chi_8^2)$	8.77	10.29	11.35	16.28	7.43	6.84
n_{00}/n_0	74.9	64.9	68.7	69.6	68.5	64.8
n_{11}/n_1	59.4	58.2	61.3	63.2	60.4	53.4
$(n_{00}+n_{11})/(n_0+n_1)$	69.7	64.1	67.2	65.3	66.4	62.0
$\hat{\alpha}_L$	-3.249*	-2.038*	-1.997*	-3.045*	-2.991*	-2.401*
β	0.203**	0.106**	0.192**	0.145**	0.160**	0.183**
R25						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-383.991	-394.036	-338.87	-286.2135	-242.76	-271.493
$LR(\chi_1^2)$	3.909	12.264	5.43	8.0865	3.24	3.107
R_N^2	0.015	0.042	0.034	0.0285	0.041	0.024
$HL(\chi_8^2)$	12.106	18.760	8.44	9.38	6.27	4.38
n_{00}/n_0	10.6	64.7	54.1	37.65	42.6	37.9
n_{11}/n_1	90.9	62.7	80.1	76.8	70.5	60.3
$(n_{00}+n_{11})/(n_0+n_1)$	21.3	64.5	59.2	42.9	45.7	40.2
$\hat{\alpha}_L$	-2.016*	-2.292*	-2.345	-2.154	-3.34	-4.02
β	0.227*	0.554*	0.667	0.3905	0.478	0.613

Table A3: Leverage Ratios

	94	95	96	97	98	99
R26						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-383.035	-399.47	-338.065	-288.53	-242.604	-270.34
$LR(\chi_1^2)$	4.865	6.83	6.235	5.77	3.396	4.26
R_N^2	0.018	0.024	0.025	0.022	0.019	0.023
$HL(\chi_8^2)$	4.672	6.701	5.601	6.712	9.978	8.719
n_{00}/n_0	67.3	73.2	53.1	64.1	72.7	54.6
n_{11}/n_1	43.9	64.1	63.2	66.8	36.8	56.
$(n_{00}+n_{11})/n$	64.2	69.2	54.3	64.9	69.1	55.1
$\hat{\alpha}_L$	-2.708*	-2.917*	-3.006*	-3.615*	-3.142*	-4.327
β	1.387*	1.426*	1.695*	1.562*	1.556*	1.830
R4						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-386.7	-404.9	-343.54	-292.06	-244.02	-272.25
$LR(\chi_1^2)$	1.2	1.4	0.76	2.24	1.98	2.35
R_N^2	0.018	0.024	0.025	0.022	0.019	0.023
$HL(\chi_8^2)$	4.672	6.701	5.601	6.712	9.978	8.719
n_{00}/n_0	67.3	73.2	53.1	64.1	72.7	54.6
n_{11}/n_1	43.9	64.1	63.2	66.8	36.8	56.
$(n_{00}+n_{11})/(n_0$	64.2	69.2	54.3	64.9	69.1	55.1
$\hat{\alpha}_L$	-2.708*	-2.917*	-3.006*	-3.615*	-3.142*	-4.327
β	1.387*	1.426*	1.695*	1.562*	1.556*	1.830

Table A4: Liquidity Ratios

	94	95	96	97	98	99
R7						
$2 \ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2 \ln L_p$	-375.678	-389.46	-323.501	-276.37	-233.57	-264.74
$LR(\chi_1^2)$	12.222	16.84	20.799	17.93	12.43	9.86
R_N^2	0.045	0.080	0.082	0.091	0.038	0.084
$HL(\chi_8^2)$	7.924	4.661	4.011	4.495	19.794	3.68
n_{00} / n_0	62.9	61.600	52.4	61.200	76.9	51.60
n_{11} / n_1	51.5	65.575	71.9	63.925	52.6	68.6
$(n_{00} + n_{11}) / n$	61.3	63.500	57.4	67.050	74.4	64.5
$\hat{\alpha}_L$	-0.891*	-0.639	-0.512**	-0.744	-1.209*	-0.723
β	-0.743*	-1.864	-1.207*	-2.252	-0.741*	-1.984
R22						
$2 \ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2 \ln L_p$	-376.13	-393.97	-323.002	-278.74	-238.37	-264.49
$LR(\chi_1^2)$	11.77	12.33	21.298	15.56	7.63	10.11
R_N^2	0.043	0.0565	0.085	0.07	0.0775	0.060
$HL(\chi_8^2)$	4.866	4.759	5.968	4.652	5.31	5.08
n_{00} / n_0	71.5	73.75	65.6	76.0	70.8	71.15
n_{11} / n_1	48.5	51.9	63.2	55.3	59.25	53.87
$(n_{00} + n_{11}) / n$	68.4	71.15	65.3	73.9	69.6	69
$\hat{\alpha}_L$	-1.598*	-1.714	-1.700*	-1.830	-1.765	-1.70
β	-2.496*	-3.019	-3.498*	-3.542	-3.52	-3.01

Table A5: Growth Ratios

	94	95	96	97	98	99
R1						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-380.3	-397.6	-332.76	-288	-240.1	-380.3
$LR(\chi_1^2)$	7.6	8.7	11.54	6.3	5.9	7.6
R_N^2	0.014	0.033	0.046	0.03	0.09	0.03
$HL(\chi_8^2)$	11.524	9.74	12.974	9.73	0.395	12.249
n_{00}/n_0	90.7	80.34	40.0	30	91.1	65.35
n_{11}/n_1	19.7	45.94	73.7	56.2	13.2	46.7
$(n_{00}+n_{11})/n$	81.2	76.55	44.1	33.07	83.2	62.65
$\hat{\alpha}_L$	-1.84	-1.94	-1.930*	-2.33	-2.166	-3.34
β	-0.463	-0.891	-1.427*	-0.823	-0.664	-1.723
R20						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-377.316	-401.4	-339.6	-288.2	-240.6	-268.7
$LR(\chi_1^2)$	10.584	4.9	4.7	6.1	5.4	5.9
R_N^2	0.039	0.06	0.059	0.0415	0.023	0.0595
$HL(\chi_8^2)$	10.497	8.584	9.673	6.957	5.330	9.1285
n_{00}/n_0	86.4	67.9	68.4	67.65	67.4	68.15
n_{11}/n_1	31.8	56.1	57.7	62.25	68.4	56.9
$(n_{00}+n_{11})/n$	79.1	66.5	67.1	67	67.5	66.8
$\hat{\alpha}_L$	-1.811*	-1.975	-1.871*	-2.0585	-2.142	-1.923
β	-6.001*	-10.011	-8.217*	-8.177	-6.343	-9.114
R28						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-383.194	-404.2	-342.5	-293.24	-242.97	-271.81
$LR(\chi_1^2)$	4.692	2.1	1.8	1.06	3.03	2.79
R_N^2	0.017	0.010	0.012	0.009	0.014	0.011
$HL(\chi_8^2)$	20.508	20.508	20.508	20.508	20.508	20.508
n_{00}/n_0	33.6	33.6	33.6	33.6	33.6	33.6
n_{11}/n_1	71.2	71.2	71.2	71.2	71.2	71.2
$(n_{00}+n_{11})/n$	38.6	38.6	38.6	38.6	38.6	38.6
$\hat{\alpha}_L$	-4.521*	-4.521*	-4.521*	-4.521*	-4.521*	-4.521*
β	0.524*	0.524*	0.524*	0.524*	0.524*	0.524*

Table A6: Activity Ratios

	94	95	96	97	98	99
R21						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-386.7	-405.2	-343.52	-292.96	-244.56	-273.49
$LR(\chi_1^2)$	1.2	1.1	0.78	1.34	1.44	1.11
R_N^2	0.035	0.035	0.035	0.035	0.035	0.035
$HL(\chi_8^2)$	14.343	14.343	14.343	14.343	14.343	14.343
n_{00}/n_0	34.3	34.3	34.3	34.3	34.3	34.3
n_{11}/n_1	78.8	78.8	78.8	78.8	78.8	78.8
$(n_{00}+n_{11})/n$	40.2	40.2	40.2	40.2	40.2	40.2
$\hat{\alpha}_L$	-1.856	-1.856	-1.856	-1.856	-1.856	-1.856
β	-3.982	-3.982	-3.982	-3.982	-3.982	-3.982
R18						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-378.31	-402.6	-341.5	-292.4	-242.9	-271.63
$LR(\chi_1^2)$	9.59	3.7	2.8	1.9	3.1	2.97
R_N^2	0.035	0.035	0.035	0.035	0.035	0.035
$HL(\chi_8^2)$	14.343	14.343	14.343	14.343	14.343	14.343
n_{00}/n_0	34.3	34.3	34.3	34.3	34.3	34.3
n_{11}/n_1	78.8	78.8	78.8	78.8	78.8	78.8
$(n_{00}+n_{11})/(n_0+n_1)$	40.2	40.2	40.2	40.2	40.2	40.2
$\hat{\alpha}_L$	-1.856	-1.856	-1.856	-1.856	-1.856	-1.856
β	-3.982	-3.982	-3.982	-3.982	-3.982	-3.982
R19						
$2\ln L_0$	-387.9	-406.3	-344.3	-294.3	-246.0	-274.6
$2\ln L_p$	-386.67	-405.16	-342.52	-292.95	-243.98	-272.93
$LR(\chi_1^2)$	1.23	1.14	1.78	1.35	2.02	1.67
R_N^2	0.035	0.035	0.035	0.035	0.035	0.035
$HL(\chi_8^2)$	14.343	14.343	14.343	14.343	14.343	14.343
n_{00}/n_0	34.3	34.3	34.3	34.3	34.3	34.3
n_{11}/n_1	78.8	78.8	78.8	78.8	78.8	78.8
$(n_{00}+n_{11})/(n_0+n_1)$	40.2	40.2	40.2	40.2	40.2	40.2
$\hat{\alpha}_L$	-1.856	-1.856	-1.856	-1.856	-1.856	-1.856
β	-3.982	-3.982	-3.982	-3.982	-3.982	-3.982

Chapter 4 – Measuring the Market Risk of Equities: A Review of Alternative Approaches and Some Empirical Tests on the Their Assumptions.

4.1 Introduction

The role of market risk in financial decision making is the central issue in financial economics. Financial theory defines the relationship between reward and risk but, there is little agreement as to how risk should be measured. The seminal work of Markowitz, [see Markowitz (1991)] established the variance as the appropriate measure of risk and made explicit the trade-off between risk and reward in the context of a portfolio of financial assets. Extensions by Sharpe (1964), Lintner (1965), and Ross (1976), have used equilibrium arguments to develop asset pricing models such as the capital asset pricing model (CAPM) and the arbitrage pricing theory (APT), relating the expected return of an asset to a set of risk factors. A common theme of these models is the assumption of normally distributed returns.

However, financial asset returns often possess distributions with tails heavier than those of the normal distribution and therefore models based on the assumption of normally distributed returns should be seen as approximations, whose success is a matter of empirical testing. Many alternative probability distributions have been proposed as more suitable representations of the underlying asset returns but few asset pricing theories are based on those alternative distributional assumptions.

The measurement of risk is important of course on its own merit not only as part of a theory for the pricing of assets. Following the established financial theory paradigm, and accepting the variance as the appropriate measure of risk presented two difficulties to risk managers. The first one arose from the need to have a measure that would express the potential losses in monetary terms rather than as deviations for the mean return. Using concepts from the insurance industry where losses were expressed in monetary terms, a number of financial institutions such as J.P. Morgan, and Bankers Trust, in the early 1990*s*, proposed a new risk measure to quantify by a single number an institution's aggregate exposure to market risk. This measure, commonly known today as *Value- at- Risk* (*VaR*), is now used to measure not only market risk but also other forms of risk such as liquidity and credit risk. The second reason that the variance became a problematic measure of risk was the fact that it could only be justified theoretically by making the assumption that returns are normally distributed, an assumption that was not supported by the evidence.

In its original application Value-at-Risk was also based on normally distributed returns, making it a simple multiple of the standard deviation, with the multiple being the value of a chosen quantile from the standardised normal distribution. An important improvement in the measure of risk using the normally distributed returns assumption was the introduction of time-varying volatility by RiskMetrics (1996). Since the original application the research has expanded exponentially in two areas. One is the modelling of volatility; the second is the specification of the appropriate model either for the entire probability distribution or for the tails of the distribution.

This chapter serves as a critical review of the literature on the measurement of the market risk for equities, concentrating in particular on the issues of implementation of one particular measure the Value-at-Risk. There have been recently a large number of books and articles that provide useful overviews of the various approaches to Value at Risk methods and calculation issues and reviews of these developments can be found in Duffie and Pan (1997), Dinenis (2002a) and Dowd (2002).

The chapter is organised in eight sections as follows. In Section 1, we introduce, in a common statistical framework, the most popular risk measures such as the variance, semi variance, value at risk and expected shortfall. Since the objective is the measurement of risk we only consider quantitative measures of risk.

In the second section we look at some of the issues that need to be tackled in order to apply any of the measures and in particular value-at-risk. The issues we examine relate to the appropriate distribution of returns, to the existence of the moments of the distribution, to the dynamic specification of the return process and issues of portfolio risk, and time aggregation. Finally following the discussion in Artzner, Delbean, Eber and Heath (1999) we examine whether the various risk measures are coherent.

In Section 3 we test the hypothesis of normality for a set of 132 stocks quoted on the Athens Stock Exchange. We examine the properties of daily, weekly and monthly returns and test the hypothesis of normality using a variety of tests. The assumption of normally distributed returns was introduced by Bachelier but early on Mandelbrot (1963) recognized the heavy-tailed, highly peaked nature of certain financial time series. Since then many models have been proposed to model heavy-tailed and semi-heavy distributions of returns of financial assets. These techniques are increasingly drawn not only from finance and economics but also from other areas. So in Section 4 we test the empirical performance of two general classes of non-Gaussian distributions. We fit to the set of our 132 companies a generalised skewed –t distribution, and a stable distribution. The performance of each distribution is tested using both likelihood function and non-parametric criteria.

In Section 5, we model the dynamics of equity returns using a low order autoregressive model for the mean and a t- GARCH (1, 1) model to model volatility. Volatility clustering has been proposed as one explanation for the existence of fat tails in asset returns, so testing for the existing of serially correlated volatility is an important issue. Also given the critical role that volatility forecasting plays in the calculation of VaR, it is important to test in order to know what the appropriate volatility model is. Using the standardised residuals we can test whether the fat tails of the return distributions are due to volatility clustering.

In Section 6, we review the methods of estimating tail probabilities directly rather than through the estimation of the entire distribution. The various approaches to estimating tails such as the block method or the method of generalised Pareto Distribution is critically reviewed and the problems of implementation that arise. In the same section we estimate the lower tail of the 132 stocks of our sample. The implication that returns of financial assets have a heavy-tailed distribution may be profound to a risk manager in a financial institution or indeed to an investor. For example, three standard deviation events may occur with a much larger probability when the return distribution is heavy-tailed than when it is normal. Quantile based measures of risk, such as value at risk, may also be drastically different if calculated for a heavy-tailed distribution. This is especially true for the highest quantiles of the distribution associated with very rare but very damaging adverse market movements.

In Section 7, we review the Cornish-Fisher approximation to the quantiles of an arbitrary distribution. The approximation adjusts the quantile of the normal distribution by adding the effect of nonzero skewness and kurtosis. The empirical performance is evaluated by comparing it to empirical quantiles.

Finally in Section 8 we present the conclusions of this chapter with some suggestions for further research.

4.2 Measures of Market Risk for Equities

The measurement of risk in equity markets, indeed in all financial markets, has been the subject of countless efforts long before financial theory acquired its rigorous foundation. A rigorous approach to modelling risk starts with the description of the equity return series $\{R\}$ as a random variable with density and distribution functions $f(r)$ and $F(r)$ respectively. Knowledge of the distribution function allows a complete characterisation of the risk profile of an investment or financial position. The most common distribution in finance is the normal distribution, but as we have mentioned above other distributions have been use to model equity returns.

The number of distributions that have been proposed to model the density of returns is huge (see e.g. Andreou, Pittis and Spanos (2001) for a recent review of the various distributions employed in finance) and still increasing. The statistical measures of risk are classified into two broad categories. Moment-based and quantile based. For some distributions e.g. the normal distribution, this classification is irrelevant as one approach is equivalent to the other. However there are cases of distributions where the two approaches produce different results.

4.2.1 Moment –Based Risk Measures

Moment-based measure of risk use the moments of a distribution to express information about the probability distribution. The qth moment of a distribution is defined as

$$E(r^q) = m_q = \int_{-\infty}^{\infty} r^q f(r)dr \tag{4.2.1}$$

The first order moment or the mean is defined for $q = 1$ and is given by

$$E(r) = m_1 = \mu = \int_{-\infty}^{\infty} rf(r)dr$$

Moment –based measures of risk consider risk as the dispersion of return values around the mean. The statistical measures of this category are based on the central moments of the distribution defined as

$$c_q(r) = E(r-\mu)^q = \int_{-\infty}^{\infty} (r-\mu)^q \phi(r)dr$$

Markowitz (1952) established the second central moment, $\sigma^2 = E(r-\mu)^2 = \int_{-\infty}^{\infty} (r-\mu)^2 f(r)dr$

known as the variance of a random variable as the main measure of risk in finance. The variance measures both positive and negative deviations from the mean as risk. This measure has dominated developments in financial theory and is the "risk" component in the "risk return trade-off" that financial models such as the CAPM and the APT are trying to explain. Markowitz defined the reward of a portfolio of assets as the weighted average of the rewards on the individual assets. The expected return is therefore a linear function of the returns on the individual assets. The variance of a portfolio of assets is not however a weighted average of the variances of the assets in the portfolio. It is instead a nonlinear function and it is this nonlinearity that allows for reduction in portfolio risk due to diversification. The portfolio risk is given by

$$\sigma_P^2 = \mathbf{w}'\Sigma\mathbf{w}$$

where $\mathbf{w}$ is the vector of weights in the portfolio and Σ is the variance-covariance matrix of returns. The property of variance of linear combination of random variables to become small is the fundamental building block of financial theory and risk management.

Despite its dominance as a measure of risk in finance, the use of variance as a measure of total risk of an asset has been seriously questioned since it is based on some assumptions which are not supported by the empirical evidence. There are two assumptions that justify the variance as a measure of risk. The first assumption is the assumption that returns follow a normal distribution. This is clearly at odds with a large body of empirical research that shows that returns are not normally distributed.

The mean variance theory of Markowitz as well as the CAPM and the APT rely either explicitly or implicitly on the assumption of normally distributed returns.

The second assumption that can justify the use of variance is when we assume that investors have a quadratic utility function. Such a utility is not however consistent with observed investor behaviour since it implies negative marginal utility above a certain level of wealth.

The third moment also plays an important role. The third central moment

$$c_3 = E(r-\mu)^3 = \int_{-\infty}^{\infty} (r-\mu)^3 f(r)dr$$

of a distribution measures the asymmetry of a distribution and it can be seen also as a measure of risk. Its standardised version known as the coefficient of skewness is given by

$$\gamma_3 = \frac{E(r-\mu)^3}{\sigma^3}$$

A non zero value for the coefficient of skewness means that there are more observation on one side of the mean than on the other side. Thus deviations from the mean do not convey the same information. There may be cases of two distributions with equal variance but one having positive skewness and the other having negative skewness. The probability of a large negative return is higher in the case of the distribution with negative skewness and therefore from a risk management point of view, this asset is riskier. Ignoring skewness will produce the wrong result.

In a similar fashion the fourth central moment of a distribution

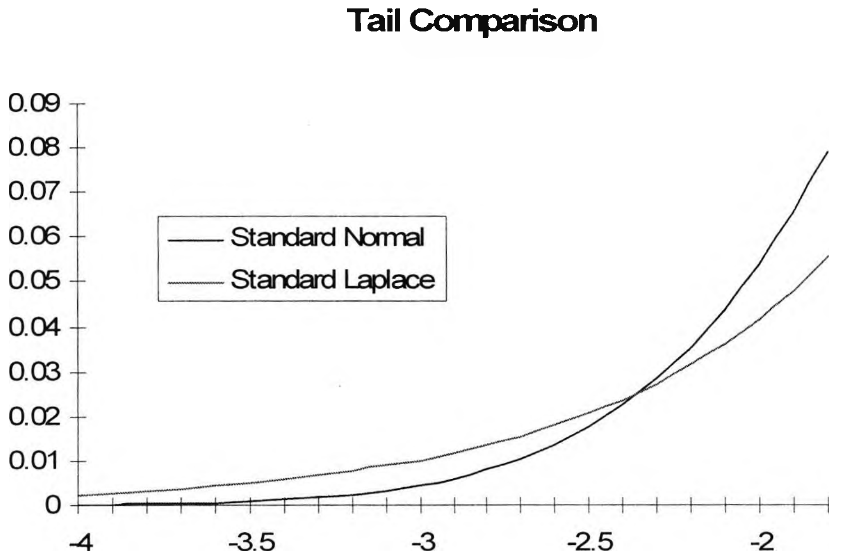
$$c_4 = E(r-\mu)^4 = \int_{-\infty}^{\infty} (r-\mu)^4 \phi(r)dr$$

which measures kurtosis (i.e. the peakedness and the fatness of a distribution) can make the classification of assets on the basis of the variance completely useless. The concept of kurtosis is best understood in relation to the normal density function whose coefficient of kurtosis is 3. The coefficient of excess kurtosis is given by

$$\gamma_4 = \frac{E(r-\mu)^4}{\sigma^4} - 3$$

Distributions with tails fatter than the normal imply a larger concentration of observations in the tails of the distribution. Thus distributions which exhibit excess kurtosis exhibit a higher probability of extreme observations. This can be seen by comparing in Figure 4.1 two symmetric distributions, the standardised normal distribution with a mean of zero and a variance of one, and the Laplace distribution with a mean of zero and a variance of one. The only difference between the two is that the Laplace distribution has a coefficient of kurtosis which is equal to 3 whereas the normal distribution has a kurtosis coefficient of 0.

Figure 4.1: Kurtosis and Tails



The measures we have considered so far are all based on deviations from the mean, which takes into account both positive and negative dispersion around the mean. Fishburn (1977) introduced a different class of risk measures , the lower partial moments (LPM) based again on the notion of dispersal of values but this time it only measures the dispersion of values below a target value d.

$$f_k(r) = E(d-r)^k = \int_{-\infty}^{d} (d-r)^k f(r)dr$$

Of particular interest is the semi-variance which is a special case when $d = \mu$ and $k = 2$. The semi-variance is given by

$$LPM_K(r) = \int_{-\infty}^{d} (\mu - r)^2 f(r)dr$$

The LPM measure was first suggested (as the semi-variance measure) by Markowitz (1959). An optimal algorithm for the semi-variance was first developed by Hogan and Warren(1973). Despite the early predominance of the semi variance version of LPM, Porter(1974) shows empirically that semi variance calculated from the mean return is inferior to a semi variance calculated from a target rate of return. Roy(1952) developed the target rate of return approach earlier. Fishburn(1977) and Ang and Chua(1979) provide proofs using utility theory showing that the target return should be the standard for calculating the semi variance rather than the mean return.

4.2.2 Quantile-Based Measures

Quantile-based measures are based on another characterisation of the probability distribution through its quantiles. For a random variable R with distribution function $F(r) = \Pr(R < r)$ the pth quantile is defined as that value of r , denoted as r_p for which $F(r_p) = \Pr(R < r_p) = p$. If the distribution function can be inverted, then the pth quantile can be calculated as $r_p = F^{-1}(p)$ where $F^{-1}(p)$ is the inverse function of $F(r)$.

The Value-at-Risk (VaR) at confidence level $c = 1 - p$ is defined as the pth quantile of R. That is $VaR(p) = r_p = F^{-1}(p)$. The confidence level $c = 1 - p$ is typically a large number between 0.95 and 1.

From the definition of VaR as the pth quantile of a distribution it is clear that the calculation of VaR requires the specification of a probability distribution for the returns. Thus VaR can be calculated for specific density functions like the Gaussian density function, but in general we are not able to get a closed form solution.

The calculation of the VaR is greatly simplified if we make the assumption that the distribution is symmetric and its first and second moment exists. In that case the VaR of a random variable with mean μ and variance σ can be written as $VaR = \mu + \sigma F_{0.1}^{-1}(p)$ where $F_{0.1}^{-1}(p)$ is the distribution function of a random variable with a mean of zero and a standard deviation of 1. Because of its intuitive appeal and simplicity, *VaR* has become the standard risk measure used around the world today. For example, today *VaR* is frequently used by regulators to determine minimum capital adequacy requirements. In 1995, the Basle Committee on Banking Supervision suggested that banks be allowed to use their own internal *VaR* models for the purpose of determining minimum capital reserves. The internal models approach of the Basle Committee is a ten day *VaR* at the 99% confidence level multiplied by a *safety factor* of at least 3. Thus if *VaR* = £1 , the institution is required to have at least £3 in reserve in a safe account.

The safety factor of three is an effort by regulators to ensure the solvency of their institutions. It has also been argued, see Stahl [1997] or Danielsson, Hartmann and De Vries [1998] , that the safety factor of three comes from the heavy-tailed nature of the return distribution.

Expected Shortfall

The VaR measure gives the frequency with which the quantile will be exceed but it does not tell us by how much. The last measure of risk is the Expected Shortfall defined as

$$E(r_t / r_t < -VaR_t)$$

The expected shortfall is a special case of the measure proposed by Fishburn. The expected shortfall can be calculated for particular distributions. For example if asset returns follow a normal distribution, then it can be shown that the expected shortfall can be written as

$$E(r / r < -VaR(p)) = \sigma \frac{\phi(\Phi^{-1}(p))}{p}$$

where

$VaR(p)$ is the VaR based on the pth quantile

$\Phi^{-1}(p)$ is the pth quantile of the standard normal distribution

σ is the standard deviation of return

$\phi()$ is the density function of the standard normal distribution

This concludes the risk measures that are commonly used to quantify the risk exposure of a single financial position, portfolio of assets or an entire financial institution to market risk.

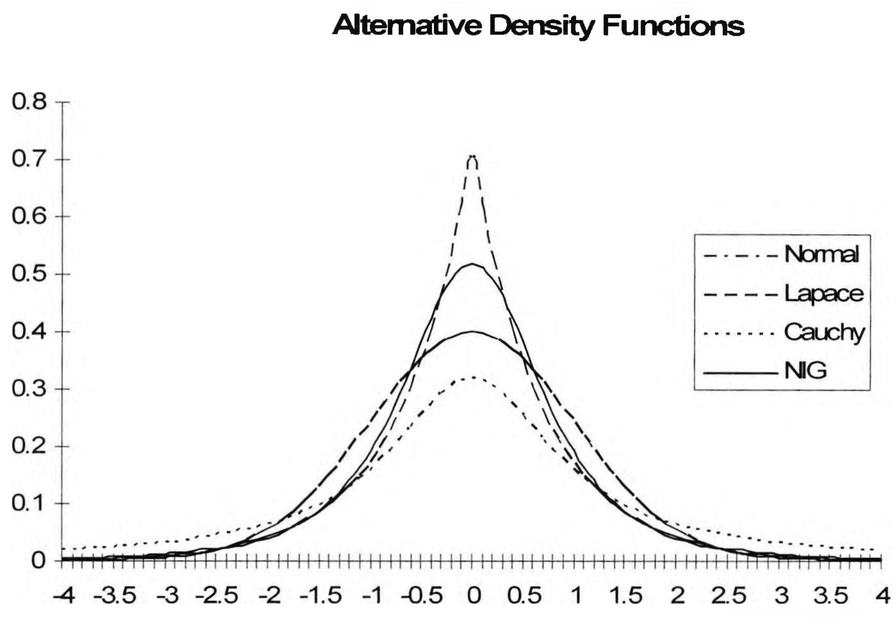
4.3 Issues Arising in the Estimation of Risk Measures

In this section we present some of the practical and theoretical problems associated with the application of the risk measures we have discussed.

4.3.1 Distributional Assumption and Model Risk

The first issue that arises in implementing any of the measures of risk we have reviewd is the decision regarding the type of probability distribution that needs to be employed for the calculation of its moments, quantiles or expected shortfall. Heidelberger and Shahabudding (2000) for example state " the central problem in risk management is estimation of the profit-and-loss distribution of a portfolio over a specific horizon. Given this distribution, the calculation of specific risk measures is relatively straightforward. Value-at-Risk, for example, is a quantile of this distribution. .... The difficulty in estimating these types of risk measures lies primarily in estimating the profit and loss distribution itself, especially the tail of this distribution associated with large losses".

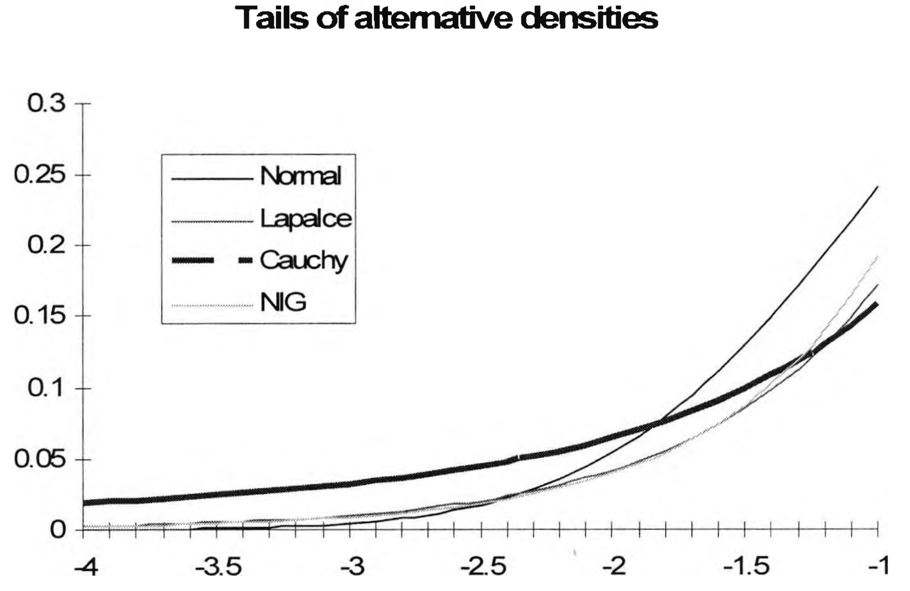
Figure 4.2: Example of Model Risk



Use of the wrong distribution will produce completely different results. This is known as model risk. By way of example we have plotted in Figure 4.2, the densities of four probability distributions all of which are symmetric with mean zero and standard deviation of one.

These distributions are different but this may not be detected easily since the bulk of the data lies between the same set of values (between -2 and +2) making it difficult to differentiate between the alternative distributions. Looking however at the tails of the distribution in Figure 4.3 reveals greater potential errors. The Cauchy distribution for example (thick line in the graph) postulates that the probability of extreme movements is significantly higher than the other three distributions. The real issue is whether there is a particular probability distribution that describes the behaviour of equity price changes and which could be adopted for the calculation of Value-at-Risk, so that the model risk is minimised. There are three distinct approaches that have been employed to estimate the quantiles of the distribution. One is to fit a probability distribution to the data. If data availability prevent the fitting of a distribution, then Monte Carlo techniques can be employed to derive the risk profile and its consistency with reality. The second approach is to fit a model to the tails of the distribution. Since we are interested in the tails of the distribution this approach may be better, because irrelevant information about the centre of the distribution does not contaminate the evidence. The last approach is a distribution-free approach. It has two main variants. The first variant is based on a historical simulation of the data. No distributional assumptions are made. The second variant is to approximate the quantiles of the distribution via the Cornish-Fisher approximation. The empirical performance of these techniques will be evaluated in subsequent sections.

Figure 4.3: Model Risk and Tails



4.3.2 Unconditional versus Conditional measure

In the previous section we discussed the effects of distributional assumptions. An important distinction for the purposes of risk measurement is between conditional and unconditional distribution. The conditional distribution is defined as the probability distribution conditional on a particular set of information available at the beginning of the horizon. For example different normal distributions may be estimated at different times, so the unconditional distribution need not be normal.

An early attempt to use conditional distributions was made by RiskMetrics by assuming that the variance of asset returns follows an exponentially declining scheme

$$\sigma_{t+1}^2 = \lambda \sigma_t^2 + (1-\lambda) r_t^2$$

where λ is the decay factor.

More sophisticated volatility models were pioneered by Engel (1982) and subsequently by Bollerslev (1986) who assumed that asset returns could be modelled as

$$r_t = \mu_t + \varepsilon_t \qquad (4.3.1)$$

where

$$\varepsilon_t = z_t \sigma_t \qquad (4.3.2)$$

The fundamental assumption of a conditional volatility model is that the volatility and the mean depend on information set Ψ_{t-1} which contains all the information available at the beginning of period t, i.e.

$\sigma_t^2 = h(\Psi_{t-1})$ and $\mu_t = g(\Psi_{t-1})$

An example is the ARCH(1) model introduced by Engel (1982) where

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \qquad (4.3.3)$$

Another example is the GARCH (q,p) model introduced by Bollerslev (1986) which is

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \qquad (4.3.4)$$

The conditional mean is usually expressed as a (low) order autoregressive model.

$$\mu_t = \phi_0 + \sum_{i=1}^{m} \phi_i r_{t-i} \qquad (4.3.5)$$

The variable z_t is an independently and identically distributed (i.i.d.) process, independent of Ψ_{t-1}, with $E(z_t) = 0$ and $Var(z_t) = 1$; A common assumption is that it is normally distributed, i.e. $z_t \sim N(0,1)$. Another popular assumption is that it follows the student –t distribution, i.e. $z_t \sim t(0,1,v)$.

An attractive feature of the GARCH model is that even though the conditional distribution is assumed to be normal, the unconditional distribution is non-normal with tails fatter than the normal distribution. Eagle and Gonzalez-Rivera (1991) have shown that the unconditional kurtosis for the conditional distribution in (4.3.4) is given by

$$\gamma_4 = \frac{E(\varepsilon_t)}{[E(\varepsilon_t^2)]^2} - 3 = \frac{E[E(\varepsilon_t^4 / \psi_{t-1})]}{[E[E(\varepsilon_t^2 / \psi_{t-1})]]^2} - 3 = \frac{E(3\sigma_t^2)}{[E(\sigma_t)]^2} - 3 = 3\frac{E(\sigma_t^2)}{[E(\sigma_t)]^2} - 3 \geq 0 \qquad (4.3.6)$$

For the ARCH(1) model with $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$ and normal density, Engle (1982) showed that whereas the kurtosis of $z_t \sim N(0,1)$ is 3, the kurtosis of the rate of return is

$$\gamma_4(r_t) = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} - 3 > 0 \quad \text{for } \alpha_1 > 0, 3\alpha_1^2 < 1$$

Similar expressions can be derived when $z_t \sim t(0,1,v)$. In this case we have

$$\gamma_4(r_t) = 3\frac{1-\alpha_1^2}{1-\theta\alpha_1^2} - 3 > 0 \quad \text{with } \theta = 3(v-2)/(v-4)$$

For the GARCH (1,1) model

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

we have from Bollerslev (1986) that

$$\gamma_4(r_t) = 3\frac{(1-\beta_{t-1}^2 - 2\alpha_1\beta_1 - \alpha_1^2)}{(1-\beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)} - 3$$

which exceeds 0 even further for $\alpha_1 > 0, \beta_1 > 0$ and $\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1$. From (4.3.6) it seen that the extend to which the kurtosis will exceed 0 is related to

$$E(\sigma_t^2) - [E(\sigma_t^2)]^2 = Var(\sigma_t^2)$$

The variance of σ_t^2 in turn depends on the unconditional fourth moment of ε_t. From this it can be seen that the heavier the tail of the conditional density of ε_t, the higher the degree of excess kurtosis that can be incorporated into the model.

A GARCH type model has a number of advantages. First the VaR can be calculated using the simple expression

$$VaR_{t/t-1} = \mu_{t/t-1} + \sigma_{t/t-1}\Phi_{0,1}^{-1}(p)$$

and the quantiles of the standardised normal distribution. The forecast of the future standard deviation is simply the estimate of volatility based on the GARCH model.

The empirical issue that arises in respect of the conditional distribution of asset returns is the appropriate model for the mean and the variance. A plethora of models has been proposed to model volatility dynamics and the risk manager has to choose the correct model in order to get the right conditional distribution and appropriate quantiles. A related issue is whether the conditioning process should be applied to the mean and the variance only or whether the dynamics of other moments should also be taken into account. For example if $\varepsilon_t / \Psi_{t-1}$ follows a t-distribution, the question is whether we should specify dynamic models for the skewness and kurtosis not just for variance as in Hansen (1994).

The choice between conditional and unconditional models for forecasting VaR is not clear-cut. For, whereas it is accepted that conditional models are superior for short-term forecasts, their value vanishes as the time horizon increases. Christoffersen and Diebold (2000) argue that the recent history of data series has little to tell about the probability of events occurring in the future. This applies especially to the prediction of rare events which are assumed to be stochastically independent. Therefore Danielsson and de Vries (2000) recommend to derive predictions about extreme events from unconditional distributions.

4.3.2 Moment Existence

In the previous sections we have defined and described various statistical measures of risk based on the moments of a distribution. We have discussed the conditions under which the variance can be interpreted as a legitimate measure of risk consistent with utility maximising risk-averse individuals. During our discussion we made the implicit assumption that these moments exist. This is certainly true for some of the distributions we have already discussed such as the normal distribution, but it is not true that they exist for all distributions.

A necessary condition for the qth moment of a distribution to exist is that the density function of the distribution $f(r)$ should decay faster than $1/|r|^{q+1}$ for $|r|$ going towards infinity, otherwise the integral in equation (4.2.1) would diverge for large $|r|$. In the case of distributions with densities of the type

$$f(r) \sim \frac{qC^{\alpha}}{|r|^{\alpha+1}}$$

then all the moments such that $q \geq \alpha$ are infinite. A distribution, for which $\alpha \leq 2$ for example, would have no finite variance. An example of a distribution with the tail behaviour described above is the standard Cauchy distribution with density function[1]

$f(r) = \frac{1}{\pi(1+r^2)}$ and probability function $F(r) = 0.5 + \frac{\arctan(r)}{\pi}$ then it is well known that the moments of this distribution are not defined. In fact for the class of stable distributions of which the Cauchy distribution is a special case the most popular measure of risk, the variance, is not defined. The class of distributions that are characterised by such tails as well as the issues in estimating the parameter $\alpha \leq 2$ is dealt with in detail in the section on Stable Paretian distribution in Section 4.5. Similarly, for a Student –t distribution, the second moment exist only if the degrees of freedom parameter is larger than 2, whereas the degree of kurtosis exists if and only if the degrees of freedom parameter is larger than 4. That is, once we consider models that exhibit fat tails there is no guarantee that moments based risk measures exist.

4.3.3 Risk Aggregation

So far we have reviewed models of risk for individual equities. The issue that arises in practice is how to characterise the risk not of a single individual asset, though this could also be required, but that of a portfolios of assets. The issue in particular is whether information on the statistical distribution of individual assets, together with measures of dependence amongst themselves can lead to the characterisation of the risk of portfolios of assets.

The answer is no, unless the assets are normally distributed. In the normal case the portfolio VaR can be simply calculated in the same way as the single asset VaR. The only difference is that the portfolio standard deviation is now used instead of the single asset standard deviation.

Consider the case where the vector of returns $\mathbf{r} = \{r_1, ..., r_k\}'$ follows a multivariate normal distribution $\mathbf{r} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The portfolio with vector of weights $\mathbf{w} = \{w_1, ..., w_k\}'$ will have portfolio return and variance given by $r_p = \mathbf{w}'\mathbf{r}$ and $\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ respectively. The value at risk for the portfolio is given

$$VaR_p = \mu_p + \sigma_p \Phi^{-1}(\alpha)$$

where $\mu_p = \mathbf{w}'\boldsymbol{\mu}$ is the expected return of the portfolio which for daily returns is normally assumed to be zero and $\Phi^{-1}(\alpha)$ is the αth quantile of the standardised normal distribution.

The assumption of normality allows the derivation of important properties for the Value at Risk. Consider the two asset case for example for which we have with a little manipulation. For the two asset case for example it can be shown that

$$VaR_p = \sqrt{(VaR_1)^2 + (VaR_2)^2 + 2\rho_{12} VaR_1 \times VaR_2}$$

where ρ_{12} is the correlation coefficient between the two asset returns. Reflecting the properties of the variance we have that

[1] The Cauchy function is also a special case of the Student distribution with 1 degree of freedom

$$VaR_p = \sqrt{(VaR_1)^2 + (VaR_2)^2 + 2\rho_{12} VaR_1 \times VaR_2} \leq VaR_1 + VaR_2$$

As in the case of the portfolio variance the reduction in portfolio VaR depends on the correlation coefficient between the returns of the two assets in the portfolio. For example, when $\rho_{12} = 1$ then $VaR_p = VaR_1 + VaR_2$. That is the portfolio VaR is the sum of the VaRs of the individual securities. If, on the other hand, $\rho_{12} = -1$ then there is a portfolio which will produce $VaR_p = 0$. Finally if $\rho_{12} = 0$ then $VaR_p \leq VaR_1 + VaR_2$

The example above highlights that the key issue in the measurement of portfolio risk is the correlation between equity returns and it is our ability to model these correlations correctly that determines how effective our modelling of the portfolio risk is.

Leaving the aspects of parameter estimation aside, the normal distribution has the attractive properties that not only the marginal distributions of the multivariate normal are normal, but also a linear combination of normally distributed returns is also normal. Thus evidence from univariate distributions can be used to characterise the behaviour of multivariate distributions and thus the behaviour of portfolio of assets. Similarly, a linear combination of random variable that follow a stable distribution also follows a stable distribution (see e.g. Khindanova, Rachev and Schwartz (1999) for a discussion or Mandelbrot (1997)). However other distributions which describe the behaviour of returns better do not have these attractive properties. The student-t distribution for example, which captures the fat tails of asset returns, can be the marginal of a variety of other distributions (Tong (1990)). However a linear combination does not produce a t-distributed random variable. To overcome this problem many variants of multivariate t-distributions have been proposed which retain the attractive properties of the univariate distributon. The multivariate t-distribution for example which has been used by Glasserman, Heidelberger and Shahabuddin (2000) as a way of modelling portfolio risk is given by

$$t(\mathbf{r}, v) = \frac{\Gamma\left(\frac{v+d}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} v^{v/2} \left|\mathbf{\Sigma}\right|^{-\frac{1}{2}} \pi^{-d/2} \left(v + (\mathbf{r} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu})\right)^{-(v+d)/2}$$

Where d is the number of assets and v is the degrees of freedom parameter. This distribution has marginal distributions which are also t-distributions but it imposes the restriction that the degrees of freedom parameter v (see e.g. Anderson (1984), or Tong (1990)) is the same in all stocks an assumption that is not necessarily supported by the data. If the assumption of common degrees of freedom is violated then there are problems in using these model in the measurement of portfolio risk.

The discussion above has shown that apart from the normal distribution, univariate tests of the distributional assumptions do not provide information about the multivariate distribution and consequently the multivariate distributions themselves have to be the subject of empirical testing in order to assess the right distribution for the calculation of portfolio Value at Risk. These difficulties are partly the explanation for the use of the normal distribution in portfolio analysis.

4.3.4 Multi-period Risk Measures

We have shown so far how to calculate the value at risk for a single investment horizon which we have taken to be a single trading day. There are two approaches to extending the Value-at-Risk concept to more than one trading period. The first approach is to calculate asset returns for the desired investment horizon, e.g. weekly, 10-day or monthly asset returns and calculate the Value-at-Risk on the basis of those returns. The second approach is to find a way of extrapolating the daily VaR to longer time horizons.

The first approach has the serious drawback that the number of observations at the disposal of the researcher fall dramatically. For example with 10 years of data, a 10-day VaR will only leave 250 observations on which to calculate the empirical VaR at 99 percent confidence level. This may not be enough to allow estimation of the salient characteristics of the underlying distribution.

The second method avoids this problem. However, under this method, the calculation of Value –at-Risk for longer investment horizons requires assumptions about the independence or lack of it, of returns across time.

In the second approach we assume that returns are identically and independently distributed. The h-day return is given by

$$R(h) = \sum_{i=1}^{h} R_i$$

The Variance of the h-day return is, assuming independence, given by

$$\sigma^2(h) = \sum_{i=1}^{h} \sigma_i^2 = h\sigma^2$$

and so the standard deviation of the h-day return is given by

$$\sigma(h) = \sum_{i=1}^{h} \sigma_i^2 = \sigma\sqrt{h}\,.$$

Now if we define the 1-day VaR as $VaR(1) = \sigma\Phi^{-1}(p)$ and the h-day VaR as

$VaR(h) = \sigma(h)\Phi^{-1}(p)$ then it is obvious that the h-day VaR is given by, the famous square root rule:

$$VaR(h) = VaR(1)\sqrt{h}$$

Diebold et all (1997) point out that the correctness of the square root rule relies on three conditions. First, the structure of the considered portfolio may not change over time. When we talk about a single asset this is not a serious problem. Secondly, the returns must be identically and independently distributed and thirdly they must be normally distributed. To see what happens when the iid assumption is not fulfilled consider the case where we want to calculate the VaR over a period of two days. If we make the assumption that the returns are identically distributed but not independent then the variance of the two-day return is given by

$$\sigma^2_{(1+2)} = 2\sigma^2_D(1+\rho)$$

where

σ_D^2 = daily variance of returns

ρ = the correlation coefficient between the returns

Only when the returns are uncorrelated, that is when $\rho = 0$ in which case

$\sigma_{(1+2)} = \sqrt{2}\sigma_D$ we have the square root rule. In general the volatility of the h-day return will be given by

$$\sigma_h = \sqrt{h}\sigma_D$$

and the Value at Risk will be derived by simply multiplying the daily VaR by the square root of time elapsed.

$$VaR_h = \sigma_D\sqrt{h}\Phi^{-1}(\alpha)$$

The reliability of VaR estimates over longer investment horizons and how the distributional assumptions may impact on the measurement of risk has been discussed extensively. Christoffersen, Diebold, and Schuermann (1998) discuss the issues involved in calculating VaRs at different horizons and Diebold, Hickman, Inoue, and Schuermann (1998) study the problems arising from simple scaling rules of volatility across horizons. Christoffersen and Diebold (2000) investigate the usefulness of dynamic volatility models for risk management at various forecast horizons.

As we have said the assumption of normality may not be an appropriate way of describing the return distribution as all the available evidence supports the view that the tails of the empirical distribution are fatter than the tails of the Gaussian normal. We look at the implication for time aggregation of the two main ways through which kurtosis is introduced, that is the cases where the return follows a GARCH model and the case when the distribution of returns is described by a fat tail distribution.

In the case of a GARCH(1,1) model

$$\sigma_{t+1}^2 = \omega + \alpha r_t^2 + \beta\sigma_t^2$$

Drost and Nijman (1993) have shown that the h-period volatilities can be calculated as follows:

$$\sigma_{t+1}^2(h) = \omega(h)^2 + \alpha(h)r_t^2 + \beta(h)\sigma_t^2(h)$$

with

$$\omega(h) = h\omega\frac{1-(\alpha+\beta)^h}{1-(\alpha+\beta)}$$

$$a(h) = (\alpha+\beta)^h - \beta(h)$$

and $|\beta(h)| < 1$ as solution of the quadratic equation

$$\frac{\beta(h)}{1+\beta^2(h)} = \frac{a-(\alpha+\beta)^h - b}{a[1+(\alpha+\beta)^{2h}]-2b}$$

The coefficients a and b are defined as:

$$a = h(1-\beta)^2 + 2h(h-1)\frac{(1-\alpha-\beta)^2(1-2\alpha\beta-\beta^2)}{(k-1)[1-(\alpha+\beta)^2]}$$

$$+4\frac{[(h-1-h(\alpha+\beta)+(\alpha+\beta)^h][\alpha-\alpha\beta(\alpha+\beta)]}{1-(\alpha+\beta)^2}$$

$$b = [\alpha-\alpha\beta(\alpha+\beta)]\frac{1-(\alpha+\beta)^{2h}}{1-(\alpha+\beta)^2}$$

where k is the coefficient of kurtosis of the return distribution.

A comparison of the square root rule with the GARCH multi-period volatility expression reveals systematic differences which become larger with increasing the horizon h. If h goes to infinity, then α and β converge to zero and hence the stochastic terms vanish, whereas the first deterministic term increases. That means that the average levels of the h-period volatility coincide in both cases, but the square root rule magnifies the fluctuations of the volatility, while they actually become smaller with increasing time horizon. Diebold et all (1997) illustrate the magnitude of the difference of both methods of volatility forecasting by means of simulation experiments.

We now turn to the see the implication of fat tails on the multi-period VaR. Extreme value theory can also be used to link short term volatility to long-term volatility. Assume that the tail decay of a single period return is given by the power law

$$P\left(|X| > x\right) = Cx^{-\alpha}$$

For the h-period return, we have (Danielson and de Vries (2000)):

$$P\left(X_1 + X_2 + \cdots + X_h > x\right) = hCx^{-\alpha}$$

The above relation holds due to the linear additivity of the tail risks of fat tailed distributions. It follows that a multi-period VaR forecast of a fat tail return distribution under the iid assumption is given by

$$VaR(h) = VaR(1)h^{1/a}$$

If the returns have finite variances then $\alpha > 2$ and thus a smaller scaling factor applies than postulated by the square –root-rule (Danielson, Hartman and de Vries (1999)]. Suppose for example, that $\alpha = 3$, then the four day will be given by $VaR(4) = 1.587 \times VaR(1)$ whereas according to the square root rule we would have $VaR(4) = 2 \times VaR(1)$. The above results show that the square root rule can be misleading not only when the iid assumption is violated but also when the distribution is leptokurtic.

4.3.5 Coherence of Risk Measures

The measures we have examined so far have been criticised on practical grounds. For example, measures based on standard deviation are criticized based on their inability to describe rare events and *VaR* is criticized because of its inability to aggregate risks in a logical manner. In two, by now famous papers, Artzner, Delbean, Eber and Heath [1997] and Artzner, Delbean, Eber and Heath [1999] on financial risk, the authors propose a set of properties any reasonable risk measure should satisfy. Any risk measure which satisfies these properties is called *coherent*. If X and Y are the future values of two risky positions, a risk measure $\rho(.)$ is said to be coherent if it satisfies the following properties:

$\rho(X) + \rho(Y) \leq \rho(X+Y)$	(sub-additivity)
$\rho(tX) = t\rho(Y)$	(Homogeneity)
$\rho(X) \geq \rho(Y)$ if $X \leq Y$	(Monotonicity)
$\rho(X+n) = \rho(X) - n$	(risk-free condition)

for any number n and positive number t.

The authors show that neither the standard deviation nor the semi variance is coherent measures. VaR is not coherent either unless we assume that the returns are normally (or more generally elliptically) distributed. Only expected shortfall satisfies the criteria of coherence. The primary reason for the failure of VaR to be a coherent measure in general is because sub-additivity is not satisfied by VaR.
The lack of the coherence property is disturbing, but it is not fatal to the use of VaR. Jorion (2000) counters the critique of Artzner, Delbean, Eber and Heath by stating that the conditions under which lack of coherence may affect the validity of VaR are very rarely met in practice and despite its shortcomings VaR is now the main risk management framework.

4.4 Are Returns Normally Distributed?

The assumption of normally distributed returns allows the derivation of neat equilibrium pricing models and allows the portfolio decision problem to be described by only two parameters, the expected return and the standard deviation of the portfolio return. Similarly in the area of derivatives valuation, both the classic Black and Scholes (1973) option pricing theory and the Merton (1973) model assume that the return distribution of the underlying asset is normal. The problem with these valuation models both for cash as well as for derivative instruments, is that they do not always conform with the empirical evidence which shows conclusively that returns do not follow a normal distribution.
In risk management the assumption of normality also allows the derivation of easy rules for portfolio risk as well as for multi-period risk. The assumption of normality also ensures that popular risk measures such as the standard deviation and VaR are coherent.
However the use of a Gaussian distribution to describe the probability distribution of the equity return series is problematic. First of all, as it can be seen from the density function of the normal distribution, $f(r) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{r^2}{\sigma^2}\right)$ this distribution is suitable only for variables that do not take extreme values. The quantity $\exp\left(-r^2/\sigma^2\right)$ that determines the shape of the distribution, decays so quickly for large values of r that makes the realisation of values a few standard deviations from the mean virtually impossible. A Gaussian variable departs from its more probable value by more than 2σ only 5% of the time, of more than 3σ in 0.2% of the times, whereas a fluctuation of 10σ has a probability of less than 2×10^{-23}. Yet there is ample empirical evidence that suggests that we observe returns of 3 or 4 standard deviation much more often than it is suggested by the normal distribution. One such example is the 1987 stock market crash when the market moved 20 standard deviations in one day. Using data on the aggregate stock market as it is measured by the Athens Stock Exchange General Index we estimated the daily standard deviation o logarithmic returns to be 0.0184. In a sample size was of 3559 daily observations we had 67 observations that were more than 3 standard deviations and 6 observations that were more than 5 standard deviations. The observations exceed 3σ 1.8% of the time instead of the 0.2% of the normal model, that is nine times more than the prediction of the normal model. In addition we had an observation in excess of 5σ once every 2.5 years whereas the normal model predicts one observation every 7250 years.

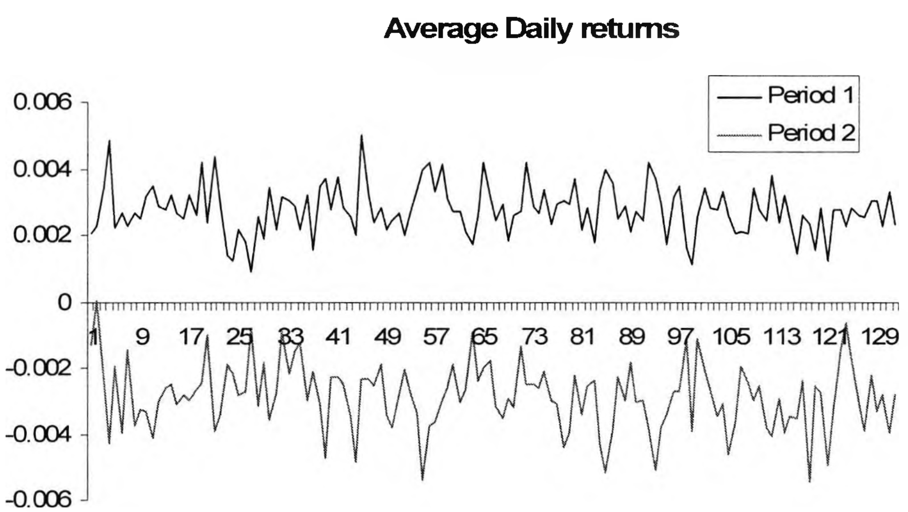
The empirical evidence against the normal distribution is overwhelming (see e.g. Andreou, Pittis and Spanos (2001)). The Normal – VaR, together with estimates of the parameters of the distribution was the original metric suggested by J.P. Morgan RiskMetrics, and has since been since the most widely used risk metric.

As in the case of valuation theories that are based on the assumption of normality, the performance of Normal –VaR will depend on the how close the normality assumption is to the data. If the returns distribution deviates too much from the normal distribution, then the Normal- VaR will produce misleading forecasts of the potential losses from price movements.

The importance or not of the assumption on normality depends on how crucial is the existence of fat tails. For high confidence level say at 99 percent, the importance of fat tails may be greater than say at 95 percent level. Normal-Value-at Risk may therefore be less of a problem in the latter case. Pafka and Kondor (2001) make precisely this point.

The literature on testing the normality assumption is vast. In this section we test to see whether there is evidence of normality for daily weekly and monthly returns of 132 shares traded on the Athens Stock Exchange. This is a sample of companies of those traded and the selection criterion was data availability over the period 2/1/1996-31/3/2003. The Athens Stock exchange experienced a period of rapid growth up until September 1999, which was followed by a collapse in share prices. To allow for the possible different behaviour, the data were split into two periods: 1/1/1996-19/9/1999 and 20/9/1999-31/3/2003.

Figure 4.4: Average Daily Returns



All the empirical results for the individual companies are shown to the Appendix to this chapter. In this section we present summaries of the results. The average daily return of all companies for the two periods is depicted in Figure 4.4. For all the companies the average daily return was positive in the first period and negative in the second.

Looking at the daily standard deviation in the two periods, it is obvious that the volatility of all companies increased in the second period. It is unusual for volatility to increase in falling prices, and an explanation might be in the trading mechanism that was in operation in the two periods.

For the duration of the first period and a few months of the second period, there was a limit up and limit down restriction in daily trading. The increased volatility may be due to the relaxation of this restriction.

Figure 4.6 shows the maximum and minimum return for each period for all the companies which possibly reflects this fact.

Figure 4.5: Daily Volatility

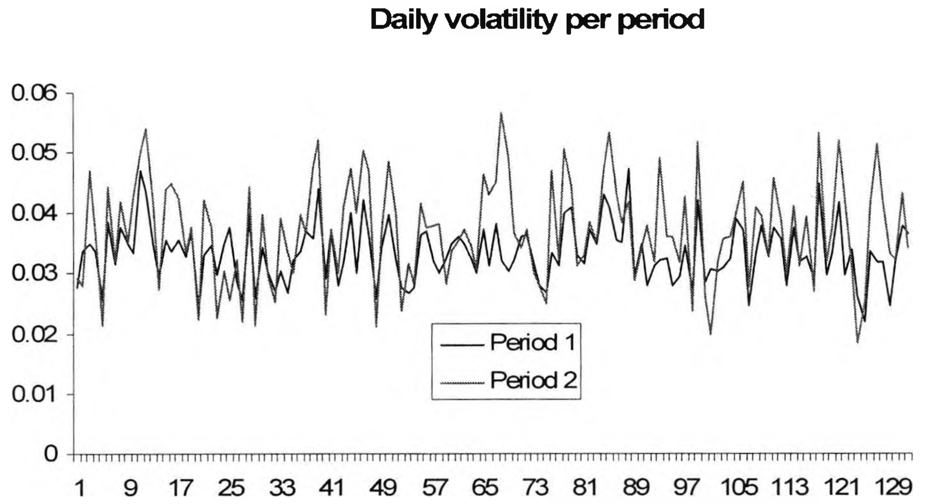
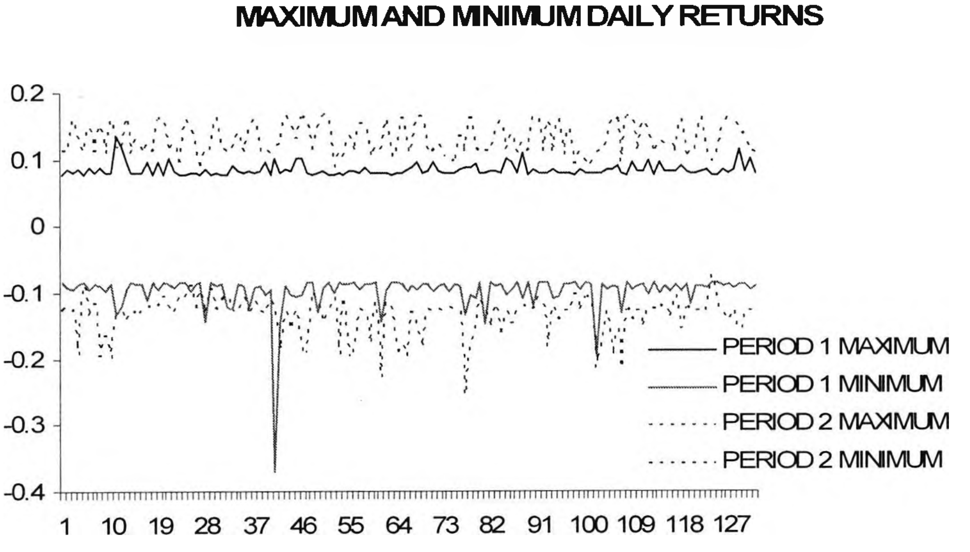


Figure 4.6: Maximum and Minimum Values of Daily Returns



The skewness and kurtosis coefficients are depicted in Figures 4.7 and 4.8. From a first glance of the data it may appear that the degree of skewness has decreased in the second period, whereas the coefficient of kurtosis, has increased. It would appear that the tails of the stocks have become fatter.

It is a common assumption in the literature on risk management that the daily return has a mean value of zero and the drift can therefore be ignored in calculating Value-at-Risk. We have tested this hypothesis for the whole sample and for the two sub periods. In all cases the return are not statistically significant from zero.

Next we test for the existence of skewness. Under the null hypothesis of normality, the skewness statistic is normally distributed with standard errors $se(\gamma_3) = \sqrt{6/T}$ where T is the sample size. The results for skewness show that 60 percent of the stocks are positively skewed, 20 percent are negatively skewed and the rest do not show any skewness.

Figure 4.7:Skewness Coefficient

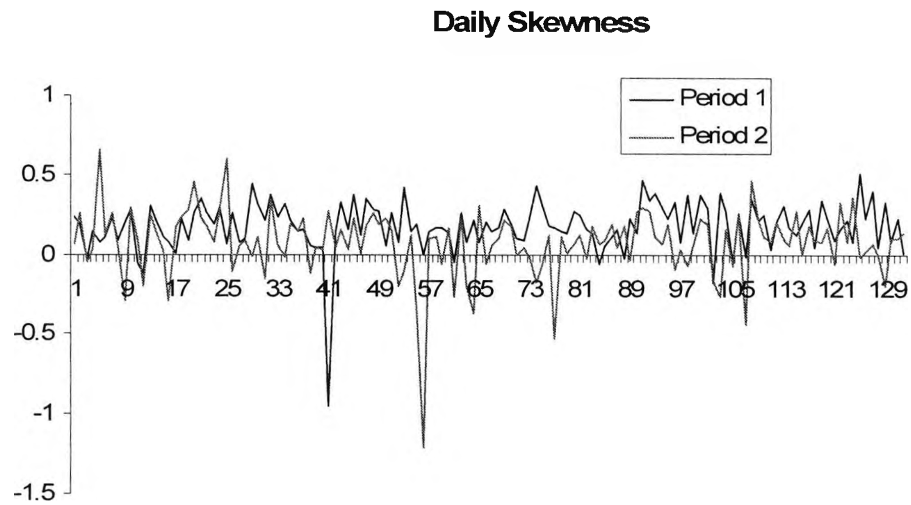
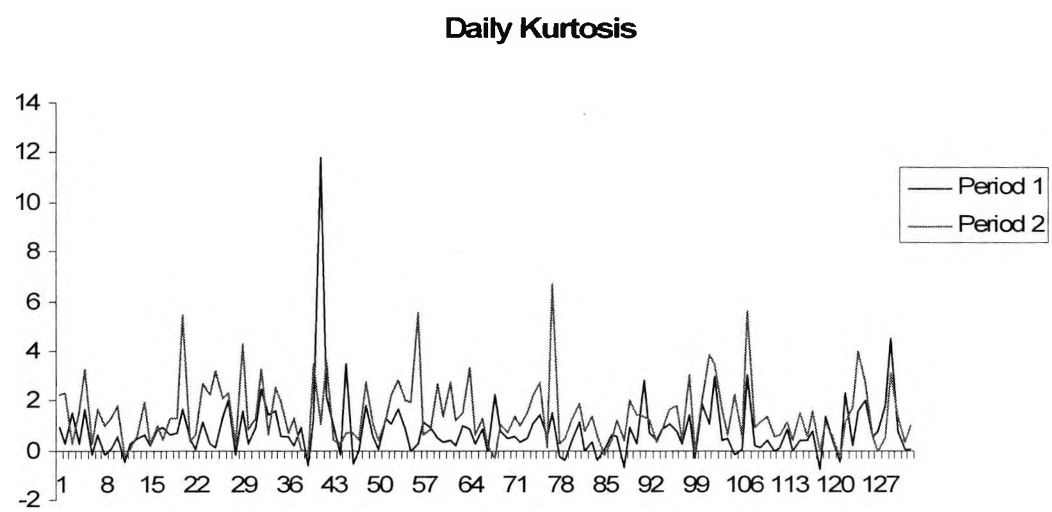


Figure 4.8: Excess Kurtosis Coefficient



It has been suggested that the investment horizon is an important determinant of the statistical properties of equity returns. We have tested therefore the hypothesis that different investment horizons produce different skewness characteristics.

Table 4.1 Summary of Results: Skewness- Whole sample

	Daily	Weekly	Monthly
Positive	63	103	76
Negative	11	1	2
Zero	58	28	54
Total Number of Companies	132	132	132

The results are shown in Table 4.1. The complete results for every company are shown in the Appendix to this chapter. In the table we show the proportion of companies with positive, negative or zero skewness (i.e. statistically insignificant coefficient of skewness) in the sample.

The results show that for daily returns, the companies with positive skewness are very close to the number of companies with negative skewness. Next we have tested for the two sub-periods.

Table 4.2 Summary of Results: Skewness- First Period

	Daily	Weekly	Monthly
Positive	74	114	94
Negative	2	0	1
Zero	56	18	37
Total Number of Companies	132	132	132

Most of the stocks are either positively skewed or there is no skewness at all.

Contrary to the results in the first period, now the majority of stocks do not exhibit any significant degree of skewness at all, whereas the number of stocks with negative skewness, although still a minority, has increased seven fold.

Table 4.3 Summary of Results: Skewness- Second period

	Daily	Weekly	Monthly
Positive	43	47	21
Negative	14	8	6
Zero	75	77	105
Total Number of Companies	132	132	132

Next we look for the whole sample results on kurtosis. The standard error of the coefficient of kurtosis is given by $se(\gamma_4) = \sqrt{24/T}$ where T is the sample size. The result for kurtosis implies that nearly all the r_t series exhibit kurtosis larger than that of the normal distribution. For the daily return series only 10 companies did not show any degree of kurtosis, with the overwhelming majority showing positive excess kurtosis. Only in the case of monthly returns the coefficient of excess kurtosis is zero for a large enough number (40 out of 132).

Table 4.4 Summary of Results: Kurtosis-Whole sample

	Daily	Weekly	Monthly
Positive	119	132	92
Negative	3	0	0
Zero	10	0	40
Total Number of Companies	132	132	132

Comparing the results for the sub-sample we see that whereas the daily and weekly kurtosis does not prevent a great variation between the two sub-periods, in the case of monthly data we have a significant change, with the majority of the stocks exhibiting no kurtosis in the second part of the sample.

Table 4.5 Summary of Results: Kurtosis- First Period

	Daily	Weekly	Monthly
Positive	119	131	77
Negative	3	0	0
Zero	10	1	55
Total Number of Companies	132	132	132

Table 4.6 Summary of Results: Kurtosis- Second Period

	Daily	Weekly	Monthly
Positive	119	124	35
Negative	3	0	0
Zero	10	8	97
Total Number of Companies	132	132	132

The presence of strong kurtosis and extensive skewness in the data means that the normality assumption is not a suitable assumption to describe the equity returns of the companies quoted on the Athens Stock Exchange.

Having tested the significance of Skewness and Kurtosis in the individual stocks we now move to test whether the distribution is normal. We use the Bera-Jarque (BJ) statistic for detecting departures from normality which is calculated using the formula

$$BJ = T(\gamma_3^2 / 6 + \gamma_4^2 / 24).$$

Under the null hypothesis of normality, the BJ statistic is distributed as $\chi^2(2)$ with 2 degrees of freedom. All BJ values are greater than their critical value even at at the one-percent level of 9.21, indicating that the r_t series are non-normal. The results are not affected by splitting the sample into two sub-samples. Normality is still rejected for the two sub-samples. The only exception is in the case of monthly returns were we can reject the assumption of normality for only 36 of the 132 stock returns.

Table 4.7 : Bera –Jaque Normality Test- Number of companies rejecting the hypothesis of normality

	Daily	Weekly	Monthly
Period 1	132	132	93
Period 2	132	132	36
Whole Period	132	132	99

This result, although consistent with other studies that show that normality cannot be rejected for monthly returns, may be affected by the small size of the sample, as we have only 41 monthly observations in the second period.

Given the central role of the assumption of normality in the pricing and measurement of risk, it was important that we tested this hypothesis for the Greek stock market. The tests we have performed have shown conclusively that the distribution is not normal except perhaps in the case of monthly data in the second sub-period of our sample.

4.5 Alternative Probability Models for Return

The parametric approach to VaR assumes that the return process has a specific distribution, usually normal, with the parameters of the distribution, primarily the variance, being estimated either from historical data or from implied option prices.

Although the normal distribution of equity returns is the underlying assumption for valuation models of cash and derivative instruments, there have been many studies that have modelled equity returns on the basis of other probability distributions. These alternative probability distributions include the stable Paretian [Mandelbrot (1963), Fama (1965), Officer (1972), Clark (1973)], the student –t [Praetz (1972), Blattberg and Gonedes (1974), Kon (1984), Gray and French (1990), Peiro (1994), Apparicio and Estrada (2001), Kim and Kon (1994)], the Box-Tiao or power exponential [Hsu(1982), Gray and French (1990), Peiro (1994), Aparicio and Estrada (2001)], the logistic [Smith (1981), Peiro (1994), Aparicio and Estrada (2001)] a discrete mixture of normal distributions [Ball and Torous (1983), Kon (1984), Peiro(1994), Kim and Kon (1994)] and a Poisson mixture of moments [Press (1967), Kim and Kon (1994)]. Several other studies have estimated general models that encompass the specific distribution that we have mentioned so far. These models include Harris and Kucukozmen (2001) who estimated Exponential Generalised Beta distributions and Skewed Generalised –t distributions.

Given the overwhelming rejection of normality as a distribution for daily equity returns, we turn our attention now to two alternative classes of models. The first class of models is the skewed generalised Student-t distribution which is an extension of the symmetric case. The second class of model is the class of stable models. The first of the two classes represent probability distributions with tails heavier than the normal, but not as heavy as the stable distributions.

4.5.1 Student T and Skewed Generalised T-Student

The problem with the normal distribution as we have already explained is that the coefficients of skewness and kurtosis are zero a property that is clearly at odds with the empirical evidence. Kurtosis can be introduced in a simple way if we assume that the equity returns follow the Student-t distribution. The probability density function of the Student-t distribution with degree of freedom ν is given by

$$f(r,\mu,\nu)=\frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi}\Gamma(\nu/2)}\left[1+\frac{(r-\mu)^2}{\nu}\right]^{-\frac{1+\nu}{2}}$$

this is symmetric about μ. It has polynomial tails and the weight of the tails is controlled by the parameter ν. If R has the Student – t distribution then

$$P(R>r)=\text{constant}\times r^{-\nu} \quad \text{as} \quad r\to\infty$$

In contrast if Z has a standard normal distribution then

$$P(Z>r)=\text{constant}\times\frac{e^{-r^2/2}}{r} \quad \text{as} \quad r\to\infty$$

so the tails are qualitatively different, especially for small values of ν as the tails of the distribution decay much more slowly.

Empirical support for modelling univariate returns with a Student – t distribution or Student t-like tails can be found in Blattberg and Gonedes [8], Danielsson and de Vries [12], Hosking et al. [26], Huisman et al. [27], Hurst and Platen [28], Koedijk et al. [33], and Praetz [44]. Application of the t-distribution in the calculation of VaR is found in Goolbergh and Vlaar (1999) who found that whereas the t-distribution performed better than the normal, it nevertheless failed to capture the lower confidence levels. The Value-at-Risk for this distribution can be expressed as

$$VaR = \mu + \sigma F_v^{-1}(p)$$

where σ is the standard deviation of the t-distribution defined as $\sigma = v/(v-2)$. The variance of the distribution is therefore defined only if $v > 2$. In the special case of $v = 1$ we have the Cauchy distribution. Application to the measurement of portfolio risk can be found in Glasserman et all (2000). The square root rule is difficult to apply since the sum of t-distributed random variables is not t-distributed The t-distribution as a candidate for representing the distribution of returns accounts for kurtosis, but it imposes symmetry. In this way it cannot be used as a distribution to explain returns when there is evidence of skewness in the data. Theodosiou (1998) introduced the concept of a skewed t-distribution (SGT) which solves this problem. The probability density function of the SGT distribution is given by

$$f(r/\mu,k,v,\lambda,\sigma^2) = \begin{cases} f_1 & \text{for } r < \mu \\ f_2 & \text{for } r \geq \mu \end{cases}$$

where

$$f_1 = \gamma\left(1 + \frac{k}{v-2}\theta^{-k}(1-\lambda)^{-k}\left|\frac{r-\mu}{\sigma}\right|^k\right)^{-(1+v)/k}$$

$$f_2 = \gamma\left(1 + \frac{k}{v-2}\theta^{-k}(1+\lambda)^{-k}\left|\frac{r-\mu}{\sigma}\right|^k\right)^{-(1+n)/k}$$

$$\gamma = \frac{1}{2\sigma}kB\left(\frac{1}{k},\frac{v}{k}\right)^{-3/2} B\left(\frac{3}{k},\frac{n-2}{k}\right)^{1/2} \varsigma$$

$$\theta = \frac{1}{\varsigma}\left(\frac{k}{v-2}\right)^{1/k} B\left(\frac{1}{k},\frac{v}{k}\right) B\left(\frac{3}{k},\frac{v-2}{k}\right)^{-1/2}$$

$$\varsigma = \left[1 + 3\lambda^2 - 4\lambda^2 B\left(\frac{2}{k},\frac{v-1}{k}\right)^2 B\left(\frac{1}{k},\frac{v}{k}\right)^{-1} B\left(\frac{3}{k},\frac{v-2}{k}\right)^{-1}\right]^{1/2}$$

$B(a,b)$ is the beta function with parameters a,b and with the following restrictions on the value of the parameters.

$$k > 2, v > 0, -1 < \lambda < 1, \sigma > 0$$

The parameter μ determines the location of the ransom variable r, whereas σ is a scale parameter. The parameters κ and v control the height and tails of the density and thus its degree of kurtosis. The

parameter λ determines the skewness of the density. A symmetric distribution will have $\lambda = 0$. The parameter v can be interpreted as the degrees of freedom parameter when $\lambda = 0$ and $k = 2$.
Many distributions are nested as special cases within the SGT. For example when $k > 2, n > 0, -1 < \lambda < 1, \sigma > 0$ we have the generalized --t distribution of McDonald and Newey (1988), when $k = 2$, the skewed t-diststribution of Hansen (1994), when $k = 2, \lambda = 0$ the standard t-distribution with v degrees of freedom, when $k = 2, \lambda = 0, v \to \infty$, the normal distribution, when $k = 2, \lambda = 0, v = 1$ the Cauchy, when $\lambda = 0, v \to \infty$ the power exponential , when $k = 1, \lambda = 0, v \to \infty$, the Laplace and finally when $k \to \infty, \lambda = 0, v \to \infty$ the uniform distribution.
Empirical application of the above distribution include Theodosiou (1998) for US data and Kuzumeglou and Harris (2001), Adcock and Mead (2003) for UK data.
The parameters of the SGT distribution were estimated by maximum likelihood (ML) using the BFGS algorithm with a convergence criterion of 0.0001 applied to the log likelihood function value. The sample mean and standard deviation were used as starting values, whereas starting values for the shape parameters were chosen on the basis of empirical results of other studies (Bookstaber and McDonald (1987), McDonald and Hu (1995), McDonald (1996), Theodosiou (1998) , Kuzumeglou and Harris (2001), and Adcock and Mead (2003)).
One of the problems faced with the estimation of the GST distribution is the fact that the parameters take values over a specific range. To ensure that the parameter values are restricted in that range we have followed Kuzumeglou and Harris (2001) and have used the transformed parameter w^* instead of the original parameter $w = k, v, \lambda, \sigma$ where the transformed parameter is defined as

$$w^* = l + \frac{u - l}{1 + e^{-w}}$$

where u and l are the upper and the lower bounds of the acceptable values for the parameter w. Once the maximisation of the likelihood function has been done with respect to the transformed parameters $w^* = k^*, v^*, \lambda^*, \sigma^*$ and parameter estimates have been obtained the original estimates are obtained form

$$\hat{w} = -\ln\left(\frac{u - l}{\hat{w}^* - l} - 1\right)$$

The empirical results are summarised in Section 4.5.3 with the detailed results for each company in the Appendix.

4.5.2 Stable Distributions

The class of stable distributions were developed by Cauchy (1853) and Levy (1925) and were used by Mandelbrot (1963) to model the distribution of equity returns. Applications of stable models for the modelling of asset returns include the seminal studies of Mandelbrot (1963), Fama (1965), Fama and Roll (1972), Blattberg and Gonedes (1974), and the more recent studies of Embrechts (1977) and Rachev and Mittnik (2000), Kindarhova, Rachev and Scwartz (2001).
Two reasons have been proposed for the use of stable distributions to describe the distribution of asset returns. The first is that the overwhelming empirical evidence shows that the distribution of asset returns

has tails which are fatter that those of the normal distribution and are asymmetric. Stable distributions allow for asymmetry and kurtosis. Thus stable distributions are potentially consistent with the empirical description of asset returns.

The second reason is the Generalised Central Limit Theorem which states that the only possible non-trivial limit of normalised sums of independent identically distributed terms is stable. Since it can be argued that the price of a stock is the sum of independent innovations, a stable model should be used to describe such a model.

A problem with stable distributions is that the density function cannot be written down explicitly, except for three special cases. Thus, instead of using a density function, stable distributions are described through their characteristic function.

Stable distributions can be defined in a number of alternative ways (see e.g. Nolan (1997) Kindarova, Rachev and Schwartz (2001) and Johnson, Kotz and Balakrishnan (1998) for the alternative definitions).

A random variable r is stable only if $r \overset{d}{=} aZ + b$ where $o < \alpha \leq 2$, $-1 < \beta \leq 1$, $a > 0$, $b \in \Re$ and Z is a random variable with characteristic function

$$E\exp(iuZ) = \begin{cases} \exp\left(-|u|^{\alpha}[1 - i\beta \tan\frac{\pi\alpha}{2}(signu)]\right) & \alpha \neq 1 \\ \exp\left(-|u|[1 - i\beta\frac{\pi}{2}(signu)\ln|u|]\right) & \alpha = 1 \end{cases}$$

The sign function is defined as

$$\text{sign}u = \begin{cases} -1 & u < 0 \\ 0 & u = 0 \\ 1 & u > 0 \end{cases}$$

The above definition shows that a stable distribution is identified by four parameters (α, β, a, b) where α is an index of stability or characteristic exponent, β is a skewness parameter (a), is a scale parameter and (b) is a location parameter.

When $\beta = 0$ and $b = 0$ the stable distribution is symmetric around zero and the characteristic function of αZ has the simpler form

$$\phi(u) = \exp\left(-\alpha^{\alpha}|u|^{\alpha}\right)$$

The three special cases when the density function can be written explicitly is when $\alpha = 2$, $\beta = 0$ $a = \sigma^2 / 2$ and $b = \mu$ in which case the stable distribution reduces to the Gaussian density.

$$G(r / \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(r-\mu)^2}{2\sigma^2}\right)$$

The second special case is when $\alpha = 1$, $\beta = 0$ in which case the stable reduces to the Cauchy distribution with density

$$C(x/a,b)=\frac{1}{\pi}\frac{a}{a^2+(x-b)^2}$$

The third special case is when $\alpha = 1/2$, $\beta = 1$ in which case we have the Levy distribution

$$L(r/a,b)=\sqrt{\frac{a}{2\pi}}\frac{a}{(r-b)^{3/2}}\exp\left(-\frac{a}{2(r-b)}\right)$$

Stable distributions with $\alpha < 2$ are leptokurtic and have heavy tails. One consequence of heavy tails is that not all moments exist. In finance, the first moment and variance are normally used to describe the distribution of returns. These are not generally useful for heavy tailed distributions, because the integral expressions for these expectations may diverge. In their place it is sometimes useful to use fractional absolute moments defined as

$$E|X|^p = \int |x|^p f(x)dx$$

where p is any real number. It can be shown (see e.g. Nolan (1997)) that for $\alpha < 2$ $E|X|^p$ is finite for $0 < p < a$ and that $E|X|^p = +\infty$ for $p \geq a$. Thus when $0 < \alpha < 2$, $E|X|^p = EX^2 = +\infty$ and the stable distributions do not have finite second moments or variances. This of course causes tremendous problems to the measurement of risk as it invalidates the variance of asset returns as a measure.

When $\alpha < 2$, Levy (1925) has shown that the tails of a non-Gaussian stable distribution is asymptotically equivalent to a Pareto law. Specifically, if X is a standardised stable random variable with characteristic exponent $\alpha < 2$ and skewness parameter β, then as $x \to \infty$[2]

$$P(X > x) \sim (1+\beta)C_\alpha x^{-\alpha}$$

where

$$C_\alpha = \frac{1}{\pi}\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)$$

Note that the lower the value of $\alpha < 2$, the slower the decay of the tails. Thus the class of stable distributions present a flexible class which is also theoretically plausible.

4.5.3 Empirical Results

The two classes of probability distributions were fitted to the 132 share prices of the companies quoted on the Athens Stock Exchange. The estimation period was for the period 2/1/1996-31/3/2003 and we used a total of 1790 daily observations. A maximum likelihood approach was used in both cases.[3]

We employed three statistical criteria to choose the appropriate distribution for each stock return. The first one is the Schwartz Bayesian Criterion which has been explained in Chapter 3. The criterion penalises for the different number of parameters that a model has and it is given by

---

[2] Property 1.2.15 Samorodnitsky and Taqqu (1994)

$$SBC = \ln L - 0.5 \times m \times \ln T$$

where m is the number of parameters to be estimated, $\ln L$ is the logarithm of the likelihood function and T is the number of observations. The assumption is that the more parameters a model have the better the fit, but this could be spurious.

The second criterion is the Kolmogorov distance

$$KD = 100 \times \sup_{r \in \Re} \left| F_s(r) - \hat{F}(r) \right|$$

where $\hat{F}(r)$ denoted the cumulative distribution function of the estimated parametric density and $F_s(r)$ is the empirical sample distribution., i.e. $F_s(r) = \frac{1}{T}\sum_{t=1}^{T} I_t(-\infty, r)r_t$, where $I_t(-\infty, r)$ is the indicator function. The properties of this statistic is discussed in DeGroot(1986) and D'Agostino and Stephens (1986). It is a robust measure in the sense that it focuses only on the maximum deviation between the sample and fitted distributions.

The third measure is the Anderson –Darling statistic (Anderson and Darling (1952)), which weights the absolute deviations $\left| F_s(r) - \hat{F}(r) \right|$ by the reciprocal of the standard deviation of $F_s(r)$, $\sqrt{\hat{F}(r)\left(1-\hat{F}(r)\right)}$, i.e.

$$AD = \sup_{r \in \Re} \frac{\left| F_s(r) - \hat{F}(r) \right|}{\sqrt{\hat{F}(r)\left(1-\hat{F}(r)\right)}}$$

The use of this statistic allows discrepancies in the tails of the distribution to be appropriately weighted. On the other hand, KD emphasizes deviations around the median of the fitted distribution.

The results for each individually company are shown in the Appendix to this Chapter. In Table 4.8 we summarize the results of our distribution fitting. In the second column we show the companies that had the highest SBC for each distribution. Thus the skewed distribution produced a higher SBC in 112 of the 132 companies. According to this criterion, 85 percent of the companies support the skewed –t distribution. In column 3 and 4 we present the results from the other two criteria. According to the KD criterion 105 stocks had a smaller distance for the skewed –t distribution so that 80 percent of the stocks seem to have a distribution closer to the skewed-t distribution. Similar results are obtained for the Anderson Darling test.

Table 4.8: Statistics for alternative distributions

Distribution	SBC	KD	AD
Skewed T	112	105	107
Stable	20	27	25

The empirical results of this section show that the GST seems to be a better description of the distribution of returns although the stable family of distributions fits well at least some of the share prices (up to 20 %

[3] In the case of the stable distribution we used the programme Stable kindly supplied by Professor John Nolan of the American University, Washington. The algorithm for the case of the GST distribution was provided by Professor Elias Dinenis of Cass Business School, City University, London.

of the stock returns). In implementing a risk management system one is therefore left with an inconclusive advise as to what is the correct model to use.

The empirical unconditional distribution reflects the statistical properties of the returns and can be used to specify the risk associated with the underlying asset. In practice, one is typically more interested in a conditional risk assessment. However the unconditional sample distribution is of interest because any dynamic model used for that purpose has to be compatible with the unconditional distributions of the data at hand.

4.6 Modelling Mean and Volatility Dynamics of Equity Returns

The parametric approach to estimating Value-at-Risk requires the estimation of the moments of the distribution in order to have a full characterization of the probability density function and therefore of the quantiles of the distribution. Even in the simplest of the probability distributions we have examined, the normal distribution we need to estimate the mean and the volatility of returns. There is general agreement that the mean of the daily returns is not statistically different from zero, and so the only parameter that needs to be estimated in the Normal – VaR model is the volatility.

There are two general approaches to estimating the volatility. The first approach is to assume that the volatility is constant. This is the standard assumption for the Brownian model of asset prices and has been used extensively not only for risk measurement but in valuation as well (Black and Scholes (1973)). The second approach is to assume that the volatility is not constant but it evolves according to a particular process. There are two classes of non-constant volatility models; (a) the stochastic volatility model and (b) the class of general autoregressive conditionally heteroskedastic (GARCH) models which we shall review in this section.

As we have seen in the discussion of the stable processes that fat tails entail the non-existence of moments. Because of the undesirable assumption of the infinite variance in non-Gaussian stable models, several authors have proposed alternative models for observed heavy tailed and skewed data sets. These models include ARCH and GARCH models with normal innovations, mixture models etc. Thus the family of ARCH-GARCH model introduced by Engel (1982) and Bollerslev (1986) apart from providing estimates of the volatility parameter, represent another attempt to introduce kurtosis and fat tails into the distribution of asset returns. In doing so they exploit a stylized fact, namely the fact that there are volatility clusters which are not captured by the static models we examined far. The variance is of course one measure of spread of a distribution and it is not appropriate for all problems. From a risk measurement point of view what is more important is a model that captures the shape of the distribution of asset returns.

In the GARCH (1,1) class of models we assume that the return is given by

$$r_t = \mu_t + \varepsilon_t$$

where $\varepsilon_t = \sigma_t z_t$ with

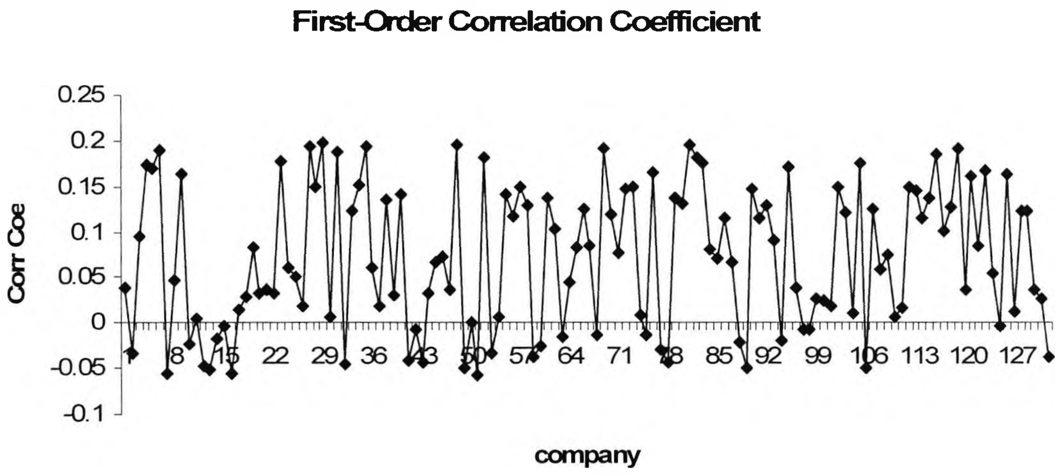
$$z_t / \Psi_{t-1} \sim t(0,1,v)$$

$$\mu_t = \phi_0 + \phi_1 r_{t-1}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

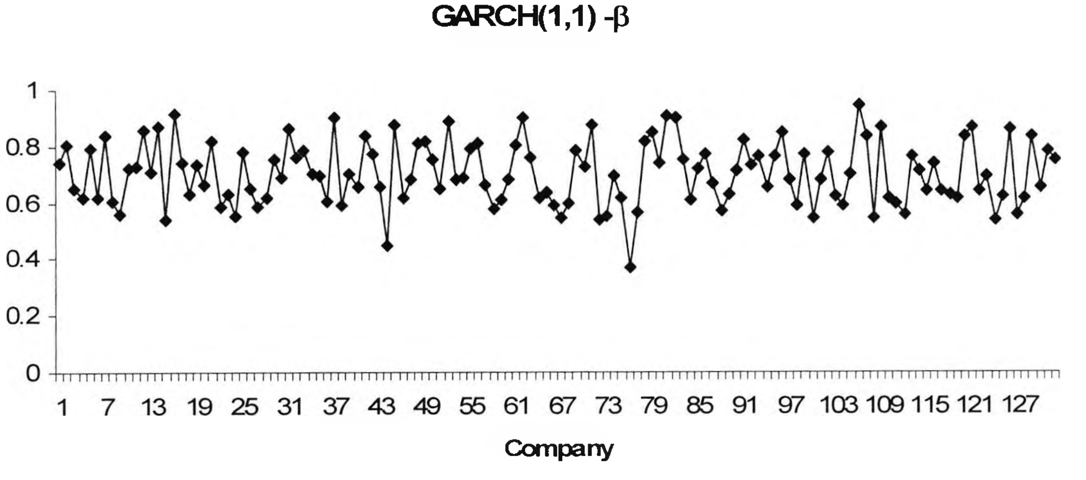
The empirical results for the fitting of the T-GARCH(1,1)[4] model are presented in the form of a series of graphs which summarise the parameter values for each share.

Figure 4.9 Autocorrelation Coefficient



Starting with the estimation of the conditional mean model we have found the about 40 stocks exhibit significant positive correlation whereas the rest showed no significant first order correlation. Figure 4.9 shows the values of the individual first order coefficients. We estimated higher order autoregressive models but in no case we found significant autoregressive coefficients.

Figure 4.10 ß Coefficient



Looking now at the parameters of the volatility process we report the values of the β coefficient that shows the speed of adjustment in the volatility

A test of how well the dynamic model has fitted the data is to look at the standardised residuals defined as

$$z_t = \frac{r_t - \hat{\mu}_t}{\hat{\sigma}_t}$$

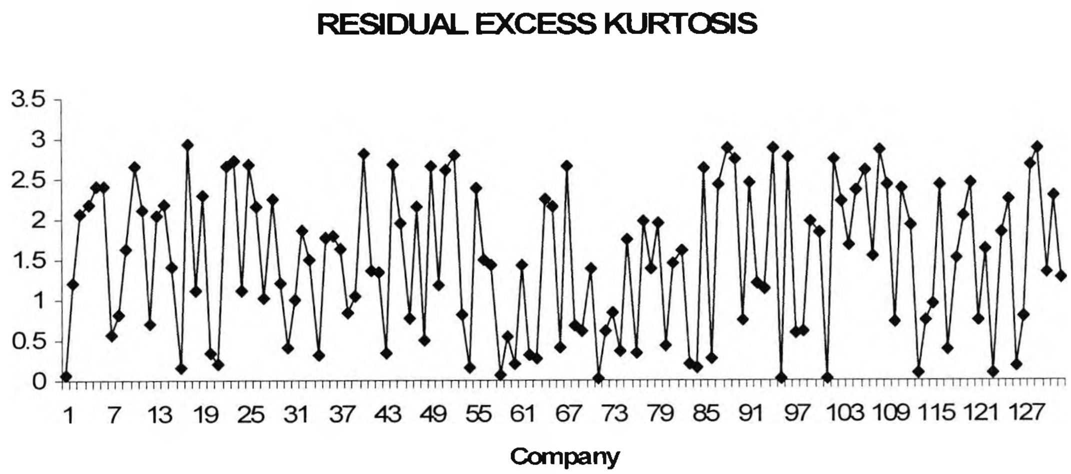
If the model fits the data well then the residual should exhibit no serial correlation either in the level or their square values. It should also eliminate kurtosis if the reason for the kurtosis was volatility clustering.

[4] The estimation of the T-GARCH(1,1) model was done using the software package PC GIVE.

Looking at the standardised residuals we see that the application of a T-GARCH(1,1) model has not eliminated the kurtosis in the data. The data still exhibit smaller than the unconditional but nevertheless statistically significant excess kurtosis in the majority of the companies.

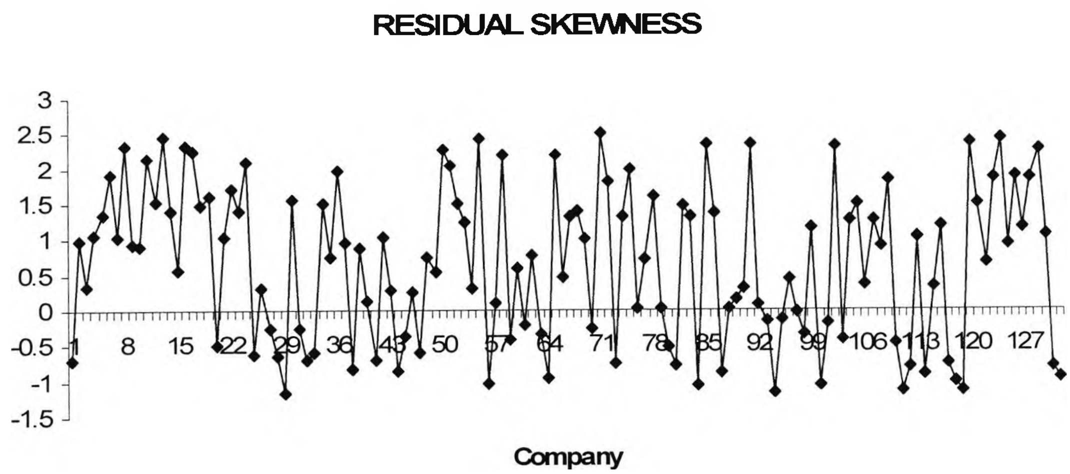
The results on skewness are not surprising given the fact that the innovation distribution we have used in the GARCH model is not an asymmetric one. Thus the coefficient of skewness for about 20 percent of the companies remains.

Figure 4.11 Residual Excess Kurtosis



The empirical results from the modelling of the return dynamics can be summarised as follows. A significant number of stocks exhibit first order serial correlation. No stock had conditional mean dynamics of higher order. The existence of serial correlation creates problems for time aggregation and invalidates the square root rule. With respect to the modelling of volatility, we have used a GARCH(1,1) model with innovation generated from a symmetric t-distribution. The resulting residuals however exhibit skewness and kurtosis for about 20 percent of the companies. For those companies, fat tails are not the result of volatility clustering.

Figure 4.12 Residual Skewness



In terms of modelling the volatility dynamics it also means that more sophisticated models that incorporate asymmetries and fat tails should be considered for the modelling of innovations.

4.7 Direct Estimation of Tail Probabilities

The approaches we have reviewed so far were based on some explicit assumptions about the probability distribution of the returns. In this section and the next, we are assessing two approaches that can be considered as distribution-free approaches.

The statistical distributions reviewed in the previous sections attempt to introduce fat tails and skewness in the distribution of returns to provide a more realistic representation of the equity return probability distribution. The potential problem with this approach is that by specifying an alternative characterisation for the entire probability distribution we may be fitting well to the data parts of the distribution we are not interested in. More specifically the goodness of fit criterion may be influenced by a good fit of the middle part of the distribution and by a poor fit of the tails. Since we are interested in the estimation of tail quantiles, we need to pay more attention to the tails of the distribution.

The extreme value theory concentrates exclusively on the tails of a probability distribution and ignores the central part of the distribution. This method will be reviewed in the following chapter.

Let $\{R_1,\dots R_n\}$ be the logarithmic daily returns observed on days $\{1,2,\dots n\}$ with density function f_R and cumulative distribution function F_R. As we have explained, the basic idea is to study the tail of a distribution, that is to study the statistical properties of observations that come from the tail of distribution. These observations are called extreme observations and the purpose of this section is to describe the behaviour of these extreme observations. Extreme observations are defined in two ways, both of which are related. In the classical extreme value theory, extremes are defined as the maxima and minima of non overlapping samples of n random variables. For example we may split the sample of daily observations into blocks of five observations and define the maximum or minimum weekly return. Then the set of extreme observations is made up of the weekly maximum or minimum values. Alternatively one may define as extreme values the maximum or minimum values of monthly observations. In other areas, such in meteorology where data are available for many years, the set of extreme observations is made up of annual maximum or minimum values. In the second approach extreme observations are classified as those observations which are above or below a certain value, which is called the threshold value.

4.7.1 Block Method

Let $R_{\max,n}$ denote the highest daily return (the maximum) observed over n trading days that is $R_{\max,n} = Max(R_1,\dots R_n)$. Similarly let $R_{\min,n}$ denote the lowest daily return (the minimum) observed over n trading days defined as $R_{\min,n} = Min(R_1,\dots R_n)$. Note that $R_{\min,n} = Min(R_1,\dots R_m) = -Max(-R_1,\dots,-R_m)$ so that one only needs analysing the behaviour of the one of the extreme statistics.

If the random variables $\{R_1,\ldots R_n\}$ are statistically independent and drawn from the same distribution, then the probability distribution of the maximum and minimum $R_{\max,n}$ and $R_{\min,n}$ are a function of F_R, the parent distribution, and the length of the selection period *n*.

$$\begin{aligned} F_{R_{\max,n}}(r) &= P(R_{\max,n} \le r) = P[Max(R_1,\ldots R_n) \le r] \\ &= P[R_1 \le r,\ldots,R_n \le r] \\ &= P[R_1 \le r]\times P[R_2 \le r]\times\ldots\times P[R_n \le r] \\ &= [F_R(r)]^n \end{aligned} \quad (4.7.1)$$

To calculate the distribution function for the minima we first show that

$$\begin{aligned} P(R_{\min,n} > r) &= P[Min(R_1,\ldots R_n) > r] \\ &= P[R_1 > r,\ldots,R_n > r] \\ &= [1-P(R_1 \le r)]\times\ldots\times[1-P(R_n \le r)] \\ &= [1-F_R(r)]^n \end{aligned}$$

The distribution function of the minimum is then easily derived and is given below

$$F_{R_{\min,n}}(r) = P(R_{\min,n} \le r) = 1 - P(R_{\min,n} > r) = 1-[1-F_R(r)]^n \quad (4.7.2)$$

It follows therefore that if the distribution of returns is known, then the distribution of the extreme values is also known. However there are many cases in which the distribution of returns is not known and inference based on Equations (4.7.1) and (4.7.2) is not possible. One possibility is to estimate the distribution $F_R(r)$ form the data and then use Equations (4.7.1) and (4.7.2) to calculate the distribution of the minima or maxima. However as Coles (2001) notes, such an approach may lead to unreliable results since, small discrepancies in the estimation of $F_R(r)$ may lead to large discrepancies in the estimation of $[F_R(r)]^n$.

The second approach is to derive the asymptotic distribution of $[F_R(r)]^n$ as the sample size *n* increases and derive the limiting distribution which then we can use to describe the probability function of maxima or minima. Unfortunately the limit distribution of the extreme values as derived above is degenerate because as the sample size increases the limit becomes 0 or 1 which is shown below

$$\lim_{n\to\infty} F_{R_{\max,n}}(r) = 1 \text{ if } F_R(r) = 1 \quad (4.7.3)$$

and

$$\lim_{n\to\infty} F_{R_{\max,n}}(r) = 0 \text{ if } F_R(r) < 1 \quad (4.7.4)$$

In order to find a non-degenerate limiting distribution the maximum is reduced by a scale parameter and a location parameter so that we are finding the asymptotic distribution of the standardised maximum

$$R^*_{\max,n} = \frac{R_{\max,n} - b_n}{a_n}$$

The distribution function $R^*_{\max,n}$ is given below

$$F_{R^*_{\max,n}}(r) = \mathrm{P}(R^*_{\max,n} < r) = P\left(\frac{R_{\max,n} - b_n}{a_n} < r\right) = \left[F(ra_n + b_n)\right]^n \qquad (4.7.5)$$

This distribution function is not degenerate, but it converges to a function $G_{\max}(r)$ as $n \to \infty$. This result is based on the assumption that there numbers b_n and a_n such that

$$\lim_{n\to\infty}\left[F(ra_n + b_n)\right]^n = G_{\max}(r) \qquad (4.7.6)$$

An example will highlight the derivation of the limiting distribution. Suppose that the random variables $\{R_1, \ldots R_n\}$ follow an exponential distribution with distribution function

$$F_R(r) = 1 - e^{-r}$$

In this case we have

$$F_{R_{\max,n}}(r) = P(R_{\max,n} \le r) = [F_R(r)]^n = [1 - e^{-r}]^n$$

Selecting $a_n = 1$ and $b_n = \ln n$ we can find the limiting distribution of the standardised maximum $R^*_{\max,n}$

$$\mathrm{P}(R^*_{\max,n} < r) = \left[F(ra_n + b_n)\right]^n = \left[F(r + \ln n)\right]^n$$
$$= \left[1 - e^{-(r+\ln n)}\right]^n = \left[1 - \frac{1}{n}e^{-r}\right]^n$$

The limiting distribution of above function as $n \to \infty$ is $G_{\max}(r) = \exp(-e^{-r})$.

It should be clear from the above example that the limiting distribution depends on the distribution function $F_R(r)$. It could be shown for example that for the distribution function $F_R(r) = \exp(-1/r)$ for $r > 0$ and with $a_n = n, b_n = 0$, we have $G(r) = e^{-1/r}$. Similarly for $F_R(r) = r$, that is the distribution function of a uniform distribution and and with $a_n = 1, b_n = \ln n$ we have limiting function $G(r) = e^r$

The above analysis may give the impression that the limiting functions $G(r)$ could take a variety of forms. However Fisher and Tippet (1928) and Gnedenko (1943) showed that if there is a non degenerate limiting distribution G for $R^{*}_{\max,n}$ then this limiting distribution will belong to one of the three members of the family of extreme value distributions known as

Type I or Gumbel distribution

$$G_{\max}(r) = \exp\left(-\exp\left(-\frac{r-b}{a}\right)\right) \tag{4.7.7}$$

Type II or Fréchet distribution:

$$G_{\max,\xi}(r) = \begin{cases} 0 & r \le b \\ \exp\left(-\left(\frac{r-b}{a}\right)^{-1/\xi}\right) & r > b, \xi > 0, a > 0 \end{cases} \tag{4.7.8}$$

Type III or Weibull distribution:

$$G_{\max,\xi}(r) = \begin{cases} \exp\left\{-\left[-\left(\frac{r-b}{a}\right)^{1/\xi}\right]\right\} & r < b, \xi > 0, a > 0 \\ 1 & r \ge b \end{cases} \tag{4.7.9}$$

Each family of distribution functions is defined by the location parameter b the scale parameter a and the shape parameter ξ. Note that the shape parameter in the Gumbel distribution is equal to zero, i.e. $\xi = 0$. The location parameter b is the left endpoint of the Fréchet distribution function and the right endpoint of the Weibull distribution function.

In order to understand the behaviour of the limiting distributions better, we study the properties of the density functions which correspond to the three distribution functions.

Given that we have three ways of representing the limiting distribution of extremes, the question that arises is how to determine which one is the appropriate distribution for a particular sample of data. One way to choose the limiting distribution is to look at the statistical properties of returns. When the distribution function of the original series is known, then it can be used to determine the type of limiting extreme-value distribution. The exponential distribution, the normal distribution, discrete mixtures of normal distributions and mixed diffusion processes lead to a Type I limiting distribution, because all moments exist. On the other hand, a Student-t distribution obeys a Type II distribution with shape parameter equal to the degrees of freedom (Mood et all, (1974)). An ARCH process where the conditional variance changes over time also satisfies equation (4.7.8) as does a heavy – tailed stable distribution introduced by Mandelbrot (1963), with the shape parameter ξ in this cases equal to the characteristic exponent.

Jansen and de Vries (1991) argue that the Type I limiting distributions do not adequately describe the tails of the returns because the tails are too thin compared with the tails of the empirical distributions . The Type II distribution on the other hand has fat tails and is more consistent with the underlying distribution functions.

When we are unsure about the true probability distribution of returns, then we can use the generalised extreme value (GEV) derived by von Mises (1936) and Jenkinson (1955) which encompasses the three limiting distributions as special cases.

The cumulative distribution function of the generalized extreme value distribution is given by

$$G(r) = \exp\left\{-\left[1+\xi\left(\frac{r-b}{a}\right)\right]^{-1/\xi}\right\} \qquad (4.7.10)$$

The three extreme value distributions are derived as special cases for different values of the shape parameter ξ. The Gumbel distribution corresponds to the case $\xi \to 0$, with the distribution function being derived using the fact that $(1+\xi r)^{-1/\xi} \to \exp(r)$ as $\xi \to 0$. The Fréchet distribution corresponds to the case when $\xi > 0$ whereas the Weibull distribution corresponds to the case $\xi < 0$.

The extreme value distributions derived so far are very useful for the estimation of tail probabilities and quantiles. Since we have an explicit distribution function for the extreme observations in the form of GEV in equation (4.7.10) we can calculate any quantile of the distribution of extreme values. Suppose, for instance, we want to find the value r_p, such that this value is exceed by $R^*_{\max,n}$ with a given probability p, i.e.

$$\mathrm{P}(R^*_{\max,n} > r_p) = p$$

The above probability can be expressed in term of the generalised extreme value distribution as

$$1 - G(r_p) = p$$

Using equation (4.6.10) we have

$$1 - p = \exp\left\{-\left[1+\xi\left(\frac{r_p-b}{a}\right)\right]^{-1/\xi}\right\}$$

Solving the above equation yields the pth quantile

$$r_p = b + \frac{a}{\xi}\left\{[-\ln(1-p)]^{-\xi} - 1\right\} \qquad \text{for } \xi \neq 0$$

$$r_p = b - a\ln[\ln(1-p)] \qquad \text{for } \xi = 0$$

So if we have estimates of the parameters b, a and ξ we can estimate the quantiles of the generalised extreme value distribution and thus the Value - at – Risk corresponding to probability level p.

4.7.2 Peaks over Threshold Method

In practice modelling all block maxima is wasteful if other data on extreme values are available. Therefore a more efficient approach is to model the behaviour of extreme values above a high threshold. This method known as the peaks over threshold (POT) is easier to estimate and easier to calculate VaR estimates.

Consider again a sample of n iid random $\{R_1, \ldots R_n\}$ with a common distribution function F whose functional form is unknown. In this approach we define extreme observations as those observations which are larger than a value u. That defines a subset of n, say k observations which are characterised as extremes and whose distribution we aim to establish.

The conditional distribution function of those observations R which lie above u is give by

$$F_u(r) = P(R \leq r / R > u) = \frac{P(R \leq r, R > u)}{P(R > u)} = \frac{F(r) - F(u)}{1 - F(u)}, \qquad r \geq 0$$

If we formulate the problem in terms of the excesses $R_i - u$ then the pertaining excess distribution function at u s given by

$$F_u(r) = P(R - u \leq r / R > u) = \frac{F(r+u) - F(u)}{1 - F(u)}, \qquad r > 0$$

Pickands (1975) has shown that if F is in the maximum domain of attraction of G then as u approaches the end value of the distribution (normally infinitive) then the conditional probability function converges asymptotically to a function H

$$\lim_{u \to \infty}\left[F_u(r)\right] = H(r)$$

Where the function $H(r)$ is given by

$$H(r) = \begin{cases} 1 - \left(1 + \dfrac{\xi}{\sigma + \xi(u - \mu)} r\right)^{-\frac{1}{\xi}} & \xi \neq 0 \\ 1 - \exp\left(-\dfrac{1}{\sigma} r\right) & \xi = 0 \end{cases}$$

where $\sigma + \xi(u - \mu) > 0, r \geq 0$ when $\xi \geq 0$ and $0 \leq r \leq -(\sigma + \xi(u - \mu))/\xi$ when $\xi < 0$.

The function $H(r)$ is known as the Generalised Pareto Distribution (GPD). A simple proof of the relationship between the GEV and the Generalised Pareto Distribution is given below (Coles (2001)).

$$[F_R(r)]^n \approx \exp\left\{-\left[1+\xi\left(\frac{r-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

or

$$n\ln[F_R(r)] \approx -\left[1+\xi\left(\frac{r-\mu}{\sigma}\right)\right]^{-1/\xi}$$

For large values of r, a Taylor expansion implies that

$$1-F_R(u)=\frac{1}{n}\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}$$

for large u. Similarly for r > 0.

$$1-F_R(u+r)=\frac{1}{n}\left[1+\xi\left(\frac{u+r-\mu}{\sigma}\right)\right]^{-1/\xi}$$

Hence

$$P(R>r+u/R>u)\approx\frac{n^{-1}\left[1+\xi(u+r-\mu)/\sigma\right]^{-1/\xi}}{n^{-1}\left[1+\xi(u-\mu)/\sigma\right]^{-1/\xi}}\frac{1}{2}$$

$$=\left[1+\frac{\xi(u+r-\mu)/\sigma}{\xi(u-\mu)/\sigma}\right]^{-1/\xi}$$

$$=\left[1+\frac{\xi}{\sigma+\xi(u-\mu)}r\right]^{-1/\xi}$$

The above relationship shows that whether we estimate tail probabilities based on the GEV or the GPD should produce the same results. Indeed for a given value of u the parameters ξ, μ, σ of the GEV distribution determine the parameters of the GPD. In particular the shape parameter ξ is independent of u and it is the same for both GEV and GPD distributions. Choosing a different, but still large, block size *n*, would affect the values of the GEV parameters but not those of the corresponding Generalised Pareto Distribution of threshold excesses.

The Generalised Pareto distribution contains a number of special cases. If $\xi \geq 0$, F is the Fréchet family and H is a Pareto distribution

$$H_p(r)=1-cr^{-1/\xi}$$

If $\xi = 0$, F is in the Gumbel family and H is an exponential distribution

$$H(r) = 1 - \exp\left(\frac{-r}{\sigma}\right)$$

If $\xi < 0$, F is in the Weibull family and H is a Pareto type II distribution.

In most applications of risk management in finance, the data comes from a heavy-tailed distribution, so that $\xi \geq 0$.

In order to estimate the tails of the loss distribution, we use the conditional distribution function which we can write now as

$$F(r+u) = F_u(r)[1 - F(u)] + F(u)$$

or after using the fact that $F_u(r) \approx H(r)$ from above theorem

$$F(r+u) = H(r)[1 - F(u)] + F(u)$$

The function $F(u)$ can be estimated non-parametrically using the empirical distribution function

$$\hat{F}(u) = \frac{n-k}{n}$$

where k represents the number of exceedances over the threshold u. After some manipulation we have

$$\begin{aligned} \hat{F}(u+r) &= \frac{k}{n}\left[1 - \left(1 + \frac{\xi}{\beta} r\right)^{-1/\xi}\right] + \frac{n-k}{n} \\ &= \frac{k}{n} - \frac{k}{n}\left(1 + \frac{\xi}{\beta} r\right)^{-1/\xi} + 1 - \frac{k}{n} \\ &= 1 - \frac{k}{n}\left(1 + \frac{\xi}{\beta} r\right)^{-1/\xi} \end{aligned}$$

where $\beta = \sigma + \xi(u - \mu)$

Denoting $x = u + r$ we have

$$\hat{F}(x) = 1 - \frac{k}{n}\left(1 + \frac{\xi}{\beta}(x-u)\right)^{-1/\xi}$$

and

$$1 - \hat{F}(x) = \frac{k}{n}\left(1 + \frac{\xi}{\beta}(x-u)\right)^{-1/\xi}$$

The Value-at-Risk is computed by first setting $1 - \hat{F}(x) = p$ and then solving for the unknown value of $x_p = \hat{F}^{-1}(p)$ that corresponds to probability level p.

$$x_p = u + \frac{\beta}{\xi}\left[\left(\frac{p}{k/n}\right)^{-\xi} - 1\right]$$

The Value-at-Risk $VaR(p) = x_p$ calculated in this way uses only extreme values and given that it is derived from a description of the tails only it should provide a more accurate estimate of the true Value-at-Risk. Most importantly the functional form of the distribution function is not ad hoc but it is based on the asymptotic theory of extreme values. As in all cases where asymptotic theory is invoked, the results hold when only when the assumptions on which the asymptotic results are obtained also hold. The crucial assumption on which the asymptotic results are based is that the random variables $\{R_1,... R_n\}$ are iid.

4.7.3 Estimation of the Generalised Extreme Value Function

Maximum likelihood estimators of the parameters of Equation () have been discussed by a number of authors including Jenkinson (1969), Prescott and Walden (1980,1983), Hosking (1985) and Macleod (1989). Hosking, Wallis and Wood (1985) have discussed the method of probability-weighted moments for the estimation of parameters. The estimation of the Gumbel distribution is covered extensively in Gumbel (1958) and Galambos (1978). A review of these methods is presented in Johnson, Kotz and Balakhrisnan (1995). Goodness of fit tests for the generalised extreme value distribution have been examined by Chowdhury, Stedinger and Lu (1991).

4.7.4 Estimation of the Generalised Pareto Distribution

There are a number of ways for the estimation of the parameters of the generalised Pareto distribution. The estimation of the tail parameter in the threshold model is relatively straightforward, but it depends on the specification of the threshold above or bellow which we classify observations as extreme.

The first one is based on maximum likelihood principles. The density function of the generalised Pareto distribution can be derived by differentiation of the distribution function with respect to r. This produces the following density function.

$$h(r) = \frac{\partial H(r)}{\partial r} = \frac{1}{\sigma + \xi(u-\mu)}\left(1 + \frac{\xi}{\sigma + \xi(u-\mu)} r\right)^{-\frac{1}{\xi}-1} = \frac{1}{\beta}\left(1 + \frac{\xi}{\beta} r\right)^{-\frac{1}{\xi}-1}$$

where $\beta = \sigma + \xi(u - \mu)$. Once the density function has been derived, we can write down the likelihood function for the sample which has the form

$$l(\beta,\xi) = -k \ln \beta - \left(1 + \frac{1}{\xi}\right)\sum_{i=1}^{k} \ln\left(1 + \frac{\xi r_i}{\beta}\right)$$

The MLE of the parameters are obtained from the first order conditions

$$\frac{\partial l(\beta,\xi)}{\partial \beta} = 0 \text{ and } \frac{\partial l(\beta,\xi)}{\partial \xi} = 0$$

The parameters of the likelihood function can be derived using numerical optimisation. The asymptotic properties of the MLE of the GPD parameters have been derived by Smith (1987) and Joe (1987) who showed that the asymptotic distribution of parameter estimates is bivariate normal, i.e.

$$\sqrt{k}\begin{bmatrix} \hat{\beta}/\beta - 1 \\ \hat{\xi} - \xi \end{bmatrix} \to N\left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Omega\right\} \text{ with } \Omega = \begin{bmatrix} (1+\xi)^2 & -(1+\xi) \\ -(1+\xi) & 2(1+\xi) \end{bmatrix}$$

The asymptotic distribution can be used to construct confidence intervals which can in turn be used to evaluate hypotheses. For example the 95% confidence intervals for the parameters β and ξ are obtained as $\hat{\beta} \pm 1.96\sqrt{\frac{2(1+\hat{\xi})\hat{\beta}^2}{k}}$ and

$\hat{\xi} \pm 1.96\sqrt{\frac{(1+\hat{\xi})^2}{k}}$ respectively.

In the case $\xi = 0$ the log-likelihood is

$$l(\beta) = -k \ln \beta - \frac{1}{\beta}\sum_{i=1}^{k} r_i$$

in the latter case the estimate of β is given by $\hat{\beta} = \frac{1}{k}\sum_{i=1}^{k} r_i$.

A second approach is the non-parametric estimator of the tail suggested by Pickands (1975, p.121) and which is given by

$$\hat{\xi} = -\frac{1}{\ln 2}\ln\left(\frac{r_{k/2:k}}{r_{3k/4:k} - r_{k/2:k}}\right)$$

where $r_{k/2:k}$ and $r_{3k/4:k}$ are order statistics in sample of size k. The normalized Picands's statistic defined as

$$(\hat{\xi} - \xi)\sqrt{k} \to N(0, \zeta^2)$$

where

$$\hat{\zeta}^2 = \hat{\xi}^2(2^{-2\hat{\xi}+1} + 1)/[2(2^{-\hat{\xi}} - 1)\ln 2]^2$$

A third method is the method of moments (Castillo and Hadi (1997) which produces parameter estimates given by

$$\hat{\xi}_{MOM} = (\bar{r}^2 / s^2 - 1)/2$$

$$\hat{\beta}_{MOM} = \bar{r}(\bar{r}^2 / s^2 + 1)/2$$

where $\bar{r}$ and s^2 are the sample mean and the sample variance.

Finally Hoskins and Wallis (1987) have suggested the method of probability –weighted moments (PWM). The PWM estimates of the parameters are

$$\hat{\xi}_{PWM} = \bar{r}/(\bar{r} - 2t) - 2$$

and

$$\hat{\beta}_{PWM} = 2\bar{r}t/(\bar{r} - 2t)$$

where

$$t = \frac{1}{k}\sum_{i=1}^{k}(1 - p_{i:k})r_{i:k}$$

where

$p_{i:k} = (i - 0.35)/k$ and $r_{i:k}$ is the ith order statistic in a sample of size k.

4.7.5 Estimation in the special case of the Pareto Distribution

A simplified method can be employed to estimate the tail index of a distribution that is simpler than is a special case of the GPD framework described so far. This approach is based on the fact that the tail of the distribution is described by a Pareto distribution

$$P(R - u \leq r / R > u) = 1 - cr^{-1/\xi}$$

It is well known that if a fat-tailed distribution function F is in the maximum domain of attraction of a Fréchet distribution, then the tail of the F satisfies

$$1 - F(r) = L(L)r^{-1/\xi}$$

where $1 - F(r) = L(L)r^{-1/\xi}$ is a slowly regularly varying at infinity function such that

$$\lim_{r\to\infty} \frac{L(tr)}{L(r)} = 1 \quad \text{for } t > 0$$

That the tail distribution can be approximated as

$$1 - F(r) \approx cr^{-1/\xi}$$

Given this approximation and using the conditional density function

$$\frac{f(r_i)}{1 - F(u)} = -\frac{1}{\xi}\frac{cr_i^{-1/\xi-1}}{cu^{-1/\xi}}$$

we can write the likelihood function for all observations r_i larger than the threshold u, as

$$l(c,\xi) = \prod_{i=1}^{k} \frac{f(r_i)}{1 - F(u)} = \prod_{i=1}^{k} -\frac{1}{\xi}\frac{cr_i^{-1/\xi-1}}{cu^{-1/\xi}}$$

so that the log likelihood function is

$$\ln l(c,\xi) = -\sum_{y=1}^{k} - \ln(\xi) - \left(\frac{1}{\xi} + 1\right)\ln(r_i) + \frac{1}{\xi}\ln(u)$$

Taking derivatives with respect to ξ and setting it to zero yields the simple Hill (1972) estimator.

$$\hat{\xi} = \frac{1}{k}\sum_{i=1}^{k} \ln(r_i / u) = \frac{1}{k}\sum_{i=1}^{k} \ln r_i - \ln u$$

The c parameter can be estimated by ensuring that the fraction of observations beyond the threshold is accurately captured by the density as in

$$F(u) = 1 - k/n$$

From the definition of F(u) we can write

$$1 - \hat{c}u^{-1/\hat{\xi}} = 1 - k/n$$

Solving this equation for c yields the estimate of c

$$\hat{c} = \frac{k}{n} u^{1/\hat{\xi}}$$

where k is an integer which determines the number of extreme observations considered in estimating the tail index. Goldie and Smith (1987) show that

$$(\hat{\xi} - \xi)\sqrt{k} \rightarrow N(0, \xi^2)$$

The last result allows again the statistical significance of the estimated coefficient. The value at risk is easily calculated in this case and it is given by

$$r_{EP} = (p / \hat{c})^{-\xi} = u(pn / k)^{-\xi}.$$

The Choice between the Generalised Pareto Distribution and the Pareto distribution is a matter of choice for the researcher. As McNeil and Frey (2000) state, the Pareto distribution describes only heavy-tailed data whereas the Generalised Pareto Distribution is able to represent data that are light-tailed or even short-tailed. If there are periods when the distribution of financial returns appears to be light-tailed then the GPD will be a more appropriate model. Against that, one of course has to set the increased computational burden that this more general procedure requires.

4.7.6 Empirical Results

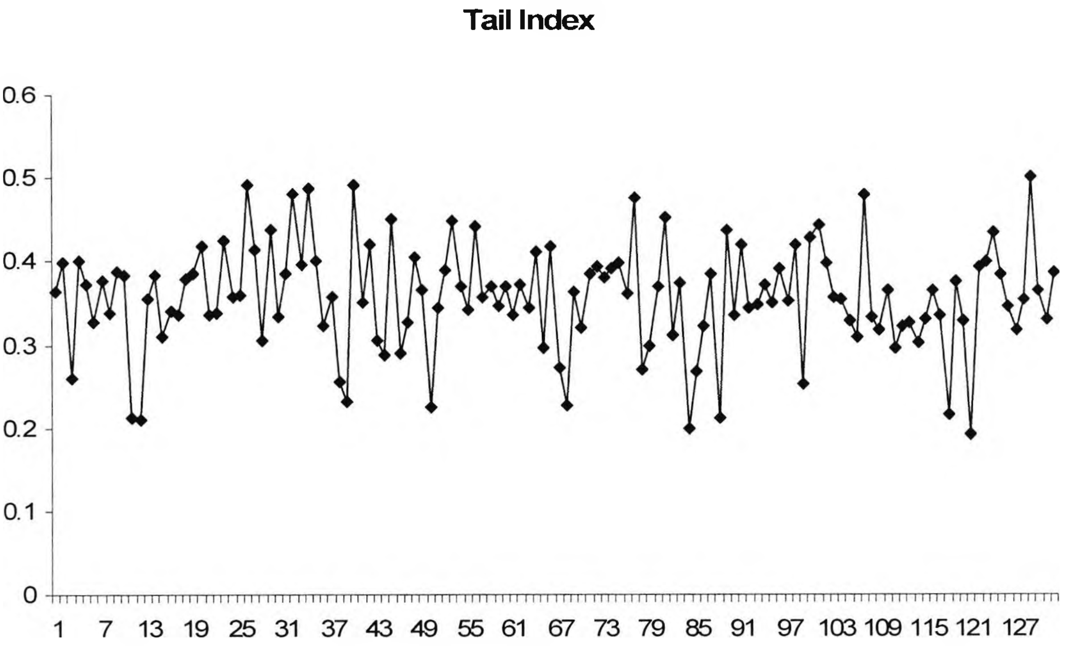
We have used the Pareto distribution version of the tails to estimate the lower tail of the returns of the 132 Greek stocks of our sample. In fact we have estimated the lower tail for both the raw data as well as for the standardised residuals.

Before we report the results it is necessary to discuss one important issue that is considered a major weakness of this approach, namely the choice of the sample size. All the estimators derived in the previous section require the choice of value for k that essentially determines which observations lie in the tail of the empirical distribution. The size of the sample from which the estimates of the tail parameter are drawn plays an important role. The main problem is that estimates of ξ seem to be affected by the choice of k as they become larger for larger values of k. The dilemma the researcher faces is whether to use a large or a small value for k. A small value of k produces large standard errors for ξ and makes the statistical inference unreliable. A large value for k on the other hand, improves the asymptotic properties of ξ (Pagan 1996) but it also increases the probability that the sample contains observations that are not extreme, increasing the bias in ξ.

There very little guidance as to what is the appropriate value for k, although DuMouchel (1983) recommends that k should not exceed 10% of the sample. Similarly Brock and de Lima (1996) report that "for reasonably sized samples, Loretan's (1991) simulations indicate that ξ is a robust estimator of ξ if k does not exceed 10% of the sample size". Also Loretan and Phillips (1994) advised that ξ should be estimated for a variety of values of k.

We have selected as our extreme sample the top 10% of our negative returns. We have also experiments with sample sizes of 5% to 9%. The results are relatively robust. The results for the raw negative returns are shown in Figure 4.13. The tail index takes values between 0.19 and 0.5, a set of values consistent with other empirical studies. As we have discussed the lower the value of the ξ, the close the tails are to the tails of the normal distribution, with the case $\xi = 0$ corresponding to normal tails.

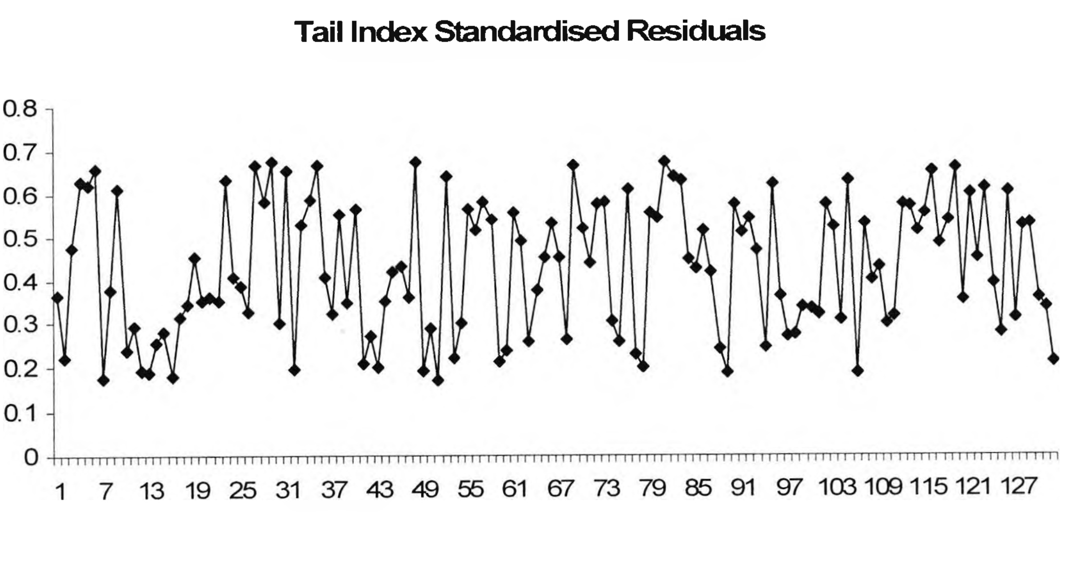
Figure 4.13 ξ Coefficient Estimates Raw Series



Since the value of $\xi \leq 0.5$, it means that the corresponding tail index of a stable distribution α must be larger than or equal to 2. This is an important finding because the empirical results support the view that the second moment (i.e. the variance) exists. The overwhelming majority of the values of also indicate that the third as well as the forth moment of the distribution may also be defined.

The application of extreme value theory presupposes that the returns are iid an assumption that is not supported by the data since serial correlation and volatility clustering are present. Whereas it is claimed by some researchers that only in the case of strong correlation is the use of extreme value theory inappropriate, we have nevertheless have estimated the lower tails of the standardised residuals, which meet this assumption. The results are shown in Figure 4.14.

Figure 4.14 ξ Coefficient Estimates Residuals



Summarising the results from this section we can say that the results provide further evidence about the fat tails of the distribution. The individual company results show a wide range of fatness but all have fatter tails than the normal. The results also provide evidence that the variance of the returns does exist.

4.8 Cornish – Fisher Approximations

The alternatives to the normal distribution considered so far consisted of (a) distributions that exhibit kurtosis or both skewness and kurtosis (b) models with time varying second moment and (c) models that approximate the tail of a distribution. In the spirit of the last approach, a different approach is instead of using a specific distribution to "correct" the normal distribution quantile estimates using the so-called Fischer Cornish approximation given by

$$z_{CF} = z_N + \frac{\gamma_3}{6}\left(z_N^2 - 1\right) + \frac{\gamma_4}{24}\left(z_N^3 - 3z_N\right) - \frac{\gamma_3^2}{36}\left(2z_N^3 - 5z_N\right)$$

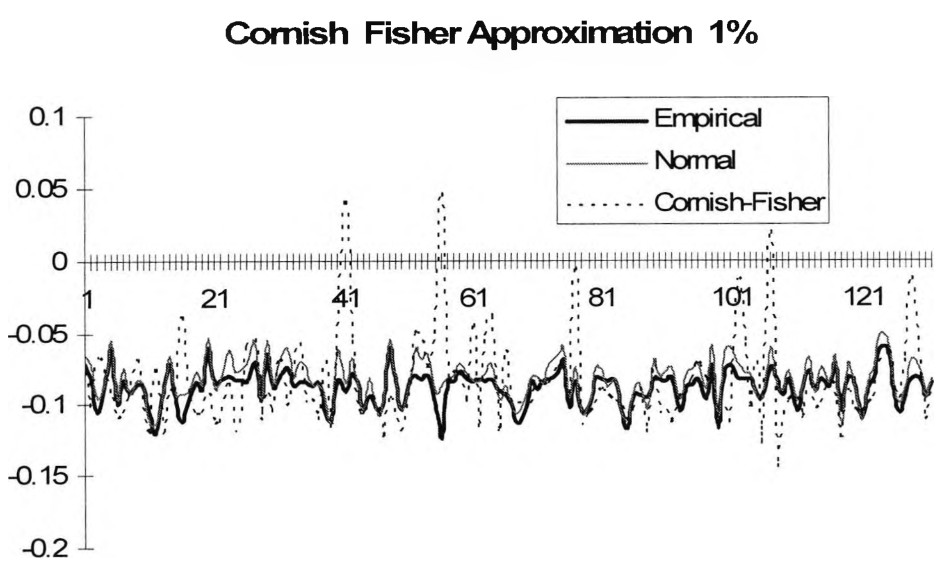
where:

z_N = the critical value for the standard normal distribution at a given percentile

γ_3 = Skewness coefficient , γ_4 = Excess kurtosis coefficient

In order to asses the usefulness of the approximation we have calculated the Cornish- Fisher based quantile for all the companies a compared it to the corresponding quantile of the empirical and the normal distribution. All the moments are estimated for whole sample. The results are shown in graphs 4.15 and 4.16.

Figure 4.15 1% Quantile Estimation via Cornish –Fisher Approximation

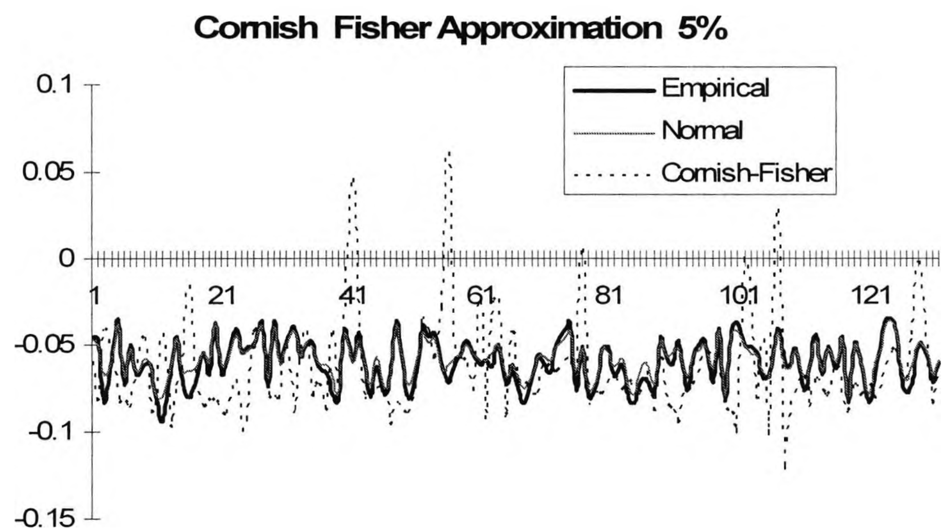


Not surprisingly the normal distribution seem to underestimate the empirical 1% quantile for nearly all the companies. The Cornish-Fisher approximation quantile, on the other hand, tends to fall on either side of the empirical one and thus with the exception a few outliers does not represent any systematic departure.

The results for the 5% quantile are similar. The normal quantile tracks down the empirical quantile very closely whereas the Cornish –Fisher approximation quantile is again much too volatile.

The above results are not conclusive evidence that the Cornish-Fisher approximation is not a useful way to modify some of the problems of the normal distribution. It merely highlights here that in this particular sample it makes the quantiles too volatile and is therefore of limited use.

Figure 4.16 5% Quantile Estimation via Cornish –Fisher Approximation



4.9 Conclusions

The purpose of this chapter has been to review the various measures that have been proposed for the measurement of the market risk of equities and to address some of the practical issues that are involved in the implementation of those measures.

The "Amendment to the Capital Accord to Incorporate Market Risk" recommends the calculation of the value at risk of a portfolio by assessing the loss that occurs in the 1 percent worst cases within a specified period of time. It does not specify how this 1 percent of worst cases will be estimated. The most obvious way is to estimate the 1 percent quantile is to derive it from the empirical frequency distribution. Another way would be to assume that returns follow a particular probability distribution which can be used to estimate the required quantile. To implement the Value-at-Risk measure under this approach we need to know with some degree of certainty the probability distribution of returns. A variation of this parametric approach is to fit a model to the tails of the distribution rather than to the entire distribution. A third general approach is to simulate the portfolio returns in order to gain a better understanding of the process that generated equity returns and to estimate the quantiles of the distribution from the simulated issues. The latter again requires some distributional assumptions to be made.

There is no general agreement as to which of the general approaches mentioned above is the most effective and acceptable. In a survey by the Financial Services Authority revealed that in estimating their market risk models 42 percent of UK banks used a variance-covariance (i.e. parametric approach), 31 percent used historical simulation and 23 percent used Monte Carlo methods. [see Berkowitz and O'Brien (2001)].

The difficulties of implementing the Value-at-Risk approach for the measurement of market risk has been the subject of this chapter. Starting with the issue of distributional assumptions, using a sample of 132 companies quoted on the Athens Stock Exchange, we have found that the normality assumption is not supported by the data. Using a variety of test we have found that all companies have unconditional distributions that cannot be described as normal. However trying to identify alternative distributions is not an easy task. We have examined the performance of two classes of probability distributions, namely the generalised t-distribution and the class of stable distributions. Although the skewed t-distribution performed better for the majority of the stocks it was not a suitable model for all the companies. The fact that different distributions describe the individual stocks has severe implications for the measurement of portfolio risk. For the portfolio return distribution may not be adequately be captured by one of the multivariate versions of the single probability distributions.

The next aspect that we have examined was the dynamic specification of asset returns. Finding the right model for the conditional mean and volatility is important for two reasons. First, by modelling the conditional mean we can decide whether the conventional square root method for time aggregation of risk is correct. We have found that for a large number of returns there is significant first order correlation. Thus for those companies the square root of time rule is wrong. Modelling the volatility using a GARCH(1,1) model with a t-distribution for the innovations seems to be the most common way of modelling volatility (McNeil and Frey (2000)). We have used this technique to model volatility for each of the 132 share returns. We have found that the model fits the data well in terms of the conventional statistical criteria. However, the standardised residuals still exhibit kurtosis, which is an indication that a either a distribution with heavier tails should be used or that extreme value theory should be applied to the data. As a corollary, we have found that the degrees of freedom parameter differs from share to share so that the return of a portfolio made up of these shares will not follow a multivariate t-distribution.

In concluding this chapter we will agree with the statement of Khindarova Rachev and Schwartz (2001) who summarize the state of the main approaches as follows: "The existing methods do not provide satisfactory evaluation of VaR The main drawback is the lack of convincing unified model for VaR capturing the following phenomena generally observed in financial data such as asset returns, interest rates, exchange rates and equities:

- Heavy tails of the marginal distribution of the process of financial returns
- Time-varying volatility
- Short and long-term dependence"

We have tried to highlight some of the implications and the problems of the above observations and critically evaluate their usefulness in terms of the assumptions they make and in terms of whether they incorporate the salient empirical characteristics..

Risk managers, regulators and investors have developed a variety of methods to model equity returns and measuring their equity price risks. Understanding and properly measuring risk is necessary for a successful investment management strategy or a risk management system.

Our review has shown that although the range of statistical techniques employed in the estimation of financial risk has increased exponentially with techniques such as extreme value theory brought in from other fields such as hydrology and meteorology, the nature of economic time series make statistical prediction unreliable. Some of the problems relate to small sample that are available to financial series research, whereas other problems relate to the estimation error problems. We have not addressed in this review these types of problems which are examined by Figlewski (2002).

APPENDIX TO CHAPTER 4: STATISTICAL RESULTS FOR INDIVIDUAL COMPANIES

Table A1: Average Returns and Volatility

	Average Returns			Volatility		
	Whole	Period 1	Period 2	Whole	Period 1	Period 2
ABK	0.0004	0.0020	-0.0013	0.0285	0.0277	0.0293
AVAX	0.0012	0.0023	0.0000	0.0285	0.0336	0.0280
AAAK	0.0007	0.0034	-0.0023	0.0285	0.0350	0.0471
ALTEC	0.0004	0.0048	-0.0043	0.0285	0.0338	0.0369
ALFA	0.0002	0.0022	-0.0019	0.0285	0.0256	0.0213
ATTIK	-0.0005	0.0027	-0.0040	0.0285	0.0384	0.0443
AKTOR	0.0005	0.0023	-0.0015	0.0285	0.0317	0.0320
ALATK	-0.0004	0.0027	-0.0038	0.0285	0.0376	0.0418
AEGEP	-0.0003	0.0025	-0.0032	0.0285	0.0349	0.0352
AXON	0.0000	0.0031	-0.0033	0.0285	0.0335	0.0425
BIOSK	-0.0002	0.0035	-0.0041	0.0285	0.0471	0.0501
BIOSP	0.0000	0.0029	-0.0030	0.0285	0.0433	0.0541
BIOT	0.0002	0.0028	-0.0026	0.0285	0.0348	0.0416
BIOXK	0.0005	0.0032	-0.0025	0.0285	0.0295	0.0274
BISK	-0.0001	0.0027	-0.0031	0.0285	0.0356	0.0442
BISP	-0.0001	0.0025	-0.0028	0.0285	0.0338	0.0449
GEBKA	0.0002	0.0032	-0.0030	0.0285	0.0356	0.0424
GEK	0.0001	0.0026	-0.0027	0.0285	0.0330	0.0334
GENAK	0.0010	0.0042	-0.0024	0.0285	0.0370	0.0376
GOOD	0.0007	0.0024	-0.0010	0.0285	0.0238	0.0223
DIAS	0.0004	0.0043	-0.0039	0.0285	0.0331	0.0421
DIEKA	-0.0001	0.0029	-0.0034	0.0285	0.0347	0.0376
DK	-0.0002	0.0014	-0.0019	0.0285	0.0298	0.0226
EBZ	-0.0004	0.0012	-0.0022	0.0285	0.0350	0.0305
EGNAK	-0.0002	0.0022	-0.0028	0.0285	0.0376	0.0257
EDRA	-0.0004	0.0018	-0.0027	0.0285	0.0288	0.0321
EEEK	0.0001	0.0009	-0.0008	0.0285	0.0255	0.0219
EKTER	-0.0002	0.0026	-0.0031	0.0285	0.0399	0.0442
ELAIS	0.0001	0.0019	-0.0019	0.0285	0.0258	0.0214
ELBE	0.0001	0.0034	-0.0036	0.0285	0.0345	0.0397
ELKA	-0.0002	0.0022	-0.0028	0.0285	0.0303	0.0294
ELL	0.0012	0.0032	-0.0009	0.0285	0.0271	0.0255
ELMEK	0.0005	0.0030	-0.0021	0.0285	0.0306	0.0391
ELMPI	0.0007	0.0029	-0.0015	0.0285	0.0267	0.0336
ELTEX	0.0005	0.0022	-0.0013	0.0285	0.0324	0.0300
ELTK	0.0002	0.0032	-0.0030	0.0285	0.0339	0.0398
ELYF	-0.0002	0.0016	-0.0021	0.0285	0.0372	0.0368
ELFK	0.0003	0.0035	-0.0031	0.0285	0.0358	0.0475
EMDKO	-0.0003	0.0037	-0.0047	0.0285	0.0439	0.0521
EMP	0.0003	0.0028	-0.0023	0.0285	0.0294	0.0232
EXEL	0.0008	0.0037	-0.0023	0.0285	0.0372	0.0374
EPERA	0.0003	0.0028	-0.0025	0.0285	0.0281	0.0300
EPILK	-0.0004	0.0026	-0.0035	0.0285	0.0320	0.0411
ERGAS	-0.0013	0.0020	-0.0049	0.0285	0.0400	0.0473
ERMES	0.0015	0.0050	-0.0023	0.0285	0.0300	0.0400
ESK	0.0005	0.0032	-0.0023	0.0285	0.0422	0.0502
ESHA	0.0000	0.0024	-0.0025	0.0285	0.0360	0.0470

ETE	0.0005	0.0028	-0.0019	0.0285	0.0261	0.0211
ETLES	-0.0005	0.0022	-0.0034	0.0285	0.0345	0.0367
ETMAK	-0.0006	0.0025	-0.0038	0.0285	0.0399	0.0485
ZAMPA	0.0000	0.0026	-0.0028	0.0285	0.0322	0.0402
HERACL	0.0000	0.0020	-0.0021	0.0285	0.0278	0.0239
THEMEL	0.0001	0.0028	-0.0029	0.0285	0.0269	0.0318
IATR	0.0001	0.0034	-0.0034	0.0285	0.0276	0.0282
INTEK	-0.0005	0.0040	-0.0054	0.0285	0.0365	0.0416
INTEP	0.0004	0.0042	-0.0037	0.0285	0.0371	0.0376
INTER	0.0000	0.0033	-0.0036	0.0285	0.0320	0.0380
INTET	0.0007	0.0041	-0.0030	0.0285	0.0301	0.0383
INTKA	0.0003	0.0031	-0.0026	0.0285	0.0324	0.0284
IONA	0.0005	0.0027	-0.0019	0.0285	0.0349	0.0339
IONE	0.0000	0.0027	-0.0030	0.0285	0.0361	0.0357
KALSK	-0.0002	0.0021	-0.0026	0.0285	0.0351	0.0374
KARELIA	0.0005	0.0017	-0.0009	0.0285	0.0327	0.0347
KATSK	0.0002	0.0026	-0.0024	0.0285	0.0301	0.0308
KEKROPS	0.0012	0.0042	-0.0020	0.0285	0.0374	0.0465
KERAL	0.0008	0.0032	-0.0018	0.0285	0.0315	0.0432
KERK	-0.0003	0.0024	-0.0032	0.0285	0.0383	0.0452
KORFK	-0.0002	0.0029	-0.0035	0.0285	0.0324	0.0567
KREKA	-0.0005	0.0018	-0.0029	0.0285	0.0305	0.0495
LABI	-0.0002	0.0026	-0.0032	0.0285	0.0323	0.0368
LAMDA	0.0007	0.0027	-0.0014	0.0285	0.0361	0.0345
LAMPSA	0.0010	0.0042	-0.0025	0.0285	0.0365	0.0373
LOYLI	0.0003	0.0029	-0.0025	0.0285	0.0317	0.0301
LYKOS	0.0001	0.0026	-0.0026	0.0285	0.0281	0.0280
MAIK	0.0007	0.0033	-0.0021	0.0285	0.0268	0.0251
MAXIM	-0.0002	0.0023	-0.0030	0.0285	0.0336	0.0471
MARFIN	0.0000	0.0029	-0.0031	0.0285	0.0315	0.0322
MEAGA	-0.0005	0.0030	-0.0044	0.0285	0.0400	0.0508
MESOX	-0.0004	0.0029	-0.0040	0.0285	0.0410	0.0442
METK	0.0008	0.0037	-0.0022	0.0285	0.0332	0.0313
MHXK	-0.0005	0.0022	-0.0034	0.0285	0.0317	0.0332
MINE	0.0002	0.0028	-0.0026	0.0285	0.0375	0.0387
MOUZK	-0.0002	0.0018	-0.0024	0.0285	0.0350	0.0351
MOYLT	-0.0003	0.0034	-0.0043	0.0285	0.0431	0.0473
MOYR	-0.0005	0.0040	-0.0052	0.0285	0.0411	0.0533
MOXLOS	0.0000	0.0036	-0.0039	0.0285	0.0356	0.0435
MPENK	0.0002	0.0025	-0.0023	0.0285	0.0352	0.0387
MPOKA	0.0001	0.0029	-0.0030	0.0285	0.0472	0.0418
MPSTK	0.0002	0.0021	-0.0018	0.0285	0.0296	0.0291
MPTK	-0.0001	0.0027	-0.0030	0.0285	0.0349	0.0344
MRFKO	-0.0002	0.0025	-0.0030	0.0285	0.0280	0.0380
MYTIL	0.0003	0.0042	-0.0038	0.0285	0.0313	0.0319
NAOYK	-0.0005	0.0037	-0.0051	0.0285	0.0323	0.0491
NEL	-0.0003	0.0029	-0.0038	0.0285	0.0326	0.0361
NHR	-0.0008	0.0017	-0.0034	0.0285	0.0280	0.0363
NIKAS	0.0003	0.0031	-0.0027	0.0285	0.0296	0.0320
XYLK	0.0005	0.0035	-0.0027	0.0285	0.0348	0.0427
PAPAK	0.0003	0.0016	-0.0011	0.0285	0.0271	0.0238
PARN	-0.0013	0.0011	-0.0039	0.0285	0.0421	0.0518
PEILH	0.0008	0.0026	-0.0011	0.0285	0.0287	0.0262
PEIR	0.0008	0.0034	-0.0020	0.0285	0.0307	0.0199
PEPA	0.0002	0.0028	-0.0027	0.0285	0.0306	0.0318
PETZK	-0.0002	0.0028	-0.0035	0.0285	0.0310	0.0359
PLATH	0.0002	0.0033	-0.0031	0.0285	0.0325	0.0361

PLAS	-0.0009	0.0026	-0.0046	0.0285	0.0392	0.0398
PRD	-0.0007	0.0020	-0.0036	0.0285	0.0374	0.0451
PROOD	0.0002	0.0021	-0.0019	0.0285	0.0247	0.0278
RILKE	-0.0001	0.0021	-0.0025	0.0285	0.0334	0.0409
RINTE	0.0003	0.0034	-0.0030	0.0285	0.0381	0.0397
ROKKA	0.0002	0.0027	-0.0026	0.0285	0.0329	0.0328
SAIKL	-0.0006	0.0025	-0.0038	0.0285	0.0377	0.0459
SANYO	0.0000	0.0038	-0.0041	0.0285	0.0358	0.0390
SAR	-0.0002	0.0024	-0.0029	0.0285	0.0280	0.0300
SATOK	-0.0003	0.0032	-0.0040	0.0285	0.0377	0.0413
SELMK	-0.0004	0.0025	-0.0035	0.0285	0.0320	0.0314
SELO	-0.0009	0.0015	-0.0035	0.0285	0.0328	0.0394
SIDE	0.0002	0.0026	-0.0024	0.0285	0.0294	0.0271
STALK	-0.0014	0.0023	-0.0054	0.0285	0.0450	0.0533
STRIK	-0.0004	0.0016	-0.0026	0.0285	0.0299	0.0320
TERNA	0.0002	0.0028	-0.0027	0.0285	0.0334	0.0378
TEXN	-0.0017	0.0013	-0.0049	0.0285	0.0420	0.0521
TZKA	-0.0002	0.0028	-0.0033	0.0285	0.0299	0.0428
THLET	0.0007	0.0027	-0.0015	0.0285	0.0340	0.0310
TITK	0.0009	0.0023	-0.0006	0.0285	0.0263	0.0183
TSIP	0.0006	0.0028	-0.0017	0.0285	0.0220	0.0254
YALKO	-0.0001	0.0026	-0.0031	0.0285	0.0338	0.0413
FANKO	-0.0006	0.0025	-0.0039	0.0285	0.0318	0.0517
FINTO	0.0005	0.0030	-0.0022	0.0285	0.0318	0.0422
FOIN	0.0000	0.0030	-0.0033	0.0285	0.0246	0.0336
FRLK	-0.0002	0.0023	-0.0028	0.0285	0.0323	0.0322
HALYB	-0.0002	0.0033	-0.0040	0.0285	0.0381	0.0435
XATZK	-0.0001	0.0024	-0.0028	0.0285	0.0367	0.0343

Table A2: Maximum and Minimum Values of Returns

	MAX Period 1	MAX Period 2	MIN Period 1	MIN Period 2
ABK	0.0778	0.1118	-0.0844	-0.1265
AVAX	0.0855	0.1121	-0.0918	-0.1199
AAAK	0.0787	0.1606	-0.0948	-0.1278
ALTEC	0.0846	0.1272	-0.0853	-0.1917
ALFA	0.0773	0.1115	-0.0837	-0.0855
ATTIK	0.0886	0.1499	-0.0938	-0.1350
AKTOR	0.0800	0.1127	-0.0867	-0.1147
ALATK	0.0870	0.1542	-0.0892	-0.1957
AEGEP	0.0808	0.1102	-0.0960	-0.1618
AXON	0.0800	0.1630	-0.0864	-0.1977
BIOSK	0.1335	0.1178	-0.1335	-0.1288
BIOSP	0.1178	0.1158	-0.1178	-0.1278
BIOT	0.0931	0.1643	-0.0924	-0.1421
BIOXK	0.0805	0.1100	-0.0843	-0.1234
BISK	0.0795	0.1316	-0.0861	-0.1342
BISP	0.0794	0.1133	-0.0874	-0.1282
GEBKA	0.0953	0.1234	-0.1112	-0.1252
GEK	0.0785	0.1118	-0.0849	-0.1148
GENAK	0.0953	0.1625	-0.0953	-0.1200
GOOD	0.0774	0.1623	-0.0838	-0.1032
DIAS	0.1018	0.1132	-0.0870	-0.1252
DIEKA	0.0837	0.1267	-0.0910	-0.1259
DK	0.0786	0.0902	-0.0839	-0.1026
EBZ	0.0782	0.1632	-0.0839	-0.1054
EGNAK	0.0789	0.1480	-0.0980	-0.0852
EDRA	0.0797	0.1118	-0.0860	-0.1280
EEEK	0.0772	0.0845	-0.0836	-0.0950
EKTER	0.0849	0.1126	-0.1439	-0.1306
ELAIS	0.0773	0.1123	-0.0829	-0.0924
ELBE	0.0800	0.1654	-0.0883	-0.1239
ELKA	0.0786	0.1190	-0.0859	-0.1115
ELL	0.0775	0.1126	-0.1220	-0.1040
ELMEK	0.0902	0.1155	-0.1252	-0.1274
ELMPI	0.0831	0.1438	-0.0863	-0.1283
ELTEX	0.0805	0.1137	-0.0878	-0.1049
ELTK	0.0822	0.1415	-0.1216	-0.1321
ELYF	0.0800	0.1623	-0.0913	-0.1108
ELFK	0.0829	0.1130	-0.0892	-0.1271
EMDKO	0.0976	0.1160	-0.1027	-0.1289
EMP	0.0774	0.1095	-0.0938	-0.1084
EXEL	0.1018	0.1171	-0.3697	-0.1247
EPERA	0.0792	0.1241	-0.1475	-0.1932
EPILK	0.0846	0.1664	-0.0899	-0.1285
ERGAS	0.0838	0.1499	-0.1022	-0.1542
ERMES	0.1018	0.1306	-0.1054	-0.1235
ESK	0.1018	0.1641	-0.1018	-0.1939
ESHA	0.0794	0.1562	-0.0845	-0.1911
ETE	0.0774	0.1105	-0.0836	-0.0981
ETLES	0.0795	0.1463	-0.1290	-0.1164
ETMAK	0.0817	0.1685	-0.0910	-0.1435
ZAMPA	0.0776	0.1653	-0.0835	-0.1311
HERACL	0.0777	0.0819	-0.1004	-0.1058

THEMEL	0.0787	0.1112	-0.0841	-0.1915
IATR	0.0785	0.1007	-0.0852	-0.1043
INTEK	0.0834	0.1374	-0.0877	-0.1967
INTEP	0.0839	0.1133	-0.0849	-0.1965
INTER	0.0800	0.1526	-0.0935	-0.1249
INTET	0.0870	0.1547	-0.0852	-0.1269
INTKA	0.0796	0.1040	-0.0852	-0.1763
IONA	0.0791	0.1193	-0.0841	-0.1137
IONE	0.0796	0.1123	-0.1435	-0.2285
KALSK	0.0812	0.1620	-0.0945	-0.1115
KARELIA	0.0775	0.1049	-0.0838	-0.1272
KATSK	0.0786	0.1107	-0.0849	-0.1949
KEKROPS	0.0798	0.1649	-0.0853	-0.1768
KERAL	0.0855	0.1135	-0.0968	-0.1957
KERK	0.0870	0.1134	-0.0893	-0.1278
KORFK	0.0953	0.1635	-0.0953	-0.1542
KREKA	0.0810	0.1652	-0.0854	-0.1808
LABI	0.0816	0.1110	-0.0862	-0.1244
LAMDA	0.0953	0.1135	-0.0972	-0.1268
LAMPSA	0.0820	0.1226	-0.0878	-0.1257
LOYLI	0.0800	0.1054	-0.0858	-0.1268
LYKOS	0.0796	0.1048	-0.0976	-0.1230
MAIK	0.0796	0.0948	-0.0865	-0.1234
MAXIM	0.0849	0.1372	-0.0903	-0.1272
MARFIN	0.0892	0.1100	-0.1319	-0.2557
MEAGA	0.0870	0.1652	-0.1027	-0.1744
MESOX	0.0924	0.1154	-0.1082	-0.1585
METK	0.0811	0.1116	-0.0838	-0.1082
MHXK	0.0791	0.1128	-0.1454	-0.1262
MINE	0.0830	0.1133	-0.0846	-0.1521
MOUZK	0.0819	0.1413	-0.0893	-0.1163
MOYLT	0.0804	0.1602	-0.0853	-0.1666
MOYR	0.1018	0.1152	-0.1018	-0.1278
MOXLOS	0.0972	0.1405	-0.0942	-0.1470
MPENK	0.0809	0.1108	-0.0839	-0.1270
MPOKA	0.1092	0.1276	-0.1092	-0.1257
MPSTK	0.0786	0.1115	-0.0839	-0.1173
MPTK	0.0849	0.1641	-0.1190	-0.1242
MRFKO	0.0805	0.1642	-0.0847	-0.1058
MYTIL	0.0792	0.1129	-0.0838	-0.1212
NAOYK	0.0793	0.1594	-0.0838	-0.1823
NEL	0.0861	0.1094	-0.1078	-0.1199
NHR	0.0798	0.1606	-0.1054	-0.1404
NIKAS	0.0792	0.1071	-0.0864	-0.1278
XYLK	0.0800	0.1477	-0.0864	-0.1273
PAPAK	0.0780	0.0950	-0.0832	-0.1011
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

TEXN	0.0810	0.1128	-0.0880	-0.1280
TZKA	0.0841	0.1651	-0.0887	-0.1274
THLET	0.0855	0.1106	-0.0931	-0.1278
TITK	0.0772	0.0972	-0.0834	-0.0725
TSIP	0.0780	0.1123	-0.0842	-0.1228
YALKO	0.0854	0.1506	-0.0899	-0.1289
FANKO	0.0792	0.1647	-0.0866	-0.1418
FINTO	0.0849	0.1596	-0.0931	-0.1233
FOIN	0.1153	0.1447	-0.0870	-0.1980
FRLK	0.0827	0.1342	-0.0865	-0.1265
HALYB	0.1018	0.1132	-0.0953	-0.1262
XATZK	0.0806	0.1106	-0.0887	-0.1278

Table A3: Coefficients of Skewness and Kurtosis

	SKEWNESS			KURTOSIS		
	Whole	Period 1	Period 2	Whole	Period 1	Period 2
ABK	0.1376	0.2335	0.0692	1.6465	0.9057	2.2698
AVAX	0.2345	0.1870	0.2607	1.0169	0.2503	2.3443
AAAK	-0.1064	-0.0317	-0.0499	0.8908	1.5118	0.2891
ALTEC	0.0447	0.1393	0.0309	1.0369	0.2669	1.5987
ALFA	0.3250	0.0761	0.6546	2.2293	1.6791	3.2825
ATTIK	0.0874	0.1148	0.1259	0.1406	-0.1964	0.2729
AKTOR	0.2451	0.2372	0.2597	1.1057	0.6277	1.6370
ALATK	0.0387	0.0836	0.0500	0.5408	-0.1616	0.9838
AEGEP	-0.0393	0.2036	-0.2906	0.8025	0.1152	1.3848
AXON	0.2313	0.2832	0.2909	1.5832	0.5238	1.7673
BIOSK	0.0030	-0.0599	0.0870	-0.3592	-0.4944	-0.2254
BIOSP	-0.2102	-0.1169	-0.1971	0.2927	0.2599	0.0652
BIOT	0.2367	0.3080	0.2513	0.6765	0.4586	0.6562
BIOXK	0.2005	0.2126	0.1411	1.1688	0.6170	1.9294
BISK	0.0070	0.1124	0.0183	0.3901	0.1785	0.2725
BISP	-0.2373	0.0779	-0.2938	1.3232	0.8805	0.9959
GEBKA	0.0646	0.0115	0.1654	0.6710	0.9109	0.4448
GEK	0.2259	0.2292	0.2342	0.9469	0.6389	1.2999
GENAK	0.1745	0.0872	0.2772	0.9438	0.7252	1.2762
GOOD	0.3579	0.2614	0.4567	3.2014	1.6671	5.4484
DIAS	0.1883	0.3520	0.2198	0.5656	0.4813	0.3716
DIEKA	0.1932	0.2597	0.1792	0.3908	0.0390	0.6389
DK	0.2129	0.1952	0.0733	1.8069	1.1056	2.7018
EBZ	0.3127	0.2964	0.2817	1.0489	0.2939	2.2447
EGNAK	0.2807	0.0690	0.6032	1.0598	0.1314	3.1564
EDRA	0.0257	0.2593	-0.1117	1.8739	1.3969	2.0682
EEEK	0.0821	0.0733	0.0526	2.2278	2.0109	2.3455
EKTER	0.0726	0.0959	0.0917	0.0795	-0.1829	0.2254
ELAIS	0.3229	0.4377	-0.0182	2.6086	1.5462	4.2758
ELBE	0.1533	0.3180	0.1067	0.7121	0.2545	0.8663
ELKA	0.0478	0.2089	-0.1545	1.1281	0.9155	1.2925
ELL	0.3722	0.3684	0.3521	2.7982	2.4631	3.2835
ELMEK	0.0595	0.2336	0.0485	1.0831	1.4409	0.6269
ELMPI	0.0506	0.3193	-0.0209	2.5042	1.5857	2.5137
ELTEX	0.2136	0.2136	0.1864	1.1444	0.5785	1.9245
ELTK	0.0995	0.1386	0.1405	0.7049	0.5725	0.6730
ELYF	0.1873	0.1531	0.2239	0.7038	0.2102	1.2888
ELFK	-0.1348	0.0535	-0.1244	0.5179	0.8917	0.0173
EMDKO	-0.0015	0.0454	0.0414	-0.2568	-0.6076	-0.1841
EMP	0.0984	0.0374	0.0257	2.0936	1.2956	3.4897
EXEL	-0.3585	-0.9567	0.2693	6.3006	11.8372	1.0945
EPERA	0.0716	0.1438	0.0448	2.9206	2.1280	3.6022
EPILK	0.1550	0.3254	0.1565	0.7554	0.9726	0.3877
ERGAS	0.0359	0.1506	0.0264	0.1979	-0.1811	0.2604
ERMES	0.1839	0.3735	0.2193	1.7245	3.4767	0.7378
ESK	0.0189	0.1211	0.0039	0.3521	-0.5578	0.6871
ESHA	0.1927	0.3517	0.1771	0.5365	-0.0055	0.4408
ETE	0.3345	0.2824	0.2570	2.2593	1.7878	2.7479
ETLES	0.2108	0.2654	0.1944	0.8339	0.5431	1.0687
ETMAK	0.1209	0.0541	0.2250	0.3963	0.0669	0.4385

ZAMPA	0.1429	0.2532	0.1559	1.4097	1.3000	1.2020
HERACL	0.0068	0.0761	-0.2035	1.5614	1.0363	2.2648
THEMEL	0.0521	0.4230	-0.1125	2.6016	1.6610	2.8473
IATR	0.1227	0.1450	0.1216	1.3616	0.8157	1.9830
INTEK	-0.1584	0.1899	-0.3354	1.3343	-0.0560	1.9595
INTEP	-0.5917	0.0008	-1.2136	3.0461	0.2452	5.5188
INTER	0.0655	0.1439	0.0983	0.9090	1.1619	0.6406
INTET	0.0559	0.1619	0.1106	1.0766	0.9546	0.8704
INTKA	0.1125	0.1663	-0.0666	1.3501	0.5002	2.6527
IONA	0.1533	0.1372	0.1613	0.7802	0.3051	1.3817
IONE	-0.1443	-0.0451	-0.2633	1.4991	0.3855	2.7348
KALSK	0.2354	0.2525	0.2469	0.7374	0.2073	1.1818
KARELIA	-0.0794	0.0742	-0.2035	1.2768	0.9845	1.4718
KATSK	-0.0809	0.2126	-0.3642	2.1796	0.8536	3.3562
KEKROPS	0.1761	0.0728	0.3046	0.6256	0.2914	0.6346
KERAL	-0.0426	0.2047	-0.0603	1.6005	0.8338	1.3000
KERK	0.0564	0.1434	0.0533	0.2229	0.0036	0.2132
KORFK	0.0161	0.1711	0.1021	0.7116	2.2556	-0.3129
KREKA	0.1620	0.2848	0.2108	1.8332	0.7886	1.0770
LABI	0.1502	0.1939	0.1771	0.6568	0.4975	0.6930
LAMDA	0.0625	0.0994	-0.0005	0.9057	0.5399	1.3601
LAMPSA	0.0542	0.0860	0.0357	0.6412	0.3082	0.9805
LOYLI	0.1314	0.2250	-0.0155	0.9788	0.5180	1.5125
LYKOS	0.1398	0.4273	-0.1721	1.6915	1.0940	2.1895
MAIK	0.1673	0.3156	-0.0697	2.0251	1.4369	2.7227
MAXIM	0.0753	0.1833	0.1159	0.5757	0.5741	0.1462
MARFIN	-0.1885	0.1645	-0.5340	4.1689	1.5004	6.6848
MEAGA	0.0652	0.1472	0.1078	0.2693	-0.2482	0.2586
MESOX	0.0474	0.1346	0.0067	0.1196	-0.3723	0.4589
METK	0.1910	0.2653	0.0639	0.8052	0.3807	1.2993
MHXK	0.1695	0.2509	0.1202	1.5129	1.1209	1.8670
MINE	0.0547	0.1509	-0.0250	0.3970	-0.0447	0.7785
MOUZK	0.1507	0.1279	0.1781	0.8224	0.3500	1.3617
MOYLT	-0.0153	-0.0598	0.0637	0.1605	-0.3755	0.5570
MOYR	0.0077	0.0568	0.0914	0.0536	0.0193	-0.1656
MOXLOS	0.1058	0.1047	0.1951	0.5460	0.6053	0.3875
MPENK	0.0741	0.1514	0.0446	0.9716	0.5436	1.2349
MPOKA	0.0757	-0.0273	0.1756	-0.2722	-0.6857	0.4075
MPSTK	0.1032	0.2213	-0.0376	1.4450	0.9500	1.9835
MPTK	0.2015	0.1374	0.2704	0.7572	0.2546	1.4084
MRFKO	0.2777	0.4646	0.2904	2.0558	2.8274	1.3383
MYTIL	0.2857	0.3333	0.2684	0.9273	0.7155	1.1818
NAOYK	0.0511	0.3852	0.1049	0.8825	0.4041	0.3503
NEL	0.1343	0.2961	0.0612	0.8890	0.8265	0.8451
NHR	0.1414	0.2290	0.1840	1.7042	1.0410	1.6229
NIKAS	0.0751	0.3304	-0.0981	1.4001	0.7440	1.7867
XYLK	-0.0010	0.0810	0.0271	0.5064	0.2841	0.4168
PAPAK	0.2220	0.3735	-0.0728	2.1103	1.4106	3.0524
PARN	0.0462	0.1279	0.0499	-0.0270	-0.3323	-0.0612
PEILH	0.3292	0.3775	0.2206	1.9498	1.8480	1.9919
PEIR	0.4124	0.2974	0.1944	2.1753	1.0417	3.8263
PEPA	-0.1780	-0.1788	-0.1601	3.1744	2.9465	3.4786
PETZK	-0.0428	0.3800	-0.2672	1.3616	0.4004	1.6535
PLATH	0.1745	0.2704	0.1568	0.6018	0.4881	0.6227
PLAS	-0.0681	-0.0592	-0.0718	1.0159	-0.1759	2.2728
PRD	0.2200	0.2624	0.2532	0.4532	0.0346	0.5343
PROOD	-0.2795	-0.0170	-0.4335	4.6664	3.0223	5.5929

RILKE	0.3844	0.3482	0.4578	0.7716	0.2152	0.8929
RINTE	0.1973	0.2103	0.2094	0.6495	0.1314	1.1440
ROKKA	0.1803	0.2432	0.1139	0.8440	0.3855	1.3359
SAIKL	0.0147	0.0311	0.0740	0.4736	-0.0159	0.5521
SANYO	0.1675	0.1959	0.1980	0.4089	0.1086	0.6396
SAR	0.1682	0.2989	0.0931	1.0385	0.8674	1.1310
SATOK	0.0811	0.1685	0.0553	0.2395	-0.0576	0.4044
SELMK	0.1968	0.1256	0.2724	0.8576	0.3757	1.5160
SELO	0.0341	0.1866	-0.0008	0.6570	0.4175	0.5923
SIDE	0.2525	0.2758	0.1763	1.1243	0.7518	1.6051
STALK	0.0300	0.0399	0.0888	-0.1986	-0.7307	-0.0091
STRIK	0.1823	0.3356	0.0747	1.2862	1.3585	1.1715
TERNA	0.1574	0.2216	0.1615	0.5420	0.4224	0.5532
TEXN	-0.0468	0.0909	-0.0582	-0.2020	-0.4593	-0.2905
TZKA	0.2075	0.1606	0.3289	1.8059	2.2824	1.1442
THLET	0.1695	0.2078	0.0737	0.8279	0.1914	1.7134
TITK	0.2194	0.0810	0.3637	2.5685	1.5952	3.9640
TSIP	0.1600	0.5122	-0.0164	2.6278	2.0210	2.7836
YALKO	0.0546	0.2250	0.0236	0.6972	0.5868	0.5319
FANKO	0.0353	0.3984	0.0621	0.7471	0.7398	0.0023
FINTO	-0.0483	0.0378	-0.0029	1.0698	1.7668	0.4505
FOIN	-0.1370	0.3257	-0.1955	4.0349	4.4932	3.1181
FRLK	0.0835	0.0737	0.0951	1.0461	0.7765	1.3856
HALYB	0.1172	0.2222	0.1033	0.2471	0.0093	0.3022
XATZK	0.0735	0.0057	0.1304	0.4584	0.0701	1.0292

Table A4: Test Statistics for the Significance of the Coefficients of Skewness and Kurtosis

	SKEWNESS				KURTOSIS	
	whole	Period 1	Period 2	Whole	Period 1	Period 2
ABK	2.373	2.900	0.828	14.199	2.270	13.587
AVAX	4.044	2.322	3.121	8.770	2.344	14.033
AAAK	-1.834	-0.393	-0.598	7.682	0.289	1.731
ALTEC	0.771	1.730	0.370	8.942	1.599	9.570
ALFA	5.605	0.945	7.837	19.225	3.283	19.650
ATTIK	1.508	1.426	1.508	1.212	0.273	1.633
AKTOR	4.227	2.945	3.109	9.536	1.637	9.799
ALATK	0.668	1.038	0.599	4.664	0.984	5.889
AEGEP	-0.679	2.527	-3.479	6.921	1.385	8.289
AXON	3.990	3.517	3.483	13.654	1.767	10.580
BIOSK	0.051	-0.744	1.042	-3.097	-0.225	-1.349
BIOSP	-3.626	-1.452	-2.359	2.524	0.065	0.390
BIOT	4.083	3.824	3.009	5.834	0.656	3.928
BIOXK	3.459	2.640	1.689	10.080	1.929	11.550
BISK	0.120	1.396	0.219	3.365	0.272	1.631
BISP	-4.092	0.967	-3.518	11.411	0.996	5.962
GEBKA	1.115	0.143	1.981	5.787	0.445	2.663
GEK	3.897	2.846	2.804	8.166	1.300	7.781
GENAK	3.010	1.083	3.318	8.139	1.276	7.640
GOOD	6.172	3.246	5.468	27.609	5.448	32.615
DIAS	3.248	4.370	2.632	4.878	0.372	2.224
DIEKA	3.333	3.225	2.145	3.370	0.639	3.825
DK	3.672	2.423	0.878	15.583	2.702	16.173
EBZ	5.394	3.681	3.373	9.046	2.245	13.437
EGNAK	4.841	0.857	7.222	9.140	3.156	18.894
EDRA	0.444	3.219	-1.338	16.161	2.068	12.381
EEEK	1.417	0.910	0.630	19.212	2.345	14.040
EKTER	1.252	1.190	1.098	0.686	0.225	1.349
ELAIS	5.570	5.435	-0.218	22.496	4.276	25.595
ELBE	2.644	3.949	1.278	6.142	0.866	5.186
ELKA	0.824	2.594	-1.849	9.728	1.293	7.737
ELL	6.420	4.574	4.216	24.132	3.284	19.656
ELMEK	1.026	2.901	0.580	9.341	0.627	3.753
ELMPI	0.873	3.964	-0.251	21.596	2.514	15.047
ELTEX	3.685	2.652	2.232	9.870	1.924	11.520
ELTK	1.717	1.721	1.682	6.079	0.673	4.029
ELYF	3.231	1.900	2.681	6.070	1.289	7.715
ELFK	-2.325	0.664	-1.489	4.467	0.017	0.103
EMDKO	-0.026	0.564	0.496	-2.215	-0.184	-1.102
EMP	1.697	0.464	0.307	18.055	3.490	20.890
EXEL	-6.183	-11.878	3.224	54.337	1.094	6.552
EPERA	1.235	1.786	0.536	25.187	3.602	21.563
EPILK	2.673	4.040	1.874	6.515	0.388	2.321
ERGAS	0.620	1.869	0.316	1.707	0.260	1.559
ERMES	3.172	4.638	2.625	14.872	0.738	4.416
ESK	0.327	1.503	0.047	3.037	0.687	4.113
ESHA	3.323	4.367	2.121	4.627	0.441	2.639
ETE	5.769	3.506	3.077	19.484	2.748	16.449
ETLES	3.635	3.295	2.328	7.192	1.069	6.397
ETMAK	2.085	0.671	2.693	3.417	0.439	2.625
ZAMPA	2.465	3.144	1.866	12.157	1.202	7.195

HERACL	0.117	0.945	-2.437	13.465	2.265	13.557
THEMEL	0.898	5.253	-1.347	22.437	2.847	17.044
IATR	2.116	1.801	1.455	11.742	1.983	11.870
INTEK	-2.733	2.357	-4.016	11.507	1.960	11.730
INTEP	-10.207	0.010	-14.529	26.270	5.519	33.036
INTER	1.130	1.786	1.177	7.839	0.641	3.835
INTET	0.964	2.011	1.325	9.285	0.870	5.210
INTKA	1.940	2.065	-0.798	11.643	2.653	15.879
IONA	2.645	1.704	1.931	6.729	1.382	8.271
IONE	-2.489	-0.560	-3.153	12.928	2.735	16.371
KALSK	4.060	3.135	2.956	6.360	1.182	7.074
KARELIA	-1.369	0.921	-2.437	11.011	1.472	8.811
KATSK	-1.395	2.639	-4.361	18.797	3.356	20.091
KEKROPS	3.037	0.904	3.647	5.395	0.635	3.799
KERAL	-0.735	2.541	-0.722	13.803	1.300	7.782
KERK	0.973	1.780	0.639	1.922	0.213	1.276
KORFK	0.278	2.125	1.223	6.137	-0.313	-1.873
KREKA	2.794	3.537	2.524	15.810	1.077	6.447
LABI	2.591	2.407	2.120	5.664	0.693	4.148
LAMDA	1.078	1.235	-0.006	7.811	1.360	8.142
LAMPSA	0.934	1.067	0.427	5.529	0.980	5.869
LOYLI	2.267	2.794	-0.186	8.442	1.513	9.054
LYKOS	2.411	5.305	-2.061	14.588	2.190	13.107
MAIK	2.886	3.918	-0.835	17.465	2.723	16.298
MAXIM	1.298	2.276	1.387	4.965	0.146	0.875
MARFIN	-3.251	2.043	-6.393	35.953	6.685	40.016
MEAGA	1.125	1.828	1.290	2.323	0.259	1.548
MESOX	0.818	1.672	0.080	1.032	0.459	2.747
METK	3.295	3.294	0.765	6.945	1.299	7.778
MHXK	2.923	3.115	1.438	13.047	1.867	11.176
MINE	0.944	1.873	-0.300	3.424	0.779	4.660
MOUZK	2.599	1.587	2.132	7.093	1.362	8.151
MOYLT	-0.264	-0.743	0.762	1.384	0.557	3.334
MOYR	0.132	0.706	1.094	0.462	-0.166	-0.991
MOXLOS	1.825	1.300	2.336	4.709	0.388	2.320
MPENK	1.278	1.879	0.534	8.379	1.235	7.392
MPOKA	1.306	-0.340	2.103	-2.347	0.407	2.439
MPSTK	1.780	2.747	-0.450	12.462	1.983	11.873
MPTK	3.475	1.705	3.238	6.530	1.408	8.431
MRFKO	4.790	5.769	3.477	17.729	1.338	8.011
MYTIL	4.927	4.138	3.213	7.997	1.182	7.074
NAOYK	0.881	4.783	1.256	7.611	0.350	2.097
NEL	2.317	3.676	0.733	7.667	0.845	5.059
NHR	2.438	2.844	2.202	14.697	1.623	9.715
NIKAS	1.295	4.102	-1.174	12.075	1.787	10.695
XYLK	-0.017	1.006	0.324	4.367	0.417	2.495
PAPAK	3.830	4.637	-0.871	18.199	3.052	18.272
PARN	0.797	1.588	0.597	-0.233	-0.061	-0.366
PEILH	5.677	4.687	2.641	16.815	1.992	11.924
PEIR	7.113	3.692	2.328	18.760	3.826	22.905
PEPA	-3.070	-2.220	-1.916	27.376	3.479	20.823
PETZK	-0.738	4.719	-3.199	11.743	1.654	9.898
PLATH	3.009	3.357	1.877	5.190	0.623	3.728
PLAS	-1.175	-0.735	-0.860	8.761	2.273	13.605
PRD	3.794	3.258	3.031	3.909	0.534	3.199
PROOD	-4.821	-0.212	-5.190	40.244	5.593	33.479
RILKE	6.630	4.324	5.481	6.654	0.893	5.345

RINTE	3.404	2.611	2.507	5.601	1.144	6.848
ROKKA	3.110	3.020	1.363	7.279	1.336	7.997
SAIKL	0.253	0.387	0.885	4.084	0.552	3.305
SANYO	2.890	2.432	2.371	3.526	0.640	3.829
SAR	2.900	3.712	1.114	8.956	1.131	6.770
SATOK	1.398	2.093	0.662	2.066	0.404	2.421
SELMK	3.395	1.559	3.262	7.396	1.516	9.075
SELO	0.589	2.316	-0.010	5.666	0.592	3.545
SIDE	4.356	3.425	2.110	9.696	1.605	9.608
STALK	0.518	0.496	1.063	-1.713	-0.009	-0.055
STRIK	3.144	4.167	0.894	11.093	1.171	7.013
TERNA	2.715	2.751	1.934	4.674	0.553	3.312
TEXN	-0.807	1.128	-0.697	-1.742	-0.290	-1.739
TZKA	3.580	1.994	3.938	15.574	1.144	6.849
THLET	2.924	2.580	0.882	7.140	1.713	10.257
TITK	3.784	1.006	4.354	22.151	3.964	23.729
TSIP	2.759	6.360	-0.197	22.662	2.784	16.663
YALKO	0.941	2.793	0.282	6.013	0.532	3.184
FANKO	0.609	4.947	0.744	6.443	0.002	0.014
FINTO	-0.833	0.469	-0.035	9.226	0.450	2.696
FOIN	-2.363	4.044	-2.341	34.797	3.118	18.665
FRLK	1.440	0.915	1.138	9.022	1.386	8.294
HALYB	2.022	2.759	1.236	2.131	0.302	1.809
XATZK	1.268	0.071	1.561	3.953	1.029	6.161

Table A5: Bera -Jarque Normality test

ABK	207.25	ETLES	64.94	XYLK	19.07
AVAX	93.27	ETMAK	16.03	PAPAK	345.88
AAAK	62.38	ZAMPA	153.87	PARN	0.69
ALTEC	80.55	HERACL	181.33	PEILH	314.98
ALFA	401.04	THEMEL	504.21	PEIR	402.53
ATTIK	3.74	IATR	142.36	PEPA	758.87
AKTOR	108.80	INTEK	139.88	PETZK	138.43
ALATK	22.20	INTEP	794.29	PLATH	35.99
AEGEP	48.36	INTER	62.73	PLAS	78.14
AXON	202.35	INTET	87.14	PRD	29.67
BIOSK	9.60	INTKA	139.33	PROOD	1642.79
BIOSP	19.52	IONA	52.27	RILKE	88.24
BIOT	50.70	IONE	173.33	RINTE	42.96
BIOXK	113.57	KALSK	56.93	ROKKA	62.65
BISK	11.33	KARELIA	123.12	SAIKL	16.74
BISP	146.96	KATSK	355.27	SANYO	20.78
GEBKA	34.73	KEKROPS	38.33	SAR	88.62
GEK	81.87	KERAL	191.06	SATOK	6.22
GENAK	75.30	KERK	4.64	SELMK	66.22
GOOD	800.35	KORFK	37.74	SELO	32.45
DIAS	34.34	KREKA	257.77	SIDE	112.99
DIEKA	22.46	LABI	38.80	STALK	3.20
DK	256.31	LAMDA	62.18	STRIK	132.93
EBZ	110.93	LAMPSA	31.45	TERNA	29.21
EGNAK	106.98	LOYLI	76.40	TEXN	3.69
EDRA	261.37	LYKOS	218.61	TZKA	255.36
EEEK	371.12	MAIK	313.35	THLET	59.53
EKTER	2.04	MAXIM	26.34	TITK	504.97
ELAIS	537.11	MARFIN	1303.19	TSIP	521.20
ELBE	44.71	MEAGA	6.66	YALKO	37.04
ELKA	95.32	MESOX	1.73	FANKO	41.88
ELL	623.55	METK	59.08	FINTO	85.82
ELMEK	88.30	MHXK	178.78	FOIN	1216.44
ELMPI	467.16	MINE	12.61	FRLK	83.47
ELTEX	110.99	MOUZK	57.06	HALYB	8.63
ELTK	39.90	MOYLT	1.99	XATZK	17.24
ELYF	47.28	MOYR	0.23		
ELFK	25.35	MOXLOS	25.50		
EMDKO	4.91	MPENK	71.84		
EMP	328.87	MPOKA	7.22		
EXEL	2990.72	MPSTK	158.47		
EPERA	635.92	MPTK	54.71		
EPILK	49.59	MRFKO	337.28		
ERGAS	3.30	MYTIL	88.22		
ERMES	231.25	NAOYK	58.70		
ESK	9.33	NEL	64.14		
ESHA	32.45	NHR	221.95		
ETE	412.93	NIKAS	147.48		

Table A6: Parameter Estimates of the Skewed Generalised –T Distribution

	k	n	σ	μ	λ	lnL	SBC
ABK	2.8255	7.9129	0.0081	0.0003	1.4137	6041.0	1553.5
AVAX	2.2544	3.6869	0.0210	0.0012	2.0451	7133.9	2646.4
AAAK	3.2609	4.7864	0.0305	0.0007	0.5769	5869.2	1381.7
ALTEC	3.8752	3.0649	0.0320	0.0004	0.2748	6786.9	2299.4
ALFA	3.8458	4.1843	0.0344	0.0002	1.9263	6862.6	2375.1
ATTIK	3.9957	6.6149	0.0344	-0.0004	0.2802	6926.2	2438.7
AKTOR	2.0794	5.2892	0.0139	0.0005	2.3235	7868.5	3381.0
ALATK	2.8771	2.9736	0.0166	-0.0003	0.1700	6713.9	2226.4
AEGEP	3.8024	6.8658	0.0258	-0.0003	-0.2743	5963.3	1475.8
AXON	2.3313	5.7255	0.0372	0.0000	1.2387	7362.3	2874.8
BIOSK	2.5474	3.4874	0.0311	-0.0001	1.3396	8017.8	3530.3
BIOSP	2.1423	5.1621	0.0154	0.0000	2.5147	8018.8	3531.3
BIOT	2.1157	8.2896	0.0303	0.0003	1.1453	7978.9	3491.4
BIOXK	2.3832	8.2385	0.0318	0.0005	2.6804	6630.1	2142.6
BISK	2.4958	3.5978	0.0233	-0.0001	-0.4289	6268.5	1781.0
BISP	2.0847	9.5500	0.0093	0.0000	3.0588	6358.9	1871.4
GEBKA	2.6286	9.6415	0.0401	0.0003	1.4334	6139.4	1651.9
GEK	2.7465	3.4699	0.0199	0.0001	0.3841	6815.3	2327.8
GENAK	3.1739	3.2143	0.0332	0.0011	1.4067	7915.3	3427.8
GOOD	2.7755	10.7604	0.0114	0.0006	0.7170	6636.8	2149.3
DIAS	2.8048	9.0034	0.0097	0.0004	2.1846	5922.3	1434.8
DIEKA	2.7719	4.0739	0.0372	0.0000	0.0118	6083.2	1595.7
DK	3.8979	8.1016	0.0381	-0.0002	0.3857	7932.9	3445.4
EBZ	2.9960	10.6456	0.0198	-0.0003	-0.3199	7465.6	2978.1
EGNAK	2.9151	7.8012	0.0374	-0.0003	1.8191	7899.5	3412.0
EDRA	2.6669	6.2970	0.0316	-0.0004	0.6240	7325.3	2837.8
EEEK	4.0307	6.4729	0.0190	0.0001	-0.0189	7234.5	2747.0
EKTER	3.6875	6.3344	0.0327	-0.0003	0.3017	8142.6	3655.1
ELAIS	4.0622	4.7872	0.0211	0.0000	1.5387	6836.0	2348.5
ELBE	2.5702	4.8095	0.0122	0.0001	0.9274	7617.9	3130.4
ELKA	3.9819	11.4672	0.0186	-0.0002	2.5964	7644.1	3156.6
ELL	2.1585	5.6106	0.0283	0.0011	1.6223	8024.0	3536.5
ELMEK	3.4812	11.1114	0.0242	0.0004	1.8495	6195.0	1707.5
ELMPI	3.7075	4.4029	0.0112	0.0007	1.0624	7993.2	3505.7
ELTEX	4.0242	9.9571	0.0272	0.0005	1.0153	7528.3	3040.8
ELTK	2.9966	5.4615	0.0276	0.0003	0.1751	6385.7	1898.2
ELYF	2.6594	4.0702	0.0258	-0.0002	2.9580	6542.5	2055.0
ELFK	3.5729	9.4863	0.0168	0.0002	0.0483	7875.1	3387.6
EMDKO	2.7635	3.7192	0.0192	-0.0003	1.0827	6480.0	1992.5
EMP	3.6246	11.3843	0.0389	0.0003	0.6584	7159.0	2671.5
EXEL	2.2009	7.7128	0.0228	0.0007	2.3321	6433.0	1945.5
EPERA	2.4528	10.9840	0.0224	0.0003	1.7374	6428.8	1941.3
EPILK	2.1800	6.8629	0.0113	-0.0004	0.6560	6504.5	2017.0
ERGAS	2.7774	7.5123	0.0374	-0.0014	1.2755	7058.5	2571.0
ERMES	3.0387	3.1351	0.0294	0.0014	2.7322	6300.7	1813.2
ESK	3.0878	3.1157	0.0161	0.0005	0.2734	7207.8	2720.3
ESHA	2.8081	5.1592	0.0317	-0.0001	0.9259	8242.5	3755.0
ETE	4.0514	6.6910	0.0130	0.0005	2.0894	5987.0	1499.5
ETLES	2.1326	11.2666	0.0372	-0.0005	2.1656	7121.9	2634.4
ETMAK	2.5184	6.7004	0.0206	-0.0005	1.5871	6461.9	1974.4
ZAMPA	2.0601	7.4599	0.0367	0.0001	0.6242	6929.4	2441.9
HERACL	3.9301	4.6658	0.0388	0.0000	2.8334	5914.5	1427.0

THEMEL	2.2536	8.0303	0.0166	0.0001	0.9084	6693.3	2205.8
IATR	2.5711	11.2081	0.0092	0.0001	0.9630	5840.5	1353.0
INTEK	3.6246	5.8652	0.0341	-0.0004	1.8979	6338.8	1851.3
INTEP	3.4297	10.4602	0.0243	0.0003	2.1221	7548.5	3061.0
INTER	3.6925	6.0901	0.0235	0.0000	0.6867	7466.1	2978.6
INTET	3.5204	9.7750	0.0083	0.0008	2.3316	7678.9	3191.4
INTKA	2.2227	9.1911	0.0137	0.0002	0.2100	6126.5	1639.0
IONA	2.3185	11.4706	0.0099	0.0005	0.9074	7951.3	3463.8
IONE	3.5850	6.0146	0.0234	0.0000	2.0453	6751.5	2264.0
KALSK	3.3218	5.0446	0.0111	-0.0002	2.9589	6496.9	2009.4
KARELIA	2.4025	9.3455	0.0105	0.0004	1.6021	7930.5	3443.0
KATSK	2.8717	10.9903	0.0326	0.0001	0.2817	5960.4	1472.9
KEKROPS	3.1712	5.0415	0.0316	0.0013	0.4709	7875.6	3388.1
KERAL	3.4944	5.4256	0.0121	0.0008	0.0698	6356.4	1868.9
KERK	3.1770	10.8015	0.0371	-0.0003	-0.3579	7960.0	3472.5
KORFK	2.4178	5.2843	0.0151	-0.0002	0.1142	6244.8	1757.3
KREKA	4.0197	6.7961	0.0143	-0.0005	1.8615	5850.9	1363.4
LABI	3.4445	9.8942	0.0231	-0.0002	1.3174	6247.4	1759.9
LAMDA	3.1235	8.5491	0.0077	0.0008	2.7114	6646.9	2159.4
LAMPSA	3.6673	3.5640	0.0143	0.0011	-0.4592	6677.5	2190.0
LOYLI	3.6855	5.3925	0.0169	0.0002	-0.3319	5856.4	1368.9
LYKOS	2.5823	9.5199	0.0117	0.0001	0.9911	7791.5	3304.0
MAIK	2.4112	10.3569	0.0270	0.0008	0.2733	8021.3	3533.8
MAXIM	3.8095	9.5246	0.0114	-0.0002	2.1549	6676.2	2188.7
MARFIN	2.2831	7.5341	0.0296	0.0000	-0.1811	5915.7	1428.2
MEAGA	2.1708	3.8435	0.0230	-0.0004	2.1599	7144.9	2657.4
MESOX	3.5960	5.8966	0.0293	-0.0004	2.4606	7349.8	2862.3
METK	3.5489	7.6636	0.0123	0.0007	1.4409	6404.3	1916.8
MHXK	4.0522	5.9100	0.0237	-0.0006	2.9946	6992.8	2505.3
MINE	3.9290	8.7455	0.0256	0.0002	2.9618	8139.1	3651.6
MOUZK	3.8848	4.3394	0.0097	-0.0002	1.5799	6228.9	1741.4
MOYLT	3.1563	10.4859	0.0094	-0.0004	0.2542	7382.6	2895.1
MOYR	3.0666	10.4563	0.0369	-0.0004	1.2421	6538.3	2050.8
MOXLOS	3.4202	10.9806	0.0105	0.0000	1.7557	6224.3	1736.8
MPENK	3.0443	8.6162	0.0346	0.0001	0.7935	6854.9	2367.4
MPOKA	2.3459	3.6502	0.0397	0.0001	-0.1484	6579.8	2092.3
MPSTK	2.1271	7.8274	0.0383	0.0002	0.4185	6669.0	2181.5
MPTK	3.6664	9.2331	0.0159	-0.0001	1.2271	7357.1	2869.6
MRFKO	3.4133	4.9946	0.0349	-0.0001	2.2563	7415.1	2927.6
MYTIL	3.5353	8.3756	0.0209	0.0003	1.3569	6851.2	2363.7
NAOYK	3.2366	7.1875	0.0203	-0.0005	1.6906	7800.5	3313.0
NEL	2.3590	10.9633	0.0397	-0.0004	0.6428	6754.5	2267.0
NHR	3.8611	7.9737	0.0079	-0.0008	1.6855	6992.8	2505.3
NIKAS	2.8170	3.6678	0.0386	0.0003	2.4832	6226.3	1738.8
XYLK	2.4570	8.3032	0.0140	0.0005	0.9011	6595.8	2108.3
PAPAK	2.4677	4.6433	0.0144	0.0002	0.0280	6444.7	1957.2
PARN	2.7279	7.2127	0.0295	-0.0013	1.7654	7436.9	2949.4
PEILH	2.7100	5.0011	0.0281	0.0007	-0.3619	6162.5	1675.0
PEIR	2.6594	7.1044	0.0078	0.0008	0.9089	6008.4	1520.9
PEPA	3.6784	8.5790	0.0383	0.0003	1.8205	6389.4	1901.9
PETZK	3.4631	11.1159	0.0325	-0.0003	0.3607	5802.1	1314.6
PLATH	2.6011	7.6512	0.0264	0.0002	0.0741	6194.0	1706.5
PLAS	3.8852	3.5396	0.0338	-0.0009	1.0731	7004.1	2516.6
PRD	2.1294	9.7576	0.0366	-0.0007	1.0127	6219.7	1732.2
PROOD	3.4893	8.0235	0.0247	0.0002	2.3432	6699.9	2212.4
RILKE	2.9710	5.4127	0.0396	-0.0001	-0.3917	7844.7	3357.2
RINTE	3.0988	6.6325	0.0347	0.0004	2.6691	6985.7	2498.2

ROKKA	2.5708	5.5918	0.0157	0.0001	0.2997	6571.9	2084.4
SAIKL	2.6413	5.2664	0.0341	-0.0007	0.0844	6278.8	1791.3
SANYO	3.6784	7.6546	0.0290	-0.0001	-0.2652	7553.6	3066.1
SAR	3.6517	10.1587	0.0086	-0.0002	1.6853	6539.7	2052.2
SATOK	3.4222	4.9432	0.0158	-0.0004	1.2033	7636.2	3148.7
SELMK	3.5833	9.9868	0.0182	-0.0004	0.5386	6740.7	2253.2
SELO	3.9755	7.7367	0.0348	-0.0009	1.4547	6411.9	1924.4
SIDE	3.3074	11.1578	0.0119	0.0001	0.5097	7640.7	3153.2
STALK	3.5158	3.7294	0.0246	-0.0015	0.4070	6229.3	1741.8
STRIK	4.0151	8.8996	0.0303	-0.0005	0.2903	8184.9	3697.4
TERNA	2.7983	6.1399	0.0350	0.0003	2.3253	6600.8	2113.3
TEXN	3.7764	10.3394	0.0159	-0.0017	2.6750	7393.1	2905.6
TZKA	3.1808	6.2498	0.0257	-0.0002	0.5177	6249.9	1762.4
THLET	3.8223	8.4736	0.0085	0.0008	1.0483	7118.1	2630.6
TITK	2.9453	6.9675	0.0281	0.0010	-0.4247	7547.6	3060.1
TSIP	2.4920	6.4705	0.0327	0.0006	0.3297	7411.1	2923.6
YALKO	3.7938	4.4480	0.0096	0.0000	2.6036	6865.7	2378.2
FANKO	2.6191	3.7216	0.0163	-0.0006	-0.2769	7558.0	3070.5
FINTO	3.4777	5.2725	0.0374	0.0006	0.3074	6185.1	1697.6
FOIN	3.4860	9.6997	0.0397	0.0001	2.3271	5953.4	1465.9
FRLK	2.8129	3.4337	0.0224	-0.0002	0.6463	7936.3	3448.8
HALYB	2.7193	5.7947	0.0332	-0.0003	1.8868	7540.6	3053.1
XATZK	2.2245	4.1279	0.0217	-0.0002	1.5912	6938.5	2451.0

Table A7: Parameter Estimates of the Stable Distribution

	α	lnL	SBC		α	lnL	SBC
ABK	1.42	5881.53	1453.02	KERK	1.57	7589.34	3238.99
AVAX	1.18	7049.74	2593.37	KORFK	1.25	6559.14	1955.32
AAAK	1.60	5610.00	1218.41	KREKA	1.92	5744.90	1296.63
ALTEC	1.86	6491.63	2113.36	LABI	1.68	6076.46	1652.21
ALFA	1.84	6764.25	2313.13	LAMDA	1.54	6642.20	2156.44
ATTIK	1.91	6631.60	2253.10	LAMPSA	1.77	6294.77	1948.88
AKTOR	1.11	7817.53	3348.89	LOYLI	1.78	6224.01	1600.51
ALATK	1.44	6406.18	2032.54	LYKOS	1.32	7581.67	3171.81
AEGEP	1.83	5602.59	1248.55	MAIK	1.25	7725.88	3347.69
AXON	1.22	7182.04	2761.25	MAXIM	1.83	6747.31	2233.52
BIOSK	1.31	7849.56	3424.33	MARFIN	1.20	5566.10	1207.95
BIOSP	1.14	7990.65	3513.57	MEAGA	1.15	7074.40	2612.97
BIOT	1.13	7787.45	3370.78	MESOX	1.74	7315.16	2840.47
BIOXK	1.24	6621.74	2137.35	METK	1.72	6248.05	1818.35
BISK	1.28	5889.38	1542.16	MHXK	1.93	7021.87	2523.62
BISP	1.11	6395.61	1894.53	MINE	1.88	8164.20	3667.40
GEBKA	1.34	5982.31	1552.93	MOUZK	1.86	6089.26	1653.43
GEK	1.39	6533.08	2150.00	MOYLT	1.56	7084.88	2707.53
GENAK	1.57	7755.07	3326.87	MOYR	1.52	6718.24	2164.17
GOOD	1.40	6394.32	1996.55	MOXLOS	1.67	6105.64	1662.05
DIAS	1.41	5854.79	1392.27	MPENK	1.51	6621.52	2220.38
DIEKA	1.40	5756.56	1389.90	MPOKA	1.22	6234.07	1874.48
DK	1.86	7650.93	3267.78	MPSTK	1.13	6390.92	2006.32
EBZ	1.49	7099.45	2747.42	MPTK	1.77	7175.39	2755.11
EGNAK	1.46	8010.59	3481.98	MRFKO	1.66	7356.08	2890.40
EDRA	1.36	7071.75	2678.08	MYTIL	1.71	6685.01	2259.01
EEEK	1.92	6904.26	2538.95	NAOYK	1.59	7674.10	3233.37
EKTER	1.78	7850.57	3471.12	NEL	1.23	6503.16	2108.66
ELAIS	1.93	6691.49	2257.47	NHR	1.85	6865.79	2425.29
ELBE	1.32	7400.44	2993.39	NIKAS	1.42	6194.34	1718.65
ELKA	1.90	7625.69	3145.00	XYLK	1.27	6816.34	2247.23
ELL	1.14	7889.44	3451.73	PAPAK	1.27	6120.05	1752.67
ELMEK	1.69	6087.49	1639.76	PARN	1.38	7319.43	2875.41
ELMPI	1.79	7791.85	3378.84	PEILH	1.37	5791.34	1441.17
ELTEX	1.92	7321.37	2910.43	PEIR	1.35	5788.78	1382.55
ELTK	1.49	6078.55	1704.69	PEPA	1.77	6278.42	1831.97
ELYF	1.35	6567.23	2070.59	PETZK	1.68	6087.14	1494.19
ELFK	1.73	7552.88	3184.61	PLATH	1.33	5874.83	1505.43
EMDKO	1.40	6678.33	2117.44	PLAS	1.86	6804.05	2390.57
EMP	1.75	6909.51	2514.33	PRD	1.13	6012.46	1601.64
EXEL	1.16	6383.03	1914.01	PROOD	1.70	6651.31	2181.80
EPERA	1.27	6307.99	1865.21	RILKE	1.48	8219.36	3593.22
EPILK	1.15	6254.70	1859.62	RINTE	1.53	6975.92	2492.03
ERGAS	1.40	6882.63	2460.22	ROKKA	1.32	6279.60	1900.24
ERMES	1.51	6298.47	1811.79	SAIKL	1.34	5960.89	1591.02
ESK	1.53	7503.22	2906.42	SANYO	1.77	7913.26	3292.71
ESHA	1.41	8024.93	3617.94	SAR	1.76	6412.62	1972.14
ETE	1.93	5908.11	1449.79	SATOK	1.67	7451.68	3032.46
ETLES	1.13	7052.15	2590.47	SELMK	1.73	6476.90	2087.00
ETMAK	1.29	6323.16	1887.00	SELO	1.90	6257.37	1827.05
ZAMPA	1.10	6675.86	2282.18	SIDE	1.62	7373.42	2984.80
HERACL	1.88	5924.32	1433.18	STALK	1.71	5949.87	1565.78
THEMEL	1.18	6473.62	2067.40	STRIK	1.91	7891.54	3512.60

IATR	1.32	5627.32	1218.69	TERNA	1.41	6550.06	2081.34
INTEK	1.75	6237.05	1787.18	TEXN	1.81	7384.04	2899.89
INTEP	1.67	7473.48	3013.72	TZKA	1.57	6516.15	1930.13
INTER	1.78	7220.00	2823.56	THLET	1.83	6915.10	2502.71
INTET	1.71	7728.85	3222.85	TITK	1.47	7168.92	2821.52
INTKA	1.17	5823.58	1448.18	TSIP	1.28	7699.84	3105.52
IONA	1.21	7731.48	3325.31	YALKO	1.82	6848.16	2367.17
IONE	1.74	6667.38	2211.01	FANKO	1.34	7196.99	2843.07
KALSK	1.63	6521.69	2025.02	FINTO	1.69	5893.70	1514.00
KARELIA	1.25	7793.56	3356.74	FOIN	1.69	5902.90	1434.10
KATSK	1.44	5665.95	1287.38	FRLK	1.42	7685.39	3290.74
KEKROPS	1.56	7603.79	3216.88	HALYB	1.38	7437.61	2988.23
KERAL	1.70	6036.67	1667.45	XATZK	1.17	6800.18	2363.84

Table A8: Estimates of the Tail Parameter

Company	ξ	Company	ξ	Company	ξ
ABK	0.3647	ESK	0.2913	MRFKO	0.4192
AVAX	0.3981	ESHA	0.3264	MYTIL	0.3437
AAAK	0.2590	ETE	0.4041	NAOYK	0.3487
ALTEC	0.4003	ETLES	0.3654	NEL	0.3717
ALFA	0.3724	ETMAK	0.2248	NHR	0.3512
ATTIK	0.3266	ZAMPA	0.3448	NIKAS	0.3910
AKTOR	0.3758	HERACL	0.3889	XYLK	0.3531
ALATK	0.3377	THEMEL	0.4477	PAPAK	0.4208
AEGEP	0.3885	IATR	0.3695	PARN	0.2533
AXON	0.3840	INTEK	0.3415	PEILH	0.4281
BIOSK	0.2119	INTEP	0.4420	PEIR	0.4443
BIOSP	0.2107	INTER	0.3577	PEPA	0.3982
BIOT	0.3542	INTET	0.3711	PETZK	0.3573
BIOXK	0.3833	INTKA	0.3462	PLATH	0.3552
BISK	0.3107	IONA	0.3709	PLAS	0.3291
BISP	0.3397	IONE	0.3363	PRD	0.3105
GEBKA	0.3359	KALSK	0.3725	PROOD	0.4814
GEK	0.3781	KARELIA	0.3450	RILKE	0.3330
GENAK	0.3846	KATSK	0.4117	RINTE	0.3184
GOOD	0.4177	KEKROPS	0.2958	ROKKA	0.3668
DIAS	0.3356	KERAL	0.4182	SAIKL	0.2965
DIEKA	0.3380	KERK	0.2735	SANYO	0.3235
DK	0.4234	KORFK	0.2264	SAR	0.3281
EBZ	0.3572	KREKA	0.3644	SATOK	0.3026
EGNAK	0.3595	LABI	0.3195	SELMK	0.3313
EDRA	0.4920	LAMDA	0.3845	SELO	0.3649
EEEK	0.4136	LAMPSA	0.3950	SIDE	0.3361
EKTER	0.3055	LOYLI	0.3808	STALK	0.2176
ELAIS	0.4371	LYKOS	0.3914	STRIK	0.3778
ELBE	0.3336	MAIK	0.3977	TERNA	0.3302
ELKA	0.3845	MAXIM	0.3621	TEXN	0.1917
ELL	0.4813	MARFIN	0.4758	TZKA	0.3936
ELMEK	0.3960	MEAGA	0.2717	THLET	0.4012
ELMPI	0.4882	MESOX	0.2983	TITK	0.4362
ELTEX	0.4006	METK	0.3709	TSIP	0.3858
ELTK	0.3227	MHXK	0.4536	YALKO	0.3475
ELYF	0.3582	MINE	0.3113	FANKO	0.3177
ELFK	0.2565	MOUZK	0.3747	FINTO	0.3556
EMDKO	0.2324	MOYLT	0.1994	FOIN	0.5035
EMP	0.4909	MOYR	0.2681	FRLK	0.3660
EXEL	0.3513	MOXLOS	0.3233	HALYB	0.3303
EPERA	0.4201	MPENK	0.3848	XATZK	0.3885
EPILK	0.3044	MPOKA	0.2126		
ERGAS	0.2890	MPSTK	0.4375		
ERMES	0.4506	MPTK	0.3366		

Chapter 5 - Evaluation of the Forecasting Performance of Alternative Value- at-Risk Models.

5.1 Introduction

Risk measurement using Value at Risk entails the estimation of quantiles or tail probabilities of a distribution. There are many approaches to the estimation of tail probabilities. The most common approach is to assume that the loss distribution function has a particular form, e.g. normal, student, hyperbolic, stable etc whose parameters can be estimated from a particular sample. Since this approach allows a complete characterisation of the distribution, then tail probability estimation is quite simple. A more parsimonious approach compared to the first approach is to modify the quantiles of the normal distribution assuming either time varying volatility or using Fisher Cornish type modifications. A host of models are derived as a result of the different volatility models.

Extreme value theory is also a parametric method but it differs from the above approaches in the sense that it deals with the statistical properties of sample observations that are in the tails of a distribution. Extreme value theory uses asymptotic results, much in the same way as the central limit theory to derive the distribution function of extreme observations. These results are derived without any reference to the distribution function that may characterise the entire sample of random variables.

The purpose of this chapter is to evaluate the forecasting performance of the various approaches to estimating VaR. The approach we adopt is to evaluate the out-of –sample performance of the various models as the primary purpose of a VaR model is to forecast future losses.

The chapter is organised as follows. In Section 1 we present a framework for the evaluation of forecast performance. Kupiec (1995) suggested a statistic for evaluating the performance of a model based on the proportion of prediction failure. Evaluation based on the proportion of correct forecasts ignores conditioning or time variation in the data which may introduce serial correlation in the prediction failure. Christoffersen (1998) developed a test that takes this aspect into account when a model is evaluated and this is the test we have used in this chapter.

In Section 2 we review the data we have used to compare the performance of the alternative models. More specifically we explore the statistical properties of returns of the Athens Stock Exchange General Price Index. We test for normality and compare three distributions in terms of quantile prediction, namely the normal, the Laplace and the Student-Distribution. Estimation of the density function is useful on its own right but it also gives us insights as to the right density for the error term when we model the dynamics of the returns.

In Section 3 we estimate the tails of the distribution using the Pareto model paying particular emphasis to the specification of the correct threshold for the characterisation of the extreme observations.

In Section 4 we tackle the issue of heteroskedasticity in returns by modelling the volatility dynamics through a GARCH model. The implications of applying extreme value theory to the heteroskedasticity - corrected series are further explored by calculating the Value-at-Risk for the standardised residuals.

In Section 5 we present the specification of the alternative models that we compare. The approach we have adopted is to estimate each model over a certain sample and then to forecast the next day VaR producing a total of 2000 forecasts. In total we compare eight different models which are differentiated in terms of their distributional assumptions and in terms of their conditioning methods. The first group of models assumes that the parameters remain fixed within the sample on which estimation is conditioned. In the second group of models we assume that volatility varies over time and is not fixed within the sample.

In Section 6 we present the results of the comparison using the metrics we have described in Section 2, and finally in Section 7 we present the conclusions to this chapter.

5.2 A Framework for Assessing the Forecasting Performance of Alternative VaR Models

The basic problem in market risk measurement is how on the basis of information up to time t we can form a prediction of the Value-at-Risk for the following day, i.e.

$$\hat{VaR}_{t+1/t}(p) = F^{-1}_{t+1/t}(p) \tag{5.1.1}$$

Where $F^{-1}_{t+1/t}(p)$ is the predicted pth quantile and is the inverse of the distribution function. In the special case where the first two moments of the distribution are finite the predicted VaR may be expressed in terms of the predicted moments as follows

$$\hat{VaR}_{t+1/t}(p) = \mu_{t+1/t} + F_z^{-1}(p) \times \hat{\sigma}_{t+1/t} \tag{5.1.2}$$

where $\mu_{t+1/t}$, $\hat{\sigma}_{t+1/t}$ are the predicted values of the mean and standard deviation based on the information up to time t and $F_z^{-1}(p)$ is the pth quantile of the random variable $z = (R - \mu)/\sigma$. The models we shall employ will differ in terms of whether we estimate directly the quantile function as in (5.1.1) or whether we are assuming a particular standardised distribution function and we estimate the moments (i.e. the standard deviation since the mean is close to zero and of little consequence). A large number of models can be examined in this way depending on the assumed distribution function or on the way that volatility is modelled.

The evaluation of the predictive performance of any forecasting system requires the adoption of a metric. In order to introduce the appropriate framework, consider the hit variable I_{t+1} which takes the value of 1 when the predicted VaR is below the actual loss, i.e. when the loss exceeds the predicted VaR and zero if it does not.

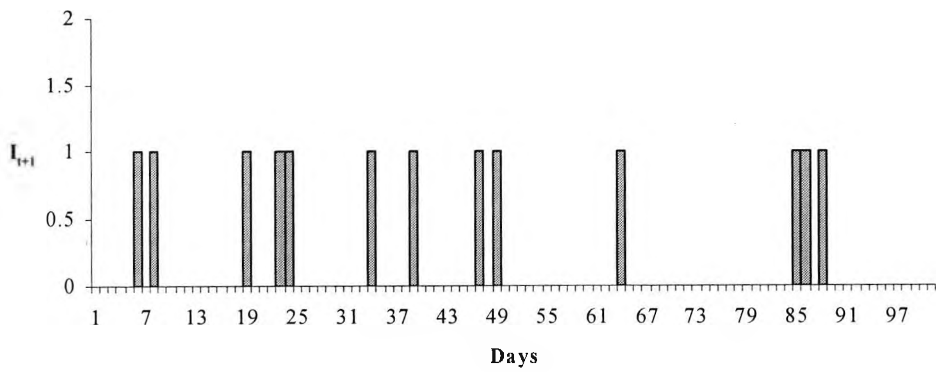
$$I_{t+1} = \begin{cases} 1 & if\ r_{t+1} < VaR_{t+1/t} \\ 0 & otherwise \end{cases}$$

If the perfect forecast VaR forecast model was used, then the forecasting error $\eta_{t+1} = r_{t+1} - VaR_{t+1/t}$ would be unpredictable in the sense that we would not be able at time t to make a prediction as to whether the predicted VaR level would be violated. Our forecast of a VaR violation should be 100p% every day. If we could predict violations, then that information could be used to formulate a better VaR prediction level. The assumption of unpredictable forecasting errors implies that the sequence $\{I_{t+1}\}$ should be independently and identically distributed as a Bernoulli random variable which takes values.

$$I_{t+1} = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

An example of a particular realisation of the hit variable when $p = 0.10$ is given in Figure 5.1

Figure 5.1: Hit Variable values



If the probability of a violation is p, the Bernoulli density function is given by

$$f(I_{t+1}; p) = (1-p)^{1-I_{t+1}} p^{I_{t+1}}$$

Christoffersen (1998) presents a number of tests for the evaluation of the forecasting performance of VaR model. The first test which he calls the unconditional coverage test, tests if the fraction of violations obtained for a particular VaR forecasting model, denoted by π is significantly different from the predicted fraction of violations p. The likelihood function of an iid Bernoulli random variable with probability of violation equal to π is given by

$$L(\pi) = \prod_{t=1}^{T} (1-\pi)^{1-I_{t+1}} \pi^{I_{t+1}} = (1-\pi)^{T_0} \pi^{T_1}$$

where

T_1=Number of 1's in the indicator series

T_0 = Number of 0's in the indicators series

The Maximum Likelihood Estimator of π is $\hat{\pi} = \frac{T_1}{T_0 + T_1}$ which is just the observed proportion of violations in the sample sequence.

The value of the likelihood function given the estimates $\hat{\pi}$ is therefore

$$L(\hat{\pi}) = (1-\hat{\pi})^{T_0} \hat{\pi}^{T_1} = \left(1 - \frac{T_1}{T_0 + T_1}\right)^{T_0} \left(\frac{T_1}{T_0 + T_1}\right)^{T_1}$$

Under the null hypothesis that $\pi = p$ the likelihood function is given by

$$L(p) = \prod_{t=1}^{T} (1-p)^{1-I_{t+1}} p^{I_{t+1}} = (1-p)^{T_0} p^{T_1}$$

The null hypothesis can therefore be tested using the likelihood ratio test

$$LR_{UC} = -2\ln\left[\frac{L(p)}{L(\hat{\pi})}\right]$$

$$= -2\ln\left[\frac{p^{T_1}(1-p)^{T_0}}{\hat{\pi}^{T_1}(1-\hat{\pi})^{T_0}}\right] \sim \chi^2_{(1)}$$

$$= -2\ln\left[p^{T_1}(1-p)^{T_0}\right] + 2\ln\left\{\left(1-\frac{T_1}{T_0+T_1}\right)^{T_0}\left(\frac{T_1}{T_0+T_1}\right)^{T_1}\right\}$$

Asymptotically, the Likelihood Ratio Statistic follows a chi-square distribution with 1 degree of freedom. That is $LR_{UC} \sim \chi^2_{(1)}$

The unconditional coverage test is equivalent to the test suggested by Kupiec (1995). However, the fact that the number of observed violations is close to the expected number is not sufficient to classify a model as a good model. A second desirable property of a good VaR forecasting system should be the independence of violations. That is, a good model should not produce violations which are correlated, since a correlated sequence of violations increases the risk of a position for a financial institution. Christoffersen (1998) has proposed a test that rejects a VaR model with clustered violations.

In order to derive the test let us define the probability of a particular combination of values for two successive periods say t and t+1 as

$$\pi_{ij} = P(I_t = i / I_{t+1} = j) \text{ for } (i, j = 0,1)$$

the four possible combinations are

$$\pi_{11} = P(I_t = 1 / I_{t+1} = 1)$$

$$\pi_{10} = P(I_t = 1 / I_{t+1} = 0)$$

$$\pi_{01} = P(I_t = 0 / I_{t+1} = 1)$$

$$\pi_{00} = P(I_t = 0 / I_{t+1} = 0)$$

Similarly we define T_{ij}, for $(i, j = 0,1)$, the number of observations with a j following an i. The likelihood function of T observations is given by

$$L(\pi_{01}, \pi_{11}) = (1-\pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1-\pi_{11})^{T_{10}} \pi_{11}^{T_{11}}$$

The Maximum Likelihood Estimator of the parameters is obtained by setting the first derivatives of the likelihood function equal to zero and solving for the parameters. The estimates are given by

$$\hat{\pi}_{01} = \frac{T_{01}}{T_{01}+T_{00}}$$

$$\hat{\pi}_{11} = \frac{T_{11}}{T_{10}+T_{11}}$$

the other probabilities can be estimated using the identities

$$\hat{\pi}_{00} = 1 - \hat{\pi}_{01}$$

$$\hat{\pi}_{10} = 1 - \hat{\pi}_{11}$$

Allowing for dependence in the violations is equivalent to allowing π_{01} to be different from π_{11}. If on the other hand the violations are independent we should have $\pi_{01} = \pi_{11} = \pi$

The Likelihood Ratio for the test of independence is given by

$$LR_{IND} = -2\ln\left[\frac{\hat{\pi}^{T_1}(1-\hat{\pi})^{T_0}}{(1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{T_{10}}\hat{\pi}_{11}^{T_{11}}}\right] \sim \chi^2_{(1)}$$

where $\hat{\pi}$ are the estimates obtained for the unconditional coverage test. High values of LR_{CC} means that the forecasting errors are not independent.

Christoffersen (1998) has also developed a test for the joint hypothesis of unconditional coverage and independence. The Likelihood Ratio statistic for the test of "correct conditional coverage" is given by

$$LR_{CC} = -2\ln\left[\frac{p^{T_1}(1-p)^{T_0}}{(1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{T_{10}}\hat{\pi}_{11}^{T_{11}}}\right] \sim \chi^2_{(2)}$$

The test of correct conditional coverage uses the null hypothesis from the unconditional test and the alternative hypothesis from the independence test. If we condition on the first observation, then these Likelihood ratio statistics are related by the identity

$$LR_{CC} = LR_{UC} + LR_{IND}$$

Christoffersen's basic framework is limited in that it only deals with first order dependence in the $\{I_t\}$ series. It would fail to reject an $\{I_t\}$ series which does not have first order Markov dependence but does exhibit some other kind of dependence structure (e.g. higher order Markov dependence or periodic dependence). Recently Christoffersen and Diebold (2000) have generalised this approach and suggested that a regression of the $\{I_t\}$ series on its own lagged values and some other variables of interest, such as day-dummies or the lagged observed returns, can be used to test for the existence of various form of dependence structures that may be present in the $\{I_t\}$ series. Under this framework, conditional efficiency of the $\{I_t\}$ process can be tested by testing the joint hypothesis :

$$H: \theta = 0, \alpha_0 = p$$

where $\theta = [\alpha_1, ..., \alpha_s, \mu_1, ..., \mu_s]'$

in the regression

$$I_t = \alpha_0 + \sum_{s=1}^{S-1}\alpha_s I_{t-s} + \sum_{s=1}^{S-1}\mu_s D_{s,t} + \varepsilon_t$$

for $t = S+1, S+2, ...T$ and where $D_{s,t}$ are the explanatory variables.

5.3 Data Description and Estimates of the Unconditional Density Function

The data employed in this chapter to asses the performance of the various VaR models is the Athens Stock Exchange All Share Price Index. We have used daily data over the period 2/1/1989-24/4/2003. Throughout the study we use logarithmic daily returns defined as $r_t = \ln S_t - \ln S_{t-1}$ where S_t and S_{t-1} denote closing prices for day t and $t-1$ respectively.

The purpose of this section is to investigate the statistical properties of the return series for the index. The results of the statistical analysis are summarised in Table 5.1 whereas the actual distribution of returns compared to the normal distribution is shown in Figure 5.2.

The average daily return is not statistically different from zero as it was expected. The usual assumption of zero daily return for the calculation of VaR is therefore justified. The daily volatility is on average around 1.8% which completely dominates the daily return.

Table 5.1: Descriptive Statistics for the General Index Return Series

Statistic	Sample Estimate	Standard Error
Daily mean	0.0005	0.0003
Volatility	0.0184	
Skewness	0.0144	0.041
Kurtosis	4.862	0.082
Max Value	0.14	
Min Value	-0.11	
Number of observations	3559	
Bera-Jarque test	3505	

The coefficient of skewness of the series is slightly positive. Under the null hypothesis of normality, the skewness coefficient is normally distributed with standard error $se(\gamma_3) = \sqrt{6/T}$. A test of the hypothesis $H_0 : \gamma_3 = 0$ against $H_1 : \gamma_3 \neq 0$ can be performed using the statistic

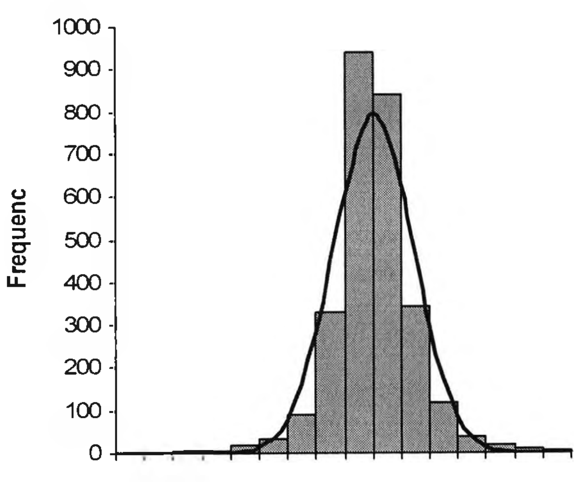
$$\frac{\hat{\gamma}_3}{se(\hat{\gamma}_3)} \sim N(0,1)$$

The value of the test statistic was 0.35 indicating that the coefficient of skewness is not significantly different from zero. However the coefficient of excess kurtosis, using the test statistic

$$\frac{\hat{\gamma}_4}{se(\hat{\gamma}_4)} \sim N(0,1)$$

where $se(\hat{\gamma}_4) = \sqrt{24/T}$ was found statistically different from zero since the test statistic was 59.3 . So the result for kurtosis implies that the r_t series exhibit excess kurtosis beyond that of the normal distribution.

Figure 5.2: Frequency Distribution of Daily Returns of the General Index



A visual inspection of the frequency distribution shown in Figure 5.2 provides further evidence that the normal distribution may not be a good fit.

To test whether the distribution is normal we have also used the Bera-Jarque (BJ) statistic for detecting departures from normality which is calculated using the formula

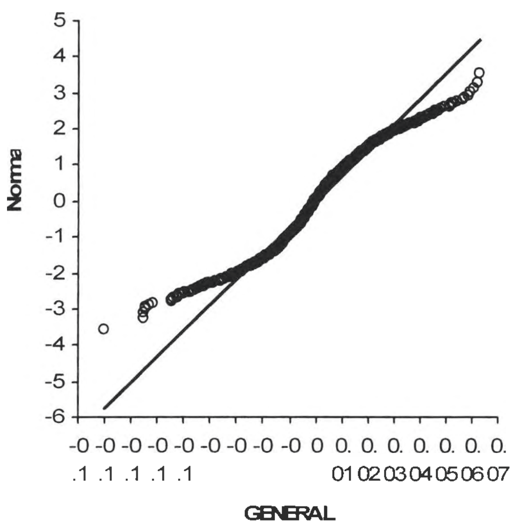
$$BJ = T(\hat{\gamma}_3^2 / 6 + \hat{\gamma}_4^2 / 24).$$

This is a test of the joint rejection of both zero skewness and zero excess kurtosis. Under the null hypothesis of normality, the BJ statistic is distributed as $\chi^2(2)$ with 2 degrees of freedom. The value of the test statistic is 3505 which is significantly greater than their critical value of 5.99 at the 95 percent confidence interval, accepting the hypothesis that the r_t series are non-normal.

We have also employed QQ plots to test the normality hypothesis. The plot is shown in Figure 5.3. The Q-Q plot is linear for a normally distributed random variable, but in our case it obvious from the graph that the high and low quantiles do not behave like the normal distribution.

The statistical analysis of the general index returns for the Athens Stock Exchange show conclusively that the daily return series exhibit excess kurtosis but not skewness. The evidence from the empirical results is that the assumption of normality is not supported by the data over the time period we have studied. That means that the tails of the distribution are significantly fatter than those predicted by the normal distribution.

Figure 5.3 QQ Plot for the general index return



Given this findings the obvious next step is to try and fit alternative density probability functions and we do this by using two symmetric distributions. We have restricted the range of parametric distributions we have used to symmetric only distributions because of the conclusive evidence against the existence of skewness in the return series. Thus the alternative distributions we have used to the Normal distribution are the Student-t distribution, and the Laplace distribution.

The Laplace density function is given by

$$p(r / \theta, \phi) = \frac{1}{2\phi} e^{-|r-\theta|/\phi} \quad \text{with } -\infty < r < \infty, \phi > 0$$

whereas its cumulative distribution function is

$$F(r) = \begin{cases} \frac{1}{2\phi} e^{-\frac{\theta - r}{\phi}} & r \le \theta \\ 1 - \frac{1}{2\phi} e^{-\frac{r-\theta}{\phi}} & r \ge \theta \end{cases}$$

The mean of the distribution is given by

$\mu = \theta$

and the standard deviation by

$\sigma = \phi\sqrt{2}$

For this probability distribution the coefficient of skewness is given by

$\gamma_3 = 0$

whereas the coefficient of excess kurtosis is given by

$\gamma_4 = 3$

Thus the double exponential is symmetric and has fatter tails than the normal distribution as it is indicated by the value of the kurtosis coefficient.

In order to compare the double exponential with the normal distribution we can express the density function in terms of the mean and the variance as follows:

$$p(r/\mu,\sigma) = \left(\sigma\sqrt{2}\right)^{-1} \exp\left[-\frac{|r-\mu|\sqrt{2}}{\sigma}\right]$$

From the distribution function we have

Lower quartile $\theta - \phi \log 2 = \theta - 0.693\phi$ or $\mu - \frac{\sigma}{\sqrt{2}} \log 2 = \mu - 0.490\sigma$

Upper Quartile $\theta + \phi \log 2 = \theta + 0.693\phi$ or $\mu + \frac{\sigma}{\sqrt{2}} \log 2 = \mu + 0.490\sigma$

The corresponding values for the normal distribution are

Lower quartile $\mu - 0.674\sigma$

Upper Quartile or $\mu + 0.674\sigma$

This difference reflects the sharp peak in the Laplace distribution. For quantiles however further in the tails the comparison is reversed because the Laplace density function decreases as

$$\exp\left[-\frac{|r-\mu|\sqrt{2}}{\sigma}\right]$$

whereas the normal density decreases as

$$\exp\left\{-\frac{1}{2}\left[\frac{r-\mu}{\sigma}\right]^2\right\}$$

For example the upper and lower 1% points of the Laplace distribution are $\mu \pm 2.722\sigma$ compared with $\mu \pm 2.326\sigma$ for the normal distribution.

The estimation of the parameters is straightforward by the method of maximum likelihood. Given observed values of n independent random variables $R_1,\ldots,R_n$ each with a Laplace density function, we can write the likelihood function as

$$L(r/\theta,\phi) = -n\ln(2\phi) - \frac{1}{\phi}\sum_{i=1}^{n}|R_i - \theta|$$

The maximum likelihood estimators of θ and ϕ are

$$\hat{\theta} = \tilde{R}$$

$$\hat{\phi} = \frac{1}{n}\sum_{i=1}^{n}|R_i - \theta|$$

where $\tilde{R}$ is the median value for R. The statistical significance of $\hat{\theta}$ and $\hat{\phi}$ can be tested by using the fact that $\hat{\theta}$ is (approximately) normally distributed with mean θ and variance

$$V(\hat{\theta}) = \frac{4\phi n!}{[(n-1)/2]!}\sum_{j=0}^{(n-1)/2}(-1)^j\left[j!\left(\frac{n-1}{2}-j\right)!2^{j+(n+1)/2}\left\{\frac{1}{2}(n+1)+j\right\}^3\right]^{-1}$$

The statistical significance of $\hat{\phi}$ can be tested by using the fact that $\hat{\phi}$ is distributed (when θ is known) as $(2n)^{-1}\phi\chi^2_{2n}$ where χ^2_{2n} is the chi-square distribution with $(2n)$ degrees of freedom. The limits of a $100(1-\alpha)\%$ confidence interval for ϕ are then

$$2\sum_{i=1}^{n}\frac{|R_i - \theta|}{\chi^2_{2n,1-\alpha/2}} \quad \text{and} \quad 2\sum_{i=1}^{n}\frac{|R_i - \theta|}{\chi^2_{2n,\alpha/2}}$$

The parameter estimates and the corresponding Value at Risk (quantiles) are given below in the Table 5.2.

Table 5.2: Parameter Estimates for the Laplace Distribution

	Estimates	Standard errors
$\hat{\theta}$	-0.00013	0.00008
$\hat{\phi}$	0.012712	0.0035

The second symmetric distribution we have used is the Student –t distribution with density function

$$\tau(r/\nu) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\nu}\Gamma(\nu/2)\Gamma(1/2)}\left[1+\frac{r^2}{\nu}\right]^{-\frac{1+\nu}{2}}$$

where ν is the degrees of freedom parameter. The parameter ν can be estimated by differentiating the log-likelihood function

$$\ln(L) = \sum_{t=1}^{T} \ln(\tau_t) = \sum_{t=1}^{T} \ln\left[\frac{\Gamma[(v+1)/2]}{\sqrt{v}\Gamma(v/2)\Gamma(1/2)} \left[1 + \frac{r_t^2}{v} \right]^{-\frac{1+v}{2}} \right]$$

with respect to the parameter of interest v and then using some numerical methods to solve the first order conditions.

Although the only parameter that needs to be estimated is the degrees of freedom parameter v, which can be achieved by setting the first derivative of the log-likelihood function equal to zero and solving the resulting nonlinear equation, it is nevertheless a difficult exercise. An alternative way to estimate the degrees of freedom parameter is using the methods of moments. The latter method is computationally much more efficient and given that we have to perform thousand of estimations, computationally efficiency is of paramount importance.

The odd moments of a t-distribution are not defined but the even moments are easily estimated. The moments of a t-distribution are given by

$$m_k(t_v) = v^{k/2} \frac{\Gamma[(v+1)/2]\Gamma[(v-k)/2]}{\Gamma[1/2]\Gamma[v/2]}$$

$$= v^{k/2} \frac{1 \cdot 3 \cdots (v-1)}{(v-k)(v-k+2)\cdots(v-2)}$$

The second moment (variance) is

$$m_2 = E(r^2) = \frac{v}{v-2}$$

whereas the fourth moment is given by

$$m_4 = \frac{3v^2}{(v-2)(v-4)}$$

The coefficient of excess kurtosis is given by

$$\gamma_4 = \frac{m_4}{(m_2)^2} - 3 = \frac{\dfrac{3v^2}{(v-2)(v-4)}}{\dfrac{v}{v-2}} - 3 = \frac{3v^2(v-2)}{(v-2)(v-4)v} - 3 = \frac{6}{(v-4)}$$

the parameter v can therefore be estimated in terms of the sample moments as

$$v = \frac{6}{\gamma_4} + 4$$

Using the estimate of the kurtosis from Table 5.1 the degrees of freedom parameter is estimated as

$$\hat{v} = \frac{6}{4.862} + 4 = 5.234$$

The tail of the student-t distribution can be calculated using

$$1 - F(r) \sim \frac{v^{(v-2)/2}}{B(1/2, v/2)} r^{-v}$$

where $B(a,b)$ is the Beta function with parameters a and b. The pth quantile can be calculated from the distribution function and is given by

$$r_p \sim \left[\frac{\Gamma(1/2)\Gamma(v/2)p}{\Gamma((v+1)/2)v^{(v-2)/2}} \right]^{-\frac{1}{v}}$$

Alternatively one can estimate the quantiles of the t-distribution using the EXCEL function TINV[1] Although the assumption of normally distributed returns is rejected by the data, we have nevertheless for comparison purposes estimated the quantiles under the assumption of normality using the formula for the pth quantile

$$r_{NP} = \hat{\mu} + \hat{\sigma}\Phi^{-1}(p)$$

where the parameters $\hat{\mu}$ and $\hat{\sigma}$ are the MLE estimators of the population parameters and take the values $\hat{\mu} = 0.0002$ and $\hat{\sigma} = 0.017$. The estimated quantiles are shown in the first column of Table 5.3 which brings together the estimates of the quantiles from the distributions we have used and a comparison is made with the empirical quantiles given in the fourth column of the table.

Table 5.3 shows that the normal distribution constantly underestimates the quantiles of the empirical distribution whereas the Laplace overestimates the quantiles. The Student –t is closer to the empirical quantiles and is therefore the better model of the three in terms of matching the tails of the empirical distribution.

Table 5.3 : Alternative quantile estimates predictions

Quantile	normal	Student	Laplace	Empirical Distribution
0.01	-0.037	-0.050	-0.060	-0.052
0.02	-0.032	-0.041	-0.049	-0.040
0.03	-0.028	-0.036	-0.044	-0.034
0.04	-0.025	-0.032	-0.039	-0.030
0.05	-0.023	-0.029	-0.036	-0.026
0.06	-0.021	-0.027	-0.034	-0.025
0.07	-0.019	-0.025	-0.031	-0.023
0.08	-0.018	-0.023	-0.030	-0.021
0.09	-0.016	-0.022	-0.028	-0.020
0.1	-0.015	-0.021	-0.026	-0.018

5.4 Modelling the Tails of Index Returns

An alternative to estimating the entire distribution as a method of obtaining VaR estimates we have estimated the lower tail of the distribution directly by assuming that the tail, that is, observations above a level u of the return distribution, is described by a Pareto distribution. The Pareto distribution specifies that the tails are slowly decaying under the so-called power law which is represented for a random variable X as

[1] The EXCEL function $r_p = TINV(p,v)$ calculates the pth quantile for a two tail distribution. To calculate the pth quantile for a one-tailed distribution we use the following formula $r_p = TINV(2p,v)$. The function also gives the probabilities for a standardized variable. We have therefore subtracted from each observation the mean and divided the difference by the standard deviation before we calculated the probability. The derived quantile was then multiplied by the standard deviation and added to the mean to produce the Value at Risk estimate.

$P(X > x) = cx^{-1/\xi}$ for $x > u$

In the estimation procedure we have used losses, that is negative returns defined as $X = -R$. There is no information lost in doing this since the highest loss is the lowest return. In the previous chapter we derived the Hill (1972) estimator and the estimate of the constant parameter c given by

$$\hat{\xi} = \frac{1}{k}\sum_{i=1}^{k} \ln(X_i / u) = \frac{1}{k}\sum_{i=1}^{k} \ln X_i - \ln u$$

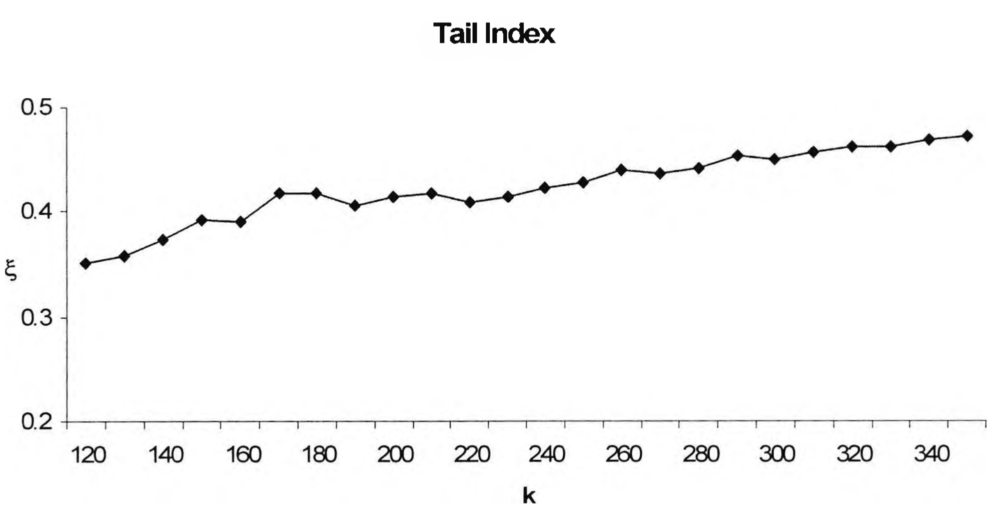
$$\hat{c} = \frac{k}{n} u^{1/\hat{\xi}}$$

The value x_{EP} for which $P(X > x_{EP}) = p$ is $x_{EP} = (p/\hat{c})^{-\xi} = u(pn/k)^{-\xi}$.

The main problem in estimating the above model as we have already discussed in chapter 4, is the determination of the threshold, or in other words the determination of the observations that can be described as extreme. The estimated shape parameter for different values of k is shown in Figure 5.4. In order to find the optimal value of k we have followed the methodology suggested by Huisman, Koedjik, Kool and Palm (2001). Their suggestion is to regress the different values of the estimated parameter on the variable representing the size of the extreme sample.

$$\hat{\xi}(k) = \beta_0 + \beta_1 k + \varepsilon(k) \qquad \text{for } k = 1,\ldots,k_0 \qquad (5.3.1)$$

Figure 5.4 Estimated ξ for different k



Once the above equation has been estimated the optimal $\hat{\xi}(k^*)$ is given by $\hat{\xi}(k^*) = \hat{\beta}_0$, that is the optimal $\hat{\xi}(k)$ is equal to the constant term of the equation.

There are some econometric considerations that need to be taken into account when the parameters of equation (5.3.1) are estimated. There are two main problems which make the use of Ordinary Least Squares an inappropriate method. First, the error term is heteroskedastic (see Huisman, Koedjik, Kool and Palm (2001)) Therefore a weighted least squares methodology was used to correct for this problem. To explain the methodology we write the equation in matrix form

$$\xi = \mathbf{Z\beta} + \varepsilon \tag{5.3.2}$$

where

$$\xi = \begin{bmatrix} \xi(1) \\ \vdots \\ \xi(k) \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} 1 \cdots 1 \\ \vdots \quad \vdots \\ 1 \cdots k \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \text{ and } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(k) \end{bmatrix}$$

The weighting matrix to be used in view of (5.3.2) is the following

$$\mathbf{W} = \begin{bmatrix} \sqrt{1} & 0 & \cdots & 0 \\ 0 & \sqrt{2} & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sqrt{k} \end{bmatrix}$$

Transformation of (5.3.2) through pre-multiplication with matrix $\mathbf{W}$ yields the following weighted least squares estimate for the vector of the regression coefficients

$$\hat{\boldsymbol{\beta}}_{WLS} = (\mathbf{Z'W'WZ})^{-1}\mathbf{Z'W'W}\xi$$

The estimated tail index is the first element of the vector $\hat{\boldsymbol{\beta}}_{WLS}$.

The second problem with the estimation of the regression equation (5.3.2) is that the estimates of $\xi(k)$ are auto-correlated for different *k* due to the use of common observations. The WLS method proposed above to deal with Heteroskedasticity does not take this problem into account and as a result the standard errors produced are not correct. To calculate the correct standard errors we have used the formulae derived by Huisman, Koedjik, Kool and Palm (2001).

The estimated weighted least squares regression equation is given below, with the correct t-statistics in brackets.

$$\xi(k) = \underset{(47.13)}{0.3996} + \underset{(17.20)}{0.0004\,4k}$$

Table 5.4 : Alternative VaR predictions

Quantile	normal	Student	Laplace	Extreme Value	Empirical Distribution
0.01	-0.037	-0.050	-0.060	-0.052	-0.052
0.02	-0.032	-0.041	-0.049	-0.039	-0.040
0.03	-0.028	-0.036	-0.044	-0.033	-0.034
0.04	-0.025	-0.032	-0.039	-0.029	-0.030
0.05	-0.023	-0.029	-0.036	-0.027	-0.026
0.06	-0.021	-0.027	-0.034	-0.025	-0.025
0.07	-0.019	-0.025	-0.031	-0.023	-0.023
0.08	-0.018	-0.023	-0.030	-0.022	-0.021
0.09	-0.016	-0.022	-0.028	-0.021	-0.020
0.1	-0.015	-0.021	-0.026	-0.020	-0.018

The optimal value of k is therefore the value that correspond to a value of $\xi = 0.3996$. The corresponding value of k is 150. Note that the optimal value of ξ is not very sensitive for values of k between 150 and around 240. The estimated quantiles based on the above value of ξ are given below in Table 5.4 where we have presented for comparison reasons the estimates of the other distributions. The performance of the extreme value model seems superior to all other models.

More specifically as it can be seen from the table the extreme value approach produces the closest fit to the empirical quantiles. Second best is the student distribution whereas the Laplace and Normal distributions always overestimate and underestimate the empirical tail distribution.

5.5 Heteroskedatisticity –Corrected Extreme Value VaR

The existence of serial correlation in both the returns and the squared returns is one of the most commonly observed empirical phenomena. The existence of serial correlation and volatility clustering, makes, at least in theory, the application of extreme value theory problematic as the asymptotic results that underpin this theory are based on the independence of returns. Indeed Resnick (1997) suggests two steps in fitting a heavy tailed model to a data set. In the first step the investigator decides if the data could possibly be explained by a heavy tailed model. The rejection of normality and the existence of kurtosis obviously support the view that a heavy tailed model is appropriate. The tests that we have employed to test the existence of heavy tails such as the presence of kurtosis and the Q-Q plots have established without doubt that the tails of the distribution of the stock market returns have thick tails. In the second stage the investigator should try to assess if there is dependency in the data. In this section we shall concentrate on tests of independence in the returns and how to deal with the possible problem of temporal dependence.

The motivation for this section is the investigation of the properties of the Hill estimator and the performance of alternative VaR measures when applied to non-independent data. Jansen and de Vries (1991) note that ξ is still a consistent estimator of ξ in the case of non independent variables as long as the dependence is not too strong. More recently, Resnick and Starica (1996) show the consistency of the Hill estimator when applied to data following an ARCH process. Also Pagan (1996) notes that large data sets are required to compute ξ accurately due to the departure of the data from the i.i.d. case.

In order to test the existence of serial correlation in the return series we have calculated the partial autocorrelation functions for up to 10 lags and tested for the statistical significance of the autocorrelation functions by sing Box-Lujng portmanteau statistic defined as

$$Q = T(T+2)\sum_{j=1}^{k}\frac{\rho_j^2}{T-j}$$

where T is the number of observations and ρ_j is the estimated autocorrelation of order j for r_t^2, defined as

$$\rho_j = \frac{\sum_{j=2}^{T}\varepsilon_t^2\varepsilon_{t-1}^2}{\sum_{j=2}^{T}\varepsilon_j^4}$$

Q is asymptotically distributed as χ_k^2. The results are given in Table 5.5, whereas a plot of the autocorrelation function is given in Figure 5.6.

Table 5.5: Autocorrelation Functions and Q statistic.

Lag	Autocorrelation coefficient
1	0.1667
2	-0.0088
3	-0.0092
4	-0.0004
5	-0.0116
6	0.0130
7	0.0162
8	0.0244
9	-0.0022
10	0.0043
Q(10)	81.77

From the value of the Q statistic we cannot reject the hypothesis that the returns are serially correlated. The critical value of the chi-square distribution with 10 degrees of freedom being 18.3.

Figure 5.5: Autocorrelation Functions of Returns

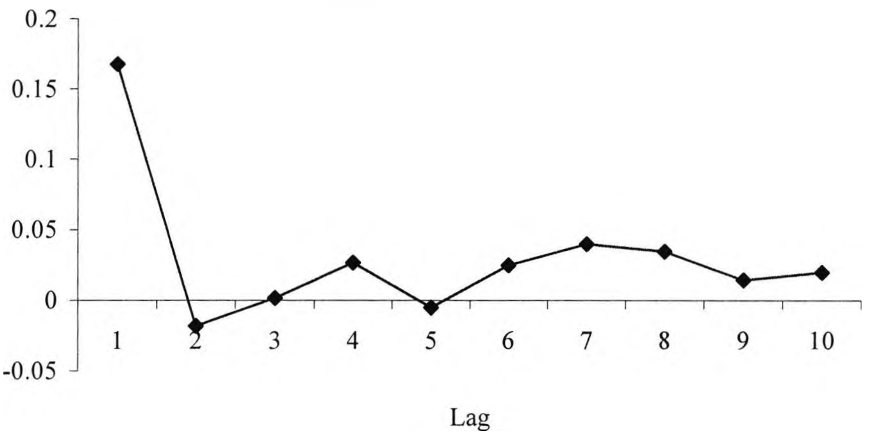
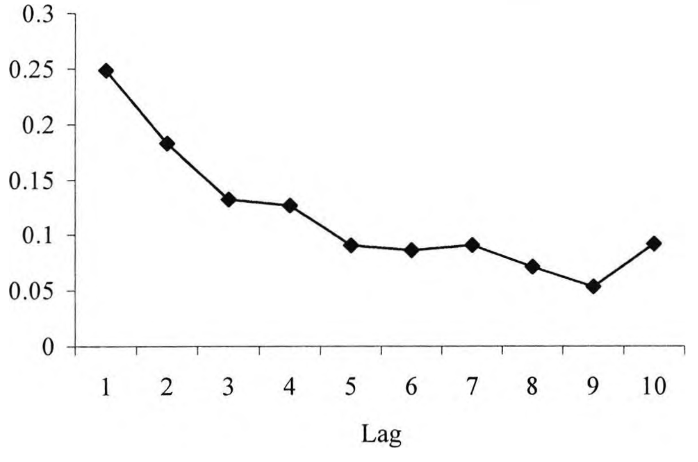


Figure 5.6: Autocorrelation Functions of Squared Returns



The plot of the autocorrelation functions in Figure 5.5 also indicates the existence of at least first order correlation given the size of the first order partial correlation coefficient. A formal modelling procedure will be undertaken to derive an appropriate process capturing the dynamics of asset returns.
The next test is to test whether the square of returns is a serially correlated process. The existence of volatility clustering is a universal empirical phenomenon in almost every stock market. Tests for autocorrelation in the squared returns can be performed using again the Ljung-Box Q-statistics.
The partial autocorrelation functions, as it can be seen from Figure 5.6 are quite high and a formal statistical test using the Q-statistic produced a value Q(1) = 11.7 and Q(10) = 47.3. So we can reject the hypothesis that the squared series are not correlated.
Having established, on the basis of the statistical sets that both the level and the square of returns exhibit serial correlation the next step is to identify the correct model to capture the dynamics of return and volatility. We postulate a model for the dynamics of the $\{R_t\}$ series in the form

$$R_t = \mu_t + \varepsilon_t \text{ with } \varepsilon_t = \sigma_t z_t$$

where the innovations z_t follow an iid strict white noise process with zero mean, unit variance and marginal distribution function $F_z(z)$. We assume that the conditional mean is given by a low order autoregressive model.

$$\mu_t = \phi_0 + \phi_1 R_{t-1} + \phi_2 R_{t-2} + \ldots + \phi_q R_{t-q}$$

Further we assume that the conditional variance of the mean-adjusted series $\varepsilon_t = R_t - \mu_t$ is given by a GARCH(1,1) model[2]

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad \text{where } \omega > 0, \quad \alpha > 0, \quad \beta > 0.$$

The mean-adjusted series ε_t is strictly stationary if $E[\log(\beta + \alpha z_{t-1}^2)] < 0$ which is equivalent to $\beta + \alpha < 1$. [McNeil and Frey (2000)]. This condition also ensures that the marginal distribution $F_R(r)$ or the random variable $\{R\}$ has a finite second moment.
To arrive at an appropriate model for the conditional mean we examined the correllogram of return series in order to see what is the appropriate lag length. We have found that a second order autoregressive model

$$R_t = \phi_0 + \phi_1 R_{t-1} + \phi_2 R_{t-2} + \varepsilon_t$$

captures adequately the dynamics of the return series. The estimates of the GARCH(1,1) conditional variance model was estimated under two assumption for the marginal distribution of $F_z(z)$. First we assume that z follows a normal distribution

[2] Although we have started with a GARCH(1,1) model this is not unduly restrictive. There is ample evidence that the model outperforms other volatility models. In the specific case of the Greek stock market the empirical evidence provided by Dinenis and Priniotakis ( ) supports the choice of a GARCH model.

$$l_t(z_t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(r_t - \mu_t)^2}{2\sigma_t^2} \right)$$

which with the volatility model

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

produces a joint likelihood or the entire sample given by

$$L = \prod_{t=1}^{T} l_t = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(r_t - \mu_t)^2}{2\sigma_t^2} \right)$$

whereas the log-likelihood function is given by

$$\ln(L) = \sum_{t=1}^{T} \ln(l_t) = \sum_{t=1}^{T} \ln\left[\frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(r_t - \mu_t)^2}{2\sigma_t^2} \right) \right]$$

$$= \sum_{t=1}^{T} \left(-\frac{1}{2} \right) \left[\ln(2\pi) + \ln\sigma_t^2 + \frac{(r_t - \mu_t)^2}{2\sigma_t^2} \right]$$

The Maximum likelihood estimator of the model parameters $(\omega, \alpha, \beta, \phi_0, \phi_1, \phi_2)$ can be obtained by taking the first derivative of the above likelihood function and setting them equal to zero[3].

In the second case we assume that the innovations z follow a student –t distribution with density function

$$\tau(r_t / \sigma_t, v) = \frac{\Gamma[(v+1)/2]}{\sigma_t\sqrt{(v-2)\pi}\Gamma(n/2)\Gamma(1/2)} \left[1 + \frac{(r_t - \mu_t)^2}{\sigma_t^2(v-2)} \right]^{-\frac{1+v}{2}}$$

where v is the degrees of freedom parameter. The parameters $(\omega, \alpha, \beta, \phi_0, \phi_1, \phi_2, v)$ can be estimated by differentiating the log-likelihood function of the entire sample, given below,

$$\ln(L) = \sum_{t=1}^{T} \ln(\tau_t) = \sum_{t=1}^{T} \ln\left[\frac{\Gamma[(v+1)/2]}{\sigma_t\sqrt{(v-2)\pi}\Gamma(n/2)\Gamma(1/2)} \left[1 + \frac{(r_t - \mu_t)^2}{\sigma_t^2(v-2)} \right]^{-\frac{1+v}{2}} \right]$$

with respect to the vector of parameters and then using some numerical methods to solve the first order conditions.

The Maximum likelihood estimates of the parameters of the two models are shown for each index return in Tables 5.6

Both models have statistically significant parameters and in order to select the correct model we have resorted to information criteria based on the value of the likelihood function. More specifically we have used both the Akaike information criterion as well as the Schwartz criterion. Both criteria favour the model with the Student –t marginal density function.

[3] The parameter estimates were estimated using the software package PC GIVE

Table 5.6 : Parameter Estimates of GARCH(1,1) models

	Marginal T		Marginal Normal	
	Parameter Estimates	T-Statistics	Parameter Estimates	T-Statistics
ϕ_1	0.197706	9.95	0.1959	8.79
ϕ_2	-0.07598	3.59	-0.05614	2.35
ϕ_0	-0.00028	1.31	-0.00013	0.563
ω	0.00001	3.73	0.00001	5.36
α	0.2459	6.96	0.19773	8.01
β	0.7373	20.1	0.7814	33.7
ν	5.8409	9.01		
$\alpha+\beta$	0.9832		0.97913	
LLF	7890.7		7804.3	

Once we have estimated the parameters $\phi_0, \phi_1, \phi_2, .., \phi_q, \omega, \alpha, \beta$ we have estimates of the conditional mean $\hat{\mu}_t$ and standard deviation $\hat{\sigma}_t$ and we calculate the standardised residuals

$$\hat{z}_t = \frac{R_t - \hat{\mu}_t}{\hat{\sigma}_t}$$

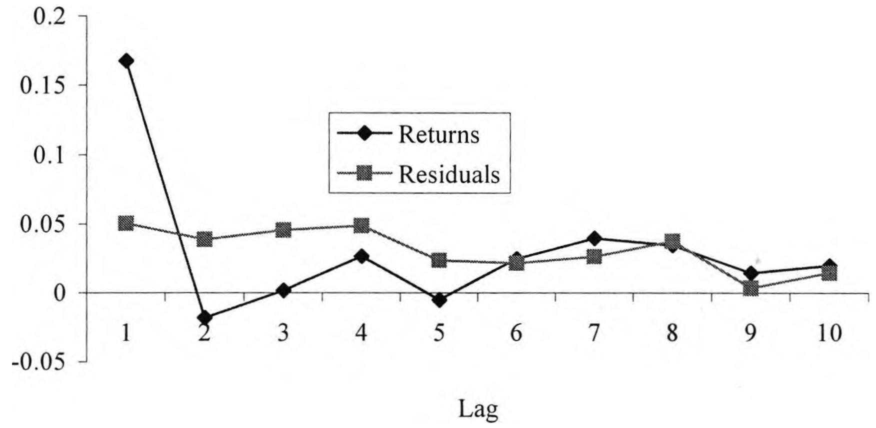
These residuals should be an iid white noise process, if the fitted model is correct. We can then estimate the tails of the innovations using extreme value theory that shall be applied on the residual series $\hat{z}_t$. So in order to see that we have produced a statistically adequate model for the conditional mean and variance, we now examine whether the standardized residuals based on the estimated model are independent and identically distributed processes. The statistical properties are shown in the Table 5.7. The standardised residuals follow a distribution with zero mean and standard deviation of 1. Moreover the Q statistic for 1 and 10 lags show that there is very little evidence of serial correlation left in the residuals and therefore they could be described as iid.

Table 5.7: Descriptive Statistics for standardised T- residuals

Average	0.01478
Volatility	1.0329
Skewness	0.22785
Kurtosis	2.9823*
Max Value	7.8279
Min Value	-4.9189
Q(1)	2.698
Q(10)	14.899

Figure 5.7 shows the autocorrelation functions of the standardised residuals which confirm the iid property. Similarly there is very little correlation left in the squared residuals as it can be seen from Figure 5.8.

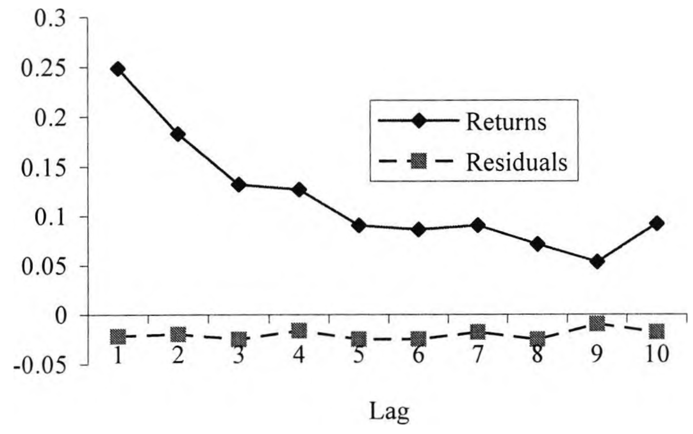
Figure 5.7: Autocorrelation Functions



However the excess kurtosis of the standardised residuals although lower than that of the unconditional returns is still significantly different from zero. The lowering of kurtosis through the standardisation process is a well known empirical fact as it is reported from instance in Anderson, Diebold and (1999). Lower kurtosis in the standardised returns is also evidence that the conditional model is not miss-specified.

The existence of kurtosis in the residuals means that fat-tailed distributions such as the student –t cannot capture the tail behaviour and justifies the use of extreme value theory for the estimation of tail quantiles. The results also show that the thickness of the tail is not due to volatility clustering but is an intrinsic property of the return series.

Figure 5.8 Autocorrelation Functions of Squared Series



A GARCH (1,1) model seems to capture the dynamics of the general index asset returns and it will be used in the next section in order to forecast VaR.

As we said at the beginning of the section one of the objectives was to see the impact of serial correlation on the performance of the tail index. The raw series whose tails we modelled earlier are clearly serially correlated. The standardised residuals we have produced are not. To assess the impact of serial correlation we have applied the extreme value theory to the estimation of the standardised residuals.

The results of our exercise are shown in Table 5.8 and Figure 5.9 below.

Again the extreme value quantiles are much closer to the quantiles of the empirical distribution. However there is no improvement in the fit compared to that of the unconditional distribution. . The percentage error in each case is given by

$$e_s = \frac{x_X^s - x_E^s}{x_E^s} \text{ and } e_D = \frac{x_X^D - x_E^D}{x_E^D}$$

where

x_E^s = empirical quantiles of return series

x_X^s = unconditional extreme value estimated quantiles of return series

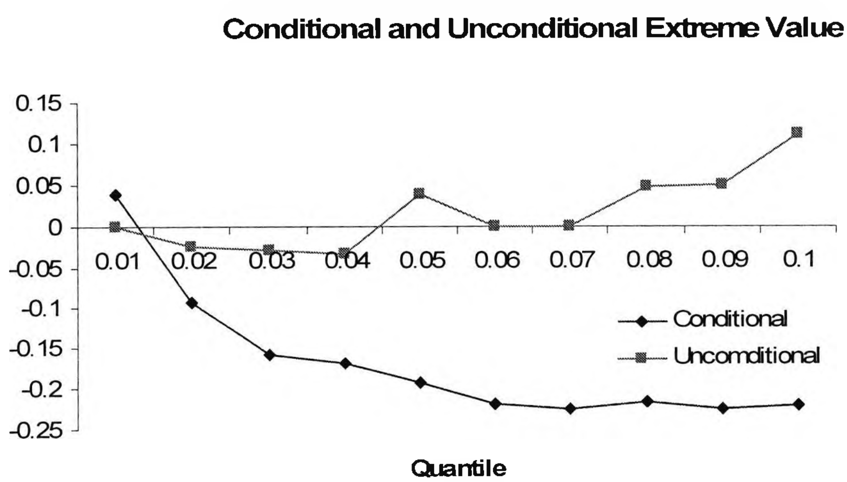
x_E^D = empirical quantiles of residual series

x_X^D = extreme value estimated quantiles of residuals

Table 5.8: Quantile Estimation Based on Standardised Residuals

	Normal	Student	Extreme	Empirical
0.01	-2.3902	-3.9510	-3.0281	-2.9186
0.02	-2.0356	-3.2390	-2.1876	-2.4113
0.03	-1.8063	-2.8471	-1.8086	-2.1509
0.04	-1.6311	-2.5772	-1.5803	-1.9029
0.05	-1.4866	-2.3713	-1.4233	-1.7630
0.06	-1.3618	-2.2047	-1.3066	-1.6725
0.07	-1.2511	-2.0645	-1.2154	-1.5697
0.08	-1.1506	-1.9433	-1.1416	-1.4592
0.09	-1.0580	-1.8363	-1.0803	-1.3939
0.1	-0.9717	-1.7404	-1.0282	-1.3191

Figure 5.9: Comparison of the Conditional and Unconditional Extreme Value Quantiles



The graph shows that the performance of both models deteriorates as they move towards higher percentiles but the deviation for the conditional model is much faster and underestimates the empirical distribution by up to 20 percent for the 0.07 percentile. This may be due to the fact that we have eliminated most of the kurtosis from the distribution of the residuals and thus the residuals have thinner tails the original raw series.

5.6 VaR Forecasting Models

In this section we use the metrics derived in the previous section to evaluate the predicted performance of the various forecasting models.

We have divided the models we use for evaluation into groups. In the first group we have models that ignore the heteroskedastic nature of the underlying series. We assume that the variance, or indeed the mean of the series, remains fixed within the sample that is used to estimate those parameters. This is tantamount to assuming that in calculating next day's volatility forecast, all the observations in the specific sample used have the same weight. In the second group of models we take into account the heteroskedastic nature of returns and the predicted volatility is still a function of return observations but this time their importance in forming a prediction is not uniform. We have used three ways of modelling volatility. First we have used the exponentially weighted moving average (EWMA) model used by RiskMetrics. The predicted volatilities are used with the assumption of normally distributed returns to produce estimates of VaR according to 5.12. The second model we have used is a GARCH(1,1) model with a normally distributed innovation term and the third model in this group is a GARCH(1,1) with an innovation term that follows the Student-t distribution. We have also used two models based on the extreme values of the distribution. The first model is the model applied to the return series which ignores the heteroskedastic nature of the series. In the second model we use the residuals form the Student-GARCH model and estimate the tails of the normalised residuals.

The methodology we have followed is to estimate each model over a certain period and then to produce a VaR forecast for the following day. In this way we have produced 2000 daily prediction of value-at-risk for each model. Thus we create a time series of predicted VaRs and on the basis of the time series properties of these forecasts we evaluate the performance of each model. For each model we have calculated VaR forecasts at both the 99 percent and 95 percent level.

Model 1 - The Naïve Normal Model

The first model to be evaluated is the normal model in which we assumed that the returns are iid and follow a normal distribution with mean μ and standard deviation σ. That is we assume

$R_t \sim N(\mu_t, \sigma_t)$. The predicted one period ahead Value –at-Risk is given by

$$VaR_{t+1}^{NORMAL}(p) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\Phi(p)^{-1}$$

where $\Phi^{-1}(p)$ is the pth quantile of the standard normal distribution. We will calculate the Value-at Risk for $p = 0.01$ and $p = 0.05$. For these two values the quantiles of the standard normal distribution are $\Phi^{-1}(0.01) = -2.33$ and $\Phi^{-1}(0.05) = -1.64$. The parameters $\hat{\mu}$ and $\hat{\sigma}$ were estimated from a sample of size k.

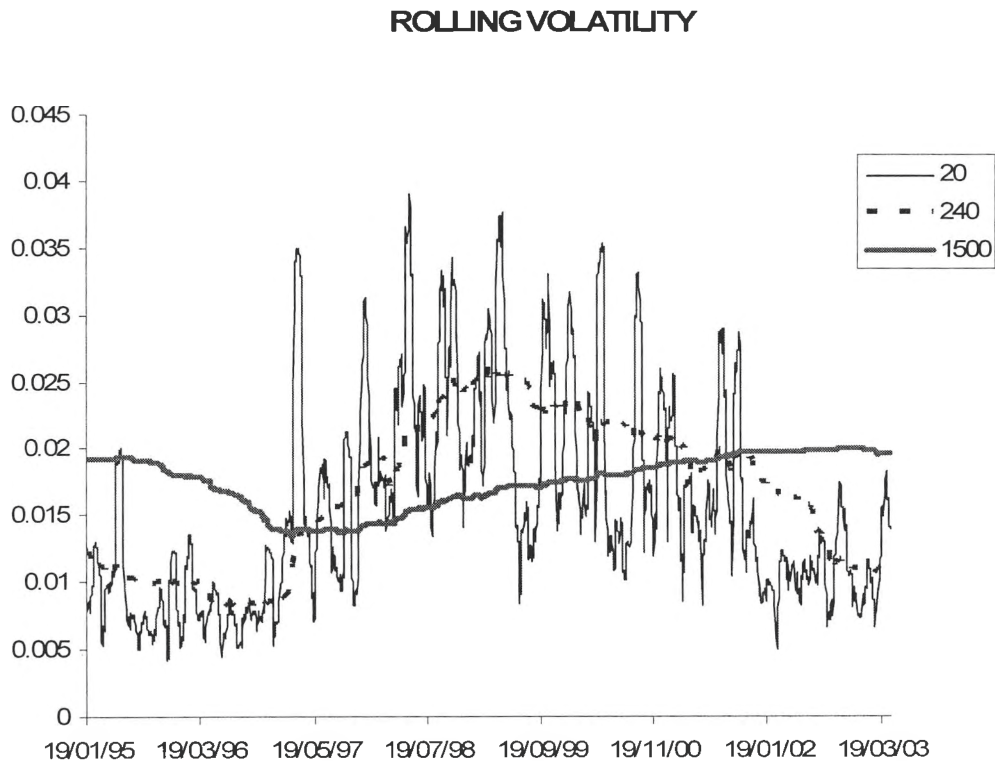
$$\hat{\mu}_{t+1/t} = \frac{1}{k}\sum_{j=0}^{k-1} r_{t-j} \quad \text{and} \quad \hat{\sigma}_{t+1/t} = \sqrt{\frac{1}{k}\sum_{j=0}^{k-1}[r_{t-j} - \hat{\mu}_{(k)}]^2}$$

The estimation of μ presents little difficulty as this quantity is close to zero irrespective of sample size. The estimation of σ on the other hand presents serious difficulties as its behaviour depends crucially on the size of the sample from which it is estimated. Figure 5.10 shows the estimates of σ for 2000 samples of 20, 240 and 1500 observations.

The 20-day standard deviation is too volatile and it will produce volatile VaR estimates. The 1500-day volatility estimate is too smooth and it may misses short-term changes in volatility. The 240-day volatility as it is expected shows more sensitivity to short –term fluctuations.

Given the dependence of the estimate of σ on sample size we have calculated the VaR for a series of sample sizes. More specifically we have used the following size samples k=20,60,120,240,600,1200, 1500 for the empirical studies.

Figure 5.10: Rolling Volatility estimates based on different sample sizes



Note that regardless of the window width, in all cases we have exactly the same number of "out of sample" VaR estimates.

Restricting the returns sample to a fixed window only makes sense if it is thought that the more recent data give a better estimate of the a moving parameter. However a fixed window imposes an artificial weighting scheme, with weights of 1 for the observations in the window and zero for the observations outside the window.

Model 2- The Student –T Distribution

Model 2 fits the Student – t distribution on a sample of size k to estimate the single parameter v and then calculates the Value- at- Risk. The parameter v can be estimated by MLE, using its density function

$$\tau(r_t / v_t) = \frac{\Gamma[(v_t + 1)/2]}{\sqrt{v_t}\Gamma(v_t/2)\Gamma(1/2)}\left[1 + \frac{r_t^2}{v_t}\right]^{-\frac{1+v_t}{2}}$$

as in the previous section but this time for each of the 2000 samples. Computationally this is quite time consuming and for this reason the degrees of freedom parameter was estimated from the sample moments.

The one –day ahead value of VaR is given by

$$VaR_{t+1}^{STUDENT}(p) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} F(p)^{-1}$$

where the pth quantile is calculated from

$$\tau(r_t / v_t) = \frac{\Gamma[(v_t + 1)/2]}{\sqrt{v_t}\,\Gamma(v_t/2)\Gamma(1/2)} \left[1 + \frac{r_t^2}{v_t}\right]^{-\frac{1+v_t}{2}}$$

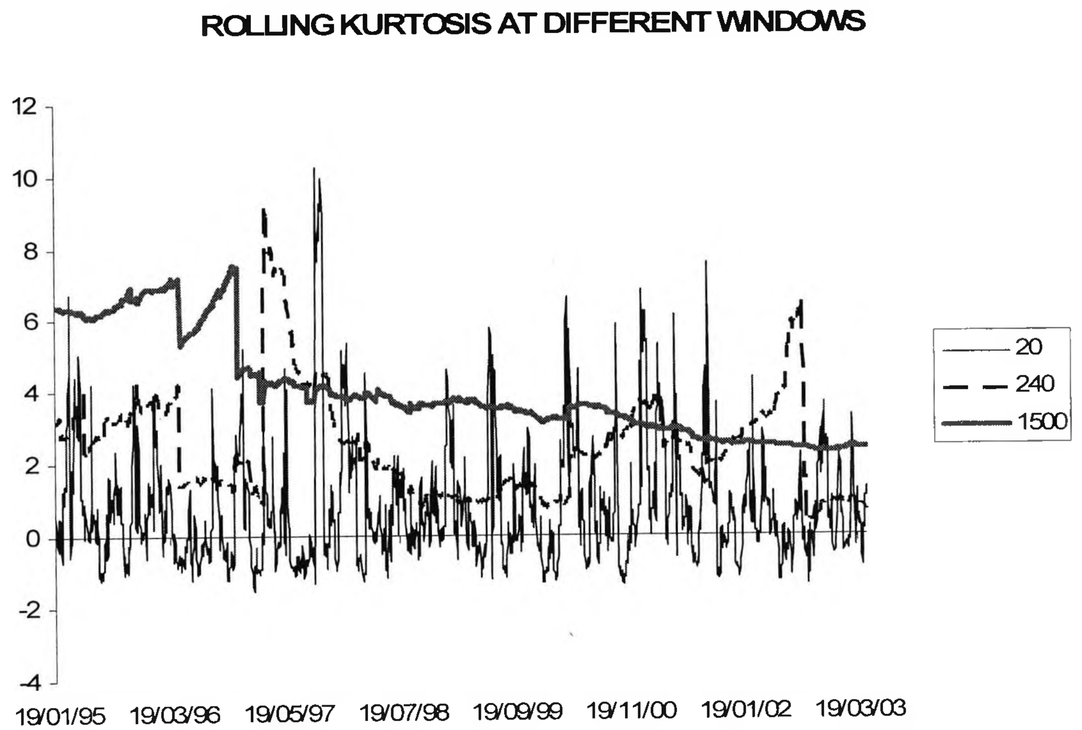
and the degrees of freedom parameter is estimated from

$$v_t = \frac{6}{\hat{\gamma}_{3,t}} + 4$$

The kurtosis parameter has been estimated using

$$\hat{\gamma}_{3,t+1/t} = \sqrt{\frac{\frac{1}{k}\sum_{j=0}^{k-1}[r_{t-j} - \hat{\mu}_{t-j}]^4}{\hat{\sigma}_{t+1/t}}}$$

As in the case of the normal model where we have forecasted volatility based on different sample sizes we have also experimented with different sample sizes for the estimates of skewness and kurtosis. Figure 5.11 Sample Size and Kurtosis Estimates



As it can be seen from Figure 5.11 kurtosis exhibit different behaviour for different sample sizes. As it was expected skewness and kurtosis is very volatile for short sample sizes and very smooth for larger sizes. The optimal size is an empirical question.

The kurtosis parameter and the one-day ahead forecasts are based on samples of different sizes. As in the case of the previous models we have calculated on-day VaR forecasts based on the following values of k.

k=20,60,120,240,600,1200,1500

Model 3 The Cornish Fisher Correction to Normal Quantile

In the modified normal model we use a Cornish-Fisher approximation to change the quantile of the normal distribution by taking into account the degree of skewness and kurtosis of the series.

$$VaR_{t+1}^{CF}(p) = VaR_{t+1}^{NORMAL}(p) + \frac{\hat{\gamma}_3}{6}\left(\left(VaR_{t+1}^{NORMAL}\right)^2 - 1\right) + \frac{\hat{\gamma}_4}{24}\left(\left(VaR_{t+1}^{NORMAL}\right)^3 - 3\left(VaR_{t+1}^{NORMAL}\right)\right)$$
$$-\frac{\hat{\gamma}_3^2}{36}\left(2\left(VaR_{t+1}^{NORMAL}\right)^3 - 5\left(VaR_{t+1}^{NORMAL}\right)\right)$$

The modified normal was employed to alternative models of fixed window volatility estimators described earlier.

Model 4 Static Extreme Value

In this model we have estimated the parameter ξ and the predicted Value-at-Risk based on a sample of m observations. We have kept the sample window fixed to 1990 observations. For the first window we estimated the shape parameter and the value at risk estimate. We added a new observation and we dropped the first one so we stayed with the same sample size. We estimated again the parameter and the VaR. In total we estimated 2360 extreme value parameters and an equal number of VaR forecasts. The VaR forecast was calculated using the formula

$$VaR_{t+1}^{EXTREME}(p) = u\left(p\frac{n}{k}\right)^{-\xi}$$

$$n = 1990$$

$$k = 150$$

$$p = 0.01, p = 0.05$$

A new threshold was calculated for each sample, based on the value of 151^{st} lowest return in the sample of 1500 observations.

Model 5 - Normal Model EWMA (RISKMETRICS)

In the models that follow we have allowed time varying second moment of the probability distribution of returns. The first model employed to obtain forecasts for σ_{t+1} is the exponentially weighted moving average or EWMA procedure used by J.P. Morgan in the RiskMetrics model. EWMA specifies the following period's variance to be a weighted average of the current variance and the current actual return

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)r_t^2$$

where λ is the decay factor. By recursive substitution, the conditional variance given by the EWMA model can be written as

$$\sigma_{t+1}^2 = (1-\lambda)r_t^2 + \lambda\sigma_t^2 = (1-\lambda)r_t^2 + \lambda[(1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2]$$
$$= (1-\lambda)r_t^2 + \lambda(1-\lambda)r_{t-1}^2 + \lambda^2\sigma_{t-1}^2 = (1-\lambda)r_t^2 + \lambda(1-\lambda)r_{t-1}^2 + \lambda^2(1-\lambda)r_{t-2}^2 + ...$$
$$= (1-\lambda)\sum_{i=0}^{\infty}\lambda^i r_{t-i}^2$$

Thus, the EWMA can be seen as an infinite weighted average of past squared returns, incorporating information from all the past shocks to the squared returns, but with exponentially declining weights. The EWMA, in contrast to the simple estimation of the standard deviation over a specific sample assumes that the importance of shocks declines rapidly. In practice we need to use a finite number of terms for the calculation of the expected variance and we have therefore the following version

$$\hat{\sigma}_{t+1} = \sqrt{(1-\lambda)\sum_{j=t-k}^{t-1}\lambda^{t-j-1}[r_j - \bar{r}]^2}$$

For the model to become operational we need estimates of the decay parameter λ, which is unknown. There are three solutions to the problem of finding values for this parameter. The first one is to rely on the RiskMetrics recommendation and use the value λ. RiskMetrics uses the value $\lambda = 0.94$. The second is to use a range of values for λ and select the one that produces the best forecasts. Ranges that have been suggested in the literature are in the range of $\lambda \in [0.85 - 0.99]$. The third way is to use Maximum Likelihood Estimation methods to obtain the value of λ. In this approach the decay coefficient λ can be estimated using the maximum likelihood approach where the sample observations are weighted. The Log-Likelihood Function when the returns are normally distributed is given by

$$\ln(L) = \sum_{t=1}^{T}(1-\lambda)\lambda^{t-1}\ln(l_t) = \sum_{t=1}^{T}(1-\lambda)\lambda^{t-1}\ln\left[\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp\left(-\frac{r_t^2}{2\sigma_t^2}\right)\right]$$
$$= \sum_{t=1}^{T}(1-\lambda)\lambda^{t-1}\left(-\frac{1}{2}\right)\left[\ln(2\pi) + \ln\sigma_t^2 + \frac{r_t^2}{2\sigma_t^2}\right]$$

The factor $(1-\lambda)$ ensures that the weights of the log-likelihood function sum to unity. Maximization of the above function, leads to MLE of the three parameters.

Model 6 - The Normal GARCH(1,1) model

The conditional forecast model is based on the Value at Risk formulation which makes explicit the separation between the distributional assumptions and the information set used.

$$VaR_p = \mu_{t+1/t} + \sigma_{t+1/t}F_Z^{-1}(p)$$

In this model we have used the basic framework we have used so far, that is,

First, we postulate a model for the dynamics of the $\{r_t\}$ series in the form

$$r_t = \mu_t + \sigma_t z_t$$

where the innovations z_t follow an iid strict white noise process with zero mean, unit variance and marginal distribution function $F_z(z)$. We assume that the conditional mean is given by

$$\mu_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2}$$

$$\hat{\mu}_{t+1/t} = \hat{\phi}_0 + \hat{\phi}_1 r_t + \hat{\phi}_2 r_{t-1}$$

In this model we have assumed that the conditional variance of the mean-adjusted series $\varepsilon_t = r_t - \mu_t$ is the normal GARCH(1,1) model which has been estimated in the earlier sections of this chapter. The model consists of

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \varepsilon_t$$

$$\varepsilon_t = \sigma_t z_t \qquad z_t \sim N(0,1)$$

$$\sigma_{t+1}^2 = \omega + \alpha r_t^2 + \beta \sigma_t^2 \text{ with } \alpha + \beta < 1$$

Notice that the RiskMetrics model can be viewed as a special case of the GARCH(1,1) model if we assume that $\alpha + \beta = 1$ and further $\omega = 0$. However, although the two models appear to be similar, there is an important difference which comes out from the definition of the unconditional or long-run average variance which is defined as

$$\sigma^2 = E(\sigma_{t+1}^2) = \omega + \alpha E(r_t^2) + \beta E(\sigma_t^2) = \omega + \alpha\sigma^2 + \beta\sigma^2$$

so that

$$\sigma^2 = \frac{\omega}{1 - \alpha - \beta}$$

The special case of GARCH(1,1) which results into the RiskMetrics model, i.e. $\alpha + \beta = 1$, implies that the long-run variance is infinite. The RiskMetrics model therefore ignores the empirical fact that the long-run variance tends to be relatively stable over time. The GARCH (1,1) model on the other hand implicitly relies on σ^2. This can be seen by using $\sigma^2 = \frac{\omega}{1-\alpha-\beta}$ to replace ω in the GARCH equation producing the alternative formulation

$$\sigma_{t+1}^2 = (1 - \alpha - \beta)\sigma^2 + \alpha r_t^2 + \beta\sigma_t^2 = \sigma^2 + \alpha(r_t^2 - \sigma^2) + \beta(\sigma_t^2 - \sigma^2)$$

Under this reformulation, tomorrow's variance is a weighted average of the long-run variance, today's squared return and today's variance. Or alternatively, tomorrow's variance is the long-run average variance with something added (subtracted) if today's squared return is above (below) its long-run average, and something added (subtracted) if today's variance is above (below) its long-run average.
A key advantage of GARCH(1,1) model for risk management is that the one-day forecast of volatility, $\sigma_{t+1/t}^2$ is given directly by the model as $\sigma_{t+1/t}^2$.

The sum $(\alpha + \beta)$ is usually called the persistence of the model. A high persistence, that is a value of $(\alpha + \beta)$ close to 1, implies that shocks which push volatility away from its long-run average will persist for a long time, but eventually the long horizon forecast will be the long-run average variance σ^2.

The model was implemented as follows. The maximum likelihood function

$$\ln(L) = \sum_{t=1}^{T} \ln(l_t) = \sum_{t=1}^{T} \ln\left[\frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{r_t^2}{2\sigma_t^2}\right)\right]$$

$$= \sum_{t=1}^{T}\left(-\frac{1}{2}\right)\left[\ln(2\pi) + \ln\sigma_t^2 + \frac{r_t^2}{2\sigma_t^2}\right]$$

was maximised based on a fixed data window of 1500 days. The Maximum likelihood estimator of the model parameters were obtained by taking the first derivative of the above likelihood function and setting them equal to zero.

$$\frac{\partial \ln(l_t)}{\partial\gamma} = \left(-\frac{1}{2}\right)\left[\frac{1}{\sigma_t^2} - \frac{r_t^2}{(\sigma_t^2)^2}\right]\frac{\partial\sigma_t^2}{\partial\gamma} = \frac{1}{2}\left(\frac{1}{\sigma_t^2}\right)\frac{\partial\sigma_t^2}{\partial\gamma}\left[\frac{r_t^2}{\sigma_t^2} - 1\right]$$

where $\gamma = (\phi_0, \phi_1, \phi_2, \omega, \alpha, \beta)$. Based on the parameter estimates we forecast the volatility for the following day and then derive the VaR forecast. This is repeated for a total of 1500 new observations.

Model 7 - The T- GARCH(1,1) model

The third conditional forecasting model is the T- GARCH(1,1) model which has been estimated in the earlier sections of this chapter. The model consists of

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \varepsilon_t$$

$$\varepsilon_t = \sigma_t z_t \qquad z_t \sim t(0,1,v)$$

$$\sigma_{t+1}^2 = \omega + \alpha r_t^2 + \beta\sigma_t^2 \text{ with } \alpha + \beta < 1$$

This time however we assume that the innovation variable has a t distribution with a unit standard deviation given by $v/(v-2)$.

The second assumption we make about the distribution of returns is that they follow the student –t distribution with degrees of freedom v. We simply replace the normal distribution by the T-distribution.

$$t(r/\mu,\sigma,\nu) = \frac{\Gamma[(v+1)/2]}{\sigma_t\sqrt{(\nu-2)\pi}\Gamma(n/2)\Gamma(1/2)}\left[1 + \frac{(r-\mu)^2}{\sigma_t^2(\nu-2)}\right]^{-\frac{1+\nu}{2}}$$

The above model, allows the variance to depend on time, but it makes the assumption that the degrees of freedom parameter and therefore the kurtosis parameter are fixed over time. We have not pursued the issue of time varying kurtosis in an explicit manner, although the degrees of freedom parameter is estimated every time for forecasting purposes. Thus, in terms of the kurtosis parameter the model is an unconditional one. There is very little empirical evidence to guide us in terms of the dynamics of kurtosis, i.e. whether a constant weight model is better in terms of forecasting rather than a model with a exponentially weighted observations or even a GARCH type model.

Model 8 - The GARCH Extreme Value Model

In the extreme value model we use the standardised residuals form the T-GARCH model and derive the Hill estimator of the tail index for a given sample. Once the tail estimator has been estimated, then we can calculate the Value-at Risk from the formula we derived earlier

$$VaR_{t+1} = \mu_{t+1} + \sigma_{t+1} Z_X(p)$$

The procedure was followed for the same number of periods as in the other cases. That is the tail index estimator, the conditional mean and the conditional volatility model were estimated over a sample of 1500 observations for a total of 1200 days. A VaR forecast was derived in this way for every single day.

5.7 Empirical Results

Having presented the forecasting models we report now the performance of these various models. The detailed results are shown in the Appendix to this chapter. Here we shall provide a summary of the findings. We start with the unconditional forecast test first. This test show how close the forecasts have been to the empirical quantile. Acceptance of the model is indicated with the letter Y whereas rejection of the model is indicated with the letter X. Tables 5.9 and 5.10 show the results for the two confidence levels.

Table 5.9– Test of unconditional performance 1%

	k							
	20	60	120	240	600	1200	1500	
Model 1	X	X	X	X	Y	Y	X	
Model 2	X	Y	Y	Y	Y	Y	Y	
Model 3	X	X	X	X	Y	Y	Y	
Model 4								Y
Model 5								X
Model 6								X
Model 7								Y
Model 8								Y

Model 1 is accepted at 1% percent level for sample periods of 600 and 1200 days only and is rejected for any other sample period on which the estimates are based. Model 2 on the other hand is accepted for all samples apart for the 20-day one. Model 3 is accepted only when the estimates are based on long samples. The unconditional extreme value model is also accepted. Of the dynamic models the RiskMetrics model (model 5) and the normal GARCH are not accepted but the t-GARCH and the conditional extreme value model are accepted.

The results for 5% VaR tests of unconditional coverage are given in Table 5.10 Models 1 and 3 are accepted for large samples. Of the conditional models though only the two GARCH models produce acceptable results.

Although the normal model was expected to perform poorly, Cornish –Fisher adjustment which takes into account the skewness and the kurtosis of the distribution also is disappointing. Theoretically such an adjustment should result in a more successful model. However the improvement depends very much on the quality of the sample estimates of the kurtosis and the skewness parameters.

The results for independence of VaR violations are though disappointing. This is true for models 1,2 and 3 at both 1% and 5% probability level. The unconditional extreme value is accepted at 1% but not at 5%.

Table 5.10 – Test of unconditional performance 5%

	k							
	20	60	120	240	600	1200	1500	
Model 1	X	X	X	Y	Y	Y	X	
Model 2	Y	Y	X	X	X	X	X	
Model 3	X	X	Y	Y	Y	Y	Y	
Model 4								X
Model 5								X
Model 6								Y
Model 7								Y
Model 8								X

Table 5.11 – Test of independence 1%

	K							
	20	60	120	240	600	1200	1500	
Model 1	X	X	X	X	X	X	X	
Model 2	X	Y	X	X	X	X	X	
Model 3	X	X	X	X	X	X	X	
Model 4								Y
Model 5								Y
Model 6								X
Model 7								Y
Model 8								Y

Table 5.12 – Test of independence 5%

	k							
	20	60	120	240	600	1200	1500	
Model 1	X	X	X	X	X	X	X	
Model 2	X	X	X	X	X	X	X	
Model 3	X	X	X	X	X	X	X	
Model 4								X
Model 5								X
Model 6								X
Model 7								X
Model 8								X

As a result of the failure of the independence test, the test of conditional coverage reject all the models except the extreme value model and the t-GARCH.

Table 5.13– Test of conditional coverage 1%

	K							
	20	60	120	240	600	1200	1500	
Model 1	X	X	X	X	X	X	X	
Model 2	X	Y	X	X	X	X	X	
Model 3	X	X	X	X	X	X	X	
Model 4								Y
Model 5								X
Model 6								X
Model 7								Y
Model 8								Y

Rejection of the models other than those that allow for fat tails or semi-fat tails and volatility clustering, means that risk measurement models that are based on the assumption of normality fail to capture the characteristics of the equity returns and result in poor assessments of the risks to which an equity position is exposed.

Table 5.14– Test of conditional coverage 5%

	K							
	20	60	120	240	600	1200	1500	
Model 1	X	X	X	X	X	X	X	
Model 2	X	X	X	X	X	X	X	
Model 3	X	X	X	X	X	X	X	
Model 4								X
Model 5								X
Model 6								X
Model 7								X
Model 8								X

5.8 Conclusions

In this chapter we have dealt with an important issue in the measurement of market risk for equities, that is the evaluation of the performance of alternative VaR models.

First we have explored the statistical properties of the Athens General Price Index. We have estimated alternative densities and modelled the dynamics of the returns via a GARCH model. Then we estimated the tail characteristics and finally we used the series to assess the performance of various models.

We have tested a number of models in terms of three criteria. First we have tested their unconditional coverage, that this, their ability to predict correctly the number of VaR violations. We have found that models based on the normal distribution do not perform well, even when we allow for volatility clusters like the RiskMetrics model or the normal GARCH. In general the T-distribution and the extreme value approach perform better. Secondly we have tested them in terms of the serial correlation of forecasting errors. A good VaR model should not exhibit serially correlated forecasting errors, as the existence of a cluster of VaR violations may lead to the bankruptcy of a financial institution. We have found that extreme value theory and conditional models that take into account the existence of volatility clusters perform much better than unconditional models which assume that the distribution parameters remain fixed.

Table A1-Model 1: Normal Distribution for daily returns

	20 Day		60 DAY		120	
	5%	1%	5%	1%	5%	1%
T_0	1904	1968	1917	2010	831	82
T_1	157	93	144	51	1230	1979
T_{00}	1771	1889	1800	1966	819	49
T_{01}	133	79	117	44	12	33
T_{10}	133	79	117	44	12	33
T_{11}	24	14	27	7	1218	1946
π	0.0761	0.0451	0.0698	0.0247	0.596798	0.9602
π_{01}	0.069	0.0401	0.0610	0.0218	0.01444	0.4024
π_{11}	0.152	0.1505	0.1875	0.1372	0.990244	0.983
LR_{uc}	25.80	138.088	15.327	32.091	4675.333	17539.4
LR_{ind}	11.49	16.427	24.066	13.903	2518.888	243.292
LR_{cc}	37.29	154.515	39.393	45.995	7194.22	17782.7
Significance						
LR_{uc}	Reject	Reject	Reject	Reject	Reject	Reject
LR_{ind}	Reject	Reject	Reject	Reject	Reject	Reject
LR_{cc}	Reject	Reject	Reject	Reject	Reject	Reject

Table A1-Model 1: Normal Distribution for daily returns (continued)

	240		600		1200		1500	
	5%	1%	5%	1%	5%	1%	5%	1%
T_0	1949	2029	1960	2042	1966	2043	978	978
T_1	112	32	101	19	95	18	1083	1083
T_{00}	1856	2002	1879	2026	1895	2028	561	561
T_{01}	93	27	81	16	71	15	417	417
T_{10}	93	27	81	16	71	15	417	417
T_{11}	19	5	20	3	24	3	666	666
π	0.0543	0.0155	0.049	0.00921	0.04609	0.0087	0.525	0.525
π_{01}	0.0477	0.0133	0.041	0.00783	0.03611	0.0073	0.426	0.426
π_{11}	0.1696	0.1562	0.198	0.15789	0.25263	0.1666	0.614	0.614
LR_{uc}	0.7967	5.4411	0.043	0.13044	0.67894	0.3487	7142.6	7142.6
LR_{ind}	20.807	15.446	30.862	12.2920	51.8086	12.969	73.69	73.69
LR_{cc}	21.603	20.887	30.905	12.4224	52.4875	13.3182	7216.3	7216.3
Significance								
LR_{uc}	Don't Reject	Reject	Don't Reject	Don't Reject	Don't Reject	Don't Reject	Reject	Reject
LR_{ind}	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
LR_{cc}	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

Table A2-Model 2: t-distribution for daily returns

	20-Day		60-DAY		120-DAY	
	5%	1%	5%	1%	5%	1%
T_0	1942	2028	1969	2035	1981	2043
T_1	119	33	92	26	80	18
T_{00}	1841	2000	1889	2010	1912	2029
T_{01}	101	28	80	25	69	14
T_{10}	101	28	80	25	69	14
T_{11}	18	5	12	1	11	4
π	0.058	0.016	0.045	0.013	0.039	0.009
π_{01}	0.052	0.014	0.041	0.012	0.035	0.007
π_{11}	0.151	0.152	0.130	0.038	0.138	0.222
LR_{uc}	2.481	6.364	1.292	1.315	5.860	0.349
LR_{ind}	14.766	14.833	11.443	0.921	13.748	20.002
LR_{cc}	17.247	21.196	12.735	2.236	19.608	20.351
Significance						
LR_{uc}	Don't Reject	Reject	Don't Reject	Don't Reject	Reject	Don't Reject
LR_{ind}	Reject	Reject	Reject	Don't Reject	Reject	Reject
LR_{cc}	Reject	Reject	Reject	Don't Reject	Reject	Reject

Table A2-Model 2: t-distribution for daily returns (continued)

	240-DAY		600-DAY		1200-DAY		1500-DAY	
	5%	1%	5%	1%	5%	1%	5%	1%
T_0	1999	2048	2001	2045	1994	2042	1992	2043
T_1	62	13	60	16	67	19	69	18
T_{00}	1944	2037	1951	2032	1940	2026	1937	2028
T_{01}	55	11	50	13	54	16	55	15
T_{10}	55	11	50	13	54	16	55	15
T_{11}	7	2	10	3	13	3	14	3
π	0.030	0.006	0.029	0.008	0.033	0.009	0.033	0.009
π_{01}	0.028	0.005	0.025	0.006	0.027	0.008	0.028	0.007
π_{11}	0.113	0.154	0.167	0.188	0.194	0.158	0.203	0.167
LR_{uc}	19.953	3.267	22.135	1.128	15.070	0.130	13.336	0.349
LR_{ind}	9.163	9.543	20.883	14.470	28.699	12.292	31.502	12.969
LR_{cc}	29.116	12.809	43.018	15.598	43.769	12.422	44.838	13.318
Significance								
LR_{uc}	Reject	Don't Reject	Reject	Don't Reject	Reject	Don't Reject	Reject	Don't Reject
LR_{ind}	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
LR_{cc}	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

Table A3-Model 3: Cornish-Fischer approximation

	20		60		120	
	5%	1%	5%	1%	5%	1%
T_0	1904	1968	1917	2010	1940	2024
T_1	157	93	144	51	121	37
T_{00}	1771	1889	1800	1966	1840	1990
T_{01}	133	79	117	44	100	34
T_{10}	133	79	117	44	100	34
T_{11}	24	14	27	7	21	3
π	0.076	0.045	0.070	0.025	0.059	0.018
π_{01}	0.069	0.040	0.061	0.022	0.052	0.017
π_{11}	0.152866	0.151	0.188	0.137	0.174	0.081
LR_{uc}	25.804	138.08	15.327	32.092	3.125	10.652
LR_{ind}	11.490	16.427	24.067	13.904	21.374	4.685
LR_{cc}	37.295	154.51	39.394	45.996	24.499	15.337
Significance						
LR_{uc}	Reject	Reject	Reject	Reject	Don't Reject	Reject
LR_{ind}	Reject	Reject	Reject	Reject	Reject	Reject
LR_{cc}	Reject	Reject	Reject	Reject	Reject	Reject

Table A3-Model 3: Cornish-Fischer approximation (continued)

	240		600		1200		1500	
	5%	1%	5%	1%	5%	1%	5%	1%
T_0	1949	2029	1960	2042	1966	2043	1962	2041
T_1	112	32	101	19	95	18	99	20
T_{00}	1856	2002	1879	2026	1895	2028	1887	2024
T_{01}	93	27	81	16	71	15	75	17
T_{10}	93	27	81	16	71	15	75	17
T_{11}	19	5	20	3	24	3	24	3
π	0.054	0.016	0.049	0.009	0.046	0.009	0.048	0.010
π_{01}	0.048	0.013	0.041	0.008	0.036	0.007	0.038	0.008
π_{11}	0.170	0.156	0.198	0.158	0.253	0.167	0.242	0.150
LR_{uc}	0.797	5.441	0.043	0.130	0.679	0.349	0.170	0.018
LR_{ind}	20.807	15.446	30.862	12.292	51.809	12.969	47.866	11.656
LR_{cc}	21.604	20.888	30.905	12.422	52.488	13.318	48.035	11.675
Significance								
LR_{uc}	Don't Reject	Reject	Don't Reject	Don't Reject	Don't Reject	Don't Reject	Don't Reject	Don't Reject
LR_{ind}	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
LR_{cc}	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject

Table A4-Model 4 – Extreme value applied to return series

	5%	1%
T_0	1451	1556
T_1	124	19
T_{00}	1353	1538
T_{01}	98	18
T_{10}	98	18
T_{11}	26	1
π	0.0787	0.0121
π_{01}	0.0675	0.0115
π_{11}	0.2096	0.0526
LR_{uc}	23.475	0.6355
LR_{ind}	23.5094	1.4692
LR_{cc}	46.9847	2.1047
Significance		
LR_{uc}	Reject	Don't Reject
LR_{ind}	Reject	Don't Reject
LR_{cc}	Reject	Don't Reject

Table A5-Models 5-6

	Model 6		Model 5	
	Normal GARCH		EWMA	
	5%	1%	5%	1%
T_0	1964	2031	1931	2022
T_1	97	30	130	39
T_{00}	1877	2003	1820	1985
T_{01}	87	28	111	37
T_{10}	87	28	111	37
T_{11}	10	2	19	2
π	0.0470	0.0145	0.0631	0.0189
π_{01}	0.0442	0.0137	0.0575	0.0183
π_{11}	0.1030	0.0666	0.1462	0.0513
LR_{uc}	0.3810	3.7885	6.8759	13.1335
LR_{ind}	5.4804	3.1297	12.3707	1.5488
LR_{cc}	5.8614	6.9182	19.2467	14.6823
Significance				
LR_{uc}	Don't Reject	Reject	Reject	Reject
LR_{ind}	Reject	Reject	Reject	Don't Reject
LR_{cc}	Reject	Reject	Reject	Reject

Table A6-Models 7-8

	Model 7		Model 8	
	T-GARCH		T-GARCH & Extreme Value	
	5%	1%	5%	1%
T_0	1961	2043	1451	1550
T_1	100	18	120	20
T_{00}	1872	2026	1353	1538
T_{01}	89	17	60	15
T_{10}	89	17	20	13
T_{11}	11	1	30	15
π	0.0485	0.0087	0.076	0.012
π_{01}	0.0453	0.0083	0.041	0.009
π_{11}	0.11	0.0555	0.25	0.75
LR_{uc}	0.0959	0.3487	19.96	1.094
LR_{ind}	6.5434	2.0960	256.62	0.542
LR_{cc}	6.6394	2.4448	276.58	1.63
Significance				
LR_{uc}	Don't Reject	Don't Reject	Reject	Don't Reject
LR_{ind}	Reject	Don't Reject	Reject	Don't Reject
LR_{cc}	Reject	Don't Reject	Reject	Don't Reject

References and bibliography

ALTMAN, E. (1968): "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". *Journal of Finance*, September: pp. 589-609.

ALTMAN, E. (1984): "The success of Business Failure Prediction Models: An International Survey". *Journal of Banking and Finance* 8 pp 171-198

ALTMAN, E. (2000): "Forecasting Financial Distress of Companies: Revisiting the Z-Score and the Zeta Models" Working Paper, Stern School, July

ALTMAN, E., R.HALDEMAN AND P.NARAYANAN (1977): "Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations," *Journal of Banking and Finance*, 1

ANDERSON, T.W (1984), *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

ANDREOU, E. PITTIS, N. and A. SPANOS (2001) "On Modelling Speculative Prices: The Empirical Literature" Journal *of Economic Surveys,* Volume 15, Issue 2, Page 187-220.

APARICIO, F. and J. ESTRADA (2001) "Empirical Distributions of Stock Returns: European Securities Markets, 1990-95", *European Journal of Finance*, Vol. 7 pp 1-21.

ARTZNER, P., DELBAEN, F., EBER, J. & HEATH, D. (1997). "Thinking Coherently" *Risk*, 10 (November).

ARTZNER, P., DELBAEN, F., EBER, J. & HEATH, D. (1999). "Coherent Measures of Risk" *Mathematical Finance*, 9 (November), 203-228..

BARONE-ADESI, G., GIANNOPOULOS, K. & VOSPER, L. (1999). "VaR without Correlations for Portfolios of Derivative Securities", *Journal of Futures Markets*, 19, 583-602.

BARVIN, R., AGARWAL, A., and LEACH, R. (2002), "Predicting Bankruptcy Resolution" *Journal of Business Finance and Accounting* 29(3)&4, April/May, pp. 497-520

BEAVER, W. (1966): "Financial Ratios as Predictors of Failure", *Journal of Accounting Research*, pp 77-111.

BERA, A. K., and HIGGINS, M. L. (1993), "ARCH Models: Properties, Estimation and Testing," *Journal of Economic Surveys*, Vol. 7, No. 4, 307-366

BERKOWITZ, JEREMY AND JAMES O'BRIEN, (2001) "How Accurate are Value-at-Risk Models at Commercial Banks?", *Journal of Finance*, Vol 57, No 3, pp. 1093–1112.

BILDERBEEK, J. (1977), Financiele ratio analyse (Stenfert-Kroese, Leiden)

BIS (2003) "The New Capital Accord", Basel Committee on Banking Supervision, Bank for International Settlements, April.

BLACK, F. and M. SCHOLES (1973) "The Pricing of Options and Corporate Liabilities" *Journal of Political Economy*, Vol. 81, pp 637-59.

BLATTBERG, R and N.GONEDES (1974) " A Comparison of Stable and Student Distributions as Statistical Models for Stock Prices", *Journal of Business*, Vol. 47, pp244-80.

BOLLERSLEV, T., (1986). "Generalised Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

BOOKSTABER,R., and J. McDonald (1987) " A general Distribution for Describing Security Price Returns" *Journal of Business*, Vol. 60 pp 401-24.

BROCK and de LIMA (1996) "Nonlinear Time Series, Complexity Theory and Finance" in G. S. MADDALA AND C. R. RAO (eds.), *Handbook of Statistics, Volume 14: Statistical Methods in Finance* Amsterdam: Elsevier Science Publishers, pp. 317-361

BURGSTAHLER D, JIAMBALVO, J and NOREEN, E. (1989): "Changes in the probability of bankruptcy and equity value" *Journal of Accounting and Economics* 11, 207-224

BUY, R., KAMINSKI, V., PINNAMANENI, K., and SHANBHOGUE V. (1998) "Actively Managing Corporate Credit Risk" in *Credit Derivatives, Risk Books.*

CARTY and LIEBERMAN, (1996): Moody's Investors Service Global Credit Research.

CHEN, K., AND T.SHIMERDA, "An Empirical Analysis of Useful Financial Ratios" Financial Management, Spring 1981.

COLES, S. (2001), An Introduction to Statistical Modelling of Extreme Values, Springer.

COX, D.R., AND SNELL, E.J. (1989). The Analysis of Binary Data. Second Ed.. London, Chapman and Hall.

COTTER, J., (1999). "Margin Exceedences for European Stock Index Futures using Extreme Value Theory", preprint, University College Dublin.

COTTER, J. (1998). "Testing Distributional Models for the Irish Equity Market", *Economic and Social Review*, 29, 257-269.

COTTER, J. & MCKILLOP, D.G. (2000). "The Distributional Characteristics of Selection of Contracts Traded on the London International Financial Futures Exchange" *Journal of Business Finance and Accounting*, pp 15-27

COSSIN, D. and PIROTTE, H., (2000): "*Advanced Credit Risk Analysis*" Wiley.

CHRISTOFFERSEN, P. F., (1998), "Evaluating Interval Forecasts", *International Economic Review*, 39, 841-862.

CHRISTOFFERSEN, P. AND F. DIEBOLD (2000), "How Relevant is Volatility Forecasting for Financial Risk Management?", *Review of Economics and Statistics*, 82, 12-22.

CHRISTOFFERSEN, P., F. DIEBOLD, AND T. SCHUERMANN (1998), "Horizon Problems and Extreme Events in Financial Risk Management", *Economic Policy Review*, Federal Reserve Bank of New York, October, 109-118.

CRAMER, J. "Predictive Performance of the Binary Logit Model in Unbalanced Samples." *Journal of the Royal Statistical Society, Series D (The Statistician)* 48, 1999, pp.85-94.

CROUCHY, M. D. GALAI AND R. MARK, (2000): "A comparative Analysis of Current Credit Risk Models" *Journal of Banking and Finance* 24 pp59-117.

CROUCHY, M., D. GALAI AND R. MARK , (2001): "Risk Management" McGrawHill.

DAMBOLENA, I.G. and KHOURY S.J. "Ratio Stability and Corporate Failure" *The Journal of Finance*, Vol. XXXV, No 4, September, 1017-1026.

DANIELSSON, J., AND C.G. DE VRIES, C.G. (2000). "Value at Risk and Extreme Returns" *Annals d' Economie et de Statistique* 60 :239-269.

DANIELSSON, J., AND C.G. DE VRIES, C.G. (1997a) "Tail index and quantile estimation with very high frequency data". *Journal of Empirical Finance* 4:241-257.

DANIELSSON, J., AND C.G. DE VRIES, C.G. (1997b). "Value at risk and extreme returns". Financial Markets Group Discussion Paper no. 273, London School of Economics.

DIEBOLD, F.X. A. HICKMAN, A., A. INOUE, A. AND T. SCHUERMANN. (1998), Scale Models, *Risk*, 11, 104-107.

DINENIS, E (2002a): "Lecture Notes on the Measurement of Market Risk", Cass Business School, City University London.

DINENIS, E (2002b): "Lecture Notes on Bankruptcy Prediction and Credit Risk Measurement" Cass Business School, City University London.

DINENIS E (2003a): "Spotting Failing Companies: Evidence on the usefulness of financial ratios from Greek SMEs" Cass Business School, City University London.

DINENIS E (2003b): "Developing a Credit Quality Index for Greek Companies" Cass Business School, City University London.

DINENIS, E and T. PRINIOTAKIS (2002) "Alternative Statistical Distributions for the Greek Stock Market Returns: An empirical investigation" Mimeo.

DINENIS, E and T. PRINIOTAKIS (2003): "Alternative Volatility Models for the Greek Stock Market Index: An empirical evaluation" Mimeo.

DING, Z., & GRANGER, C. W. J. (1993). "A Long Memory Property of Stock Market Returns and a New Model", *Journal of Empirical Finance*, 1, 83-106.

DING, Z., AND GRANGER, C. W. J., (1996). "Modelling Volatility Persistence of Speculative Returns: A new approach", *Journal of Econometrics*, 73, 185-215.

DA SILVA , A.L.C., AND B.V. DE MELO MENDES (2003): "Value at risk and Extreme Returns in Asian Stock Markets" . *International Journal of Business* 891) pp 17-40

DEAKIN, E. B. (1976), *"Distributions of Financial Accounting Ration: Some Empirical Evidence"*, The Accounting Review, *January, pp. 90-96*

DIMITRAS, A.I, ZANAKIS, S.H., and ZOPOUNIDIS, C, (1996): "A survey of business failures with an emphasis on prediction methods and industrial applications" *The European Journal of Operational Research*, 90, pp 487-513.

DOWD, K. (2000) "Assessing VaR Accuracy" *Derivatives Quarterly*, Spring, pp 61-63.

DROST, F.C., AND T.E. NIJMAN (1993) "Temporal Aggregation of GARCH Processes" *Econometrica* 61, 909-927.

DUFFIE, D., AND PAN, J. 1997. An overview of value at risk. *Journal of Derivatives*, Vol. 4, No. 3 (Spring) pp. 7–49.

DUMOUCHEL, W. H. (1983). Estimating the Stable Index □ in order to Measure Tail Thickness: A Critique, *Annals of Statistics*, 11, 1019-1031.

EFRON, B (1975) "The efficacy of logistic regression compared to normal discriminant analysis" *Journal of the American Statistical Association* 70, 892-898.

EISENBEIS, R.A. *(1977)* "Pitfalls in the application of discriminant analysis in business, finance and economics" *The Journal of Finance, Vol. XXXII, No. 3 June. pp. 875-900.*

EMBRECHTS, P., KLUPPELBERG, C., & T. MIKOSCH, T. (1997). Modelling Extremal Events, Berlin: Springer Verlag.

EMBRECHTS, P., RESNICK, S. & SAMORODNITSKY, G. (1998). Extreme Value Theory as a Risk Management Tool, *North American Actuarial Journal*, 3, 12-27.

ENGEL R. F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". *Econometrica* 50:987-1007.

ESPAHBODI, H and ESPAHBODI, P. *(2003) "Binary Choice Models and Corporate Takeover".* Journal of Banking and Finance, *27, pp. 549-574.*

EZZAMEL, M., BRODIE, J. AND MAR-MOLINERO, C. (1987), "Financial Patterns of UK Manufacturing Companies", *Journal of Business Finance & Accounting*, Vol. 14, No. 4

FAMA, E. (1965) "The behaviour of Stock Market Prices" *Journal of Business*, Vol. 38, pp 34-105.

FELLER, W., (1971). An Introduction to Probability Theory and its Applications, New York: John Wiley.

FISHBURN, P.C.(1977). "Mean-Risk Analysis with Risk Associated with Below-Target Returns." *American Economic Review, 67:116-126.*

FISHER, R. A., and TIPPETT, L. H. C. 1928. "Limiting forms of the frequency distribution of the largest or smallest member of a sample." *Proc. Cambridge Philos. Soc.* 24:180-190. ..

FOREMAN, R.D. (2003) "A logistic analysis of bankruptcy within the US local telecommunication industry" *Journal of Economics & Business* 55, pp. 135-166.

FOSTER, G. (1986) "Financial Statement Analysis" Second Edition, Prentice-Hall, Engelwood Cliffs, New Jersey

FRESKA., T. J. AND HOPWOOD, W. S. (1983), "The Effects of Outliers on the Cross-Sectional Distributional Properties of Financial Ratios*",* *The Accounting Review*, Vol. LVIII, No 1, pp. 115-127

FLURY, B. (1997): A first course in multivariate statistics, Springer.

FULMER, J.G. JR., MOON, J.E., GAVIN, T. A., ERWIN, M.J., "A Bankruptcy Classification Model For Small Firms". *Journal of Commercial Bank Lending* (July 1984): pp. 25-37.

HAWKINS, D.M.and McLACHLAN, G.J., "High-breakdown linear discriminant analysis.*" Journal of the American Statistical Association 92 (437), pp 136-143.*

HENDERSON and VELLEMAN *(1981)* "Building multiple regression models interactively*". Biometrics, 37:391-411*

HOSMER DW AND S LEMESHOW, "*Applied Logistic Regression*" Wiley 1989

HOUGHTON, K.A. (1984): "Accounting data and the prediction of business failure: the setting of priors and the age of data" *Journal of Accounting Research* Vol.22 Spring, 361-368

HUBERTY, C.J. (1994): *"Applied Discriminant Analysis*" Wiley.

GALAMBOS, J. 1978. *The Asymptotic Theory of Extreme Order Statistics*. New York: John Wiley & Sons.

GARMAN AND KLASS (1980) "On the estimation of Security Price Volatilities from Historical data" *Journal of Business*, 53, 67-78

GIOT PIERRE AND SÉBASTIEN LAURENT, "Value-at-Risk for Long and Short Trading Positions", *Journal of Applied Econometrics*, Vol. 18, No. 6, 2003, pp. 641-664.

GLASSERMAN, P, HEIDELBERGER, P and SHAHABUDDIN, P (2000) " Portfolio Value-at-Risk with Heavy-Tailed Risk Factors" Columbia University Working Paper.

GLUBOS, G and GRAMMATIKOS, T. (1988): "The success of bankruptcy prediction models in Greece", *Studies in Banking and Finance* 7, 37-46.

GNEDENKO, B.V. (1943). "Sur la distribution limite du terme maximum d'une serie aleatoire". *Annals of Mathematics* 44:423-453.

GORDY, M. (2000): "A comparative anatomy of credit risk models" *Journal of Banking and Finance,* 24, 119-49.

GUERMAT, C. and HARRIS, R.D.F., (2002) "Forecasting Value at Risk Allowing for Time Variation in the Variance and Kurtosis of Portfolio Returns" International *Journal of Forecasting* 18, 409-419.

GREEN W.H. (2000): "Econometric Analysis" Fourth Edition, Prentice Hall.

GUMBELL, E. J., (1958). *Statistics of Extremes*, New York: Columbia University Press.

IZAN, H.Y. (1984): "Corporate distress in Australia" *Journal of Banking and Finance* 8, 303-320

HALL, P. (1982). On some simple Estimates of an Exponent of Regular Variation, *Journal of the Royal Statistical Society*, Series B, 44, 37-42.

HANSEN, B. (1994) "Autoregressive Conditional Density Estimation" *International Economic Review*, 35, 705-730.

HAUCK W.W., and A. DONNER (1977) "Wald's test as applied to hypotheses in logit analysis" *Journal of the American Statistical Association*, 72: 851-853

HENDRICKS, D. 1996. Evaluation of value-at-risk models using historical data. *Economic Policy Review Federal Reserve Bank of New York*, April:39-69.

HILL, B. M. (1975). "A Simple General Approach to Inference About the Tail of a Distribution" *Annals of Statistics* 3: 1163-1174.

HOPPER, G. P. 1996. Value at risk: A new methodology for measuring portfolio risk. *Business Review, Federal Reserve Bank of Philadelphia*, July/August:19-29.

HUISMAN, R, C. COEDIJK, C.KOOL and F. Palm (2001) "Fat tails in small samples" *Journal of Business and Economic Statistics*, 3, 208-216.

HULL, J, AND WHITE, A. 1998. "Value at risk when daily changes in market variables are not normally distributed". *Journal of Derivatives* Spring:9-19.

HULL, J., & WHITE, A. (1998). Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk, *Journal of Risk*, 1, 5-19.

HARRIS, R.D.F AND C.C. KUCUKOZMEN (2001): "The Empirical Distribution of UK and US Stock Returns" *Journal of Business Finance and Accounting*, 28(5) & (6), June/July, pp715-740.

HO, L-C, BURRIDGE, P, CADDLE J, AND M. THEOBALD (2000); "Value at Risk: Applying the extreme value approach to Asian markets in the recent financial turmoil" *Pacific-Basin Finance Journal*, 8, pp 249-275

JACKSON, P.; MAUDE, D. J.; AND PERRAUDIN, W. 1997. Bank capital and value at risk. *Journal of Derivatives* Spring:73-90.

JANSEN, D.W., AND DE VRIES, C.G. (1991). "On the frequency of large stock returns: Putting booms and busts into perspectives". *Review of Economics and Statistics* 73:18-24.

JENKINSON, A. F. 1955. "The frequency distribution of the annual maximum (or minimum) values of meteorological elements". *Quarterly Journal of the Royal Meteorology Society* 87:145-158.

JENKINSON, A. F. 1(969). "Statistics of extremes". In *Estimation of Maximum Flood*, Technical Note 98, World Meteorological Organization, Geneva, 183-227

JOHNSON, R.A. (1978), "The Cross-Sectional Stability of Financial Patterns", *Journal of Business Finance and Accounting*, Vol. 5, No 2 (Summer 1978), pp. 207-214

JOHNSON, N., KOTZ, S., and N. BALAKHRISNAN (1994) " Continuous Univariate Distributions" Wiley .

JOHNSON, R.A AND WICHERN, D.W. (1998), "Applied Multivariate Statistical Analysis" Fourth Edition, Prentice Hall, Upper Saddle River, New Jersey

JONES, F. (1987): "Current Techniques in Bankruptcy Prediction" *Journal of Accounting Literature*, 6, 131-164.

JUDD CM. AND MCCLELLAND GH. (1989) *"Data Analysis: A Model Comparison Approach"* Harcourt Brace Jovanovich

KEARNS, P., & PAGAN, A. (1997). Estimating the Density Tail Index for Financial Time Series, The Review of Economics and Statistics, 79, 171-175.

KOEDIJK, K. & KOOL, C. J. M. (1992). Tail Estimates of East European Exchange Rates, Journal of Business and Economic Statistics, 10, 83-96.

KENNEDY, D., LAKONISHOK, J. AND SHAW, W.H. (1992), "Accommodating Outliers and Nonlinearity in Decision Models", *Journal of Accounting, Auditing & Finance*, Vol. 7, No. 2, pp. 161-190

KENT J.T. (1983) "Information gain and a general measure of correlation" *Biometrika* 70, 163-173

KELHOFER, Portfolio Management of Default Risk. Document Number: 999-0000-033, *KMV Corporation*, 1997

KON, S. (1984) "Models of Stock Returns-A Comparison" *Journal of Finance,* Vol. 39, pp 147-65.

KOYLUOGLU, H.U and HICKMAN, A., (1998): "Reconciling the Differences" *Risk,* 11(10) pp. 56-62.

KRAUSE, ANDREAS (2003), "Exploring the Limitations of Value at Risk: How Good is it in Practice?", *Risk Finance*, Winter , pp 19-28.

KUPIEC, P. H. (1995). "Techniques for verifying the Accuracy of Risk Measurement Models" *Journal of Derivatives* 3, 73-84.

KUPIEC, P. H. (1999). Risk capital and VaR. *Journal of Derivatives* Winter:41-52.

KHINDAROVA, I., RACHEV, S. and SCHWARTZ, E. (2001). "Stable Modelling of Value at Risk", *Mathematical and Computer Modelling* 34: 1223-1259.

LAITINEN, T and KANKAANPAA, M. (1999) " Comparative Analysis of Failure Prediction Methods: the Finish Case" The *European Accounting Review* 8:1, pp. 67-92.

LAMBERT P. and J.K. LINDSEY (1999) "Analysing financial returns by using regression models based on non-symmetric stable distributions" *Journal of Applied Statistics*, 48, 409-424.

LANDWEHR, J.M., PREGIBON, D., SHOEMAKER, A.C. (1984) "Graphical Methods for Assessing Logistic Regression Models" *Journal of the American Statistical Association*, 79, 61-71.

LANE, W., LOONEY, S., and WANSLEY, J. (1986) " An application of the Cox proportional hazards model to bank failure" *Journal of Banking and Finance*, 10: 511-31.

LEMESHOW, S. AND HOSMER, D.W. (1982) "The use of goodness-of-fit statistics in the development of logistic regression models." *American Journal of Epidemiology*, 115, 92-106.

LEVY, P. (1925). *Calcul des Probabilites*, Gauthier Villars.

LAURENT, C. R. (1979), "Improving the Efficiency and Effectiveness of Financial Ratio Analysis" *Journal of Business Finance and Accounting*, Vol. 6, No 3 (Autumn 1979), pp. 401-413

LINCOLN, MERVYN, (1984): "An empirical study of the usefulness of accounting ratios to describe levels of insolvency risk" *Journal of Banking and Finance,* 8, 321-340.

LO, A. (1986): "Logit versus Discriminant Analysis: A Specification Test and Application to Corporate Bankruptcies," *Journal of Econometrics*, 31, 151-78.

LONGIN, F.M. (1996): "The Asymptotic Distribution of Extreme Stock Market Returns" *Journal of Business*, Vol. 69, No. 3, pp 383-408

LONGIN, F.M. (2001): "Stock market crashes: Some quantitative results based on extreme value theory" *Derivative Use, Trading &Regulation*, Vol. 7, No. 3, pp 197-205.

LONGIN, F. M., (1999a). From Value at Risk to Stress Testing, The Extreme Value Approach, *Journal of Banking and Finance*.

LONGIN, F. M., (1999b). Optimal Margin Levels in Futures Markets: Extreme Price Movements, *Journal of Futures Markets*, 19, 127-152.

Lopez, J. A. 1998. "Methods for evaluating value-at-risk estimates." *Economic Policy Review, Federal Reserve Bank of New York* October:119-124.

LORETAN, M. & PHILLIPS, P. C. B. (1994). Testing the Covariance Stationarity of Heavy-tailed Time Series, *Journal of Empirical Finance*, 1, 211-248.

LUX, T. (2001) "The limiting extremal behaviour of speculative returns: an analysis of intra-daily data from the Frankfurt Stock Exchange" *Applied Financial Economics*, 11, 299-315.

MCDONALD, J (1996), "Probability Distributions for Financial Models" in G.S. MADDALA AND C.R. RAO (eds.) *Handbook of Statistics*, Vol. 14, pp 427-61.

MCDONALD, J and Y.YU (1995) " A generalisation of the Beta Distribution with Applications" *Journal of Econometrics,* Vol. 66, pp. 133-52.

MCNEIL, and R.FREY (2000) "Estimation of Tail-Related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach" *Journal Empirical Finance*, 7, 271-300.

MANDELBROT, B. (1963). "The variation of certain speculative prices". *Journal of Business* 36: 394 -419.

MANDELBROT, B. (1997) "Fractals and Scaling in Finance", Springer.

MARAIS, M.L., PATELL, J.M., and WOLFSON, M.A. (1984): "The experimental design of classification models: An application of recursive partitioning and bootstrapping to commercial bank loan classifications", *Journal of Accounting Research 22,* 87-113

MARKOWITZ, H.M.(1991). *Portfolio Selection: Efficient Diversification of Investments*, Second Edition, Cambridge, MA: Basil Blackwell.

MARTIKAINEN, T., PUHALAINEN, K., YLI-OLLI, P., AND GUNASEKARAN, A. (1995), "On the Time-Series Stability of Industry-Relative Financial Ratio Patterns", Vol. 26, No 9, pp. 1701-1713

MENSAH, Y.M. (1984): "An examination of the stationarity of multivariate bankruptcy prediction models: a methodological study" *Journal of Accounting Research,* Vol. 22 No 1, Spring, 380-395
MERTON, R., "On the pricing of Corporate debt: The risk structure of interest rates" *Journal of Finance* 28, 449-470.
MICHA, B. (1984): "Analysis of Business Failures in France" *Journal of Banking and Finance* 8, 281-291
MICHALOPOULOS, M., ZOPOUNIDIS, C., and KOURIS, I. (1993): "An integrated DSS for the assessment of the risk of failure", in *Applied Stochastic Models and Data Analysis*, World Scientific Publishing, New Jersey, 606-625.
MITTNIK , S and S. RACHEV (1993) "Modelling Asset Returns with alternative stable distributions" *Econometric Reviews* 12, 261-330.
MOODY'S PUBLIC FIRM RISK MODEL, (2000): "A hybrid Approach to Modelling Short Term Default Risk", March.
MOODY'S INVESTOR SERVICES, (2000): "RiskCalc for Private Companies: Moody's Default Model", May.
MORRIS, R. (1997): "Early Warning Indicators of Corporate Failure", Ashgate Publications.
NEFTCI, S.N. (2000) "Value at Risk Calculations, Extreme Events, and Tail Estimation" *The Journal of Derivatives*, Spring, pp 1- 15
NICKELL, P., W. R. PERRAUDIN, AND S. VAROTTO (2001): "Ratings- Versus Equity-Based Credit Risk Models: An Empirical Study," *Bank of England Quarterly Bulletin*, 41, pp: 215 - 1
NOLAN, J. P. (1997). Numerical Calculation of Stable Densities and Distribution Functions, *Communications in Statistics - Stochastic Models* 13: 759-774.
OHLSON, J. (1980) "Financial Ratios and the Probabilistic Prediction of Bankruptcy" *Journal of Accounting Research* 19, 109-131
ONG, M.K., (1999): "Internal Credit Risk Models: Capital allocation and performance Measurement" *Risk Books.*
PAGAN, A. R., (1996). The Econometrics of Financial Markets, *Journal of Empirical Finance*, 3, 15-102.
PALEPU, K. (1986) "Predicting Takeover Targets: A Methodological and Empirical Analysis" *Journal of Accounting and Economics*, 8, 3-35.
PINCHES, G.E., MINGO, K.A. AND CARUTHERS, K.J. (1973), "The Stability of Financial Patterns in Industrial Organizations", *Journal of Finance*, Vol. 28, pp. 369-396
PLATT,H., and M. PLATT (1990): "Development of a Class of Stable Predictive Variables: The Case of Bankruptcy Prediction," *Journal of Business, Finance and Accounting*, 17(1), 31-51.
PORTER, R.B.(1974). "Semivariance and Stochastic Dominance: A Comparison." *American Economic Review,* 64:200-204.
PREGIBON, D. (1981). "Logistic regression diagnostics" *Annals of Statistics,* 9, pp 705-724.
PRITSKER, M. (2001), The Hidden Dangers of Historical Simulation, *Federal Reserved Board.*
PARKINSON, M. 1980. The extreme value method for estimating the variance of the rate of returns. *Journal of Business* 53:61-65.

PHOA, W. 1999. Estimating credit spread risk using extreme value theory. *Journal of Portfolio Management* Spring:69-73.
PICKANDS, J. (1975). " Statistical inference using extreme order statistics". *Annals of Statistics* 3:119-131.
PRAETZ, P. D. (1972). The distribution of share price changes. *Journal of Business* 45:49-55.
PRESCOTT, P., AND WALDEN, A.T. 1980. Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution. *Biometrica* 67:723-724.
PRESCOTT, P., and WALDEN, A.T. 1983. "Maximum-likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution from censored samples" *Journal of Statistical Computation Simulation* 16:241-250.
RACHEV, S. AND MITTNIK, S. (2000). *Stable Paretian Models in Finance*, Wiley.
REISS, R.D. AND M. THOMAS (2000): " Statistical Analysis of Extreme Values" 2nd edition. *Birkhause Verlag.*
RISKMETRICS, (1996): "Technical Document" *J.P. Morgan* 3rd edition.
RICHARD, J.F., (1975), "A note on the information matrix of the multivariate normal distribution" *Journal of Econometrics* 3, 57-60.
RICHARDSON, F. M. AND DAVIDSON, L.F. (1984), "On Linear Discrimination with Accounting Ratios", *Journal of Business Finance and Accounting*, Vol. 11, No 4 (Winter 1994), pp. 511-525
SAMORODNITSKY, G. AND TAQQU, M. S. (1994). *Stable Non-Gaussian Random Processes*, Chapman & Hall.
STEPHENS, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons, *Journal of the American Statistical Association* 69: 730-737.
SANTOMERO, A., AND J.VINSO (1977): "Estimating the Probability of Failure for Commercial Banks and the Banking System," *Journal of Banking and Finance* 1.
SALMI, T., VIRTANEN, I. AND YLI-OLLI, P. (1990), "On the Classification of Financial Ratios. A Factor and Transformation Analysis of Accrual, Cash Flow, and Market-Based Ratios" Acta Wasaensia. Universitas Wasaensis. Vaasa
SAUNDERS, A. "Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms" Wiley 1999.
SCOTT, J. (1981): "The probability of bankruptcy: a comparison of empirical predictions and theoretical models" *Journal of Banking and Finance* 5, 317-344.
SHARMA, S (1996), *Applied Multivariate Techniques*, Wiley.
SHUMWAY, T. (2001): "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," *Journal of Business*, 74(1), 101-124.
SMITH R., (1987) "Estimating tails of probability distributions" *The Annals of Statistics* , 15, 1174-1207
STANDARD AND POOR'S (2000): "Adjusted Key U.S. Industrial Financial Ratios" Credit Week, 20th September.
SRINIVASAN, V and KIM, Y.H. (1988): "Designing expert financial systems: a case study of corporate credit management", *Financial Management*, Autumn, 32-44.

SUDARSANAM, P.S. AND TAFFLER, R.J. (1995), "Financial Ratio Proportionality and Inter-temporal Stability: an Empirical Analysis", Journal of Banking & Finance, Vol. 19, pp. 45-60

TAFFLER, R. J. AND SUNDARSANAM, P.S. (1979), "Auditing the Board: A New Approach to the Measurement of Company Performance", *Management Finance*, Vol. 5, No 2 (1979), pp. 127-147

TAFFLER, R.J: (1984) "Empirical models for the monitoring of UK corporations" *Journal of Banking and Finance* 8 pp 199-227

TAMARI, M. (1984) "The use of a bankruptcy forecasting model to analyse corporate behaviour in Israel", *Journal of Banking and Finance* 8, pp 293-302.

TATSUOKA, M.M. (1988), *Multivariate Analysis*, New York : Macmillan

THEODOSSIOU, P. (1998) "Financial Data and the Skewed Generalized-t Distribution", *Management Science*, Vol. 44, pp. 1650-61.

TONG, Y.L. (1990) *The Multivariate Normal Distribution*, Springer, New York.

WATSON, COLLIN J. (1990), "Multivariate Distributional Properties, Outliers, and Transformation of Financial Ratios", *The Accounting Review*, Vol. 65, No. 3, pp. 682-695

WESTGAARD, S AND VAN DER WIJST, N.(2001): "Default Probabilities in a corporate bank portfolio: A logistic model approach", *European Journal of Operational Research*, 135, pp 338-349.

WILSON (1997a) "Portfolio Credit Risk, Part I", *Risk*, September , pp. 111-17

WILSON (1997b): "Portfolio Credit Risk, Part II", *Risk*, October , pp. 56-61

ZAVGREN, C.Y., (1983) "The Prediction of Corporate Failure: The State of the Art" *Journal of Accounting Literature*.

ZAVGREN, C.Y., (1985) "Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Analysis" *Journal of Business Finance and Accounting*.

ZMIJEWSKI, M. (1984): "Methodological Issues Related to the Estimation of Financial Distress Prediction Models," *Journal of Accounting Research*, Supplement, 59-86.

VILASUSO, J. AND D. KATZ (2000): "Estimates of the likelihood of extreme returns in international stock markets" *Journal of Applied Statistics*, Vol. 27, No. 1, pp 119-130