STATISTICAL ASPECTS OF THE PORTFOLIO CONSTRUCTION PROBLEM

ALEXANDER BELGORODSKY

Ph.D. Thesis 2007
STATISTICAL ASPECTS OF THE PORTFOLIO CONSTRUCTION PROBLEM

Alexander Belgorodski

Centre for Operational Research and Applied Statistics,
Management and Management Sciences Research Institute,
Salford Business School,
University of Salford, Salford, UK

Submitted in Partial Fulfilment of the Requirements of the Degree of Doctor of Philosophy, October 2007
Table of Contents

List of Figures and Tables ........................................................................................................ iv

ACKNOWLEDGEMENTS ........................................................................................................ vi

The following publications and conference papers emerged from the study as a part of this thesis .......................................................................................................................... viii

Abstract ..................................................................................................................................... xi

CHAPTER 1 Introduction ........................................................................................................... 1
  1.1 Importance of the research .............................................................................................. 2
  1.2 Research objectives ....................................................................................................... 6
  1.3 Structure of the thesis .................................................................................................. 10

References .............................................................................................................................. 12

CHAPTER 2 The background to security analysis and portfolio theory ............................. 14
  2.1 Introduction .................................................................................................................. 15
  2.2 Implementation of Markowitz's mean-variance strategy .......................................... 21
  2.3 Limitations of the Markowitz approach .................................................................... 27
  2.4 The inputs to portfolio analysis ................................................................................ 30
  2.5 An overview of single-index model .......................................................................... 33
  2.6 International diversification of an investment ......................................................... 38
  2.7 International diversification, exchange rate risk is incorporated into consideration .... 44
  2.8 The capital asset pricing model ................................................................................ 48
  2.9 Efficient market hypothesis ....................................................................................... 50
  2.10 Failure of CAPM: irrational pricing or risk ............................................................ 59
  2.11 Fama and French three factor model ................................................................. 63
  2.12 Arbitrage pricing theory ......................................................................................... 65
  2.13 Regression-based approach to forecasting return ................................................ 66

References .............................................................................................................................. 68

CHAPTER 3 Transforming predictor variables to near normality .................................... 72
  3.1 Introduction ................................................................................................................ 73
  3.2 Transforming to near normality ............................................................................... 73
  3.3 Programming .......................................................................................................... 79
  3.4 Analysis of the results ............................................................................................ 81

References .............................................................................................................................. 88
CHAPTER 4 Improving the accuracy of return predictions .............................................89
4.1 Introduction..............................................................................................................90
4.2 Adjusting historical estimates of beta......................................................................92
4.3 Bayesian adjustments as applied to the regression-based model described in
section 2.13...............................................................................................................96
4.4 Detailed implementation of the maximization scheme: prior distribution of
δ is Gaussian........................................................................................................... 103
4.5 Detailed implementation of the maximization scheme: prior distribution of
δ is Student t........................................................................................................... 110
4.6 Programming methodology................................................................................... 113
References.................................................................................................................... 118

CHAPTER 5 Results: shrunken regression model.......................................................... 119
5.1 Introduction............................................................................................................ 120
5.2 Rival methods........................................................................................................120
5.3 Results.................................................................................................................... 124
5.4 Assessing prediction performance: shrunken regression.......................................131
References.................................................................................................................... 137

CHAPTER 6 Predicting stock returns and assessing prediction performance ............... 138
6.1 Introduction............................................................................................................ 139
6.2 Methodology.......................................................................................................... 141
6.3 Correlation analysis............................................................................................... 144
6.4 Metrics................................................................................................................... 149
6.5 Results: regression model...................................................................................... 162
6.6 Results: more complex models (improving the modelling of return).................... 166
References.................................................................................................................... 170

CHAPTER 7 Conclusions and further research .............................................................. 172
References.................................................................................................................... 179

APPENDICES.................................................................................................................180
Appendix 1: Exporting data from Perfect Analysis..................................................... 181
Appendix 2: The structure of the data file................................................................... 181
Appendix 3: Exponential utility and the certainty equivalent..................................... 183
Appendix 4: Covariance of returns between securities - international
diversification of an investment.............................................................................. 185
Appendix 5: A derivation of an analytical solution for the Markowitz investment
strategy – international diversification of an investment when exchange rate
risk is taken into consideration ............................................................................... 186
Appendix 6: Updating formula for the variance.......................................................... 188
Appendix 7: Results of transforming predictive variables to approximate
normality................................................................................................................. 190
Appendix 10: The listing of a fortran program that has been developed by the author

Appendix 11: Computation of PL - reducing rounding error

Appendix 12: Proofs of the properties of the new measure of investment performance (the value of risk aversion that makes investment valueless)
List of Figures and Tables

1 Figure 2.1 A typical scatter plot: $R_m - R_F$ versus $R_f - R_F$ with the fitted straight line from equation (2.5). The high scatter of points and the presence of outliers are typical. .................................................................35

2 Figure 3.1 An untransformed predictor variable: 2-monthly return December 1994-November 2004 with superimposed normal curve .......................................................82

3 Figure 3.2 A transformed predictor variable: 2-monthly return December 1994-November 2004 after Johnson’s transformation with superimposed normal curve ....83

4 Figure 3.3 An untransformed predictor variable: market capitalisation December 1994-November 2004 with superimposed truncated normal curve ................84

5 Figure 3.4 A transformed predictor variable: market capitalisation December 1994-November 2004 after Box-Cox transformation with superimposed normal curve .....85

6 Figure 3.5 A transformed predictor variable: market capitalisation December 1994-November 2004 after Johnson’s transformation with superimposed truncated normal curve (for different values of an upper bound for the variation of $c$) ........................................87

7 Figure 4.1 Convergence of the market capitalisation regression coefficient ..........101

8 Figure 4.2 Graphical representation of the maximisation scheme .....................102

9 Figure 4.3 The structure of the data. All possible situations of positional relationship between the data for the company and the data required for computing parameters and making a forecast are modelled. The aeries marked with blue represent the data utilised in computing regression coefficients and estimates of hyperparameters....115

10 Table 5.1 Regression coefficients delta according to four rival methods ........125

11 Figure 5.1 Estimates of $\delta_k$, $k = 1, \ldots, p$ according to the four methods .......126

12 Table 5.2 Estimates of the parameters $\beta$, $\sigma^2$ for the conventional model, market model and ridge regression approach .................................................................129

13 Figure 5.2 Estimates of beta for the conventional model and the ridge regression. The black bars correspond to the prior mean (0.88), which is the same for all the companies ........................................................................................................130
14 Figure 5.3 Percentage of time other methods perform better than conventional approach

15 Table 5.3 Percentage better: different methods are used as a benchmark

16 Figure 5.4 Percentage best: percentage of time when a given method outperforms all other methods

17 Table 5.4 MSE for different methods across the months

18 Figure 5.5 Difference in MSE of the methods in relation to ridge regression approach

19 Table 5.5 Median SE for different methods across the months

20 Table 6.1 Average annual Spearman correlation coefficients between predictor variables and return, from 1956 to the end of 2005. ‘Winter’ is unity from October through to May inclusive, otherwise zero. The ratio of average correlation to average absolute value of correlation is an indicator of the stability of the predictor.

21 Figure 6.1 Annual Spearman correlations between 5-year return and monthly return, with error bars, and for annual dividend yield with monthly return (error bars suppressed for clarity). Both predictors are comparatively stable with time.

22 Table 6.2 Regression coefficients for minimum-BIC rolling regression to the end of 2005. $R_m$ denotes market return for the month being forecast. Variables have been transformed to be approximately standard normal. The last column shows the z-value for inclusion of the variable.

23 Figure 6.2 Certainty-equivalent gains from different investment types as a function of risk aversion, using the exponential utility function. The ‘no regression line’ corresponds to setting $\delta = 0$ in (6.1). This type of plot could be a useful summary of investment performance.

24 Table 6.3 Out-of-sample performance of various regression methods, showing the average gain per unit estimated risk, its risk (as a standard deviation), the Sharpe ratio, a new statistic $\xi_0$ defined in (6.13), and the Sharpe ratio using below-target semivariance. Under 'comments', $n$ is the number of firms used in computing regression coefficients, and ‘trans’ denotes whether variables were transformed.
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my parents Vladimir Belgorodskiy and Elena Belgorodskaya who have sacrificed so much of their lives for me. I cannot even begin to thank them for all they contributed to my life.

I would like to express my special gratitude and respect to Professor Rose Baker for her precious time and attention, patience at the early stages of the research, constructive advice and criticism, and willingness to share her wisdom. This thesis would not be possible without Professor Baker's supervision. Finally I thank her for her optimistic view and remarkable ability to cheer up others and for an exquisite sense of humour.

I would also like to express my gratitude to Dr. Philip Scarf for giving me the opportunity to pursue a PhD study. My most sincere acknowledgements go to Dr. David Percy and Roland Fox for their continuous assistance.

I wish to express my appreciation to the Centre for Operational Research and Applied Statistics (CORAS) within the University of Salford for the help and support. Special thanks to the director of CORAS, Professor Khairy Kobbacy and research manager Susan Sharples. Without them, my life would have been much more difficult.
I am deeply indebted to Meena Patel, Artur Akbarov and Pushpa Rajaguru for helping me out through some rough times. Thanks are due also to my lab-mates Michael Quigley and Y. Xiao for their great company.

I have much pleasure in recording my gratitude to Professor Leonid Kossovich for his constant inspiration.

Last but not least, I would like to acknowledge the enormous contribution of my girlfriend Xiang Li for her love and patience especially during the last months of thesis writing. Her everlasting encouragement and faith gave me strength and inspiration to go through the challenges that I faced during the PhD.
The following publications and conference papers emerged from the study as a part of this thesis:

- **Publications**

  a) **Refereed Publications**


  b) **Other Publications**


- **Conference presentations**

  a) **Keynote presentations**

b) Presentations


Abstract

The area of finance poses many challenging problems to the decision maker. One of them is the modelling of the expected return on stocks and the covariance matrix of returns.

This thesis approaches the decision problem of choosing an optimum portfolio of stocks in which to invest from the point of view of statistical decision theory. We use regression methods to predict the expected monthly return on stocks and the covariance matrix between returns, the predictor variables being a company's 'fundamentals', such as dividend yield and the history of previous returns. Predictions are evaluated out of sample for shares traded on the London Stock Exchange from 1976 to 2005. Many modelling and inferential approaches are examined and evaluated, the main ones being shrinkage of regression coefficients, and transforming predictor variables to near normality.

It is important to use suitable statistics to make a fair comparison of the out-of-sample performance of rival methodologies. We review well-known measures of assessing investment performance, including Sharpe, Sortino and Omega ratios, and derive a new statistic from the exponential utility function. We also suggest a graphical aid which could be used as a useful summary of investment performance.
CHAPTER 1

Introduction
1.1 Importance of the research

The problem of portfolio construction is one of the most debated topics in financial economics. In general the portfolio problem is how to 'optimally' spread available capital over a number of financial instruments. These instruments may be stocks, bonds, commodities etc. Our focus is on investment in the stock market. One of the widely accepted approaches (which is described in section 2.2) declares that the optimal selection of stocks could be achieved by maximising the value of the investor's expected utility function at the end of the investment period. This is the approach used here. While the methodology assumes that the values of expected return on stocks and the covariance matrix of returns are available, in fact this is not the case. These quantities are not known precisely and must be estimated from data. Therefore the problem has a statistical aspect and predicting the expected monthly return on stocks as well as the covariance matrix between returns is essential for investment decisions.

The approach presented here is intended to shed light on the optimal investment problem. Contrary to the opinion prevailing among professional fund managers and disseminated widely among academics (e.g. Fama and French, 2004) that rather than to adopt any rational strategy investors should simply follow benchmarks such as market indices like the FTSE, we show that investments based on our model outperform the market and predicting return is worthwhile. We model returns on stocks using a regression on predictor variables and involve the single-index model to estimate covariance between returns. We investigate a vast number of modelling and inferential choices. The
A regression model can be fitted by least squares, with the option of fitting all predictors (listed in appendix 2) or of carrying out a minimum Akaike information criterion (AIC) or minimum Bayesian information criterion (BIC) regression. Instead of using least squares, ‘ridge’ type regression, where shrinkage of regression coefficients is introduced, can be adopted. We also explore the effect of transformation to normality of predictor variables.

Another important issue addressed in the thesis is how to evaluate the performance of regressions. To compare historical out-of-sample predictions of various models (including market tracking) we set up the portfolios based on model predictions. Unfortunately, there is no single statistic available to compare competing types of investment in the sense that the measures suggested in literature can lead to the controversial and often misleading results. One of the most used measures of “risk-adjusted returns” – the Sharpe ratio (Sharpe, 1966), defined as the ratio of average excess return over the risk-free rate to the standard deviation of returns, has an obvious weakness since it uses standard deviation as a measure of risk. Bernardo and Ledoit (2000) demonstrate this drawback by giving a “striking” example of the lottery ticket costing one cent today which guarantees a 10% probability of receiving a prize of $50 billion next year and nothing otherwise. Clearly, any risk-averse investor would regard this gamble as an attractive opportunity. The lottery however gives a Sharpe ratio of only 0.33, which is below the U.S. stock market index. The contradiction is related to the fact that the Sharpe ratio rests on mean-variance theory which requires either a quadratic utility function or normally distributed returns. Since the return distribution is skewed (e.g. Arditti, 1967;
Samuelson, 1970), variance (or standard deviation) cannot be regarded as a sufficient measure of risk. In this case higher moments such as skewness and kurtosis also need to be taken into account.

Using a suitable statistic is crucial, and we develop a new measure of performance that apart from avoiding the shortcoming shown above has a number of desirable properties and is based on the use of an exponential utility. We also suggest a graphical tool for the purpose of assessing performances of different risky assets and portfolios.

We forecast returns on stocks one month ahead. As a result, the outcomes of the research may be particularly interesting for investors who are prepared to regularly adjust their portfolios. Some investors would wish to forecast returns over a longer period, such as a year, if following a ‘buy and hold’ strategy. Our methodology could then still be used although the need to forecast some of the predictor variables makes this nontrivial.

This research is focused on the UK stock market. There is a particular reason why analysing UK data is interesting. Most studies are based on the US stock market, therefore the UK stock market represents a largely fresh dataset. This ameliorates to some extent the inauspicious effect known as ‘data snooping’. A large body of research (e.g. Cooper and Gulen, 2006, p. 1264) shows that many researchers “collectively condition their studies on existing empirical regularities with the unintended consequence of snooping the data”. The overuse of data both to suggest useful variables to include in
models and to fit the models is a dangerous practice which may invalidate the assumptions underlying conventional statistical inference.

"Data snooping occurs when a given set of data is used more than once for purposes of inference or model selection. When such data reuse occurs, there is always the possibility that any satisfactory results obtained may simply be due to chance rather than to any merit inherent in the method yielding the results" (White, 2000, p. 1097).

What is troublesome is that we may choose ‘good’ predictive variables (for return) such as market capitalisation or dividend yield with the hindsight, based on previous successful papers that were to some extent contemporaneous to the studies from which we collected the data. The problem of data snooping is also related to look-ahead bias. By definition, look-ahead bias occurs when the information which is not available during the time period being analysed is used in the analysis. In this research this bias was largely absent, because only information available at any time was used to forecast following month’s return. However, as White (2000, p. 1097) pointed out this difficulty is endemic and impossible to be removed completely since in “in the analysis of time-series financial data, as typically only a single history measuring a given phenomenon of interest is available for analysis”. The effect can only be alleviated by employing a fresh dataset in simulation.
1.2 Research objectives

This research aims to forecast returns (on individual companies) as well as possible, and also their joint distribution, summarised by the covariance matrix. Company 'fundamentals' such as dividend yield and history of previous returns are used as predictor variables. The thesis aims to contribute in terms of the evaluation of many modelling and inferential approaches for which contradictory claims have been made in literature. We do not however seek to build up a complete strategy for portfolio construction. This would require further analysis such as incorporating the fact that shares have a discrete nature and cannot be bought or sold in any amount.

It is not our aim to address the vexed and seemingly insoluble question of market efficiency, but rather to develop a pragmatic approach. However our claim that such an approach outperforms the market does have limited negative implications for the concept of market efficiency.

The data for research are obtained from LSPD 2003, 2005 (the London Share Price Database, maintained by the London Business School) and Perfect Analysis (HYDRA). The LSPD2005 gives complete listing of firms from 1975 to 2005, and a random sample before that back to 1955. When we commenced this research in October 2004 LSPD 2005 was not available. To test the scheme by making real rather than simulated investments we had to combine LSPD 2003 with the 'Perfect analysis' database, which makes share price data available up to the present day ('Perfect analysis' is updated at the
end of each month). Appendix 1 describes the methodology of exporting data from Perfect Analysis. The results presented in Chapter 5 are based on this combined dataset. However, it was no longer relevant to combine databases after the release of LSPD 2005.

The databases supply month-end share prices and the monthly returns we aim to predict. The predictive variables used were either present in the database, such as market capitalisation, or dividend or could be computed from the information that was given. For instance, annual dividend yield was calculated as a ratio of the annual dividend income per share received from a company to its current share price.

The survivorship bias problem (e.g. Haugen and Baker, 1996) is absent in this research since dead companies as well as live companies were included into consideration. In fact the analysis presented in Chapter 6 assumes that we are restricting ourselves to investment in larger companies (such as FTSE 350) only. This helps to avoid the problems associated with the liquidation of companies. In more detail when companies liquidate, the database might supply an inaccurate return for the last month of life because liquidation dividends are sometimes paid late and often it is difficult to determine exactly what an investor would receive. Investing in larger firms helps to ameliorate this problem since it is very unlikely that a firm in a sample will liquidate. It would drop out of the FTSE350 long before it did so. A model fitted to firms that are currently in the FTSE350 would lead to a survivorship bias problem since it would predict that shares in small companies were a good investment because many firms that were small in the past are now large enough to be included in the sample. To avoid this difficulty, we must also
include in the sample those companies that decreased in size and maybe even liquidated.

The study presented in Chapter 6 of the thesis avoids survivorship bias by including only the largest $m$ firms each month in the regression data sample and making predictions for the largest $n$ companies (typically, 350).

Appendix 2 outlines the structure of the data file and gives a detailed account of the predictor variables used.

The choice of the predictive variables is one of the central points in the problem of return prediction. Studies normally employ a fixed set of three to five variables. De Bondt and Thaler (1985), Jagadeesh and Titman (1993) suggested the use of previous return history. A number of studies (e.g. Dimson and Marsh, 1999) utilise company's market capitalisation as a predictive variable. Other prominent variables in literature include various financial ratios such as dividend yield, book-to-market ratio and earning-price ratio (Lewellen, 2004). Non-traditional variables include raininess and snowiness, distance of a trader from the corporate headquarters of the traded stock etc. Forecasting return is a challenging statistical problem. As Dangl and Halling (2006) pointed out "yet no consensus exists on the fundamental questions: whether predictability exists and which variables show best predictive performance". This research does not claim to answer this question. There are very many possible predictors of return, but they are very poor, so that $R^2 \approx 2\%$ at best (where $R$ is the correlation between the predicted and actual returns). The number of predictors can be substantially increased by including
interactions between predictors. Moreover, although some variables, such as, for example, the book-to-market ratio, are suggested to be good predictors of return (Lawellen, 2004), they are quite difficult to obtain.

Another problem is that many researchers in the financial area focus their attention on demonstrating the value of some predictor by constructing portfolios from deciles of companies ranked by it. There is however a regression-based approach to forecasting return (Kandel and Stambaugh, 1996). This work aims to add something to the literature of return prediction, through the use of the regression-based approach.

Some of the predictor variables such as market capitalisation have a distribution with very high skewness and kurtosis. They could not be used in their 'raw' form since a very skew-distributed predictor can only influence predicted returns for the largest companies which leads to the effective loss of this variable. Foster et al. (1997, p. 593) drawn attention to the fact that, “there are limitless possible linear and nonlinear transformations of these variables”. In this thesis transforming predictor variables to approximate normality was tried.

We have used Fortran 95 together with the NAG (Numerical Algorithms Group) mathematical software library, Microsoft Excel 2000 and statistical package Minitab version 13 to explore the data and forecast return on stocks. In fact, for the purpose of analysing the data, a series of Fortran 95 programs has been written. These programs involve the NAG library to solve the problems of searching the minimum, in a given
finite interval, of a function of a single variable, calculating the solution of a set of linear
equations and minimising the function of several variables. Microsoft Excel 2002 and
Minitab ver.13 were employed to check the results of the programs and plot the diagrams.

1.3 Structure of the thesis

This thesis is organised as follows.

Chapter 2 presents a literature review which seeks to put this work into an appropriate
academic context. Here we describe the previous work done in related areas. It starts with
the implementation of portfolio theory. It also addresses the question of market
efficiency and outlines basic models for return prediction (including regression-based
approach). This is not however simply a review chapter – it mixes up well-known theory
and some new research, presented in sections 2.6 and 2.7 (international diversification of
investment).

Chapter 3 is concerned with the problem of transforming predictor variables to
approximate normality. The Box-Cox transformation and Johnson’s arcsinh
transformation and their application to the data are discussed here.
Chapter 4 provides the methodology of estimating Bayesian adjusted model parameters (regression coefficients). We present the mathematics involved and summarize the programming methodology.

Chapter 5 reports the results obtained on applying the methodology described in Chapter 4 to the stock market data. We present an out-of-sample evaluation of return predictions. The performance of shrunken regression (Bayesian approach) is compared with several (Non -Bayesian) rival methods. Measures of forecasting accuracy utilised in M3 forecasting competitions are used for this analysis.

While Chapters 4 and 5 deal with shrunken estimators only, Chapter 6 examines many more modelling approaches including minimum AIC and BIC regressions. Here, as compared with Chapter 5, we also explore more inferential approaches. Using suitable statistics to enable the out-of-sample performance of competing methodologies to be compared is crucial. For this reason, in this chapter we develop some new measures of portfolio performance that, unlike the measures introduced in Chapter 5, are related to the financial cost of making a bad prediction. The chapter concludes the PhD thesis by evaluating many modelling and inferential procedures for which conflicting claims have been made in the literature.

Chapter 7 outlines conclusions and identifies fruitful areas for further research.
As this is a wide canvas, occasionally a symbol is used for more than one purpose. The meaning will, however, be clear from the context.

References


CHAPTER 2

The background to security analysis and portfolio theory
2.1 Introduction

In finance, a portfolio is a collection of investments held by an institution or a private individual. In real life, a portfolio of assets is owned by the majority of people. Constituents of this portfolio may range from real assets (different kinds of property) to financial assets (such as stocks and bonds). Portfolio formation may be a result of a series of accidental and unrelated decisions or it may be a result of careful planning. In this chapter the fundamental principles underlying rational asset allocation are discussed. Confining our attention to financial assets we focus on investment in a stock market. With a thousand assets in a stock market, any investor faces a choice from among an enormous number of shares. The decision problem looks extremely difficult when one considers the number of possible assets and the various possible fractions in which each can be included into the final portfolio. Portfolio theory deals with the problem of constructing for a given collection of assets the investment with optimal characteristics. A general solution for the portfolio problem was introduced by Harry Markowitz with his paper “Portfolio selection” published in the 1952 *Journal of Finance*. In 1990 he shared a Noble Prize in Economics with M. Miller and W. Sharp for their pioneering work in the theory of financial economics\(^1\). The ideas proposed by the authors have gained such wide acceptance that thousands of papers have subsequently been written.

---

\(^1\) According to the press release from The Royal Swedish academy of science issued on 16 October 1990:

"Harry Markowitz is awarded the Prize for having developed the theory of portfolio choice; William Sharpe, for his contributions to the theory of price formation for financial assets, the so-called, *Capital Asset Pricing Model* (CAPM); and Merton Miller, for his fundamental contributions to the theory of corporate finance".
Before the work done by Markowitz saw the light of the day, in constructing their portfolios investors focused on assessing the risks and returns of individual assets. Based on the principle “Don’t put all your eggs into one basket”, investors knew intuitively that it was clever to diversify their portfolios. As early as 1920 Pigou wrote:

“if out of a hundred people, each of whom has £100 to invest, every one divides his investment among a hundred enterprises, the aggregate amount of uncertainty-bearing undertaken by the group is smaller than it would have been had every investor concentrated on a single enterprise” (Pigou, 1960, p. 778).

Markowitz formalised this intuition, turning it into the idea that investors hold portfolios of shares and therefore their focus should be upon portfolio risk and portfolio return, not on the return and risk of individual assets. He was the first to quantify risk and express quantitatively why and how portfolio diversification works to reduce risk and optimise return for investors. It does not however mean that the investors were irrational before Markowitz’ MPT (modern portfolio theory). As McGoun (1992, p. 164) commented: “while investors may always have been rational, their rationality may have been constrained by their knowledge and ability”. Markowitz accepted certain beliefs of the time regarding rational investment behaviour and developed a quantitative approach to portfolio selection. He believed that quantifying existing practices “would be more effective at maximising investors utility”. McGoun (1992, p. 174) actually raised doubts whether MPT deserves a Noble prize: “there simply were no “good” reasons for the
award”. It is not our aim to address the question of whether Markowitz and his successors deserve such a prize. The reasons for prize awarding appeared in the official press release from the Royal Swedish academy of science on 16 October 1990: “Markowitz's primary contribution consisted of developing a rigorously formulated, operational theory for portfolio selection under uncertainty - a theory which evolved into a foundation for further research in financial economics”.

Markowitz has also introduced the concept of an “efficient portfolio”. According to his definition, the portfolio which has the largest expected return for a given level of risk (or smallest risk for a given level of expected return) is called efficient.

James Tobin (1958) expanded Markowitz's mean-variance model by incorporating risk-free borrowing and lending into the analysis.

"Tobin showed that under certain conditions Markowitz's model implies that the process of investment choice can be broken down into two phases: first, the choice of the unique optimal combination of risky assets; and second, a separate choice concerning the allocation of funds between such a combination and a single riskless asset" (Sharpe, 1964, p. 426).

This result is known as Tobin's separation theorem.
In this chapter we begin by outlining the Markowitz mean-variance strategy (section 2.2) and discussing the limitations of the approach (section 2.3).

The subsequent two sections are fully dedicated to the implementation of portfolio theory. Section 2.4 discusses the amount and type of input data needed to solve a portfolio problem, while section 2.5 outlines a simplification of the computational procedure needed to calculate optimal portfolios. We examine the oldest and most widely used simplification of the portfolio structure: the single-index model. Here the idea is introduced that the problem of portfolio construction has a statistical nature.

The sixth and seventh sections deal with the problem of international diversification of the investment. The method proposed is a modification of a well-known one-factor model developed under assumption that international diversification of investment is allowed. Although, as we realized later, we reinvented the model proposed by Solnik in 1974 there is still an original contribution. The original work here is the derivation of the analytical solution for the Markowitz optimum investment strategy when both national and international factors are taken into account. We also look at the problem of exchange risk variation.

All the sections preceding section 2.8 have been concerned with how an individual or institution could select an optimum portfolio. As the behaviour of an individual investor can be modelled, it should not be a problem to determine how the aggregate of investors will behave. The purpose of building the equilibrium models consists in determination of
the appropriate measure of risk and the relationships between expected return and risk for any asset.

Section 2.8 deals with the simplest form of an equilibrium model, called the standard capital asset pricing model (CAPM), or one-factor capital asset pricing model, developed in parallel by Sharpe (1964) and Lintner (1965). The assumptions of the model imply that markets are efficient, although not explicitly. In particular the CAPM inherits basic assumptions made for portfolio theory in general: investors care only about mean and variance, they are risk averse and seek to maximise the expected utility at the end of the period. McGoun (1992, p. 157) identifies four interpretations of the CAPM: (1) the "normative" CAPM, (2) the "instrumental" CAPM, (3) the "positive" CAPM, and (4) the "useful" CAPM. He concludes that the CAPM asserts that "pricing according to the CAPM will maximise utility" or at least "will yield higher utility than some unspecified alternative" depending on the interpretation. CAPM implies market efficiency, but simply if markets are not efficient then there are strategies other than mean and variance to maximise returns.

The question of whether the stock market is efficient has attracted academics and practitioners for a long time. In spite of extensive research, there is still no agreement about this question in the academic world (e.g. Fama and French, 2004, Haugen and Baker, 1996). Section 2.9 addresses the problem of market efficiency.
Although the CAPM is widely advocated as the principal model for estimating the cost of capital for firms and evaluating the performance of managed portfolios, a number of empirical studies revealed that the CAPM has serious problems. As Fama and French (2004, p. 25) pointed out "The CAPM's empirical problems may reflect theoretical failings, the result of many simplifying assumptions". Section 2.10 is concerned with explaining the CAPM failure and introducing the idea of multifactor efficiency. In the next two sections the idea of multifactor efficiency and multifactor models is examined in more detail. They deal with the Fama and French (1993) three-factor model and with arbitrage pricing theory.

In the final section of this chapter we consider the regression-based approach to forecasting return. This approach is an indirect test of market efficiency in that any variation in risk adjusted returns is a possible indication of inefficiency. The expected return on a dot-com share will be higher than that on a property share only by virtue of high beta and not because the market is inefficient. It is pointed out that the method proposed does not correspond to the efficient market concept since abnormal returns generated by the model are certain returns, i.e. returns adjusted for risk. The claim that such an approach outperforms the market has negative implications for market efficiency. The analysis presented in this thesis is largely based on the regression method introduced in this section.
2.2 Implementation of Markowitz's mean-variance strategy

Investors and fund managers form their portfolios by spreading available funds between a number of different asset classes. One criterion is to maximize the expected utility of the portfolio of assets at the end of the investment period. As mentioned earlier, the classical mean-variance model, which involves maximizing the portfolio return and minimizing the risk, was proposed by Markowitz in 1952. Since this time, there has been continual interest in the problem, with increasingly sophisticated mathematical models being proposed. In the most basic setting the planning horizon is just a single period, and transaction costs are ignored. This allows some of the fundamental ideas to be discussed but substantially limits the applicability of the models.

As R. Baker (2004, p. 92) points out, "financial mathematics is an interesting and lucrative area of research. The problem of choosing a portfolio of shares to maximise profit and minimise risk has statistical and forecasting aspects. There we shall seek to minimise the variance on a weighted mean of forecasts. In the portfolio problem we seek to maximise the expected (forecast) return on a portfolio of shares and to minimise the risk of a poor return. Currently, there is not unfortunately agreement as to exactly how to define 'risk' in this context. Investors probably want not so much to minimise the variance of their return, but to have a very low probability of a 'bad' outcome".
Assume that an individual's total amount of money is $T$, of which $X$ is invested for the period $\Delta t$ in a stock market. Let the asset classes be indexed $\{1,..., N\}$ and let $x_i$ be the amount of money invested in the $i$-th of $N$ possible stocks. After period $\Delta t$, this sum of money has become $x_i(1+R_i)$, where $R_i$ is the return on the $i$-th stock. The amount not invested has become $(T-X)(1+R_F)$, where $R_F$ is the return on riskless asset.

A statistical decision theory approach to portfolio construction seeks to maximise the expected utility of an investment. An exponential utility function may be used to represent a decision maker's attitude towards risk. The general form of the exponential utility function is

$$U(X) = \frac{1-\exp(-\eta X)}{\eta},$$

where $U(X)$ is the utility of an amount $X$ of money, and $\eta > 0$ represents the degree of risk tolerance which determines the curvature of the utility function. As $\eta$ becomes larger, the utility function displays more risk aversion. This function has a unique property of exhibiting constant absolute risk aversion. The Pratt (1964)- Arrow (1965) measure of absolute risk aversion defined as $-U'(X)/U(X)$ is equal to $\eta$ and is constant with respect to $X$, so that the money one is prepared to risk does not depend on one's capital. Exponential utility allows investment decisions to be independent of the total capital owned by an investor. Consider two uncertain alternatives $X_1$ and $X_2$ with the base wealth given by $\omega$. Then alternative 1 (or investment strategy) is to be preferred to 2 if and only if $EU(\omega + X_1) > EU(\omega + X_2)$. Bell (1995) showed that if this preference holds
irrespective of wealth $\omega$ and the utility function is continuous then the choice is limited to linear or exponential utility functions. To justify the use of the latter the additional assumption is needed that an investor is risk averse. Only exponential utility has this property. In fact, in some situations using relative risk aversion would be more realistic. In the real world investor's decision depends on his or her wealth. Decreasing risk aversion is an intuitively sensible property as it implies that risk becomes less of a burden as $\omega$ increases.

In the light of the latter assumption linear plus exponential utility

$$U(x) = ax - be^{-cx}$$

where $a \geq 0$, $b$, $c > 0$ is recommended by academics (e.g. Bell (1995), Bell and Fishburn (2001)) since it has a number of desirable virtues. Certain properties such $U' > 0$ - more wealth is always preferred to less and that $U'' < 0$ - the risk is bad are always imposed. Bell, 1995a showed that if additionally we require the utility function to satisfy a contextual uncertainty condition (CUC) then only linear plus exponential utility remains viable. CUC conveys the intuitively sensible notion that it is more advantageous to resolve uncertainties in $\omega$ with greater spread rather then uncertainties with a smaller spread. More fundamentally, the equation above belongs to a small class of utility functions consistent with the one-switch condition introduced in Bell, 1988. According to the one-switch rule with an increment in wealth preference for one of two gambles can switch to the other but not back again. In addition linear plus exponential utility is decreasingly risk-averse i.e. $U''U'' < U'U'''$. 

23
In spite of the beauty of linear plus exponential utility, the functional form
\( U(X) = \frac{1 - \exp(-\eta X)}{\eta} \) is used more often. The reason why researchers use the
simple exponential function instead of the more attractive alternative lies in problems
associated with finding a suitable parametric form of the model and the impossibility of
measuring the organisation's wealth precisely.

Assuming a multivariate normal (MVN) distribution of returns \( r \sim MVN[\mu, \Sigma] \)
\[ E[\exp(-\eta r^T x)] = \exp(-\eta x^T \mu + \frac{1}{2} \eta^2 x^T \Sigma x), \quad (2.1) \]
where \( x = (x_1, ..., x_N)^T \) is a vector of proportions of the capital \( X \), invested in each
security.

Therefore, if we denote \( r_F = R_F e \), \( e = (1, ..., 1)^T \) then, taking into account expression
(2.1), the expected utility of the total capital \( T \) possessed after period \( \Delta t \) is
\[ U_{\Delta t} = \frac{1 - \exp(-\eta (T + R_F) + (\mu - r_F)^T x - (\eta / 2) x^T \Sigma x))}{\eta}, \quad (2.2) \]

The certainty equivalent \( D \) of some risky sum \( Y \) is the sure payment for which the
decision maker remains indifferent to the gamble, i.e. \( U(D) = E(U(Y)) \). For the

---

2 See Appendix 3 for derivation of the expression below
exponential utility function it is given by \( E(\exp(-\eta Y)) = \exp(-\eta D) \). Here the increase in \( D \) after \( \Delta t \) resulting from investing rather than saving at the risk-free rate is

\[
D = (\mu - r_f)^T x - (\eta / 2) x^T V x. \tag{2.3}
\]

Maximization of the certainty equivalent rather than expected utility will not make any difference to the optimal portfolio choice, because the latter is an increasing function of the former. Then the optimal strategy is one with the highest certainty equivalent.

Therefore \( D \) may be maximized with respect to \( x \), by setting \( \frac{\partial D}{\partial x_i} = 0 \), for \( i = 1, \ldots, N \). This gives us a system of equations

\[
\mu - R_f = \eta V x, \tag{2.4}
\]

from which \( x \) can be found.

Since the value of \( \eta \) for an individual investor is unknown, it can be imputed from the amount \( X \) that the person is willing to invest, irrespective of the value of \( T \). Then the individual need only decide how much to invest, and \( \eta \) is then determined. Thus from equation (2.4)

\[
\eta = \frac{e^T V^{-1}(\mu - R_f)}{X}.
\]

\(^3\) Note that the first component of this formula is simply a weighted average of the expected return on individual assets minus the risk-free interest rate. So the first item is the expected return on a portfolio \( R_p \) and the second item \( x^T V x \) is the variance of a portfolio \( \sigma_p^2 = E(R_p - E(R_p))^2 \).
The formula given above is only valid when holding a negative position in the security (short-selling) is not allowed. We now examine short selling in more detail. Many investors in the stock market are making a profit by selling securities they do not own. This investing technique is called short selling. It is possible to earn money on a decline in an individual stock by taking a negative position in the security. When an investor sells a security, a security is physically sold. Since the investor does not own it, he/she has to borrow this asset from another trader through an investment bank. As soon as the security is sold the company is not eligible to pay the dividends to the investor, whose share was borrowed. Now the dividends on the share contribute to the profit of the purchaser. Therefore, the individual who sold the security short has an obligation of supplying the funds to the individual whose share was borrowed, so that he/she receives any dividends paid on the stock.

At a future time the short seller buys the share back and simply replaces the share that was borrowed. If the price of the security has declined, he/she is able to repurchase it at a price lower than at which it was sold, so that he can make some profit. At the same time, if the price is higher then the opposite occurs. As Elton et al (2003, p. 26) pointed out, it is important to understand that “the capital gains and losses are equal in magnitude but opposite in sign for a short seller compared to a purchaser of the shares. Since the short seller pays dividends to the person whose shares were borrowed, and the capital gains or losses to a short seller are exactly opposite those of purchaser, the return to the short seller is minus the return of the purchaser”.
The reasons for using short sales as a trading strategy are

- **Speculation**, when a short seller tries to benefit from the decline in the value of the share.
- **Hedging**, when the short-seller wants to decrease the sensitivity of the portfolio to market movements. A portfolio with negative investments as well as with long purchases reduces the sensitivity to market fluctuations. This happens because the return on a long purchase is the opposite of the return on short sales.

However, when short selling is allowed, the net amount invested does not relate simply to an individual's risk aversion. The initial outlay can even be zero. Hence that in the situation when investments can be positive as well as negative it is better to fix \( \eta \) by specifying the acceptable standard deviation of monthly cash flow, \( (x^TVx)^{1/2} \).

### 2.3 Limitations of the Markowitz approach

In the previous section we discussed the basic logic of the Markovitz investment scheme as implemented via exponential utility functions (the analysis however is not dependent on exponential utility). The naïve scheme presented neglects a huge number of complications. The Markowitz approach has some severe limitations:
1. It is obvious that estimating expected return is crucial for investment decisions. Thus we have to use "sub models" to estimate the vector of expected returns $\mathbf{\mu} = (E(R_1),...,E(R_N))$, and the covariance matrix of returns $\mathbf{V} = (V_{ij})_{N \times N}$. With the single-index model (see Section 2.5 for an overview) we shall obtain the concrete formula for expected returns and for the covariance matrix. Even then, empirical Bayes adjustments could be made to the model parameters and the utility function need to be integrated over the distribution of model parameters. This problem is discussed in more detail in the next section.

2. There are two alternative definitions of return on an investment, and this has implications for the analysis. In evaluating the expectation of the utility function (see Appendix 3), we utilised the concept of the "arithmetic" return $R_A = (S_f - S_o) / S_o$, where $S_o$ and $S_f$ are initial and final share prices, and assumed that arithmetic returns are normally distributed. In fact, the continuously-compounded return $R_c = \log(1 + R_A)$ has a distribution much closer to normality. The models developed for estimating expected return, such as single-index model, use continuously-compounded returns to obtain estimates of model parameters. Therefore, since the arithmetic return is needed in equation (2.2), a back transformation $R_A = \exp(R_c) - 1$ should be carried out before the fractions of the optimal investment are computed.
3. What $R_F$ should be used? For institutional investors, the 90-day treasury rate, but for individuals, the decision is to take money out of a savings account, and then $R_F$ is the rate of interest. Anyway, this is of course assumed constant.

4. In its original form, the model proposed by Markowitz ignores transaction costs associated with buying and selling shares. However, it soon became apparent that the costs incurred when constructing a new portfolio or rebalancing an existing portfolio must be incorporated in any realistic analysis. Kamin (1975, p. 1263) argues that “the introduction of the transaction cost implies investor behaviour which is systematically different from that implied by the no-transaction-cost dynamic portfolio selection models”. Transaction costs and the need for discrete numbers of shares to be bought and sold will modify the elegant solution obtained in the previous section. On adding transaction costs $C$, which are a function of $x$, to expression (2.3) we have

$$ D = (\mu - R_F)^T x - (\eta / 2)x^T V x - C(x). $$

5. Does maximizing expected utility to the next time point where the portfolio will be balanced maximize expected utility over a long time period? Also, how about using non-exponential utility functions that allow absolute risk aversion to decrease with capital, rather then adjusting the risk aversion parameter?
6. Tax has been ignored. This means, for example, that the individual is indifferent to the form (dividends or capital gains) in which the return on the investment is received. To maximize utility after tax, add a term to the utility function.

This list contains the major problems associated with the Markowitz investment scheme, however it is not exhaustive. A number of limitations of the simplistic mean-variance framework have been discussed in the literature (e.g. Harald et al (2007)). A variety of reasons other than those taken into account by the mean-variance analysis and those mentioned above drive the actual decisions about the allocation of assets. Montezuma (2004), for example, highlights that the presence of liabilities in the decision process is not considered in the modern portfolio theory framework. This is not completely realistic since “institutional investors must also tailor their assets to meet liabilities”.

Our research is focused on the problem of getting the best predictions of returns and their covariance structure (problem 1). This is one of the most important aspects of the problem worth to be considered deeply in order to enable good investments.

2.4 The inputs to portfolio analysis

According to the Markowitz mean–variance approach, optimal portfolio construction is a decision problem of applied finance. In this analysis we see the construction of an
optimum mean-variance portfolio as an operational research problem. However, from equation (2.2) we see that we need estimates of the expected return on each security that is a candidate for inclusion in our portfolio, and we also need the estimates of the correlation between each possible pair of stocks under consideration.

The challenge is described as follows:

"It seems unlikely that analysts will be able to directly estimate correlation structures. Their ability to do so is severely limited by the nature of feasible organizational structures and the huge number of correlation coefficients that must be estimated. Recognition of this has motivated the search for the development of the models to describe and predict the correlation structure between securities" (Elton et al, 2003, p. 132).

Therefore implementing Markowitz's scheme requires a great deal of model choice and parameter estimation. Hence nearly all the nontrivial parts of the problem are statistical in nature.

The two main alternatives available for the purpose of estimation of the expected return are a single-factor model and multi-index models.
Sharpe (1964) and Lintner (1965) introduced the one-factor model that has been widely used to explain and estimate the correlation structure of security returns: Capital Asset Pricing Model (CAPM). At the heart of the CAPM is an old and common observation – the return on financial assets increase with risk. The main assumption behind this model is that stocks move together only because of a common co-movement with the market.

Multifactor models of security returns offer an alternative to the Sharpe (1964) and Lintner (1965) equilibrium-based CAPM in terms of measuring the expected returns on risky assets. Multi-index models attempt to capture some of nonmarket influences that cause securities to move together. These models are based on a number of explanatory factors influencing the returns on shares rather than on one solitary variable that is considered in CAPM. An important alternative to CAPM is the three factor model suggested by Fama and French (1993). In this model size and book to market factors are included, in addition to a market index, as explanatory variables. Another alternative that needs to be considered is Arbitrage Pricing Theory (APT) which also permits factors other than beta to explain share return.

Several subsequent sections of this chapter are devoted to the exploration of these models in greater detail.
2.5 An overview of single-index model

In this section we describe methods for selecting optimal portfolios that are appropriate when the single-index model is accepted as a description of the covariance structure between securities, and present the rules for optimal portfolio selection when the investor is allowed to short sell securities. The section takes notation from chapter 7 of Elton et al (2003).

Observation of the stock market reveals that when the market goes down most stocks decrease in price and when it goes up most stocks have a tendency to increase in price (the widely available stock market indices are usually used as proxies of market changes). This suggests that the influence on returns can be decomposed into a systematic factor (as measured by the return on the market index) and firm-specific factors. Common response to the market changes is one of the reasons stock returns are correlated, and the measure of this correlation can be found by relating the return on a security to the return on a stock market index.

One-factor model assumes that the excess return on any security is a linear function of the excess return on the domestic market index.

\[ R_i - R_f = a_i + \beta_i (R_m - R_f), \]

where

\( a_i \) is the component of security \( i \)'s return that is independent of the market's performance.
\( R_m \) is the return on the world market portfolio (return on the market).

\( \beta_i \) is a systematic risk of asset \( i \) relative to the world market portfolio (beta).

The expression given above decomposes the return on any security into two independent components, the part insensitive to the market changes and the part due to the market. Beta, the coefficient in a linear equation, measures the degree of sensitivity of the return on a particular stock to the market movements.

Now the world market portfolio, which consists of all assets in the world is not observable so it is necessary to use a proxy instead. As Arnold (2005, p. 341) pointed out, “ideally, when referring to the market portfolio, we should include all assets, ranging from gold through to bonds, property and shares. In practice to make the CAPM workable, we use a proxy for the market portfolio, usually a broadly based index of shares such as the FTSE Actuaries All-Share Index which contains about 700 shares”.

The component of return insensitive to the return on the market index is represented by the first term \( \alpha_i \). Breaking this term into two is practically helpful. Let us denote the expected value of \( \alpha_i \) by \( \alpha_i \) and let \( e_i \) stand for the uncertain component of \( \alpha_i \), so that \( E(e_i) = 0 \). Then

\[
R_i - R_F = \alpha_i + \beta_i (R_m - R_F) + e_i .
\] (2.5)
An estimate for $\beta_i$ is typically obtained by running the time-series regression: excess return on the asset $i$ versus the return on the index used as a proxy for the world market portfolio. Typical returns $R_{it} - R_{ft}$ for the firm (actually for HBOS shares) plotted against the adjusted market return $R_{mt} - R_{ft}$ for the period $t=1$ to $t=60$ with the fitted line from equation (2.5) are shown in Figure 2.1.

![Figure 2.1: A typical scatter plot: $R_{it} - R_{ft}$ versus $R_{mt} - R_{ft}$ with the fitted straight line from equation (2.5). The high scatter of points and the presence of outliers are typical.](image)

Bartholdy and Peare (2004) argued that in spite of the existence of a large academic literature which discusses the implementation of single-index model, in particular in relation to estimation of the key parameter beta, little research has been done in relation to how a best estimate should be obtained. There is a lack of agreement concerning what time frame, index, and data frequency provide the best estimate of this parameter. As a
result, different beta providers may suggest different estimates of beta for the same company. This may potentially lead to significantly different expected returns for an individual company and therefore to conflicting financial decisions.

The results obtained by Bartholdy and Peare (2004) suggest that five years of monthly data are appropriate time period and data frequency for the estimation of beta. However they also found that an equal-weighted index, as opposed to the commonly recommended value-weighted index, provides better estimates.

Turning back to equation (2.5), it is important to highlight that $e_i$ and $R_m$ are random variables. Therefore they each have a probability distribution and a mean and a standard deviation. Let us denote their standard deviations by $\sigma_i$ and $\sigma_m$, correspondingly.

It is convenient to assume that there is no correlation between residuals and return on the market index. Formally, this means that

$$\text{cov}(e_i, R_m) = E[(e_i)(R_m - E(R_m))] = 0.$$  

As mentioned earlier the parameter values $\alpha$, $\beta$, and $\sigma^2_i$ are typically estimated from time series-regression analysis. Regression analysis is one procedure that assumes that the noise term $e_i$ is uncorrelated with $R_m$, at least over the time interval to which the equation has been fit.
Notice that no simplifying assumptions have been made so far. The crucial simplification of the one-factor model is that for all values of \(i\) and \(j\), \(e_i\) is uncorrelated with \(e_j\) or, more formally, \(E(e_i e_j) = 0\). This suggests that the only explanation for the effect that the stock prices move in the same direction, systematically, is a common co-movement with the market.

Then simply the expected excess return on a security is

\[
E(R_i) - R_F = \mu_i - R_F = \alpha_i + \beta_i (E(R_m) - R_F),
\]

and the covariance of returns\(^4\) between securities \(i\) and \(j\) is

\[
V_{ij} = \beta_i \beta_j \sigma_m^2 + \delta_{ij} \sigma_i^2,
\]  \hspace{1cm} (2.6)

where

\[
\delta_{ij} = \begin{cases} 
0; i \neq j \\
1; i = j
\end{cases}
\]

As shown in Chapter 6 (Elton et al., 2003), if the investor wishes to assume a riskless lending and borrowing rate and short sales are allowed, then he or she can obtain an optimum portfolio by solving a system of simultaneous equations. The system of equations the investor solves is

\[
\mu_i - R_F = Z_i V_{ii} + \sum_{j=1}^{N} Z_{ij} V_{ij}, \quad i = 1, \ldots, N,
\]  \hspace{1cm} (2.7)

\(^4\) The derivation of this formula can be found in Elton and Gruber (2003), pp. 134-135
where \( Z_i \) is proportional to the amount invested in security \( i \).

The efficient set is determined as (see Elton et al. (2003), Appendix 3 of Chapter 9)

\[
Z_i = \frac{\beta_i}{\sigma_i^2} \left[ \frac{\mu_i - R_F}{\beta_i} - C^* \right], \quad i = 1, \ldots, N, \quad (2.8)
\]

where \( C^* \) is given by

\[
C^* = \frac{\sigma_m^2 \sum_{j=1}^{N} (\mu_j - R_F) \beta_j}{1 + \sigma_m^2 \sum_{j=1}^{N} \frac{\beta_j^2}{\sigma_j^2}}.
\]

Expression (2.8) can be used to determine the relative investment in each share. In order to ensure the full investment the weights on each security should be scaled so that they sum to one. Taking this into account, the proportion invested in each security can be computed as

\[
x_i = \frac{Z_i}{\sum_{i=1}^{N} Z_i}, \quad (2.9)
\]

2.6 International diversification of an investment

In this section, we deviate from the conventional analysis given by Elton et al. (2003) and propose a modification of the single-index model extending the idea of diversification to
an international setting. In the next section we address some of the many issues that arise when investing internationally.

Lintner (1965) highlighted that the return on the portfolio of shares is a composite product made up of the returns and risks of the market index on the one hand, and the independent returns and risks (the residual variances) on the other. Risk-averter investors "seek to minimize the risks associated with any given expected return". The gains from diversification come from (a) the fact that residual variances are not equal to zero and correlations with market index and other assets are consequently not perfect, and (b) the fact that some stocks are negatively correlated with other stocks and stock market indexes.

Since it is possible to reduce the risk of the investment by diversifying available funds within the boundaries of one country, it is reasonable to assume that further risk reduction is available by spreading investment internationally. Many researchers showed that the benefits of international investment are significant and that it is a sensible strategy for investors to use. As pointed out by Solnik (1974, p. 372) "An investor unaware of international investment opportunities will still base his decisions on sound measures of relative risk since the excess return he can expect should be proportional to the national systematic risk or beta. However, even if he gets a perfect diversification on his (national) investment he is still left with some unique country risk which he could have diversified away internationally". In fact researchers find that there is a 'home bias' (the tendency to invest in a large amount of domestic assets) towards much investment (e.g. French and Poterba, 1991). The possible gains from diversifying abroad greatly depend on the
correlation coefficient across markets, the risk of each market, and the returns in each
market. In other words, the gain from international diversification would be little if the
stock markets in different countries move together. Fortunately, markets are not perfectly
correlated.

Let us return to the simple single-index model and modify the solution, taking into
account, that international diversification is allowed.

For the purpose of mathematical convenience here and thereafter let us assume that the
risk-free rate (return on a riskless asset) has already been deducted from the values of
share and market index return. So that now \( R_i \) denotes excess return for the \( i \)th company
and \( R_m \) is the corresponding excess return on the market.

International diversification is implemented by using the relevant index for beta
calculations for each foreign stock. We consider the approach where the return on a share
is affected by the international factor through the national market index. The covariance
matrix \( V \) then changes in an obvious way because of the correlation between the LSE
index and the foreign index.

Modelling international stocks using a global beta, we have

\[
R_i = \alpha_i + \beta_i R_m + e_i
\]

as usual for a UK company, and
for a foreign company, where $R_f$ is the return on the foreign market index, e.g. the CAC, DAX, Dow-Jones, Nikkei etc. We can model

$$R_m = B_m R_g + \rho_m ,$$

$$R_f = B_f R_g + \rho_f ,$$

where $B_m, B_f$ are global betas and $R_g$ the return on the global market index (if only European countries were included, this would be the European index, and so on), $\rho_m$, $\rho_f$ are the components of local market return that are independent of the global market performance (the expected values of $\rho_m$, $\rho_f$ are zeros).

Hence, we can model return on a stock, using expression (2.10)

$$R_i = \alpha_i + \beta_i (B_m R_g + \rho_m) + e_i ,$$

(2.10)

e etc. We assume that $\rho_m, e_i$ are uncorrelated. Then simply the covariance between any two assets is$^5$

$$V_{ij} = \beta_i \beta_j B_m B_f \sigma^2_g + \beta_i \beta_j \sigma^2_m \delta_{mf} + \sigma^2_i \delta_{ij} ,$$

(2.11)

where $\sigma^2_g$ is the variance of returns to the global market $\sigma^2_g = E(R_g - E(R_g))^2$, $\sigma^2_m = E(\rho_m^2)$ and $\sigma^2_i = E(e_i^2)$.

$^5$ See Appendix 4 for derivation of this formula
This can be hierarchical, with sector betas, country betas and world betas. The benefit of such modeling would be that similar companies’ returns would be more highly correlated than others, so that a utility-maximizing or Markowitz portfolio selection procedure would minimize risk by picking stocks balanced between different sectors and countries.

For just one country, \( B_m = 1 \) and \( \rho_m = 0 \). In general we have \( N \) stocks in \( n \) countries.

The question is: can the Markowitz optimum investment strategy be solved analytically with hierarchical betas? The answer is: yes – the efficient set can be determined. To find the efficient set we must solve the system of simultaneous equations (2.7) again just as for the ordinary single-index model.

Substituting for covariance between any two securities from the previous formulae now yields the system of equations

\[
\mu_i = \sum_{j=1}^{N} \left( \delta_i \sigma_i^2 + \beta_i \beta_j \sigma_j m f + \beta_i \beta_j B_m B_j \sigma_j^2 \right) Z_j, \quad i = 1, \ldots, N. \tag{2.12}
\]

We can reexpress equation (2.12) in terms of

\[
D_m = \sum_{j \in m} \beta_j Z_j,
\]

and

\[
C = \sum_{m=1}^{n} B_m D_m, \tag{2.13}
\]

where \( j \in m \) means that the sum runs over all companies in country \( m \).
Then we get a more friendly looking system of equations (2.14)

\[ \mu_i = \sigma_i^2 Z_i + \beta_i \sigma_m^2 D_m + \sigma_g^2 B_m \beta_i C, \quad i = 1, \ldots, N \]  

(2.14)

and solving (2.14) for \( Z_i \) yields

\[ Z_i = \frac{\mu_i - \sigma_g^2 B_m \beta_i C - \beta_i \sigma_m^2 D_m}{\sigma_i^2}. \]  

(2.15)

To simplify further calculations let us introduce

\[ S_m = \sum_{j \in m} \frac{\mu_j \beta_j}{\sigma_j^2}, \]

\[ W_m = \sum_{j \in m} \frac{\beta_j^2}{\sigma_j^2}. \]

Then solving expression (2.15) for \( D_m \) gives

\[ D_m = \frac{S_m - \sigma_g^2 C B_m W_m}{1 + \sigma_m^2 W_m}, \]  

(2.16)

and on using equation (2.13) \( C \) can be found

\[ C = \frac{\sum_{m=1}^{n} \frac{S_m B_m}{1 + \sigma_m^2 W_m}}{1 + \sigma_g^2 \sum_{m=1}^{n} \frac{B_m^2 W_m}{1 + \sigma_m^2 W_m}}. \]

Hence \( D_m \) is found on using equation (2.16) and \( Z_i \) from equation (2.15). The percentage of investors' capital to be invested in each security can be computed by substituting equation (2.15) to (2.9).
Once the optimal fractions are determined, the variance of a portfolio of stocks can be calculated as

\[ \sigma_p^2 = x^T V x = \sum_{i=1}^{N} \sigma_i^2 x_i^2 + \sum_{m=1}^{n} \sigma_m^2 (\sum_{i \in m} \beta_i x_i)^2 + \sigma_g^2 (\sum_{m=1}^{n} B_m \sum_{i \in m} \beta_i x_i)^2. \]

In spite of the fact that risk can be reduced by spreading investments internationally, some risk remains even for the broadest portfolio. Furthermore, the correlation between national stock markets is increasing because of an increasing degree of economic integration around the world. The correlation between European and Japan equity market returns is still relatively low because their economies are not closely linked, while the correlation between the USA and Canada is quite strong.

Note, that the return on an international investment is affected not only by the return on a stock within its own market, but it is also affected by the change in the exchange rate between the securities own currency and the currency of the purchaser's country of residence. In the next section we focus on the problem of international diversification taking into account the risk associated with exchange rate variation.

2.7 International diversification, exchange rate risk is incorporated into consideration

Finally, exchange rate risk can be considered. Exchange rate risk is the risk that a business operation or an investment's value will be affected by changes in exchange rates.
For example, if money must be converted into a different currency to make a certain investment, changes in the value of the currency relative to the British pound will affect the total loss or gain on the investment when the money is converted back. Therefore the return on the international investment can be different from the return in the asset's own market. The variation can be substantial and is subject to the residence of the purchaser. So that changes in exchange rates are one of the main sources of risk in a foreign investment. This risk usually affects businesses, but it can also affect individual investors who make international investments.

Capital $S_0$ becomes $S_i$ after an investment period, giving a continuously-compounded return of $R = \log(S_i/S_0)$. If the investment is in a foreign currency, we have $R' = \log(S_iE_i/S_iE_0)$, where $E_0$ and $E_i$ are exchange rates. Hence, $R' = R + \log(E_i/E_0)$. If exchange rates follow the geometric random walk, $E(R') = E(R)$, and the only effect of exchange rate variation is to add to the equation for covariance a term $\nu_m^2 \delta_{mf}$, where $\nu_m^2$ is the exchange rate volatility (to pounds sterling) of country $m$, and is therefore zero for the UK. This also assumes that changes in exchange rates to sterling are uncorrelated.

Therefore now we have that covariance between any two securities can be written as

$$V_{ij} = \beta_i \beta_j B_i B_j \sigma_g^2 + \beta_i \beta_j \sigma_m^2 \delta_{mf} + \sigma_i^2 \delta_{ij} + \nu_m^2 \delta_{mf}.$$
The optimal amount of money to be invested in each security can be determined by substituting $V_{ij}$ from the equation above to formula (2.7). Then the system of simultaneous equations (2.7) is solved with respect to $Z_i$ (the amount the capital to be invested in security $i$). It needs to be said, that the computational procedure (to determine $Z_i$) is analogous to the one described in the previous section. For this reason the solution of the problem is derived in appendix 5. The result is

$$Z_i = \frac{\mu_i - \sigma^2 B_m \beta_i C - \beta_i \sigma^2 D_m - \nu^2 E_m}{\sigma_i^2},$$

where

$$D_m = \frac{S_m - B_m \sigma^2 W_m C - \nu^2 E_m G_m}{1 + \sigma^2 W_m},$$

$$E_m = \frac{1}{1 + \nu^2 f_m} \left[ N_m - B_m \sigma^2 C G_m - \frac{\sigma^2 G_m S_m + \sigma^2 G_m B_m \sigma^2 W_m C}{1 + \sigma^2 W_m} \right],$$

$$C = \frac{\sum_{m=1}^{n} \frac{B_m S_m}{1 + \sigma^2 W_m} - \sum_{m=1}^{n} \frac{\nu^2 G_m}{1 + \sigma^2 W_m} \times \frac{N_m + N_m \sigma^2 W_m - \sigma^2 G_m S_m}{(1 + \nu^2 f_m)(1 + \sigma^2 W_m) - \sigma^2 \nu^2 G_m^2}}{1 + \sigma^2 \sum_{m=1}^{n} \frac{B_m^2 W_m}{1 + \sigma^2 W_m} - \sum_{m=1}^{n} \frac{\nu^2 G_m^2 B_m \sigma^2}{1 + \sigma^2 W_m} \times \frac{1 + 2 \sigma^2 W_m}{(1 + \nu^2 f_m)(1 + \sigma^2 W_m) - \sigma^2 \nu^2 G_m^2}}.$$
\begin{align*}
N_m &= \sum_{j \in m} \frac{\mu_j}{\sigma_j^2}, \quad \text{and} \quad I_m = \sum_{j \in m} \frac{1}{\sigma_j}.
\end{align*}

These equations look complex, but they are relatively easy to solve by the substitution of parameters. The only aim of taking these cumbersome equations into consideration is to show that the solution for the problem can be found even in the case when exchange rate risk is considered together with international diversification. However the computational procedure is far more complicated.

The modification of the basic single-index model with international diversification can substantially improve portfolio investment decisions because international diversification reduces the risk (variance of portfolio). Local companies’ returns might be more highly correlated, and then if the investment is without international diversification it would increase the risk. From equation (2.11) we can see that if securities \( i \) and \( j \) are not from the same country \( m \) the second term in this formula will disappear. It means that if the portfolio selection process is used with international diversification then the risk would be minimized by picking stocks balanced between different countries.

Moreover if international diversification is allowed it means that the investment should be in a foreign currency, and in this case exchange rate volatility also affects the variance of portfolio, which is the risk of an investment. Thus further modification of the model,
when exchange rate risk is taken into consideration can considerably improve the applicability of the model.

2.8 The capital asset pricing model

The CAPM is an equilibrium relationship. By construction CAPM (or any equilibrium model) allows us to determine the relevant measure of risk for any asset and the relationship between expected return and risk for any asset when the market is in equilibrium. The central point of the model is that systematic risk is the only factor affecting the level of return required on a share for a completely diversified investor.

There are a number of simplifying assumptions behind CAPM:

- There are no transaction costs.
- Assets are infinitely divisible.
- There is no tax.
- No single investor can affect the price by an individual action
- Returns are distributed normally
- Unlimited short sales, lending and borrowing at the riskless rate are allowed.
- Investors have homogeneous beliefs regarding market conditions. They are assumed to have identical expectations with respect to necessary inputs to the
portfolio decision. The efficient frontier is the same for all investors and they all hold some proportion of the same market portfolio.

The main conclusion of CAPM is that the only portfolio of risky assets that any investor will own is the market portfolio. By definition, the market portfolio is a portfolio in which the proportion invested in any asset is equal to the market value of that asset divided by the market value of all risky assets. Each investor will hold a combination of the market portfolio and riskless asset in accordance to his or her risk preferences.

The equilibrium relationship for CAPM is given by

\[ E(R_i) - R_F = \beta_i \left( \frac{E(R_m) - R_F}{\sigma_m^2} \right) \text{cov}(R_i, R_m) \]

or

\[ E(R_i) - R_F = \beta_i (E(R_m) - R_F). \]

Here \( (E(R_m) - R_F)/\sigma_m^2 \) is described as the market price of risk and \( \text{cov}(R_i, R_m) \) as the measure of the risk associated with security \( i \). Notice that asset returns are proportional to the contribution of each asset to the portfolio (betas), which can be seen as a slope coefficient relating the return on asset \( i \) to the excess return on market portfolio. This relationship is often called the security market line.

The representation given above is nothing more or less than a description of single-index model presented in section 2.5. CAPM is the explanation of the expected returns that can
be derived when returns are generated by one-factor model meeting the assumptions defined at the beginning of this section. The contribution of CAPM is in demonstrating how one can go from a single-index model to a description of equilibrium.

The Sharpe-Lintner CAPM says that the expected value of an asset's excess return is completely explained by its expected CAPM risk premium (its beta times the expected value of the excess return on the market). This means that the intercept term $\alpha_i$ in the time-series regression equation (2.5) is zero for each individual asset. The implication here is that there is no opportunity of making abnormal return for investor. The absence of abnormal profit possibilities, in its turn, indicates the efficiency of the market (the next section of the chapter is fully devoted to in depth examination of the market efficiency concept).

### 2.9 Efficient market hypothesis

This section plays an important role in the thesis and aims to put our work in the appropriate academic context. Since market inefficiency was one of the implications emerged from our approach, it would be useful to consider the concept of market efficiency firstly.

In finance, the efficient market hypothesis (EMH), summarized by Eugene Fama in 1970, declares that all available information is already incorporated into the prices on traded
assets, e.g. stocks, bonds or property. Efficiency therefore asserts that the prices are accurate in the sense that they fully reflect the collective beliefs of all investors about future prospects.

Systematic overvaluing or undervaluing of assets does not occur in an efficient market – it is not possible to consistently outperform the market by using any information that is already available. As a result it is not feasible to build up trading rules which will “beat the market”. At the same time, an inefficient market allows us to exploit gainful trading opportunities, since it regularly prices shares imperfectly. Believers in the doctrine of market efficiency reject any trading strategy which a trader may believe he has discovered to choose winning shares. If the scheme really worked, they say, someone would have exploited it before. Any benefits from exploiting imperfections will be short lived.

EMH further suggests that the future flow of news is random and unknowable in the present (Samuelson, 1965). Also, if new information is revealed about a company, it will be assimilated into the share price rationally and very quickly so that none of the investors have an opportunity of making an abnormal profit. In other words, a trader can make a return on a share that is greater than the fair return for the risk associated with that share only by chance (or luck). Obviously people cannot have good luck all the time.

Moreover EMH dictates that security price changes follow a random walk. The prices of shares move in a random fashion, so that one can not forecast the current day’s price
change by looking at the previous day’s price movement. There is an explanation for this kind of behaviour. The price of the share at any one time reflects all existing information and it will only change if new information is obtained. Since the next item of information will be independent of the last item of information, price movements will be independent. Furthermore, it is not known whether the next piece of the news is going to be optimistic or pessimistic.

Efficiency for CAPM implies informational rather than allocative efficiency, that is the precision with which the market values the assets relative to their attributes, that is, how rapidly investors respond to fresh incoming news. If the relevant information is not processed quickly enough and delays in the investors reaction occur, then the share prices may stay away from their intrinsic or fundamental values and, therefore, the market is inefficient.

It needs to be pointed out that market efficiency does not mean that security prices are always equal to the true values. It only states that the errors made in pricing shares are unbiased.

Fama (1970) made a distinction between three forms of EMH, according to the nature of information which is incorporated into the share prices. Each form has its own implications for how markets work.
1. The **strong form** suggests that security prices reflect all available information, even private information. In a strong form efficient market no one (not even insiders) can earn excess returns. Seyhun (1986, 1998) provides evidence that insiders profit from trading on information not already incorporated into prices. There is no evidence that the strong form consistently holds though insider trading is seen as a constant threat by exchanges.

2. The **semi-strong form** of EMH asserts that security prices reflect all publicly available information. Analysing the information such as past price movements, earnings, dividend announcements, technological breakthrough and so on will not be able to consistently produce superior returns, because this information has already been reflected in the prices. Security prices adjust instantaneously when the new information is released. Semi-strong form efficiency implies that no fundamental analysis techniques will be able to reliably produce abnormal returns.

3. The **weak form** of the hypothesis suggests that no excess returns can be earned by using investment techniques based on analysing past price movements. There is no point in analysing historical share prices, because the market has already absorbed the information contained in past price movements into the present share price. Weak form efficiency implies that current share prices are the best, unbiased, estimate of the value of the security and no technical analysis techniques will be able to generate above-average returns by interpreting charts of the past history of share prices (except by chance). Hull (2000) emphasised that the weak form of market efficiency is
consistent with the assumption that the stock prices follow a Markov process. This form of efficiency, unlike the semi-strong form, does not however invalidate fundamental analysis.

An increasing number of empirical test findings have challenged the paradigm of market efficiency. Many observations from developed stock markets seem to be inconsistent with the EMH and indicate market inefficiency (Jegadeesh and Titman (1993), Lakonishok, Shleifer and Vishny (1994), Banz (1981), DeBondt and Thaler (1985)). The EMH became controversial especially after the detection of certain anomalies in the capital markets. In the course of the research in the area of market efficiency the following phenomena have been discovered (the effects are grouped according to the form of efficiency they reject):

- **Weak form**

  ➢ **Return reversal effect**

  De Bondt and Thaler (1985) found that assets that have given the lowest (highest) returns over the previous three to five years usually go on to outperform (underperform) the stock market over the subsequent three to five years. A possible explanation of this effect is that the market is overreacting to the bad news and undervalues the shares causing thereby significant weak form inefficiencies.
Momentum effect

Jagadeesh and Titman (1993) showed that stocks that outperformed (underperformed) the average stock returns in the past three to twelve months tended to continue to perform well (poorly) over the subsequent few months. The US-based research of Jagadeesh and Titman was followed up with studies examining the profitability of utilising the momentum investment strategy in stock markets around the globe. Rouwenhorst (1998) investigated the momentum phenomenon in twelve developed country stock markets. His study has demonstrated that the effect is not only a feature of the US market. Two main hypotheses have been submitted to explain the effect. The first asserts that investors underreact to new incoming information. If the announcement is made the market could take some time to absorb new information into the share price. This means that the share price does not approach a new efficient level instantly because of investor's under reaction. The second hypothesis assumes that investors are in fact overreacting during the test period. If the share price steadily goes up then after a series of months investors still keep on buying this share expecting the price to rise even higher. That is pushing the share prices of winners above the efficient level. At the same time, selling off the losers, which is not always reasonable, pushing their prices to irrationally low level.
• Semi-strong form

➤ Seasonal and calendar effects

A number of studies found cyclical anomalies in share price movements. One of the anomalies is the so-called January effect. It was documented that shares tend to give higher returns in January as compared to other months. Also the research in this area identified the weekend effect, it refers to the fact that there appears to be a systematic drop in the daily rate of return on stocks between the Friday closing and Monday opening. A trading strategy, which would be profitable in this case, would be to buy stocks on Monday and sell them on Friday.

➤ Small-firm effect

The 'small firm effect', which is also known as the 'size effect', is a theory which states that companies with smaller market capitalisation (i.e. smaller firms) outperform larger companies. It is believed that smaller companies have more growth opportunities than larger companies. However, a strategy based on these findings would not necessarily be more profitable than buying and holding a well-diversified portfolio. Usually it is more costly to trade in small companies' shares: if transaction costs are included, the net return of trading in small-cap firm assets decreases.
Empirical work, started in the late 1970s, revealed that financial ratios which measure stock prices relative to fundamentals are good predictors of future returns. The three most commonly used financial ratios are \( D/Y \) (dividend yield), \( B/M \) (book-to-market ratio) and \( E/P \) (earning-price ratio). Basu (1977) provided the evidence which appears to confront with the semi-strong form of EMH. He claimed that, when common stocks are sorted on price earning-price ratios, future returns on high \( E/P \) stocks are higher than predicted by CAPM. Many studies (e.g. Stattman, 1980) have concluded that stocks with high book-to-market ratios have high average returns that are not captured by their betas. In other words, a share price which is low relative to the balance sheet assets is an attribute of an undervalued share. Finally, it was documented that shares offering higher dividend yield tend to give higher returns. More recently Lewellen (2004) focused primarily on \( D/Y \), he tested the predictability of aggregate stock returns with financial ratios. The paper utilises the following regression model

\[
R_t = \alpha + \beta x_{t-1} + \epsilon_t,
\]

where \( R_t \) is the return in month \( t \) and \( x_{t-1} \) is the dividend yield (or other financial ratio) known at the beginning of the month.

\( D/Y \) is assumed to follow a stationary AR1 process:

\[
x_t = \phi + \rho x_{t-1} + \mu_t,
\]

where \( \rho < 1 \).
Basing his argument on the empirical study Lewellen (2004) provides strong evidence of stock return predictability with the dividend yield. He also concludes that although B/M and E/P have some forecasting ability, the evidence is less reliable than for DY.

Robertson and Wright (2006, p. 91) highlight that nowadays "an increasing body of research has cast doubt on the earlier evidence of predictability, attributing it to data mining or other statistical problems". The authors see the weaknesses of the nonfinancial dividend yield (ratio of dividend per share to price) as a predictor in mismeasurement and suggest using a new cash-flow yield as a predictor. The distinguishing feature of cash flow yield is that it includes not only dividend but also nondividend cash flows to the shareholder.

- **Strong form**

For the strong form efficiency to hold there must not be consistent evidence of abnormal price reaction before the announcement of information to the market (e.g. interim results). The idea of a strong-form efficient market seems impracticable when the topic of trading on inside knowledge is introduced. It is well known that it is possible to trade shares on the basis of information which is not publicity available and thus make abnormal profits. Trading on inside information is considered as a criminal offence in many countries. In US, for example, the Securities Exchange Act of 1934 substantiates the regulation of insider trading on the ground that acquiring the information intended to be available
solely for the internal use causes unfairness to outside investors. Such an activity also leads to the material losses for company stockholders.

"The insiders have acquired the information at the expense of the enterprise, and for the purpose of conducting the business for the collective good of all stockholders, entirely apart from personal benefits from trading in its securities. There is no reason for them to be entitled to trade for their own benefit on the basis of such information, particularly when, as we have noted, permission to do so will enable, if not indeed tempt, them to disserve the corporation and all its stockholders" (Brudney, 1979, p. 344).

Acts of this type, however, can not guarantee complete absence of this phenomenon in real world since trading on inside information is very lucrative.

2.10 Failure of CAPM: irrational pricing or risk

Summarising, Sharpe (1964) and Lintner (1965) introduced the first asset-pricing models based on the Markowitz portfolio theory and EMH assumptions. Their work resulted in the CAPM. Many authors (e.g. Fama and French, 2004) highlighted the advantages of the model.
"The attraction of CAPM is that it offers powerful and intuitively pleasing predictions about how to measure risk and the relation between expected return and risk" (Fama and French, 2004, p. 25).

Unfortunately, the model does not appear to adequately describe the variation in stock returns. CAPM assumes that the difference in market beta is the only explanation for the differences in expected return across the securities. The assumption implies that other variables such as, for example, the company’s fundamentals add nothing to the description of expected return. The CAPM’s main argument is that a security’s return has nothing to do with company and industry specific events for these sources of risk are simply immaterial as they are easily diversified away by traders. Empirical studies show that the CAPM has potentially fatal problems since low beta stocks may offer higher returns than the model would predict. Fama and French (1992) provide evidence on the empirical failure of the CAPM. They based their study on using the cross-section regression approach. It was confirmed that size, earning-price and book-to-market ratios add to the explanation of expected return provided by market beta.

The crash of CAPM has fuelled an ongoing dispute over the correct theory of asset pricing. While searching for new models researchers often turn to the explanations of the failure. Two explanations of this phenomenon are debated in the literature.

The first theory is supported by behavioralists, De Bondt and Thaler (1987) and Lakonishok et al. (1994) are among them. The proponents of behavioural finance argue
that the prices of shares can be pushed away from their fundamental values (for considerable periods of time) because of the systematic errors frequently made by investors. As Fama and French (2004) highlighted, from the behavioralist’s point of view sorting companies on book-to-market ratios indicates investor overreaction to good or bad times. Investors overextrapolate past performance. This results in stock prices that are too high for growth firms (with low B/M ratio) and too low for distressed companies (with high B/M ratio). The outcome of this effect is high return for value stocks and low return for growth stocks once the pricing anomaly is corrected. This has been a strong attack on EMH since EMH assumes that all investors are rational, or, even if there are some irrational investors in a stock market, then the rational investors will eliminate mispricing through arbitrage.

However, the problem that often clouds such evidence is that the estimated abnormal returns could be nothing more than a premium for some risk that the researcher failed to identify and measure accurately. If so, such evidence does not indicate a violation of market efficiency. It rather indicates imperfection in our ability to quantify the risks that affect asset pricing.

CAPM explains why the expected returns on some shares are better then on others. The shares have particular attributes, but that does not mean that the market is inefficient. In fact, efficient markets can exist without the CAPM as such. To indicate market inefficiency one should provide evidence that average or expected risk adjusted returns
are higher for identifiable shares (that would create an opportunity to earn an abnormal return).

The efficient market protagonists advocate EMH by saying that higher returns usually come with higher risks. For example, a reasonable explanation for the outperformance of small firms was suggested: it might be that the researchers had not adequately allowed for the additional risk of small firms – in particular the risk arising from lower liquidity. Believers in the doctrine of market efficiency also found the explanation for the low price-earning anomaly. They build up their line of defence by saying that the supposed outperformers are more risky than the average share, thus the efficient market should allow them to deliver superior returns. Lakonishok, Shleifer and Vishny (1994) investigated this and found that low price-earnings securities are in fact less risky than the average. In spite of the results of Lakonishok’s examination, the alternative theory highlights that all we need is a more sophisticated (than CAPM) asset pricing model. If so, market beta cannot be interpreted as an exhaustive description of the risk associated with the security, and one should not be surprised to discover that differences in beta can not fully capture the differences in expected return. According to this view, the findings should turn to asset pricing models that do a better job explaining average returns. This leads us to the idea of multifactor efficiency which implies a relation between expected return and beta risks, but to explain expected return it requires additional betas, along with the market beta.
2.11 Fama and French three factor model

The three factor model developed by Fama and French (1993, 1996) provides an alternative to CAPM for estimation of expected return. In the model two extra factors are added to explain excess return: size and book-to-market ratio. It is attempting to pick up systematic risk factors not captured by the simple CAPM.

Fama and French (2004, p. 38) draw attention to the fact that “though size and book-to-market equity are not themselves state variables, the higher average returns on small stocks and high book-to-market stocks reflect unidentified state variables that produce undiversifiable risks (covariances) in returns that are not captured by the market return and are priced separately from market betas.” To support this idea, the authors show that returns on high book-to-market stocks covary more with one another than with returns on low book-to-market stocks, and returns on the small-cap firm stocks covary more with one another than with returns on the stocks of large firms. Rather than explaining abnormal return by irrational underpricing of small companies and those with high book-to-market ratio Fama and French interpret these phenomena as additional risk factors that require compensation in the form of higher returns.

Thus for stock, \( i \), to estimate expected return, beta estimates for each of the factors are obtained from the time-series regression

\[
R_i - R_F = \beta_{sf}(R_m - R_F) + \beta_{s}(SMB) + \beta_{b}(HML).
\]
In this equation, $SMB$ (small minus big) is the return on the portfolio of small stocks minus the return on the portfolio of large stocks and $HML$ (high minus low) is the difference between the returns on diversified portfolios of high and low book-to-market stocks.

According to the three-factor model, higher returns are regarded as a compensation for taking on higher risk. In particular that means that if returns increase with book/price, than stocks with a high book/price ratio must be more risky than average. At the same time, traditional financial analysts would give exactly the opposite advice. This contradiction comes from the fact that the latter approach (the one which business analysts usually utilise) rejects the efficient market hypothesis, while three-factor model does not. For someone who does not believe in EMH high book/price ratio is a sign of a good buying opportunity, because the stock looks cheap. EMH offers a different explanation: low-cost assets can only be cheap for a good reason, namely that investors believe they are more risky.

Before we go on to the next section it needs to be pointed out that the Fama and French three factor model is nothing more than a special case of arbitrage pricing theory (APT) developed primarily by Ross (1976). In the next section of this chapter we give a brief overview of APT.
2.12 Arbitrage pricing theory

Multifactor models of share returns such as APT offer an alternative to the Sharpe (1964) and Lintner (1965) equilibrium-based CAPM in terms of measuring the expected returns on risky assets. In fact, the explanation of equilibrium offered by APT is more general than that provided by CAPM. APT was developed as a one-period model in which investors believe that the stochastic properties of returns are consistent with the factor structure. Ross (1976) points out that the expected returns on the assets are approximately linearly related to the factor loadings in the case when equilibrium prices offer no arbitrage opportunities over static portfolios of assets.

The returns on a share under the APT are found through the following equation:

\[ R_i - R_F = \beta_{i1}(R_1 - R_F) + \beta_{i2}(R_2 - R_F) + ... + \beta_{in}(R_n - R_F) + \epsilon_i, \]

where \( \beta_j, j = 1, ..., n \) is the sensitivity of stock \( i \)'s return to the \( j \)th index. The terms in brackets are the risk premiums for each of the factors in the model and \( \epsilon_i \) is a random error term with mean equal to zero and variance equal to \( \sigma_i^2 \). We see that each share has a different degree of sensitivity to each of the risk factors, investors will only accept the extra risk if they are rewarded with higher return.

The main problem with APT is that, unlike the three factor model, it does not identify the systematic risk factors, which should be included in the model. Nowadays there is a lack of agreement about the key variables since the identified factors vary from study to study.
Many researchers have struggled to recognize the most commonly encountered systematic risk factors. As a result, a number of research papers have demonstrated these to be changes in the macroeconomic variables such as interest rates, industrial production levels, inflation etc (Arnold, 2005). It is sensible to assume that future earnings are greatly influenced by the economic situation. Most firms' profits will rise if the economy as a whole is growing and decrease if the economy is distressed. Interfacing with the outer world, all the companies are affected by the variations in the macroeconomic environment. However some companies are more sensitive to these variations than others – this is measured by the relevant betas.

2.13 Regression-based approach to forecasting return

As mentioned earlier, the efficient market hypothesis states that it is impossible for an investor to systematically outperform the stock market and consistently earn risk adjusted profits above the risk free rate. If so, an investor should rather aim to track the market as accurately as possible (passive investment). There is also the more pragmatic tradition of empirical finance, which does not rely heavily on theory, but rather seeks investment strategies that work in practice. Following this tradition, we develop and assess a regression model of stock returns. Our aim is to forecast returns as well as possible, and also their joint distribution, summarised by the covariance matrix $V$. That is, we forecast
those quantities required by an active investor. This is already a challenging statistical problem.

Although predicting return on stocks is crucial for investors, and although several determinants of high return have been identified, there is comparatively little in the financial literature about regression-based approaches to forecasting return. This lack of research may be because many workers in finance prefer to demonstrate the value of some predictor by constructing portfolios from deciles of firms ranked by it. There is however a regression-based strand. Thus Kandel and Stambaugh (1996) model returns on stocks using a regression on predictor variables. They find that even poor predictions can drastically influence investment decisions.

The simple regression model

$$R_i - R_F = \alpha_i + \beta_i (R_m - R_F) + \delta^T x_i + \varepsilon_i$$

(2.17)

for returns of the $i$th company can be used, where $x_i$ is a $p$-fold vector of predictors (such as dividend yield and market capitalization) and $\varepsilon_i \sim N[0, \sigma_i^2]$. Since this is a one-factor model the covariance matrix can be estimated with formula (2.6).
References


CHAPTER 3

Transforming predictor variables to near normality
3.1 Introduction

Although the regression model presented in section 2.13 assumes a linear relationship between excess return and predictive variables, in general, the predictor variables (elements of vector \( x \)) can take various forms, and a standard regression methodology is still appropriate. As there is no theory to dictate the functional form of the model, transforming predictor variables to approximate normality seems to be a sensible idea to improve modelling of the return. This chapter is concerned with this problem and includes the sections covering the theory of the Box-Cox (Box and Cox, 1964) and Johnson's arcsinh (Johnson and Kotz, 1970) transformations (sections 3.2), structure and algorithm of a Fortran 95 program which produces the optimal transformation parameters using the data from 'lsedata.dat' (section 3.3). Section 3.4 describes the results obtained on using Johnson's transformation (for all predictor variables) and Box-Cox transformation (applied to positive variables only). It also provides the comparison between these two transformations in terms of their ability to reduce the skewness and kurtosis of the distribution of original variable.

3.2 Transforming to near normality

Many studies (e.g. Cook and Weisberg, 1999) highlight the importance of utilising transformations in the regression analysis. A great number of them are concerned with developing the methodology for achieving approximate normality. The principal aim of these studies is to transform the random variable \( X \) so that the resulting distribution is
approximately normal. To achieve this, the parametric family of transformations is used with high emphasis placed on estimating the parameters.

The valuable technique for trying to normalise the data set was suggested by Box and Cox in 1964. Their work has been the starting point of many investigations in this area. As Yeo and Johnson (2000, p. 954) pointed out “a major step towards an objective way of determining a transformation was made by Box & Cox”. This power family is used most often and defined by

$$y^{bc}(\omega, x) = \begin{cases} (x^{\omega} - 1)/\omega, & \text{if } \omega \neq 0 \\ \log(x), & \text{if } \omega = 0 \end{cases}$$

where $x$ is a list of $n$ strictly positive numbers and superscript $BC$ is used to denote Box-Cox transformation.

The Box-Cox family is consistent with the family of power transformation, so the meaning of the parameter $\omega$ is straightforward to understand. That makes this family extremely useful since it includes the important special cases of untransformed, logarithmic, inverse, and square and cube root. The Box-Cox family can be used for many purposes, in particular for choosing the transformations for the response or dependent variable and for transforming a set of predictive variables toward multivariate normality.

6 taking the square root or logarithm of a data set is extremely useful for the data that show moderate right skewness.
However, it has a major drawback since it is not valid for negative $x$. Several attempts to define transformation families variables $x$ that include negative values have been made. One possibility is to consider transformations of the form (3.1), where a shift parameter $\lambda$, which is sufficiently large to ensure that $(x + \lambda)$ is strictly positive, is introduced to handle situations where the response is negative but bounded below.

$$y^{bc}(\omega, \lambda, x) = \begin{cases} \frac{(x + \lambda)^{\omega} - 1}{\omega}, & \text{if } \omega \neq 0, \\ \log(x + \lambda), & \text{if } \omega = 0. \end{cases}$$

(3.1)

Atkinson (1985, pp. 195-199) nevertheless argued that since the range of the distribution is determined by an unknown shift factor the standard asymptotic results of maximum likelihood theory may not apply.

In theory, the estimates of $(\omega, \lambda)$ can be obtained simultaneously, although in reality estimates of $\lambda$ are subject to high variation. Alternatively, other families of transformations such as the folded power family (see Cook and Weisberg, 1999, p. 330) have been suggested. However, since the transformed distributions have poor properties this family is not often used.

Johnson and Kotz (1970) have proposed an $\text{arcsinh}$ transformation to induce normality that can be used without any restrictions on $x$. This two-parameter transformation is defined by

$$y^J(a, c, x) = \sinh^{-1} \left\{ c(x - a) \right\} / c,$$

(3.2)

where the superscript $J$ is used to denote Johnson's transformation.
The remainder of this section is concerned with the problem of transforming a random sample from a parent distribution to approximate normality. The procedure of estimating transformation parameters is described below. We firstly consider Johnson's transformation which allows negative as well as positive variable values.

Johnson and Kotz (1970) obtain parameter estimates via calculating the moments of the distributions of the untransformed variables in terms of the transformation parameters. However, the method of moments, as applied to this problem, has a serious drawback. Namely, since the distributions are long-tailed the variance of the distributions is subject to a large sample error. As a result, the parameter estimates might not be accurate. To circumvent this problem, we use the maximum likelihood approach.

Let \( X_1, \ldots, X_n \) be a sample from an untransformed distribution with the probability density function \( f(\cdot) \). Denote the transformed values by \( y(a,c,X_1), \ldots, y(a,c,X_n) \). Parameters \( a \) and \( c \) are unknown and need to be estimated from data by maximising the log-likelihood function (3.3) with respect to them.

\[
\ell(\theta|x) = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(y(a,c,x_i) - \mu)^2}{2\sigma^2} + \log|J(a,c)|. \tag{3.3}
\]

Here \( \theta = (a, c, \mu, \sigma^2)^T \), \( x = (x_1, \ldots, x_n)^T \) and \( J(a,c) \) - the Jacobian of the transformation is defined as a matrix of first derivatives of \( y \) by \( x \),

\[
J(a,c) = \left[ \frac{\partial y(a,c,x_i)}{\partial x_j} \right]_{ij}.
\]
Notice that the likelihood (3.3) is derived under the assumption that, for some values of $a$ and $c$, the transformed observations follow a normal distribution with some arbitrary mean $\mu$ and variance $\sigma^2$.

Recall that a two-parameter transformation is given by

$$y' = y' (a, c, x_i) = \sinh^{-1}\left\{c(x_i - a)\right\}/c.$$  \hspace{1cm} (3.4)

Let us denote $y_i = y' (a, c, x_i)$. Then from (3.4),

$$c(x_i - a) = \sinh(cy_i).$$

By definition $\sinh(cy_i) = (e^{cy_i} - e^{-cy_i})/2$, hence

$$c(x_i - a) = \frac{e^{cy_i} - e^{-cy_i}}{2}.$$  \hspace{1cm} (3.5)

By multiplying both sides of the equation above by $2e^{cy_i}$ and rearranging the terms we obtain a quadratic equation (3.5)

$$e^{2cy_i} - 2c(x - a)e^{cy_i} - 1 = 0.$$  \hspace{1cm} (3.5)

If we denote $t = e^{cy_i}, t > 0$, the equation (3.5) can be written as

$$t^2 - 2c(x - a)t - 1 = 0.$$  \hspace{1cm} (3.6)

Solving (3.6) for $t$ yields

$$t = c(x_i - a) \pm \sqrt{c^2(x_i - a)^2 + 1},$$

therefore
\[
\gamma_i = \frac{\log\left\{c(x_i - a) + (c^2(x_i - a)^2 + 1)^{1/2}\right\}}{c}.
\]

Coming now to the computation of the Jacobian for the model we note that

\[
\frac{\partial y_i}{\partial x_i} = \left(\frac{\partial x_i}{\partial y_i}\right)^{-1}, \quad i = 1, \ldots, n.
\]

Expressing \(x_i\) in terms of \(y_i\) from the equation (3.4) and differentiating it with respect to \(y_i\) yields

\[
\frac{\partial x_i}{\partial y_i} = \cosh(cy_i).
\]

Therefore the Jacobian is given by

\[
\frac{\partial y_i}{\partial x_i} = \frac{1}{\cosh(cy_i)}, \quad i = 1, \ldots, n
\]

(3.7)

Recall that by definition \(\cosh^2(cy_i) = 1 + \sinh^2(cy_i)\), thus the equation above is equivalent to

\[
\frac{\partial y_i}{\partial x_i} = (1 + c^2(x_i - a)^2)^{-1/2}.
\]

(3.8)

Hence on using (3.8) equation (3.3) can be written as

\[
\ell(\theta|x) = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma^2} + \sum_{i=1}^{n} \log\left\{\frac{1}{(1 + c^2(x_i - a)^2)^{1/2}}\right\},
\]

Holding \(a\) and \(c\) fixed, we initially maximize \(l(a,c,|x)\), yielding

\[
\hat{\mu}(a,c) = \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y(a,c,x_i),
\]

\[
\hat{\sigma}^2(a,c) = \frac{1}{n} \sum_{i=1}^{n} \{y(a,c,x_i) - \hat{\mu}(a,c)\}^2.
\]
The maximum likelihood estimates, \( \hat{a} \) and \( \hat{c} \), of \( a \) and \( c \) are obtained by maximizing the profile likelihood function (3.9) and then \( \hat{\theta} = (\hat{a}, \hat{c}, \hat{\mu}(\hat{a}, \hat{c}), \hat{\sigma}^2(\hat{a}, \hat{c}))^T \).

\[
\ell(\theta|x) = -\frac{n}{2} \log \left( 2\pi \frac{\sum_{i=1}^{n} (y_i - \hat{\mu})^2}{n} \right) - \frac{n}{2} + \frac{1}{2} \sum_{i=1}^{n} \log \left( \frac{1}{(1 + c^2(x_i - a)^2)^{1/2}} \right). \tag{3.9}
\]

For variables that can take only positive and zero values it is reasonable to utilize the Box-Cox transformation (3.1). By analogy with the computations given above (for Johnson's method) the profile likelihood function for Box-Cox transformation can be written as

\[
I(\theta_x) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \left( \frac{\sum_{i=1}^{n} (y^{\text{BC}}(\omega, \lambda, x_i) - \hat{\mu}_p)^2}{n} \right) - \frac{n}{2} + (\omega - 1) \sum_{i=1}^{n} \log(x_i + \lambda), \tag{3.10}
\]

where \( \mu_p, \sigma_p^2 \) - are the transformed distribution mean and variance respectively and \( \theta_p = (\omega, \lambda, \mu_p, \sigma_p^2)^T \).

The optimal values of transformation parameters \( \hat{\omega} \) and \( \hat{\lambda} \) are obtained by maximising the likelihood (3.10) with respect to \( \omega \) and \( \lambda \).

### 3.3 Programming

A Fortran 95 program has been written to compute optimal transformation parameters. It reads the data from the source file 'lsedata.dat' variable by variable and calculates
optimal transformations for each variable. Some of the predictors can only take positive and zero values while the others can also take negative values. For this reason only Johnson's method is utilised by the program in the latter case. However, in the former case both Box-Cox and Johnson's arcsinh transformations are feasible.

In both cases optimal transformations were found by varying the parameters over the grid of values to maximise the profile likelihood function (equations 3.9 and 3.10 for Johnson's and Box-Cox transformations respectively). For Johnson's arcsinh transformation the scale of variation for \( a \) and \( c \) is given by the standard deviation of \( x \) because that is the only sensible measure of the scale of \( x \) we have. Parameter \( c \) must be in inverse units to \( x \), so we took \( c \) to go from 0 to \( 5/\sigma_x \), \( a \) must be in the same units as \( x \), thus we took that to go from \(-2\sigma_x\) to \(2\sigma_x\).

For the Box-Cox transformation the scale of \( \lambda \) is also geared to the standard deviation of \( x \), while \( \omega \) varies between -1 and 1. When the absolute value of \( \omega \) is small it would be troublesome for the computations to employ formula (3.1) directly. Therefore the following approximation is used instead

\[
\frac{(x + \lambda)^\omega - 1}{\omega} \approx \log(x + \lambda) + (1/2)\omega^2 \log^2(x + \lambda) + ...
\]

All transformed variables were finally given an affine transformation to zero mean and unit standard deviation.
The computational time can be greatly reduced by using the recursive equation (3.11) to calculate the variance.

\[ \sigma_{n+1}^2 = \frac{n}{n+1} \left( \sigma_n^2 + \frac{\bar{x}_n - x_{n+1}}{n+1} \right), \quad (3.11) \]

where \( \sigma_n^2 \), \( \sigma_{n+1}^2 \) are the variances of populations of sizes \( n \) and \( n+1 \) respectively, and \( \bar{x}_n \) is the mean of the population of size \( n \). The derivation of this equation is given in Appendix 6.

3.4 Analysis of the results

As mentioned before we are interested in transforming the random variable \( X \) so that the transformed distribution is approximately normal. Having provided a brief introduction to the theory of transformations, we close this chapter by outlining the results obtained on using real data.

1. Consider a first group of predictors: "monthly return", "2-monthly return", "3-monthly return", "4-monthly return", "6-monthly return", "annual return", "2-years return", "3-years return", "4-years return", "5-years return".
Figure 3.1. An untransformed predictor variable: 2-monthly return December 1994-November 2004 with superimposed normal curve

The variables listed above are allocated to a separate group since they have got an intrinsically similar nature and their resulting distributions (after applying the transformation) look quite similar. The value of return, calculated over the different periods, can be negative, hence Johnson's transformation is needed.

Without loss of generality, here we confine ourselves to the examination of 2-monthly return. Figure 3.1 shows an untransformed variable with a skewness of 2.1 and a kurtosis\(^7\) of 67.1. Compared to the normal, the histogram has a stronger peak near the

\(^7\) Some older works define kurtosis as \(\frac{\mu_4}{\sigma^4}\), where \(\mu_4\) is the forth moment about the mean and \(\sigma\) is a standard deviation. Nowadays a different definition of the kurtosis is more frequently used. It is defined as \((\mu_4/\sigma^4) - 3\) and is known as excess kurtosis. The "minus 3" in this formula can be explained as an adjustment to make the kurtosis of the normal distribution equal to zero (since the kurtosis for the standard normal distribution is three). The letter definition is the one which is used to calculate kurtosis in Minitab ver. 13 and many other statistical packages. It also has been utilized in this work. Here and thereafter we refer to excess kurtosis simply as kurtosis.
mean, more rapid decay and heavier tails. This is indicated by the high value of the kurtosis. The histogram for transformed data set is shown on Figure 3.2. It has a more rounded peak with wider "shoulders". The skewness is now 0.07 and the kurtosis is 0.14.

3 Figure 3.2. A transformed predictor variable: 2-monthly return December 1994-November 2004 after Johnson's transformation with superimposed normal curve

It can be seen that after applying Johnson's arcsinh transformation together with an affine transformation the distribution of transformed variable looks approximately standard normal, however it still has a spike at zero, probably caused by rounding down small returns to zero. The optimal values of parameters are:
\( a \approx 0.03, c \approx 9.69 \). Table A7.2 given in Appendix 7 shows full details of applying Johnson’s transformation to predictive variables of the group

2. Consider the second group of predictors. It contains: “share price in pounds”, “average trading volume in millions of shares per month”, “market cap in millions of pounds”, “variance of beta”, “dividend yield”, “dividend yield, over last 2 years”, “dividend yield, over last 3 years”. These variables are incorporated into the same group since all of them can take only positive or zero values. Here both Johnson’s and Box-Cox transformation can be applied and now we are interested to compare the goodness of two methods. Figure 3.3 shows the distribution of market capitalisation, with a skewness of 8.5 (indicating that the data are skewed right) and kurtosis of 96.3, caused by a few very large companies such as Shell or BP.

Figure 3.3. An untransformed predictor variable: market capitalisation December 1994-November 2004 with superimposed truncated normal curve
Firstly, let us introduce the results of using the Box-Cox transformation. From Figure 3.4 it could be seen that the resulting distribution is not in fact normal, being bimodal. Nevertheless the transformation gave a skewness of only 0.0123 and a kurtosis of only -0.1 and the variable is now usable as a predictor. The optimal value of the likelihood function is approximately -439367. The results for all variables of the group are shown in Table A7.1 given in Appendix 7. Furthermore, the transformation is almost logarithmic since $\omega = 0.08$ and $\lambda = 0$.

Secondly, consider the performance of Johnson’s method for positively defined predictor variables. Referring back to section 3.3, optimal transformations are found by varying the parameters over the grid of values to maximize the profile likelihood function. Now let lower and upper bounds for the variation of $a$ be fixed and run the program with different values for the upper bound of $c$.

![Figure 3.4. A transformed predictor variable: market capitalisation December 1994- November 2004 after Box-Cox transformation with superimposed normal curve](image-url)
The outcome of the sequence of program executions for all positively defined predictive variables is presented in Tables A7.3-A7.9 of Appendix 7. Figures 3.5a, 3.5b, 3.5c and 3.5d show the results of applying Johnson’s transformation to market capitalisation. The results are obtained on using different values of an upper bound for the variation of c. In more detail, one after another we have chosen the following values of the upper limit: 

(a) $5/\sigma_x$, (b) $50/\sigma_x$, (c) $500/\sigma_x$, (d) $5000/\sigma_x$, where $\sigma_x \approx 8599.85$. Optimal values of $c$ with corresponding optimal values of the log-likelihood function (3.9) are approximately given by: (a) $5.63E-04$ and $-496480$, (b) $5.62E-03$ and $-460475$, (c) $5.62E-02$ and $-446982$, (d) $5.62E-01$ and $-441110$.

We can see from above that the optimal value of $c$ coincides with the value of an upper limit for $c$ in each of these 4 cases. Several attempts to avoid this have been made by further increase in the upper limit, however the results were similar. This suggests that if the upper limit for the variation of $c$ and consequently the optimal value of $c$ are tending to infinity, the optimal value of the log-likelihood for Johnson’s transformation is tending to the optimal value of the log-likelihood for the Box-Cox transformation. It also explains why Figures 3.4 and 3.5(d) look quite similar.
Figure 3.5. A transformed predictor variable: market capitalisation December 1994- November 2004 after Johnson’s transformation with superimposed truncated normal curve (for different values of an upper bound for the variation of c)

The discussion given above can analogously be repeated for any other variable from this group. Appendix 8 provides the results of transforming variance of beta to near normality.
References


88
CHAPTER 4

Improving the accuracy of return predictions
4.1 Introduction

In the second chapter we outlined three approaches to model return on a stock: the market model (section 2.5), the Fama and French three-factor model (section 2.11) and the regression-based approach (section 2.13). All these models are parameter dependant, but contain different numbers of parameters.

Recall that the Markowitz strategy is implemented by investing so as to maximise expected (exponential) utility (see section 2.1). The problem is that we only have the estimates of model parameters, not the true values. In fact, the expected utility of an investment is decreased by model parameter uncertainty. Customarily, parameters are estimated from past data by least-squares regression procedures. It is well known that maximum-likelihood regression predictions are very 'noisy'. Better investment may be achieved by estimating parameters more accurately.

In section 2 of this chapter we turn to the well-known market model discussed in section 2.5. The section describes the adjustments that can be utilised to improve the accuracy of the estimate of beta. Several methods are examined here, but particular emphasis is placed on the Vasicek (1973) correction. With thousands of companies in the stock market, prior distributions of model parameters can be estimated accurately from monthly return data, and so empirical Bayes (EB) methods can be used to provide parameter estimates. Vasicek (1973) gave what we would now regard as an EB estimate of beta.
In section 4.3 we develop the idea of adopting an EB approach for the model introduced in section 2.13 (where the vector of parameters $\hat{\delta}$ is given in addition to the parameters of market model). We parallel the work done by Vasicek and assign the prior distributions to the model parameters. We then outline the scheme for generating Bayesian estimates of the regression coefficients. We do not however simply adopt Vasicek’s approach for our model. The original work here and in the subsequent sections is the methodology of generating Bayesian adjusted coefficients, which is different from the one utilised by Vasicek.

Different formulations of prior distributions will lead to different estimates of the parameters. Section 4.4 deals with the ridge regression approach as applied to the parameters $\delta_k$, $k = 1, \ldots, p$. Regression coefficients are made comparable by standardising the corresponding variables to have zero mean and unit variance. The method can be motivated in a Bayesian or empirical Bayes context by giving regression coefficients independent prior normal distributions with zero mean and common variance. In section 4.5 we describe the procedure of generating parameter estimates when the prior distribution of $\hat{\delta}$ is the student $t$-distribution.

Finally, in the last section, we describe the programming methodology for the ridge regression. The programs are written by the author and are based on the scheme described in section 4.4.
It needs to be pointed out that for all computations presented in this chapter we assume that predictive variables were already transformed to near normality and then given an affine transformation. Chapter 3 is fully devoted to the methodology of transforming.

4.2 Adjusting historical estimates of beta

Section 5 of chapter 2 presents the market model which linearly relates returns on the security and market returns. Least-squares regression analysis is usually used to estimate the alpha, beta and specific risk in the equation for return. However, regression analysis only provides us with the estimates of the model parameters. The true values are not known and the estimates are subject to error. Many empirical studies have shown that the accuracy of beta may be considerably improved via applying different adjustment techniques. One of them has been suggested by Blume (1975). Basing his reasoning on the empirical evidence he came to the conclusion that “estimated beta coefficients tend to regress towards the grand mean of all betas over time”. In order to capture this tendency adjustments to the estimates of beta obtained from historical data can be made. As a result, the beta forecast error can be substantially reduced. The modification scheme proposed by Blume utilises simple linear regression. Initially, the betas for individual securities or/and portfolios are computed for two adjacent but non overlapping time periods. The betas for the later period are then regressed against the betas for the earlier period. Following this procedure the regression tendencies from the earlier to later period
can be determined. Then the regression coefficients obtained from foregoing regression are used to forecast beta for the third subsequent time period.

However, there is another, and perhaps more pragmatic representation of Blume’s correction. As discussed in section 4 of Chapter 2 we are interested in estimating expected return on stocks. It can be shown that the following equation for return emerges from Blume’s method.

\[ R_i = c_i + \hat{\alpha}_i + c_2 \hat{\beta}_i R_m + \epsilon_i, \]

where \( \hat{\alpha}_i \) has already been estimated. In the equation above beta is estimated from the time-series regression and is used in cross-sectional regression to estimate next month’s return afterwards.

An alternative adjustment method has been proposed by Vasicek (1973). In his paper the Bayesian approach to beta estimation is discussed in comparison with conventional sampling-theory procedures (least-squares technique). Conversely to classical inference which assumes that parameters (such as beta) take unique values, although these values are unknown, in Bayesian approach the parameters are treated as random variables. The essence of the Bayesian approach consists in assigning probability distributions not only to data variables but also to the parameters (regression coefficient beta in our case). The information available prior to sampling expressed in the form of a prior distribution is a central point in Bayesian inference. One of the fundamental assumptions behind Vasicek’s model is that the cross-sectional distribution of betas is approximately normal.
with hyperparameters $\bar{\beta}, \sigma_{\beta}^2$. In other words, $\bar{\beta}$ is the average beta over all companies in the stock market and $\sigma_{\beta}^2$ is the variance of the distribution of betas. Let $\hat{\beta}$ be the estimate of beta obtained from regressing return on the share versus return on the market and the variance of the estimate $\sigma_{\hat{\beta}}^2$.

Utilising Bayes' Theorem Vasicek has shown that the posterior distribution of the parameter $\beta$ is approximately normal with mean $\hat{\beta}$, where

$$\hat{\beta} = \frac{\bar{\beta} / \sigma_{\bar{\beta}}^2 + \hat{\beta} / \sigma_{\hat{\beta}}^2}{1 / \sigma_{\bar{\beta}}^2 + 1 / \sigma_{\hat{\beta}}^2}.$$ (4.1)

The Bayesian adjusted estimate $\hat{\beta}$ is the weighted average of sample and prior estimates. The adjustment makes $\hat{\beta}$ shrink towards the prior mean $\bar{\beta}$, which is approximately unity. The amount of shrinkage depends the strength of prior information and on the accuracy with which beta has been measured from the time-series regression. It is proportional to the precisions $h = 1 / \sigma_{\bar{\beta}}^2, \quad h = 1 / \sigma_{\hat{\beta}}^2$ of the prior distribution and sample estimate respectively. Formula 4.1 shows that observations with large standard errors pull further towards the mean than observations with small standard error.

The variance of the weighted estimate is given by $\sigma_{\hat{\beta}}^2 = \sigma_{\beta}^2 (\sigma_{\beta}^2 + \sigma_{\hat{\beta}}^2)$ which is smaller than either $\sigma_{\beta}^2$ and $\sigma_{\hat{\beta}}^2$. It can be seen that, the precision of the posterior distribution
\[ h^* = \frac{1}{\sigma_{\hat{\beta}}^2} + \frac{1}{\sigma_\beta^2} \] is the sum of the precision of \( \hat{\beta} \) and the precision of the prior distribution.

Vasicek’s method is an attempt to incorporate available prior knowledge together with the sample information so that modified estimates are more accurate. The author highlights the importance of adopting Bayesian approach for the problem of estimating betas of stocks, since in this case “the prior information is usually sizeable”.

Klemosky and Martin (1975) and Eubak and Zumwalt (1979) investigated the usefulness of the adjustment techniques. They utilised decomposed mean squared error as a measure of forecast error.

“MSE was chosen over alternative measures of forecast error because of its statistical tractability and because it can be easily partitioned into three components (bias, inefficiency and random error)” (Klemosky and Martin, 1975, p. 1124).

Both of the studies indicate that Blume and Vasicek adjustments lead to more accurate forecasts of future betas than did the unadjusted betas, however the Bayesian technique has a slight tendency to outperform Blume’s method.

The beta coefficients are also affected by infrequent trading. If the share price does not change between two trades the return over this period is simply zero. This leads to an
estimate of beta which is too low as compared to the true beta. This is an especially serious problem in small security markets. Scholes and Williams (1977) and Dimson (1979) proposed corrections for asynchronous trading. These kinds of adjustments are usually used when the daily data are analysed. However, with monthly data they may still be needed. The current study does not use corrections for thin trading since market capitalisation is included in regression as one of the variables.

4.3 Bayesian adjustments as applied to the regression-based model described in section 2.13

Vasicek (1973, p. 1233) recommended the technique “for generating Bayesian estimates of the regression coefficient of rates of return of a security against those of a market index”. Following Vasicek’s logic we utilise a Bayesian method to compute the estimates of parameters for the model described in section 13 of chapter 2. It can be seen that this model contains a vector of parameters delta in addition to single coefficient beta. An attempt to improve the accuracy of return prediction by applying Bayesian adjustments not only to the betas but also to the elements of vector delta was made. However, the approach of computing the adjusted estimates we use is different from the one adopted by Vasicek. While Vasicek derives explicit formulas for the estimates assuming that the values of the hyperparameters (parameters of the prior distributions) are known (see formula 4.1) we obtain the estimates of hyperparameters from the data.
Adjusted estimates of model parameters and estimates of hyperparameters are computed by maximising the likelihood (4.2) with respect to them. The approach is widely accepted and practiced in the statistical community (e.g. Tibshirani, 1996).

\[
L = \prod_{i=1}^{N} \left\{ (2\pi \sigma_i^2)^{-n_i/2} \exp\left(\sum_{j=1}^{n_i} \frac{(R_{ij} - \alpha_i - \delta_j x_{ij} - \beta_i R_{mj})^2}{2\sigma_i^2}\right) \right\} \\
\times f_\alpha (\alpha_i | \theta_\alpha) \times f_\beta (\beta_i | \theta_\beta) \times f_\sigma (\sigma_i | \theta_\sigma) \\
\times \prod_{k=1}^{p} f_\delta (\delta_k | \theta_\delta),
\]

(4.2)

where \( N \) is a number of companies under observation and company \( i \) has \( n_i \) previous months’ data to be used, there are \( p \) predictors, and \( x_{ij} \) denotes the vector of transformed and normalised predictor variables for the \( i \)-th company and \( j \)-th month.

Note that the function to be maximised is simply the conventional likelihood for the model multiplied by the prior probability density functions assigned to each of the model parameters. Here the prior pdfs depend on the vectors of hyperparameters given by \( \theta_\alpha \), \( \theta_\beta \), \( \theta_\sigma \) and \( \theta_\delta \). As mentioned earlier the elements of these vectors are also unknown and need to be estimated from the data.

The assumptions behind our model are:

- both \( f_\alpha \) and \( f_\beta \) are Gaussian pdfs with vectors of parameters \( \theta_\alpha = (\alpha, \sigma_\alpha) \)

and \( \theta_\beta = (\beta, \sigma_\beta) \) respectively (taking normal distribution as a prior for beta
is a reasonable choice, it is well known that beta approximately follows a Gaussian distribution with the mean 1 and standard deviation of 0.2).

- $\sigma_i^2$ follows inverse Gamma pdf with parameters $\Theta_\sigma = (\alpha, \beta)^T$.

- EB prior for vector $\delta$ assumes that the parameters are iid, and have a mean value of zero.

In order to reduce the computational burden it is convenient to have a normal distribution as a prior for each element of vector $\delta$. Chapter 9 of O'Hagan (1994) deals with the Bayesian analysis for the normal linear model under various formulations of prior distributions. The general behaviour of the ridge regression estimator is considered as one of the special cases. If the elements of vector $\delta$ are independent and each of them is normally distributed with the mean of zero then, as O'Hagen (1994, p. 263) points out “The zero prior means cause the posterior estimates to be gradually shrunk towards zero, relative to the classical least squares estimates”. The smaller the prior variance, the more influence the prior has and the shrinkage towards the origin is greater.

At the same time, O'Hagan (1994) argues that, although it is convenient to assume a normal prior distribution for the parameters, this often gives unrealistically thin tails. He provides the following example to show that prior beliefs are often better represented by a heavier-tailed distribution than the normal.

---

8 if the random variable $\xi$ follows a Gamma distribution, then $1/\xi$ is distributed inverse Gamma

98
"With a normal prior distribution, the prior probability that the parameter will lie more than, say, two and a half prior standard deviations from its prior mean is very small (0.0124), and in practice we wish to allow somewhat more probability to the event of the prior mean being far from the true value". (O'Hagan (1994). p. 273)

Very commonly Student t-distributions are used as an appropriate family of heavy-tailed distributions.

Nevertheless, in addition to three assumptions made earlier let us assume, for the moment, that the prior distributions for each of the elements of \( \delta \) are Gaussian pdfs with parameters \( \Theta_s = (0, \sigma_s) \). However, there is still a practical problem with maximising likelihood (4.2). Since the likelihood is nonlinear in model parameters, we will be maximising it as a function of very many parameters indeed (e.g., there are 3 per company, plus some others).

Luckily (in this context) the term \( \delta^T \mathbf{x} \) is 'small' and hence ignoring it we can get 'good' estimates of all the other model parameters such as \( \alpha_i, \beta_i \) and \( \sigma_i^2 \) for \( i = 1, \ldots, N \).

The following four-step maximisation scheme can be proposed:

1. Perform regression analysis for each company separately, ignoring the \( \mathbf{x} \) term. Initially we maximise the likelihood (4.2) with respect to \( \alpha_i, \beta_i \) and \( \sigma_i \).
\( i = 1, \ldots, N \) assuming a known value for \( \delta \) (e.g., zero) and hyperparameters values (elements of vectors \( \Theta_\alpha, \Theta_\beta, \Theta_\sigma \)).

2. Carry out the regression for delta. Plug the values \( \hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i, i = 1, \ldots, N \) estimated on a previous step back into the likelihood function and get \( \hat{\delta} \), assuming that \( \Theta \) parameters are known.

3. Cycle round these steps until convergence, using the new value of \( \delta \). The \( \alpha_i, \beta_i \) and \( \sigma_i \) will change slightly, but not much, because the term \( \delta^T x \) is small.

4. Determine hyperparameters \( \Theta \) - the prior distribution parameters for alpha, beta, sigma and delta. Plug the estimates obtained on step 3 back into the likelihood function to obtain a profile likelihood that can be then maximised for the hyperparameters \( (\overline{\alpha}, \sigma_\alpha, \overline{\beta}, \sigma_\beta, \alpha, \beta \) and \( \sigma_\delta \)).

The convergence for \( \delta \) on step 3 of the algorithm is slow. Figure 4.1 shows the convergence for the market cap regression coefficient (for the arbitrary choice of hyperparameters).
Figure 4.1 Convergence of the market capitalisation regression coefficient

Figure 4.2 presents the pictorial representation of the algorithm given above. It can be seen that it is necessary to fix $\delta$ and maximise the likelihood function for the $\alpha_i$, $\beta_i$, and $\sigma_i$, then to maximise for $\delta$, and to alternate these steps until convergence for each choice of the hyperparameters. The likelihood function was then maximised for the hyperparameters.
The task of maximising likelihood (4.2) with respect to hyperparameters (which corresponds to the step 4 of the scheme) consists in computing the likelihood for the different choices of hyperparameters. For each choice of the hyperparameters we perform the following algorithm:

1. **Particular choice of hyperparameters**
   - \( \alpha, \sigma_a, \overline{\beta}, \sigma_\beta, \alpha, \beta, \sigma_\delta \)

2. Delta is plugged back until convergence is achieved.
   - Set each element of vector delta to zero: \( \delta_k = 0, k = 1, \ldots, p \)

3. Assume that \( \delta \) is fixed and calculate \( \alpha_i, \beta_i, \sigma_i, i = 1, \ldots, N \) (for each company) correspond to the step 1.

4. Given the values of \( \alpha_i, \beta_i, \sigma_i, i = 1, \ldots, N \), compute \( \delta \) correspond to the step 2.

5. Calculate the value of the log-likelihood function.

---

8 Figure 4.2. Graphical representation of the maximisation scheme
In the light of the discussion about the choice of the prior distribution for $\delta$, it needs to be pointed out that Gaussian distribution is a convenient choice because in this case the maximisation with respect to the $\delta_k, k = 1, \ldots, p$ gives a set of linear equations that are relatively easy to solve. On the other hand, if $f_\delta$ is long-tailed the problem is nonlinear and we need to iterate for the optimal values of $\delta$. In any case, if the t-distribution is taken, the Gaussian prior values of $\delta$ and the degrees of freedom (infinity) are starting values for the full iteration.

4.4 Detailed implementation of the maximization scheme: prior distribution of $\delta$ is Gaussian.

Recall that we are maximizing the likelihood function (4.2) with respect to unknown model parameters ($\alpha_i, \beta_i$ and $\sigma_i$ for each company $i = 1, \ldots, N$ and $\delta_k, k = 1, \ldots, p$) and hyperparameters ($\bar{\alpha}, \sigma_\alpha, \bar{\beta}, \sigma_\beta, \alpha, \beta$ and $\sigma_\delta$).

Now the optimization scheme presented in the previous section for the case when the prior distribution of $\delta$ is Gaussian is discussed in more detail. Taking into account the assumptions made earlier, the likelihood function for the model is given by
\begin{equation}
L_z = \prod_{i=1}^{N} \{ (2\pi \sigma_i^2)^{-n_i/2} \exp\left(-\sum_{j=1}^{n_i} \frac{(R_{yj} - \alpha_i - \delta_i' \mathbf{x}_{yj} - \beta_i R_{mj})^2}{2\sigma_i^2} \right) \\
\times (2\pi \sigma_a^2)^{-1/2} \exp\left(-\frac{(\alpha - \bar{\alpha})^2}{2\sigma_a^2}\right) (2\pi \sigma_{\beta}^2)^{-1/2} \exp\left(-\frac{(\beta - \bar{\beta})^2}{2\sigma_{\beta}^2}\right) \alpha^\beta (\sigma_i^2)^{-1-\beta} \exp\left(-\frac{\alpha}{\sigma_i^2}\right) \}
\times \prod_{k=1}^{p} \left(2\pi \sigma_k^2\right)^{-1/2} \exp\left(-\frac{\delta_k^2}{2\sigma_k^2}\right),
\end{equation}

where the subscript \( z \) is used to indicate that normal distribution has been chosen as a prior for \( \delta_k, \, k = 1, \ldots, p \).

Let us recall that we have \( N \) companies under consideration with \( n_i \) previous months of data for the company \( i \). The observed data are monthly return on the shares \( R_y \) with corresponding monthly return on the market index \( R_{mj}, \, i = 1, \ldots, N, \, j = 1, \ldots, n_i \). Plus there are \( p \) predictor variables \( x_y^k \) such as market capitalisation in a previous month, returns over various periods, 1-year dividend yield and so on. All the predictive variables are transformed to near normality before they are used in regression.

Usually instead of the likelihood, its logarithm is maximized since working with sums is easier than working with products. This replacement is possible because the logarithm is a strictly monotonically increasing function and the likelihood \( L \) is positive. The derivative of the likelihood and the derivative of the logarithm of the likelihood will have exactly the same roots since
\[
\frac{\partial (\log L)}{\partial \gamma} = \frac{\partial L}{\partial \gamma} / L,
\]

where \( \gamma \) is the model parameter.

Finding the logarithm of \( L \) yields

\[
l_z = \log L_z = \sum_{i=1}^{N} \left\{ -\frac{n_i}{2} \log(2\pi \sigma_i^2) - \sum_{j=1}^{n_i} \frac{(R_{ij} - \alpha_i - \delta_j^T \mathbf{x}_j - \beta_i R_{nj})^2}{2\sigma_i^2} - \frac{1}{2} \log(2\pi \sigma_{\alpha_i}^2) \right\}
\]

\[
- \frac{(\alpha_i - \bar{\alpha})^2}{2\sigma_{\alpha_i}^2} - \frac{1}{2} \log(2\pi \sigma_{\beta_i}^2) - \frac{(\beta_i - \bar{\beta})^2}{2\sigma_{\beta_i}^2} + \beta \log \alpha - \log(\Gamma(\beta)) - (1 + \beta) \log \sigma_i^2 - \frac{\alpha_i}{\sigma_i^2}
\]

\[
+ \sum_{k=1}^{P} \left\{ -\frac{1}{2} \log(2\pi \sigma_{\delta_k}^2) - \delta_k^2 \right\}.
\]

(4.3)

We now outline the mathematics involved on each step of the algorithm

- **Step 1:**

The first step of the optimization algorithm is to maximize \( l_z \) with respect to \( \alpha_i \), \( \beta \), and \( \sigma_i^2 \) assuming the other parameters to be fixed. To find out where the function has its extreme values, just find out where its derivatives are equal to zero, thus we need to differentiate (4.3) with respect to unknown model parameters. The derivatives of function \( l_z \) w.r.t. \( \alpha_i \), \( \beta \), are given by
The way to calculate a maximum point itself now comprises three stages. The first stage is to find where function derivatives (4.4) are equal to zero, and that will give us a system of two simultaneous equations in 3 unknown variables. In spite of the fact that this is not enough to get a unique solution, we can easily express optimal values of $\alpha_i$ and $\beta_i$ in terms of $\sigma_i^2$.

Let us introduce the following notation:

a) these quantities are calculated directly from the observed data

$$R_{mi} = \frac{1}{n_i} \sum_{j=1}^{n} R_{mj}, \quad R_{mi}^2 = \frac{1}{n_i} \sum_{j=1}^{n} R_{mj}^2, \quad \bar{R}_m = \frac{1}{n_i} \sum_{j=1}^{n} R_{ij}, \quad \bar{R}_i = \frac{1}{n_i} \sum_{j=1}^{n} R_{ij}^2,$$

$$N_i = \frac{1}{n_i} \sum_{j=1}^{n} R_{ij} R_{mj}.$$

b) in this step we assume that the values of hyperparameters are given, so that the quantities

$$\nu_\alpha = 1/\sigma_\alpha^2, \quad \nu_\beta = 1/\sigma_\beta^2$$

can be calculated.
c) it is also assumed that the vector \( \delta \) is given, thus combining data with the elements of vector \( \delta \) we get

\[
E_i = \sum_{j=1}^{n_i} \delta^T x_{ij} / n_i, \quad Q_i = \sum_{j=1}^{n_i} \delta^T x_{ij} R_{mj} / n_i,
\]

\[
Z_i = \sum_{j=1}^{n_i} \delta^T x_{ij} R_{ji} / n_i, \quad P_i = \sum_{j=1}^{n_i} (\delta^T x_{ij})^2 / n_i.
\]

Then (See appendix 9 for a derivation of expressions (4.5) and (4.6))

\[
\hat{\alpha}_i = \frac{R_{mi}(N_i - Q_i + \bar{\beta} \nu_i \sigma^2 / n_i) + (E_i - \bar{R}_i - \bar{\alpha} \nu_i \sigma^2 / n_i)(R_{mi}^2 + \nu_i \sigma^2 / n_i)}{(R_{mi})^2 - (1 + \nu_i \sigma^2 / n_i)(R_{mi}^2 + \nu_i \sigma^2 / n_i)}, \quad (4.5)
\]

\[
\hat{\beta}_i = \frac{N_i + \bar{\beta} \nu_i \sigma^2 / n_i - \hat{\alpha} \bar{R}_{mi} - Q_i}{R_{mi}^2 + \sigma_i^2 \nu_i / n_i}, \quad (4.6)
\]

In the second stage we plug the estimates (4.5) and (4.6) back into (4.3). This gives us a function of only one variable \( \sigma_i^2 \). By maximizing (4.7) with respect to \( \sigma_i^2 \) for each company \( i = 1, \ldots, N \) the optimal values of \( \sigma_i^2, \hat{\sigma}_i^2 \) could be obtained

\[
\ell_z = \frac{n_i(R_i^2 - 2\alpha_i \bar{R}_i - 2Z_i - 2\beta_i N_i + \alpha_i^2 + 2\alpha_i E_i + 2\alpha_i \beta_i \bar{R}_{mi} + P_i + 2\beta_i Q_i + \beta_i^2 \bar{R}_{mi}^2)}{2\sigma_i^2} - \frac{n_i + 2\beta + 2}{2} \log(\sigma_i^2) - \frac{(\alpha_i - \bar{\alpha})^2}{2\sigma^2_a} - \frac{(\beta_i - \bar{\beta})^2}{2\sigma^2_{\beta}} - \frac{\alpha}{\sigma_i^2} + \text{const}. \quad (4.7)
\]

In the final stage, having optimal values of \( \sigma_i^2, \ i = 1, \ldots, N \), we plug them back to equations (4.5) and (4.6). This gives us the optimal values of parameters \( \alpha_i \) and \( \beta_i \) for each company \( i = 1, \ldots, N \).
On the second step of the optimization algorithm, having the optimal values of $\alpha$, $\beta$, and $\sigma^2$, we wish to adjust the parameters $\delta_k, k = 1, ..., p$ to give the maximum likelihood.

To find the extreme values $\hat{\delta}_k$ of $\delta_k$, we take the derivatives of (4.3) with respect to these parameters. This gives us a system of $p$ linear simultaneous equations in $p$ unknowns.

$$\sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{(R_{ij} - \alpha_i - \hat{\delta}_k^T x_{ij} - \beta_i R_{ij} x_{ij}^k)}{\sigma_i^2} \hat{\delta}_k - \frac{\hat{\delta}_k^2}{\sigma_\delta} = 0 , \quad k = 1, ..., p , \quad (4.8)$$

where $x_{ij}^k$ is the $k$-th element of vector $x_{ij}$ and $\delta_k$ is the $k$-th element of vector $\delta$.

In order to simplify further calculations let us introduce the following notation:

$$W_i^k = \sum_{j=1}^{n_i} R_{ij} x_{ij}^k, \quad V_i^k = \sum_{j=1}^{n_i} x_{ij}^k, \quad U_i^k = \sum_{j=1}^{n_i} R_{ij} x_{ij}^k \quad \text{and} \quad G_i^{k,l} = \sum_{j=1}^{n_i} x_{ij}^l x_{ij}^k .$$

Using the notation given above, system (4.8) can be rewritten as

$$\sum_{l=1}^{p} \delta_l \left( \sum_{i=1}^{N} \frac{G_i^{k,l}}{\sigma_i^2} + \frac{1}{\sigma_\delta^2} \delta_l \right) = \sum_{i=1}^{N} \left\{ \frac{W_i^k - \alpha_i V_i^k - \beta_i U_i^k}{\sigma_i^2} \delta_l \right\} , \quad k = 1, ..., p , \quad (4.9)$$

where $\delta_l$ is Kronecker delta.

Furthermore, we can rewrite the system of equations (4.9) in matrix form which makes the representation easier to understand. We can define a 'data vector' $b$ and a 'measurement matrix' $A$ with components as follows:
\[ b_k = \sum_{i=1}^{N} \left\{ \frac{1}{\sigma_i^2} \left( W_i^k - \alpha_i V_i^k - \beta_i U_i^k \right) \right\}, \quad A_{kl} = \sum_{i=1}^{N} \frac{G_i^{k,l}}{\sigma_i^2} + \frac{1}{\sigma_\delta^2} \delta_{lk}. \]

Note that \( A \) is symmetric. The conditions (4.9) are then the matrix equation

\[ A \delta = b \]

with the solution \( \delta = A^{-1} b \).

- **Step 3:**

Note that plugging \( \hat{\delta} \) back to (4.3) leads to new values for \( \hat{\alpha}_i, \hat{\beta}_i \) and \( \hat{\sigma}_i \), which in turn generates a new maximum likelihood estimate for \( \delta \).

- **Step 4:**

After the above, we have the estimates of \( \alpha_i, \beta_i, \sigma_i^2 \) for each company \( i = 1, \ldots, N \) and the estimates of the components of vector \( \delta \) available. However, we still need to estimate seven hyperparameters since we just assumed them to be fixed on the previous steps. Applying the method of profile likelihood again on the final step of the optimization algorithm we plug \( \hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i^2 \) and \( \hat{\delta} \) back to (4.3) and iterate for the optimal values of hyperparameters by maximizing the log-likelihood w.r.t. them. Once we have got the optimal values of hyperparameters it is possible to use them and recalculate all the other parameters obtained in steps 1 and 2 above.
4.5 Detailed implementation of the maximization scheme: prior distribution of $\delta$ is Student $t$

As mentioned in section 4.2 prior beliefs are often better represented by a heavier-tailed distribution than the normal. In this context a $t$-distribution might be considered as a useful specimen of a heavy-tailed distributional family.

In this section we are concerned with the implementation of the maximisation algorithm when the prior distribution of $\delta$ is $t$-distribution with $\nu$ degrees of freedom. To estimate the model parameters in this case a procedure similar to the algorithm given in the previous section can be performed. But now we have a different objective function, to maximize $(4.10)$ instead of $(4.3)$, and we need to iterate for the optimal values of $\delta$ on step 2 since the analogue of system $(4.9)$ is not linear any more.

Applying the $t$-distribution to the problem results in the form of log-likelihood function given below

$$l_i = \sum_{i=1}^{N} \left\{ -\frac{n_i}{2} \log(2\pi\sigma_i^2) - \sum_{j=1}^{n_i} \frac{(R_{ij} - \alpha - \delta^T x_{ij} - \beta_i R_{ij})^2}{2\sigma_i^2} - \frac{1}{2} \log(2\pi\sigma_i^2) \right\}$$

$$- \frac{(\alpha_i - \bar{\alpha})^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2_p) + \beta \log \sigma_i \log(\Gamma(\beta)) - (1 + \beta) \log \sigma_i^2 - \frac{\alpha}{\sigma_i^2}$$

$$+ \sum_{k=1}^{p} \left\{ -\log \sigma_\delta + \log \Gamma \left( \frac{\nu+1}{2} \right) - \log \Gamma \left( \frac{\nu}{2} \right) - \frac{1}{2} \log(\pi\nu) - \frac{\nu+1}{2} \log \left( 1 + \frac{\delta^2_k}{\nu\sigma_\delta^2} \right) \right\}, (4.10)$$
where \( \nu \) is the number of degrees of freedom and subscript \( t \) is used to indicate that a \( t \)-distribution has been chosen as a prior for \( \delta_k, k = 1, \ldots, p \).

Those changes in the form of log-likelihood function do not affect the implementation of step 1 of the optimisation algorithm given in section 4.4 since the last term in formula (4.10) does not contain any of the parameters \( \alpha_i, \beta_i, \sigma_i^2 \), whereas step 2 suffers from some considerable changes. We now consider it (step 2) in detail.

- **Step 2:**

Before finding the logarithm and carrying out the differentiation with respect to \( \delta_k, k = 1, \ldots, p \) in the likelihood function (4.2) it is convenient to replace the prior distribution of \( \delta_k - f_\delta'(\delta_k|\theta_\delta) \) by its equivalent \( f_\delta'(\delta_k|\theta_\delta) = f_{\delta t}^*(\delta_k|\theta_\delta) f_{\delta z}^*(\delta_k|\theta_\delta) \). We denote by an upper index \( t \) and \( z \) Student and Gaussian pdfs respectively.

Substituting the ratio

\[
\frac{f_\delta'(\delta_k|\theta_\delta)}{f_\delta'(\delta_k|\theta_\delta)} = \frac{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{2\exp\left(\frac{(\nu-2)\delta_k^2}{2\nu\sigma_\delta^2}\right)}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu-2\left(1+\frac{\delta_k^2}{\nu\sigma_\delta^2}\right)}}
\]

likelihood with respect to unknown parameters \( \delta_k, k = 1, \ldots, p \) yields the system of nonlinear simultaneous equations

\[
\sum_{i=1}^{N} \sum_{j=1}^{n} \frac{R_{ij} - \alpha_i - \hat{x}_i y_j - \beta_i R_{ij} x_j}{\sigma_i^2} = \frac{(\nu-2)\hat{x}_i}{\nu \sigma_\delta^2} + \frac{(\nu-2)\hat{y}_i}{\nu \sigma_\delta^2} - \frac{(\nu+1)\hat{\delta}_k}{\nu \sigma_\delta^2 + \delta_k^2} = 0, \  k = 1, \ldots, p. \quad (4.11)
\]
In deriving the ratio given above, we assumed that \( \text{var}(\delta_k) = \frac{\nu}{\nu-2} \sigma_\delta^2 \) for the normal pdf.

Note, that if the number of degrees of freedom tends to infinity \( \nu \to \infty \), then

\[
\frac{(\nu-2)\delta_k}{\nu\sigma_\delta^2} - \frac{(\nu+1)\delta_k}{\nu\sigma_\delta^2 + \delta_k^2} = 0
\]

and, therefore, this case is identical to the case considered in the previous section when the normal prior was used for the elements of vector \( \delta \).

Turning back to matrix notation note that the ‘data vector’ \( b \) does not get affected by applying the \( t \)-distribution (for \( \delta \)) instead of the Gaussian distribution while ‘measurement matrix’ \( A \) changes in an obvious way and its components now are

\[
A_{ij} = \sum_{i=1}^{N} \frac{G_{ij}}{\sigma_i^2} + \frac{(\nu-2)}{\nu\sigma_\delta^2} \delta_{ik} = A_{ik}.
\]

We also define a \( p \times p \) diagonal matrix \( C(\delta) \) with the elements \( C_k(\delta_k), k = 1, \ldots, p \).

\[
C_k(\delta_k) = \frac{(\nu-2)}{\nu\sigma_\delta^2} - \frac{(\nu+1)}{\nu\sigma_\delta^2 + \delta_k^2}.
\]

The conditions (4.11) are then the matrix equation

\[
A\delta = b + C(\delta)\delta
\]  

(4.12)

The following approximating scheme might be proposed to find the solution of the equation given above.

To get sensible starting values on the first iteration we assume that \( \nu \to \infty \) and therefore, \( C(\delta) = C_0(\delta) = 0_{p \times p} \). Then matrix equation (4.12) simplifies to equation
(4.13), which is linear and corresponds to the case when we use a Gaussian pdf as a prior for \( \delta \)

\[
A\delta^{(0)} = b.
\]  

(4.13)

Solving (4.13) for \( \delta^{(0)} \) gives values of \( \delta^{(0)} = A^{-1}b \) that can be used as initial values for \( \delta \) in the following recurrence formula

\[
\delta^{(n+1)} = A^{-1}b + A^{-1}C(\delta^{(n)})\delta^{(n)} = \delta^{(0)} + A^{-1}C(\delta^{(n)})\delta^{(n)}.
\]  

(4.14)

On each subsequent iteration we use the updating equation (4.14) to obtain more accurate values for \( \delta \), where the upper index \( n \) denotes the number of iteration.

### 4.6 Programming methodology

A series of FORTRAN 95 programs has been written by the author (the listing of the program is given in appendix 10). These programs use the scheme given in section 4.3 to estimate the regression coefficients and hyperparameters by maximizing the log-likelihood function (4.2). It is significant to point out that in order to simplify the programming technique the assumption was made that the prior distribution of \( \delta \) is Gaussian.

The programs use the data file “lsedata.dat” as a source of input information. As mentioned earlier, this data sample contains monthly financial data for the period from 1955 up to 2004. It is supposed that at any time point by means of our programs we can
get sensible estimates of $\alpha_i$, $\beta_i$, $\sigma_i^2$ for each company $i = 1, \ldots, N$, as well as the vector of parameters $\delta$ and a set of 7 hyperparameters for all the companies. To get optimal values of the model parameters we just use ten years' data back from the current time point and assume that $\alpha_i$, $\beta_i$, $\sigma_i^2$, $\delta$ and hyperparameters were constant over that period. The estimated values of regression coefficients can be used in predicting future returns.

In more detail, programming methodology comprises the following steps:

1. Before carrying out the regression analysis, the predictor variables need to be transformed to approximate normality. In chapter 3 we examined the theory of transformations for achieving approximate normality and discussed the application of two transformation methods. The data file has to be read once for each variable to be transformed. Thus, for each variable we take all the company-months in the ten-year period for which this variable is defined and calculate the optimal transformation. This gives us 4 parameters, including the affine transformation.

2. Read the file 'params.con' that contains starting values for the hyperparameters and for the elements of vector $\delta$. The initial values of $\delta$ should be specified, since $\delta$ is assumed to be fixed when we do separate regressions for the companies.
Figure 4.3. The structure of the data. All possible situations of positional relationship between the data for the company and the data required for computing parameters and making a forecast are modelled. The aeries marked with blue represent the data utilised in computing regression coefficients and estimates of hyperparameters.
3. Iterate for the optimal values of the prior distribution parameters (hyperparameters) using NAG FORTRAN Library Routine E04UCF. This routine is designed to minimize an arbitrary smooth function of several variables subject to constraints using a sequential quadratic programming method. Note that this step corresponds to the fourth step of the optimization algorithm considered in section 4.4. In our particular case we have got a function of 7 variables (4.2) to be minimized without any constraints. To ensure that the solution we get is a global optimum, several random restarts from the maximum might be carried out. In E04UCF the objective function is defined by subroutine FUNCT which has to be supplied by user (FUNCT is one of the parameters for E04UCF that is needed to be imputed when the routine is called). The log-likelihood function (4.2) depends not only on hyperparameters but also on the values of parameters $\alpha_i$, $\beta_i$, $\sigma_i^2$, $\delta$ that have to be evaluated within the subroutine FUNCT. Therefore, within FUNCT we have to carry out the following steps:

a) Read the data file company by company. For each company pick out all the available data that we have in this ten-year period (see Figure 4.3) and transform the predictor variables using the transformation parameters obtained before. If there is less than five years’ data in the ten-year period available we do not use that company. However, the company need not still be “alive” at the end of the period to be included for consideration. For each company, we use a maximum of 120 months’ data to calculate $\bar{R}_{m}$, $\bar{R}_{m}^2$, $N_i$, $\bar{R}_i$, $\bar{R}_i^2$, $Q_i$, $Z_i$, $P_i$, $E_i$, $W_i^k$, $V_i^k$,
$U_k^i$ and $G_i^{kl}$, $k = 1, \ldots, p$, $l = 1, \ldots, p$. Assuming that the elements of the vector $\delta$ are known (initially, they all are zeros), we can evaluate the optimal values of $\alpha_i$, $\beta_i$, $\sigma_i^2$. The mathematics introduced in step 1 of the optimization algorithm is involved in computing these parameters. NAG FORTRAN Library Routine E04ABF is used to maximize function (4.7) with respect to $\sigma_i^2$. E04ABF searches for a minimum, in a given finite interval, of a continuous function of a single variable, using function values only.

b) This step corresponds to step 2 of the optimization algorithm. Now we have got $\alpha_i$, $\beta_i$, $\sigma_i^2$, we calculate the deltas by solving the system of linear simultaneous equations (4.9). The NAG routine that enables those computations to be done is F04ATF. It is designed to calculate the accurate solution of a set of real linear equations with a single right-hand side, using an LU factorization with partial pivoting, and iterative refreshment.

Once we get $\delta$ we can recalculate the values of $\alpha_i$, $\beta_i$, $\sigma_i^2$ obtained in step a). Cycling round these steps using non-zero values for the elements of vector $\delta$ makes $\alpha_i$, $\beta_i$, $\sigma_i^2$ change slightly. Only slightly, because the elements of $\delta$ are small. It is probably best to have a fixed number of cycles when we recalculate $\alpha_i$, $\beta_i$, $\sigma_i^2$ and $\delta$ alternately.
We will do this calculation for each month when we want to forecast the next month's return.

References


CHAPTER 5

Results: shrunken regression model
5.1 Introduction

In chapter 4 we proposed the original model for predicting returns. This model utilises a ridge regression approach. We now present the results for ridge regression approach and compare them with the estimates obtained using other methods. We start by outlining several rival methods involved in comparison procedures. We then discuss the results in section 5.3.

Section 5.4 deals with the evaluation of regression performance which was done entirely out-of-sample. Measures of forecast accuracy often used in forecasting competitions such as M3 (Makridakis and Hibon, 2000), like symmetric mean absolute percentage error, average ranking, percentage better, median symmetric absolute percentage error and median relative error are utilised.

5.2 Rival methods

The ridge regression model has already been examined in detail. We now briefly review another four methods: conventional approach (least-squares regression), vague prior for \( \delta \) (no ridge regression for the elements of \( \delta \), other parameters are still shrunken), market model and best subsets regression.

---

9 In fact, EB adjustments (shrinkage) were applied to all the parameters. Ridge regression is a special case of shrinkage which was utilised only for the elements of vector \( \delta \). Other parameters were not shrunken to zero. In spite of this, in order to avoid confusing terminology, here and thereafter we refer to the model discussed in the previous chapter as ridge regression model.
Conventional approach.

We assume that all the hyperparameters (for the distributions of alpha, beta and deltas) correspond to vague prior parameters. This case is equivalent to the case when we do not have any priors in the log-likelihood function (4.2) at all and a conventional least-squares method is used to estimate the regression coefficients. None of the coefficients, even betas, are shrunken. The programming methodology (for shrunken regression) described in section 4.6 can still be used. In terms of programming it means that we do not iterate for the optimal values of hyperparameters, just keep them all fixed at their starting values while the starting values correspond to the vague prior parameters. The parameters for the vague priors could be chosen as follows:

1. for the distribution of $\alpha_i$: $\bar{\alpha} = 1.0E-05$, $\sigma_\alpha = 1.0E+05$;
2. for the distribution of $\beta_i$: $\bar{\beta} = 1.0E-05$, $\sigma_\beta = 1.0E+05$;
3. for the distribution of $\sigma_i^2$: $\alpha = 1.0E-05$, $\beta = 1.0E-05$;
4. for the distribution of $\delta$: $\sigma_\delta = 1.0E+05$.

Assuming these values for the hyperparameters, we simply calculate alphas, betas and sigmas for each company separately, determine the values for deltas and finally plug deltas back to the log-likelihood to compute new values of alphas, betas and sigmas (cycle round these steps several times). Notice that all the variables we have got (in the data file ‘lseddata.dat’) are included in the regression model.
• Vague prior for $\delta$

In this case we do not apply shrinkage to the elements of vector $\delta$, while other parameters are still adjusted. When running the programs we take a very huge number as a starting value for $\sigma_\delta$ and keep $\sigma_\delta$ fixed. At the same time, we still iterate for the optimal values of the hyperparameters for alpha, beta and sigma. So that Vasicek’s correction for betas is utilised.

• Market model

For a short description of this model see section 2.5, while for greater details see Elton et al (2003). One of the fundamental assumptions behind this model is that the changes in return on the stock can be fully explained only by changes in the return on the market (delta term in regression equation 2.17 is equal to zero). Then the regression equation for return on a stock $i$ is

$$R_i = \alpha_i + \beta_i R_m + \epsilon_i.$$  

This method is the crudest among the methods we use for comparison. To obtain the estimates of the regression coefficients $\alpha_i$ and $\beta_i$ we carry out a simple regression: return on the share versus return on the market. Again, in this case, we shrink none of the regression coefficients. In terms of programming it means that we fix all the hyperparameters at their starting values, while starting values correspond to vague prior
parameters. The only difference between this method and conventional approach (from the point of view of the programming) is that, since the delta term is zero, we do not plug the deltas back in.

- **Best subsets regression**

First of all, the value $T_y = R_y - \alpha_i - \beta_i R_{my}$ needs to be computed for each company-month, where the market model described above is used to determine $\alpha_i$ and $\beta_i$ for each company $i = 1, \ldots, N$. Again, we do not shrink the coefficients. Then a Minitab best subsets regression ($T_y$ versus the vector of transformed predictors $x$) is carried out to calculate the estimates of deltas.

The benefits of choosing the models described above for comparison are:

1) We compare the shrunken regression approach with completely non-Bayesian procedures (conventional model, market model, best subsets regression).

2) The effect of ridge regression shrinkage can be demonstrated by comparing the estimates of delta obtained from ridge regression with the estimates obtained using conventional approach (or approach with the vague prior for $\delta$).
3) A Vasicek correction for betas can easily be demonstrated by comparing shrunken estimates of betas with the estimates obtained on using the conventional model.

4) With best subsets regression we have got the flexibility of including or excluding predictor variables with the object of finding the best model.

5.3 Results

We firstly consider the estimates obtained for deltas. Table 5.1 shows the results for the conventional method, approach with the vague prior for $\delta$, best subsets and shrunken (ridge) regression. November 2004 was chosen as the current time point; ten years' data back from this point were used to estimate regression coefficients for each of the methods.
<table>
<thead>
<tr>
<th>Name of predictor</th>
<th>Conventional model</th>
<th>Vague prior for delta</th>
<th>Ridge regression</th>
<th>Best subsets regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. monthly return</td>
<td>-6.23E-03</td>
<td>-5.92E-03</td>
<td>-5.84E-03</td>
<td>-7.91E-03</td>
</tr>
<tr>
<td>2. market return</td>
<td>6.26E-03</td>
<td>6.36E-03</td>
<td>6.32E-03</td>
<td>1.31E-02</td>
</tr>
<tr>
<td>3. share price</td>
<td>7.49E-03</td>
<td>6.65E-03</td>
<td>5.63E-03</td>
<td>5.01E-03</td>
</tr>
<tr>
<td>4. dividend</td>
<td>-1.24E-04</td>
<td>-1.27E-04</td>
<td>-1.19E-04</td>
<td>0</td>
</tr>
<tr>
<td>5. trading volume</td>
<td>1.11E-02</td>
<td>1.14E-02</td>
<td>1.05E-02</td>
<td>1.25E-02</td>
</tr>
<tr>
<td>6. market capitalization</td>
<td>-2.30E-02</td>
<td>-2.41E-02</td>
<td>-2.23E-02</td>
<td>-1.55E-02</td>
</tr>
<tr>
<td>7. 2-monthly return</td>
<td>-8.85E-04</td>
<td>-1.08E-03</td>
<td>-1.04E-03</td>
<td>0</td>
</tr>
<tr>
<td>8. 3-monthly return</td>
<td>3.47E-03</td>
<td>3.07E-03</td>
<td>2.96E-03</td>
<td>-3.48E-03</td>
</tr>
<tr>
<td>9. 4-monthly return</td>
<td>-5.14E-03</td>
<td>-4.84E-03</td>
<td>-4.69E-03</td>
<td>-3.69E-03</td>
</tr>
<tr>
<td>10. 6-monthly return</td>
<td>4.81E-03</td>
<td>4.87E-03</td>
<td>4.80E-03</td>
<td>6.43E-03</td>
</tr>
<tr>
<td>11. annual return</td>
<td>-5.14E-04</td>
<td>-1.92E-04</td>
<td>-2.42E-04</td>
<td>3.15E-03</td>
</tr>
<tr>
<td>12. 2 years return</td>
<td>-8.27E-04</td>
<td>-5.48E-04</td>
<td>-5.73E-04</td>
<td>-5.02E-03</td>
</tr>
<tr>
<td>13. 3 years return</td>
<td>-1.92E-03</td>
<td>-2.01E-03</td>
<td>-1.75E-03</td>
<td>0</td>
</tr>
<tr>
<td>14. 4 years return</td>
<td>8.26E-04</td>
<td>7.74E-04</td>
<td>9.68E-04</td>
<td>0</td>
</tr>
<tr>
<td>15. 5 years return</td>
<td>5.17E-03</td>
<td>4.53E-03</td>
<td>2.89E-03</td>
<td>1.24E-02</td>
</tr>
<tr>
<td>16. alpha</td>
<td>-1.14E-02</td>
<td>-1.07E-02</td>
<td>-9.38E-03</td>
<td>-1.60E-02</td>
</tr>
<tr>
<td>17. beta</td>
<td>-5.55E-04</td>
<td>-1.72E-03</td>
<td>-1.59E-03</td>
<td>-1.50E-03</td>
</tr>
<tr>
<td>18. specific risk</td>
<td>1.59E-02</td>
<td>1.49E-02</td>
<td>1.20E-02</td>
<td>2.69E-02</td>
</tr>
<tr>
<td>19. variance of beta</td>
<td>-1.16E-02</td>
<td>-1.08E-02</td>
<td>-8.14E-03</td>
<td>-2.48E-02</td>
</tr>
<tr>
<td>20. dividend yield</td>
<td>3.88E-03</td>
<td>4.61E-03</td>
<td>4.72E-03</td>
<td>0</td>
</tr>
<tr>
<td>over last 2 years</td>
<td>9.03E-03</td>
<td>7.85E-03</td>
<td>6.30E-03</td>
<td>1.40E-02</td>
</tr>
<tr>
<td>21. dividend yield</td>
<td>-2.74E-03</td>
<td>-1.70E-03</td>
<td>-4.35E-04</td>
<td>-8.48E-03</td>
</tr>
<tr>
<td>over last 3 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Notice that the coefficients for the selected predictors from subset selection tend to be larger in magnitude than the conventional (full) model values. As Tibshirani (1996) pointed out, this is common with positively correlated predictors. However, the ridge regression shows the opposite effect, as it shrinks the coefficients from their full model values. At the same time, it can be seen that there is not much difference between conventional model estimates and the coefficients produced by the approach with the vague prior for $\mathbf{\delta}$ (see column 1 and 2 of Table 5.1). This is not surprising since both methods do not utilise ridge regression, or, to be more precise, there is no shrinkage for the elements of vector $\mathbf{\delta}$. Figure 5.1 shows the estimates of delta for different methods.

11 Figure 5.1. Estimates of $\delta_k$, $k = 1, ..., p$ according to the four methods
Ridge regression is a continuous process that shrinks the coefficients towards zero. We expect that ridge regression will scale the coefficients by a constant factor, the results, however, show that this is not always the case. The explanation for this effect comes from Tibshirani (1996, p. 272). He argues that "the form of ridge regression shrinkage depends on the correlation of the predictors". Ridge regression does proportional shrinkage if and only if the predictors are not correlated. Moreover, the signs of the ridge regression estimates can be different from those of the conventional model estimates. Still it can be seen that most of the coefficients are shrunken to zero by ridge regression.

Secondly, for the ridge regression model the parameters of the prior distributions (hyperparameters) were estimated as follows:

1. for the distribution of $\alpha$: $\bar{\alpha} = 1.0E-05, \sigma_\alpha = 1.0E+05$;
2. for the distribution of $\beta$: $\bar{\beta} = 0.8817, \sigma_\beta = 0.2651$;
3. for the distribution of $\sigma^2$: $\alpha = 0.0045, \beta = 0.8332$;
4. for the distribution of $\delta$: $\sigma_\delta = 0.0078$.

The results for the prior distribution of beta broadly agree with the prior parameters suggested by Vasicek (1973). As Vasicek points out, the choice of the parameters of the prior distribution greatly depends on the prior information available. The prior parameters might be approximately taken as $\bar{\beta} = 1$ and $\sigma_\beta = 0.5$ for the New York Stock Exchange.
The results for the approach with the vague prior for the elements of vector $\delta$ are not reported since the estimates are very close to the results presented above, with the only difference that $\sigma_\epsilon$ is not estimated.

Vasicek’s correction for betas can finally be demonstrated. Let us point out that, although the method described in the previous chapter does not utilise ridge regression for the betas, we still refer to this method as ridge regression. It helps to avoid confusing terminology since this approach is the only one, among the methods used in comparison, which utilises ridge regression for some parameters (namely for the elements of the vector $\delta$ only).

Since EB adjustments (but not ridge regression) are still utilised for the betas we expect betas for the individual companies to shrink towards their prior mean (as compared to the conventional model estimates). Table 5.2 presents the estimates of $\beta_i$ and $\sigma_i^2$ for the first 15 companies for the conventional model, the market model (where the term $\delta^T \mathbf{x}$ is ignored), and ridge regression. The results for the approach with the vague prior for $\delta$ are not provided since the estimates are very similar to the ones obtained from the ridge regression.
Table 5.2. Estimates of the parameters $\beta_i$, $\sigma_i^2$ for the conventional model, market model and ridge regression approach

<table>
<thead>
<tr>
<th>Company number</th>
<th>Conventional model</th>
<th>Market model</th>
<th>Ridge regression (with Vasicek's adjustment for beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_i$</td>
<td>$\sigma_i^2$</td>
<td>$\beta_i$</td>
</tr>
<tr>
<td>1.</td>
<td>0.556</td>
<td>9.92E-03</td>
<td>0.570</td>
</tr>
<tr>
<td>2.</td>
<td>0.908</td>
<td>3.36E-02</td>
<td>0.934</td>
</tr>
<tr>
<td>3.</td>
<td>0.561</td>
<td>3.61E-03</td>
<td>0.567</td>
</tr>
<tr>
<td>4.</td>
<td>1.511</td>
<td>3.88E-03</td>
<td>1.521</td>
</tr>
<tr>
<td>5.</td>
<td>1.301</td>
<td>8.74E-03</td>
<td>1.312</td>
</tr>
<tr>
<td>6.</td>
<td>1.748</td>
<td>7.56E-03</td>
<td>1.737</td>
</tr>
<tr>
<td>7.</td>
<td>0.980</td>
<td>1.39E-02</td>
<td>1.002</td>
</tr>
<tr>
<td>8.</td>
<td>0.705</td>
<td>2.73E-03</td>
<td>0.716</td>
</tr>
<tr>
<td>9.</td>
<td>0.570</td>
<td>6.26E-03</td>
<td>0.558</td>
</tr>
<tr>
<td>10.</td>
<td>0.662</td>
<td>9.38E-03</td>
<td>0.690</td>
</tr>
<tr>
<td>11.</td>
<td>1.539</td>
<td>3.34E-02</td>
<td>1.527</td>
</tr>
<tr>
<td>12.</td>
<td>0.938</td>
<td>7.09E-03</td>
<td>0.945</td>
</tr>
<tr>
<td>13.</td>
<td>2.443</td>
<td>7.71E-03</td>
<td>2.479</td>
</tr>
<tr>
<td>14.</td>
<td>1.346</td>
<td>7.99E-03</td>
<td>1.359</td>
</tr>
<tr>
<td>15.</td>
<td>0.569</td>
<td>4.16E-03</td>
<td>0.562</td>
</tr>
</tbody>
</table>

As we expected, ridge regression shrinks the betas towards the prior mean, which is approximately 0.88 (see columns 2 and 6 of the Table 5.2). The degree of the adjustment depends on the specific risk $\sigma_i$ for the individual company$^{10}$. The bigger specific risk, the

---

$^{10}$ Standard error on the coefficient beta is proportional to $\sigma_i$.
greater the shrinkage to the prior mean. Figure 5.2 shows the estimates of betas for the conventional model and the ridge regression.

![Figure 5.2. Estimates of beta for the conventional model and the ridge regression. The black bars correspond to the prior mean (0.88), which is the same for all the companies](image)

It can also be seen that there is only a tiny difference between the results for the conventional model and market model. This confirms the assumption made earlier that the inclusion of $\delta \neq 0$ in regression hardly changes $\beta$, and $\sigma$ (the predictor variables give a very poor prediction with $R^2 \approx 2\%$). The results just presented (for betas, sigmas, deltas and hyperparameters) are stationary over the time.

In previous sections and in the first part of this section we examined the ridge regression model. Now we must think about the performance of this method taking into account that better predictions are the primary purpose of any forecasting model. Pure theory and
sophisticated methods are of little practical value unless they can contribute to improving the accuracy of post-sample predictions. But how do we know that a forecasting method we introduced is performing adequately? To answer this question we have to compare the performance of the method we use with several rival methods. The absolute accuracy of the various methods is not as important as how well these methods perform relative to some benchmark. Different measures of forecasting accuracy could be used to make this comparison. Makridakis and Hibon (2000) described the five accuracy measures utilized in M3-Competition: symmetric MAPE, average ranking, percentage better, median symmetric APE, median RAE. To analyze the performance of various methods for our case we adopt some of the methods introduced by Makridakis plus some others. Before outlining the methodology of comparison it should be pointed out that the out of sample forecast of return for each company has been made for several time points (months). Therefore, since we know the true value of return of each company, for each month the forecasting error can be computed easily. The following methods were used to analyze the goodness of each model:

5.4 Assessing prediction performance: shrunken regression

- Percentage better

The percentage better measures counts and reports the percentage of time that a given method has a smaller absolute forecasting error than other methods. Each forecast made,
regardless of company and month, is given equal weight. Our comparison in Figure 5.3 uses conventional approach as the benchmark to present the percentage of time that this method does better or worse than the others. It can be seen that, although the use of ridge regression does give a slightly better performance according to this measure, the effect is small and may be not very stable. Table 5.3 shows the results for this measure of accuracy when different methods are used as a benchmark.

![Graph showing percentage comparison](image)

14 Figure 5.3 Percentage of time other methods perform better than conventional approach

15 Table 5.3. Percentage better: different methods are used as a benchmark

<table>
<thead>
<tr>
<th>Benchmark method</th>
<th>Conventional approach</th>
<th>Best subsets regression</th>
<th>Ridge regression</th>
<th>Vague prior for delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional approach</td>
<td>-</td>
<td>53.34%</td>
<td>54.47%</td>
<td>52.42%</td>
</tr>
<tr>
<td>Best subsets</td>
<td>46.66%</td>
<td>-</td>
<td>49.43%</td>
<td>49.23%</td>
</tr>
<tr>
<td>Ridge regression</td>
<td>45.53%</td>
<td>50.57%</td>
<td>-</td>
<td>46.87%</td>
</tr>
<tr>
<td>Vague prior for delta</td>
<td>47.58%</td>
<td>50.77%</td>
<td>53.13%</td>
<td>-</td>
</tr>
</tbody>
</table>
**Percentage best**

Percentage best is similar to percentage better with the only difference that this method counts the percentage of time that a given method has a smaller forecasting error than all the other methods that we have under consideration. Each forecast made is given equal weight. Figure 5.4 reports the results: ridge regression gives the best forecast in 13.5% of cases, conventional model in 27.3%, best subset regression in 45.0%, and vague prior for delta in 14.2%.

![Figure 5.4. Percentage best: percentage of time when a given method outperforms all other methods](image)

Therefore, as far as percentage better is concerned, the accuracy of the ridge regression is worse than that of any other method.
• **Percentage worst**

This method is analogous to the method described above, but now we measure the percentage of time that a given method has a bigger absolute forecasting error than all the other methods. The results are: conventional approach gives the worst forecast in 34.1%, best subset regression in 41.4%, ridge regression model in 8.5%, and vague prior for delta – in 16.0% of trials. It can be seen that when percentage worst is used as a measure of regression performance the results obtained are contrary to the ones obtained for percentage best. While best subsets regression gives the best prediction in 45% of trials it also gives the worst forecast in 41.4% of trials. This demonstrates that it is not enough to utilise a single measure while assessing regression performance.

• **Correct sign**

Another measure of accuracy that is useful for financial time series data (since the investor makes a decision based on the sign of the change in share price) is the percentage of times that the forecast predicts the correct sign. The methods involved predict correct signs in approximately 67% of trials, ridge regression performs slightly better than the rival approaches.
Mean squared error

MSE is found across the companies for each method and for each month we have done the forecast for. Then the average MSE is calculated for each method across the months. The results are presented in Table 5.4. In terms of MSE the ridge regression outperforms all other methods except for the best subsets regression.

<table>
<thead>
<tr>
<th>month number</th>
<th>Best subsets regression</th>
<th>Conventional approach</th>
<th>Ridge regression</th>
<th>Vague prior for delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 2004</td>
<td>2.857E-03</td>
<td>2.886E-03</td>
<td>2.912E-03</td>
<td>2.922E-03</td>
</tr>
<tr>
<td>November 2004</td>
<td>4.731E-03</td>
<td>4.902E-03</td>
<td>4.675E-03</td>
<td>4.672E-03</td>
</tr>
<tr>
<td>October 2004</td>
<td>4.050E-03</td>
<td>4.154E-03</td>
<td>4.196E-03</td>
<td>4.207E-03</td>
</tr>
<tr>
<td>September 2004</td>
<td>2.943E-03</td>
<td>2.901E-03</td>
<td>2.910E-03</td>
<td>2.914E-03</td>
</tr>
<tr>
<td>Average across the months</td>
<td>3.645E-03</td>
<td>3.711E-03</td>
<td>3.674E-03</td>
<td>3.679E-03</td>
</tr>
</tbody>
</table>

Figure 5.5 shows the differences in MSE between the ridge regression and other methods involved for comparison.
Median squared error

Median SE is found (across the companies) for each method and for each month we have done the forecast for. Then the average Median SE is calculated for each method across the months. Such a measure is not influenced by extreme values and is more robust than MSE. The results are given in Table 5.5.

<table>
<thead>
<tr>
<th>month number</th>
<th>Best subsets regression</th>
<th>Conventional approach</th>
<th>Ridge regression</th>
<th>Vague prior for delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>November 2004</td>
<td>1.342E-03</td>
<td>1.543E-03</td>
<td>1.461E-03</td>
<td>1.430E-03</td>
</tr>
<tr>
<td>October 2004</td>
<td>1.491E-03</td>
<td>1.624E-03</td>
<td>1.779E-03</td>
<td>1.798E-03</td>
</tr>
<tr>
<td>Average across the months</td>
<td>1.161E-03</td>
<td>1.233E-03</td>
<td>1.270E-03</td>
<td>1.278E-03</td>
</tr>
</tbody>
</table>

Average ranking

For each company the average rankings are computed by sorting the median SE (calculated for each company across the months) of each method from the smallest (taking the value of 1) to the largest. Consequently, once the ranks for all the companies have been determined, the mean rank is calculated over all series (companies). The forecasting method with the smallest average rank is regarded as the best method. The ranks are: 2.42 for the ridge regression, 2.65 for the conventional model, 2.42 for the best subsets, and 2.51 for vague prior for delta approach.
References


CHAPTER 6

Predicting stock returns and assessing prediction performance
6.1 Introduction

In chapter 4 of this thesis we developed a new approach which utilises Bayesian decision theory to model return on a stock. Later, in chapter 5, we outlined the results obtained for the proposed methodology and made an evaluation of regression predictions delivered by our model as compared to several conventional (rival) techniques. Although it is well-known that using the Bayesian approach can substantially improve maximum-likelihood regression predictions, the comparisons have shown that it does not seem feasible to obtain any great improvement over the existing methods. A possible explanation would be that, with the large sample size available, we expect Bayesian, Empirical Bayes and likelihood-based frequentist approaches to behave similarly. That does not however detract from the work that has been done. A great deal of the complexity of the model considered is imposed by the fact that deltas and other parameters such as betas are computed simultaneously. The original contribution here consists in obtaining the evidence that the inclusion of $\delta \neq 0$ hardly changes $\beta_i$ and $\sigma_i^2$. Due to this fact it is possible to greatly simplify the computations by estimating $\alpha_i$, $\beta_i$, and $\sigma_i^2$ separately from $\delta$. This simplification allows us to explore many more modelling possibilities for which conflicting claims have been made in the literature.

Another problem is that measures of forecast accuracy employed earlier in section 5.4 have a drawback of not being closely related to the financial cost of making a bad prediction. Although we do not want to address the full problem of portfolio construction
for an investor, the only meaningful way to compare the predictions for the various models is to set up a portfolio based on the model predictions. One can then use the methodologies available for investment appraisal to assess the different regression models via the performance of ‘optimal’ portfolios based on them. This is the subject of section 6.4.

This portfolio is slightly an idealised one, in that transaction costs are ignored, and shares can be bought and sold in any amount which is required, ignoring their discrete nature. As the portfolio weights can be negative, it is also assumed that shares can be short-sold. In fact, this ‘frictionless’ world is not very idealised, and the modern stock market allows short selling even to individuals who can, e.g. indulge spread betting over the internet. Here one can mimic the process of buying or short-selling shares even in small numbers, and transaction costs are extremely low.

The chapter is organised as follows. In section 6.2 we outline the methodology. In section 6.3 we show the outcomes of the correlation analysis. In section 6.4 existing measures of regression performance are reviewed and a new statistic which is a function of the investment gain is proposed. Section 6.5 and 6.6 compares the results for different model choices including shrunken regression approach.
6.2 Methodology

Turning back to the single-index model discussed in section 2.5 we note that parameters $\alpha$ and $\beta$ appearing in equation (2.5) depend on the financial ‘fundamentals’ of a company, such as dividend yield, and on statistics of its previous performance, such as annual return. One way to improve the prediction of returns is to make $\alpha_i$, the excess return of a company’s stock above what we would expect from its gearing to the market return, into a function of $p-1$ predictors, so that now

$$ R_y = \sum_{k=1}^{p-1} \delta_k x_{yk} + \beta_i R_{my} + \varepsilon_y, $$

where the $x_{yk}$ are predictor variables, and the $\delta_k$ regression coefficients. One of the $x_{yk}$ is unity so that there is still an actual intercept term. However, predictor variables can be products of predictors such as market capitalisation with market return $R_{my}$. If there are $t$ such products, we can write the model as

$$ R_y = \sum_{k=1}^{p-1} \delta_k x_{yk} + \sum_{t=1}^{f} \gamma_t q_{yt} R_{my} + \varepsilon_y $$

(6.1)

where the predictors are written as $q_{yt} R_{my}$ and the $\gamma_t$ are regression coefficients. There are now $p$ predictors, because one of the $q_{yt}$ is $\hat{\beta}_i$, the most recent estimate of the ‘beta’ for company $i$ known at month $j-1$. Besides giving a better estimate of $\alpha_i$, the regression on history variables such as dividend yields and previous long-term returns, and fundamentals such as market capitalisation for the $i$th company also gives a better
estimate of the company's beta. So that if \( q_{t1} = \hat{\beta}_{i,t-1} \), the most recent EB estimate of beta, the effective value of beta becomes

\[
\hat{\beta}_{i,\text{eff}} = \hat{\gamma}_1 \hat{\beta}_{i,t-1} + \sum_{t=2}^{i} \hat{\gamma}_t q_{t1}.
\]

The regression scheme adopted was to carry out a rolling 5-year regression of continuously-compounded returns for each company against the market return, yielding estimates of \( \alpha_i, \beta_i, \) and \( \sigma_i^2 \), to apply EB correction to the estimates of \( \beta_i \) and \( \sigma_i^2 \), and finally to carry out a rolling 10-year regression for monthly returns that included all companies in the LSPD and had the form (6.1). In equation (6.1) among the predictors of return are \( \alpha_i \) and \( \beta_i \), so that these are scaled by an element of \( \delta \) or \( \gamma \) before use.

This sort of regression for betas is often done (e.g. Rosenberg 1985). We have assumed that the beta \( \beta_i \) and the specific risk \( \sigma_i \) are firm-dependent, but that the dependence on predictor variables is common across all firms. Some variables such as share price and market capitalisation were divided by the retail price index to make them comparable across the 10 years of the data sample.

The inclusion of \( R_{mj} \), the market return for the month for which we are trying to predict returns, is unusual in forecasting. It adds nothing to the forecast for return, because we must replace \( R_{mj} \) by its expected value when making forecasts, but it allows \( V \) to be estimated.
This model was initially fitted by least-squares, with the option of fitting all shortlisted predictors, or by carrying out minimum-AIC or minimum BIC regression.

Recall that the Akaike information criterion (AIC) proposed in Akaike (1974), is a measure of the goodness of fit of an estimated statistical model. The AIC is the technique of compromising the complexity of the model against how well the model fits the data. The criterion to be minimised is given by

\[ AIC = 2k - 2\ln L, \]

where \( k \) is the number of parameters and \( L \) is the likelihood function. While the goodness of fit is improved by increasing the number of the parameters, AIC is an increasing function of the number of estimated parameters. The first term penalises the inclusion of new parameters and discourages overfitting.

The AIC has been criticised as still allowing models to have too many parameters. Schwarz (1978) adopted another statistical criterion for model selection. The Bayes information criterion (BIC), penalises free parameters more strongly than does the AIC. The preferred model is the one with the lowest BIC value. The BIC is

\[ BIC = -2\ln L + k \ln(n), \]

where \( n \) is the number of observations.

Needless to say, both the form of the model and the method of inference beg many questions, and very many ‘improvements’ are possible, some of which we explore. First, we justify the use of a model with this functional form.
6.3 Correlation analysis

A Spearman rank correlation analysis was performed. The Spearman correlation coefficient $\rho$ has the interpretation which is similar to that of the sample correlation coefficient but it was developed to measure the association between two variables when ordinal (rather than interval) data are used. Using this statistic can be beneficial because, unlike the Pearson correlation, it does not require the assumption that the relationship between the variables is linear. The following procedure is normally used to compute $\rho$.

The raw scores are transformed into ranks, and the differences $d$ between the ranks of each observation on the two variables are calculated, $\rho$ is then given by

$$
\rho = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}, \quad d_i = x_i - y_i
$$

where $n$ is the number of pairs of values, $x_i$ is the rank of item $i$ with respect to the variable and $y_i$ is the rank of item $i$ with respect to the second variable.

Table 6.1 shows average Spearman correlations between predictor variables and return. Correlations were computed using all the company-months of data in each calendar year, and combined by weighting each year's correlation by the sample size (because the variance of the correlation is approximately the inverse of sample size). Table 6.1 also shows the stability of correlations with predictor variables with return over time. If the ratio of average correlation to average modulus of correlation is low, the correlation changes sign often from year to year.
Table 6.1 Average annual Spearman correlation coefficients between predictor variables and return, from 1956 to the end of 2005. ‘Winter’ is unity from October through to May inclusive, otherwise zero. The ratio of average correlation to average absolute value of correlation is an indicator of the stability of the predictor.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Average correlation</th>
<th>Average modulus of correlation</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Next months market return</td>
<td>0.392477</td>
<td>0.3925</td>
<td>1.0000</td>
</tr>
<tr>
<td>monthly market return</td>
<td>0.039768</td>
<td>0.1162</td>
<td>0.3421</td>
</tr>
<tr>
<td>monthly return</td>
<td>-0.008145</td>
<td>0.0662</td>
<td>0.1231</td>
</tr>
<tr>
<td>2-monthly return</td>
<td>-0.033745</td>
<td>0.0619</td>
<td>0.5453</td>
</tr>
<tr>
<td>3-monthly return</td>
<td>-0.038979</td>
<td>0.0575</td>
<td>0.6781</td>
</tr>
<tr>
<td>4-monthly return</td>
<td>-0.036260</td>
<td>0.0567</td>
<td>0.6399</td>
</tr>
<tr>
<td>6 monthly return</td>
<td>-0.042449</td>
<td>0.0707</td>
<td>0.6003</td>
</tr>
<tr>
<td>annual return</td>
<td>-0.012069</td>
<td>0.0534</td>
<td>0.2259</td>
</tr>
<tr>
<td>2 years return</td>
<td>0.001201</td>
<td>0.0538</td>
<td>0.0223</td>
</tr>
<tr>
<td>return 6-24 months ago</td>
<td>0.030857</td>
<td>0.0556</td>
<td>0.5552</td>
</tr>
<tr>
<td>3 years return</td>
<td>-0.011527</td>
<td>0.0441</td>
<td>0.2616</td>
</tr>
<tr>
<td>4 years return</td>
<td>-0.017301</td>
<td>0.0423</td>
<td>0.4093</td>
</tr>
<tr>
<td>5 years return</td>
<td>-0.018558</td>
<td>0.0427</td>
<td>0.4349</td>
</tr>
<tr>
<td>share price in pounds</td>
<td>0.000261</td>
<td>0.0334</td>
<td>0.0078</td>
</tr>
<tr>
<td>dividends</td>
<td>-0.042095</td>
<td>0.0803</td>
<td>0.5244</td>
</tr>
<tr>
<td>P/E ratio (from 1979)</td>
<td>0.006731</td>
<td>0.0279</td>
<td>0.2411</td>
</tr>
<tr>
<td>monthly trading volume (from 1990)</td>
<td>0.009956</td>
<td>0.0226</td>
<td>0.4401</td>
</tr>
<tr>
<td>market capitalisation</td>
<td>-0.000897</td>
<td>0.0325</td>
<td>0.0276</td>
</tr>
<tr>
<td>alpha</td>
<td>-0.001674</td>
<td>0.0288</td>
<td>0.0581</td>
</tr>
<tr>
<td>Beta</td>
<td>-0.007624</td>
<td>0.0287</td>
<td>0.2658</td>
</tr>
<tr>
<td>specific risk</td>
<td>-0.023701</td>
<td>0.0338</td>
<td>0.7004</td>
</tr>
<tr>
<td>annual dividend yield</td>
<td>0.032732</td>
<td>0.0501</td>
<td>0.6532</td>
</tr>
<tr>
<td>2 yearly dividend yield</td>
<td>0.029587</td>
<td>0.0507</td>
<td>0.5834</td>
</tr>
<tr>
<td>3 yearly dividend yield</td>
<td>0.029025</td>
<td>0.0511</td>
<td>0.5684</td>
</tr>
<tr>
<td>Winter</td>
<td>0.092197</td>
<td>0.0922</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Figure 6.1 shows yearly Spearman correlations for dividend yield and 5-year return reversal. These effects are consistently present, but are variable.

Table 6.2 shows typical values of the regression coefficients for the various predictors, which have been standardised to unit standard deviation and zero mean. In fact, in Table 6.2 the predictor variables have also been transformed to near normality. Table 6.2 is broadly consistent with table 6.1 and both of them are consistent with Table 5.1, which shows regression coefficients for shrunken regression introduced in chapter 5.
Table 6.2: Regression coefficients for minimum-BIC rolling regression to the end of 2005. $R_m$ denotes market return for the month being forecast. Variables have been transformed to be approximately standard normal. The last column shows the z-value for inclusion of the variable.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Variable</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.1035E-02</td>
<td>intercept</td>
<td>1.0268</td>
</tr>
<tr>
<td>+0.7043E-02</td>
<td>monthly return on the market</td>
<td>9.2326</td>
</tr>
<tr>
<td>+0.5407E-02</td>
<td>6 monthly return</td>
<td>6.7236</td>
</tr>
<tr>
<td>-0.3680E-02</td>
<td>3 years return</td>
<td>4.1471</td>
</tr>
<tr>
<td>+0.3765E-02</td>
<td>3 yearly dividend yield</td>
<td>4.3917</td>
</tr>
<tr>
<td>+0.4230E-02</td>
<td>Winter</td>
<td>5.5235</td>
</tr>
<tr>
<td>+0.3145E-02</td>
<td>3-monthly return$^2$</td>
<td>5.3902</td>
</tr>
<tr>
<td>-0.5968E-02</td>
<td>6 monthly return$^2$</td>
<td>7.8740</td>
</tr>
<tr>
<td>+0.3167E-02</td>
<td>annual dividend yield$^2$</td>
<td>4.0347</td>
</tr>
<tr>
<td>-0.9806E-02</td>
<td>2 yearly dividend yield$^2$</td>
<td>5.5983</td>
</tr>
<tr>
<td>+0.8128E-02</td>
<td>3 yearly dividend yield$^2$</td>
<td>6.0692</td>
</tr>
<tr>
<td>+0.9862E-01</td>
<td>monthly trading volume* $R_m$</td>
<td>5.3173</td>
</tr>
<tr>
<td>-0.9674E-01</td>
<td>2-monthly return* $R_m$</td>
<td>4.8004</td>
</tr>
<tr>
<td>-0.1769</td>
<td>annual return* $R_m$</td>
<td>7.4635</td>
</tr>
<tr>
<td>+0.1402</td>
<td>3 years return* $R_m$</td>
<td>5.7839</td>
</tr>
<tr>
<td>-0.9933E-01</td>
<td>$R_m$</td>
<td>4.8775</td>
</tr>
<tr>
<td>+0.1655</td>
<td>specific risk* $R_m$</td>
<td>7.9236</td>
</tr>
<tr>
<td>-0.9701E-01</td>
<td>annual dividend yield* $R_m$</td>
<td>4.8763</td>
</tr>
<tr>
<td>-0.1270</td>
<td>Winter* $R_m$</td>
<td>6.5859</td>
</tr>
<tr>
<td>+0.9518</td>
<td>coef of beta* $R_m$</td>
<td>45.5195</td>
</tr>
</tbody>
</table>
The stock market anomalies were described in section 9 of chapter 2. The signs and magnitudes of the coefficients in Table 6.1 agree broadly with what is known about determinants of stock return (e.g., Arnold, 2005). Empirical results we obtained confirms

a) that dividend yield is a good and stable predictor of high return;

b) there is the characteristic pattern of oscillation in returns: a short oscillation over a few months, where high returns quickly reverse;

c) presence of the momentum effect, which is a yearly effect where a high return over the last year sometimes predicts higher future return;

d) presence of the return reversal effect, whereby companies with low (high) returns over the last 4 or 5 years subsequently outperform winners (underperform losers). Unlike the momentum effect this effect is quite stable.

e) presence of so-called ‘size-effect’, where small firms often deliver higher returns. There is a small negative correlation between market capitalisation and return, even when we restrict ourselves to FTSE350 companies.

The LSPD does not contain book to market ratio, often quoted as a good predictor of return, but the other accounting variable often cited, price to earnings ratio (e.g. Basu, 1977), has poor predictive ability in our data range.
Variables such as 'winter' reflect the return experienced during winter months. This, although the correlation is high and stable, is not very useful, because it does not differentiate between stocks. It serves to tilt the portfolio towards buying stocks in winter months and selling in summer. Share price is not thought to be a useful predictor, and is included to show that poor predictive variables are indeed characterised by low correlations and low stability.

6.4 Metrics

There are many measures of prediction accuracy used to compare forecasting methods out-of-sample, e.g. those used in forecasting competitions. In this case, predictors have very poor predictive power, $R^2$ is about 2% at the best. Nevertheless, portfolios chosen in accordance with regression predictions can outperform the market, and predicting return is worthwhile. In measuring regression performance, it seems reasonable therefore to use some function of the investment gain as a statistic.

If we assume that the covariance matrix of returns $V$ is approximately known, we can measure the monetary loss incurred through imperfect predictions as follows.

The gain from investing monies $x$, in the $i$th of $N$ stocks is $r^T x$, where $r$ is a vector of excess returns over the risk-free rate. The Markowitz strategy consists in maximising the ratio of expected gain to risk, $E(r)^T x / (x^T V x)^{1/2}$. The solution is $x \propto V^{-1} E(r)$.

149
This solution also follows from an attempt to maximise the expected utility of an investment, when returns are assumed normally distributed (see section 2.2).

Imagining that the expected arithmetic returns $\hat{\mathbf{r}}$ predicted by the regression are to be used in this investment strategy, we would invest monies $\mathbf{x} \propto \mathbf{V}^{-1}\hat{\mathbf{r}}$ with gain $\mathbf{r}_0^T \mathbf{V}^{-1}\hat{\mathbf{r}}$, whereas a perfect regression model that could miraculously deliver the observed returns $\mathbf{r}_0$ in the guise of expected returns would lead to an investment $\mathbf{x} \propto \mathbf{V}^{-1}\mathbf{r}_0$. The drop in gain to risk ratio resulting from suboptimal investment is

$$D = \left( \mathbf{r}_0^T \mathbf{V}^{-1}\mathbf{r}_0 \right)^{1/2} - \frac{\mathbf{r}_0^T \mathbf{V}^{-1}\hat{\mathbf{r}}}{\left( \hat{\mathbf{r}}^T \mathbf{V}^{-1}\hat{\mathbf{r}} \right)^{1/2}}.$$  

It is simplest to use $PD$, the proportional drop in gain per risk, as a measure of loss resulting from inaccurate regression prediction. We have that

$$PD = 1 - \frac{\mathbf{r}_0^T \mathbf{V}^{-1}\hat{\mathbf{r}}}{\left( \mathbf{r}_0^T \mathbf{V}^{-1}\mathbf{r}_0 \right)^{1/2} \left( \hat{\mathbf{r}}^T \mathbf{V}^{-1}\hat{\mathbf{r}} \right)^{1/2}}. \quad (6.2)$$

From the definition of scalar products of two vectors the proportional loss function is $1 - \cos \theta$, where $\theta$ is an angle between the vectors $\mathbf{V}^{-1/2}\mathbf{r}_0$ and $\mathbf{V}^{-1/2}\hat{\mathbf{r}}$. This is a new statistic.
To evaluate $PD$, the (estimated) $V^{-1}$ is needed. The single-index model formula for $V_{ij}$ in (2.6) can be inverted analytically.

$$V_{ij}^{-1} = \delta_{ij} / \sigma_i^2 - \kappa \beta_i \beta_j / (\sigma_i^2 \sigma_j^2), \quad (6.3)$$

where

$$\kappa = \sigma_m^2 \left( 1 + \sigma_m^2 \sum_{j=1}^{N} \beta_j^2 / \sigma_j^2 \right). \quad (6.4)$$

Utilising formula (6.3) the components of $PD$, $r_0^T V^{-1} r_0$, $r_0^T V^{-1} r_0$ and $r V^{-1} r$ can be computed as

$$r_0^T V^{-1} r_0 = \sum_{i=1}^{N} \frac{R_{0i}^2}{\sigma_i^2} - \sigma_m^2 \left( 1 + \sigma_m^2 \sum_{i=1}^{N} \frac{\beta_i^2}{\sigma_i^2} \right) \left( \sum_{i=1}^{N} \frac{\beta_i R_{0i}^2}{\sigma_i^2} \right)^2, \quad (6.5)$$

$$\hat{r}^T V^{-1} \hat{r} = \sum_{i=1}^{N} \frac{\hat{R}_i^2}{\sigma_i^2} - \sigma_m^2 \left( 1 + \sigma_m^2 \sum_{i=1}^{N} \frac{\hat{\beta}_i^2}{\sigma_i^2} \right) \left( \sum_{i=1}^{N} \frac{\beta_i \hat{R}_i^2}{\sigma_i^2} \right)^2, \quad (6.6)$$

and

$$r_0^T V^{-1} \hat{r} = \sum_{i=1}^{N} \frac{R_{0i} \hat{R}_i}{\sigma_i^2} - \sigma_m^2 \left( 1 + \sigma_m^2 \sum_{i=1}^{N} \frac{\hat{\beta}_i^2}{\sigma_i^2} \right) \left( \sum_{i=1}^{N} \frac{\beta_i R_{0i}}{\sigma_i^2} \right) \left( \sum_{i=1}^{N} \frac{\beta_i \hat{R}_i}{\sigma_i^2} \right). \quad (6.7)$$

It is troublesome however to evaluate $PD$ by using the equations given above. The authors faced the problem that the value of this statistic is not the same when Microsoft Excel or, alternatively, fortran95 are involved for the computations. The difficulty is caused by the rounding errors which arise when fortran95 is used. Employing Microsoft Excel as a tool removes the difficulty since it stores and calculates with 15 significant
digits of precision. Using Excel is tempting, but fortran programming gives us more flexibility since, for example, we do not need to convert a program output file to the format accepted by Excel and import the data to Excel spreadsheet. To reduce the rounding error we propose the following approach.

Let us write \( R_{oi} = p\beta_i + \delta_i \) and \( \hat{R}_i = q\beta_i + \xi_i \). We choose \( p \) and \( q \) so that

\[
\sum_j \left( R_{oj} - p\beta_j \right) \beta_j / \sigma_j^2 = 0 \quad \text{and} \quad \sum_j \left( \hat{R}_j - q\beta_j \right) \beta_j / \sigma_j^2 = 0.
\]

From the equations above

\[
p = \frac{\sum_j R_{oj} \beta_j / \sigma_j^2}{\sum_j \beta_j^2 / \sigma_j^2} \quad \text{and} \quad q = \frac{\sum_j \hat{R}_j \beta_j / \sigma_j^2}{\sum_j \beta_j^2 / \sigma_j^2}.
\]

Utilising the above notation three components of \( PD \) can be written as (see Appendix 11)

\[
r_0^T V^{-1} r_0 = \frac{\left( \sum_{i=1}^N R_{oi} \beta_i / \sigma_i^2 \right)^2}{\left( \sum_{i=1}^N \beta_i^2 / \sigma_i^2 \right) \left( 1 + \sigma_m^2 \sum_{i=1}^N \beta_i^2 / \sigma_i^2 \right)} + \sum_{i=1}^N \frac{\delta_i^2}{\sigma_i^2}, \quad (6.8)
\]

\[
\hat{r}_0^T V^{-1} \hat{r}_0 = \frac{\left( \sum_{i=1}^N \hat{R}_i \beta_i / \sigma_i^2 \right)^2}{\left( \sum_{i=1}^N \beta_i^2 / \sigma_i^2 \right) \left( 1 + \sigma_m^2 \sum_{i=1}^N \beta_i^2 / \sigma_i^2 \right)} + \sum_{i=1}^N \frac{\xi_i^2}{\sigma_i^2}, \quad (6.9)
\]

and

\[
r_0^T V^{-1} \hat{r}_0 = \frac{\left( \sum_{i=1}^N R_{oi} \beta_i / \sigma_i^2 \right) \left( \sum_{i=1}^N \hat{R}_i \beta_i / \sigma_i^2 \right)}{\left( \sum_{i=1}^N \beta_i^2 / \sigma_i^2 \right) \left( 1 + \sigma_m^2 \sum_{i=1}^N \beta_i^2 / \sigma_i^2 \right)} + \sum_{i=1}^N \frac{\delta_i \xi_i}{\sigma_i^2}. \quad (6.10)
\]
Formulas (6.8) – (6.10) - to ameliorate but do not remove the rounding errors problem. Still it is safer to evaluate the terms in (6.2) by solving the linear equations numerically.

$PD$ gives a measure of financial loss resulting from wrong regression predictions on the scale from 0 to 2, where the maximum loss results from predicting all returns correctly but with the wrong sign. When $PD \geq 1$, the gain from the investment is zero or negative relative to saving the money at the risk-free rate, so basing investment on regression predictions would be useless. The measure is calculated on a monthly basis, but may be cumulated over the several investment periods by summing its three components, $r_0^T V^{-1} \hat{f}$, $r_0^T V^{-1} r_0$ and $\hat{f} V^{-1} \hat{f}$. Typically, cumulated $PD$ was approximately 0.97 for the regressions in this study.

Although $PD$ is useful for evaluating predictions using only a few out-of-sample periods, it ignores errors in estimating the covariance matrix, and when there are sufficient out-of-sample data statistics even more closely related to investment performance can be used. There is an initial problem of normalising investment sizes. Equation $x \propto V^{-1} r$ gives optimal portfolios $x$ for each month, where $V$ is now taken as the monthly covariance matrix of returns, if the successive monthly returns are uncorrelated so that the full covariance matrix is block diagonal. However, empirically a more constant gain $Z$ is obtained by standardising the investment size by the predicted risk, so that

$$Z = \frac{r_0^T V^{-1} \hat{f}}{(\hat{f}^T V^{-1} \hat{f})^{1/2}}.$$
When there are many investment periods, the Sharpe measure (Sharpe, 1966) is a commonly used statistic for evaluating investment performance. This is the ratio of average excess return over the risk-free rate to its risk $s$ (sample standard deviation). This measure is often referred to as an excess return to variability measure. A modified version of this is quoted in Table 6.3. When, as here, short selling is allowed (investment can be negative as well as positive) the concept of return becomes problematical. For instance, initial outlay might be zero. The Sharpe measure was therefore adapted by replacing return by the excess gain on an investment that had been standardised to have unit predicted risk, so that $S = \frac{\bar{z}}{s}$.

The Sharpe measure $S$, however, has one unfortunate property as a measure of investment performance and a criticism of this statistic has been raised in financial literature:

"Clearly, due to asymmetric distributions the Sharpe ratio is not necessarily an adequate statistic to guide investors" (Annaert et al., 2006, p. 24).

Use of the excess return to variability ratio may lead to deceptive conclusions, since if one of the investment gains were to be drastically increased, the measure could decrease, because the denominator would increase faster than the numerator. Aumann and Serrano (2006) provide an example to demonstrate that using standard deviation to measure risk has some drawbacks. They consider two gambles: the first gamble guarantees the gain of 1 dollar while the second offers a 50% chance to receive 3 dollars or 1 dollar otherwise. Clearly, any investor would choose the second one because it maximises his/her expected
utility. However, since the second gamble has higher variance, the Sharpe measure is lower. Thus an investment strategy that would surely be preferred when viewed in retrospect would appear inferior. Formally, as Annaert et al (2006) points out, the Sharpe measure does not respect first-order stochastic dominance.

The concept of stochastic dominance is utilised in the situations where the decision needs to be made about the preferences regarding outcomes of competing methodologies. First-order stochastic dominance is a canonical case of stochastic dominance, defined as follows: the random variable $X$ first-order stochastically dominates the random variable $Y$ iff $\Pr[X \geq a] \geq \Pr[Y \geq a]$ for all $a$. In other words, if the CDFs of $X$ and $Y$ are given by $F(a)$ and $G(a)$, respectively, then $X$ first-order stochastically dominates $Y$ if $F(a) \leq G(a)$ for all $a$. To demonstrate the drawback of the Sharpe ratio consider an example. Historical returns obtained from utilising investment strategy A are given by $<1,4,7,3>$, while the outcomes from investing according to strategy B are $<1,4,7,100>$. If an investor preferred more to less, then investment strategy B is preferable to investment strategy A, and, according to the definition, strategy A is first-order stochastically dominated by strategy B. However, strategy B gives a lower Sharpe ratio (of only 0.67) as compared to strategy A which yields 1.73. This is of course is the case also when carrying out significance tests using a $z$ or $t$ statistic, for example when comparing two drugs. The assumption of normality there resolves the paradox of how a drug that performs better for all patients could have the lower $z$ score. Here however we do not expect normality, and this property is a drawback of this measure.
A solution adopted in the investment literature has been to replace the risk by 'downside risk'. While the Sharpe measure takes into consideration any volatility associated with return on a share, its extension, which is often referred to as Sortino ratio, differentiates volatility due to up and down movements. There is no penalty for the upside volatility, since up movements are considered advantageous. The standard deviation is replaced by the square root of the 'below target semivariance', so that

$$S' = \frac{\bar{Z}}{\sqrt{\sum_{i=1}^{n} (\text{Max}(T - z_i, 0))^2 / n}},$$

where $n$ is the number of investment gains, and $T$ a target, here taken as zero (Markowitz, 1959). For recent work on this see Sortino and Forsey (1996).

The mean-target semivariance model presented above is a special case of the $\alpha - t$ model introduced by Fishburn (1977). According to Fishburn (1977, p. 116) the $\alpha - t$ model is "a mean-risk dominance model in which risk is measured by a probability-weighted function of deviations below a specified target return $T$" and defined by the two-parameter function

$$F_\alpha(T) = \int_{-\infty}^{T} (T - z)^\alpha dF(z), \quad \alpha > 0,$$

where $F(z)$ is the probability of getting a return not exceeding $z$.

It can be seen that when $\alpha = 2$ equation (6.12) becomes the downside semi-standard deviation which is used as a measure of risk in equation (6.11).
Fishburn (1977, p. 116) argues that “there is no compelling a priori reason for taking $\alpha = 2$”, and “different values of $\alpha$ can approximate a wide variety of attitudes towards the risk of falling below the target return”. $\alpha$ in equation (6.12) can be interpreted as a “measure of the relative impact of large and small deviations” below the target. Assuming a given value of parameter $T$, $\alpha$ indicates the feelings of the decision maker of failing to achieve the threshold return $T$. The small value of $\alpha$ should be chosen when attaining the specified target has the primary importance and the amount of loss when failure occurs does not have such an importance. Contrarily, the larger values of $\alpha$ are appropriate when the amount of loss when falling below the target is the main concern of the decision maker.

The Theorem formulated by Fishburn (1977, p. 119) says that “$\alpha = 1$ is the point which separates risk-seeking from risk-averse behaviour with regard to returns below target, if $\alpha < 1$ then the risky option dominates the sure-thing, and if $\alpha > 1$ then the sure-thing dominates the risky option”. The methods for estimating $\alpha$ are presented in the paper.

The central point of the work done by Fishburn is the idea that performance measures based on downside risk are congruent with the utility functions. However he did not derive $S'$ from the corresponding utility function. Doing so is instructive.
For one-parameter utility functions, the parameter (say $\eta$) is a measure of risk aversion. The better the performance of the investment, the more risk aversion an investor requires to make it valueless. We therefore propose $\eta_0$, the value of $\eta$ that makes the certainty-equivalent gain zero, as a measure of investment performance. Note that to obtain a dimensionless measure of performance that does not depend on the amount invested, the risk aversion parameter must often change with investment size. For example, writing $\eta = \xi^a \mu^{1-a}$, where $\alpha > 0$, $\mu$ is the mean gain and $\xi$ the dimensionless risk aversion parameter, we can obtain (6.11) from utility

$$u(z) = \begin{cases} 
  z & \text{for all } z \geq T \\
  z - \xi^a \mu^{1-a} (T - z)^a & \text{for all } z \leq T
\end{cases}.$$ 

Equating the sample mean utility $\bar{u}$ to zero and replacing $\mu$ by the sample mean $\bar{z}$ yields $\xi_0$ as

$$\xi_0 = \frac{\bar{z}}{\left(\sum_{i=1}^{n} \max(T - z_i, 0)^a / n\right)^{1/a}},$$

a general class of measures of which (6.11) is the most widely used.

Another performance measure, called Omega, was suggested by Keating and Shadwick (2002). It is defined by

$$\Omega = \frac{\int_1^\infty (1 - F(z))dz}{\int_{-\infty}^1 F(z)dz}.$$
The Omega ratio is widely used and claimed to have an advantage since "unlike other measures of performance, Omega was developed with the intention to take the entire return distribution into account" (Kazemi et al, 2007).

It can be seen that Omega is also congruent with the utility function given above when \( \alpha = 1 \). Setting the expected utility to the minimum acceptable gain \( T \) gives the metric \( \Omega = 1 \).

This analysis suggests that, besides calculating \( \xi_0 \), we could usefully assess investments by plotting \( \bar{u} \) or its certainty equivalent \( C(\xi) \) against \( \xi \). Besides displaying the performance measures as the abscissa, such plots also show mean gain as \( C(0) \). They could be a useful decision aid.

We here obtain another measure \( M \) also having the property \( \partial M / \partial z_1 > 0 \) for any investment gain \( z_1 \), by applying the same logic, but now to more conventional exponential (Pratt-Arrow) utility function \( u(z) = (1 - \exp(-\eta z))/\eta \). The certainty-equivalent gain is \( C(\eta) = -\log(\sum_{i=1}^n \exp(-\eta z_i)/n)/\eta \), \( \partial C / \partial \eta < 0 \) (Appendix 12). It can be shown that \( \partial \eta_0 / \partial z_1 > 0 \), so that this measure also has a desired property (Appendix 12). Defining \( \xi \) by \( \eta = 2\xi^2 / \mu \) and replacing \( \mu \) by the sample estimate \( \bar{z} \) we obtain a dimensionless measure as before. Figure 6.2 shows plots of \( C \) versus \( \xi \) for several regression methods. As \( \xi \to \infty \), \( C \to z_{(1)} \), the lowest gain.
Figure 6.2. Certainty-equivalent gains from different investment types as a function of risk aversion, using the exponential utility function. The 'no regression line' corresponds to setting $\delta = 0$ in (6.1). This type of plot could be a useful summary of investment performance.

This measure $\xi_0$ must be computed by solving $C(\eta_0) = 0$, which reduces to the nonlinear equation

$$\frac{1}{n} \sum_{i=1}^{n} \exp\left(-2\xi_0^2 z_i / \bar{z}\right) = 1,$$

and does not have such a simple form as (6.11). However, we can explore its properties. Since $C(\eta)$ is closely related to the cumulant generating functions (this would be $\eta C(-\eta)$), we have that $C = \mu - \sigma^2 \eta / 2 + \sigma^3 \gamma \eta^2 / 6 - \sigma^4 \kappa \eta^3 / 24 + \ldots$, where
$\mu$, $\sigma^2$, $\gamma$ and $\kappa$ are the mean, variance, skewness and kurtosis respectively. If the distribution of return were normal, we would have $\eta_0 = 2\mu/\sigma^2$.

When higher cumulants are zero, $\xi_0 = \mu/\sigma$, giving the Sharpe measure. When skewness is nonzero, to first order in skewness and kurtosis,

$$\xi_0 \approx \mu/\sigma \left\{ 1 + (\mu/\sigma)^3 \gamma/6 - (\mu/\sigma)^4 \kappa/12 \right\},$$

showing that positive skewness increases this measure of investment value while high kurtosis decreases it. This approach thus properly takes all the moments into account. The statistic $\xi_0$ from equation (6.13) is given in Table 6.3. It is generally lower than the Sharpe measure in our results, reflecting the high kurtosis of gains.

The drawback of below-target semivariance or our measure as measures of investment performance is that, unlike the Sharpe ratio, they are undefined if all gains are above target, or, in our case, are all positive. As $C \to z_{(1)}$ as $z \to \infty$, if $z_{(1)} > 0$, $C > 0$ always. No degree of risk aversion can then render the investment valueless. Sortino and Forsey (1996) address this problem by calculating the semivariance from a fitted distribution.

It can also be seen from equation (6.13) that the losses have a greater effect on the measure proposed $\xi_0$ than the gains (the resulting exponent is positive for the negative values of $z_i$, and negative for the gains). This, in fact, shows very typical investor's
behaviour when decision maker is much more concerned about potential losses rather than potential gains. Nevertheless the gains are still appreciated in our measure. This addresses the problem raised by Koekebakker and Zakamouline (2007).

6.5 Results: regression model

Table 6.3 shows the (monthly) Sharpe measure and other performance statistics for the market portfolio, an equally weighted market portfolio, and portfolios derived from regressions. If fluctuations in stochastic processes from one period to the next are independent volatility increases with the square root of the unit of time. Annual performance therefore is $\sqrt{12}$ times the figures given. Regressions were used to predict returns, and so to derive an ‘optimal’ portfolio. Such portfolios were constructed monthly from January 1976 to the end of 2005. Regression coefficients were used for one year out of sample, before being updated, while betas and specific risks were calculated using data up to the month for which returns were forecast. In Table 6.3, the covariance matrix of returns was calculated using the single-index model. The table shows results for various modelling choices. The largest companies only were included in the computation of regression coefficients, but the number of these could be varied. Predictor variables could be transformed to normality or not (see Chapter 3) and regressions could use 10 or 20 years’ data.
It can be seen that regression-based portfolios, with Sharpe measure of around 0.4, considerably outperform the market portfolio, with a Sharpe measure of 0.095 and the equally-weighted portfolio, which includes all firms in the LSE. This latter portfolio also does better than the market, because small firms tend to give higher returns than large ones (Banz, 1981).

24 Table 6.3: Out-of-sample performance of various regression methods, showing the average gain per unit estimated risk, its risk (as a standard deviation), the Sharpe ratio, a new statistic $\xi_0$ defined in (6.13), and the Sharpe ratio using below-target semivariance. Under 'comments', $n$ is the number of firms used in computing regression coefficients, and ‘trans’ denotes whether variables were transformed.

<table>
<thead>
<tr>
<th>Years</th>
<th>Type</th>
<th>Comments</th>
<th>Mean</th>
<th>Risk</th>
<th>Sharpe</th>
<th>$\xi_0$</th>
<th>Semi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>market</td>
<td></td>
<td>0.10</td>
<td>1.00</td>
<td>0.095</td>
<td>0.092</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>equally weighed market</td>
<td></td>
<td>0.21</td>
<td>1.05</td>
<td>0.197</td>
<td>0.184</td>
<td>0.185</td>
</tr>
<tr>
<td>10</td>
<td>full</td>
<td>trans, $n=350$</td>
<td>1.07</td>
<td>2.67</td>
<td>0.400</td>
<td>0.457</td>
<td>0.742</td>
</tr>
<tr>
<td>10</td>
<td>none</td>
<td>$n=1000$, just constant</td>
<td>0.31</td>
<td>2.72</td>
<td>0.112</td>
<td>0.139</td>
<td>0.245</td>
</tr>
<tr>
<td>10</td>
<td>AIC</td>
<td>trans, $n=350$, equally</td>
<td>1.09</td>
<td>2.90</td>
<td>0.377</td>
<td>0.438</td>
<td>0.669</td>
</tr>
<tr>
<td></td>
<td></td>
<td>weighted market</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>AIC</td>
<td>trans, $n=350$</td>
<td>0.99</td>
<td>2.25</td>
<td>0.440</td>
<td>0.436</td>
<td>0.693</td>
</tr>
<tr>
<td>10</td>
<td>BIC</td>
<td>trans, $n=350$</td>
<td>1.04</td>
<td>2.41</td>
<td>0.432</td>
<td>0.430</td>
<td>0.700</td>
</tr>
<tr>
<td>10</td>
<td>shrink</td>
<td>trans, $n=350$</td>
<td>1.07</td>
<td>2.86</td>
<td>0.375</td>
<td>0.442</td>
<td>0.724</td>
</tr>
<tr>
<td>10</td>
<td>AIC</td>
<td>$n=1000$</td>
<td>1.03</td>
<td>3.00</td>
<td>0.344</td>
<td>0.385</td>
<td>0.622</td>
</tr>
<tr>
<td>10</td>
<td>AIC</td>
<td>no trans, $n=1000$</td>
<td>1.31</td>
<td>4.19</td>
<td>0.312</td>
<td>0.419</td>
<td>0.674</td>
</tr>
<tr>
<td>10</td>
<td>AIC</td>
<td>no trans, $n=all$</td>
<td>1.30</td>
<td>2.99</td>
<td>0.436</td>
<td>0.461</td>
<td>0.739</td>
</tr>
<tr>
<td>10</td>
<td>AIC</td>
<td>trans, $n=all$</td>
<td>1.29</td>
<td>3.44</td>
<td>0.375</td>
<td>0.464</td>
<td>0.763</td>
</tr>
<tr>
<td>20</td>
<td>AIC</td>
<td>trans, $n=1000$</td>
<td>1.08</td>
<td>2.30</td>
<td>0.470</td>
<td>0.423</td>
<td>0.676</td>
</tr>
<tr>
<td>10</td>
<td>BIC</td>
<td>trans, $n=1000$</td>
<td>1.09</td>
<td>3.35</td>
<td>0.324</td>
<td>0.394</td>
<td>0.639</td>
</tr>
<tr>
<td>20</td>
<td>AIC</td>
<td>no trans, $n=350$</td>
<td>1.09</td>
<td>2.99</td>
<td>0.363</td>
<td>0.317</td>
<td>0.589</td>
</tr>
</tbody>
</table>
With the predictor variables used, there is a steadily linear downward trend in the performance of regression-based portfolios. The gain per unit expected risk decreases steadily with time (correlation $\rho = -0.13$, $p = 0.014$, Spearman correlation $\rho_s = -0.217$, $p < 0.001$). The Sharpe measure decreases to roughly a quarter of its 1976 value. The expected gain for unit risk also decreases, in fact this decreases faster ($\rho_s = -0.469$). These effects date from around 1975. A tentative explanation would be that, as investors become more knowledgeable about such determinants of high return as annual dividend yield, low market capitalisation, and return reversal, their actions are removing these mispricings and are driving the market to become more efficient. Although it is not our aim to speculate on this, it would be surprising if investors completely ignored such information. There is some evidence that they do not, but that learning takes time. Levy and Yagil (2005, p. 1200) investigated how the “magnitude of the mispricing changes with the additional knowledge investors gather from publication of relevant scientific articles”. Their study examines the mispricing phenomenon of Shell Transport and Royal Dutch stocks. This anomaly is one of the examples for the case of Siamese twin stocks which is well documented in the literature (e.g., Rosental and Young, (1990), Froot and Dabora, (1999)). It was demonstrated that the mispricing gradually disappeared with the publication of relevant mispricing articles. These findings imply that although “market participants are sometimes unaware of theoretical price relationships that should prevail, they react in the right direction when such mispricing is brought to their attention via the publication of relevant empirical studies”. However, as Levy and Yagil (2005) concluded, this anomaly violates the concept of market efficiency since the mispricing problem may not be rectified immediately (e.g. the Royal Dutch/Shell Transport price disparity disappeared after 9 years).
Various attempts to improve performance were made. There are number of modelling problems and choices:

1. There are very many possible predictors of return, but they are very poor. Should model choice be made by minimising criteria such as AIC, BIC, or by using a shrunken estimator?

2. The distributions of both firm-specific risk and market return are long-tailed.

3. Some of the predictor variables are very long tailed. Maybe they should be transformed before use to approximate normality.

4. There are alternative definitions of ‘market return’ that could be used in (6.1)

5. Choice of sampling frame: if investing only in FTSE350 companies, we can include many more firms of smaller market capitalisation in the data sample used to estimate regression coefficients. Will doing this improve predictions?

The next sections address some of these more complex modelling approaches. When predicting returns for FTSE350 companies, one can include data from smaller market capitalisation companies to estimate regression coefficients. Although smaller firms may behave differently from large ones, and introduce bias, their inclusion would reduce statistical error. Table 6.3 suggests that, if anything, their inclusion serves only to slightly worsen performance measures.

Regarding the duration of the rolling regression period in (6.1), using 5 years’ data gave somewhat worse performance than using 10 years’ data, and increasing to 20 years if anything slightly worsened performance. Hence using 10 years’ data seems optimal. Unlike the methodology presented in chapter 4 in this approach, betas are calculated *en*
route to regression predictions, and we have the freedom to define betas with respect to any measure of the market we choose. Taking market returns with an equally weighted market index had little effect on performance, although it has been claimed to produce benefits (Bartholdy and Peare, 2004).

6.6 Results: more complex models (improving the modelling of return)

Transforming predictor variables to approximate normality was tried. Chapter 3 of the thesis is fully dedicated to the theory of transformations to near normality. Table 6.3 shows that, rather surprisingly, omitting to transform predictor variables had little effect on investment performance. This may be because, in this study, we consider investment only in large companies. Market capitalisation is the only predictor variable to have a really skewed distribution, and this is not a major determinant of expected return.

Turning to model choice, it is well known that maximum-likelihood regression predictions are too 'noisy', and can be improved by shrinkage towards zero (e.g. Copas, 2000). Ridge regression is an empirical method that has worked well in practice. Regression coefficients $\delta$ are made comparable by standardising the corresponding variables to have zero mean and unit variance. The method can be motivated in a Bayesian and Empirical Bayes context by giving regression coefficients independent prior normal distributions with zero mean and common variance $\psi^2$. The unknown variance can be estimated by maximising the posterior probability $f(r|\sigma^2)$. We have that
\[ f(r|\sigma^2) = (2\pi \psi^2)^{-1/2} (2\pi \sigma^2)^{-n/2} (2\pi \psi^2)^{-(p-1)/2} \]
\[ \times \int \exp\left\{ -\frac{(r - \alpha - X\delta)^T (r - \alpha - X\delta)}{2\sigma^2} - \frac{\delta^T \delta}{2\psi^2} \right\} d\alpha d\delta, \]

where \( \alpha = \alpha \mathbf{1}, \) and \( \psi^{-2} \) is the very large prior variance of \( \alpha. \)

Doing the trivial integration over \( \alpha, \)
\[ f(r|\sigma^2) = \psi^{-1} (2\pi \sigma^2)^{-n/2} (2\pi \psi^2)^{-(p-1)/2} \]
\[ \times \int \exp\left\{ -\frac{(r - \bar{\mathbf{r}} - X\delta)^T (r - \bar{\mathbf{r}} - X\delta)}{2\sigma^2} - \frac{\delta^T \delta}{2\psi^2} \right\} d\alpha d\delta, \]

Writing \( \psi^2 = \sigma^2 / s \) and completing the square with respect to \( \delta \) in the argument of the exponential yields
\[ f(r|\sigma^2) = \frac{(2\pi)^{(p-1)/2}}{\psi (2\pi \sigma^2)^{(n+p-2)/2}} \exp(-a) \]
\[ \frac{s^{(p-1)/2}}{|V^{-1}|^{1/2}}, \]

where \( V^{-1} = (1/\sigma^2)(X^T X + s\mathbf{I}), \)
\[ a = (r - \bar{r})^T (r - \bar{r}) / 2\sigma^2 - (1/2)\gamma^T V^{-1} \gamma, \]

and
\[ \gamma = VX^T (r - \bar{r}) / \sigma^2 = \hat{\delta}. \]  \hfill (6.14)

The variance \( \sigma^2 \) is estimated by maximising the log-likelihood
\[ \log \{ f(r|\sigma^2) \} = -a - \frac{n + p - 2}{2} \log(\sigma^2) - \frac{1}{2} \log|V^{-1}| + \text{const}. \]
By the properties of determinants we have that

$$\log\{f(r|\sigma^2)\} = -a - \frac{n-1}{2} \log(\sigma^2) + \text{const}.$$ 

Taking the derivative with respect to $\hat{\sigma}^2$ and equating it to zero yields the estimate of $\sigma^2$

$$\hat{\sigma}^2 = \frac{\mathbf{r} - \bar{\mathbf{r}})^T (\mathbf{r} - \bar{\mathbf{r}}) - \gamma^T \mathbf{X}^T (\mathbf{r} - \bar{\mathbf{r}})}{n-1}.$$ 

Expanding the brackets out and completing the square in the denominator gives us

$$\hat{\sigma}^2 = \frac{SS + \gamma^T \mathbf{X}^T (\mathbf{r} - \bar{\mathbf{r}}) - \gamma^T \mathbf{X}^T \mathbf{X} \gamma}{n-1}. $$

Taking (6.14) into account, equation above can be rewritten as

$$\hat{\sigma}^2 = \frac{SS + \gamma^T (\mathbf{X}^T \mathbf{X} + sI) \gamma - \gamma^T \mathbf{X}^T \mathbf{X} \gamma}{n-1}. $$

Thus

$$\hat{\sigma}^2 = \frac{SS + s \gamma^T \gamma}{n-1},$$

where SS is the residual sum of squares.

It is straightforward to compute $\hat{\mathbf{b}}$ from (6.14) as the solution of a set of $p-1$ linear equations, if $s$ is known; this is the only nonlinear parameter. The estimation of $s$ was done by Newton-Raphson iteration, after diagonalising the matrix $\mathbf{X}^T \mathbf{X}$, which greatly
simplified subsequent computations. Write $X^T X = \Lambda D \Lambda^T$, where $D$ is diagonal, and let $\phi = \Lambda^T X^T (r - \bar{r})$.

Then the profile log-posterior pdf

$$l(r, \hat{\sigma}^2) = \left( \frac{p - 1}{2} \right) \log(s) - \left( \frac{1}{2} \right) \log|X^T X + sI|$$

$$- \left( \frac{n - 1}{2} \right) \log\left\{ (r - \bar{r})^T (r - \bar{r}) - \gamma^T X^T (r - \bar{r}) \right\} + \text{const}$$

becomes

$$l(r, \hat{\sigma}^2) = \left( \frac{p - 1}{2} \right) \log(s) - \left( \frac{1}{2} \right) \sum_{k=1}^{p-1} \log(D_{kk} + s)$$

$$- \left( \frac{n - 1}{2} \right) \log\left\{ (r - \bar{r})^T (r - \bar{r}) - \sum_{k=1}^{p-1} \phi_k^2 / (D_{kk} + s) \right\} + \text{const}$$

from which $s$ is found by Newton-Raphson iteration, starting from e.g. $s = 1$.

Table 6.3 and Figure 6.2 show that using shrunken estimator rather than using minimum AIC or BIC as model choice criteria made little difference. There is a suggestion that minimum AIC regressions and shrunken regressions outperform minimum BIC regressions, but there is little to choose between the different models, despite the fact that minimum-AIC models include many more variables (typically 35 as opposed to 16). Shrunken regressions behave very like minimum AIC regressions. An unshrunken regression including all available variables performed slightly worse. Presumably, there is so much data here that the statistical error of the regression coefficients is not a major problem.
References


CHAPTER 7

Conclusions and further research
Forecasting returns is a challenging statistical problem. The great deal of complexity imposed by the unknown functional form of the regression equation is complicated by highly skew distributed predictors. Although there is a substantial amount of data available, the predictor variables are very poor, meaning many are needed. There is also a choice between different modelling approaches: minimum-AIC/BIC regression can be used or an empirical Bayes approach where regression coefficients are shrunk towards zero rather than being either simply included or omitted as in stepwise regression can be used instead.

We have shown that, although prediction of stock returns is very inaccurate, it can be done accurately enough to enable an investor to substantially outperform the market, at least in the frictionless world of zero transaction costs. The regression coefficients exemplify well-known effects, such as that firms with smaller market capitalization outperform larger firms, that high annual dividend yield is a good predictor of future return, and show the momentum and return-reversal effects. A minimum AIC or BIC 10-year rolling regression are hard to improve on.

The work is largely devoted to an examination of the ridge regression approach. The method can be used to improve the modelling of return and consists in shrinking regression coefficients towards zero. The shrinkage is introduced in a Bayesian context by assigning independent prior distributions with zero mean and common variance to each of the regression coefficients.
Summarising the approach developed in chapter 4 the natural way to generate Bayesian adjusted regression coefficients is to write down the likelihood function and to put prior distributions on the three firm-specific variables $\alpha_i$, $\beta_i$ and $\sigma_i^2$. Regression coefficients $\delta$ are estimated simultaneously by maximising the product of the likelihood function and the prior pdfs. The estimates of $\beta_i$ and $\sigma_i$ changes when the regression parameters $\delta$ are not fixed at zero. Finally, the likelihood function is maximised to find the hyperparameters. It is necessary to fix $\delta$ and maximize the likelihood function for the $\alpha_i$, $\beta_i$ and $\sigma_i^2$, then to maximize for $\delta$ and to alternate these steps until convergence, for each choice of hyperparameters.

The results given in chapter 5 revealed that shrunken regression gives only a slight improvement in predictive ability as measured by forecast accuracy measures such as median squared error, percentage worst, percentage better. This motivated further research which addresses more modelling and inferential approaches such as the use of AIC or BIC as a model choice criterion. In chapter 5 it was shown (see table 5.2) that due to the poor predictive ability of predictor variables the inclusion of $\delta \neq 0$ hardly changes $\beta_i$ and $\sigma_i$. This is a key result which enables us to greatly simplify the computations by estimating $\delta$ and firm-specific parameters separately. Based on this notion, we adopt the approach discussed in chapter 6. Unlike the methodology presented in chapter 4 (for the ridge regression), the approach examined in chapter 6 has more flexibility and allows more ideas to be evaluated. In addition, in chapter 4 we demonstrated the use of
Vasicek's (1973) adjustment for beta which aims to improve the accuracy of historical betas by taking account of the tendency of betas to regress towards one.

We also investigated the effect of transforming predictor variables to approximate normality on predictive ability. We used the Box-Cox transformation (Box and Cox, 1964) for positively defined predictor variables and Johnson's arcsinh transformation (Johnson et al., 1994, Chapter 1) for predictors that could take zero or negative values. Transforming predictor variables to approximate normality seems prudent, but does not give any marked improvement here, probably because market capitalization is the only really skew variable and we restricted investment to the 350 largest companies. However, many other modelling and inferential possibilities that were explored gave no improvement out of sample or worsened performance. The result presented in this thesis was also given in Baker and Belgorodskiy (2007). The paper explores more modelling approaches including the use of weighted regression, discounted regression and alternative covariance matrix models, such as a two-index model and shrinkage model. It was however shown that use of discounted regression which allows regression parameters to change with time worsened regression performance, as also did the use of a constant correlation model for covariance.

There is no ideal measure of investment performance. We introduced a statistic linked to the loss resulting from a poor regression prediction that could be useful when there is little out-of-sample data. Since computing the statistic in a straightforward manner resulted in rounding error problems the procedure was adopted by rearranging the basic
formulas. Because the Sharpe measure can decrease when gains increase, we also suggest using the degree of risk aversion needed to reduce the value of the investment to zero as an alternative option to using a Sharpe-like statistic based on the semi-variance. We also show the congruence between performance measures based on semi-variance and utility functions and derive the Sortino ratio from the corresponding utility function. However, the best decision aid is probably to plot the certainty equivalent gain as a function of risk aversion (Figure 6.2). Such plots can reveal the complex behaviour of investment strategies where high kurtosis can cause a strategy preferable at low risk aversion to become less favoured under higher risk aversion.

There is a steady decline in the performance of optimal investments made using the regressions since 1975, gain per unit risk declining to less than half of its 1975 value. Our findings show that as investors become more knowledgeable about the determinants of high return their actions are removing these mispricings slowly. It is tempting to speculate that this might be caused by an increase in market efficiency resulting from the kind of investment strategy described in this paper. If so, to succeed in beating the market in future, investors will need ever more detailed models.

Diversifying investment across national boundaries is often regarded as a beneficial financial strategy which may potentially lead to an effective reduction of risk (e.g Mathur and Hanagan, 1983). The advantage lies in the fact that the level of correlation between returns on international assets is lower as compared to the level of correlation between domestic asset returns. Evidence exists that diversifying a portfolio internationally
decreases risk more than industry-wide national diversification (Solnik, 1974). To model international diversification in the context of the single-factor model we develop the approach based on the use of hierarchical betas (see section 2.6). In this approach national stock indices assumed to be linearly related to a proxy world market factor. We also adopted the Markowitz optimum investment strategy to account for international portfolio diversification and obtained analytical formulae for computing the composition of optimum portfolios. The model implies no impediments exist to investment in foreign assets. In fact that is not the case. As Mathur and Hannagan (1983, p. 136) point out "investors may face a variety of barriers to successful international diversification through acquisition of financial instruments issued in foreign countries". One of the principal barriers identified by the authors is the risk associated with the variation of the foreign exchange rate. Section 2.7 addresses this problem and implements the Markowitz strategy when exchange rate risk is taken into account. We have shown that an analytical solution for optimal asset allocation can still be found.

The thesis aims to shed light on the optimal investment problem through the use of mainstream statistical methods. The original contribution here consists in the evaluation of many modelling and inferential methodologies.

There is indeed great scope for much further research. The problem of portfolio evaluation is not only essential for the decision-making process but should be an integral part of it. Unfortunately, there is no single statistic available to compare competing types of investment in the sense that the measures suggested in the literature can lead to
controversial and often misleading results (the problem is addressed in section 6.4). Although, we suggest the degree of risk aversion needed to reduce the value of the investment to zero as a metric of investment performance, which overcomes the problem associated with the Sharpe measure, substantial research still needs to be carried out in order to define the measure for the case when all the gains are above the threshold. This area of the research may be particularly fruitful since the measures of performance like the one suggested in this thesis can be utilised not only for the problem of assessing different types of investment but whenever rival methodologies are to be ranked and when each outcome of a competing methodology is expressed as a number. This covers a wide variety of circumstances such as to enable performance evaluation both include finance and outside, for example evaluation of different approaches to education or different medical interventions.
References


APPENDICES
Appendix 1: Exporting data from Perfect Analysis

To download files from Perfect Analysis, go to the prices screen and use the Export command under the file menu. The FTSE 350 constituents can be selected, so that batch mode need not be invoked. The user can create and save (his own) new portfolios on the portfolios screen, so that constituents of these portfolios are available to export along with FTSE 350. To download HTML files containing other information, go to the perfect information screen, use the batch command under the files menu, and choose Syncard views and the default template. Select HTML file output.

Appendix 2: The structure of the data file

1. Company sequence number;
2. month number (Jan 1955 is month 1);
3. month of year;
4. year;
5. country code;
6. sector code;
7. LSPD company number;
8. current risk free rate;
9. next month's return on the market (continuously compounded) minus next month's risk free rate;
10. next month's return (continuously compounded) minus next month's risk free rate;

11. current RPI (retail price index)

12. monthly return;

13. monthly return on the market;

14. share price in pounds;

15. dividends in pounds;

16. ratio of adjusted to unadjusted prices;

17. total trading volume in millions of shares per month;

18. market cap in millions of pounds;

19. 2-monthly return;

20. 3-monthly return;

21. 4-monthly return;

22. 6 monthly return;

23. annual return;

24. 2 years return;

25. 3 years return;

26. 4 years return;

27. 5 years return;

28. alpha;

29. beta;

30. specific risk (as standard deviation);

31. variance of beta;
32. Annual dividend yield (value of dividends over last year as a percentage of current share price);

33. dividend yield over last 2 years;

34. ditto, over 3 years;

35. total trading volume (last 3 months);

36. total trading volume (last 6 months);

37. total trading volume (last year);

38. total trading volume (last 2 years);

39. total trading volume (last 5 years);

40. EB corrected beta;

41. EB corrected specific risk (as standard deviation);

42. share price over RPI;

43. market cap over RPI;

44. P/E ratio from LSPD (can't be used for real investments)

45. market cap rank (1 is largest, -1 if missing);

46. Forecast month is winter (1) or summer (0).

**Appendix 3: Exponential utility and the certainty equivalent**

In this appendix we derive expression (2.1) assuming that the vector of returns has a multivariate normal distribution with parameters $\mu$ and $V$. 
Thus,

\[ E[\exp(-\eta r^T x)] = \frac{1}{(2\pi)^\frac{W}{2}|\mathbf{V}|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp[-\eta r^T x - \frac{1}{2}(r - \mu)^T \mathbf{V}^{-1}(r - \mu)]dr. \]

Notice that

\[ \eta r^T x + \frac{1}{2}(r - \mu)^T \mathbf{V}^{-1}(r - \mu) = \eta r^T x - r^T \mathbf{V}^{-1} \mu + \frac{1}{2} \mu^T \mathbf{V}^{-1} \mu + \frac{1}{2} r^T \mathbf{V}^{-1} r \]

and

\[ \rho (r - \omega)^T \mathbf{V}^{-1} (r - \omega) + \kappa = \rho r^T \mathbf{V}^{-1} r - 2 \rho \omega^T \mathbf{V}^{-1} r + \rho \omega^T \mathbf{V}^{-1} \omega + \kappa, \]

we can equate the trio with

\[ \rho = 1/2, \]

\[ \eta x^T - \mu^T \mathbf{V}^{-1} = -\omega^T \mathbf{V}^{-1}; \text{ so } \omega = \mu - \eta \mathbf{V} \]

\[ \frac{1}{2} \omega^T \mathbf{V}^{-1} \omega + \kappa = \frac{1}{2} \mu^T \mathbf{V}^{-1} \mu, \]

or \( \kappa = \frac{1}{2} \mu^T \mathbf{V}^{-1} \mu - \frac{1}{2} (\mu^T - \eta x^T \mathbf{V}) \mathbf{V}^{-1} (\mu - \eta \mathbf{V} x) \)

\[ = \frac{1}{2} \mu^T \mathbf{V}^{-1} \mu - \frac{1}{2} \mu^T \mathbf{V}^{-1} \mu + \eta x^T \mu - \frac{1}{2} x^T \mathbf{V} x \]

\[ = \eta x^T \mu - \frac{1}{2} x^T \mathbf{V} x. \]

Therefore

\[ E[\exp(-\eta r^T x)] = \frac{1}{(2\pi)^\frac{W}{2}|\mathbf{V}|^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp(-\kappa) \int_{-\infty}^{\infty} \exp[-\frac{1}{2} (r - \omega)^T \mathbf{V}^{-1} (r - \omega)]dr, \]

\[ E[\exp(-\eta r^T x)] = \exp(-\eta x^T \mu + \frac{1}{2} x^T \mathbf{V} x). \]
Appendix 4: Covariance of returns between securities – international diversification of an investment

The expected return on a security $i$ is

$$E(R_i) = E[\alpha_i + \beta_i(B_m R_g + \rho_m) + e_i].$$

Since the expected value of the sum of the random variables is the sum of the expected values we have

$$E(R_i) = E(\alpha_i) + E(\beta_i B_m R_g) + E(\beta_i \rho_m) + E(e_i),$$

$\alpha_i$, $\beta_i$ and $B_m$ are constants and by construction the expected values of $e_i$ and $\rho_m$ are zeros. Thus

$$E(R_i) = \alpha_i + \beta_i B_m E(R_g).$$

The covariance between any two securities can be written as

$$V_{ij} = E[(R_i - E(R_i))(R_j - E(R_j))].$$

Substituting for $R_i$, $E(R_i)$, $R_j$ and $E(R_j)$ yields

$$V_{ij} = E[(\alpha_i + \beta_i(B_m R_g + \rho_m) + e_i - \alpha_i - \beta_i B_m E(R_g))(\alpha_j + \beta_j(B_j R_g + \rho_j) + e_j - \alpha_j - \beta_j B_j E(R_g))]$$

Simplifying by cancelling the $\alpha$'s and combining the terms involving $\beta$'s yields

$$V_{ij} = E[(\beta_i(B_m R_g + \rho_m - B_m E(R_g))) + e_i)(\beta_j(B_j R_g + \rho_j - B_j E(R_g)) + e_j)].$$

Carrying out the multiplication and recalling that by assumption

$$\text{cov}(e_i R_m) = 0, \text{cov}(e_j R_m) = 0,$$
yields

\[
V_{ij} = \beta_i \beta_j B_m B_f \sigma_g^2 + \beta_i \beta_j \sigma_m^2 \delta_{mf} + \sigma_i^2 \delta_{ij},
\]

where \( \sigma_m^2 \) is the variance of \( \rho_m \) (\( \sigma_m^2 = E(\rho_m^2) \)), \( \sigma_g^2 \) is the variance of \( R_g \) (\( \sigma_g^2 = E(R_g - E(R_g))^2 \)) and \( \sigma_i^2 \) is the variance of \( \epsilon_i \) (\( \sigma_i^2 = E(\epsilon_i^2) \)).

Appendix 5: A derivation of an analytical solution for the Markowitz investment strategy – international diversification of an investment when exchange rate risk is taken into consideration

Utilising the formula for covariance given in the main body the optimal selection of stocks might be found by solving the system of simultaneous equations

\[
\mu_i = \sum_{j=1}^{N} \left( \delta_{ij} \sigma_i^2 + \beta_i \beta_j \sigma_m^2 \delta_{mf} + \beta_i \beta_j B_m B_f \sigma_g^2 + \nu_m^2 \delta_{mf} \right) Z_j,
\]

\[
i = 1, ..., N.
\]

Let

\[
D_m = \sum_{j=m}^{n} \beta_j Z_j \quad \text{(A5.1)}, \quad E_m = \sum_{j=m}^{n} Z_j \quad \text{(A5.2)},
\]

\[
C = \sum_{m=1}^{n} B_m D_m \quad \text{(A5.3)}.
\]
Hence

\[ \mu_i = \sigma_i^2 Z_i + \beta_i \sigma_m^2 D_m + \beta_i \sigma_\gamma^2 B_m C + \nu_m^2 E_m, \quad i = 1, \ldots, N, \]

so that

\[ Z_i = \frac{\mu_i - \beta_i \sigma_m^2 D_m - B_m \sigma_\gamma^2 \beta_i C - \nu_m^2 E_m}{\sigma_i^2}. \tag{A5.4} \]

Let

\[ S_m = \sum_{j \in m} \frac{\mu_j \beta_j}{\sigma_j^2}, \quad W_m = \sum_{j \in m} \frac{\beta_j^2}{\sigma_j^2}, \quad G_m = \sum_{j \in m} \frac{\beta_j}{\sigma_j^2}. \]

Then from (A5.4),

\[ D_m = \frac{S_m - B_m \sigma_\gamma^2 W_m C - \nu_m^2 E_m G_m}{1 + \sigma_m^2 W_m}, \tag{A5.5} \]

and on using equation (A5.3)

\[ C = \frac{\sum_{m=1}^{n} B_m S_m}{1 + \sigma_\gamma^2 W_m} - \frac{\sum_{m=1}^{n} \nu_m^2 E_m G_m}{1 + \sigma_m^2 W_m}. \tag{A5.6} \]

Let

\[ N_m = \sum_{j \in m} \frac{\mu_j}{\sigma_j^2}, \quad \text{and} \quad I_m = \sum_{j \in m} \frac{1}{\sigma_j}. \]

Then from (A5.4)

\[ E_m = \frac{N_m - \sigma_m^2 D_m G_m - B_m \sigma_\gamma^2 C G_m}{1 + \nu_m^2 I_m}. \tag{A5.7} \]
Substituting (A5.5) to the equation above gives

\[
E_m = \frac{1}{1 + \nu_m^2 I_m} \left[ N_m - B_m \sigma_g^2 c G_m - \frac{\sigma_m^2 G_m S_m + \sigma_m^2 G_m B_m \sigma_g^2 W_m C}{1 + \sigma_m^2 W_m} \right].
\]

Substituting for \( E_m \) from the equation above to (A5.6) gives

\[
C = \frac{\sum_{m=1}^{n} \frac{B_m S_m}{1 + \sigma_m^2 W_m} - \sum_{m=1}^{n} \frac{\nu_m^2 G_m}{1 + \sigma_m^2 W_m} \times \frac{N_m + N_m \sigma_m^2 W_m - \sigma_m^2 G_m S_m}{1 + \nu_m^2 I_m)(1 + \sigma_m^2 W_m) - \sigma_m^2 \nu_m^2 G_m^2}}{1 + \sigma_g^2 \sum_{m=1}^{n} \frac{B_m^2 W_m}{1 + \sigma_m^2 W_m} - \sum_{m=1}^{n} \frac{\nu_m^2 G_m^2 B_m \sigma_g^2}{1 + \sigma_m^2 W_m} \times \frac{1 + 2 \sigma_m^2 W_m}{(1 + \nu_m^2 I_m)(1 + \sigma_m^2 W_m) - \sigma_m^2 \nu_m^2 G_m^2}}.
\]

**Appendix 6: Updating formula for the variance**

In general, the population variance of a finite population of size \( n \) is given by

\[
\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}_n^2,
\]

where \( \bar{x}_n \) is the sample mean.

Analogously, for the population of size \( n+1 \) the variance is

\[
\sigma_{n+1}^2 = \frac{1}{n+1} \sum_{i=1}^{n+1} x_i^2 - \bar{x}_{n+1}^2.
\]
Let us assume that the population of size $n+1$ is obtained as a result of inclusion of a new element $X_{n+1}$ into the population of size $n$. We are interested in expressing the variance of the population of size $n+1$ in terms of the mean and variance of the population of size $n$ and the new element $X_{n+1}$.

The sample mean $\bar{x}_{n+1}$ can be written as

$$\bar{x}_{n+1} = \frac{1}{n+1} \left( x_{n+1} + n\bar{x}_n \right).$$

Substituting $\bar{x}_{n+1}$ from the expression above yields

$$\sigma^2_{n+1} = \frac{1}{n+1} \left( \sum_{i=1}^{n} x_i^2 + x_{n+1}^2 - (x_{n+1} + n\bar{x}_n)^2 / (n+1) \right). \quad (A6.1)$$

Adding and subtracting $n\bar{x}_n^2 / (n+1)$ to the right-hand side of equation (A6.1) and expanding the square we obtain the following equation

$$\sigma^2_{n+1} = \frac{n}{n+1} \sigma_n^2 + \frac{1}{(n+1)^2} \left( n(n+1)\bar{x}_n^2 + (n+1)x_{n+1}^2 - x_{n+1}^2 - 2n\bar{x}_n x_{n+1} - n^2\bar{x}_n^2 \right).$$

Rearranging the terms in the above equation gives

$$\sigma^2_{n+1} = \frac{n}{n+1} \sigma_n^2 + \frac{n}{(n+1)^2} \left( \bar{x}_n^2 + x_{n+1}^2 - 2\bar{x}_n x_{n+1} \right).$$

Completing the square

$$\sigma^2_{n+1} = \frac{n}{n+1} \left( \sigma_n^2 + \frac{(\bar{x}_n - x_{n+1})^2}{n+1} \right).$$
Appendix 7: Results of transforming predictive variables to approximate normality

Table A7.1. Results of applying Box-Cox transformation to positively defined predictive variables

<table>
<thead>
<tr>
<th>Variable name</th>
<th>For an untransformed data set</th>
<th>Optimal transformation parameters</th>
<th>Optimal value of the log-likelihood</th>
<th>After applying Box-Cox transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard deviation $\sigma_s$</td>
<td>skewness $s$</td>
<td>kurtosis</td>
<td>$\phi$</td>
</tr>
<tr>
<td>share price in pounds</td>
<td>4.113</td>
<td>3.071</td>
<td>15.575</td>
<td>0.28</td>
</tr>
<tr>
<td>average trading volume in millions of shares per month</td>
<td>266.123</td>
<td>22.406</td>
<td>634.452</td>
<td>0.10</td>
</tr>
<tr>
<td>market cap in million of pounds</td>
<td>9454.640</td>
<td>8.528</td>
<td>96.307</td>
<td>0.08</td>
</tr>
<tr>
<td>specific risk</td>
<td>0.090</td>
<td>2.751</td>
<td>10.844</td>
<td>-0.52</td>
</tr>
<tr>
<td>variance of beta</td>
<td>0.549</td>
<td>6.827</td>
<td>67.237</td>
<td>-0.05</td>
</tr>
<tr>
<td>dividend yield</td>
<td>51.858</td>
<td>26.971</td>
<td>907.169</td>
<td>0.02</td>
</tr>
<tr>
<td>dividend yield over last 2 years</td>
<td>98.910</td>
<td>25.633</td>
<td>804.037</td>
<td>0.02</td>
</tr>
<tr>
<td>ditto, over 3 years</td>
<td>150.740</td>
<td>23.548</td>
<td>679.567</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Table A7.2. Results of applying Johnson’s transformation to the predictive variables

<table>
<thead>
<tr>
<th>Variable name</th>
<th>For an untransformed data set</th>
<th>Optimal transformation parameters</th>
<th>Optimal value of the log-likelihood</th>
<th>After applying Johnson’s transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>standard deviation $\sigma_x$</td>
<td>skewness $\gamma$</td>
<td>kurtosis $\kappa$</td>
<td>$a$</td>
</tr>
<tr>
<td>monthly return</td>
<td>0.158</td>
<td>3.144</td>
<td>126.901</td>
<td>0.000</td>
</tr>
<tr>
<td>monthly return on the market</td>
<td>0.040</td>
<td>-0.900</td>
<td>1.018</td>
<td>0.048</td>
</tr>
<tr>
<td>dividends in pounds</td>
<td>0.147</td>
<td>20.689</td>
<td>647.728</td>
<td>-0.018</td>
</tr>
<tr>
<td>2-monthly return</td>
<td>0.218</td>
<td>2.127</td>
<td>67.144</td>
<td>0.030</td>
</tr>
<tr>
<td>3-monthly return</td>
<td>0.264</td>
<td>1.604</td>
<td>43.586</td>
<td>0.036</td>
</tr>
<tr>
<td>4-monthly return</td>
<td>0.303</td>
<td>1.211</td>
<td>32.837</td>
<td>0.041</td>
</tr>
<tr>
<td>6-monthly return</td>
<td>0.370</td>
<td>0.793</td>
<td>22.667</td>
<td>0.101</td>
</tr>
<tr>
<td>annual return</td>
<td>0.535</td>
<td>0.269</td>
<td>12.879</td>
<td>0.148</td>
</tr>
<tr>
<td>2 years return</td>
<td>0.784</td>
<td>0.752</td>
<td>17.278</td>
<td>0.216</td>
</tr>
<tr>
<td>3 years return</td>
<td>1.016</td>
<td>1.787</td>
<td>24.547</td>
<td>0.273</td>
</tr>
<tr>
<td>4 years return</td>
<td>1.245</td>
<td>2.500</td>
<td>29.505</td>
<td>0.326</td>
</tr>
<tr>
<td>5 years return</td>
<td>1.478</td>
<td>2.771</td>
<td>30.194</td>
<td>0.385</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.024</td>
<td>2.624</td>
<td>28.468</td>
<td>0.000</td>
</tr>
<tr>
<td>Beta</td>
<td>0.648</td>
<td>-0.140</td>
<td>4.013</td>
<td>0.964</td>
</tr>
</tbody>
</table>
Table A7.3. Results of applying Johnson's transformation to share price

<table>
<thead>
<tr>
<th>Parameter</th>
<th>the range of variation of $c$ is bounded above by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$5 / \sigma_x$</td>
</tr>
<tr>
<td></td>
<td>(1.23E+00)</td>
</tr>
<tr>
<td>optimal value of $a$</td>
<td>-5.41E-01</td>
</tr>
<tr>
<td>optimal value of $c$</td>
<td>1.19E+00</td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-130634</td>
</tr>
<tr>
<td>skewness of the</td>
<td>3.44E-02</td>
</tr>
<tr>
<td>resulting distribution</td>
<td></td>
</tr>
<tr>
<td>kurtosis of the</td>
<td>-5.70E-01</td>
</tr>
<tr>
<td>resulting distribution</td>
<td></td>
</tr>
</tbody>
</table>

Table A7.4. Results of applying Johnson's transformation to trading volume

<table>
<thead>
<tr>
<th>Parameter</th>
<th>the range of variation of $c$ is bounded above by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$5 / \sigma_x$</td>
</tr>
<tr>
<td></td>
<td>(2.07E-02)</td>
</tr>
<tr>
<td>optimal value of $a$</td>
<td>-3.21E+01</td>
</tr>
<tr>
<td>optimal value of $c$</td>
<td>2.01E-02</td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-285296</td>
</tr>
<tr>
<td>skewness of the</td>
<td>2.71E+00</td>
</tr>
<tr>
<td>resulting distribution</td>
<td></td>
</tr>
<tr>
<td>kurtosis of the</td>
<td>9.86E+00</td>
</tr>
<tr>
<td>resulting distribution</td>
<td></td>
</tr>
</tbody>
</table>
Table A7.5. Results of applying Johnson's transformation to market capitalisation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\sigma_x$ (5.8E-04)</th>
<th>$\sigma_x$ (5.8E-03)</th>
<th>$\sigma_x$ (5.8E-02)</th>
<th>$\sigma_x$ (5.8E-01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal value of parameter $a$</td>
<td>-4.09E-12</td>
<td>-4.09E-12</td>
<td>-4.09E-12</td>
<td>-4.09E-12</td>
</tr>
<tr>
<td>Optimal value of parameter $c$</td>
<td>5.63E-04</td>
<td>5.79E-03</td>
<td>5.62E-02</td>
<td>5.62E-01</td>
</tr>
<tr>
<td>Optimal value of the log-likelihood</td>
<td>-496480</td>
<td>-460212</td>
<td>-446982</td>
<td>-441110</td>
</tr>
<tr>
<td>Skewness of the resulting distribution</td>
<td>2.08E+00</td>
<td>6.64E-01</td>
<td>-1.40E-01</td>
<td>-4.30E-01</td>
</tr>
<tr>
<td>Kurtosis of the resulting distribution</td>
<td>4.61E+00</td>
<td>-3.40E-02</td>
<td>-5.10E-01</td>
<td>-2.70E-01</td>
</tr>
</tbody>
</table>

Table A7.6. Results of applying Johnson's transformation to variance of beta

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\sigma_x$ (9.4E+00)</th>
<th>$\sigma_x$ (9.4E+01)</th>
<th>$\sigma_x$ (9.4E+02)</th>
<th>$\sigma_x$ (9.4E+03)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal value of parameter $a$</td>
<td>-2.22E-16</td>
<td>-2.22E-16</td>
<td>-2.22E-16</td>
<td>-2.22E-16</td>
</tr>
<tr>
<td>Optimal value of parameter $c$</td>
<td>9.08E+00</td>
<td>9.36E+01</td>
<td>9.08E+02</td>
<td>9.08E+03</td>
</tr>
<tr>
<td>Optimal value of the log-likelihood</td>
<td>19478</td>
<td>33811</td>
<td>34291</td>
<td>34299</td>
</tr>
<tr>
<td>Skewness of the resulting distribution</td>
<td>1.54E+00</td>
<td>4.66E-01</td>
<td>2.27E-01</td>
<td>2.20E-01</td>
</tr>
<tr>
<td>Kurtosis of the resulting distribution</td>
<td>2.29E+00</td>
<td>4.38E-01</td>
<td>6.05E-01</td>
<td>6.15E-01</td>
</tr>
</tbody>
</table>
### Table A7.7. Results of applying Johnson’s transformation to dividend yield

<table>
<thead>
<tr>
<th>the range of variation of $c$ is bounded above by</th>
<th>$5 / \sigma_x$</th>
<th>$50 / \sigma_x$</th>
<th>$500 / \sigma_x$</th>
<th>$5000 / \sigma_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1.07E-01)</td>
<td>(1.07E+00)</td>
<td>(1.07E+01)</td>
<td>(1.07E+02)</td>
</tr>
<tr>
<td>optimal value of parameter $a$</td>
<td>3.55E-14</td>
<td>3.55E-14</td>
<td>3.55E-14</td>
<td>3.55E-14</td>
</tr>
<tr>
<td>optimal value of parameter $c$</td>
<td>1.03E+01</td>
<td>1.07E+00</td>
<td>1.03E+01</td>
<td>1.03E+02</td>
</tr>
<tr>
<td>optimal value of the log-likelihood</td>
<td>-173077</td>
<td>-130050</td>
<td>-101162</td>
<td>-63299</td>
</tr>
<tr>
<td>skewness of the resulting distribution</td>
<td>4.32E+00</td>
<td>9.12E-01</td>
<td>-5.70E-02</td>
<td>-3.20E-01</td>
</tr>
<tr>
<td>kurtosis of the resulting distribution</td>
<td>2.38E+01</td>
<td>1.51E+00</td>
<td>-1.16E+00</td>
<td>-1.54E+00</td>
</tr>
</tbody>
</table>

### Table A7.8. Results of applying Johnson’s transformation to dividend yield over 2 years

<table>
<thead>
<tr>
<th>the range of variation of $c$ is bounded above by</th>
<th>$5 / \sigma_x$</th>
<th>$50 / \sigma_x$</th>
<th>$500 / \sigma_x$</th>
<th>$5000 / \sigma_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(5.61E-02)</td>
<td>(5.61E-01)</td>
<td>(5.61E+00)</td>
<td>(5.61E+01)</td>
</tr>
<tr>
<td>optimal value of parameter $a$</td>
<td>-2.49E+14</td>
<td>-2.49E+14</td>
<td>-2.49E+14</td>
<td>-2.49E+14</td>
</tr>
<tr>
<td>optimal value of parameter $c$</td>
<td>5.42E-02</td>
<td>5.59E-01</td>
<td>5.42E+00</td>
<td>5.42E+01</td>
</tr>
<tr>
<td>optimal value of the log-likelihood</td>
<td>-213462</td>
<td>-170137</td>
<td>-141103</td>
<td>-103931</td>
</tr>
<tr>
<td>skewness of the resulting distribution</td>
<td>3.84E+00</td>
<td>8.90E-01</td>
<td>-5.40E-02</td>
<td>-3.30E-01</td>
</tr>
<tr>
<td>kurtosis of the resulting distribution</td>
<td>1.82E+01</td>
<td>1.19E+00</td>
<td>-1.15E+00</td>
<td>-1.51E+00</td>
</tr>
</tbody>
</table>
Table A7.9. Results of applying Johnson's transformation to dividend yield over 3 years

<table>
<thead>
<tr>
<th>the range of variation of ( c ) is bounded above by</th>
<th>( 5 / \sigma_x )</th>
<th>( 50 / \sigma_x )</th>
<th>( 500 / \sigma_x )</th>
<th>( 5000 / \sigma_x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 5 / \sigma_x )</td>
<td>( 3.68E-02 )</td>
<td>( 6.39E-14 )</td>
<td>( 6.39E-14 )</td>
<td>( 6.39E-14 )</td>
</tr>
<tr>
<td>( 50 / \sigma_x )</td>
<td>( 3.68E-01 )</td>
<td>( 6.39E-14 )</td>
<td>( 6.39E-14 )</td>
<td>( 6.39E-14 )</td>
</tr>
<tr>
<td>( 500 / \sigma_x )</td>
<td>( 3.68E+00 )</td>
<td>( 6.39E-14 )</td>
<td>( 6.39E-14 )</td>
<td>( 6.39E-14 )</td>
</tr>
<tr>
<td>( 5000 / \sigma_x )</td>
<td>( 3.68E+01 )</td>
<td>( 6.39E-14 )</td>
<td>( 6.39E-14 )</td>
<td>( 6.39E-14 )</td>
</tr>
</tbody>
</table>

| optimal value of parameter \( a \) | \( 6.39E-14 \) | \( 6.39E-14 \) | \( 6.39E-14 \) | \( 6.39E-14 \) |
| optimal value of parameter \( c \) | \( 3.56E-02 \) | \( 3.67E-01 \) | \( 3.56E+00 \) | \( 3.56E+01 \) |
| optimal value of the log-likelihood | \( -239530 \) | \( -194828 \) | \( -165503 \) | \( -128955 \) |
| skewness of the resulting distribution | \( 3.61E+00 \) | \( 9.31E-01 \) | \( -2.60E-02 \) | \( -3.20E-01 \) |
| kurtosis of the resulting distribution | \( 1.58E+01 \) | \( 1.13E+00 \) | \( -1.12E+00 \) | \( -1.48E+00 \) |
Appendix 8: Transforming predictor variable: variance of beta 1996-2005 to near normality

Figure A8.1 A distribution of “variance of beta” 1996-2005 (a) before transformation and (b) after Box-Cox transformation

Figure A8.2 A transformed predictor variable: “variance of beta” 1996-2005 after Johnson’s transformation
Appendix 9: The derivation of the explicit formulas for alpha and beta

In the appendix we derive formulae (4.5) and (4.6). The system of simultaneous equations to solve is given by

\[ \sum_{j=1}^{n_t} \frac{R_y - \alpha_i - \delta^T x_j - \beta_i R_{m_j} - \alpha_i - \bar{\alpha}}{\sigma_i^2} = 0 \]

and

\[ \sum_{j=1}^{n_t} \frac{(R_y - \alpha_i - \delta^T x_j - \beta_i R_{m_j}) R_{m_j} - \beta_i - \bar{\beta}}{\sigma_i^2 \sigma_\beta^2} = 0. \]

By rearranging the terms of the equations above

\[ n_i \hat{\alpha}_i + \hat{\beta_i} \sum_{j=1}^{n_t} R_{m_j} + \sum_{j=1}^{n_t} \delta^T x_j = \sum_{j=1}^{n_t} R_y - \hat{\alpha}_i - \bar{\alpha} = 0, \quad (A9.1) \]

\[ \hat{\alpha}_i \sum_{j=1}^{n_t} R_{m_j} + \hat{\beta_i} \sum_{j=1}^{n_t} R_{m_j} + \sum_{j=1}^{n_t} \delta^T x_j R_{m_j} = \sum_{j=1}^{n_t} R_y R_{m_j} - \hat{\beta}_i - \bar{\beta} = 0. \quad (A9.2) \]

Bearing in mind that the system of equations needs to be solved with respect to parameters \( \alpha_i \) and \( \beta_i \), we rearrange equations (A9.1) and (A9.2). This gives

\[ n_i \hat{\alpha}_i + \hat{\beta_i} \sum_{j=1}^{n_t} R_{m_j} + \sum_{j=1}^{n_t} \delta^T x_j = \sum_{j=1}^{n_t} R_y + \frac{\hat{\alpha}_i - \bar{\alpha}}{\sigma_i^2}, \quad (A9.3) \]

and

\[ \hat{\alpha}_i \sum_{j=1}^{n_t} R_{m_j} + \hat{\beta}_i \sum_{j=1}^{n_t} R_{m_j} + \sum_{j=1}^{n_t} \delta^T x_j R_{m_j} = \sum_{j=1}^{n_t} R_y R_{m_j} + \frac{\hat{\beta}_i - \bar{\beta}}{\sigma_i^2} \sigma_\beta^2. \quad (A9.4) \]

We can eliminate \( \hat{\beta}_i \) by multiplying the left and right hand sides of equation (A9.3) by

\[ \left( \sum_{j=1}^{n_t} R_{m_j} \right)^{-1} \left( \sum_{j=1}^{n_t} R_{m_j} R_{m_j} + \frac{\sigma_i^2}{\sigma_\beta^2} \right) \]
and subtracting the resulting equation from equation (A9.4). That is

\[ \hat{\alpha}_i \bar{R}_{m_i} + Q_i - \frac{\hat{\alpha}_i}{R_{m_i}^2} (1 + \nu \sigma^2_i / n_i) \left( \bar{R}_{m_i}^2 + \nu \sigma^2_i / n_i \right) - \frac{E_i}{R_{m_i}^2} (\bar{R}_{m_i}^2 + \nu \sigma^2_i / n_i) = N_i - \]

\[ - \frac{\beta \nu \sigma^2_i / n_i}{R_{m_i}^2} (\bar{R}_{m_i}^2 + \nu \sigma^2_i / n_i) - \frac{\bar{\alpha}_i \nu \sigma^2_i / n_i}{R_{m_i}^2} (\bar{R}_{m_i}^2 + \nu \sigma^2_i / n_i), \]

where

\[ \bar{R}_{m_i} = \sum_{j=1}^{n} R_{m_j} / n_i, \quad \bar{R}_{m_i}^2 = \sum_{j=1}^{n} R_{m_j}^2 / n_i, \quad \bar{R}_i = \sum_{j=1}^{n} R_{ij} / n_i, \]

\[ \bar{R}_i^2 = \sum_{j=1}^{n} R_{ij}^2 / n_i, \quad N_i = \sum_{j=1}^{n} R_{ij} R_{mj} / n_i, \quad E_i = \sum_{j=1}^{n} \delta^T x_j / n_i, \]

\[ Q_i = \sum_{j=1}^{n} \delta^T x_y R_{mj} / n_i, \quad Z_i = \sum_{j=1}^{n} \delta^T x_y R_{ij} / n_i, \quad P_i = \sum_{j=1}^{n} (\delta^T x_y)^2 / n_i, \]

\[ \nu_\alpha = 1/\sigma^2_\alpha, \quad \nu_\beta = 1/\sigma^2_\beta. \]

Thus

\[ \hat{\alpha}_i = \frac{\bar{R}_{m_i} (N_i - Q_i + \bar{\beta}_i \nu \sigma^2_i / n_i) + (E_i - \bar{R}_i - \bar{\alpha}_i \sigma^2_i / n_i) (\bar{R}_{m_i}^2 + \nu \sigma^2_i / n_i)}{(\bar{R}_{m_i}^2 - (1 + \nu \sigma^2_i / n_i) (\bar{R}_{m_i}^2 + \nu \sigma^2_i / n_i))}, \]

and from equation (A9.4)

\[ \hat{\beta}_i = \frac{N_i + \bar{\beta}_i \nu \sigma^2_i / n_i - \hat{\alpha}_i \bar{R}_{m_i} - Q_i}{\bar{R}_{m_i}^2 + \sigma^2_i \nu / n_i}, \]

198
Appendix 10: The listing of a fortran program that has been developed by the author

The program computes EB adjusted regression coefficients and iterates for the optimal values of hyperparameters.

```fortran
module stuff
  implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
  doubleprecision, parameter:: zero=0.d+00,half=0.5d+00,one=1.d+00,two=2.d+00,three=3.d+00,four=4.d+00
  integer, parameter :: max_params=7
  character(80):: infile ! variable for input file (using for 'lsedata.dat' or its modification and 'params.con1)
  character(80):: file_stem
  integer: :n_restarts,ibest,ipar
  real*8 :: start_val(max_params), best_vals(max_params)
  integer, dimension(max_params):: to_be_iterated,to_be_trans
  integer,parameter:: nits=1000
end module stuff

module forf04atf
  integer, parameter :: nmax=25,ia=nmax,iaa=nmax !for f04atf
  integer:: ifail,np !
  real*8:: ar(ia,nmax),aa(iaa,nmax),b(nmax),wksl(nmax),wks2(nmax) !forf04atf
  integer:: extra_term(22)
end module forf04atf

module specific
  implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
  ! stores information specific to the problem at hand...
  integer :: predictdate
  integer, parameter :: maxmonth=700 ! maximum number of observations for each company
  integer, parameter :: maxcomp=5000 ! maximum number of companies
  integer :: jnumber_comp(0:maxmonth) ! this array stores the first column of lsedata.dat
  integer :: month(maxmonth)
  :: year(maxmonth)
  :: n_monthly_ret(maxmonth) ! next month return on the market minus next month's risk free rate
  :: n_market_ret(maxmonth) ! next month's return minus next month's risk free rate
  :: market_ret(maxmonth) ! monthly return on the market
  :: share_price(maxmonth) ! share price in pounds
  :: divid(maxmonth) ! dividends in pounds
  :: t_volume(maxmonth) ! average trading volume in millions of shares
  :: m_cap(maxmonth) ! market capitalization in millions of pounds
  :: two_m_r(maxmonth) ! 2-monthly return
  :: three_m_r(maxmonth) ! 3-monthly return
  :: four_m_r(maxmonth) ! 4-monthly return
  :: six_m_r(maxmonth) ! 6-monthly return
  :: one_y_r(maxmonth) ! annual return
  :: two_y_r(maxmonth) ! 2 years return
  :: three_y_r(maxmonth) ! 3 years return
  :: four_y_r(maxmonth) ! 4 years return
  :: five_y_r(maxmonth) ! 5 years return
  :: six_y_r(maxmonth) ! 6 years return
  :: spec_risk(maxmonth) ! specific risk
  :: var_of_beta(maxmonth) ! variance of beta
  :: div_yield(maxmonth) ! dividend yield
  :: div_yield2y(maxmonth) ! dividend yield over last two years
  :: div_yield3y(maxmonth) ! ditto, over last three years
  :: s(maxcomp)
  :: k(maxcomp)
end module specific
```
program minimise
  use stuff
  use specific
  use params

  ! optimal values of parameters
  (optim_alpha, optim_beta are regression coefficients)
  integer :: line, line1
  integer :: n_month(maxcomp) ! number of observations for particular company
  integer :: n_comp ! store number of companies
  integer :: j
  integer :: num_month(maxcomp)
  integer :: i,ti,schl,sch2,out
  real*8:: w(22), v(22),ve(22,22)
  real*8:: u(22)
  real*8:: vect_of_predx(22)
  real*8:: coefsv(22),coef_delta(22,22)
  character(20):: outp,outp_,outp_,outpreginf,outpsterr ! stores names of the output files
               ! outp - for 'test2.dat', outpreginf - for 'reginfo.dat'
end module specific

module params

! stores information about parameters of prior distributions
! and vector of regression coefficients delta
  real*8:: alpha
  real*8:: sigma_alpha
  real*8:: beta
  real*8:: sigma_beta
  real*8:: alpha
  real*8:: beta
  real*8:: v_alpha
  real*8:: v_beta
  real*8:: delta(22)
  real*8:: v_delta,sigma_delta
end module params

module fortransformation

  implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
  real*8:: hl ! m_cap 0.005
  real*8 :: h2,h ! m_cap 50
  real*8, parameter:: pi=3.14159265358
  real*8, parameter:: xmiss=-10.D+00
  real*8, parameter:: extra_val=l.D+00
  integer:: year_tr,month_tr,k_tr
  integer:: i_tr,j_tr,n_tr,l_tr,num_miss,lim
  real*8:: a_temp(30)
  real*8:: x_tr(150000)
  real*8:: y_tr(150000)
  real*8:: junk_tr
  real*8:: mean_x,disp_x,st_div_x, mean_y,disp_y,opt_mean_y,opt_stddiv_y
  real*8:: mean_y_BC,disp_y_BC, opt_mean_y_BC, opt_stdiv_y_BC
  real*8:: c_tr,a_tr
  real*8:: alpha_tr, lambda_tr
double precision :: jacobian, likelih

  double precision :: jacobian_BC, likelih_BC
  real*8:: opt_c, opt_a, opt_l
  real*8:: opt_alpha, opt_lambda, opt_1_BC
  real*8:: par1(8:30),par2(8:30),par3(8:30),par4(8:30)
  integer :: var_name(8:30)
end module fortransformation

program minimise
  use stuff
  use specific
  use params

  ! This program

  ! (1) reads parameters of prior distribution (starting values) from file 'params.con'
(2) reads starting values of regression coefficients delta from 'params.con'
(3) reads datafile lsedata.dat (or lsedat_ch.dat etc) company by company
(4) calculates values of S,K,N,L,M,R,Q,Z,P and n_month for each company
(5) minimizing minus log-likelihood (as a function of one variable xc, which is the variance of residual)
(6) writes the information about optimal solution (for each company) into the file 'test2.dat'
(7) writes regression coefficients (optim_alpha, optim_beta) and sigsq for each company into the file 'reginfo.dat'
(8) calculates values of variables W,V,U,Vel...Ve22 for each company and for each element of
the vector of predictors delta
(9) calculates coefficients before deltas in the system of simultaneous equations
(10) calculates delta_1 .... delta22 (see subroutine 'delta calculation')

NB: Before starting the program it is necessary to add line of '1' to the end of file lsedata.dat

External subroutines from Nag Library: e04abf searches for a minimum, in a given finite interval,
of a continuous function of a single variable, f04arf calculates the solution of a set of real linear equations

date 10/06/2005

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
open(9, file=infile, status='old')! opening lsedata.dat or another data file

end subroutine recognition

subroutine iter_comp_by_comp
  use fortransformation
  use stuff
  use specific
  use params
  implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
  real*8 :: junk
  logical :: flag
  integer, parameter :: nout=6
  double precision :: a !
  double precision :: b !
  double precision :: eps ! for e04abf
  double precision :: f !
  double precision :: t !
  double precision :: x !
  integer :: count !sch - number of predictor
  integer :: ifail ! for e04abf
  integer :: maxcal !
  external e04abf, funct
  outp='optim_sigma.dat' ! an output file, which contains the information about optimal solution (optimal value of
  sigma(j)**2)
  open (19,file=outp,status='replace')
  outpreginf='regression.dat' ! an output file for regression coefficients
  open (23, file=outpreginf,status='replace')
  open (3, file='forecast_shrunken_reg.dat', status='replace')
  i=1
  rewind(9)
  write (nout,*) 'Minimizing minus log-likelihood'
  do
    coefsv(sch1)=zero
    do sch2=1,22
      coef_delta(sch1,sch2)=zero
    end do
    end do
    end do
  read_data_val: do
    read(9,*,end=99, err=98) jnumber_comp(line), month(line), junk, year(line), junk, junk, junk,
    junk, n_market_ret(line), n_monthly_ret(line), junk, monthly_ret(line), market_ret(line), share_price(line),
    divid(line), junk, t_volume(line), m_cap(line), two_m_r(line), three_m_r(line), four_m_r(line), six_m_r(line),
    one_y_r(line), two_y_r(line), three_y_r(line), four_y_r(line), five_y_r(line), alpha_p(line),
    beta_p(line), spec_risk(line), var_of_beta(line), div_yield(line), div_yield2y(line), div_yield3y(line)
    jnumber_comp(0)=jnumber_comp(line)! to read the first line
    if (jnumber_comp(line).eq.jnumber_comp(line-1)) then
      line=line+1
    else
      n_month(i)=line-1 ! number of month for each company
      line=1
      backspace(9) ! return to the previous line of datafile
    end if
  end do read_data_val
  s(i)=zero
  k(i)=zero
  n(i)=zero
l(i)=zero
t(i)=zero
q(i)=zero
m(i)=zero
z(i)=zero
p(i)=zero
doschl=l,22
w(schl)=zero
v(schl)=zero
u(schl)=zero
dosch2=l,22
ve(schl,sch2)=zero
end do
end do
num_month(i)=0
write(19,*)'n_month(i),n_month(i)
out=0
flag=.false.
do j=1,n_month(i)
if (month(j)<predictdate-120 .or. month(j)>predictdate)cycle
do line 1=8,30
! write(*,*)Var_name',var_name(linel),par1(linel),par2(linel),par3(linel),par4(linel)
select case (var_name(linel))
case (8)
monthly_ret(j)=(log((par2(line 1))*(monthly_ret(j)-par1(line 1)))+sqrt((par2(line 1)**2)*((monthly_ret(j)-parity(line 1))**2+1)))/par2(line 1)
monthly_ret(j)=(monthly_ret(j)-par3(line 1))/par4(line 1)
case (9)
market_ret(j)=(log((par2(line 1))*(market_ret(j)-par1(line 1)))+sqrt((par2(line 1)**2)*((market_ret(j)-par1(line 1))**2+1)))/par2(line 1)
market_ret(j)=(market_ret(j)-par3(line 1))/par4(line 1)
case (10)
share_price(j)=(((share_price(j)-parity(line 1))**parity(line 1)+parity(line 1)-1)/parity(line 1))
share_price(j)=(share_price(j)-parity(line 1))/par4(line 1)
case (11)
divid(j)=(log((par2(line 1))*(divid(j)-par1(line 1)))+sqrt((par2(line 1)**2)*((divid(j)-parity(line 1))**2+1)))/par2(line 1)
divid(j)=(divid(j)-parity(line 1))/par4(line 1)
case (13)
t_volume(j)=(((t_volume(j)-parity(line 1))**parity(line 1)+parity(line 1)-1)/parity(line 1))
t_volume(j)=(t_volume(j)-parity(line 1))/par4(line 1)
case (15)
two_m_r(j)=(log((par2(line 1))*(two_m_r(j)-parity(line 1)))+sqrt((par2(line 1)**2)*((two_m_r(j)-parity(line 1))**2+1)))/par2(line 1)
two_m_r(j)=(two_m_r(j)-parity(line 1))/par4(line 1)
case (16)
three_m_r(j)=(log((par2(line 1))*(three_m_r(j)-parity(line 1)))+sqrt((par2(line 1)**2)*((three_m_r(j)-parity(line 1))**2+1)))/par2(line 1)
three_m_r(j)=(three_m_r(j)-parity(line 1))/par4(line 1)
case (17)
four_m_r(j)=(log((par2(line 1))*(four_m_r(j)-parity(line 1)))+sqrt((par2(line 1)**2)*((four_m_r(j)-parity(line 1))**2+1)))/par2(line 1)
four_m_r(j)=(four_m_r(j)-parity(line 1))/par4(line 1)
case (18)
  six_m_r(j) = \left( \frac{\log(paret2(linel)) + \sqrt{(paret2(linel))^2 + \text{par1(linel)}}}{\text{par2(linel)}} \right) / \text{par2(linel)}
  six_m_r(j) = \frac{\text{six}_m_r(j) - \text{par3(line 1)}}{\text{par4(line 1)}}

case (19)
  one_y_r(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(linel)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  one_y_r(j) = \frac{\text{one}_y_r(j) - \text{par3(line 1)}}{\text{par4(line 1)}}

case (20)
  two_y_r(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(linel)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  two_y_r(j) = \frac{\text{two}_y_r(j) - \text{par3(line 1)}}{\text{par4(line 1)}}

case (21)
  three_y_r(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(linel)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  three_y_r(j) = \frac{\text{three}_y_r(j) - \text{par3(line 1)}}{\text{par4(line 1)}}

case (22)
  four_y_r(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(linel)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  four_y_r(j) = \frac{\text{four}_y_r(j) - \text{par3(line 1)}}{\text{par4(line 1)}}

case (23)
  five_y_r(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(linel)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  five_y_r(j) = \frac{\text{five}_y_r(j) - \text{par3(line 1)}}{\text{par4(line 1)}}

case (24)
  alpha_pr(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(linel)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  alpha_pr(j) = \frac{\text{alpha}_pr(j) - \text{par3(line 1)}}{\text{par4(line 1)}}

case (25)
  beta_pr(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(linel)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  beta_pr(j) = \frac{\text{beta}_pr(j) - \text{par3(line 1)}}{\text{par4(line 1)}}

case (26)
  spec_risk(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(line 1)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  spec_risk(j) = \frac{\text{spec_risk(j)} - \text{par3(line 1)}}{\text{par4(line 1)}}

case (27)
  var_of_beta(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(line 1)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  var_of_beta(j) = \frac{\text{var_of_beta(j)} - \text{par3(line 1)}}{\text{par4(line 1)}}

case (28)
  div_yield(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(line 1)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  div_yield(j) = \frac{\text{div_yield(j)} - \text{par3(line 1)}}{\text{par4(line 1)}}

case (29)
  div_yield2y(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(line 1)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  div_yield2y(j) = \frac{\text{div_yield2y(j)} - \text{par3(line 1)}}{\text{par4(line 1)}}

case (30)
  div_yield3y(j) = \left( \frac{\log(paret2(line 1)) + \sqrt{(paret2(line 1))^2 + \text{par1(line 1)}}}{\text{par2(line 1)}} \right) / \text{par2(line 1)}
  div_yield3y(j) = \frac{\text{div_yield3y(j)} - \text{par3(line 1)}}{\text{par4(line 1)}}
\[ n(i) = n(i) + n\left(\text{monthly ret}(j)\right) \]
\[ m(i) = m(i) + n\left(\text{monthly ret}(j)\right) \]

\[ r(i) = r(i) + \delta(1)\text{monthly ret}(j) + \delta(2)\text{market ret}(j) + \delta(3)\text{share price}(j) + \delta(4)\text{divid}(j) + \delta(5)\text{t volume}(j) \]
\[ + \delta(6)\text{m cap}(j) + \delta(7)\text{two r}(j) + \delta(8)\text{three r}(j) + \delta(9)\text{four r}(j) + \delta(10)\text{six r}(j) + \delta(11)\text{one y r}(j) + \delta(12)\text{two y r}(j) + \delta(13)\text{three y r}(j) + \delta(14)\text{four y r}(j) + \delta(15)\text{fi r}(j) + \delta(16)\text{alpha pr}(j) + \delta(17)\text{beta pr}(j) + \delta(18)\text{spec risk}(j) + \delta(19)\text{var of beta}(j) + \delta(20)\text{div yield}(j) + \delta(21)\text{div yield2y}(j) + \delta(22)\text{div yield3y}(j) \]

\[ q(i) = q(i) + \delta(1)\text{monthly ret}(j) + \delta(2)\text{market ret}(j) + \delta(3)\text{share price}(j) + \delta(4)\text{divid}(j) + \delta(5)\text{t volume}(j) \]
\[ + \delta(6)\text{m cap}(j) + \delta(7)\text{two r}(j) + \delta(8)\text{three r}(j) + \delta(9)\text{four r}(j) + \delta(10)\text{six r}(j) + \delta(11)\text{one y r}(j) + \delta(12)\text{two y r}(j) + \delta(13)\text{three y r}(j) + \delta(14)\text{four y r}(j) + \delta(15)\text{fi r}(j) + \delta(16)\text{alpha pr}(j) + \delta(17)\text{beta pr}(j) + \delta(18)\text{spec risk}(j) + \delta(19)\text{var of beta}(j) + \delta(20)\text{div yield}(j) + \delta(21)\text{div yield2y}(j) + \delta(22)\text{div yield3y}(j) \]

\[ z(i) = z(i) + \delta(1)\text{monthly ret}(j) + \delta(2)\text{market ret}(j) + \delta(3)\text{share price}(j) + \delta(4)\text{divid}(j) + \delta(5)\text{t volume}(j) \]
\[ + \delta(6)\text{m cap}(j) + \delta(7)\text{two r}(j) + \delta(8)\text{three r}(j) + \delta(9)\text{four r}(j) + \delta(10)\text{six r}(j) + \delta(11)\text{one y r}(j) + \delta(12)\text{two y r}(j) + \delta(13)\text{three y r}(j) + \delta(14)\text{four y r}(j) + \delta(15)\text{fi r}(j) + \delta(16)\text{alpha pr}(j) + \delta(17)\text{beta pr}(j) + \delta(18)\text{spec risk}(j) + \delta(19)\text{var of beta}(j) + \delta(20)\text{div yield}(j) + \delta(21)\text{div yield2y}(j) + \delta(22)\text{div yield3y}(j) \]

\[ p(i) = p(i) + \delta(1)\text{monthly ret}(j) + \delta(2)\text{market ret}(j) + \delta(3)\text{share price}(j) + \delta(4)\text{divid}(j) + \delta(5)\text{t volume}(j) \]
\[ + \delta(6)\text{m cap}(j) + \delta(7)\text{two r}(j) + \delta(8)\text{three r}(j) + \delta(9)\text{four r}(j) + \delta(10)\text{six r}(j) + \delta(11)\text{one y r}(j) + \delta(12)\text{two y r}(j) + \delta(13)\text{three y r}(j) + \delta(14)\text{four y r}(j) + \delta(15)\text{fi r}(j) + \delta(16)\text{alpha pr}(j) + \delta(17)\text{beta pr}(j) + \delta(18)\text{spec risk}(j) + \delta(19)\text{var of beta}(j) + \delta(20)\text{div yield}(j) + \delta(21)\text{div yield2y}(j) + \delta(22)\text{div yield3y}(j) \]

\[ k(i) = k(i) + \left(\text{market ret}(j)\right)^2 \]
\[ n(i) = n(i) + \text{monthly ret}(j) \]
\[ l(i) = l(i) + \text{monthly ret}(j) \]
\[ m(i) = m(i) + \left(\text{monthly ret}(j)\right)^2 \]

\[ \text{ve of predx}(1) = \text{monthly ret}(j) \]
\[ \text{ve of predx}(2) = \text{market ret}(j) \]
\[ \text{ve of predx}(3) = \text{share price}(j) \]
\[ \text{ve of predx}(4) = \text{divid}(j) \]
\[ \text{ve of predx}(5) = \text{t volume}(j) \]
\[ \text{ve of predx}(6) = \text{m cap}(j) \]
\[ \text{ve of predx}(7) = \text{two r}(j) \]
\[ \text{ve of predx}(8) = \text{three r}(j) \]
\[ \text{ve of predx}(9) = \text{four r}(j) \]
\[ \text{ve of predx}(10) = \text{six r}(j) \]
\[ \text{ve of predx}(11) = \text{one y r}(j) \]
\[ \text{ve of predx}(12) = \text{two y r}(j) \]
\[ \text{ve of predx}(13) = \text{three y r}(j) \]
\[ \text{ve of predx}(14) = \text{four y r}(j) \]
\[ \text{ve of predx}(15) = \text{five y r}(j) \]
\[ \text{ve of predx}(16) = \text{alpha pr}(j) \]
\[ \text{ve of predx}(17) = \text{beta pr}(j) \]
\[ \text{ve of predx}(18) = \text{spec risk}(j) \]
\[ \text{ve of predx}(19) = \text{var of beta}(j) \]
\[ \text{ve of predx}(20) = \text{div yield}(j) \]
\[ \text{ve of predx}(21) = \text{div yield2y}(j) \]
\[ \text{ve of predx}(22) = \text{div yield3y}(j) \]

do sch1=1,22  ! calculates values of parameters for each predictor
w(sch1)=w(sch1)+n\left(\text{monthly ret}(j)\right)\text{ve of predx}(sch1)
\text{ve of predx}(sch1)=\text{ve of predx}(sch1)+\text{ve of predx}(sch1)\text{ve of predx}(sch1)\]
end do
write(19,*)'**************************
if (num_month(i)<60) cycle
write( 19,*)'num_month', num_month(i)
write(19,*)i,s(i),r(i),q(i),ve(5,2),v(4)
eps = 0.00D0 ! eps and t will be set by default
t=0.00D0
! The minimum is known lie in the range (0.00001, 0.4)
a=0.00001D0
b=0.4D0
! Allow 25 calls of funct
maxcal=25
ifail=1
call e04abf(func,eps,t,a,b,maxcal,x,f,ifail)
write(19,*)
if (ifail.eq.1) then
   write(19,*) 'Parameter outside expected range'
else
   if (ifail.eq.2) then
      write(19,*) 'Results after', maxcal, 'function evaluations are'
   end if
   write(19,*) 'Company sequence number',jnumber_comp(line),'Company number',i
   write(19,*) 'The minimum lies in the interval', a, ', b', b
   write(19,*) 'Its estimated position is', x, ','
   write(19,*) 'where the function values is', f
   write(19,*) maxcal, 'function evaluations were required'
   optim_sigsq(i)=x ! writing the result for each company into array
end if
optim_alpha(i)=(r(i)*(k(i)+v_beta*optim_sigsq(i)))-s(i)*q(i) + n(i)*s(i) + beta_*s(i)*v_beta*optim_sigsq(i)
& +&q(i) - (k(i)+optim_sigsq(i)*v_beta)*(l(i)+alpha_*v_alp& &ha)*optim_sigsq(i))/(s(i)**2 - (num_month(i)+optim_sigsq(i)*v_alpha)*(k(i)+o& &ptim_sigsq(i)*v_beta))
! optim_beta(i)=(q(i)*(num_month(i)+v_alpha*optim_sigsq(i))-r(i)*s(i)+l(i)*q(i)+beta_*s(i)*v_alpha)*optim_sigsq(i)
! &*(l(i)+beta_*v_alpha*optim_sigsq(i))/(s(i)**2 - (num_month(i)+v_alpha*optim_sigsq(i)*v_alpha)*(k(i)+o& &ptim_sigsq(i)*v_beta))
! &t=optim_sigsq(i))
optim_alpha(i)=(+i(i)+beta_*v_alpha*optim_sigsq(i)-r(i)-optim_beta(i)*s(i))/(n& &um_month(i)+v_alpha*optim_sigsq(i))
&optim_alpha(i)=(+i(i)+beta_*v_alpha*optim_sigsq(i)-r(i)-optim_beta(i)*s(i))/(n& &um_month(i)+v_alpha*optim_sigsq(i))
if (month(n_month(i)).lt.predictdate) then
   write(3,*)IDeadcomp'jnumber_comp(line)
else if (.not.flag .and. month(n_month(i)).ge.predictdate)then
   write(3,*)'Alive company, but there is a gap in the data'jnumber_comp(line)
else
   forecast(i)=optim_alpha(i)+optim_beta(i)*n_market_ret(out)+delta(1)*monthly_re& &t(out)+delta(2)*market_ret(out)+de& &lt(a3)*share_price(out)+delta(4)*divid(out)+delta(5)*v_vol& &cume(out)+delta(6)*n_cap(out)+delta(7)*two_m_r(out)+delta(8)*three& &e_m_r(out)+delta(9)*four_m_r(out)+delta(10)*six_m_r(out)+de& &lt(a11)*one_y_r(out)+delta(12)*two_y_r(out)+delta(13)*thre& &e_v_r(out)+delta(14)*five_y_r(out)+delta(15)*five_y_r(out)+de& &lt(a16)*alpha_pr(out)+delta(17)*beta_pr(out)+delta(18)*spec_r& &n+&&um(out)+delta(19)*var_of_beta(out)+delta(20)*div yields(out)+de& &lt(a21)*div yields2(out)+delta(22)*div yields3(out)
   write(3,i5,4(gl8.9))jnumber_comp(line),n_monthly_ret(out),forecast(i),optim_beta(i),optim_sigsq(i)
end if
write(23,*)
}
write(23,*)'Company sequence number ',number_comp(line),'Company number',i
write(23,*)'Period'
write(23,*)'Coeff. alpha', optim_alpha(i)
write(23,*)'Coeff. beta', optim_beta(i)
write(23,*)'Sigma squared', optim_sigsq(i)

n_comp=i ! number of companies

! calculating coefficients before delta (for each predictor - for each equation in thr system)

coefsv(schl)=coefsv(schl)+(l/optim_sigsq(i))*(w(schl)-optim_alpha(i)*v(schl)-optim_beta(i)*u(schl))
do schl=1,22
coef_delta(schl,sch2)=coef_delta(schl,sch2)+(l/optim_sigsq(i))*ve(schl,sch2)
end do

end do

i=i+1
end do

stop

99 write(*,543) n_comp
543 format('Number of companies under observation',i8)
return

98 write(*,'("Error on line ",i2," of data file ",a)')line,infile(:len_trim(infile))
stop 'failed'
end subroutine iter_comp_by_comp

subroutine get_stem(is_ok)

use stuff
implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
logical is_ok
!
! This little routine finds a filestem, ie the part before the
! postfix, if any.
!
is_ok=.false.
l=index(infile,1.)
if(l = 0)return ! no postfix, just exit with error flag set
file_stem=infile(:l-l)
!
end subroutine get_stem

subroutine read_params_of_prior_dist

! Reads all starting values
! if we assume that all deltas are equal to zero and special values for other
! parameters we can compare the results, for example, with resalts derived on using
! Minitab
!
use stuff
use specific
use params
implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
logical is_ok
do
write(*,'("Type filename which contains parameters of prior distributions: ",a)')line,infile(:len_trim(infile))
read(*,*) infile
if(infile = 'NONE') stop 'failed'
call get_stem(is_ok)
if(.not.is_ok)then
write(*,'("Filename must have .con postfix—try again or type NONE")')
cycle
end if
!
exit
end do
open(14,file=infile,status='old',err=1)
read(14,*) start_val(1)
read(14,*) start_val(2)
read(14,*) start_val(3)
read(14,*,err=98,end=99)start_val(4)
read(14,*,err=98,end=99)start_val(5)
read(14,*,err=98,end=99)start_val(6)
read(14,*,err=98,end=99)start_val(7)

line=1
do

read(14,*,err=98,end=99)delta(line) reading starting values of deltas
line=line+1

end do

write(*,'("Unable to open input file ",a)')infile(:len_trim(infile))
stop 'failed'
write(*,'("Error on line hii ",i2," of commands file ",a)')line(infile(:len_trim(infile)))
stop 'failed'
continue
close(14)
return
end subroutine read_params_of_prior_dist

subroutine funct(xc,fc)
! Routine to evaluate \(-l(\sigma(i)^2)\) at any point in \((a,b)\)
use stuff
use specific
use params
implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
double precision:: fc
double precision :: xc ! \(\sigma(i)^2\)
double precision :: alpha_it ! \(\alpha(i)\)
double precision :: beta_it ! \(\beta(i)\)
double precision :: pi
intrinsic log
pi=3.14159265358
alpha_it=(r(i)*(k(i)+v_beta*xc)-s(i)*q(i) + n(i)*s(i) + beta_*s(i)*v_beta*xc - (k(i)+xc*v_beta)*(l(i)+alpha_*v_alp&
&ha_it))(s(i)**2 - (num_month(i)+xc*v_alpha)*(k(i)+xc*v_beta))
! beta_it=(q(i)*(num_month(i)+v_alpha*xc)-r(i)*s(i)+l(i)*s(i)+alpha_*s(i)*v_alpha*xc-(num_mo&
! &nth(i)+v_alpha*xc))*(n(i)+beta*_v_beta*xc))/((s(i)**2-(num_month(i)+v_alpha*xc)*(k(i)+v_beta*xc))
! beta_it=(l(i)-r(i)*alpha_it*num_month(i))/xc**v_alpha*(diff_alpha_it)/s(i)
! alpha_it=(l(i)+alpha_*v_alpha*xc-r(i)-beta_it*s(i))/(num_month(i)+v_alpha*xc)

beta_it=(n(i)+beta_*v_alpha*xc-q(i)-alpha_it*s(i))/(l(i)+alpha_*v_alpha*xc)
! minus log-likelihood
fc=-(-num_month(i)*half*log(two*pi*xc)-half*(l/xc)*(m(i)-two*alpha_it*l(i)-two*z(i)-
two*beta_it+n(i)+two*alpha_it*alpha_it-it*s(i))/(l(i)+alpha_*v_alpha*xc)-half*log(two*pi*xc)-pi*alpha_it-it*s(i))/(l(i)+alpha_*v_alpha*xc)
! + half*log(two*pi*xc)-pi*alpha_it-it*s(i))/(l(i)+alpha_*v_alpha*xc)
return
end subroutine deltas

use stuff
use specific
use params
use forf04atf
external f04atf
outp_^='reg_deltas.dat'! an output file for deltas
open~(22,file=outp_,status='replace')
outpsterr='sterr_deltas.dat'! an output file for standart errors for deltas
open(24,file=outpsterr,status='replace')
np=22 ! number of predictors in vector x
write(23,*)'coefdelta',coef_delta(5,2)

do schl=1,22
  extra_term(schl)=0
end do
do schl=1,22
  extra_term(schl)=1
end do
ar(schl,scli2)=coef_delta(schl,scli2)+v_delta*extra_term(sch2) ! rearranging information according end do
b(sch1)=coefs(sch1)
extra_term(sch1)=0
end do
ifail=0
call f04atf(ar,ia,b,np,delta,aa,iaa,
vksl,wks2,ifail) !Calling Nag
write(22,*)'Solution of system of simultaneous equations for delta'
write(22,*)
doen 1=1,22
write(22,*)'coefficient delta',schl,delta(schl)
write(24,*)The standart error of deltas'
! call standart_err_deltas
! write(23,*)ar(l,l)
end do
end subroutine deltas

subroutine transformation
use fortransformation
use stuff
use specific
use params
implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
dok_tr=8,30 !8,30
if(k_tr.eq.12)then
cycle
end if
rewind(9)
j_tr=0
num_miss=0
do k_tr=8,30
  if (k_tr.eq.12) then
    cycle
  end if
  rewind(9)
  i_tr=0
  num_miss=0
  do
    read(9,*,end=10)junk_tr,month_tr,junk_tr,year_tr,(a_temp(i_tr),i_tr=1,30)
    if (month_tr<predictdate-120 .or. month_tr>predictdate-1) cycle
    if (abs(a_temp(k_tr)-xmiss)<1.D-06) then
      num_miss=num_miss+1
      cycle
    end if
    j_tr=j_tr+1
    x_tr(j_tr)=a_temp(k_tr)
  end do
 10 continue
  n_tr=j_tr
  lim=(0.7d+00)*(n_tr+num_miss)
  if (n_tr.le.lim) then
    cycle
  end if
  mean_x=0
! write(19,*n_mean_x,disp_x,st_div_x)
!
if (k_tr.ne.10 .and. k_tr.ne.13 .and. k_tr.ne.14 .and. k_tr.ne.26 .and. k_tr.ne.27 .and. k_tr.ne.28 .and. k_tr.ne.29 .and. k_tr.ne.30) then
  opt_tr=1.0+10
  h1=1/(6*(st_div_x))
  h2=4*st_div_x/30
do c_tr=1.06,5/st_div_x,h1
  do a_tr=-2*st_div_x,2*st_div_x,h2
    mean_y=0
    doljr=1,njr
    yjr(ljr)=log(cjr*(xjr(ljr)-a_tr)+sqrt((cjr**2)*((xjr(ljr)-a_tr)**2)+1))/c_tr
    mean_y=mean_y+yr(ljr)/njr
  end do
  disp_y=0
  jacobian=0
  doljr=1,njr
  disp_y=disp_y+((yr(ljr)-mean_y)**2)/njr
  jacobian=jacobian+log(1/(sqrt(1+(cjr**2)*((xjr(ljr)-a_tr)**2))))
end do
Hkelih=-(njr/2)*log(2*pi*disp_y)-(njr/2)+jacobian
if likelih>optj then
  optj=likelih
  opt_a=a_tr
  opt_c=cjr
  opt_mean_y=mean_y
  opt_stdiv_y=sqrt(disp_y)
end if
! write(19,*a,c,likelih)
end do
! write(*,19,*a,c,likelih)
end if
disp_y_BC = disp_y_BC + ((y_tr(l_tr) - mean_y_BC)**2)/n_tr
jacobian_BC = jacobian_BC + log(xjr(l_tr) + lambda_tr)
end do

! 5^|ih_BC = -(njr/2)*log(2*pi*disp_y_BC)-(n_tr/2)+(alpha_tr-l)*jacobian_BC
if (hkelih_BC > opt_l_BC) then
  opt_l_BC = likelih_BC
  opt_alpha = alpha_tr
  opt_lambda = lambda_tr
  opt_mean_y_BC = mean_y_BC
  opt_stdiv_y_BC = sqrt(disp_y_BC)
end if
write(19, *) alpha, lambda, likelih_BC
end do
end subroutine transformation

subroutine do_iteration
  use stuff
  use specific
  use params
  implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
  integer:: nin=5, nout=6
  integer: :nmax,nclmax,ncnmax
  parameter(nmax=max_params,nclmax=10,ncnmax=10)
  integer:: lda,ldcj,ldr
  parameter(lda=nclmax,ldcj=ncnmax,ldr=nmax)
  integer:: liwork,lwork
  parameter(liwork=100,lwork=1000)
  doubleprecision :: ajiypo(lda,nmax),bl(nmax+nclmax+ncnmax),bu(nmax+nclmax+ncnmax),c(ncnmax),cjac(ldcj,nmax),clamda(nmax+nclmax+ncnmax),objgrd(nmax),re(ldr,nmax),user(1),work(lwork),x_hypo(nmax)
  integer:: istate(nmax+nclmax+ncnmax),iuser(1),iwork(liwork)
  character(80):: text
  external e04udf,e04ucf,objfun
  print=.false.
  nclin=0
  ncnln=0
  open(10,file='hypoparams.dat',status='replace')
call compress(x_hypo,bl,bu,n_float)
do itnum= 1 ,n_restarts+1
  call perturb(x_hypo,n_float,itnum) Ichange x_hypo by random amount
  write(10,*)'Number of random restart is',itnum
  call objfun(mode,n_float,x_hypo,objf,objgrd,nstate,iuser,user)
  write(10,300)objf,n_float
  300 format('before iteration funcion is',gl4.8,'with',i3,'parameters floated1)
  if(n_float.le.nmax.and.nclin.le.nclmax.and.ncnln.le.ncnmax)then
    ifail=-l
    ifail=l
    ! call e04uefCNolist")! for no printing
    ! call e04uefCDerivative level = O1)
    ! call e04uefCMajor print level = 101)
    ! call e04uefCMajor print level = 0')! for no printing
    call e04uefCMonitoring file = 10')
call e04uef('Step limit = 0.2")
text='Major iteration limit = '
l_print=len_trim(text)
write(text(l_print+2:),'(i6)')nits
call e04uef(text)
call e04ucf (n_float, nclin, ncnln, Ida, Idcj, ldr, a_hypo, bl, bu, n_float, objfun, iter, istate, c, cjac, clamda, objf, o&
&bjgrd, re, x_hypo, iwork, liwork, work, hwork, iuser, user, ifail)

end if

call uncompress(x_hypo)
if(itnum==1)then
  fbest=objf
  best_vals=start_val
  ibest=itnum
end if
if(objf<fbest)then
  write(10,401)fbest,objf,itnum
  401 format('best function value improved from ',f12.4,' to ',f12.4,' on restart',i3)
  fbest=objf
  best_vals=start_val
  ibest=itnum
end if
end do
start_val=best_vals
objf=fbest

end subroutine do_iteration

subroutine compress(x_hypo,bl,bu,n_float)
use stuff
implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
!
compresses model variables into array X for NAG minimiser.
!

doubleprecision :: x_hypo(max_params),bl(max_params),bu(max_params)

j=0
do i=1,7
  if (to_be_iterated(i) == 1) cycle
  if(i.eq.5.or.i.eq.6.or.i.eq.2.or.i.eq.4.or.i.eq.7) then
    bl(j)=-50000.
  else
    bl(j)=-50000.
  end if
  bu(j)=50000.
end do

! transform first?
if(to_be_trans(i) == 1) then
  x_hypo(i)=log(start_val(i))
else if(to_be_trans(i) == 0) then
  x_hypo(i)=start_val(i)
end if
end do

n_float=j

end subroutine compress

subroutine uncompress(x_hypo)
use stuff
implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)

doubleprecision :: x_hypo(max_params)
!
uncompresses model variables from array X used by NAG minimiser.
j=0
do i=l,7
if(to_be_iterated(i) == 0) cycle
j=j+1
! transform first?
if(to_be_trans(i) == 1) then
start_val(i)=exp(x_hypo(j))
else if(to_be_trans(i) == 0) then
start_val(i)=x_hypo(j)
end if
end do
end subroutine uncompress

subroutine perturb(x_hypo,n,itnum)
use stuff
implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
real*8 :: x_hypo(*)
if(itnum= l) return
do k=1,n
! if(k.eq.5.or.k.eq.6.or.k.eq.2.or.k.eq.4.or.k.eq.7)then
! y_hypo=log(x_hypo(k))
! y_hypo=y_hypo+0.2d+00*(random()-0.5d+00)*abs(y_hypo)
! x_hypo(k)=exp(y_hypo)
! else
x_hypo(k)=x_hypo(k)+0.2d+00*(random()-0.5d+00)*abs(x_hypo(k))
! end if
end do
end subroutine perturb

subroutine objfun (mode, n, x_hypo, objf, objgrd, nstate, iuser, user)
! Routine to evaluate objective function and its 1st derivatives.
use stuff
implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
doubleprecision :: objf
integer :: mode, n, nstate
doubleprecision :: objgrd (n), user ( * ), xjiypo (n)
integer :: iuser (*)
call funct_hypo(n,x_hypo,objf)
end subroutine objfun

subroutine funct_hypo(n_float,x_hypo,objf)
use stuff
use specific
use params
implicit doubleprecision (a-h,o-z), integer (kind=3) (i-n)
real*8 :: x_hypo(*)
pi=3.14159265358
! This routine calculates the objective function objf
! (minus the log-likelihood)
! Note that call to uncompress, calculation of n-data
! and cycling on group variable MUST be present.
! 
! call uncompress(x_hypo) ! uncompress X back into start_val array
! compute minus log-likelihood
alpha_=start_val(1)
sigma_alpha_=start_val(2)
beta_=start_val(3)
sigma_beta_=start_val(4)
alpha=start_val(5)
beta=start_val(6)
sigma_delta=start_val(7)
do i=1,7
write(10,*)'start_val',i,start_val(i)
end do
v_alpha=1/((sigma_alpha_j)**2)
v_beta=1/((sigma_beta_j)**2)
v_delta=1/((sigma_delta)**2)
ifailgam=1
gammaf=14aaf(beta,ifailgam)
do schl=1,22
delta(schl)=zero
end do
ideltas=1
do
if (ideltas.gt.12)exit
call iter_comp_by_comp
call deltas
objf=0.0
end do
write(10,*)'objf,ideltas,objf
objf=0.0
write(10,*)'before dobavka',ideltas,objf
dobavka=0
end do
write(10,*)'after dobavka',ideltas,objf
write(10,*)'*************
end do
end subroutine funct_hypo

Appendix 11: Computation of PL - reducing rounding error

In the text of the chapter we discussed that in order to ameliorate the rounding error problem when calculating statistic $PD$, equations (6.5)-(6.7), appearing in the text, have to be rearranged. In the text we presented the rearranged formulas for the three components of $PD$. In the appendix we will derive these formulas. In addition to the notation introduced in the main body of the chapter let $w_i = p\beta_i$ and $z_i = q\beta_i$.

Then

$$ r_0 V^{-1} \hat{r} = (w + \delta)^T V^{-1} (z + \xi) $$
and expanding the brackets gives us

\[ \mathbf{r}_0 \mathbf{V}^{-1} \hat{\mathbf{r}} = \mathbf{w}^T \mathbf{V}^{-1} \mathbf{z} + \mathbf{\delta}^T \mathbf{V}^{-1} \mathbf{z} + \mathbf{w}^T \mathbf{V}^{-1} \mathbf{\xi} + \mathbf{\delta}^T \mathbf{V}^{-1} \mathbf{\xi}, \]  
(A11.1)

where \( \mathbf{w} = (w_i), \mathbf{z} = (z_i), \mathbf{\delta} = (\delta_i) \) and \( \mathbf{\xi} = (\xi_i) \).

Substituting \( \mathbf{V}^{-1} \) from equation (6.3), the first component in the right-hand side of equation (A11.1) becomes:

\[ \mathbf{w}^T \mathbf{V}^{-1} \mathbf{z} = \rho \sum_i \beta_i^2 / \sigma_i^2 - \kappa \rho \left( \sum_i \beta_i^2 / \sigma_i^2 \right)^2, \]

where \( \kappa \) is given by formula (6.4). Substituting for \( \kappa \) in equation above yields

\[ \mathbf{w}^T \mathbf{V}^{-1} \mathbf{z} = \frac{\rho \sum_i \beta_i^2 / \sigma_i^2}{1 + \sigma_m^2 \sum_i \beta_i^2 / \sigma_i^2}. \]

The second component in the right-hand side of (A11.1)

\[ \mathbf{\delta}^T \mathbf{V}^{-1} \mathbf{z} = \rho \sum_i \delta_i \beta_i / \sigma_i^2 - \kappa \rho \left( \sum_i \beta_i^2 / \sigma_i^2 \right) \left( \sum_i \beta_i \delta_i / \sigma_i^2 \right) \]

is zero since \( \sum_i \delta_i \beta_i / \sigma_i^2 = 0 \) by construction.

Analogously \( \mathbf{w}^T \mathbf{V}^{-1} \mathbf{\xi} = 0 \) and

\[ \mathbf{\delta}^T \mathbf{V}^{-1} \mathbf{\xi} = \sum_i \delta_i \xi_i / \sigma_i^2 - \kappa \left( \sum_i \beta_i \delta_i / \sigma_i^2 \right) \left( \sum_i \beta_i \xi_i / \sigma_i^2 \right) = \sum_i \delta_i \xi_i / \sigma_i^2. \]

Therefore

\[ \mathbf{r}_0 \mathbf{V}^{-1} \hat{\mathbf{r}} = \frac{\rho q \sum_i \beta_i^2 / \sigma_i^2}{1 + \sigma_m^2 \sum_i \beta_i^2 / \sigma_i^2} + \sum_i \delta_i \xi_i / \sigma_i^2. \]

Substituting for \( \rho \) and \( q \) in equation above yields
Formulas (6.8) and (6.9) can be derived by applying similar logic.

Appendix 12: Proofs of the properties of the new measure of investment performance (the value of risk aversion that makes investment valueless)

We show that the performance statistic \( \eta_0 \) has the desirable property that
\[
\frac{\partial \eta_0}{\partial z_i} > 0,
\]
so that it increases whenever one of the \( n \) investment gains increases.

We first show that \( dC(\eta) / d\eta < 0 \). It is convenient to write
\[
C = -\log E[\exp(-\eta Z)]/\eta,
\]
where \( E \) denotes expectation.

Then
\[
dC / d\eta = \frac{1}{\eta} \frac{E\{Z \exp(-\eta Z)\}}{E\{\exp(-\eta Z)\}} + \frac{1}{\eta^2} \log E\{\exp(-\eta Z)\}. \tag{A12.1}
\]

Now
\[
E\{\exp(\eta Z) \exp(-\eta Z)\} = 1
\]

By expanding \( \exp(\eta Z) \) in Taylor series
\[
\exp(\eta Z) = 1 + \eta Z + \frac{(\eta Z)^2}{2} + ...
\]

Hence
\[
E\{\exp(-\eta Z)\} + E\{\eta Z \exp(-\eta Z)\}/1! + E\{(\eta Z)^2 \exp(-\eta Z)\}/2! = 1
\]
From equation above

\[
1 / E\{\exp(-\eta Z)\} = 1 + \frac{E\{\eta Z \exp(-\eta Z)\}}{E\{\exp(-\eta Z)\}} + \frac{E\{(\eta Z)^2 \exp(-\eta Z)\}}{2E\{\exp(-\eta Z)\}} + ...
\]

By Jensen’s inequality, since

\[
\frac{E\{(\eta Z)^m \exp(-\eta Z)\}}{E\{\exp(-\eta Z)\}} \geq \left[ \frac{E\{(\eta Z) \exp(-\eta Z)\}}{E\{\exp(-\eta Z)\}} \right]^m,
\]

we have that

\[
1 / E\{\exp(-\eta Z)\} \geq 1 + \frac{E\{\eta Z \exp(-\eta Z)\}}{E\{\exp(-\eta Z)\}} + \frac{1}{2} \left[ \frac{E\{(\eta Z) \exp(-\eta Z)\}}{E\{\exp(-\eta Z)\}} \right]^2 + ...
\]

Using Maclaurin series expansion

\[
1 / E\{\exp(-\eta Z)\} \geq \exp\left[ \frac{E\{(\eta Z) \exp(-\eta Z)\}}{E\{\exp(-\eta Z)\}} \right].
\]

Finally,

\[
\log E\{\exp(-\eta Z)\} \leq - \frac{E\{(\eta Z) \exp(-\eta Z)\}}{E\{\exp(-\eta Z)\}}, \quad (A12.2)
\]

and from (A12.1), \(dC / d\eta < 0\). This is only an equality when the distribution of returns is a delta function.

Turning to the proof that \(\partial \eta_0 / \partial z_i > 0\), when \(C = 0\),

\[
f = \sum_{i=1}^{n} \exp(-\eta_0 z_i) - n = 0.
\]

Differentiating, we have that

\[
\partial \eta_0 / \partial z_i = -(\partial f / \partial z_i) / (\partial f / \partial \eta_0).
\]

This yields
\[ \frac{\partial \eta_0}{\partial z_i} = \frac{-\eta_0 \exp(-\eta_0 z_i)}{\sum_{j=1}^{n} z_j \exp(-\eta_0 z_j)} . \]

When \( E\{\exp(-\eta Z)\} = 1 \), from (A12.2) \( E\{Z \exp(-\eta Z)\} < 0 \) and hence \( \frac{\partial \eta_0}{\partial z_i} > 0 \), QED.