



UNIVERSITY OF CAPE TOWN

Department of Statistical Sciences

Thin Trading, Non-normality and the Estimation of Systematic Risk on Small Stock Markets

David Colin Bowie

A thesis prepared under the supervision of Associate
Professor D.J. Bradfield in fulfilment of the requirements
for the Degree of Doctor of Philosophy.

Copyright by the University of Cape Town
1994

The University of Cape Town has been given
the right to reproduce this thesis in whole
or in part. Copyright reserved by the author.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

To my parents and grandmother

ABSTRACT

This thesis examines and extends research into two of the most important attributes of small stock markets, namely thin trading and non-normality, and how they impact on the estimation of systematic risk. Bearing these points in mind, the primary objective of the thesis is to offer concrete suggestions for selecting estimators of beta coefficients.

In order to attain the objective outlined above, the first steps are to establish the extent and to model the characteristics of thin trading and non-normality. This is achieved in the thesis with the aid of empirical investigations using data from the Johannesburg Stock Exchange.

The second step in providing concrete recommendations for beta estimation is to contrast beta estimators with a view to assessing their relative abilities to counteract the effects of thin trading and non-normality. In the thesis the comparison of both existing and modified versions of the estimators is performed both empirically and in simulation studies. In particular, the improved efficiency of robust estimators of beta coefficients is demonstrated using a comprehensive array of robust estimators (some of which incorporate the concept of bounding the influence of outlying market returns). Furthermore, the findings presented in the thesis confirm that further efficiency in the estimation of beta coefficients can be accomplished by implementing a well-known Bayesian adjustment in conjunction with the trade-to-trade approach of avoiding the biases caused by thin trading. Using unbiasedness and efficiency as the necessary criteria of estimators of beta, the thesis concludes that a robust, bounded-influence version of the trade-to-trade beta estimator should be implemented on small markets.

A novel contribution in further modifying the proposed estimators is the consideration of flexible interval lengths over which to measure the security and market returns. The adaptive data selection procedure recommended in the thesis arises from the basic premise that there is more information in stock market prices in each month than is available at the month-end. Substantial increases in the efficiency of beta estimation are shown to emanate from the proposed procedure in an empirical evaluation using data from the JSE.

A further aspect of thin trading is also considered in the thesis, namely the effect that it has on the estimation of covariances between securities. It is argued that the covariances between the return series of thinly trading securities cannot readily be estimated using a trade-to-trade approach because of the difficulties in matching the trading times of two thinly traded securities. Therefore, an aggregated coefficients estimator is more appropriate. In order to implement the aggregated coefficients estimator efficiently without sacrificing unbiasedness completely, a practical solution to the selection of the number of leading and lagged coefficients is proposed. A complication involved in using the aggregated coefficients estimator is shown to exist when there is autocorrelation in the market returns (over and above that induced by thin trading in the component securities). An analytical expression is therefore derived to ascertain the extent to which the aggregated coefficients estimators of covariances are biased upwards by the serial correlation. The development of the analytical expression also enables a correction procedure for the upward bias to be proposed. An extension of the methodology using a continuous-time framework is developed to widen the applicability of the technique into the well traded markets where the durations of pricing delays are generally shorter than one day.

In the light of the evidence regarding the stability of beta coefficients highlighted in the thesis and in other literature, a robust, multivariate, state-space approach to beta estimation is proposed. The new information to be gained from knowing the correlations between any movements in the beta coefficients of securities from different sectors facilitates the interpretation of why and how beta coefficients vary over time. For the practitioner, the results of an empirical study conducted on the JSE

demonstrate that fresh insight into the management of systematic risk in portfolios can be gained from the multivariate approach.

CONTENTS

ACKNOWLEDGEMENTS

§1 INTRODUCTION AND SYNOPSIS	1-1
The estimation of systematic risk	1-1
Organization of the thesis	1-2
§2 A REVIEW OF SYSTEMATIC RISK ESTIMATION	2-1
The evolution of beta estimation and the Market Model	2-1
The impact of thin trading on beta estimation	2-4
Efficiency of estimation and the issue of non-normality	2-8
Improving the predictive power of estimators through the inclusion of prior information	2-11
Models and tests for time-varying beta coefficients	2-12
The random coefficients model	2-14
The random walk model	2-15
The AR(1) model	2-16
Other models which have been proposed to incorporate beta instability	2-17
Estimation of time-varying beta models	2-18
§3 THE THINLY TRADED ENVIRONMENT OF THE JOHANNESBURG STOCK EXCHANGE	3-1
Introduction	3-1
Beta estimation for infrequently traded shares	3-3
Correcting for thin trading bias	3-5
The effectiveness of the trade-to-trade estimator when weekly data are available	3-7
A simple model of thin trading on the JSE	3-8
The distribution of thin trading parameters on the JSE	3-9
A comparison of the modelled and empirical characteristics of thin trading	3-10
The month-end price ages	3-11
The number of days between the last traded days in successive months	3-12
Assessing the model of thin trading and empirical values using actual beta estimates	3-13
The case when only weekly data are available for beta estimation	3-14
Stability of measures of thin trading	3-17
Short term stability of annual trading frequencies	3-18
Long term stability of measures of thin trading	3-20
Testing for a trend in the thin trading of individual securities	3-21
Conclusion	3-23
§4 A COMPARISON OF THE EFFICIENCY OF RISK ESTIMATORS IN A THINLY TRADED STOCK EXCHANGE: A SIMULATION STUDY	4-1
Introduction	4-1
The efficiencies of thin trading correction procedures	4-2
The trade-to-trade technique when only weekly data are available	4-3
Simulating stock market data	4-4
Non-normality in stock market returns	4-5
Empirical investigation into non-normality	4-5
Modelling the interval effect in the non-normality of returns	4-8
Simulation study to compare estimators	4-10
Methodology for simulating non-normal deviates	4-10
Methodology for measures of efficiency	4-11
Results of initial efficiency comparison	4-13
Adaptive selection of leads and lags for Cohen estimator	4-18
Results using adaptive procedure	4-19
Conclusion and further comments	4-22

§5 THE STABILITY AND EFFICIENCY OF BETA ESTIMATORS: EMPIRICAL EVIDENCE FROM THE JSE	5-1
Introduction	5-1
Measuring the stability of beta coefficients	5-2
The spurious stability in OLS estimates of beta caused by thin trading	5-4
The effect of changing levels of trading frequency on the stability of OLS estimates of beta	5-7
Empirical evidence of the downward bias in estimates of financial risk caused by thin trading on the JSE	5-11
Empirical investigation into the stability of betas using mean squared forecast errors	5-17
Transition matrices in the assessment of beta stability	5-21
Use of the Chow-Fisher test to test for beta stability	5-24
Conclusions	5-26
Regression bias and the regression tendencies of betas	5-26
Introduction	5-26
Empirical evidence of the regression tendency of betas on the JSE	5-27
Correcting for the regression bias using a Bayesian adjustment	5-28
Identifying true regression after regression bias has been removed	5-31
Conclusion	5-35
Some further tests and suggestions for improving the efficiency of beta estimator	5-35
Weighting of observations	5-36
Alternative interval lengths in estimating beta	5-38
Bootstrap estimation of the variance of the Cohen estimator	5-41
Bootstrapping	5-41
A brief recap of the Cohen estimator	5-43
Empirical investigation to compare the simple and bootstrap estimates of the beta inefficiency	5-44
Conclusions	5-46
Summary and conclusion of Chapter 5	5-47
 §6 NON-NORMALITY AND ROBUST ESTIMATION OF BETA COEFFICIENTS ON THE JSE	 6-1
Introduction	6-1
Non-normality of residuals	6-2
Description of tests of non-normality	6-3
Extent of non-normality on the JSE	6-4
Reliability of sample skewness and kurtosis as measures of non-normality	6-7
A comparison of robust estimators as applied to security returns	6-10
The selection of robust estimators	6-13
Bounded influence	6-15
The data used in the investigation	6-17
Internal reliability of the various robust estimators	6-18
Reliability versus level of non-normality	6-23
Differences between actual beta estimates with respect to levels of non-normality	6-24
Stability of beta on the JSE	6-26
Empirical assessment of stability of betas	6-27
Concluding remarks and suggestions for beta estimation	6-29
 §7 IMPROVED BETA ESTIMATION USING ADAPTIVE INTERVAL SELECTION	 7-1
Introduction	7-1
The interval aggregation procedure for improving the efficiency of beta estimation	7-2
General regression results	7-3
The interval aggregation procedure for improving precision by increasing the variance of the market returns	7-4
A more formal description of the proposal	7-5
An hypothetical example of the predicted improvements due to the aggregation proposal	7-9
Empirical study to assess the interval aggregation procedure	7-11

Further discussion of results and a check on the theoretical assumptions of the interval aggregation procedure	7-13
Concluding remarks about the interval aggregation procedure	7-15
The extreme value set of adaptive interval selection procedures	7-15
Introduction and discussion of the extreme value procedure	7-16
The empirical study	7-17
Beta estimation using the proposed procedure	7-20
Adaptations to the extreme value procedure to incorporate robustness and the trade-to-trade approach	7-23
Conclusion	7-24
§8 COVARIANCE ESTIMATION IN THINLY TRADED ENVIRONMENTS	8-1
Introduction	8-1
Non-synchronous security trading and market index autocorrelation on the JSE	8-2
Theoretical development	8-4
Empirical investigation	8-7
Conclusions	8-9
On estimating the covariance structure of returns in thinly traded environments	8-9
Introduction	8-9
Theoretical background	8-12
Methodology	8-16
Results	8-17
Conclusion and summary	8-26
Beta estimation using the Cohen <i>et al</i> estimator in the presence of market inefficiency	8-27
Theoretical development	8-28
Covariance estimation in the presence of thin trading and market autocorrelation: a continuous-time framework	8-35
Introduction	8-35
The model for security returns	8-36
An expression for the covariance	8-41
A procedure for correcting covariance estimates in the presence of market autocorrelation	8-45
Conclusion and summary	8-46
Chapter summary and conclusions	8-47
§9 A MULTIVARIATE STATE-SPACE APPROACH TO ESTIMATING TIME-VARYING BETAS	9-1
Introduction	9-1
Multivariate state-space modelling of beta coefficients	9-3
Simplified versions of the general model	9-6
Estimation using the multivariate Kalman filter	9-8
Maximum likelihood estimation of the variance-covariance matrices	9-10
Robustification of the Kalman filter	9-11
The influence of an observation	9-14
Contaminated normal distribution	9-15
A Huber influence function	9-17
Some practical considerations in testing the filters	9-17
Empirical investigation	9-19
Empirical results using contaminated normal approximate filter	9-21
Statistical tests of the model	9-23
Conclusions and summary	9-25
§10 SUMMARY AND CONCLUSIONS	10-1
REFERENCES	

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Associate Professor Dave Bradfield for his patience and encouragement during the compilation of this thesis. I thank him not only for introducing me to the subject, but also for his sound advice and the many hours spent discussing various aspects of the thesis.

I am also grateful to Professor Cas Troskie for pointing out many interesting approaches to some of the problems encountered in the research, and for his invaluable assistance in gathering the stock exchange data. My appreciation also goes to all the members of the Department of Statistical Sciences at UCT (and in particular Associate Professor Tim Dunne and Mr Rob van den Honert) for being willing to help and offer advice when asked.

The financial assistance of the Centre for Science Development (HSRC, South Africa) towards this research is hereby acknowledged. Opinions expressed in this work, or conclusions arrived at, are those of the author and are not to be attributed to the Centre for Science Development.

Finally, I would like to thank Sally Anderson for her diligent proofreading and assistance in the layout of the thesis, as well as for her constant support and encouragement.

CHAPTER 1

INTRODUCTION AND SYNOPSIS

1.1 The estimation of systematic risk

One of the most important results to emerge from early Capital Market Theory was that investors can expect to be rewarded for only that part of the risk of a security which cannot be eliminated by holding a *diversified* portfolio. The estimation of this non-diversifiable (or systematic) risk component of securities has captured the attention of many researchers in the field of financial economics.

The estimation of systematic risk has long been associated with the 'market model' developed by Markowitz (1952, 1959) and later investigated more comprehensively by Sharpe (1963). The market model posits a simple linear relationship between the return on the security and the return on the market and it is well known that the slope coefficient of this linear relationship corresponds to the beta, or systematic risk, of the security. It was therefore originally anticipated that, under some apparently plausible assumptions, the beta coefficient could be estimated using simple linear regression applied to historical return series for the security and the market. However, authors in the '60s and '70s started to question some of the traditional assumptions made. In particular, they pin-pointed the normality of stock market returns and the timeousness of the pricing process as two important issues.

In his 1966 paper, Fisher was the first to point out that infrequent trading was the source of substantial delays in the price adjustment process of securities. It follows from Fisher's observation about pricing frictions that any historical return series used in the market model may not accurately reflect the true, underlying security returns. Hence risk estimation may be biased by the use of observed returns which do not take into account the 'Fisher effect'. Subsequently, the 'Fisher effect' has become more generally referred to as 'thin trading' and has attracted widespread interest in financial economics as a source of bias in risk estimation.

In response to the growing concern about the biases caused by thin trading, authors such as Scholes and Williams (1977), Marsh (1979), Dimson (1979) and Cohen *et al* (1983) proposed procedures to counter this biasing of beta estimates. It is, however, ironic that most of the empirical evaluations of the correction procedures have been performed on the large, international stock markets (e.g. the markets in New York and London) which are relatively well traded in comparison with the smaller markets, e.g. the Johannesburg Stock Exchange.

It is therefore important to note that the current state of beta estimation in the established literature cannot necessarily be translated to the smaller stock markets.

The second issue of non-normality has also been the source of substantial debate and research since Mandelbrot (1963), Fama (1965) and others questioned its appropriateness for security return distributions. Empirical evidence has gradually emerged since these early studies which suggests that the normal distribution is indeed not a suitable assumption for the return series of many of the securities on some of the well traded markets. Significantly for empiricists on small stock markets, it has also recently been argued by some researchers (e.g. Chan and Lakonishok (1992)), that it is the smaller, less well traded firms on the NYSE which have the most extreme non-normality. Flowing from this contention is the concern that the smaller stock markets (on which are listed smaller, less well traded companies) will, in general, exhibit more pronounced deviations from normality. It is essential to note that the importance of the non-normality question for beta estimation stems from the fact that traditional estimators are inefficient when the variables are not normally distributed.

The motivation for this thesis therefore arises from the fact that the issues of thin trading and non-normality have not been adequately researched on *small* stock markets and, in particular, on the Johannesburg Stock Exchange (JSE). Furthermore, the correction procedures which have been proposed on the well traded markets, will not necessarily be appropriate on the smaller markets, and may require novel modifications in order to produce estimates of beta which are sufficiently precise and efficient for practitioners and empiricists alike.

Apart from the literature review in Chapter 2 and the conclusions and recommendations outlined in Chapter 10, the investigations into beta estimation conducted in this thesis can be categorized into two main parts. The first part (chapters 3, 4 and 5) consists of a collection of studies which focus on the application of existing knowledge to data from the JSE. The emphasis in this first part of the thesis is on assessing the characteristics of beta coefficients and the thin trading phenomenon on a market which is substantially smaller than the markets in New York and London, where much of the original research into beta estimation has been conducted. The second part (chapters 6 to 9) includes the more novel modifications to the existing estimators, as well as some completely new approaches. The focus of the second part of the thesis is on improving the efficiency of beta estimators rather than on empirical assessments of the behaviour of beta coefficients (and covariances). It is important to note that some of the insights and modifications derived in the second part of the thesis have more general applicability than purely small markets. Accordingly, several of the techniques have

been extended so that they are able to encompass the characteristics of the more well traded markets. A more detailed structure of the thesis is given in the next section.

1.2 Organization of the thesis

It is intended that the results and discussions in the thesis should appeal to a broad spectrum of readers. Consequently, throughout the thesis, a style has been adopted of relegating most of the theoretical developments and analytical derivations to appendices which can be found at the end of each chapter. In addition, empirical findings which merely confirm results already presented in the main body of the chapter, or which just offer alternative insights into a point made elsewhere, are also included in the appendices. It should, however, be noted that the appendices form an integral part of the thesis. The rest of this section contains a brief overview of the contents of the remaining chapters of the thesis, as well as a few of the more important findings.

In Chapter 2 a review of the important contributions to systematic risk estimation is given. The review focuses specifically on the research which is deemed essential for the ensuing developments in the thesis. To this end contributions in the literature which deal with corrections for the biases caused by thin trading, the non-normality of stock market returns and the time-varying nature of beta coefficients constitute the major areas of discussion in the chapter.

Chapter 3 contains a detailed study of thin trading on the Johannesburg Stock Exchange (JSE) and is aimed at providing an understanding of the mechanisms of the trading process. Although thin trading *per se* is not of particular interest to practitioners, knowledge of the extent and characteristics of the trading frequencies of securities are of crucial importance for the developments in later chapters (where thinly traded environments are reconstructed). The investigation in Chapter 3 also includes an appraisal of the stability of trading frequencies over time and concludes that thin trading characteristics are, by and large, constant, particularly in the short term. Some evidence is presented which suggests that, over the longer term, there may be a drift in the trading frequencies which eventually manifests itself as a *significant* change after a period of several years. Furthermore, a model of the trading process is proposed and assessed in Chapter 3 and later implemented in a simulation study described in Chapter 4 (as well as in the theoretical development of Chapter 8 which deals with covariance estimation).

Chapter 3 thus forms an important backdrop to the investigations into systematic risk in the ensuing chapters of the thesis.

Chapter 4 is devoted to investigating the suitability of existing thin trading correction procedures for beta estimates on small stock markets. Specifically, the chapter contains the details and results of a simulation study designed to quantify and compare the efficiencies of the existing correction procedures. In particular, the two main approaches to counteracting the thin trading biases in beta estimates, namely the aggregated coefficients approaches (see, e.g. Cohen *et al* (1983)) and the trade-to-trade approach (see, e.g. Marsh (1979)), are compared. Although the two approaches are (by construction) unbiased when the only price adjustment delays are due to infrequent trading, the relative efficiencies are shown to be vastly different, especially in the very thinly traded market environments. The results of the simulation study reveal that the trade-to-trade estimator is substantially more efficient than the aggregated coefficients method, even when the rather onerous data requirements of the trade-to-trade approach have to be compromised (and weekly data instead of daily data are used). The chapter also includes a preliminary investigation into the non-normality of security returns. The investigation into the return distributions was necessary so that the simulation methodology could be made as realistic as possible (while still maintaining control of the critical parameters).

In Chapter 5 a comprehensive study into the long term stability of beta coefficients is conducted. The results of the study suggest that the beta coefficients of securities on the JSE exhibit the same, moderate degree of stability as was found in the UK and US. Further, an important preliminary investigation in this chapter confirms that uncorrected estimates of beta coefficients are shown to distort the assessment of systematic risk stability. The findings also strongly confirm that it is imperative that the biases caused by thin trading be eliminated before any empirical studies concerning beta coefficients are performed. A particular form of instability, namely the 'regression tendency' of portfolio beta coefficients, is analyzed using data from the JSE. In correcting for this regression tendency, it is established that a Bayesian adjustment procedure proposed by Vasicek (1973) substantially improves the efficiency of estimators. The chapter concludes with the recommendation that a trade-to-trade estimator in conjunction with a Bayesian adjustment be used in small markets.

In the final section of Chapter 5, a somewhat different topic is considered, viz. the computation of the variance of the Cohen *et al* (1983) aggregated coefficients estimator using 'bootstrapping'. Although the evidence from Chapter 4 and the first part of Chapter 5 suggests that the trade-to-trade estimator is the preferred beta estimator on the JSE, the Cohen estimator

is still widely used especially in markets where thin trading is less extreme. It was thus deemed important to be able to quantify the standard errors of the estimates produced by the Cohen estimator. The bootstrap methodology is used to provide a yardstick measure of the standard error of the estimate.

Chapter 6 discusses a very important consideration in beta estimation, namely the non-normality of stock market residuals and returns. It is well known that traditional least squares estimators are significantly inefficient if the underlying variables are not normally distributed, and, in particular, if the distributions are *heavy tailed* and contain any outliers. The discussion in Chapter 6 deals firstly with the extent and stability of non-normality in the residual distributions. It is shown that the residual distributions are not constant over time and that the degree of non-normality of a security differs in different sample periods. In the second part of the chapter, the improved efficiency of robust estimators is investigated in a comprehensive empirical study. A very broad range of robust estimators is applied to return series from the JSE and the efficiency and goodness-of-fit of the estimators are quantified and compared. It is clearly shown that the robust estimators offer significant improvements in efficiency when applied to securities which exhibit significant non-normality in their residuals and marginal improvements even when the hypothesis of normality cannot be rejected. The *bounded influence* L_p -norm estimator emerges as the single most efficient of the estimators. The third part of the chapter briefly returns to the issue of stability of beta coefficients, and the results of Chapter 5 are confirmed using the robust estimators developed in Chapter 6.

In Chapter 7 an innovative technique for improving the precision of beta estimators using adaptively selected data is proposed. The motivation for the proposal stems from the consideration that the trade-to-trade estimator conventionally utilizes returns measured between the last traded days in successive calendar months, whereas there is clearly more information about the price movements of the securities and the market than that available at the month ends. The proposal is based on the idea of adaptively selecting the days in the month so that the linear relationship between the market and the security is accentuated. In an empirical study conducted using JSE securities, the proposed method is demonstrated to yield considerably improved efficiency in beta estimators (for both well traded and thinly traded securities).

Chapter 8 focuses on the estimation of the covariance matrix of security returns. It is argued that covariances are essential input parameters for diversification considerations in portfolio selection and analysis, as well as for multivariate tests of financial models, and therefore warrant investigation. As it turns out, covariance estimates are also biased by thin

trading in much the same way as beta estimates and correction procedures consequently should be implemented to obtain accurate estimates. It is further contended that an adaptation of the aggregated coefficients Cohen estimator is the most appropriate methodology. In order to implement the aggregated coefficients estimator efficiently without sacrificing unbiasedness completely, a practical solution to the selection of the number of leading and lagged coefficients is proposed. A complication involved in using the aggregated coefficients estimator arises when there is autocorrelation in the market returns (over and above that induced by thin trading in the component securities). A novel analytical derivation results in an expression which gauges the extent to which the aggregated coefficients estimators of covariances are biased upwards when the security returns are serially correlated. The expression is then also able to be used to correct for this bias. The development is extended into a continuous-time framework and consequently has applicability on well traded markets where pricing delays of shorter duration are evident.

Chapter 9 presents an original strategy for modelling time-varying betas which incorporates a robust, multivariate state-space approach. Motivation for the study of time-varying betas arises from the evidence of beta instability presented in Chapters 5 and 6 of this thesis and from other literature. The state-space approach to modelling and estimating time-varying betas is outlined in the chapter and interesting new insights gained from the *multivariate* perspective are highlighted. It is argued that the correlations in the movements of the systematic risk coefficients of securities (which can be estimated in the multivariate paradigm) have useful implications for portfolio risk management. This chapter is primarily an expository chapter, showing how sophisticated time series models can offer an understanding of the reasons for why and how betas vary over time. However, a preliminary empirical investigation into the usefulness of the technique is conducted in the chapter using the multivariate state-space methodology applied to the sector betas from the JSE. The interpretation of the estimated correlation matrix of the sector betas is shown to provide very practical and fresh insights into the control of portfolio systematic risk.

Finally, Chapter 10 contains a summary and the recommendations emerging from the analysis in the thesis. The final recommendations for practical beta estimation are couched in a hierarchy of modifications to the trade estimator, with each step contributing to improved precision and efficiency in the resultant estimate.

CHAPTER 2

A REVIEW OF SYSTEMATIC RISK ESTIMATION¹

Since the development of modern portfolio theory (MPT) and the consequent emergence of asset pricing models, much effort in the finance literature has focused on accurate estimation of the parameters underpinning these models. The measure of systematic risk, or beta coefficient, has occupied a central role in asset pricing models and as such has been the focus of much attention in the canon of established literature in Financial Economics.

The initial motivation for this thesis stems from the view that efficient estimates of beta coefficients are crucial to accurate implementation of Capital Market Theory (CMT). More specifically, efficient estimates of beta are necessary not only for portfolio design and asset pricing decisions as well as cost of capital calculations, but also for the efficient testing of these financial models. Clearly the two purposes are not at odds: implementation of the models requires the confidence in them which evolves through positive testing of the models.

In this chapter some of the aspects of beta estimation which are pertinent when dealing with a small stock exchange such as the JSE are reviewed.

2.1 The evolution of beta estimation

Out of the plethora of research which has accompanied the development of MPT arises the basic hypothesis that investors are rewarded for taking on the risk of an investment as part of a diversified portfolio rather than in isolation. The so-called systematic (or non-diversifiable component of) risk is closely associated with the co-movements (or covariance) that a security's returns have with the market's movements. It is not surprising therefore that the beta coefficient which is calculated from that covariance is found at the heart of two of the most widely used models in empirical finance, namely the Market Model and the Capital Asset Pricing Model (CAPM).

The Market Model was originally considered by Markowitz (1959) whereafter considerable simplifications were discussed by Sharpe (1963). The original model is commonly referred to as Sharpe's Index Model (or the Diagonal Model), and was initially used to simplify

¹ Parts of this review have been published in Bowie and Bradfield (1993b).

estimation problems in the implementation of Markowitz's (1952) landmark portfolio selection theory. The intention was to reduce the number of input estimates by relating the co-movements in prices of individual securities through a common index.

The model can be written down algebraically as follows:

$$r_{j,t} = \alpha_j + \beta_j r_{m,t} + \epsilon_{j,t} \quad (1)$$

where $r_{j,t}$ is the return on security j in interval t ; $r_{m,t}$ is the return on the market in the same interval; α_j and β_j are the alpha and beta coefficients of security j and $\epsilon_{j,t}$ is the random or residual component of the return on security j in interval t . The residual component is generally assumed to have a stable and serially independent distribution.

The Market Model has since been used extensively for the estimation of beta coefficients. Stapleton and Subrahmanyam (1983), Markowitz (1984) and Jacob and Pettit (1984) have been among the many authors to discuss the important ties between the Market Model and the CAPM. The close links between the Market Model and the *ex ante* formulation of the CAPM lend mutual credence to their status as useful models for security price movements. Another arena where the Market Model has proved itself useful is described in Bradfield (1989). Bradfield proposed an extension of the Market Model to measure the impact of foreign exchanges on the volatility of JSE shares. In a similar vein, Bradfield and Bowie (1991) have used a modification of the Market Model to gauge the additional risk involved in hostile merger activity.

Although this thesis naturally presupposes that the estimation of beta coefficients provides useful information for financial analysis and decision making, there has been some debate around this viewpoint in the vast literature on beta estimation. Underpinning the debate is the fact that the beta coefficients is usually linked to the expected trade-off between market risk and expected return. The CAPM developed by Sharpe (1964), Lintner (1965) and Black (1972) is the most widely used of the models which represent this trade-off and has consequently served as the focus for much of the debate. Scepticism about the validity of the CAPM has often been associated with the contention that 'beta is dead'. However, one view is that much of the scepticism has arisen from over-ambitious conclusions drawn from empirical tests of the CAPM.

Early empirical tests of the CAPM undertaken by Black, Jensen and Scholes (1972) and Fama and MacBeth (1973), however, indicated support for the CAPM on the NYSE.

Subsequent researchers have found contradictory evidence in later time periods, e.g. Reinganum (1981) and Lakonishok and Shapiro (1986). More recently, Fama and French (1992) concluded that, based on empirical evidence from the NYSE over a 50-year period, the cross-sectional variation in average security returns was better captured by a two-dimensional model including the firm sizes and book-to-market equity ratios as variables as opposed to the simple CAPM. In particular they concluded that any explanatory power that had been attributed to beta in earlier studies was due to confounding with a firm size effect.

In their reaction to the empirical evidence and conclusions presented by Fama and French (1992), Roll and Ross (1994) demonstrated the extreme sensitivity of tests of the CAPM to the choice of market index. They showed that any empirical conclusions about the CAPM are conditional on the choice of market proxy and, moreover, that there existed a very close link between the choice of market proxy and the empirical results. The equivalence of the mean-variance efficiency of the market index and the linearity of the relationship between a security's expected return and beta coefficient had in fact been set out by Roll in his critique of tests of the CAPM as early as 1977. Furthermore, Roll and Ross (1994) noted that sampling error in the measured variables exacerbated the problem of model testing. In empirical evidence from the Helsinki Stock Exchange, Malkamaki (1992) has similarly noted that better estimates of the parameters improve results of tests of the CAPM substantially. A further explanation for the weak empirical evidence in support of the CAPM has been provided by Affleck-Graves and Bradfield (1992), who showed that conventional tests of the CAPM have very low power. It should be noted that, although much of the debate has centred on the CAPM itself, the value of accurate and efficient estimates of beta coefficients rests only on the existence of a systematic risk/expected return relationship; a relationship which concurs with basic investment intuition. The importance of the beta coefficient does not, therefore, rest solely on acceptance of the CAPM. Other more complex models may incorporate the risk-return trade-off and offer better explanatory power than the CAPM.

On smaller stock markets empirical investigation has to be sensitive to the particular problems associated with statistical estimation using data from those markets. Foremost among the problems is thin trading. The next section reviews how thin trading obscures accurate analyses and how researchers have attempted to combat the effects of thin trading.

2.2 The impact of thin trading on beta estimation

Fisher (1966) was one of the first to identify the phenomenon of thin trading. In particular, he pointed out that recorded prices of securities were not necessarily equal to their underlying theoretical values. The so-called 'Fisher effect' occurs because it is only at the discrete times at which transactions are conducted that the prices can be observed from the theoretically continuous process underlying the pricing of shares.

It was only later that Ball (1977), Scholes and Williams (1977) and others investigated the effects of thin trading on systematic risk estimation. They demonstrated that the ordinary least squares (OLS) regression estimates of beta are particularly susceptible to bias in thinly traded environments. As a consequence, several alternative beta estimation proposals designed to correct for the bias caused by thin trading began to emerge in the literature.

Schwert (1977), Franks *et al* (1977) and Marsh (1979) suggested a 'trade-to-trade' approach for estimating betas. In the trade-to-trade approach the intervals over which returns are measured are allowed to vary in length according to when the security has been traded. By considering the security prices only on the days that the security trades, the problem of out-of-date values is avoided. The returns are measured between the last traded days in consecutive months rather than the last calendar days in the months (on which there may not have been a trade). The measurement of returns on the index is then matched to the days on which trades in the security occurred. Symbolically the trade-to-trade estimator involves estimating the parameters from the equation:

$$r_{j,t} = \alpha D_{j,t} + \beta \tilde{r}_{m,t} + \tilde{\epsilon}_{j,t} \quad (2)$$

where

$D_{j,t}$ is the length of time between the last trades in intervals $t-1$ and t ;

$\tilde{r}_{m,t}$ is the return on the market between the last traded days in intervals $t-1$ and t ;

$\tilde{\epsilon}_{j,t}$ is the residual component over the interval between the last traded days in intervals $t-1$ and t ;

$r_{j,t}$ is the measured return on the security in interval t .²

² In the actual estimation the known values are usually adjusted to allow for the fact that the residual component ($u_{j,t}$) has a variance which depends on the length of time between last traded days (i.e. $D_{j,t}$). See Chapter 3 for more discussion of the adjustment.

From (2) it is clear that the two main requirements for implementing the trade-to-trade approach are firstly that the exact trading times must be known and secondly that a market index which does not suffer indirectly from thin trading (i.e. have a significant weighting of thinly traded component securities) must be available.

As an alternative approach Ibbotson (1975), Pogue and Solnik (1974), Schwert (1977) and Dimson (1979) have all proposed techniques for correcting for the biases in beta estimates caused by thin trading which are based on aggregating the coefficients relating security returns to lagged market returns. The aggregation is made in order to capture fully any common movements of the thinly traded security and the market index. The consideration of only lagged coefficients of the market index in these models begs the question of whether the market index itself could suffer from the effects of thin trading.³ It is therefore plausible that when market indices suffer from the effects of thin trading, very well traded securities may in fact *lead* the market index. Scholes and Williams (1977) addressed this issue by combining both the leading and the lagged market index returns together with the trade-to-trade approach to develop a consistent estimator of beta which dealt with the problem of thin trading in the index. Dimson (1979) pointed out that the Scholes and Williams (1977) approach is inefficient because it uses only returns calculated over consecutive traded intervals and consequently ignores returns measured over several interval when estimating beta. Dimson therefore presented a revised estimator based on the aggregated coefficients approach.

Cohen *et al* (1983) and Fowler and Rorke (1983) noticed a small inconsistency in the derivation of Dimson's (1979) aggregated coefficients estimator and produced a corrected version. Cohen *et al* (1983) also noted that thin trading is not the only source of friction in the price adjustment process and suggested that although the trade-to-trade technique corrects for the effects of thin trading it might not compensate for the other factors.⁴ The Cohen *et al* (1983) estimator (henceforth 'Cohen estimator') is a consistent estimator⁵ and does not rely on information regarding transaction times or on a continuously updated market proxy. The

³ Clearly market indices based on an equally weighted construction are likely to suffer to a greater extent from thin trading than indices based on a market capitalization weighted construction. This occurs because smaller companies (which are generally more thinly traded - see Dimson (1979) for evidence from the UK market) are given relatively larger weights in an equally weighted index than in a market capitalization type index.

⁴ Cohen *et al* (1983) argued that other delays might also result from (1) individual traders trading only periodically owing to information, decision and transaction costs, and (2) specialists/dealers impeding price adjustments to ensure stability or correcting inventory imbalances.

⁵ The attribute of consistency implies that the hypothetical series of estimates produced as the sample size is hypothetically increased will converge on to the true value of the parameter.

estimator can be regarded as a generalised version of the Scholes and Williams (1977) estimator which was designed to estimate the systematic risk of stocks on the NYSE (of which the majority trade at least once a day).

The Cohen estimator consists of the aggregation of the estimated coefficients of several simple regressions. Each simple regression measures the extent to which the return on the security in a particular interval is dependent on the return on the market in either a leading or lagged interval. If the number of leading intervals to be used in the estimator is denoted by l_1 and the number of lagged coefficients by l_2 , then the Cohen estimator can be written as:

$$\hat{\beta} = \frac{\sum_{l=-l_2}^{l_1} \beta^{(l)}}{\sum_{l=-l_2}^{l_1} \rho_m^{(l)}} \quad (3)$$

where

$$\beta^{(l)} = \frac{\text{Cov}[r_{j,t+l}; r_{m,t+l}]}{\text{Var}[r_{m,t+l}]}$$

and

$$\rho_m^{(l)} = \frac{\text{Cov}[r_{m,t+l}; r_{m,t}]}{\text{Var}[r_{m,t}]}$$

Cohen *et al* originally specified that l_1 and l_2 should be equal. However, as the number of lags plus leads increases so the estimator becomes more inefficient because more regressions are needed to obtain an estimate. The compromise solution to the dilemma is to use a slightly biased version of the Cohen estimator which will be more efficient; this can be achieved by selecting different numbers of lagged and leading coefficients. The numerator in (3) represents the aggregated slope coefficients from the simple regressions of the security returns against lagged and leading market returns. The denominator term of the estimator in (3) corrects for the autocorrelation introduced into the market proxy by thin trading in its components.⁶ It is therefore readily apparent that if the market comprises only well traded securities, then the

⁶ For evidence of the spurious autocorrelation introduced into the market proxy (and an attempt to measure the extent to which observed autocorrelations are explained by thin trading) the reader is referred to Atchison *et al* (1987).

denominator of equation (3) will tend to unity. If the security is well traded, then only the synchronous coefficients will be significantly larger than zero and hence the numerator of equation (3) will tend to the OLS estimate of beta.

On the JSE, Bradfield and Barr (1989) have assessed the suitability of the Cohen estimator for estimating the betas of securities listed on the JSE. They implemented a version of the Cohen estimator with five leading and five lagged monthly returns and found that although significant improvements in beta estimation were evident, even five leads and lags are not sufficient to capture all non-synchronous correlations between the security and the market for extremely thinly traded shares on the JSE. The Cohen estimator appeared to remove (*on average*) all thin trading biases, but resulted in very inefficient estimates of individual security betas. On the small Helsinki Stock Exchange, Berglund, Liljeblom and Löflund (1989) have compared the performances of the Cohen and trade-to-trade estimators. Unfortunately they used data at a daily frequency and thereby magnified the importance of intra-day delays in the trading process. Their conclusion that OLS performed nearly as well as the trade-to-trade estimator and better than the Cohen estimator stemmed from this choice of data frequency. Levhari and Levy (1977) and, more recently, Handa, Kothari and Wasley (1989, 1993) have been among those researchers who have further investigated the importance of data frequency in the estimation of beta coefficients. They came to the conclusion that longer interval lengths (even up to annual returns) provided more efficient estimates of beta coefficients.

When estimating covariances between the return series of two securities, thin trading presents an even more serious problem. In implementing a trade-to-trade type estimator there exists the very real possibility of finding only very few times at which there are trades in both securities. Therefore, a trade-to-trade estimator will become unreliable because of the very small sample of returns with which to measure the correlation. Empiricists are therefore compelled to look for an aggregated coefficients approach to covariance estimation. Shanken (1987) pointed out that one of the steps in the derivation of the Cohen estimator could be used directly for covariance estimation. However, a difficulty presents itself here as well because the two securities might both be very thinly traded or very well traded. A balance needs to be achieved between the inefficiency of too many coefficients and the bias remaining if too few are used. In deriving their estimators, Dimson (1979) and Cohen *et al* (1983) have discussed the crucial issue of selecting the appropriate number of leads and lags in the context of beta estimation.

An interesting extension of the use of lagged market returns to estimate the beta of a security has been recently presented by Luoma, Martikainen and Perttunen (1993a). They proposed lag distribution models for obtaining estimates of beta and presented evidence of its improved consistency using a comprehensive sample of securities from the Helsinki Stock Exchange. A further paper by the same authors (Luoma, Martikainen and Perttunen (1993b)) investigated an 'error-correction' model for the risk behaviour of thinly traded securities. Their empirical findings on the Helsinki Stock Exchange suggested that the error-correction model provided useful insights into the risk characteristics of thinly traded securities. They tempered these conclusions, however, by stating that the uniqueness of the Helsinki Stock Exchange may have led to results which cannot necessarily be generalized.

The next section further explores the issue of efficiency in systematic risk estimation.

2.3 Efficiency of estimation and the issue of non-normality

Much research has been directed at establishing the impact that different interval lengths have on estimates of beta. In examining various beta estimators Blume (1975), Eubank and Zumwalt (1979) and others have experimented with the effect that various interval lengths have on the predictive power of beta estimators.

In Europe, Pogue and Solnik (1974) were the first authors to measure the impact on the estimates of beta obtained by using different interval lengths. They identified an *intervalling effect* in which higher estimates of beta were obtained as longer intervals were used. Data from seven European countries were used in their study. More recently, Corhay (1992) has studied the intervalling effect in beta estimation using a comprehensive sample of securities from the Brussels Stock Exchange. The aim of his study was to measure the rate at which convergence in the series of beta estimates occurred as the interval length was increased. In an attempt to explain the intervalling effect, Cohen *et al* (1983) have argued that it is due to price-adjustment delays which are caused by thin trading and other frictions in the trading process.

In most empirical work authors have settled on using monthly intervals over a five-year period, e.g. Black, Jensen and Scholes (1972), Fama and MacBeth (1973), Blume and Friend (1973) and Affleck-Graves and McDonald (1989) to name a few. The selection of sixty months represents a trade-off between a large enough sample size to enable reasonably efficient estimation and a short enough period over which the underlying beta could be assumed to be stable.

Using the sixty months of returns as input data, the estimation technique used in most studies has been simple regression (using OLS as a measure of the goodness-of-fit). Under the assumption that the distribution of the monthly returns is stable and has finite variance, the Gauss-Markov theorem establishes OLS as the most efficient of the linear estimators. In addition, if the monthly returns (or rather the residual components of the Market Model) are normally distributed, then the OLS estimator coincides with the maximum likelihood estimator. The normality of stock returns is an assumption often made in empirical financial research. It is based on the central limit theorem which proves that the distribution of a sum of independent random variables tends to the normal distribution provided the distribution of the random variables has a finite variance. As the monthly return on a security can be represented by the sum of returns over many shorter intervals of time, it is argued that the monthly return should follow a normal distribution. Fama (1976) gives a more detailed account of the arguments for and against the use of the normal distribution in research using stock market returns.

However, the empirical evidence seems to be inconsistent with the normality assumption. The early work of Mandelbrot (1963) and Fama (1965) reported that the returns were more leptokurtic (heavy-tailed) than the normal, but later Fama (1976) and Blattberg and Gonedes (1974) considered the normal distribution to be a reasonable working approximation for monthly returns on the NYSE. Several researchers, e.g. Fama (1965) and Blume (1970), have proposed the stable Paretian family of distributions as an alternative. Blattberg and Gonedes (1974), on the other hand, argued for the use of the Student t -distribution for daily returns, while Kon (1984) found that they could be modelled with a mixture of normal distributions.⁷ Recently, Booth, Hatem, Virtanen and Yli-Olli (1992) have considered generalised autoregressive conditionally heteroscedastic (GARCH) models with the normal, the Student's t and the power exponential⁸ as the conditional distribution for the returns. They found in favour of the heavier tailed power exponential distribution for the Helsinki Stock Exchange. On another small stock market, the Dutch stock exchange, De Jong, Kemna and Kloek (1992) also concluded that stock market residuals are significantly non-normal.

Although the focus of much of the research in this area has been on the appropriate distribution for security returns, it is the *residual* distribution which is most pertinent to statistical estimation and inference. The fact that the residuals cannot be observed but can only

⁷ Some other suggestions are described in Chapter 6.

⁸ A discussion on the power exponential distribution can be found in Box and Tiao (1973).

be estimated means that the tests applied to the residuals are only approximate. Significant non-normality of the returns nevertheless implies that the residuals are also likely to be non-normally distributed.

Malkamaki (1992) in his research on the Finnish stock exchange postulated that the non-normality of share returns may be 'an even bigger problem in tests of asset pricing models' than thin trading. Previous research by Schlosberg (1976) and Affleck-Graves (1974) has shown that the distributions of JSE stock returns are also considerably non-normal.

The efficiency of conventional, static estimators of beta can be improved using robust estimation techniques. Robust techniques have long been recognised as appropriate in econometric studies; see, for example, Koenker (1982), Blattberg and Sargent (1971), Coursey and Nyquist (1983) and Smith and Hall (1972). The testing of robust estimators on financial data has only recently been performed (see Chan and Lakonishok (1992), although they by no means cover the full range of estimators available). Among some of the earlier researchers who did examine the relative efficiencies of robust and OLS estimators were Sharpe (1971) and then later Cornell and Dietrich (1978). They expressed surprise at the very small differences in the estimates of betas produced by the robust estimator and those of the usual OLS estimator.

McDonald and Nelson (1989) have compared the estimates of beta on NYSE stocks obtained using a robust partially adaptive generalised t -distribution (GT) estimator (see McDonald and Newey (1988)) and the OLS estimator. They find that the betas of those shares on the NYSE which exhibit the most marked departures from normality have the most significantly different beta estimates arising from OLS and GT. They also suggest that the high proportion of approximately normally distributed shares on the NYSE is the cause of the apparently insignificant difference between the OLS and robust estimates of the beta coefficients reported by Sharpe (1971) and Cornell and Dietrich (1978). Chan and Lakonishok (1992) have extended the investigation into robust estimators of beta on the NYSE by assessing several robust estimators on simulated and actual data. They conclude that the theoretically improved efficiency of the robust estimators in the face of fat-tailed distributions warrants closer investigation. Most of the tests and investigations described in this section have highlighted the differences between OLS estimates of betas and the robust estimates. Very little has been done on measuring the empirical efficiency of the robust estimators and researchers have relied on an indirect line of reasoning: stock returns are non-normal which implies that OLS is sub-optimal; hence a robust estimator must be more efficient.

2.4 Improving the predictive power of estimators through the inclusion of prior information

Implementation of asset pricing models presupposes the stability of the parameters estimated. Detection of beta stability is obscured by the fact that only the *estimates* of beta are observable. Unstable *estimates* of beta do not necessarily imply unstable *true* betas and, conversely, stable estimates do not imply stable underlying betas (see Dimson and Marsh (1983) for a discussion on spurious stability). Classical statistical estimation (such as OLS) relies on sampling theory, i.e. that errors in the estimates are random and not consistently biased either up or down. Consequently, estimates of betas are assumed to be equally likely to be greater than or smaller than the true beta. When no other information exists about the parameter being estimated then the assumptions of sampling theory may be forced on the estimator. However, in the case of the estimation of beta coefficients, prior information does exist in that it is known that it is unlikely for a security to have a beta outside the range between zero and two, say.

The Bayesian approach to statistical estimation explicitly allows for the incorporation of subjective or prior beliefs about the parameter which is being estimated into the sample information. The consideration of both prior beliefs and sampled data provides estimates which are optimal with respect to the minimization of the mean squared error (MSE) of predictions. In the Bayesian paradigm individual security betas can be adjusted to fit in with the overall distribution of security betas. Very low betas will be adjusted upwards towards an average security beta value, and very high betas will be 'shrunk' towards the average beta value. In essence the adjustment takes cognizance of the fact that a very low beta estimated from the data is more likely to have been under-estimated than over-estimated and vice versa.

Vasicek (1973) proposed a simple Bayesian adjustment to betas estimated using sample information. He emphasised that the adjustment both minimizes the loss of accuracy and efficiency due to mis-estimation and also enables incorporation of information additional to that available from the sample data. For example, gold shares listed on the JSE are known to be among the more risky investments and could be expected to have a beta of greater than unity.

Blume (1975) presented an alternative adjustment procedure which also modifies beta estimates by allowing for their regression tendencies over time (i.e. the observed tendency of portfolio betas to move, i.e. 'regress', from their current estimates towards an 'average' beta value). Betas which are over- or under-estimated in one period are likely to be better estimated

in the next period and a change in the individual estimates towards an average figure will occur. Consequently, a trend towards the mean beta value will be apparent in the estimates. Blume argued further that the changes in estimated betas are due not only to regression bias, but also to changes in the underlying, true betas. Blume's method involves calculating two sets of betas from two successive (but non-overlapping) periods and establishing the rate at which the betas changed from the one time period to the next. The estimates of the rate of change of the beta can then be used when forecasting the betas for the following period. Using US data, Eubank and Zumwalt (1979) reported that both the Blume and Bayesian adjustments resulted in smaller MSEs than the unadjusted betas. Further, they demonstrated that there is very little difference between the performance of the Blume and Bayesian techniques as measured by MSEs. They also concluded that longer estimation periods contribute towards more stable estimates: for example, they measured a reduction in the mean squared forecast error of unadjusted beta estimators of 81% in using a 36-month estimation period instead of a 12-month period. They reported a further reduction of 28% in using 60-month periods, but only a 14% additional improvement when 120-month periods were used. Other empirical studies using US data have been performed by Klemkosky and Martin (1975). Their conclusions were similar to those of Eubank and Zumwalt which are described above.

The concept of including information other than that obtainable from share price databases can be extended into time-varying beta estimators. This concept is discussed in the next section.

2.5 Models and tests for time-varying beta coefficients

As computing power has improved so has the ability to implement sophisticated statistical estimation techniques. One of the trade-off decisions necessary in the estimation of beta coefficients using historical data is the number of intervals over which to measure the security and market returns. Traditionally 60 months of data have been considered a reasonable compromise between the *stability* of the underlying betas and a sufficient number of data points for *efficient* estimation.

Blume (1971, 1975) initiated research into the instability of security beta coefficients over time. He emphasized that the conventional wisdom which prevailed at the time, namely that any apparent regression of beta coefficients was due only to 'order bias', was unfounded.

Indeed, he demonstrated that there were real shifts in the underlying betas which he attributed to changes in the firms' risk profiles.

Dimson and Marsh (1983) conducted exhaustive tests on beta stability on the UK market. They concluded that beta estimates show moderate stability, but that there is a slow regression in the underlying betas. An important aspect of stability testing which they elucidated is that biased estimators will lead to conclusions of spurious stability. In particular they identified thin trading as a major culprit in causing spurious stability in beta estimates.

Apart from *ad hoc* tests, a well-known test of parameter stability within a regression context is the Chow (1960) test described more succinctly in an expository note by Franklin Fisher (1970). Implementation of the Chow test necessitates defining a change-point in the time series when the parameter being estimated is hypothesized to have changed value. The quality of the fit of two separate regressions over the two parts of the data is compared with the fit of a single regression over the whole period. A significantly better fit by the two regressions indicates that the change-point heralds a significant change in beta value. Carter (1983) applied the Chow test to a small sample of securities from the JSE and found some evidence of non-stationarity in his attempts to define the most appropriate period for estimating betas. Brown, Durbin and Evans (1975) suggested the use of recursive residuals for detecting the existence of a change-point in a data sample and for assisting in locating it. Earlier, Quandt (1958, 1960) had invented a log-likelihood method for pin-pointing shift points in parameter values over time. The implementation of both of Brown *et al* and Quandt techniques are described in Hays and Upton (1986) and the Brown *et al* proposal has been applied to the Finnish stock market by Bos, Fetherston, Martikainen and Perttunen (1993). An important practical point which Hays and Upton noted about the application of the tests for instability is that they depend on normality in the underlying return distributions.

On a slightly different track, Breusch and Pagan (1979) noted that tests for non-stationarity of regression coefficients correspond to a particular type of heteroscedasticity in the residuals, viz. the residual variance is functionally related to the explanatory variable. They developed a test for such random coefficient variation which is easy to perform, but which also relies on normality in the residual distribution.

The current state of research appears to indicate that many security beta coefficients on markets around the world are not stationary. There are two approaches which can be adopted in solving the above problem of forecasting future betas. The first approach is to try to identify which model of non-stationarity best matches the observed movements in beta. In

other words the Market Model is re-formulated so that it allows for non-stationarity of the beta coefficients. The second approach seeks to identify other financial variables closely correlated with beta which permit forecasting of betas because of their leading nature.

In the spirit of the second approach, Beaver, Kettler and Scholes (1970) explored the relationship between betas and payout ratios, asset growth rates and earnings variabilities. They report a marginal superiority in predictive power of their model over historical estimates of beta. Rosenberg and Guy (1976a, 1976b) and Rosenberg and Marathe (1975, 1976) have extended research in the area by exploring prediction rules for individual security betas based on industry betas adjusted by fundamental aspects of the firm.

The modelling approach to non-stationarity has produced several alternative models to the constant beta model. While there have been attempts to rationalize these models in terms of how firms control their risk profiles, they have generally been motivated by mathematical tractability and a paucity of reliable evidence regarding firms' attitudes to risky projects. The main structural models suggested for testing beta stability have been the random coefficients model (see Theil (1971) or Hildreth and Houck (1968)), Rosenberg's (1973) AR(1)⁹ or return-to-normalcy model and, thirdly, the random walk model.

The ensuing sections briefly review each of the three models and the evidence found supporting their role as a model for beta coefficients. Some further suggestions made in the literature regarding possible models for the time-varying nature of beta coefficients are also discussed below.

2.5.1 The random coefficients model

In the random coefficients model, betas are purported to be generated in accordance with the following model:

$$\beta_{j,t} = \bar{\beta}_j + e_{j,t} \quad (6)$$

where $\beta_{j,t}$ is the beta for security j at time t which varies around an constant *average* level of beta ($\bar{\beta}_j$) by a random (and serially uncorrelated) amount, viz $e_{j,t}$.¹⁰ The random coefficients

⁹ AR(1) denotes an autoregressive process of order one.

¹⁰ The random amount is also assumed to have zero expectation.

model thus proposes that in each interval the beta of the security is randomly distributed around an *average* level of beta which remains constant over time. Because the amount by which the beta in a future interval will differ from the average beta is random, the best predictor of the future beta is the average constant value of beta.

Fabozzi and Francis (1978) tested the random coefficients model as an alternative to the constant level model for beta and reported strong evidence in favour of it. However, Alexander and Benson (1982) countered that the conclusion of Fabozzi and Francis overstated the case for the model. Later Bos and Newbold (1984) compared the random coefficients and AR(1) models for beta coefficients using NYSE data. They also found a high incidence of instability among the beta coefficients, but suggested that the power of their tests was too low to *contrast* the two models of instability effectively.

2.5.2 The random walk model

The random walk model is well-known in finance (and other areas of scientific research) and models the value of beta as being randomly incremented from one interval to the next, i.e.

$$\beta_{j,t} = \beta_{j,t-1} + e_{j,t} \quad (7)$$

The random walk model for betas effectively states that the value of the beta coefficient in an interval t is different from the beta in the previous interval, $t-1$, by a random amount. Hence the best estimate of the beta in a future interval is the current beta. Sunder (1980) and later Simonds *et al* (1986) (with more powerful statistical tests) examined the merits of the random walk model. Using the random walk model as an alternative to the constant beta model, they detected considerable instability in betas for even reasonably short periods of investigation, viz. eight years.

2.5.3 The AR(1) model

Ohlson and Rosenberg (1982) proposed a model for beta coefficients based on a variant of the AR(1) model initially considered by Rosenberg (1973). The model may be written as

$$\beta_{j,t} = \bar{\beta}_j + e_{j,t} + u_{j,t} \quad (8)$$

where $u_{j,t}$ follows an AR(1) process. The model implied is essentially an AR(1) model plus random measurement noise which can be reparameterised as an ARMA(1,1)¹¹ model. The model can be interpreted as stating that the beta coefficient in interval t is equal to an average constant beta plus a random component. However, unlike the random coefficients model, the random component comprises two separate and statistically independent parts. The first part (the $e_{j,t}$) is random and serially uncorrelated as with the random coefficients model. The second part (the $u_{j,t}$) is serially correlated. In practical terms the serial correlation implies that if the beta coefficient in interval t is much larger than the average constant beta (i.e. if $u_{j,t}$ is large and positive) then in the next interval ($t+1$) the beta coefficient will be expected to be larger than the average constant beta as well (because the $u_{j,t+1}$ is correlated¹² with $u_{j,t}$ and so will also be positive). Therefore a *best* predictor of the beta in the next period will have to combine information about the average constant beta *and* information about the current beta. The model can be described as *mean-reverting* in that the autoregressive component in the random term implies that a high beta is likely to be followed by a value which may be high, but which is *less extreme* and closer to some 'average' beta value.

Ohlson and Rosenberg found significant support for their model using NYSE data which was confirmed by Collins *et al* (1987) in a larger study. Chang and Weiss (1991) also reported evidence in support of the ARMA(1,1) model using formal ARIMA modelling of betas estimated from a series of short non-overlapping periods. They pointed out that the autoregressive nature *decreases* as the length of time chosen to estimate beta *increases* and suggested that this may be the cause for the inconsistency of results reported in the literature.

¹¹ For a fuller discussion of ARIMA modelling see Box and Jenkins (1976).

¹² In the literature it is usually assumed that the correlation is positive.

On the Dutch market De Jong, Kemna and Kloek (1992) based their study on a mean-reverting beta within a GARCH¹³ framework. They, too, rejected the hypothesis of beta stability for securities. In addition, they pointed out that the proper ARMA(1,1) specification of the model leads to estimation difficulties; therefore proper AR(1) models (which do not have the $e_{j,t}$ component) are generally tested in the literature instead of the full Ohlson and Rosenberg model.¹⁴

On the Australian market, Faff, Lee and Fry (1992) found evidence in support of an AR(1) model as an alternative to stability. In arriving at their conclusion they derived and implemented a locally best invariant (LBI) test similar to that of King (1987) (which is an extension of one developed by Watson and Engle (1985)¹⁵). Brooks, Faff and Lee (1992) subsequently augmented the research by comparing the fit of the AR(1) and random coefficients models.¹⁶ They concluded that the simpler random coefficients model was better supported by the empirical evidence.

2.5.4 Other models which have been proposed to incorporate beta instability

Fabozzi and Francis (1977) speculated that a firm's beta assumed one of two values depending on whether the market was currently bullish or bearish. They found no conclusive empirical evidence for their hypothesis, but later found some weak evidence linking security betas with the business cycle (Francis and Fabozzi, 1979). Bradfield, Affleck-Graves and Barr (1982) found similar results for the JSE.

Hays and Upton (1986) explored a discrete *regimes* model for betas: in their model the existence of a small number of discrete so-called regimes were postulated and security betas were allowed to shift from one stable regime to another. The betas were assumed to remain in each regime for some extended number of observations. The *shifting* regimes model is in contrast to the *switching* regimes model examined by Kon and Jen (1978) where betas switch back and forth between regimes with no noticeable stability in each regime. Hays and Upton

¹³ Generalized, autoregressive conditionally heteroscedastic models.

¹⁴ In fact, some authors have not always made it clear to which model they are referring.

¹⁵ For a comparison of the virtues of these various tests, see Shively (1988) who also derives an exact small sample test.

¹⁶ For details of their tests see Brooks (1991a, 1991b) and Brooks and King (1991).

(1986) focused predominantly on a methodology for detecting shifts in regimes and measuring the frequency, size and nature of them as only a secondary consideration. The regimes approach to time-varying betas differs from the other models described in the previous subsections (in which the betas are allowed to adopt *any* real number) in that it proposes a finite set of values for the betas.

The low power of the statistical tests has been widely mooted as the reason for the apparent inconclusiveness in the literature in discriminating between the various models which allow for time-varying betas. In the next section some of the developments in fitting time-varying beta models are discussed.

2.6 Estimation of time-varying beta models

Kalman (1960) introduced a recursive procedure for updating the estimate of a parameter one observation at a time. His work was derived in the engineering context, but the *Kalman filter*,¹⁷ as his technique has become known, has acquired wider applications. In the filter new observations are considered in the light of previous estimates of beta and their information content is filtered into the current estimate. An observation which is out of line with the current estimate of beta is considered to be reflecting one of two things: an outlying return driven by risks uniquely associated with the firm, or a return caused by a shift in the underlying systematic risk structure in the firm.

Filtering and smoothing techniques (often referred to as state-space models) aid in the interpretation of time series observations. They categorize the movements in the levels of observed share returns as either real shifts in the systematic risk of the firm or random fluctuations in the unique risk.

Chan (1985) was amongst the first researchers to implement dynamic beta estimation. Subsequent research includes Knif (1989), Ng (1991) and, in particular, Malkamaki (1992) who contrasts several different state-space models in estimating beta on the Helsinki Stock Exchange. Malkamaki (1992) makes the important observation that the proper testing of models involves efficient parameter estimation. Affleck-Graves and Bradfield (1992) have demonstrated in simulation studies that even when the parameters are estimated using OLS (which is the most

¹⁷ Kalman's (1960) filter involves the assumption of normality in the disturbance terms. The term *Kalman filter* has acquired a slightly broader meaning in some of the literature and has often been taken to encompass the non-linear, non-normal filters which have subsequently evolved. See Harvey (1989) or Meinhold and Singpurwalla (1983) for readable descriptions of the Kalman filter.

efficient estimator when the constant-beta, normal-residual-distribution model holds) the power of both univariate and multivariate tests of asset pricing models is extremely low. In stock markets where the betas of some shares appear to be unstable, and where the return distributions are non-normal, the power of the tests can be expected to be reduced. Malkamaki (1992) demonstrated that tests of the CAPM show far more favourable support for the CAPM when dynamic beta estimates are used. Unfortunately, the Kalman filter which is typically used to estimate state-space models is non-robust to non-normally distributed returns and rather more technical and computationally expensive techniques are required to adjust for the sensitivity of the usual Kalman filter.

Much of the research done in the field of robust state-space modelling has appeared in the engineering and process control literature.¹⁸ The ensuing discussion contains a brief description of some of the suggestions that have emerged with regard to the state-space models which incorporates non-normality.

When the assumption of normality cannot be made, the recursive nature of the Kalman filter becomes difficult to implement in closed form. Kitagawa (1987) proposed numerical integration to complete the filtering process. However, his method is computationally very laborious and therefore very time consuming even on powerful computers.¹⁹ As an alternative, Carlin, Polson and Stoffer (1992) tackled the problem by using a Monte Carlo approach based on the Gibbs sampler. Their method, however, still remains rather time-consuming to implement routinely.²⁰

In an important development in robust filtering, Harrison and Stevens (1976) invented the *multi-process* Kalman filter to model noisy time series. In their 1983 paper Smith and West gave an explicit example of how to implement the multi-process filter. As an example of the application of the filter to finance, Murray (1992) has implemented the Smith and West model on a single data series from the JSE and concluded that it offered better insights into the risk

¹⁸ De Jong, Kemna and Kloek (1992) rank among the few researchers in financial applications to have implemented a robust filter. The approach that they use is the usual Kalman filter with parameter estimation based on maximising the likelihood of a *t*-distribution.

¹⁹ The author has programmed Kitagawa's method and found, along with other researchers (see, for example, the comments following Kitagawa's 1987 paper) that the method is not suitable for routine parameter estimation. When more powerful computers are available, Kitagawa's method may indeed prove useful. Until then it appears as if empiricists who want to implement a robust filter for many securities will have to use alternative filtering procedures.

²⁰ The reader is referred to West, Harrison and Migon (1985) for a Bayesian alternative to robust Kalman filters.

regimes of the security.²¹ The emphasis of the Smith and West proposal is on calculating the probabilities that a process is in a given *state* at a particular time; e.g. in a finance context the probabilities might be that a particular observation is an outlier, or is a consequence of a change in systematic risk. Because of this emphasis the procedure appears to be better directed at on-line estimation for engineering processes (where observations are made at higher frequencies) rather than the estimation of systematic risk.

As an alternative to specifying a number of states for observations, Meinhold and Singpurwalla (1989) proposed a robustification of the Kalman filter based on an approximation involving *poly-t distributions*. They suggested replacing the normal densities in the Kalman filter with Student's *t*-distributions, which are blended together in the updating and prediction steps of the filter. They recommended that their robustification is best suited to univariate analysis.

An earlier theorem proved by Masreliez (1975) and subsequent developments by Martin (1979, 1981) have resulted in a robust state-space procedure based on *approximate maximum likelihood*. Their approach offers scope for generalisation to multivariate analysis.

The additional insights which may be offered by a multivariate approach to shifting betas have not been researched in any depth as far as the author is aware. Because the average beta of all stocks is unity²² an upward shift in one security's beta must be counter-balanced by a downward shift in the beta of at least one other security. The multivariate approach to time-varying beta estimation is considered in the last chapter of the thesis.

²¹ Murray (1992) unfortunately bases some his figures on those given in the Smith and West paper, which dealt with the monitoring of renal transplants rather than the changes in systematic risk.

²² Provided the average is a weighted one with weights corresponding to market capitalizations in the case where a market capitalization-weighted index is used as market proxy.

CHAPTER 3

THE THINLY TRADED ENVIRONMENT OF THE JOHANNESBURG STOCK EXCHANGE

3.1 Introduction

As discussed in Chapter 2 of this thesis, infrequent trading of securities on small stock exchanges constitutes one of the most severe problems for accurate and precise estimation of risk statistics. The focus in this chapter is on the thin trading which exists in a particular small stock exchange, namely the Johannesburg Stock Exchange (JSE). This introductory analysis is necessary to establish thin trading characteristics for the simulation studies for assessing betas estimation techniques that are conducted in Chapter 4. The results are also used in the analytical development of adjustment factors in covariance estimation (discussed in Chapter 8).

In the light of this and in order to enhance readability, a style has been adopted of discussing in the main body of the chapter only the most important results for use in later chapters. Many of the rather tedious empirical results required for a comprehensive investigation into thin trading have therefore been detailed in appendices at the end of this chapter.

It is well known that infrequent trading introduces bias into ordinary least squares (OLS) estimates of risk, especially market risk or beta. Several researchers (e.g. Dimson (1979), Cohen *et al* (1983), Scholes and Williams (1977)) have developed estimation procedures which correct for the bias in beta estimates. Somewhat ironically, most of the research and applications of the correction procedures have been conducted on the larger, more well traded exchanges, i.e. New York and London, and have not necessarily been designed for small and consequently very thinly traded markets. For example, the Johannesburg Stock Exchange (JSE) is a relatively small stock exchange which experiences very low trading frequencies for many shares (see Bradfield, 1989). The successful development and selection of a suitable correction procedure for beta estimation on the JSE therefore requires an accurate description of the extent of thin trading and of the stochastic basis underlying the trading process. The above-mentioned two issues are the focus of the first part of this chapter. In the second part of the chapter the issue of the stability of thin trading parameters is examined.

Research into the stability of thin trading parameters aids in the interpretation of previous studies in which the biases caused by thin trading in estimates of beta were not corrected.

This chapter extends the work of Bradfield (1989) in the field of thin trading assessment on the JSE in three areas: firstly, the view is adopted that the analysis of *daily* data is a necessary refinement of the work done on thin trading on the JSE thus far.¹ It is argued that daily data will lead to more precise estimation of risk measures. Secondly, a *model* of the trading process which accounts for the thin trading experienced by securities on the JSE is presented and graphically appraised. The third area is that of the *stability* of measures of trading frequency. Information about the stability of thin trading permits the extent to which *past* studies involving systematic risk have been biased to be assessed in the light of *current* levels of thin trading. Because data were available in weekly form only for long periods, the examination of stability could not be performed using daily data. Despite this limitation, it is felt that useful and necessary information concerning stability can be gleaned from the study.

In section 3.2 the theoretical effects that thin trading has on the traditional OLS estimator and two other estimators which have evolved in the literature are reviewed. The theoretical development enables the specific characteristics of infrequent trading which cause the biases in estimates of beta to be highlighted. In the section 3.3 a simple model for estimating the various characteristics of thin trading determined in section 3.2 is proposed. The model uses a single parameter to measure the infrequency of trading for each share. In section 3.4 the suitability of the model for describing the trading process is investigated in a graphical comparison of the modelled and empirical characteristics of thin trading. Section 3.5 contains a theoretical assessment of the consequences of using weekly input data instead of daily data. In section 3.6 the stability of thin trading parameters is probed. In sub-section 3.6.1 the statistical significance of any changes in the levels of thin trading from one year to the next are tested. In sub-section 3.6.2, the test of stability is extended by considering the measures of thin trading at points in time which are *several* years apart. In an attempt to gain additional insights into the movement of measures of thin trading, a model of the changes in thin trading is suggested in section 3.6.3 and implemented using JSE data. In the final section of this chapter, section 3.7, the conclusions of the research into thin trading are presented.

It should be emphasized that this chapter forms a background for the later chapters in which beta estimators are contrasted and assessed. The appendices also provide the results of

¹ Due to data restrictions, only weekly data were used in previous studies of thin trading on the JSE.

additional descriptions of the extent of thin trading. The results have been relegated to the appendices only because they do not have direct applicability to studies of the estimation of systematic risk.

3.2 Beta estimation for infrequently traded shares

If a share does not trade in a given day then its recorded price is usually taken as the price at which the share was last traded. The security prices may thus be several days or weeks old and therefore month-end prices are subject to possible error in that they may not reflect the current value of the security. The period of time between the last trade and the end of the month is referred to as the month-end price age and it turns out to be one of the most important characteristics of thin trading in the context of beta estimation. In order to quantify the theoretical biases in estimates of beta caused by using 'recorded' prices rather than 'true' prices, the traditional approach to risk estimation is considered. Risk statistics are traditionally estimated from the market model:

$$r_{jt} = \alpha_j + \beta_j r_{mt} + e_{jt} \quad (1)$$

where α_j and β_j are parameters unique to security j , r_{jt} and r_{mt} are the returns on security j and the market in month t and e_{jt} is an error term to account for any fluctuation in the return in the security other than that explainable by market movements.

An errors-in-variables type problem occurs in the estimation of β_j using equation (1) because for thinly traded securities, the intervals over which the returns on the security and the market are measured differ as a consequence of the month-end price age phenomenon referred to earlier. In more extreme cases of thin trading where a share may not trade for an entire month, a (false) zero return on the security may be recorded.

In the remainder of this section the biases in various estimators of beta caused by thin trading will be detailed. The reason for establishing the manner in which thin trading affects estimates of beta is the identification of the characteristics of thin trading which are of particular importance in beta estimation. It is obviously possible to measure thin trading in a vast number of ways and this investigation will highlight the measures of thin trading which are most pertinent in the context of beta estimation.

The model² that generates the security returns (r_{jt}) in equation (1) is assumed to have the following form:

$$r_{jt} = \alpha_j D_{jt} + \beta_j \tilde{r}_{mt} + u_{jt} \quad (2)$$

where D_{jt} is the length of time between the last trades in successively traded months, or set to zero if there is no trade in security j during month t ;
 $\tilde{r}_{mt}$ is the return on the market over the period between the last trades in security j in successively traded months, or set to zero if there is no trade in month t ;
 u_{jt} is a zero-mean error term with variance $D_{jt} \sigma^2$, or set to zero if there is no trade in month t .

For well traded securities there is little difference between the return generating equation in (2) and the market model in equation (1). Specifically, for well traded securities D_{jt} will be equal to unity for all t and $\tilde{r}_{mt}$ will correspond exactly to r_{mt} because the last trade in each month will occur at the end of the month. On the other hand, it is clear that in a thinly traded environment the model in equation (1) will be mis-specified and consequently estimates of the parameters will be biased. In fact it is shown in Appendix 3.A that the probability limit ('asymptotic expectation') of the OLS estimator is not equal to β_j , but instead:

$$plim \hat{\beta}_{OLS} = \Psi_j (1 - \mu_{jt}) \beta_j \quad (3)$$

where Ψ_j is the probability that security j does trade at some time in a month;
 μ_{jt} is the average month-end price age (measured in months) of security j ;
 β_j is the true beta of security j .

For thinly traded securities Ψ_j will generally be less than unity because there is a positive probability that the security will not trade for the whole month. Similarly $1 - \mu_{jt}$ will also be less than unity making the OLS estimate of β_j biased downwards.³

From equation (3) it is evident that an important characteristic of thin trading for establishing the bias in the OLS estimator of beta is the average month-end price age.

² The assumptions implicit in the development which follows is that the true underlying pricing mechanism of the security is independent of the trading process, and that the market index is affected negligibly by thin trading.

³ Scholes and Williams (1977), Dimson (1979) and Cohen *et al* (1983) showed that the bias in the estimate of beta depends on the market proxy used. Therefore (in practise) some of the estimates of beta will be biased upwards as well as downwards. In this development the assumption has been that the market returns are measurable without any inefficiencies or delays caused by thin trading in the component securities of the market index.

3.2.1 Correcting for thin trading bias

To ascertain any further characteristics of thin trading which appear in the expressions for the bias in the beta estimates, the properties of some of the more widely used beta correction procedures were examined more closely. The purpose of finding the important characteristics is to establish which values a suitable model of thin trading must be able to predict.

Several researchers have derived techniques for obtaining unbiased estimates of the parameters of the market model. The trade-to-trade estimator (discussed, for instance, by Marsh (1979) and Dimson (1979)) and the Cohen estimators (which encompass all methods which use non-synchronous coefficients, e.g. Scholes-Williams (1977), Dimson (1979), Cohen *et al* (1983)) appear to be the most widely used. Other estimators which have not yet been widely implemented have been suggested by Hui *et al* (1990), McDonald and Nelson (1989) and Luoma, Martikainen and Perttunen (1993a, 1993b, 1993c), to name a few.

In the trade-to-trade method the returns on the security and on the index are measured between the times of the last trades in successively traded months. The observed returns thus correspond exactly to those which appear in equation (2) and so the problem of thin trading is avoided from the outset.

In order to improve the efficiency of the trade-to-trade estimator it is essential to correct for the heteroscedasticity in the residual component. To achieve this, the variables in (2) are usually adjusted by dividing them by $\sqrt{D_{jt}}$ (see Dimson and Marsh (1983)). Thereafter equation (4) is used for estimation purposes:

$$\frac{r_{jt}}{\sqrt{D_{jt}}} = \alpha_j \sqrt{D_{jt}} + \beta_j \frac{\tilde{r}_{mt}}{\sqrt{D_{jt}}} + \epsilon_{jt} \quad (4)$$

On the other hand, the input data for the Cohen estimator remain the returns measured over the calendar months (as with the OLS estimator). Thin trading is corrected for by taking into account the correlations between the security's returns and the returns on the market in leading, synchronous and lagged months. Some further discussion of the Cohen estimator can be found in Chapter 2. The Cohen estimator is a generalised version of most of the methods

which incorporate the use of lagged and leading coefficients. If the number of leads and lags are denoted by l_1 and l_2 then the Cohen estimator has the following form:

$$\hat{\beta}_{\text{cohen}} = \frac{\sum_{k=-l_1}^{l_2} \hat{\beta}_{OLS}^k}{\sum_{k=-l_1}^{l_2} \rho_m^k} \quad (5)$$

where
$$\hat{\beta}_{OLS}^k = \frac{\text{Cov}[r_{jt} ; r_{m(t-k)}]}{\text{Var}[r_{m(t+k)}]};$$

$$\rho_m^k = \frac{\text{Cov}[r_{m(t+k)} ; r_{mt}]}{\text{Var}[r_{mt}]}$$

If it can be assumed that there is no (or negligible) thin trading in the market index then it is appropriate to set l_1 (the number of months by which the security is assumed to lead a market index) in equation (5) above to zero. The numerator of (5) then represents the sum of the beta coefficients between the thinly traded security and the market returns in the synchronous and preceding months. For a well traded security, the lagged coefficients all have an expected value of zero, and thus in the numerator will coincide with the usual OLS estimate. The denominator corrects for any autocorrelation in the market index which is caused by thin trading in its component securities. Of course, if the market index has no thin trading effect then the expected value of the denominator is unity.⁴

Both the trade-to-trade and Cohen estimators yield consistent estimates of beta coefficients. The trade-to-trade method is a generalised least squares estimator and, provided the data are available, is therefore the best linear unbiased estimator of beta.⁵ Cohen *et al* (1983) and Fowler and Rorke (1983) give proofs of the consistency of the Cohen estimator. Turning to the issue of data requirements, implementation of the trade-to-trade method is complicated by the need for information about the date of the last trade in the month for each share as well as the matched values of a well traded index. On the other hand, the

⁴ Assuming, of course, that there is no autocorrelation in the true, underlying market returns. See Chapter 8 for a further discussion on dropping this assumption.

⁵ The BLUness (best linear unbiasedness) of the trade-to-trade estimator depends on the error terms having zero expectation and finite variance. Of course the method is also only unbiased if thin trading is the only cause of price adjustment delays. Cohen *et al* (1980) describe several other sources of friction which might result in delays.

disadvantages to the practitioner of the Cohen estimators are that they are inefficient (because they necessitate several regressions to provide one estimated beta) and also that they require suitable values for l_2 and l_1 (the number of leading and lagged coefficients) to be specified.

In the next section, the probability limit of the trade-to-trade estimator is derived for the case when data are available at a weekly frequency rather than the daily⁶ frequency which is needed for the proper implementation of the trade-to-trade estimator. Since most previous studies on the JSE have been performed using weekly data, the results in the next section are important in examining the impact of the data restriction. The probability limit of the trade-to-trade estimator using weekly data is shown to contain a further characteristic of thin trading which affects the estimation of beta coefficients in small stock markets.

3.2.2 The effectiveness of the trade-to-trade estimator when weekly data are available

As mentioned in section 3.1, most studies on the JSE to date have been disadvantaged by the restriction that data bases covering extended periods of time are in weekly form.⁷ The restriction means that trades can at best be tracked to the last week (and not the last day) in which a trade occurred and hence trade-to-trade estimates will still be biased to some extent.⁸ In Appendix 3.B the bias in the trade-to-trade estimator due to the use of weekly data instead of daily data is established. It turns out that the probability limit of the trade-to-trade estimator of beta when only weekly data are available is:

$$\text{plim } \hat{\beta}_{TT}^w = \beta_j \left[1 - \frac{\mu_{l_j}^w}{1 + \mu_{k_j}} \right] \quad (6)$$

⁶ The trade-to-trade estimator actually requires that the value of the security and the market are available at the exact trading times of the security. Because this requirement is impracticable on the JSE (and most other stock exchanges), daily data are often considered a reasonable approximation.

⁷ This problem possibly exists on many other lesser developed exchanges.

⁸ Because the Cohen estimator does not require exact trading times for its implementation, it will be unaffected by the use of weekly information as input data.

where μ_l^w is the average proportion of the week between the time of the last trade in the week and the end of the week ('week-end price age');
 μ_k is the average number of weeks between the last traded weeks in successively traded months (measured as a proportion of the length of a month);
 β_j is the true beta of security j .

The characteristics of thin trading that emerge as important from the expression given in equation (6) are the week-end price ages (μ_l^w) and the average length of time between the last trades in each month (μ_k). Note that for well traded securities both of the values will be zero and the probability limit of the weekly trade-to-trade estimator will converge to the usual OLS estimator. The week-end price age is a very similar characteristic to the month-end price age which appears in expression (3) for the probability limit of the OLS estimator and so is not considered in any more depth here. The length of time between trades is a slightly different concept and is therefore a characteristic for which the appropriateness of the trading model described in the next section is assessed separately.

In the next section, a simple model of the daily trading process of securities is described. The model enables the thin trading of a particular security to be measured by a *single* parameter rather than by a *set* of values which correspond to the different characteristics of thin trading derived previously.

3.3 A simple model of thin trading on the JSE

The aim of proposing a model for the daily trading process is to find a single measure of thin trading which can be used to analytically assess the impact of thin trading on the estimation of beta coefficients. It is envisaged that the measure of thin trading will be the parameter of a probability mass function. The properties of the chosen probability mass function will enable the additional characteristics of thin trading (as discussed in the preceding sections) to be estimated indirectly rather than directly from the data itself.

In the context of beta estimation, the characteristics of thin trading which have been shown to be important in the previous sections are, firstly, the price ages of the security and, secondly, the lengths of time between the last traded days in successive months.

The model proposed in this chapter is based on the Bernoulli process which corresponds to the discrete analogue of a Poisson process.⁹ The Bernoulli process assumes that the trades of a given security occur on a given day with constant probability and independently of any other processes and independently of whether the security had traded on any other day.¹⁰ The process is therefore equivalent to a sequence of Bernoulli trials with a single parameter corresponding to the probability that a security is traded on a given day.

It is well known that the maximum likelihood estimate of the probability that a share is not traded on a particular day is equal to the proportion of days on which the security was not traded in a sample period. However, in order to emphasize the *thin* trading phenomenon (which is the subject of this chapter), the proportion of days on which the security was *not* traded is discussed rather than the proportion of days on which the security was traded.

In the next sub-section a brief description is given of the extent of thin trading on the JSE as measured by the proportion of days in which securities were not traded.

3.3.1 The distribution of thin trading parameters on the JSE

The data used in the investigation into thin trading on the JSE were the 367 securities which had been listed on the JSE for at least a year as at 28 February 1992. The maximum number of days for which data were available on the UCT data base at the time of the study for any one security in the investigation was 1108 days.

For later analysis it proved useful to summarize the securities into ten deciles based on their trading frequencies. Hence the ten deciles each contained an approximately equal number of securities based on the ranked trading frequencies.¹¹ The ranges of these trading frequencies (i.e. of the proportions of days not-traded) for each decile are presented in Table 3.1. From Table 3.1 it is clear that apart from the first and last deciles (the most heavily and most thinly traded respectively) the range of each of the deciles spans approximately 10% of the range of trading frequencies. The severity of thin trading is clearly evident from Table 3.1; e.g. the

⁹ On the well-traded NYSE, Scholes and Williams (1976) and Atchison *et al* (1987) have used the Poisson process to model the trading frequency of shares *within* days.

¹⁰ Because the Bernoulli process examines trading frequency on a daily basis, an implicit assumption is that the trades of all shares on a given day occur at the same time in the day. The effect of the assumption on beta estimation is small provided returns are measured over several days rather than on a daily basis.

¹¹ In fact there were 36 securities in each of the first nine deciles and 43 in the tenth (most infrequently traded) decile.

Table 3.1

Lower and upper decile limits of percentages of days not traded by shares on the JSE

Decile	Lower %	Upper %
1	0.000	3.971
2	4.061	12.184
3	12.545	23.195
4	23.556	33.213
5	33.394	40.704
6	41.245	50.181
7	50.993	62.906
8	62.996	69.404
9	70.036	82.581
10	82.942	98.285

most thinly traded securities, which constitute decile 10, do not trade on more than eight out of every ten days.¹² This can alternatively be interpreted as indicating that 10% of the securities on the JSE trade at most in two out of every ten days.

Having examined the extent of thin trading as measured by the trading frequencies of securities, the next step is to compare the characteristics of thin trading which were identified in section 3.2 (i.e. the price ages and intervals between trades) with the estimates of those characteristics based on the proposed theoretical trading process, i.e. the Bernoulli process.

3.4 A comparison of the modelled and empirical characteristics of thin trading

The Bernoulli process is simply an assumed model for the process which actually generates trades in the securities. Although the assumption of randomly distributed trades (which forms the basis of the Bernoulli process) might appear to be a plausible simplifying model for the trading process, the objective of this section is to establish the extent to which predictions based on the model correspond with empirically measured values.

In contrast to the formal statistical testing of hypotheses in the remaining chapters of the thesis, the comparison between the estimated and the observed statistics in this chapter is

¹² An histogram of the trading frequencies and some further insights into the price ages associated with the deciles are given in Appendix 3.D.

performed graphically. The reason for the rather informal comparison is twofold. Firstly, complete accuracy in the trading frequency model is unwarranted because the magnitude of the random effects arising from other processes which affect beta estimation are large relative to any differences which might exist between the modelled and actual effects in the thin trading process.¹³ Secondly, the purpose of this chapter is to obtain a broad picture of the extent and quality of thin trading rather than any formal testing of thin trading models.

Because the primary focus of this thesis is on the estimation of systematic risk, the comparison in this section is based on those characteristics of thin trading which affect it. The characteristics were identified in section 3.2 and are the price ages (μ_l) and the interval lengths between last traded days (μ_k) in successive months.

In order to obtain empirical estimates of these characteristics a series of non-overlapping four-week intervals was used as an approximate series of monthly intervals.¹⁴ The numbers of days between the last trade in the four-week interval (μ_l) and the end of the interval, as well as the numbers of days between the last traded days in successive four-week intervals (μ_k) were measured. For each security the proportion of days on which it was not traded was also calculated and was used as an estimate for the probability that the security would not trade on a particular day. This measure of thin trading is denoted by q .

3.4.1 The month-end price ages

The average month-end price ages (μ_l) are probably the most important figures in establishing the biases in OLS estimates of beta. In Figure 3.a the observed values of the month-end price ages are denoted by crosses and the theoretical 95% confidence intervals are illustrated.¹⁵ Along the horizontal axis are the measures of thin trading which are used in the model of thin trading: namely, the proportion of days not traded by the security.

It can be noted from Figure 3.a that the modelled 95% bounds cover most of the observed month-end price ages. A number of the observed values do indeed fall outside the confidence bounds, particularly for the range of thin trading measures between 0.5 and 0.7.

¹³ Note that even though the *differences* between the modelled and actual effects of thin trading on estimates of beta may be small, the *magnitude* of the biases is substantial.

¹⁴ Four-weekly intervals were used because there are a constant number of days in a four week interval as opposed to calendar intervals. The constant number of days facilitated the computation of the theoretical values.

¹⁵ The derivations in Appendix 3.C show how the confidence bands are computed.

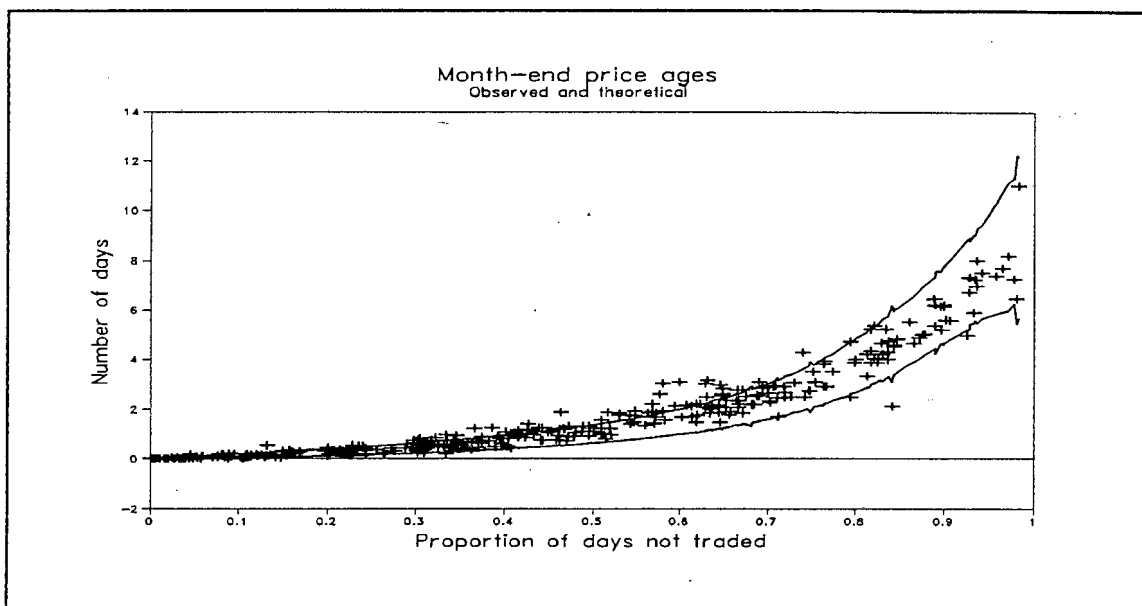


Figure 3.a

In particular, there is an indication that for this range of q values the model slightly under-predicts the price ages. Nevertheless the shape of the curve representing the relationship between q and the price age is reflected by the model and reasonable estimations can be made of the price age, especially at the higher levels of thin trading where it is arguably most important. In fact, for no values of q does the theoretically expected value differ from the observed by more than about two days.

3.4.2 The number of days between the last traded days in successive months

For the number of days between the last traded days in successive months, the model again appears to follow the shape of the observed data. (See Figure 3.b.) The reasonably good correspondence between the model and the empirical values suggests that the model is adequate for assessing the weekly trade-to-trade estimator which was discussed in section 3.2.2. The sharp rate at which the number of days between the last trades increases at the higher levels of thin trading has some interesting implications which are discussed later in section 3.5.

In the next sub-section the assessment of the model is taken a stage further by comparing the *observed* ratio of OLS beta estimates to the trade-to-trade estimates with the *expected* ratio.

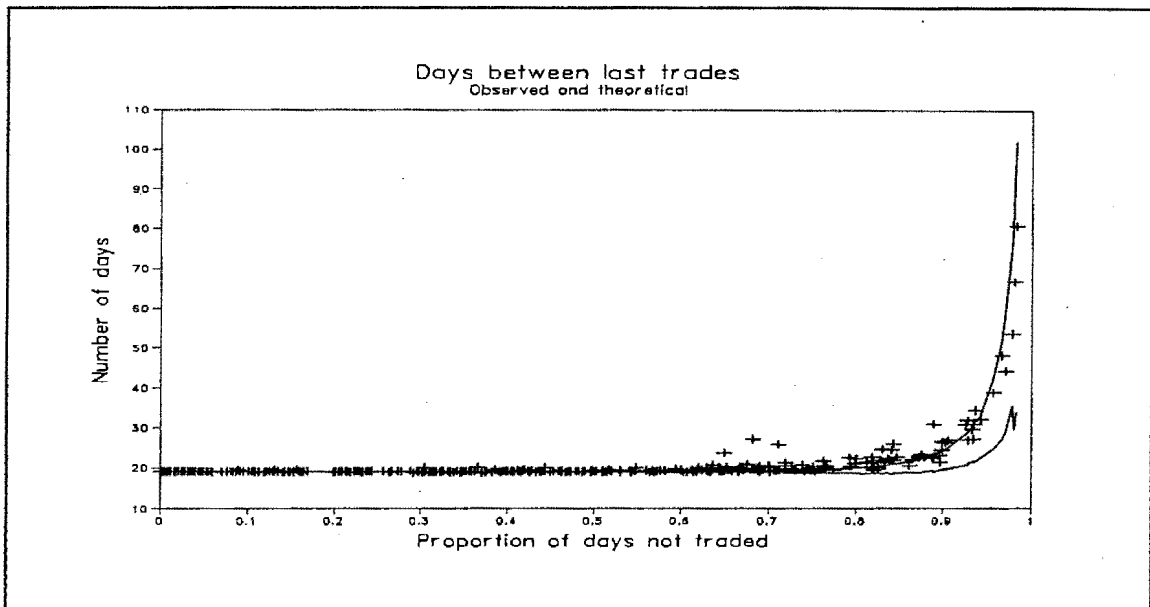


Figure 3.b

3.4.3 Assessing the model of thin trading and empirical values using actual beta estimates

While it is obvious that the Bernoulli model does not fit the observed trading process exactly, the model is reasonable particularly if the differences between modelled and observed effects are negligible in the context that they will be used in the thesis, i.e. beta estimation. Estimates of beta are subject to substantial random error relative to the size of the bias caused by thin trading. In fact, for individual securities the extent of random error will swamp any discrepancies between the model and reality.

In order to gain further insights into the effectiveness of the model in the context of beta estimation, plots were made of the ratios of the *expected* OLS betas (obtained from equation (3)) to the *expected* trade-to-trade betas (expression (6)) and the ratios of the *actual* OLS betas to the *actual* trade-to-trade betas. The same sample of shares used in the previous investigations were used. Figure 3.c shows the two plots. It is evident that the expected ratios describe the overall shape of the actual ratios and do not appear to be consistently too large or too small relative to the actual ratios. One notable feature, however, is that in the range of thin trading between $q = 0.5$ and $q = 0.8$ that the observed ratios are extremely widespread around the expected ratios (which confirms the extent of the randomness involved in beta estimation).

In the next section the model proposed for thin trading is used to investigate a practical problem which faces practitioners requiring estimates of beta using the trade-to-trade technique.

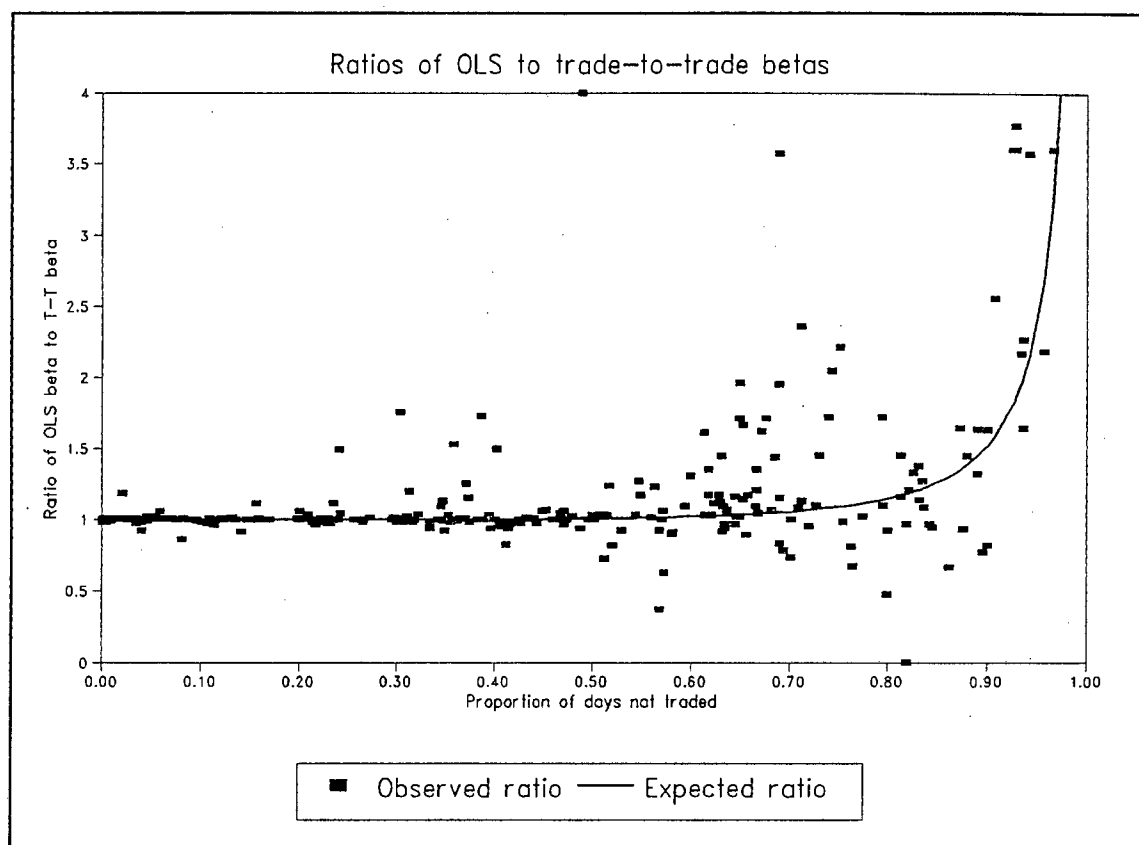


Figure 3.c

The problem is that on small stock markets data may not be available at daily frequencies (which might be regarded as a minimum requirement for implementation of the trade-to-trade technique discussed in section 3.2.1). Weekly data is often more readily available and a useful contribution for estimators of beta is knowledge about the bias which remains in the estimate of beta when weekly data have to be used as input for the trade-to-trade estimator.

3.5 The case when only weekly data are available for beta estimation

In this section two figures will be presented which have a significant bearing on beta estimation when weekly data are available rather than daily data. The figures illustrate and contrast the effectiveness of two different ways of implementing the trade-to-trade estimator using only weekly data. OLS betas are also included in the plots. There are therefore two comparisons being made. Firstly, there is the comparison between the two different ways of

implementing the trade-to-trade estimator. Secondly, there is a comparison between the trade-to-trade and OLS estimators.

In Figure 3.d the expected beta estimates of the OLS and the weekly trade-to-trade methods, calculated using equations (3) and (6) are plotted. Note that for the weekly trade-to-trade beta in this case the estimator uses as inputs the returns measured between the last traded weeks in each month. It is evident from Figure 3.d that the trade-to-trade estimator based on

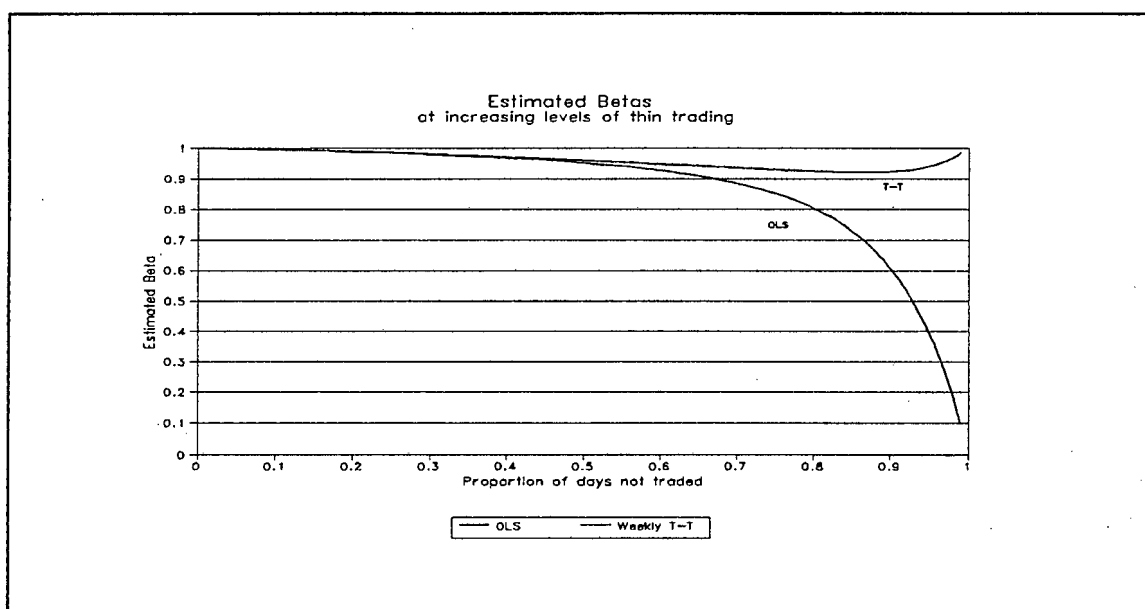


Figure 3.d

weekly data is within 10% of the true beta for all levels of thin trading. Note that if daily data were available, the theoretical trade-to-trade betas would be represented by a straight line at unity.

Another interesting fact which is illustrated in Figure 3.d is that for extremely thinly-traded shares, the trade-to-trade estimator is closer to unity than the trade-to-trade estimate of betas of moderately thinly-traded shares. In order to explain this phenomenon it must be noted from equation (6) that the bias component of the trade-to-trade estimator based on weekly data depends on the *ratio* of the price age (μ_l) of the security to the length of time between the trades (μ_k). The reduction in the bias for very thinly traded securities occurs because for those shares there are several months in which no trades occur at all. Hence the total time between last traded days is increased and makes the price age of the security relatively less important in the bias component.

A very different pair of estimators are presented in Figure 3.e. In this figure the trade-to-trade estimator is assumed to be obtained using returns measured between *successively traded weeks*. It should be emphasized that this estimator is therefore very different from measuring the returns between the *last traded weeks in successive months*. Practitioners may wish to use

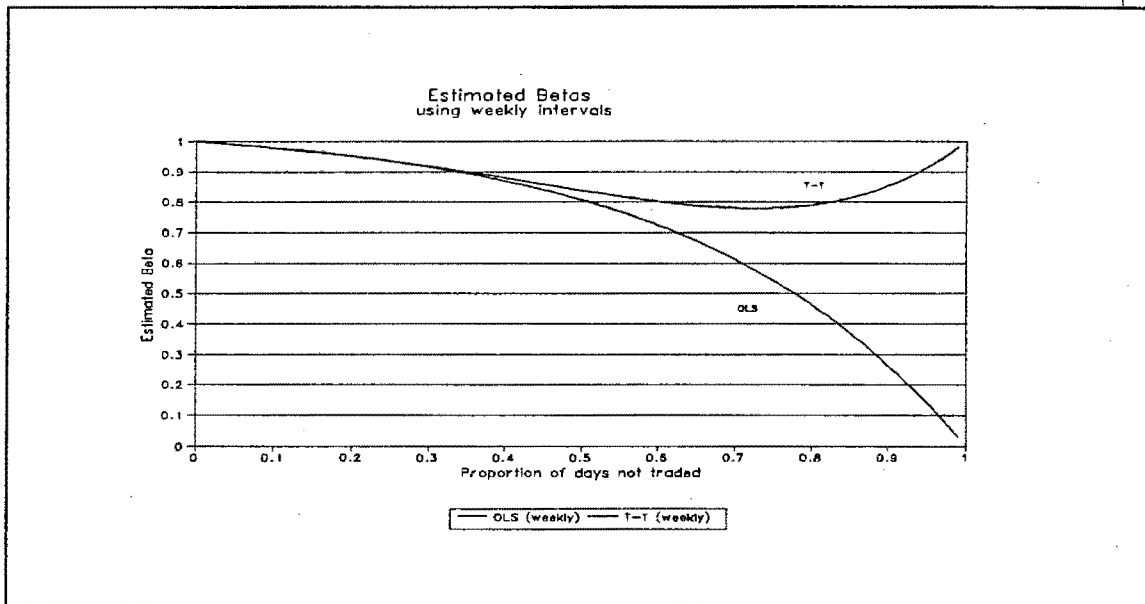


Figure 3.e

this form of the trade-to-trade estimator in order to increase the number of observations and hence the apparent efficiency of the estimator. The purpose of this study is to show that the accuracy of the estimator is compromised by not using the last traded weeks of successively traded months and using successively traded weeks. The OLS estimator presented in Figure 3.e is also different from the OLS estimator in Figure 3.d. In this version of the OLS estimator, the input data are the returns measured from the recorded prices each *week* rather than each month in the case of the OLS estimator in Figure 3.d. The difference between the trade-to-trade estimator and the OLS estimator in this case is therefore that the trade-to-trade estimator omits weeks in which no trade occurs whereas the OLS estimator does not.

The expected values of the OLS betas using weekly interval lengths are correspondingly more biased than the expected values of the OLS betas based on monthly intervals. The increase in bias caused by using the reported weekly data reflects the increased bias due to the increased number of intervals in which no trades occur at all. The increased number of intervals in which no trades occur implies that the parameter Ψ_j in equation (3) will be closer to zero and so the bias in the OLS estimate of beta will be even more severe.

The weekly trade-to-trade betas where the trades are measured between the last traded weeks in the month are also less biased than the trade-to-trade betas based on successively traded weeks. The reason for this difference is the fact that for the trade-to-trade estimator based on the last traded weeks in successive months, the time between the last trade and the end of the interval is a far smaller proportion of the total time between trades than in the case when the successively traded weeks are used.

The next section represents the start of a new part of the investigation into thin trading. The focus of the study changes from *modelling* the process which generates daily trades to *testing* the stability of the measures of thin trading. Unfortunately the only data available for extended periods of time were at a weekly frequency. Hence the study is limited to examining the stability of the estimated probabilities that a security does not trade in a given week.

3.6 Stability of measures of thin trading

In this section, the stability of trading frequency is examined. The previous sections of this chapter indicated that the estimated probability that a security does not trade in a particular interval is a suitable measure of trading frequency in the context of beta estimation. The stability of trading frequency is therefore assessed by testing for significant changes in the estimated probabilities of trades occurring. As indicated above, the data were gathered on a weekly basis due to the restricted availability of daily data.

To facilitate discussion, the estimated probability that a security j was not traded in a given week in year i is denoted by q_{ji} . The estimated probabilities are also referred to as the *annual trading frequencies* of the securities. An empirical study of the stability of annual trading frequencies was conducted using the weekly trading histories of the 246 securities which had been listed continuously from February 1973¹⁶ until December 1991.¹⁷

A preliminary overall test was conducted to assess the stability of the annual trading frequencies of each security over the entire 19 year period. Inference about the stability was based on a chi-squared test described in Appendix 3.G. The results indicate that 62 (i.e. 25%) of the shares displayed overall stability of their q_{ji} . For a large percentage (i.e. 75%) of

¹⁶ The first date available on the data base at the University of Cape Town.

¹⁷ Some descriptive cross-sectional statistics about the trading frequencies can be found in Appendix 3.H.

securities on the stock exchange, trading frequency is not constant over the years and warrants further investigation.

3.6.1 Short term stability of annual trading frequencies

In order to gain a clearer picture of the stability of security trading frequencies, the stability of thin trading measures from one year to the next was tested. Table 3.2 presents the results of the tests.

Table 3.2

Significant year-on-year changes in measures of trading frequency

Year ^a	No. of changes (out of 246)	More thinly traded		Less thinly traded		Overall
		No ^b	Med ^c	No ^d	Med ^e	
73-74	6	6	11.7	0	-	2.1
74-75	3	3	11.5	0	-	2.9
75-76	2	2	43.2	0	-	5.7
76-77	6	3	9.6	3	32.7	6.7
77-78	3	1	69.2	2	81.7	5.8
78-79	5	0	-	5	44.2	5.8
79-80	6	0	-	6	84.6	1.9
80-81	6	4	16.0	2	77.4	1.9
81-82	3	3	26.9	0	-	1.9
82-83	6	1	53.8	5	55.8	1.9
83-84	8	3	23.1	5	82.7	2.9
84-85	1	1	26.9	0	-	3.8
85-86	6	0	-	6	48.1	3.8
86-87	3	1	5.7	2	57.5	1.9
87-88	6	6	31.4	0	-	2.0
88-89	7	2	11.5	5	51.9	3.8
89-90	8	7	13.5	1	78.8	1.9
90-91	12	11	17.3	1	32.7	7.7
Ave. ^f	5.39	3.00	24.8	2.39	60.7	3.6

^a E.g. 73-74 implies that the estimates from 1973 and 1974 were contrasted.

^b Number of shares significantly more thinly traded in year after estimation.

^c Median value of q_{it} (in the year of estimation) of those shares that became more thinly traded.

^d Number of shares significantly more frequently traded in year after estimation.

^e Median value of q_{it} of those shares that became more frequently traded.

^f Median value of the q_{it} for all shares in the year of estimation.

^g Average of figures in the column above.

From Table 3.2 it is evident that the number of shares (out of a total of 246 shares) which significantly altered trading frequency from one year to the next varies from one (in

1984/85) to 12 (in 1990/91) with a mean of about 5.39 (2.20% of 246). When it is taken into consideration that the test is at an approximately 5% size, it is evident that no more shares change trading frequency from one year to the next than could be attributed to chance.

It is interesting to note that the direction (i.e. whether the change in stability was towards being *more* thinly traded or *less* thinly traded) of the majority of changes in any one year is unrelated to the median q_{ji} in each of the years; e.g. more shares become more thinly traded from 1989 to 1990 than more frequently traded and yet the median falls from 3.8% to 1.9% (a comparison which on its own would suggest that, on average, shares were more frequently traded in 1990 than in 1989).

The shares which do change trading frequency significantly tend to do so in the same direction. For example, all six of the significant changes in the trading frequency from 1982 to 1983 are towards being 'more thinly traded'. To test whether the apparent correlation between the shares experiencing significant year-on-year changes is a real feature of the data, a chi-squared statistic was calculated based on the null hypothesis of equally likely movements in either direction.¹⁸ The statistic is written as

$$\chi^2 = \sum_{k=1}^{18} \frac{(r_k - n_k \times 0.5)^2}{n_k \times 0.5^2} \quad (7)$$

where r_k and n_k are the number of changes towards less frequent trading and the total number of changes in year k , respectively.

Asymptotically χ^2 has a chi-square distribution with 18 degrees of freedom. The value of the test statistic is 56.62 which is very highly significant at the 0.005 level, indicating that the proportion of 'more thinly traded' changes is significantly different from one-half of the total number of changes in a year (the expected number of shares becoming more thinly-traded if there were no predominance of direction of change in any one year).

Another interesting feature of Table 3.2 is that the median values of the q_{ji} which do *decrease* significantly are *higher* than the median values of the q_{ji} which *increase*. This phenomenon can be noted from the last row of Table 3.2 where it is shown that the average median trading frequency of the securities which became *less* thinly traded was 60.7, compared with an average of 24.8 for those securities which became *more* thinly traded. Contrasting the

¹⁸ The test is therefore a goodness-of-fit test.

median trading frequencies with the values in the last column of Table 3.2, it is evident that, in all cases, the median values of q_{ji} for the securities which do experience a change are well above the overall median value of q_{ji} . It therefore appears as if securities which are more thinly traded are more likely to change the level of their trading frequency than securities which are well traded.

The conclusion inferred from the above results is that the annual trading frequencies of nearly all individual securities do not vary significantly from year to year. The implication which emerges is that current estimates of trading frequency are good indicators of trading frequency in adjacent years.

3.6.2 Long term stability of measures of thin trading

The previous section dealt specifically with statistically significant changes in trading frequencies from one year to the next. In this section the testing procedure is extended to test for significant changes in the levels of thin trading *several* years apart. The purpose of the extended test is to ascertain whether there exists a drift or trend in trading frequency which eventually results in a significant difference (even though there is short term stability in trading frequency).

Table 3.3 gives the average number of securities (out of 246) which experienced significant changes in trading frequency in all the pairs of years which are l , say, years apart.¹⁹ It can be noted from the table that the average number of shares with significant differences in trading frequency increases with the amount of time between the two estimates. For example, there are, on average, nearly 34 securities which experience significant changes in trading frequency over a ten year period compared with an average of just over 5 over a one year period. The rows labelled 's.d.' in Table 3.3 give the standard deviations of the numbers of significant changes at each value of l (the interval between years). The standard deviations are intended to aid in assessing the magnitude of the average number of changes. For example, over an interval of 10 years (i.e. $i = 10$) it can be seen from Table 3.3 that the average number of changes in the sample was 33.89. When compared with the standard deviation of the number of significant changes (i.e. 10.517), it is apparent that the mean number of changes

¹⁹ The full table giving the number of securities experiencing significant changes in trading frequency for each specific year is given in Appendix 3.F.

Table 3.3

The average number of significant changes at lags, l , from 1 to 18.

l	1	2	3	4	5	6
mean	5.39	11.12	15.88	20.33	21.14	22.54
s.d.	2.615	5.207	7.420	9.100	8.094	7.434
l	7	8	9	10	11	12
mean	23.58	27.27	30.40	33.89	34.88	35.00
s.d.	7.573	9.624	7.763	10.517	12.005	9.433
l	13	14	15	16	17	18
mean	36.67	37.40	39.50	41.00	46.00	59.00
s.d.	8.733	7.092	7.724	7.000	9.900	undef.

undef. - s.d. undefined because only one datum available at 18 lags

is 'significantly' greater than zero.²⁰

The inference which can be drawn from the information presented in this section is that there may be a trend in the trading frequency of individual shares which is not discernible in significant one-period changes, but which manifests itself only over several periods.

In the next section a model for formally testing the trend hypothesis is discussed and implemented.

3.6.3 Testing for a trend in the thin trading of individual securities

In this final section on thin trading a model is constructed to test firstly for *trends* in the thin trading parameters of individual shares and, secondly, for any dependency between the *differences* over time in the levels of thin trading of individual securities and the differences in the level of market activity. The second variable, market activity, was included to test whether changes in the level of thin trading were sensitive to some overall measure of thin trading. The measure of market activity for a calendar year was defined to be the average number of

²⁰ The number of securities with significant changes in trading frequencies is certainly not distributed normally and so no formal test of significance is intimated here. All that is being demonstrated by the comparison is that the average number of changes is *large* relative to the standard deviation.

securities which had not traded in each week of the year. The measure for year i is denoted by q_i^w .²¹

The measure of market activity was chosen because it was considered plausible that if the average number of securities traded each week decreased from one year to another, then the number of weeks traded by an individual security might also decrease (and vice versa). The model proposed to test for any dependencies on time differences and changes in overall market activity can be written as follows:

$$E[q_{ji} - q_{j(i+s)}] = \alpha_j s + \beta_j (q_i^w - q_{(i+s)}^w) \quad (8)$$

The dependent variable on the left hand side of equation (8) represents the expected difference in trading frequencies of a security j , over a period of s years. The first term on the right hand side is included to test for any dependence of the magnitude of the differences on the time difference (s). The second variable on the right hand side is the difference in levels of overall market activity. A significant β_j coefficient would indicate that the changes in security j 's trading frequency was dependent on changes in the overall level of market activity. Significant α_j coefficients indicate that the changes in trading frequency of security j alter linearly with time.

The model in equation (8) was fitted for the 186 securities which had shown non-negligible thin trading.²² The results indicated that none of the 186 securities had coefficients of market activity which were significantly greater than zero. The implication is that individual securities showed no signs of sensitivity to the chosen measure of overall market activity.

²¹ The behaviour of q_i^w is elaborated in Appendix 3.E where it is shown that market activity fluctuates significantly between years and between months within the year. In Appendix 3.I it is shown that the *levels* of thin trading of individual securities are predominantly unrelated to the *level* of market activity. In contrast, the model in this section tests whether the *changes* in the trading frequencies are related to *changes* in market activity.

²² The 80 shares which had traded on 99% or more of the weeks in the entire 19 year period were excluded from the analysis. The rationale for the exclusion stems from the observation in section 3.6.1 that it is only the more thinly traded securities which showed signs of instability.

A more interesting result is that 138 (i.e. 74%) of the 186 estimated²³ coefficients of s (the time differences) are significant.²⁴ It appears as if there is a linear relationship between the difference in the measures of trading frequency and the amount of time between which they are calculated. However, the magnitude of the estimated α_j coefficients are generally quite small (over 60% of the 186) are between -0.01 and 0.01). If it is noted that a coefficient of 0.01 implies that it will take 10 years before there will be an expected change of trading frequency greater than 10% and further note that the 80 securities which remained well traded throughout the period of investigation have been excluded from the testing of the model above, it is evident that the trading frequencies of most securities do not change substantially over time.

3.7 Conclusion

The purpose of this chapter was to provide insight into the extent and behaviour of thin trading on the JSE in the context of systematic risk estimation. In order to attain that objective two main investigations were undertaken. Firstly, a simple model of the trading process was proposed. Secondly, the stability of trading frequencies was measured.

The trading model suggested was based on the Bernoulli process and pin-pointed the trading frequency (measured as the proportion of days in which the security was not traded) as the key parameter in assessing thin trading. Graphical comparisons of thin trading characteristics predicted by the model with those empirically measured on the JSE confirmed the suitability of the model.

The model was then used to assess the consequences of a problem faced by many practitioners and researchers in small stock markets: namely, the limited availability of daily data. Weekly information, however, is generally more readily accessible for long periods of time. The biases caused by thin trading in estimates of beta when weekly data instead of daily data are used were derived and illustrated.

²³ The percentage of securities with significant coefficients corresponds to about 56% of the 246 securities in the whole sample.

²⁴ The values of the estimated α_j 's ranged from -.039 to .069. Of the 186 coefficients, 104 of them were negative and 82 positive which indicated a roughly even split between those securities gradually becoming better traded and those gradually becoming more thinly traded.

In the second part of the chapter the stability of trading frequencies of individual securities was investigated. Many *past* studies have incorporated beta coefficients which have not been corrected for the thin trading biases. In order to interpret the validity of the conclusions stemming from past studies, therefore, it is important to know whether they can be appraised in the light of *current* levels of thin trading. The empirical results which emerged for the JSE are that annual trading frequencies are stable in the short term. In contrast, when considering the longer term, some securities show a gradual trend in trading frequency. However, since not all the trends are in the same direction, and since there is no relationship between individual changes and changes in overall market activity, this study would indicate that current information about thin trading can be usefully applied to assess past studies.

Appendix 3.A

Probability limits of OLS parameters

This appendix contains the mathematical details of finding the probability limits for the OLS estimator of betas. It is necessary for the developments in section 3.2. Also included in this appendix is a derivation of the variance of the OLS estimator for a thinly traded security. In deriving the probability limits of the OLS estimates of alpha and beta, it is useful to introduce some further notation. The return generating equation is specified by

$$r_t = \alpha D_t + \beta \tilde{r}_m + u_t \quad (9)$$

where

$$D_t = 1 + k_t + I_{(t - k_t - 1)} - I_t$$

and

I_t is the price age in month t assuming there is a trade in month t (with mean μ_I and variance σ_I^2);

k_t is the number of untraded months between two successively traded months (with mean μ_k and variance σ_k^2);

σ_m^2 is the variance of the monthly index returns.

(Another assumption is that the trading process is stationary through time.)

The OLS beta can be written as:

$$\begin{aligned} \hat{\beta}_{OLS} &= \frac{\sum_{t=1}^n r_t (r_{mt} - \bar{r}_{mt})}{\sum_{t=1}^n (r_{mt} - \bar{r}_{mt})^2} \\ &= \frac{\sum_{t: I_t < 1} r_t (r_{mt} - \bar{r}_{mt})}{\sum_{t=1}^n (r_{mt} - \bar{r}_{mt})^2} \end{aligned} \quad (10)$$

(r_t , the reported return, will be zero whenever $I_t = 1$.)

Substituting r_t from (9) into the expression for the OLS beta and noting that because $\bar{r}_{mt}$ and r_{mt} overlap only in month t up to the date of the last trade in month t the ‘asymptotic covariance’ of the monthly market returns, r_{mt} , and the trade-to-trade market returns, $\bar{r}_{mt}$ is

$$plim (r_{mt} - \bar{r}_{mt})(\bar{r}_{mt} - \bar{r}_{mt}) = n (1 - \mu_I) \sigma_m^2 \quad (11)$$

Using the above results the derivation of the probability limit of the OLS beta follows quite easily:

$$plim \hat{\beta}_{OLS} = \Psi (1 - \mu_I) \beta \quad (12)$$

where Ψ is the probability that I_t is less than unity, i.e. the probability that there is a trade in the month.

The α parameter is estimated from

$$\hat{\alpha}_{OLS} = \bar{r}_t - \hat{\beta}_{OLS} \bar{r}_{mt} \quad (13)$$

and, because the following relationship derived from (9) also holds

$$plim r_t = \Psi (\alpha (1 + \mu_k) + \beta (1 + \mu_k) \bar{r}_{mt}) \quad (14)$$

the probability limit of α is seen to be

$$plim \hat{\alpha}_{OLS} = \Psi (\alpha (1 + \mu_k) + \beta \bar{r}_{mt} (\mu_k + \mu_I)) \quad (15)$$

The variance of beta is also affected by the existence of thin trading. To find an expression for the variance the equation for the OLS beta in (10) is used. Firstly it should be noted that the derivation is conditional on the monthly market (r_{mt}) returns being known. Also, the derivation is conditional on the trading times being known and therefore that D_t is fixed. This condition is later dropped and the additional source of variance is added back into the expression for the variance.

Noting the above conditions, it can be seen that the variance of the OLS estimator of beta depends on the variance of the observed return on the security, r_t . This variance depends

in turn (see equation (9)) on the variance of the trade-to-trade market returns ($\tilde{r}_{mt}$). It is therefore necessary to establish the variance of the trade-to-trade market returns conditional on the monthly market returns.

The conditional variance is found by defining the return on the index from the beginning of month t until the last trade in the security of interest to be r_{mt}^I and also $r_{mt}^0 = r_{mt} - r_{mt}^I$. If normality in the return distribution of the market is assumed, specifically,

$$r_{mt} \sim N(\bar{r}_{mt}; \sigma_m^2) \quad (16)$$

then it is relatively easy to show that

$$r_{mt}^0 | I \sim N((1 - I_t) \bar{r}_{mt}; (1 - I_t) I_t \sigma_m^2) \quad (17)$$

Using the above result and noting that the variances of r_{mt}^I and r_{mt}^0 are equal, it follows that the variance of the security return is

$$plim \text{Var}[r_t | I] = \beta^2 2 (1 - \mu_I) \mu_I + (1 + \mu_k) \sigma^2 \quad (18)$$

Substituting the result from (18) into (10), the variance of the OLS beta (conditional on the trading times) is

$$plim \text{Var}[\hat{\beta}_{OLS} | I] = \frac{\Psi}{n} \left[2 \beta^2 (1 - \mu_I) \mu_I + \frac{\sigma^2 (1 + \mu_k)}{\sigma_m^2} \right] \quad (19)$$

All that remains is to add back the variance caused by randomly fluctuating trading times, i.e.

$$\text{Var}[plim \hat{\beta}_{OLS} | I] = \beta^2 (1 - \mu_I)^2 \frac{\text{Var}[\Psi]}{n} \quad (20)$$

and the asymptotic variance for the OLS beta is obtained.

Appendix 3.B

Trade-to-trade betas

In this appendix the expressions for the biases in the weekly trade-to-trade beta estimators discussed in section 3.2.2 are derived. As a first step towards establishing the properties of the weekly trade-to-trade estimates, the case where use is made only of the months in which trades occur is considered. 'Gross' heteroscedasticity, i.e. that resulting from whole months not being traded, is countered by transforming the variables by the inverse of the square root of number of months over which the returns are measured. Equation (9) in Appendix 3.A then becomes

$$\begin{aligned} \frac{r_t}{\sqrt{1 + k_t}} = & \alpha \frac{(1 + k_t + I_{(t - k_t - 1)} - I_t)}{\sqrt{1 + k_t}} \\ & + \beta \frac{r_{m1}^1 + r_{m(t-1)} + \dots + r_{m(t-k_t)} + r_{m(t-k_t-1)}^0}{\sqrt{1 + k_t}} + v_t \end{aligned} \quad (21)$$

where k_t is the number of months immediately prior to month t in which no trades occurred and all the other variables are defined as in Appendix 3.A.

Estimating beta using only monthly information, i.e. using only the knowledge of whether or not the share traded in the month and the returns on the index measured over the months, involves a multiple regression, the two independent variables being

$$x_{1t} = \frac{1 + k_t}{\sqrt{1 + k_t}} = \sqrt{1 + k_t}, \text{ and}$$

$$x_{2t} = \frac{r_{m1}^1 + \dots + r_{m(t-k_t)}}{\sqrt{1 + k_t}}.$$

If the dependent variable is denoted by y_t , then the estimate of beta is given by

$$\hat{\beta} = \frac{\sum x_{2t}^2 \sum x_{1t} y_t - \sum x_{2t} y_t \sum x_{1t} x_{2t}}{\sum x_{1t}^2 \sum x_{2t}^2 - \left(\sum x_{1t} x_{2t} \right)^2} \quad (22)$$

Rewriting (21) using the new notation and expanding,

$$y_t = \alpha (x_{1t} + x_{1t}^*) + \beta (x_{2t} + x_{2t}^*) + v_t \quad (23)$$

where two new variables have been defined as

$$x_{1t}^* = \frac{I_{(t-k_t-1)} - I_t}{\sqrt{1+k_t}}$$

and

$$x_{2t}^* = \frac{r_m^2 (t-k_t-1) - r_{mt}^1}{\sqrt{1+k_t}}.$$

If the expression for the estimated beta in (22) is re-written in a form using covariances, variances and means,

$$\hat{\beta} = \frac{(\sigma_{x_1}^2 + \mu_{x_1}^2) (\sigma_{x_1 y} + \mu_{x_2} \mu_y) - (\sigma_{x_1 y} + \mu_y \mu_{x_1}) (\sigma_{x_1 x_2} + \mu_{x_1} \mu_{x_2})}{\sigma_{x_2}^2 (\sigma_{x_1}^2 + \mu_{x_1}^2) + \mu_{x_2}^2 \sigma_{x_1}^2 - \sigma_{x_1 x_2}^2 - 2 \mu_{x_1} \mu_{x_2} \sigma_{x_1 x_2}} \quad (24)$$

where, in general, σ_{ab} is the covariance between a and b , and μ_a is the mean of a , and probability limits are taken then

$$\text{plim } \hat{\beta} = \beta \left[1 - \frac{\mu_l}{1 + \mu_k} \right] \quad (25)$$

To establish the above result it should be noted that:

$$\sigma_{x_1 y} \rightarrow \beta \left[\sigma_{x_2}^2 - \frac{\mu_l \sigma_{x_2}^2}{1 + \mu_k} \right]$$

and

$$\sigma_{x_1 x_2} \rightarrow 0$$

For an estimate based on weekly data, μ_l has to be redefined to be the proportion of the week between the last trade and the end of the week, and k_t to be the number of weeks between trades (measured as a proportion of a month).

k_t is relatively easy to find on the basis that the number of days between trades follows a geometric distribution. If the probability of no trade in a day is denoted by q and the probability of no trade in a week by q^w then $q^w = q^7$.

The expected number of weeks between the last traded week in a traded month (assumed to be a four-week period, and conditional on the week being the $(4-l)$ th in the month) and the last traded week in the next traded month is given by:

$$E[k^w | l] = \sum_i \sum_{j=0}^3 [l + 4i - j](1 - q^w) q^{w \cdot 4(i-1) + j} \quad (26)$$

Therefore,

$$E[k] = \sum_{l=0}^3 \frac{(1 - q^w) q^{w \cdot l}}{1 - q^{w \cdot 4}} E[k^w | l] \quad (27)$$

Expanding and then re-simplifying (27) gives

$$E[k_t] = E[k]/4 = \left[1 + \frac{q^{w \cdot 4}}{1 - q^{w \cdot 4}} \right] \quad (28)$$

The fraction in brackets is the expected number of four-week periods (months) in which there is no trade. The expected number of months in which there are no trades between traded months will increase as the prevalence of thin trading increases. It is clear that while μ_l will increase with thin trading, it is bounded by unity. On the other hand the proportion of months not traded is not bounded and so μ_k can become very large. Referring to (25) it can be observed that at extremely high levels of thin trading when μ_k is large the bias in the beta estimate is reduced.

Appendix 3.C

Important results associated with the Bernoulli model

This appendix details some of the implications of assuming a Bernoulli process for the trading process. The results derived in this appendix are used to relate the characteristics of thin trading used in Appendix 3.A to the probability of a security trading on a given day. The results are also used in deriving the confidence intervals for graphical comparisons of the thin trading model and the empirical results in 3.4. The basic assumptions of the Bernoulli process are that a security is traded on any day, independently of all other securities and independently of whether or not it traded on any other day, with constant probability q .

The number of days between trades, x , thus follows a geometric distribution (for x is 0, 1, 2, 3, ...):

$$p(x) = (1 - q) q^x$$

The distribution of the number of days between trades is the same as that of the number of days between any arbitrary day and the next trade, using the memorylessness property of the geometric distribution. The mean and variance of the number of days until the next trade, conditional on the next trade being within k days, are:

$$k \times \mu_1 = \frac{q}{1 - q} \frac{1 - q^{(k-1)}}{1 - q^k} - (k - 1) \frac{q^k}{1 - q^k} \quad (30)$$

and

$$\begin{aligned} k^2 \times \sigma_1^2 &= \frac{2}{1 - q^k} \left[\frac{q}{1 - q} (1 - q^{(k-1)}) - (k - 1) \frac{q^k}{1 - q^k} \right] \\ &\quad - \frac{q}{1 - q} \frac{1 - q^{(k-1)}}{1 - q^k} - (k - 1)^2 \frac{q^k}{1 - q^k} - k^2 \times \mu_1^2 \end{aligned} \quad (31)$$

The proportion of intervals where there is no trade is estimable from:

$$\Psi = q^k$$

with variance:

$$Var[\Psi] = q^k (1 - q^k)$$

Appendix 3.D

More information on the distribution of daily thin trading measures

This appendix gives some summary statistics for the measures of daily trading frequency calculated in section 3.3. The proportion of days on which no trades occurred was computed for 367 companies listed on the JSE for at least year as at 28 February 1992. The maximum number of days listed for any one security on the data base at UCT at the time of

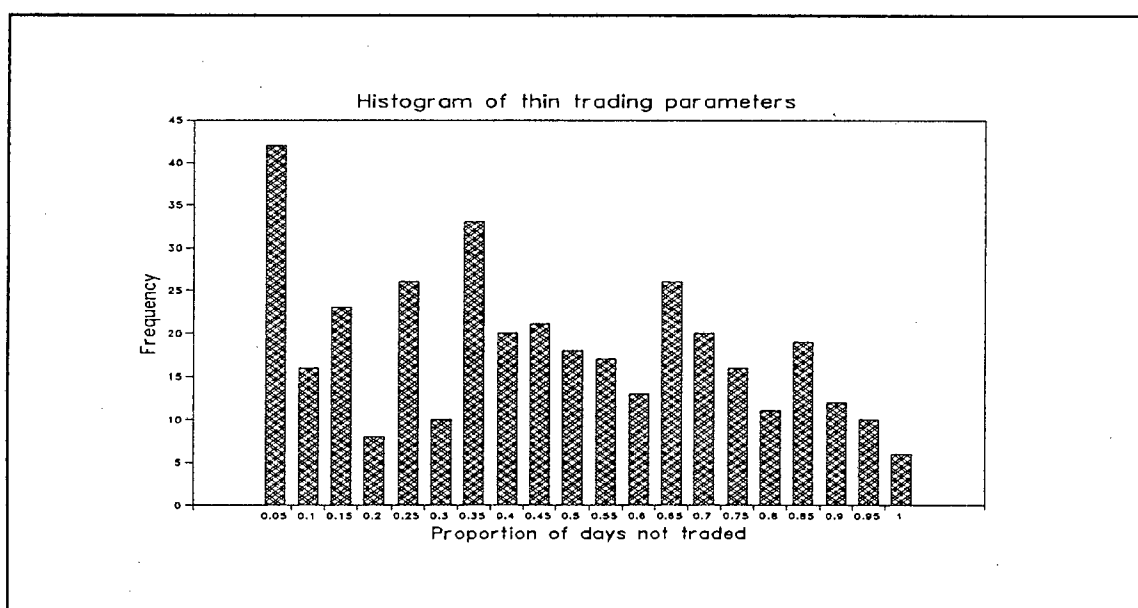


Figure 3D.1

the study was 1108 days. A histogram of these figures is displayed in Figure 3D.1.

The average proportion of days not traded is 0.431, i.e. an 'average' share trades once every $1/0.431 = 2.32$ days. The proportion of no trades range from 0.00 to 0.98 and are distributed fairly evenly over this range.

In order to gain further insights into the impact of the thin trading, the Bernoulli process assumption enabled the calculation of the *expected* number of days between trades. More information about the shape of the distribution of the number of days between trades was obtained by computing the 90th and 95th percentiles of the distribution.²⁵ Table 3D.1 gives the expected number and the 90th and 95th percentiles of the number of days between trades.

²⁵ Calculated by finding s where q^s equal 0.1 and 0.05 for the 90% and 95% ranges respectively.

The parameter values used to calculate the above figures are equal to the upper bounds of the ranges of q in each of the deciles presented in Table 3.1.

It can be ascertained, for example, from Table 3D.1 that when the proportion of days not traded is less than 0.50 (which is true for the 218 (out of the total of 367) shares in deciles one to six) then trades will be expected to be less than a day apart. Another example of a figure which can be read from Table 3D.1 is that there is a 90% probability that when the probability

Table 3D.1

The number of days expected between trades at various levels of thin trading. The 90% and 95% Ranges represent the maximum number of days that 90% and 95%, respectively, of between trade intervals will be.

Decile	q	Expected number of days between trades	90% Range (days)	95% Range (days)
1	0.04	0.04	0.71	0.93
2	0.12	0.14	1.09	1.42
3	0.23	0.30	1.58	2.05
4	0.33	0.50	2.09	2.72
5	0.41	0.69	2.56	3.33
6	0.50	1.01	3.34	4.34
7	0.63	1.70	4.97	6.46
8	0.69	2.27	6.30	8.20
9	0.83	4.74	12.03	15.65
10	0.98	57.31	133.11	173.18

of no trade on a given day is less than 0.63 (which is the case for 70% (258) of the shares) then two trades will occur within a week (5 days) of each other.

Appendix 3.E

The measure of overall market activity

In this appendix some further insights are given into the distribution of the measure of overall market activity used in section 3.6.3, namely q_i^w . The term $p_i^w (= 1 - q_i^w)$ was estimated for each year by the average number of shares traded per week in that year. The proportion of shares (of the 246 in the sample) which traded in each week of the 19-year period was

Table 3E.1

The average percentage of shares that traded in each week of the years 1973 to 1991. Also included is the change in percentage from one year to the next.

Year	Proportion traded (%)	Year-on-year change
73	89.3	-3.8
74	85.5	-3.6
75	81.9	-1.3
76	80.6	-0.1
77	80.5	2.3
78	82.8	4.0
79	86.8	3.1
80	89.9	-2.4
81	87.5	-2.7
82	84.8	1.2
83	86.0	0.3
84	86.3	-0.9
85	85.4	3.2
86	88.6	0.7
87	89.3	-4.6
88	84.7	1.7
89	86.4	-4.0
90	82.4	-3.5
91	78.9	

recorded. The sequence of proportions was regarded as an index of weekly market activity. Table 3E.1 gives the averages of these weekly proportions in each year. Similarly to the

conclusions reached in Appendix 3.H it is clear that market activity varies (from 78.9% in 1991 to 89.9% in 1980). An analysis of variance indicates that the variability in the yearly measures is significant. ($F = 24.02$ with 18 and 966 degrees of freedom).

Trading activity appears to slump in the years 1976-1977 and 1990-1991 with peaks of trading activity in 1973, 1980 and 1987. The average proportion of shares traded per week is (as might be expected) highly negatively correlated ($r = -0.89$) with the cross-sectional median levels of the thin-trading measures of individual securities which are reported in Appendix 3.H.

Variability in trading frequency between months

In testing for variation in the level of market activity from year to year as described above highly significant results were reported. In this part of the appendix the results of tests for significant variability in the levels of thin trading between months are reported.

An ANOVA similar to that for testing for significant variability in market activity between the years was performed to test for monthly variation. The results ($F = 14.21$ with 11 and 973 degrees of freedom) indicated highly significant departures from a constant level

Table 3E.2

The mean and standard deviations of the proportion of shares traded in each week of the calendar months of the year.

Month	Mean Proportion of shares traded	Std. Deviation
January	0.85	0.058
February	0.87	0.040
March	0.86	0.049
April	0.84	0.049
May	0.86	0.043
June	0.85	0.042
July	0.86	0.041
August	0.86	0.045
September	0.85	0.045
October	0.85	0.048
November	0.85	0.045
December	0.79	0.096

of market activity from month to month. On closer inspection it appeared as if the proportion traded in each month is relatively stable (as measured by the mean) except in December when activity is considerably lower. February has the highest proportion of its weeks traded. The effect of the short weeks in April (with its several public holidays) is also evident - the shorter a trading week the less likely there is to be a trade in it.

The standard deviations of the proportion of weeks traded within each month are also largely stable apart from December where it approximately doubles. The increase will be largely due to the fact that some of the trading weeks in December are very short because of the public holidays in that month whereas other weeks in the month contain a full five days.

Appendix 3.F

Table of number of shares with significant changes between years

(See section 3.6.2.)

Table 3F.1

The number of significant changes in q_{jt} between the years in the sample.

	73	74	75	76	77	78	79	80	81	82
73		6	11	21	20	14	10	7	15	20
74			3	13	7	8	11	19	22	18
75				2	5	6	16	23	23	16
76					6	9	24	34	27	23
77						3	9	33	31	24
78							5	22	19	19
79								6	14	13
80									6	8
81										3

	83	84	85	86	87	88	89	90	91
73	21	18	21	23	26	32	34	39	59
74	21	24	27	31	38	41	36	41	53
75	22	21	25	29	39	37	37	40	48
76	27	26	30	33	40	37	41	38	50
77	32	27	33	38	38	38	33	32	45
78	26	24	24	30	31	39	36	32	49
79	21	18	20	21	21	40	30	32	52
80	12	17	22	15	19	34	29	40	59
81	8	7	15	17	16	28	23	34	55
82	6	9	9	12	19	21	24	28	40
83		8	9	12	14	26	27	32	50
84			1	4	16	17	17	24	36
85				6	18	15	17	21	39
86					3	10	15	21	45
87						6	10	22	43
88							7	8	23
89								8	22
90									22

Appendix 3.G

Tests for significant changes in q between years

This appendix details the multiple comparison techniques required to test for stability in the measures of trading frequency discussed in 3.6.2. Because the differences between q_{ji} and $q_{j(i+s)}$ are tested for several different values of s , i.e. at several different lags, multiple comparisons techniques are required in order to control for the size of the tests. The tests performed were two-tailed; that is the significance of the change was not tested for in any one direction. The observed direction, either more thinly traded in year $i+s$ than in year i or less so was, however, noted.

For testing the null hypothesis $H_0: q_{j1} = q_{j2} = \dots = q_{j19}$ the customary Pearson chi-square statistic is:

$$\chi_j^2 = \sum_{k=1}^{19} \frac{(r_{jk} - n_{jk} \times \hat{q}_j)^2}{n_{jk} \times \hat{q}_j (1 - \hat{q}_j)} \quad (34)$$

where r_{jk} and n_{jk} are the number of weeks not traded by security j in year k , and the total number of weeks that j is listed in year k , respectively. The pooled estimate of trading frequency is

$$\hat{q}_j = \frac{\sum_{k=1}^{19} r_{jk}}{\sum_{k=1}^{19} n_{jk}} \quad (35)$$

Asymptotically the statistic has a chi-square distribution with 18 degrees of freedom. At a five per cent level the critical value is 28.87.

Scheffe-type projections (Kotz and Johnson, vol 5, p685) suggest declaring q_{jk} and $q_{j(k+s)}$ unequal if

$$|\hat{q}_{jk} - \hat{q}_{j(k+s)}| \geq \left(\chi_{18}^2 \right)^{\frac{1}{2}} \left[\hat{q}_j (1 - \hat{q}_j) \left(\frac{1}{n_{jk}} + \frac{1}{n_{j(k+s)}} \right) \right]^{\frac{1}{2}} \quad (36)$$

It was felt that the use of the chi-square critical value in the expression may lead to too few rejections of the null hypothesis. Because it might be considered preferable to err on the side of conservatism, i.e. falsely assess the q_{ji} to be unstable, the chi-squared value was replaced (as suggested by Kotz and Johnson) with the Tukey-Kramer critical constant: $q_{I^2, \infty}^{\alpha} / \sqrt{2}$, which is equal to approximately 3.51 at $\alpha = 0.05$, and is smaller than $(\chi_{I^2}^{2\alpha})^{1/2} = 5.37$. It should be noted that in comparison with the usual critical value of 1.96 (which corresponds to the critical value where only one pairwise comparison is made) the rejection of equality of proportions will be based on very large differences in the sample proportions.

Appendix 3.H

Extent and overall stability of weekly measures of thin trading - a cross-sectional perspective

In this part of the appendix a brief investigation is made of the changing cross-sectional distribution of thin trading in each of the years 1973 to 1991. The measures of thin trading for the individual securities are the proportion of weeks in each year not traded during the period of investigation. The measure for share j in year i is denoted by q_{ji} .

The data available for the study were the weekly trading histories of 246 shares which had been listed continuously from February 1973 (the start of the UCT database) until the end of 1991. In each year the proportion of weeks in which trading did not occur was estimated for

Table 3H.1

Cross-sectional summary statistics (minimum, lower quartile, median, upper quartile and maximum values) of q_{ji} for the whole period. Figures given as percentages.

Year	73	74	75	76	77	78	79	80	81	82
min*										
Q ₁										
Med	2.1	2.9	5.7	6.7	5.8	5.8	1.9	1.9	1.9	1.9
Q ₃	13	23	26	33	29	27	15	9.4	15	21
max	100	100	100	100	100	100	100	89	87	98

Year	83	84	85	86	87	88	89	90	91	Ave ^b
min										
Q ₁									1.9	1.0
Med	2.9	3.8	3.8	1.9	2.0	3.8	1.9	7.7	7.7	6.4
Q ₃	17	21	23	15	14	23	17	29	31	24
Max	96	96	98	83	84	94	94	94	98	90

* Zeros are omitted for clarity.

^b Ave is the proportion of weeks not traded by a share in all 19 years.

each security and cross-sectional summary statistics were calculated. Table 3H.1 gives the

statistics for the q_{ji} over the 19 year period. The minimum, maximum, median and upper and lower quartiles are presented.

The distribution of thin-trading measures across the securities is highly skewed with over a quarter of the shares trading every week in nearly every year, evidence for which is that the first quartiles in every year except 1991 are zero. The median proportion of weeks not traded ranges from 1.9% to 7.7%, which represent one and four weeks in a year, respectively.

The statistics for the overall figures²⁶ reveal that over half the shares traded on at least 93% of the weeks (which equates to 48.4 weeks in a year). About a quarter of the shares did not trade in at least one week in every four week period (i.e. a quarter of the shares have a $\bar{q}_j$ greater than 25%). (The most thinly traded share, Burlington, traded on average in only 5 weeks per year.)

The summary statistics can be compared to Bradfield's 1989 study in which he concluded 'approximately one third of JSE stocks are not traded, on average, for at least one consecutive week in every four week period'. Bradfield's study involved the 360 shares listed continuously from 1 January 1978 to 31 August 1987. The median value of Bradfield's study is 4.8%; slightly lower than the 6.4% in this investigation. The difference is a result of Bradfield's period of investigation excluding the years 1975-1977 and 1990-1991 (when the median and upper quartile values are markedly higher than in other years).²⁷

From Table 3H.1 it is apparent that the average levels of thin trading as measured by the q_{ji} do vary from year to year. There is no clear trend in the average values of the q_{ji} , but the years of more pronounced average thin trading are grouped together (1975-1978 and 1990-1991).

The spread of the q_{ji} over the range zero to unity in each year is linked closely to median value: if we measure spread with the semi-interquartile range, and note that apart from in 1991 the lower quartile is zero, we observe that the upper quartile serves effectively as a measure of spread. The correlation between the median and the upper quartile over the years is 0.91 implying a relationship between the average and spread of q_{ji} .

²⁶ Indicated by *Ave* in the table and computed as the proportions of weeks not traded by the shares over the whole 19 year period, and denoted $\bar{q}_j$ for share j .

²⁷ In fact it is only in 1976, 1990 and 1991 that the median values are higher than the overall median; and in these years with 1975, 1977 and 1978 that the upper quartiles are greater than the overall upper quartile.

Appendix 3.1

A model for the measures of trading frequency

In this appendix a logistic model for the *level* of trading frequencies of an individual securities is proposed and tested. The model stands in contrast to the model for *changes* in the levels discussed in section 3.6.3.

The model is formulated in terms of a logistic regression equation:

$$\log \frac{q_{ji}}{1 - q_{ji}} = \alpha_j + \beta_j p_i^w \quad (37)$$

where $p_i^w (= 1 - q_i^w)$ is the measure of overall market activity in interval i .

The model in equation (37) was fitted for only those shares which exhibited some degree of thin trading (defined to be those shares which had not traded on more than 1% of the weeks in the 19-year period). The restriction was included not only to avoid numerical problems in calculating the logits, but also because it has already been noted that the very well traded shares tend to stay very well traded. The exclusions meant that we used 186 shares in the consideration of the model.

The model was fitted and tested for goodness-of-fit by comparing the scaled deviance with the chi-squared critical value with 17 degrees of freedom at a nominal 5% significance level.

The test is only approximate given its asymptotic nature but it was found that for 88.7% (165 out of 186) of the shares the fit of the model was rejected at the 5% level. Furthermore the range of the q_{ji} of the shares for which the model was not rejected was from 0.01 to 0.11 whereas the overall range of q_{ji} for all the 186 shares was 0.01 to 0.91. The changes in levels of q_{ji} appear to be related to overall changes in market activity in only the better traded shares.

Nearly all (162) of the 186 β_j 's, the coefficient of market activity, were negative: marginal evidence that shares were better traded in years when market activity was higher.

Given the poor fit of the model, it was concluded that for nearly all the shares that the *levels* of thin trading are independent of the *levels* of overall market activity.

CHAPTER 4

A COMPARISON OF THE EFFICIENCY OF RISK ESTIMATORS IN A THINLY TRADED STOCK EXCHANGE: A SIMULATION STUDY

4.1 Introduction

The estimation of systematic risk on the well traded market of the NYSE has been thoroughly researched. As a result efficient beta estimators have been developed which can accommodate any small price adjustment frictions which still remain in the trading process. On the JSE and other small markets, however, thin trading is of a completely different scale. Price adjustment delays are no longer measured in hours or days, but rather weeks and months. Consequently, the suitability of the thin trading correction procedures on the JSE have to be assessed separately. Bearing this in mind, the purpose of this chapter is to compare the trade-to-trade and Cohen *et al* (1983) thin trading correction procedures in an environment which has the same thin trading characteristics as the JSE, but in which the biases and inefficiencies of the estimators can be clearly identified. The investigation therefore takes the form of a simulation study.

The comparison of systematic risk estimation techniques depends on several characteristics of the estimators. The most important of these characteristics are (a) unbiasedness, (b) efficiency and (c) the availability of the data needed for the implementation of the technique. In Chapter 3 the biases in estimators of beta coefficients caused by thin trading were discussed in predominantly analytical terms. This chapter extends the investigation into the estimation of systematic risk on the JSE by considering the aspect of efficiency of various estimators of systematic risk. Clearly, if the efficiency of an estimator is superior to another *and* there is no (or negligible) compromise in its unbiasedness, then the estimator will be the more suitable.

The assessment of efficiency in estimators of beta coefficients depends on introducing a level of realism into the investigation, while still maintaining control over many factors which would make such an investigation very difficult using actual data. The trade-off between control and realism is achieved in this chapter with a simulation study. The major variable over which

control is necessary is the 'true' beta of the security. Without knowledge of the true beta there can be no yardstick for the unbiasedness of the beta estimator while measuring its efficiency.

The simulation study in this chapter is based on a similar study by Bowie and Bradfield (1993a). The simulation methodology in this chapter is more sophisticated and incorporates more realistic assumptions about the return distributions in particular. Where appropriate, the conclusions from the earlier study are incorporated into the discussion.

The structure of the study is as follows. In the remainder of section 4.1 a brief recap is given of the thin trading correction procedures (the trade-to-trade estimator and the Cohen estimator) and some of the potential problems involved in a realistic simulation study are outlined. In section 4.2 the focus turns to some of the technical issues required to conduct the simulation study with as much realism as possible. Specifically, a preliminary empirical investigation into the non-normality of the return distributions of JSE stocks is described. An interesting feature of non-normality that is uncovered is an *interval effect* in the *extent* of non-normality of the return distributions. The interval effect manifests itself in that the longer the intervals (e.g. months rather than days) over which the return series is measured, the less divergent the distribution is from the normal distribution. Section 4.3 contains a description of the methodology and the results of the simulation study on the assessment of various estimators of systematic risk. Finally Section 4.4 contains concluding remarks and recommendations for beta estimation in a thinly traded environment.

4.1.1 The efficiencies of thin trading correction procedures

As discussed in Chapters 2 and 3, correction procedures for the bias caused by thin trading in beta estimators have developed along two lines.¹ The first approach is based on aggregating non-synchronous market model regression coefficients and is generally referred to as the 'aggregated coefficients' approach. Scholes and Williams (1977), Dimson (1979) and particularly Cohen *et al* (1983) have made notable contributions in this area. The second approach is known as the 'trade-to-trade' technique used, for example, by Marsh (1979). It is based on matching the returns of the market proxy to the times (days) when the security of interest is traded (and then using the simple least squares regression on the 'trade-to-trade'

¹ More recently, other estimators have been considered and tested by Luoma, Martikainen and Pertunen (1993a, 1993b, 1993c) using data from the Helsinki Stock Exchange. However, these estimators have not yet been implemented widely.

returns which are defined to be the returns measured between the last *traded* days in successive months rather than between the last *calendar* days in the months).

The aggregated coefficients approach uses the unadjusted monthly returns regardless of whether the security was traded on the last day of the month or not. The data are, however, combined in a more sophisticated estimation procedure involving leading and lagged coefficients. In contrast, the trade-to-trade approach first moulds the data into a form whereby the problems associated with thin trading are removed and then implements a simple estimation procedure.

As was mentioned in Chapter 3, the data requirements for implementing the trade-to-trade estimator are substantial. The actual times of the trades of the securities must be known as well as the values of a continuously updated and well traded market proxy. A case which reflects 'intermediate' data availability has been dealt with briefly in Chapter 3 and a brief reminder of this discussion is given in the next sub-section.

4.1.2 The trade-to-trade technique when only weekly data are available

As speculated in Chapter 3, data availability may be limited in smaller stock exchanges where data bases of the financial markets are not as advanced. For instance, although the trading volumes may not be obtainable on a daily basis they may well be obtainable on a weekly basis. To ascertain the potential effects of the restricted data availability a brief discussion is warranted here.

The extent of thin trading on the small stock markets is very severe and the delays between the last trade in the day and the end of the day (the *intra-day* effects) will generally be negligible when compared to delays between trades (which may last several weeks). In the simulation study the assumption has been made that all the trades made in securities on the last traded day of the month have occurred at the same time of the day. This assumption has meant that intra-day thin trading can be ignored.

Similarly, the use of weekly data as a basic data set from which to construct 'monthly' returns will result in biases from *intra-weekly* thin trading. To gauge the difference in efficiency resulting from limited data availability, the trade-to-trade estimator based on weekly data (in addition to daily data) has also been included in the simulation study.

Although a simulation study is by its nature artificial, the inferences to be drawn from it can be made more generally applicable if realism is maintained whenever possible. In the following sub-sections some of the issues involved in realistic simulation studies are discussed.

4.1.3 Simulating stock market data

In simulating the returns of hypothetical securities in order to compare the beta estimators there are two processes involved. Firstly, the distribution of trading times of the securities and, secondly, the distributions of the returns themselves. In Chapter 3 it was demonstrated that a relatively simple model based on a Bernoulli process fits the observed trading frequencies on the JSE adequately. However, the distributions of security returns remains the subject of much debate. Some researchers (e.g. Fama (1965) and Mandelbrot (1963, 1967)) have suggested the stable Paretian class of distributions as alternatives to the normal distribution while others (e.g. Blattberg and Gonedes (1974)) have proposed using the Student t -distribution.² In order to avoid linking the results of the simulation study to particular distributional assumptions for the residuals and the market proxy, a method of simulating non-normal deviates which does not require specifying a particular distribution function has been implemented.

In this study the proposal of Affleck-Graves and MacDonald (1989) is followed for simulating non-normal deviates by using a method devised by Fleishman (1978). In order to effect Fleishman's method, estimates are required of the first four moments of the distribution to be simulated (in addition to a conventional random normal variate generator). The method is easy to implement and produces deviates from an unknown, but non-normal, distribution with specified population mean, variance, skewness and kurtosis.

Bearing the requirements of Fleishman's method in mind, one aspect of non-normality that does have to be considered in some depth is the 'interval' effect in the distribution of stock market residuals. The existence of the interval effect is demonstrated in section 4.2 and can be observed in that the extent of non-normality in the return series depends on the length of the interval over which the returns are measured.

To understand why the interval effect in non-normality is an important issue in the simulation study, consider the trade-to-trade estimator. The variability in the interval lengths

² Further proposals for the distributions of returns and a far more in-depth discussion of non-normality can be found in Chapter 6.

is an integral part of the trade-to-trade estimator and necessitates the simulation of residuals and index returns over intervals of various lengths. Because of the interval effect in non-normality, the distribution of the returns over the various lengths will need to change depending on the length of the interval. Models are therefore required to predict the extent of non-normality in the return series for any given interval length. In the next section a series of models are devised for generating non-normal returns measured over various interval lengths.

4.2 Non-normality in stock market returns

In this section the extent of non-normality in stock market returns on the JSE is investigated empirically. The purpose of the investigation is to estimate the mean, variance, skewness and kurtosis of the market and residual returns so that a 'realistic' simulation study can be performed. Furthermore, the problem of catering for the interval effect (of the non-normality of returns) is addressed in the simulation study. In Appendix 4.G, a brief comparison of the levels of non-normality on the JSE and NYSE is presented. It appears from that comparison that the levels of non-normality on the JSE are substantially higher, and that therefore the incorporation of non-normality into assessments of beta estimators is an important consideration.

4.2.1 Empirical investigation into non-normality

An empirical study aimed at estimating the parameters needed for the simulation study was conducted using data from the JSE. As discussed in section 4.1.3, the parameters needed are the first four moments of the distributions of the market and residual returns.

In order to obtain suitable parameters, non-overlapping daily, two-daily, three-daily, etc, up to 20-daily returns were calculated for the JSE-Actuaries Overall Index (which is customarily used as a market proxy on the JSE). The first four sample moments were then calculated on each set of returns.

In the case of the parameters for the residual returns, residuals were computed from regressions of the returns (measured over the same set of interval lengths as for the market returns above) of 40 well traded³ shares against the market index. (The securities used to

³ 'Well traded' shares were defined to be those shares which had traded on more than 95% of the days from October 1987 until May 1992. See Appendix 4.A for the names of the securities.

obtain the estimates of the residuals were chosen to be well traded in order to avoid the biases introduced by thin trading.)

For each interval (daily, two-daily, etc) the first four sample moments of the index returns and the residuals, excluding the largest and smallest returns (residuals) in each

Table 4.1

First four moments of the returns measured over intervals of one to twenty days on the JSE-Actuaries Overall Index and of the residuals of 40 well-traded shares.

Days	Obs ^b	Overall Index				Residuals ^a		
		Mean (%)	sd (%)	b_j^c	b_k^d	sd (%)	b_j	b_k
1	1098	.070	.99	-.41	3.21	2.53	-.20	6.61
2	549	.139	1.54	-.22	1.84	3.66	.05	3.56
3	366	.204	1.99	-.26	.66	4.55	.02	2.65
4	274	.284	2.26	-.32	.69	5.14	.04	2.41
5	219	.335	2.58	-.20	.36	5.82	.16	1.44
6	183	.437	2.95	-.18	.58	6.53	-.02	1.94
7	156	.461	2.99	-.09	-.04	6.71	.22	1.12
8	137	.536	3.24	-.13	-.75	7.18	.20	.66
9	122	.586	3.54	.04	-.33	7.86	.30	.62
10	109	.694	3.53	-.15	.26	8.18	.18	.96
11	99	.769	3.92	-.20	-.52	8.71	.14	.59
12	91	.833	3.94	-.33	.06	8.86	.04	.89
13	84	.875	4.07	-.28	-.61	9.31	.14	.55
14	78	.931	4.31	-.15	-.48	9.39	.18	.40
15	73	.981	4.24	-.17	-.50	9.52	.20	.53
16	68	1.081	4.60	-.17	-.54	10.28	.17	.33
17	64	1.129	4.46	-.30	-.75	10.22	.13	.39
18	61	1.233	5.30	-.38	-.31	11.03	.04	.17
19	57	1.319	5.24	-.44	-.47	10.89	.02	.20
20	54	1.372	4.13	-.22	-.69	9.03	.03	.29

^a The averaged moments of the 40 sets of residuals. (Note that by construction the mean residuals are zero.)

^b The number of non-overlapping intervals in the period of investigation.

^c The coefficient of skewness: normal distribution has zero coefficient with approximate standard error $(6/\text{Obs})^{1/2}$.

^d The coefficient of kurtosis: normal distribution has zero kurtosis with approximate standard error $(24/\text{Obs})^{1/2}$.

regression, were calculated.⁴ The sample moments for the index and the averaged moments for the residuals are presented in Table 4.1.⁵ Judging from the large coefficients of kurtosis it is clear from Table 4.1 that returns on the index and residuals are highly non-normal, especially when measured over shorter interval lengths. When each of the columns in Table 4.1 is inspected, a structure in the moments can be identified. It appears as if the mean and standard deviation of both the residuals and the index returns increase as the length of the interval increases. For example, the mean index return for one day is 0.070 compared with a mean of 1.372 for 20 days, where it should be noted that 20 times 0.070 is approximately equal to 1.372. Similarly, the variance for one day is 0.99^2 compared with value nearly 20 times that (i.e. 4.13^2) for the variance of 20 days. On the other hand, the kurtosis decreases as the interval length increases, i.e. for the index the decrease is from 3.21 at one day to -0.69 at 20 days and for the residuals, the decrease is from 6.61 to 0.29 as the interval length is increased from one to 20 days. The coefficients of skewness do not appear to exhibit any obvious correlation with the interval length, but are generally slightly negative for the index returns and are, on average, generally positive for the residuals.

The relationship between the coefficients of kurtosis and the lengths of the intervals gives rise to the concept of an intervalling effect in the extent of non-normality. For the sake of realism it is important to incorporate the intervalling effect into the simulation. In the next sub-section, therefore, models are sought to predict the kurtosis, skewness, variance and mean of the index returns and residuals over any given interval lengths.⁶

⁴ The *ad hoc* trimming of the samples was conducted to omit the most extreme outliers in the samples since the estimation of higher order moments is highly sensitive to outliers. Data relating to 1987 were omitted to avoid the use of possibly uncharacteristic data which might have resulted from the October crash. The data therefore comprised 1098 daily returns, 549 two-daily returns, etc.

⁵ The moments are conditional on the first day of data recorded. For example, there are two sets of non-overlapping two-daily returns which can be computed from the data set. To check on the robustness of the estimates of the moments they were recomputed starting on the second trading day in 1988 instead of the first day. The means, standard deviations and coefficients of skewness were very similar, but there was some variation in the coefficients of kurtosis. For example the coefficient of kurtosis for the two-daily returns on the index was 1.01 when the set was measured from the second day (compared with a coefficient of 1.83 for the original set of two-daily returns). In general, the measures of kurtosis for the original set were higher than those for the second set of returns and residuals. The discrepancy further highlights the extreme sensitivity of higher sample moments to outliers. Attention was focused on the initial set of estimates because they suggested more extreme departures from normality and are thus a possibly conservative (in terms of making statements about the effects of departures from normality) set of estimates. For a detailed study of the distribution of returns on the JSE a much longer daily data set is needed and more robust measures required - see Chapter 6.

⁶ It is further apparent that estimates of the third and fourth moments are very sensitive to the sample selected. For example, when using the weekly data very different levels of kurtosis were estimated for the non-overlapping four-weekly returns compared with those estimated from the non-overlapping 20-daily returns (which happened to start on a different day of the week). However, the intervalling effect was apparent in both data sets and the conclusions

4.2.2 Modelling the interval effect in the non-normality of returns

Because there is no clear relationship between skewness and interval length it seems appropriate to calculate an 'average' coefficient of skewness which can be used for all lengths of interval. For the purposes of the simulation study the 'average' skewness was computed as the weighted arithmetic mean of the coefficients of skewness across all interval lengths.⁷ For the residuals the mean coefficient of skewness is 0.0702 and for the index returns the average is -0.239.

It is intuitively plausible that the variances and means of the return series increase proportionally to the length of the interval over which they are measured. For example, it would appear reasonable that the average weekly return should be approximately five times the average daily return, and that the variance should be similarly increased five-fold in going from daily to weekly returns. For the means and variances of the returns the intuitively appealing models are, therefore:

$$\bar{r}_p = p \times \bar{r} \quad (1)$$

$$\bar{\sigma}_p^2 = p \times \bar{\sigma}^2 \quad (2)$$

where $\bar{r}_p$ and $\bar{\sigma}_p^2$ are the average mean and variance of the returns over interval length p , and $\bar{r}$ and $\bar{\sigma}$ are constants interpretable as the average and variance of the daily returns respectively.

In the case of the coefficient of kurtosis, no obvious or intuitive model is apparent. Experimental plots⁸ of the sample moments against the interval lengths suggested a model of the following form might be suitable for the coefficients of kurtosis:

$$b_4^{(p)} = \alpha + \beta \frac{1}{\sqrt{p}} \quad (3)$$

where p is the interval length.

reached about the performance of the estimators were the same. See Appendix 4.C for details of the intervallling effect when using weekly data. Chapter 6 contains a more detailed study of non-normality.

⁷ The weights used were the reciprocals of the standard errors of each of the 20 estimates.

⁸ The models are designed to be used for generating reasonably realistic levels of non-normality at various interval lengths. The fact that the model may not be intuitively appealing (or even the 'best' model) is therefore to some extent immaterial.

The estimated⁹ parameters of the models for the average returns, variances and coefficients of kurtosis are given in Table 4.2. In the first panel of Table 4.2, it can be seen that the average daily return for the index is estimated from the model to be about 0.07%. The linear relationship between the return and the length of interval is nearly exact, with an R^2 of 99%. Note that the average residual will always be zero because of the use of simple OLS regression to estimate the residuals. In the second panel of Table 4.2, the model parameters for the standard deviations are given. The average daily residual standard error appears to be

Table 4.2

Estimates of the parameters in the models specified for the mean return, standard deviation of returns and the coefficient of kurtosis of the returns on daily data.

	Returns ^a		Std. Dev. ^b		Kurtosis ^c		
	Ave. daily return ($\bar{r} \times 100\%$)	R^2	Std. dev. ($\bar{\sigma} \times 100\%$)	R^2	alpha	beta	R^2
Index	.06795	.99	1.134	.89	-1.87	5.06	.92
Residuals	.00000	1	2.508	.90	-1.95	8.40	.97

^a $r_p = p \times r$

^b $\sigma_p^2 = p \times \sigma^2$

^c $b\phi_i^j = \alpha + \beta\sqrt{p}$

nearly twice that of the index (2.508% compared with 1.134% for the index). Again, the fit of the model is good, with R^2 statistics of around 90%. In the third panel of Table 4.2, the parameters in the relationship between the coefficient of kurtosis and the length of the interval are given. For both the index returns and the residuals the fits are generally excellent as revealed by very high R-squareds (92% and 97%, respectively).

In summary, the results of this section reveal considerable non-normality in the returns on the index and on residuals from the market model, particularly when measured over intervals of less than one month. Although the skewness of the index returns and residuals appear unrelated with interval length, the mean and variance may be described as ‘additive’ in the sense that their relative magnitudes are proportional to the lengths of the intervals over

⁹ The constants were estimated using weighted least squares (with weights equal to the square root of the number of observations).

intervals of less than one month. Although the skewness of the index returns and residuals appear unrelated with interval length, the mean and variance may be described as 'additive' in the sense that their relative magnitudes are proportional to the lengths of the intervals over which the returns are measured.¹⁰ Simple models relating the values of the first four moments of the index and residual distributions to the length of the interval over which they were measured were proposed and fitted. The models appeared very suitable for use in the ensuing simulation study.

4.3 Simulation study to compare estimators

In the previous section the non-normality of index returns and residuals was investigated and models derived to enable the simulation of realistic security returns. In this section the methodology and results of the simulation study are discussed. Since it is known that the thin trading correction procedures are theoretically unbiased, the focus of the investigation is on identifying the most efficient estimator.

In sub-section 4.3.1, the method used for simulating non-normal security returns is briefly described. Sub-section 4.3.2 contains a brief overview of the simulation algorithm and a description of the statistics which are to be used to assess the estimators. In 4.3.3 the results of the preliminary simulation are presented and discussed. The purpose of this preliminary simulation is to pick out those aspects of the estimators which require a more detailed investigation. One of the aspects which arises is that of the number of leads and lags of the Cohen estimator. An adaptive procedure for selecting the number of lags and leads is proposed in sub-section 4.3.4. The simulation results once the adaptive lag selection procedure has been implemented are presented in 4.3.5.

4.3.1 Methodology for simulating non-normal deviates

As discussed earlier, the method of generating non-normal deviates proposed by Fleishman (1978) and implemented by Affleck-Graves and MacDonald (1989) in their simulation study was used in this simulation study as well. In general terms the method

¹⁰ The models and parameters calculated in this section have been based on daily data. Similar results were obtained using weekly data to test for any intervallling effects over longer periods of time, and over longer interval lengths. The results are summarized in Appendix 5.C.

generates variates (denoted X) which have a specified mean, variance, skewness and kurtosis using the following form:

$$X = a + bY + cY^2 + dY^3 \quad (4)$$

where a , b , c and d are constants selected so that X has the appropriate first four moments and Y is a randomly generated standard normal deviate. Fleishman (1978) describes in detail how to select the constants and discusses the advantages and limitations of the method. The method is ideal for simulating stock market returns in that no particular alternative distribution to the normal is imposed. Implementation of the method is relatively easy and requires no more than estimates of the first four moments of the variables which are to be simulated. Once the sample moments have been set, the four constants needed for the simulation of non-normal variates described in (4) can be calculated.

Having dealt with a particular detail of the simulation methodology, the next subsection deals with the simulation methodology in more general terms. Furthermore, the measures of efficiency used to compare the beta estimators are discussed.

4.3.2 Methodology and measures of efficiency

The most important concept behind the use of a simulation study to compare estimators is that the actual values being estimated are known. Thus there exists a yardstick by which the estimators can be compared, unlike in empirical studies.¹¹ In this simulation study, security returns were simulated using the market model with a beta of unity and an alpha of zero. The choice of a unit beta and zero alpha facilitates comparisons since any estimate of beta can easily be interpreted as the proportion that the estimate is of the true value.

The simulation of the stock returns involved three random variables being generated: firstly, the return on the market, secondly, the residual value (in equation (2)) and, thirdly, an indicator for whether the hypothetical share traded on the day or not. Security returns were simulated over an hypothetical 60-month period for a large number of 'shares'. For each 'share' a new market index was simulated so that the full impact of non-normality in the

¹¹ The simulation methodology adopted in this chapter is therefore different from the 'artificial' data used by Luoma, Martikainen and Perttunen (1993b) which was created by randomly selecting a certain number of days from a well traded security and assigning them 'no trade' status. The reason for the different approach is to keep control of the value of the true, underlying beta.

distribution of the market returns and residuals could be included in the results. More inclusive details of the simulation algorithm can be found in Appendix 4.B.

The estimators considered were OLS, trade-to-trade and Cohen¹² estimators. Although the biasing effect of thin trading on OLS is well-known, it remains one of the most commonly used estimators in regression analyses and for this reason it is included both to emphasize the extent to which it is affected by thin trading and to provide a contrast for the correction techniques.

A fourth estimator computed is what has been termed a weekly trade-to-trade beta. For the weekly trade-to-trade estimator it is assumed that the data available are on a weekly basis rather than a daily basis. The weekly trade-to-trade estimator represents a compromise between the unbiasedness of the trade-to-trade technique and smaller data requirements of the OLS technique. Earlier in this thesis, in Chapter 3, the weekly trade-to-trade estimator was analyzed in theoretical terms and appeared to be a reasonable estimator when daily data were not available.

In interpreting the results of the simulation study, attention is focused on five summary statistics of the beta estimates from the simulation study: (1) the mean values of the beta estimates; (2) the average estimated standard errors of the betas provided by the regression analysis; (3) the 'cross-sectional'¹³ standard deviations of the estimates across the simulated shares; (4) the ranges of the beta estimates produced by the different methods and, finally, (5) the mean squared errors (MSE) of the estimates.

The first statistic examined, the mean estimated betas, gives an indication of the unbiasedness (or otherwise) of the estimator in the face of thin trading. In order to examine the efficiency of the estimators, the spread of the estimates around the mean beta must be measured. The 'cross-sectional' standard deviations of the estimates of beta therefore give an indication of the efficiencies of each the methods. (These 'cross-sectional' *standard deviations* of the estimated betas are compared with the *standard errors* of the beta estimates which are calculated routinely in any regression analysis. The comparison is made in order to determine whether the *standard errors* give a reasonable indication of the efficiency of the estimate despite

¹² In the first simulation only one lagged coefficient for the Cohen-type estimator was used. On page 5-11 an adaptive method for choosing the number of lags for the Cohen estimator is described.

¹³ The simulation cannot strictly be described as providing a cross-section of estimates. Rather than simulating one set of market returns and then simulating the returns of many shares based on the one set of market returns (which could then be described as yielding cross-sectional data), we simulate for each share a different set of market returns. We effectively have the estimates of one share based on many different time periods.

the thin trading and non normality.) The next pair of statistics examined are the minimum and maximum estimates of beta in the sample of simulated shares. The extreme values provide an indication of 'worst case' estimates of beta which each of the estimators may provide. Finally, the MSE is a joint measure of bias and inefficiency. The MSE is a measure of the combined effects of bias (defined to be the difference between the average estimated value and the true beta) and inefficiency, and can be calculated using the relationship (see Appendix 4.D):

$$\text{MSE} = \text{bias}^2 + \text{variance}$$

The smaller the value of the MSE for a particular estimator relative to the MSEs of other estimators the better that estimator is in comparison with the others.

In the next sub-section the statistics emerging from the first run of the simulation are discussed and some areas needing further investigation are highlighted.

4.3.3 Results of initial efficiency comparison

The evaluation of the beta estimators begins with affirming the consistency of the two correction procedures and establishing the extent of the biases in the OLS and weekly trade-to-trade estimator. Table 4.3 gives the average betas from the simulations. The expected bias in the OLS betas is clearly discernible in the first column of Table 4.3; the average OLS estimates decline from 1.00 when $q = 0$ (i.e. for the well traded shares) to 0.42 when $q = 0.95$ (i.e. the very thinly traded shares). The column headed 'TT (daily)' in Table 4.3 gives the average estimated betas using the trade-to-trade method. As can be seen, the trade-to-trade estimator is clearly unbiased with all the average betas being very close to the true value of unity. The weekly trade-to-trade estimator ('TT (weekly)') displays a small bias in the estimates as the extent of thin trading increases. For example, when $q = 0$, the average estimated value is 1.00 in contrast to an average value of 0.91 for the most thin traded securities ($q = 0.95$). One particular point of interest that is apparent from Table 4.3, is that the Cohen estimator clearly requires further lagged coefficients at the extreme levels of thin trading (where q , the probability of no trade on a day, is larger than about 0.8) in order to make it unbiased. For instance, the average betas are close to unity for values of q up to about 0.85, but then decline to 0.78 when q is 0.95.

Table 4.3

Simulation results - average estimated betas.

q^*	OLS	TT (daily)	TT (weekly)	Cohen
0.00	1.00	1.00	1.00	1.00
0.05	1.00	1.00	1.00	1.01
0.10	1.00	1.01	.99	1.01
0.15	1.00	1.01	.99	1.00
0.20	.99	1.01	.99	1.01
0.25	.98	1.00	.97	1.00
0.30	.99	1.00	.98	1.00
0.35	.97	1.00	.97	.99
0.40	.97	1.01	.97	1.01
0.45	.97	1.01	.97	1.01
0.50	.94	.99	.96	.99
0.55	.95	1.01	.95	1.02
0.60	.93	1.00	.94	.99
0.65	.92	1.01	.95	1.01
0.70	.88	.99	.93	.99
0.75	.85	1.00	.93	1.00
0.80	.82	1.00	.94	1.00
0.85	.73	.99	.92	.99
0.90	.62	1.00	.92	.95
0.95	.42	1.00	.91	.78
Average	.90	1.00	.96	.99

* q denotes the probability that a share does not trade on a day.

Before discussing the efficiency of the estimators, the 'worst case' estimates of the betas are considered. Table 4.4 shows the minimum and maximum estimated betas for a reduced¹⁴ set of q . The 'worst case' estimates are substantially different from the expected values of unity. For example, even with well traded shares, the possible trade-to-trade estimated betas can be seen in Table 4.4 to range from 0.15 to 1.84.

One aspect of thin trading which has received relatively little attention in the literature is its impact on the efficiency of estimates, as well as its effect on *estimates of the standard error*. In the simulation results the average estimated *standard errors* of the beta estimates were calculated along with the *standard deviations* of the estimates themselves. The *standard*

¹⁴ The reduced set of q is used to improve clarity, and because there were no additional insights to be gained from examining the minima and maxima from the other values of q .

Table 4.4

Range of estimated betas.

q^b	OLS	TT (daily)	TT (weekly) ^a	Cohen (1 lag)
0	.15 to 1.84	.15 to 1.84	.15 to 1.84	-.33 to 2.15
15	.07 to 1.98	.07 to 2.04	.07 to 1.78	-.37 to 2.40
30	.10 to 1.87	.09 to 1.94	.09 to 1.90	-.51 to 2.26
45	.11 to 1.90	.11 to 1.81	.24 to 1.67	-.14 to 2.35
60	.01 to 1.72	.19 to 1.77	.19 to 1.82	-.21 to 2.34
75	-.19 to 1.85	.14 to 2.05	.26 to 1.73	-.31 to 2.17
90	-.36 to 1.55	-.13 to 1.86	-.06 to 1.75	-.20 to 2.21

^a The maxima and minima for the weekly trade-to-trade betas are based on a simulation of 1000 shares. All the others are based on 2000 shares.

^b Probability that a share does not trade on a day, as a percentage.

Table 4.5

Standard deviations and estimated standard errors of estimated betas from simulation.

q	OLS		TT (daily)		TT (weekly)		Cohen (1 lag)	
	s.d.	s.e.	s.d.	s.e.	s.d.	s.e.	s.d.	s.e.
0	.261	.262	.261	.262	.261	.262	.372	.393
30	.260	.260	.258	.258	.263	.257	.375	.390
60	.265	.259	.256	.254	.258	.252	.375	.386
70	.263	.261	.252	.253	.258	.250	.362	.386
75	.260	.263	.253	.253	.244	.248	.370	.387
80	.267	.263	.249	.251	.242	.248	.375	.387
85	.261	.267	.251	.254	.246	.251	.365	.388
90	.278	.273	.268	.268	.261	.262	.372	.393
95	.296	.284	.316	.319	.299	.313	.418	.402

The standard deviations (s.d.) represent the actual efficiency of the estimator. The standard errors (s.e.) are the estimates of efficiency emerging from the regression analysis associated with the estimator.

deviation is an indicator of the true efficiency of the estimator as it measures the spread of the estimates around the mean estimate. The *standard errors*, on the other hand, are usually standard output from a regression analysis and are in fact themselves *estimators* of the efficiency. Table 4.5 gives the standard errors and the standard deviations from the simulation study over a range of thin trading levels.

From Table 4.5 it is evident that the standard deviations of the estimated betas do increase with thin trading. Considering the column of Table 4.5 headed ‘TT (daily)’, the trade-

to-trade betas can be seen to experience a 20% increase in standard deviation from 0.261 when there is no thin trading to 0.316 for shares with the most severe thin trading ($q = 0.95$). The decrease in efficiency is due to the number of observations dropping as the thin trading increases and, with the reduced trading frequency, the increased likelihood of the hypothetical security not trading for the entire month. For the trade-to-trade betas, the standard errors (in the column headed 's.e.') give reasonable estimates of efficiency as can be gauged by the closeness of the standard errors and the standard deviations.

It is further apparent from Table 4.5 that the relative increase of the variability for the OLS and Cohen estimators is smaller than that for the trade-to-trade estimator. The standard deviations of the OLS estimator rise from 0.261 to 0.296 and those of the Cohen estimator from 0.372 to 0.418. The increases occur because of the increased inefficiency involved in estimating thinly traded securities.¹⁵

By comparing the standard deviations of the trade-to-trade and OLS betas in Table 4.5 it becomes clear that the trade-to-trade estimator is more efficient than the OLS betas over most of the range of thin trading. The Cohen estimates show considerably more inefficiency, having about a 50% larger standard deviation than the OLS and trade-to-trade betas.

Further evidence from Table 4.5 is that the standard errors for the OLS betas approximate the standard deviations quite closely. The standard errors of the Cohen estimates were computed as the sum of the individual standard errors of the simple regressions used in calculating the Cohen beta.¹⁶ Judging from the last two columns of Table 4.5, the aggregated standard errors estimator slightly over-estimates the true standard deviation, except at the most thinly traded levels where the standard error under-estimates the true variance slightly.

As mentioned earlier, the MSE represents a combined measure of bias and inefficiency. A justification for its use in comparing estimators is that a slightly biased, but very much more efficient estimator may be preferred to one which is strictly unbiased. As indicated in subsection 4.3.2, the MSE can be calculated as the sum of the squared bias component and the variance of the estimator. Table 4.6 gives the values of the MSEs for the four estimators.

The results in Table 4.6 confirm that the trade-to-trade estimator (based on daily data) is the estimator of choice, both in terms of unbiasedness and efficiency. From Table 4.6 it is

¹⁵ Note that the Cohen estimator with one lagged coefficient remains biased at the very high levels of thin trading.

¹⁶ The standard error was in fact calculated as the square root of the summed squared standard errors from the individual regressions. In Chapter 5 a more detailed examination of the estimation of the Cohen estimator's efficiency is presented.

Table 4.6

Mean squared estimation errors (MSEs).

q	OLS	TT (daily)	TT (weekly)	Cohen (1 lag)
	MSE	MSE	MSE	MSE
0	.068	.068	.068	.138
30	.068	.067	.070	.141
60	.075	.066	.070	.141
70	.084	.064	.071	.131
75	.090	.064	.064	.137
80	.104	.062	.062	.141
85	.141	.063	.067	.133
90	.222	.072	.075	.141
95	.424	.100	.098	.223

clear that the MSEs of the trade-to-trade estimator are lower than the MSEs of all the other estimators (except in the case of the very well traded shares where there is no difference between the OLS and trade-to-trade estimators anyway). For example, at $q = 0$ (i.e. for well traded shares) the MSEs of the trade-to-trade estimator and the OLS estimator are both 0.068, but the gap widens to 0.100 for the trade-to-trade at $q = 0.95$ and 0.424 for OLS. Another insight to be gained from Table 4.6 is that the weekly trade-to-trade estimator has MSEs which are slightly higher than the trade-to-trade estimator based on daily data, corresponding to the slight bias in the weekly trade-to-trade estimator.

A perhaps more interesting result emerges from Table 4.6 for practitioners who may have only monthly data available (and so are therefore limited to either the OLS or the Cohen estimator). It appears that the OLS estimator outperforms the Cohen estimator in terms of MSE for all values of q up to 0.85, when the OLS MSE is 0.141 and the MSE of the Cohen estimator is 0.133. Remembering the results from Chapter 3 where the securities were split up into deciles on the basis of their trading frequencies, it is apparent that the OLS estimator will be superior to the Cohen estimator except for the approximately 10 percent of securities which are the most infrequently traded.

The results from this section indicate that the Cohen estimator is not a suitable estimator for individual securities on the JSE. However, in an earlier study (reported in Bowie and Bradfield, 1993a), the Cohen estimator was found to perform relatively more efficiently when the residual variance was small. Therefore, for example, the estimation of systematic risk

for large portfolios which have a small residual variance may be performed efficiently using the Cohen estimator. The results from the study by Bowie and Bradfield (1993) are presented in Appendix 4.E. A cautionary note about interpreting some of the results because of the design of the simulation study in Bowie and Bradfield (1993a) is given in Appendix 4.F.

In the initial simulation and comparison of estimators described in this section, the Cohen estimator was implemented with a single lagged coefficient. In the next sub-section an adaptive selection procedure is constructed for determining a more appropriate number of lags to be used for the Cohen estimator. The adaptive procedure is based on the thin trading of the simulated share.

4.3.4 Adaptive selection of leads and lags for Cohen estimator

From the previous sub-section it is clear that the performance of the Cohen estimator is dependent on the choice of the number of lagged coefficients used. Sufficient lags are needed to maintain unbiasedness, and yet the more lags included the more inefficient the estimator becomes. A possible solution to the problem involves the use of an adaptive procedure for choosing the number of lags. In particular, a security's q (the probability that the security does not trade in a day) can be used to estimate the number of months (using the model from Chapter 3) by which changes in the value of a share might be expected to lag changes in the market.

A standard result following from the assumptions regarding the trading process is that the probability that the month-end price age is greater than a value k^* is given by q^{k^*} . This expression can be used to calculate the expected number of days less than which 95%, say, of price ages are. In other words a 95% confidence interval is formed for the price ages and then the number of lags are chosen so that they cover that interval. The procedure outlined above involves solving the inequality:

$$q^{k^*} \leq 0.05 \quad (5)$$

where k^* is measured in days.

It follows that the adaptively selected number of lags (measured in 20-day intervals) should be set at:

$$\text{lag} = \left[\frac{\log(0.05)}{\log(q)} \div 20 \right] + 1 \quad (6)$$

where the square brackets denote taking the integer part of the expression enclosed.

4.3.5 Results using adaptive procedure

In order to assess the suitability of the adaptive procedure, the simulation was re-run with q varying from 0.75 up to 0.975 in increments of 0.025. The new simulation allowed a detailed investigation of the more thinly traded sector of the market. The results from Chapter 3, which dealt with the distribution of thin trading across securities in the market, indicated that about 25% of securities can be found with levels of thin trading corresponding to the levels considered in this section. It is therefore important to gain a more detailed perspective on the very thinly traded securities.

In Table 4.7 the average betas for each type of estimation and the adaptive Cohen estimator are presented. The final column in Table 4.7 in comparison with Table 4.3 indicates that the adaptive method clearly improves the unbiasedness of the Cohen betas, with all the

Table 4.7

Simulation results - average betas using adaptive Cohen estimator.

q	OLS	TT (daily)	TT (weekly)	lags	Cohen
.750	.87	1.00	.93	1	1.00
.775	.84	1.00	.93	1	.99
.800	.82	1.00	.92	1	1.00
.825	.80	1.00	.92	1	1.00
.850	.76	1.00	.91	1	.99
.875	.72	1.01	.93	2	1.02
.900	.67	1.02	.93	2	1.03
.925	.58	1.01	.92	2	.99
.950	.48	1.03	.95	3	1.01
.975	.28	1.00	.94	6	.99

mean estimates being close to unity. For example, at $q = 0.95$, the adaptive procedure in Table 4.7 produces an average Cohen beta of 1.01 compared with an average of 0.78 for the Cohen estimator with a single lag reported in Table 4.3. The second last column indicates the number of lagged coefficients which the adaptive procedure chose. For the most thinly traded securities it appears from Table 4.7 that up to six lagged coefficients are necessary.

Another point of interest which emerges from the more detailed simulation of the upper levels of thin trading in Table 4.7 is that the weekly trade-to-trade betas actually *improve*, on average, at the extreme levels of thin trading. For example, the average weekly trade-to-trade beta when $q = 0.975$ is 0.94 compared with a mean estimate of 0.91 when $q = 0.85$. The improvement is due to the fact that at the extremely high levels of thin trading the average number of months between traded months increases. Hence the proportion that the week-end price age is of the total time between trades becomes increasingly negligible.¹⁷

The maximum and minimum estimated betas are not tabled here, but it was found that as the number of lags increased so the range of estimates around the average beta increased. For example, at q equal to 0.75 the Cohen betas (with one lag) range spanned from -0.07 to 2.01, but at q equal to 0.975 (and the number of lags equal to six) the range spanned -1.86 to 5.25.

The next series of statistics to be examined are the inefficiency statistics. Table 4.8 gives the standard deviations and standard errors for the simulation which included the adaptive lag selection procedure. The results are largely similar to the previous run: (a) the standard deviations increase as thin trading increases; (b) OLS standard deviations are marginally smaller than trade-to-trade standard deviations because of the loss of observations associated with the trade-to-trade estimator at very high levels of thin trading (e.g. at $q = 0.75$, the standard deviations for the trade-to-trade and OLS estimators are similar at 0.235 and 0.248 respectively, compared with 0.312 for OLS and 0.423 for the trade-to-trade estimator at the higher level of thin trading when $q = 0.975$) and (c) Cohen standard deviations are very much larger than OLS and trade-to-trade standard deviations (e.g. the Cohen standard deviations are all above 0.3 whereas the standard deviation of the trade-to-trade estimator only becomes greater than 0.3 when $q = 0.95$).

As in the first simulation, the standard errors perform relatively well as estimators of the standard deviations, except at q equal to 0.975 where for all the estimators the average

¹⁷ This effect was predicted in Chapter 3.

Table 4.8

Standard deviations and standard errors (using adaptive Cohen estimator).

q	OLS		TT (daily)		TT (weekly)		Cohen (adaptive)	
	sd	se	sd	se	sd	se	sd	se
.750	.248	.254	.235	.244	.242	.250	.343	.377
.775	.252	.255	.242	.243	.249	.249	.358	.377
.800	.254	.255	.236	.242	.238	.250	.347	.376
.825	.250	.256	.241	.241	.250	.249	.361	.376
.850	.269	.259	.254	.244	.265	.252	.364	.379
.875	.263	.259	.247	.244	.255	.254	.456	.466
.900	.270	.264	.255	.251	.260	.262	.473	.471
.925	.274	.270	.270	.268	.283	.279	.478	.479
.950	.284	.276	.301	.304	.326	.312	.555	.560
.975	.312	.290	.423	.406	.426	.417	.865	.772

standard error is smaller than the standard deviation. The difference is relatively greater for the Cohen estimator.

The cost of making the Cohen estimator unbiased is that its levels of inefficiency are greatly increased. To assess the combined effect of the unbiasedness and increased inefficiency, Table 4.9 compares the MSEs of all the other estimators with that of the adaptive Cohen estimator. At $q = 0.75$ (when only one lagged coefficient is used for the Cohen estimator) the MSE is 0.118 compared with an MSE of 0.748 when $q = 0.975$ and the number of lags is six. It is clear from Table 4.9 that the increased number of lagged coefficients which is required to make the Cohen estimator unbiased also make it substantially more inefficient. In fact it can be seen that, except for levels of q corresponding to 0.925 and 0.950, the OLS estimator outperforms the Cohen estimator at most levels of thin trading in terms of MSE. Thus making the Cohen estimator unbiased results in it losing some of the advantages that it did have over the OLS estimator when only one lag was used.

However, the most important point to be gained from the simulation study is that the magnitude of the OLS and Cohen MSEs relative to those of the trade-to-trade estimators emphasizes the substantial advantage that the trade-to-trade approach offers. From Table 4.9 it is clear that the MSEs for the trade-to-trade estimator are approximately half of the OLS MSEs for values of q greater than about 0.80 and can be up to a third of the MSEs for the Cohen estimator. The advantage of the trade-to-trade estimator over the other estimator

Table 4.9

MSEs of the estimators (using adaptive Cohen estimator).

q	OLS	TT (daily)	TT (weekly)	Cohen (adaptive)
	MSE	MSE	MSE	MSE
.750	.078	.055	.063	.118
.775	.089	.059	.067	.128
.800	.097	.056	.063	.120
.825	.103	.058	.069	.130
.850	.130	.065	.078	.133
.875	.148	.061	.070	.208
.900	.182	.065	.073	.225
.925	.251	.073	.086	.229
.950	.351	.091	.109	.308
.975	.616	.179	.185	.748

remains, even when weekly data have to be used in conjunction with the trade-to-trade approach.

4.4 Conclusion and further comments

In this investigation the distribution of returns on the market and residual security returns have been tested for evidence of non-normality as measured by their skewness and kurtosis. In line with most markets around the world the return and residual distributions of securities on the JSE are found to be considerably non-normal predominantly because of heavier tails than the normal. An additional insight into the distributions of security returns was the existence of an intervalling effect. The longer the interval over which returns were measured, the less extreme the non-normality of the returns were. In order to include the intervalling effect into the simulation study, models were proposed in order to calculate the value of any one of the first four moments of the market returns and residuals over a specified time interval.

The models enabled a simulation study to be run which had as its objective the comparison of the efficiencies of betas estimated using various techniques designed to correct for the impact of thin trading. The results reinforce the opinion that the trade-to-trade class of

estimators is superior to the other methods tested in correcting for thin trading in terms of both unbiasedness and efficiency.

While the Cohen estimator can be adapted to be unbiased, the cost of the unbiasedness (as measured by the decrease in efficiency) makes it even less attractive than the (very biased) OLS estimator, even at high levels of thin trading. The trade-to-trade estimator's superiority over the OLS estimator arose from the lower standard errors of the trade-to-trade estimator at the low and medium levels of thin trading, and from its unbiasedness at the higher levels of thin trading (when the standard deviation of the OLS estimates around their very biased mean was lower than the standard deviation of the trade-to-trade estimates around their unbiased mean estimate). The MSE criterion enabled the joint effect of the bias and standard error to be combined in a single measure, and highlighted the overall preference for the trade-to-trade estimator.

A useful, practical result which emerged was the relative success of the weekly trade-to-trade estimator. Because much research in less developed countries depends on data which have been collected at best on a weekly level, the finding that the differences between it and the full (daily) trade-to-trade estimator are small constitutes a beneficial contribution for practical beta estimation on small stock markets.

Appendix 4.A

Shares used (* denotes not used for weekly analysis):

De Beers
Dries
Anglo-Am
Zandpan
Barlows
Deelkraal
Sasol
Rusplat
Kloof
Joel*
Elsburg
Harties
SA-Brews
ABSA*
Vaal-Reef
Gencor
Remgro
Genbel
Rembr-Beh
Lorraine
Southvaal
Beatrix*
Elands
Freegold*
Libanon
Leslie
Implats
Westn-Deep
Bankorp

Wes-Areas

Libvest*

Powtech

Sappi

Venters

Randfontn

Malbak

Nedcor

GFSA

Ergo

Minorco

Appendix 4.B**Details of simulation algorithm**

(Set q_1 , q_2 and q^*)¹

For each level of q from q_1 to q_2 in steps of q^*

(set number of lags for Cohen estimator)²

For 1000 shares

For $60 + \text{lag}^3$ months

Generate month-end price age⁴

Generate market return⁵

Generate residual⁶

Compute share returns⁷

Perform OLS, T-T and Cohen regressions⁸

Store parameters

Calculate summary statistics of estimated parameters

1. Initially these were set at 0.00, 0.95 and 0.05 respectively, but when using the adaptive Cohen estimator 0.750, 0.975 and 0.025 were used.
2. Initially the number of lags was set at one, but thereafter the adaptive number of lags described in Section 4.3.4.
3. For the Cohen estimator the returns on the market in the months prior to the first month in the estimation period are needed.
4. A random geometric variate, I^* , was needed to correspond to the price age in days. If I^* was greater than 20 (i.e. the share did not trade in the month, the price age (measured in months), I , was set at unity, otherwise $I = I^*/20$. (I was forced to be less than unity in months 0 and 60 so that trade-to-trade returns could be calculated in these months.)

5. The monthly returns were generated in three parts so that calendar monthly, trade-to-trade and weekly trade-to-trade returns could be obtained in the same loop. The three parts were the return on the market over the period from the end of the last traded week to the end of the month; from the time of the last trade to the end of the last traded week; and the return from the beginning of the month to the time of the last trade.
6. The residuals were generated for the share's return over the interval from the time of the last trade in the previously traded month and the time of the last trade in the current month.
7. If $I < 1$ then the share's observed return in the month was set equal to the trade-to-trade market return plus the residual, i.e. the market model with zero intercept and a slope (beta) of unity was assumed to hold. If $I = 1$ then the share's return was set to zero.
8. The regressions for the OLS, trade-to-trade, weekly trade-to-trade and Cohen estimators were performed using NAg routines.

The NAg random number generator was used for all normal and uniform random numbers. Geometric variates were obtained as follows:

- (a) $I^* = 0$
- (b) generate u (a random number uniformly distributed between 0 and 1)
- (c) if $u < q$ then $I^* = I^* + 1$, goto (b)
else stop

The coefficients for generating non-normal deviates for the Fleishman technique were calculated using an adaptation of a Numerical Recipes Fortran routine (*mnewt*) which uses a modification of Newton's method for root solving. The routine was tested by comparing coefficients with those published in Fleishman (1978).

The moments for the residuals and market returns were obtained using the models described in section 4.2.

Initially the simulation was run without computing the weekly trade-to-trade betas. In the second run the weekly trade-to-trade betas were computed. The results of the two runs were combined for the other estimation methods. There are thus results for 2000 shares at each level of thin trading for OLS, trade-to-trade and Cohen-type estimates, but only 1000 shares for the weekly trade-to-trade estimator.

Appendix 4.C

Simulations using the parameters estimated from the weekly data

In addition to studying the distribution of index and residual returns with daily data, weekly data were used to study non-normality over longer periods of time, and over longer intervals. The methodology used is the same as for the daily data: non-overlapping index returns and OLS residuals over one-week, two-week, etc. up to 50-week intervals were computed. The same analyses on the moments calculated in that way were performed. The results from two five-year periods of investigation (1982-1986 and 1987-1991) are presented as well as those for the whole period from January 1980 to December 1991.

Table 4C.1 shows the skewness and kurtosis over one-week, two-week, ..., five-week intervals. In all three periods the residuals show considerably more non-normality than the index returns, and in both periods the degree of non-normality decreases with the length of the interval over which the returns are measured. For example, it can be seen from Table 4C.1 that the four-weekly kurtosis of the residuals in the period 1980-1991 is 4.305, compared with a coefficient of kurtosis of -0.099 for the index. The four-weekly coefficients of kurtosis are far smaller than the weekly coefficients of kurtosis (which are 11.041 for the residuals and 2.440 for the index). Even though the returns were 'winsorised' as for the daily data (the largest and smallest were excluded when calculating the parameters), the effect of the 1987 'crash' is clear in both the 1987-1991 and 1980-1991 periods. The very large negative returns over the few weeks in October distort the estimates of the higher order moments significantly. The coefficients of kurtosis from the daily data (reported in Table 4.1) suggested less non-normality in the four-weekly returns than the weekly data in Table 4C.1. An explanation for this difference is that the weekly data (apart from the period 1982-1986) included the crash month and the daily data did not. In fact it can be seen from Table 4C.1 that the coefficient of kurtosis for the four-weekly kurtosis in the 1982-1986 period (which is 0.469) is more in line with the 20-day coefficient of 0.29 reported in Table 4.1.

Table 4C.2 shows the analysis of the results in the same way as Table 4.2. For some of the longer interval lengths there are of course very few non-overlapping observations in a five-year period. Because of this the variance of the returns at these long intervals is very large. Despite the weighting used in doing the regressions and calculating averages, performing regressions without fitting a constant leads to results which are unstable. Therefore, two results

Table 4C.1

The coefficients of skewness and kurtosis of returns measured using interval lengths of 1 to 5 weeks.

Interval (weeks)	Index		Residuals	
	skew	kurtosis	skew	kurtosis
	Period: 1980-1991			
1	-.591	2.440	-.277	11.041
2	-.533	1.390	-.389	5.783
3	-.462	.683	-.636	5.107
4	-.321	-.099	-.613	4.305
5	-.078	-.292	-.372	2.889
Period: 1982-1986				
1	-.050	.407	.153	1.170
2	-.159	-.197	.059	.734
3	-.124	-.225	-.002	.219
4	-.198	-.303	-.141	.469
5	.195	-.888	.027	-.111
Period: 1987-1991				
1	-1.262	5.376	-.680	10.014
2	-1.294	4.231	-.786	5.965
3	-.745	1.828	-1.191	7.030
4	-.794	1.057	-.307	1.601
5	-.609	.164	-.239	1.075

are sometimes reported: one based on all 50 interval lengths and the second using the data up to the maximum number of interval lengths which yielded satisfactory diagnostics after the regression analysis, (viz. a positive R^2 and a Durbin-Watson test statistic for autocorrelation near two). Some examples are now given to help clarify reading of Table 4C.2. Looking at the first row of Table 4C.2 it can be seen that the mean weekly return for the residuals in the period 1980-1991 is 0.314 and that the linear relationship between the returns and the length of interval showed a strong correlation with an R^2 of 98%. A second example is found by looking at the second row of the table which gives the model parameters for the residuals over the 1980-1991 period. The first thing to notice is that there is no need for a model for the mean residual as this will always be zero, no matter what the interval length. However, when it comes to fitting the linear model relating the standard deviation of the residuals to the lengths of the intervals, there are two results reported. The top result (which yields an R^2 of 67% and a weekly standard deviation of 5.86) is based on the fit of the model to all 50 interval lengths

Table 4C.2

Summary of moments and model parameters from weekly data.

	Weekly Returns				skew mean	Kurtosis		
	Mean	R ²	s.d.	R ²		α	β	R ²
Period: 1980-1991								
Index	.314	.98	3.18	.69	-.422	-1.24	3.60	.55
Resid			5.86	.67	-.228	-3.11	13.83	.93
Resid ^d			6.03	.93				
Period: 1982-1986								
Index	.386	.85	2.25	-.71	-.043	-1.38	1.83	.24
Index ^b			3.13	.88				
Resid			4.73	-.56	.042	-.89	2.15	.43
Resid ^c			5.58	.93				
Period: 1987-1991								
Index	.283	.49	2.59	.34	-.750	-2.63	8.18	.81
Index ^e	.301	.95	2.60	.84				
Resid			5.58	.38	-.377	-3.80	13.95	.92
Resid ^f			6.05	.92				

^b Maximum number of weeks considered: 15^c Maximum number of weeks considered: 15^d Maximum number of weeks considered: 40^e Maximum number of weeks considered: 20^f Maximum number of weeks considered: 40

examined. If, on the other hand, the first 15 interval lengths are considered, then the R² goes up to 93% and the weekly standard deviation is estimated as 6.03. The reason for the second fit is that the first fit produced a Durbin-Watson statistic indicating significant autocorrelation. The autocorrelation will have led to inefficient estimates of the weekly standard deviation and so a reduced data set was used to avoid the problem.

As with the daily analysis, the weekly returns and variances are additive; the coefficients of skewness do not show any clear trend as interval length increases and the model which related the kurtosis and interval length for the daily data also appears to fit the weekly data. In other words, the forms of the models suggested for the daily fitted the weekly data well in both five-year periods and is applicable to intervals up to nearly a year in length.

For the sake of completeness, the simulation results produced using the models estimated from the weekly data are given in Table 4C.3 and Table 4C.4. A comparison of the simulation results based on the parameters of non-normality estimated from weekly data with the simulation results used when the daily parameters were used, reveals very little difference.

Table 4C.3

Simulation results - average betas - distributional parameters derived from weekly data.

q	OLS	TT (daily)	TT (weekly)	lags	Cohen (adaptive)
.750	.88	1.00	.92	1	1.01
.775	.85	1.00	.92	1	1.00
.800	.83	1.00	.92	1	1.00
.825	.79	1.00	.92	1	.98
.850	.76	1.00	.92	1	1.00
.875	.71	.99	.91	2	1.00
.900	.66	1.00	.92	2	1.00
.925	.56	1.00	.92	2	.96
.950	.45	.99	.91	3	.98

Table 4C.4

Standard deviations and standard errors of estimated betas - distributional parameters derived from weekly data.

q	OLS		TT (daily)		TT (weekly)		Cohen (adaptive)	
	sd	se	sd	se	sd	se	sd	se
.750	.263	.254	.246	.243	.251	.249	.350	.376
.775	.267	.255	.253	.244	.261	.251	.380	.377
.800	.259	.257	.245	.244	.254	.253	.372	.379
.825	.259	.257	.246	.243	.257	.250	.363	.378
.850	.265	.260	.242	.246	.251	.254	.369	.379
.875	.260	.262	.245	.247	.263	.256	.469	.471
.900	.274	.265	.257	.253	.266	.262	.460	.474
.925	.271	.269	.260	.268	.275	.279	.467	.476
.950	.282	.264	.297	.300	.308	.309	.522	.557

Appendix 4.D**The derivation of a useful decomposition of the MSE**

If X is an estimator of θ and the expectation function is denoted by $E[X]$ then

$$\begin{aligned}\text{MSE} &= E[(X - \theta)^2] \\ &= E[(X - E[X]) + (E[X] - \theta)^2] \\ &= E[(X - E[X])^2] + (E[X] - \theta)^2 + 2(E[X] - \theta) E[(X - E[X])] \\ &= \text{Var}[X] + \text{bias}^2\end{aligned}$$

where $\text{Var}[X]$ is the variance function and bias represents the difference between the expected value of the estimator and θ , the value of the parameter being estimated.

Appendix 4.E**The standard errors of beta estimators when subject to varying levels of residual variance**

In this appendix, a table adapted from Bowie and Bradfield (1993a) is presented, giving evidence of the increased efficiency offered by the Cohen estimator when the residual standard deviation is small. The simulation study was based on simulating returns in a similar way to that described in this chapter with some possibly important differences. Firstly, the return distributions were all assumed to be normal. Secondly, instead of simulating a large number of 'securities' at a particular level of thin trading, the 'securities' were randomly assigned to deciles of thin trading. The deciles are given in Chapter 3. Once the security had been assigned to a decile, a particular level of thin trading (corresponding to a value from the range of trading frequencies in that decile) was randomly incorporated into the security's returns.

Table 4E.1 gives the standard deviations of the beta estimated from the simulated returns. The salient feature of the table is the improvement in the relative efficiency of the Cohen estimator (with one lagged coefficient) as the amount of random variation is reduced. For example, in the eighth decile, the standard deviation of the Cohen estimator declines from 0.26 (when the residual standard deviation is 0.02) to 0.06 (when the residual standard deviation is 0.0005) whereas the OLS estimator decreases from 0.18 to 0.06, which is a relatively smaller improvement.

Comparisons between standard deviations of the various estimators at the higher levels of thin trading can be made only with substantial caution. The reason for the caution stems from the fact that the simulated 'securities' in each decile were assigned a trading frequency at random from a range of trading frequencies (which define the decile). There is thus a spurious increase in the measured standard deviations caused by the variability in the amount of bias in the OLS estimator for the securities in a particular decile. Further comments and an indication of the magnitude of this effect are given in Appendix 4.F.

Table 4E.1

Cross-sectional standard deviations of betas at different levels of unique risk (adapted from Bowie and Bradfield (1993a)).

Method	OLS			TT (daily)			Cohen (1 lag)		
Unique Risk ^a	.02	0.01	0.0005	.02	0.01	0.0005	.02	0.01	0.0005
Decile									
1.	.18	0.09	0.01	.18	0.09	0.004	.23	0.14	0.01
2.	.17	0.09	0.02	.17	0.09	0.004	.25	0.13	0.02
3.	.19	0.09	0.03	.19	0.09	0.005	.26	0.13	0.03
4.	.17	0.09	0.03	.17	0.08	0.004	.23	0.13	0.03
5.	.19	0.10	0.03	.19	0.09	0.004	.26	0.13	0.04
6.	.19	0.10	0.04	.18	0.09	0.005	.26	0.14	0.04
7.	.20	0.11	0.05	.20	0.10	0.004	.26	0.15	0.04
8.	.18	0.11	0.06	.18	0.09	0.005	.26	0.13	0.06
9.	.17	0.13	0.09	.17	0.10	0.005	.22	0.17	0.08
10.	.27	0.20	0.17	.26	0.12	0.006	.34	0.22	0.18
Ave.	.19	0.12	0.07	.19	0.10	0.005	.26	0.15	0.07

^a The standard deviation of the residual component of the Market Model.

Appendix 4.F**Inflation of cross-sectional standard deviation for OLS estimator**

The average beta obtained from the simulation study in a decile can be expressed as

$$\hat{\beta} = \text{plim } \hat{\beta}_q + e_\beta$$

The variance of this average estimated beta will depend on not only the variance of e_β but also the variance of $\text{plim } \hat{\beta}_q$ with respect to q which is distributed uniformly across the decile. It can be shown (see, for example, Dimson and Marsh, 1983) that

$$E[\hat{\beta}] = (1 - q^k)(1 - \mu_I)\beta$$

where k is the number of days in the estimation interval (20 in this simulation case) and μ_I is average proportion of the interval between the last trade in the interval and the end of the interval (in those intervals in which trades occur). On the basis of the simulation (i.e. that the trades occur at random and with constant probability on each day) this parameter can be expressed in terms of q :

$$\mu_I = \left[\frac{q}{1-q} \frac{(1-q^{k-1})}{(1-q^k)} - (k-1) \frac{q^k}{(1-q^k)} \right] \div k$$

The variance of the first term $(1-q^k)$ in the expression for the probability limit of the estimated beta is quite easily obtainable as follows:

If $Z = q^k$ then $\text{Var}[Z] = \text{Var}[1 - q^k]$.

Since q is distributed $U(a, b)$, i.e. uniformly between the bounds in the decile a and b , Z will have density function

$$f_z(z) = \frac{1}{b-a} \frac{1}{k} z^{\frac{1}{k}-1} \quad \text{for } a^k \leq z \leq b^k$$

and the expectation and variance can be derived as:

$$E[Z] = \frac{1}{(b-a)(k+1)} [b^{k+1} - a^{k+1}]$$

$$\text{Var}[Z] = \frac{1}{(b-a)(2k+1)} [b^{2k+1} - a^{2k+1}] - E^2[Z]$$

yielding, for example when $a = 0.88$ and $b = 0.98$, $\text{Var}[Z] = 0.02737$.

However, the variance of μ_i is not easy to write in a closed form and the distribution of the product $(1 - q^k)(1 - \mu_i)$ will be even more difficult to ascertain. Monte Carlo techniques¹⁸ were therefore used to estimate the variance of the bias term for each decile. The results of the Monte Carlo runs are reported in Table 4F.1. The column headed 'Theoretical cross-sectional std. dev. of OLS estimates' in Table 4F.1 gives the expected measured standard

Table 4F.1

Expected cross-sectional variance for OLS betas with residual standard deviation of 0.01.

Decile	Underlying standard deviation (based on T-T sd's when the residual standard deviation was 0.01)	Additional variance in bias	Theoretical cross-sectional std. dev. of OLS estimates
6.	0.0927	0.0001	0.0932
7.	0.0966	0.0001	0.0100
8.	0.0932	0.0005	0.0958
9.	0.1023	0.0016	0.1098
10.	0.1243	0.0185	0.1843

deviation corresponding to the actual, underlying standard deviation. Thus, for example, if the

¹⁸ 20 000 uniform deviates were generated between the bounds of each decile using the random number generator in the SYSTAT package. q^k and μ_i were computed for each generated q and the variance of $(1 - \mu_i)(1 - q^k)$ then calculated.

true underlying standard deviation were 0.1243 (as in the row labelled decile 10), then the effect of the variance in the biases of the OLS estimates of beta would be expected to inflate the measured standard deviation to 0.1843, an increase of about 50%. It is clear that the inflation is greater for the more thinly traded deciles, and would appear to have a substantial effect only in decile 10.

Appendix 4.G

Comparative levels of non-normality on the JSE and the NYSE

Although this study was not specifically intended for comparison with non-normality studies in other markets it does of course provide comparable estimates of the skewness and kurtosis which have been reported for other markets.

The results on the non-normality of residuals are compared with those of Affleck-Graves and MacDonald (1989) from the NYSE. They reported the coefficients of skewness and kurtosis of the Dow-Jones 30 Industrials monthly residuals. The average coefficients of skewness and kurtosis for the 30 Dow-Jones industrials can be calculated and compared with the average coefficients of the 4-weekly¹⁹ interval residuals from this study of 35 well traded shares from the JSE.

Table 4G.1 gives the average four-weekly coefficients from the JSE together with the

Table 4G.1

Comparison of skewness and kurtosis with that from NYSE.

	NYSE	JSE		
	(1951-1968)	1980-1991	1982-1986	1987-1991
skewness	.321	-.613	-.141	-.307
kurtosis	.660	4.305	.469	1.601

results from Affleck-Graves and MacDonald for the NYSE. The period chosen by Affleck-Graves and MacDonald was that used by Fama (1976). It excludes major crashes (e.g. 1969) and there is hence no *a priori* reason to suspect that it is an uncharacteristic period. It can readily be seen that the NYSE results correspond more closely to the JSE results for 1982-1986 than the 1987-1991 and 1980-1991 results which include the 1987 crash and which are more extremely non-normal as might be expected.

In conclusion, it appears as if the residual series of JSE stocks are, on average, more significantly non-normal than NYSE securities. This finding makes it all the more important to incorporate non-normality into any analysis of financial statistics on the JSE.

¹⁹ The estimates of kurtosis and skewness derived from the weekly data.

CHAPTER 5

THE STABILITY AND EFFICIENCY OF BETA ESTIMATORS: EMPIRICAL EVIDENCE FROM THE JSE

5.1 Introduction

Although the issue of thin trading and its potential impact on systematic risk estimation has been considered for many years, it is only recently that attention has focused on the problem in South Africa. Consequently many of the studies requiring systematic risk estimation performed in the past on the JSE will have been undermined by biased estimates of beta coefficients. This chapter sets out to determine whether thin trading will indeed have affected traditional estimates of beta coefficients and further to establish how the biases will have distorted inferences about the behaviour and characteristics of systematic risk on the JSE. Particular attention is paid to the key practical issue of *stability* of beta coefficients. Estimates of *historical* betas are of possible use to researchers and finance academics, but, for the practitioner, it is the *future* beta coefficients which are required for sensible financial projections and investments. Stability of the beta coefficients therefore plays an important role in assessing the usefulness of estimates of systematic risk.

This chapter then chiefly examines the stability of measures of systematic risk on the JSE. The major focus of the investigation is on establishing whether the stability of beta coefficients in South Africa is similar to that in the UK and the USA, once the estimation biases caused by thin trading are avoided. Additional investigations discussed in this chapter provide preliminary evidence of ways for improving the efficiency of beta estimators.

In Chapter 3 thin trading *per se* was investigated empirically and the theoretical biases caused by that thin trading in estimators of beta were derived analytically. The simulation study in Chapter 4 introduced an additional level of realism into assessing and contrasting two thin trading correction procedures, namely the trade-to-trade estimator and the Cohen estimator. The conclusion reached from the controlled environment of the simulation study was that the trade-to-trade estimator was the preferred estimator in terms of its unbiasedness and efficiency. It was shown that even when its rather substantial data requirements had to be compromised (for example, by using weekly data instead of daily data) it remained superior. The empirical study

in this chapter continues the examination of thin trading estimators by assessing their applicability on *actual data* from the JSE. It is believed to be the first comprehensive empirical study of systematic risk on the JSE and is designed to act as a springboard for the improvements in beta estimation considered in the remaining chapters of the thesis.

The empirical study is split into four major sections (5.2, 5.3, 5.4 and 5.5) which can be treated separately to facilitate reading, although they are inter-related to some extent. The motivation for examining each of the subjects is given in some detail at the start of each major section. Broadly speaking, the contents of the major sections are as follows. In section 5.2 some of the classic measures of beta stability are applied to estimates of beta from the JSE. The results are compared with those reported in the UK by Dimson and Marsh (1983). Section 5.3 is closely linked with section 5.2 in that it deals with the stability of betas. However, the emphasis is focused more on a specific form of instability, namely the *regression tendency* of estimated betas. The *regression tendency* of beta coefficients describes the phenomenon observed on many stock markets¹ in which securities with high estimated beta coefficients in one period tend to have lower estimates of beta in the next period, and vice versa. Section 5.4 moves away slightly from the issue of stability of beta coefficients. Instead, it contains a discussion of a proposal for improving the efficiency of estimators of beta coefficients. The proposal is based on weighting the observations used for beta estimation in relation to the volumes traded. The effect on the efficiency and bias of beta estimators caused by using different interval lengths is also examined. The final major section, 5.5, is a fairly in-depth discussion on measuring the standard errors of the Cohen estimator in an empirical investigation. In this section a bootstrap approach is used to assess the quality of a simple estimate of the standard error.

5.2 Measuring the stability of beta coefficients

The stability of beta coefficients is clearly of critical importance to practitioners who wish to use estimates of beta coefficients to make investment projections and decisions. It should be borne in mind that it is only for predictive purposes that stability of beta coefficients may be considered a desirable phenomenon. Many firms undergo fundamental changes over time, both in operational and financial structure. The changes in structure will usually lead to

¹ For example, Blume (1971) or Dimson and Marsh (1983) provide evidence of a *regression tendency* in beta coefficients in the US and UK markets, respectively.

changes in the systematic risk of the firm. It is only for those firms in which no changes have occurred that stability of beta coefficients over time is expected and desirable. Studies of beta stability must therefore be interpreted as ascertaining to what extent securities have constant betas, thereby giving an indication of how reliable a current estimate of beta is likely to be as a predictor of the future beta given that the true future beta may change. It is therefore of considerable practical value to measure the degree of stability in systematic risk coefficients. However, the investigation into beta stability is far more onerous than might at first be thought in the case of thinly traded markets, such as the JSE. As it turns out, a problem arises in providing evidence of beta stability because its assessment can be severely distorted by the biases in estimates of betas caused by infrequent trading. The focus of this major section² is then on measuring beta stability in a way which is not affected by thin trading.

In section 5.2.1 the possible distortions that thin trading may cause when assessing beta stability using unadjusted OLS beta estimates are discussed. It is shown that the magnitude of the distortion depends on the stability of the trading frequency of the securities. Therefore, in order to gauge the expected extent of any spurious stability or instability in the estimates of the beta coefficient, the stability of trading frequencies is discussed in section 5.2.2.

In section 5.2.3 the association between beta estimates and thin trading is tested empirically. The reason for the investigation is to ascertain whether the association between thin trading and OLS estimates of beta is sufficiently strong to warrant concern about the distortions which thin trading may cause for accurate assessments of stability.

In sections 5.2.4, 5.2.5 and 5.2.6 various measures of stability are calculated. Apart from the Chow (1960) test which is performed in section 5.2.6, none of the assessments in this chapter constitutes a strict statistical test. Some effort, therefore, is expended in comparing the results with those from overseas markets in order to establish whether similar confidence can be applied to the stability of JSE betas as is applied to the betas in the UK and US.

² Much of the development in sections 5.2 and 5.3 has been adapted from a paper by Dimson and Marsh (1983) who performed a similar study on UK data. The sections where the development in this thesis differs from theirs are noted in the text.

5.2.1 The spurious stability in OLS estimates of beta caused by thin trading

Dimson and Marsh (1983) argue that the common approaches³ to investigating the stability of betas in successive periods are dependent on the biases in beta estimation caused by thin trading. The problem arises because the stability of the true (but unobserved) betas from one period to the next has to be assessed indirectly by using *estimates* of the betas. They argue that measured stability might therefore be a consequence of a stable bias in the estimates of the betas rather than stability of the true betas. To demonstrate how spurious stability arises, it is instructive to consider theoretically how the correlation between successive *estimates* of betas is related to the correlation between successive *true* betas.

Betas are typically estimated from the well-known market model:

$$r_{i,t} = \alpha_i + \beta_i r_t^m + u_{i,t} \quad (1)$$

where $r_{i,t}$ is the continuously compounded return on share i in interval t ; r_t^m is the return on the market index in interval t and $u_{i,t}$ is an independently distributed disturbance term. It is further usually assumed that the true values of beta (the β_i) are constant within an estimation period,⁴ but vary stochastically from one period to the next.

Dimson and Marsh (1983) explain that the correlation coefficient between the estimates of betas in two successive periods is biased upwards if the beta estimator is itself biased, for example by thin trading. To demonstrate the upward bias in the correlation coefficient of the estimated betas, consider a security which has an *expected* true beta of unity and a standard deviation of the true betas of σ_β . It is shown in equation (2) below that the estimated correlation coefficient between the estimates of beta in successive periods can be expressed as a weighted average of the correlation coefficients of the true beta and of the disturbance (residual) term

³ For example Sharpe and Cooper (1972) use transition matrices describing the movements of shares from one risk class to another and Blume (1971) uses the correlation coefficient between the risk estimates in successive periods.

⁴ The estimation period is defined to be the sequence of intervals from which observations of the security and market returns are taken.

in the market model (the u_{it} term in equation (1)). The weighting factor depends on the relative sizes of the standard errors of the disturbance term and the standard error of the estimated beta. If σ_e is defined as the standard error of the estimate of beta; ρ as the true correlation coefficient between successive true betas and ρ^* as the correlation coefficient between the residuals of successive estimates of beta then, following Dimson and Marsh (1983), the correlation between estimates of β_i in successive periods can be written as:

$$\hat{\rho} = (\rho \sigma_\beta^2 + \rho^* \sigma_e^2) / (\sigma_\beta^2 + \sigma_e^2) \quad (2)$$

When the estimates of beta are unbiased, ρ^* (the correlation between the errors of estimation in successive periods) will be zero. Hence with prior information about σ_β (the standard deviation of the true betas) Blume (1971) has shown that the stability of the true betas over time can be measured. On the other hand, in the case where the estimator of beta is biased the errors of estimation are serially correlated because of the consistent bias, i.e. ρ^* is not zero. Hence two shares with identical true correlation coefficients (ρ) and standard deviations (σ_β), but differing biases in the beta estimation procedures (and consequently differing correlation coefficients of standard errors) will have differing estimated correlation coefficients. The difference in the correlation coefficients between the estimated betas will result in misleading statements about the relative stability of the betas of the two shares.

To demonstrate how spurious assessments of stability arise when OLS is used without any adjustments, it should be noted that the OLS estimates of the beta of thinly traded shares are theoretically biased downwards. In fact it was shown in Chapter 3 that the bias in the OLS estimate of beta is dependent on the 'price ages' of month-end⁵ observations (i.e. the number of days between the last trade of the security's shares in the month and the end of the month)

⁵ The data usually used for OLS beta estimation are the month-end prices of the securities.

and the number of months in which no trade has occurred at all. Defining μ_t to be the mean price age of a particular share and Ψ to be the proportion of months in which that share does trade then the probability limit ('asymptotic expectation') of the OLS estimator of beta is given by the expression derived in Chapter 3, viz.:

$$\text{plim } \hat{\beta}_{OLS} = \Psi (1 - \mu_t) \beta \quad (3)$$

It can be seen from equation (3) that the OLS estimate of beta will be less than the true beta if the proportion of months in which the security trades (the Ψ) is low or if the mean price age (μ_t) is high.

Dimson and Marsh (1983) stress that the spurious stability of betas is not a result of using the correlation coefficient to measure beta stability, but rather because of the bias in OLS estimates of beta. In fact the converse might also hold; the true, underlying betas might be stable, but the degree of the bias in successive periods might vary. The variation in the bias, caused by variation in the degree of thin trading, would introduce spurious variation in the estimates of beta and *reduce* the estimated correlation coefficient, $\hat{\rho}$. Considering the expression for $\hat{\rho}$ in equation (2), the reduction would occur because ρ^* (the correlation between the estimation errors) was negative. Dimson and Marsh (1983) argue that because of the relationship between trading frequency and the usually inert size of the firm (Dimson, 1979), the correlation will generally be positive.

In conclusion, it was argued above from a theoretical perspective that OLS estimates are unsuitable for assessing the stability of systematic risk because of the bias in the OLS estimates caused by thin trading. In particular, it was contested that a consistent upward or downward bias in the estimates would lead to spurious stability of the estimates from one period to the next. On the other hand, if the biases in the estimates were negatively correlated then there would be spurious *instability* in the estimates.

In the next section the study proceeds to investigate the stability of measures of thin trading. From the above discussion it is apparent that stable measures of thin trading will lead to spurious stability of beta coefficients. Any instability in the levels of thin trading will cause spurious *instability* in the OLS estimates of beta.

5.2.2 The effect of changing levels of trading frequency on the stability of OLS estimates of beta

In section 5.2.1 the argument presented by Dimson and Marsh (1983) as to how thin trading causes biases in the estimators of beta and how the biases in turn cause false statements to be made about the stability of betas was discussed. In this section the work of Chapter 3 on the stability of thin trading is extended by examining how shifts in the level of thin trading of a share would cause spurious instability of OLS estimates of beta.

In Chapter 3 the stability (on the JSE) of a measure of thin trading, namely the proportion of intervals in which a security is not traded, was examined. On the basis of an empirical analysis it was demonstrated that the number of securities experiencing significant changes in thin trading from one year to the next could be attributed to random fluctuations in the trading frequencies. In Appendix 5.F the stability of longer term measures of thin trading is measured, i.e. the measures of thin trading based on five-year periods. From Appendix 5.F, it can be seen that the correlation between the trading frequencies in successive five-year periods averaged well over 0.9. The high correlation is similar to that reported by Dimson and Marsh (1983) for UK stocks where correlations of about 0.90 were also cited.

In this section a slightly different approach to assessing the stability of thin trading is adopted. The reason for the change is to emphasize that (in the context of this study) it is the changes in the *bias* of OLS beta estimates that are the important outcomes of changes in trading frequency. In this study, therefore, the thin trading of a security is measured by the *theoretical bias* in the OLS estimate of the security's beta. The relationship between a security's trading frequency and *theoretical bias* was derived in Chapter 3 and thus, using those results, changes in the trading frequency can be translated directly into changes in *theoretical bias*.

The empirical investigation in this section is therefore based on measuring the changes in the *theoretical bias* of securities on the JSE from one year to the next. The investigation continues as follows: initially, each security is assigned to a category which corresponds with a narrow band of *theoretical bias*. Then, in each subsequent year the number of securities which move from one class of *theoretical bias* to another will be counted. If trading frequency is stable then very few large changes in *theoretical bias* will be encountered. On the other hand, instability in trading frequency will manifest itself as a large number of changes in *theoretical bias*.

The categories of *theoretical bias* were initially defined as 0-5%, 5-10%, 10-15%, etc., of the true beta. Using the relationships between the components of the *theoretical bias* and the daily trading frequencies derived in Chapter 3,⁶ each category of *theoretical bias* was associated with a range of daily trading frequencies. Then, because weekly data were available for much longer periods of time, the daily trading frequencies were converted into weekly trading frequencies.

The categories of *theoretical bias* and the maximum levels of daily and weekly trading frequency are given in Table 5.1. It can be seen from Table 5.1, for example, that the well traded securities (i.e. those in category 1) are those securities with very little expected bias in their OLS estimates of beta (only between 0% and 5% of the true beta). Moreover, the

Table 5.1

Details of how shares were categorised according to the theoretical biases in OLS estimates of beta.

Category	Theoretical bias (as a percentage of the true beta)	Proportion of days not traded ¹	Proportion of weeks not traded ²
1	0 - 5	0.500	0.031
2	5 - 10	0.667	0.132
3	10 - 15	0.751	0.238
4	15 - 20	0.802	0.331
5	20 - 25	0.837	0.411
6	25 - 30	0.864	0.481
7	30 - 35	0.884	0.541
8	35 - 40	0.902	0.595
9	40 - 45	0.916	0.644
10	45 - 50	0.928	0.688
11	50 - 55	0.939	0.729
12	55 - 60	0.948	0.767
13	60 - 65	0.957	0.802
14	65 - 70	0.965	0.835
15	70 - 75	0.972	0.866
16	75 - 80	0.978	0.896
17	80 - 100	1.000	1.000

¹ The maximum daily trading infrequency (measured as the proportion of days not traded) associated with the theoretical bias.

² Calculated from the value in column three: q_{jp}^1

⁶ The expressions are repeated for convenience in Appendix 5.A.

probability that a security in category 1 will not trade in a day is less than 0.5, or, equivalently, the probability that a category 1 security will not trade in a week is less than 0.031.

Once the categories of thin trading had been set up as in Table 5.1, the transition matrices of the number of securities which had changed category from one year to the next could be calculated. The way in which the transition matrices were calculated is as follows. The trading frequency for the security was estimated by calculating the proportion of weeks in each year in which the security was not traded. The security could then be allocated to a category based on its *theoretical bias* for each year using the reference values in Table 5.1. The number of securities changing category from one year to the next were counted and the transition matrices built up. All 246 of the securities on the JSE which had complete data for the entire period of investigation (February 1973 to December 1991) were used.

The average of the 18 possible year-on-year transition matrices is given in Table 5B.1 in Appendix 5.B. It is a very sparse matrix and so does not convey a clear idea of the changes that occurred from year to year. A useful summary figure from the matrix is therefore the (average) percentage of securities which remained in the same category of bias from one year to the next. By counting up the number of securities on the diagonal of the matrix the percentage of securities with a stable *theoretical bias* is found to be 56%. The percentage of securities in the *same or adjacent* category is 83% which indicates relative stability of the trading frequencies.

To improve clarity the matrix was collapsed by re-categorizing the securities into four new categories which each corresponded to a wider range of trading frequencies than the original categories in Table 5.1. The four new categories were labelled 'good', 'acceptable', 'poor' and 'extremely poor'. The label names are designed to give some indication of how significant the *theoretical bias* is relative to the size of the standard error in the estimate of beta. For instance, a 'good' bias is intended to correspond to the case where the bias in the OLS estimate of beta is small relative to the standard error. Before the *theoretical biases* in the four categories can be specified, a value for the standard error of the beta estimate must be specified. For simplicity a single, typical standard error was considered. A standard error of 0.15 is a plausible, typical value for securities on the JSE (see Bradfield and Bowie, The Financial Risk Service). Based on the selected standard error, the four categories were defined as follows: 'good' corresponded to a bias of 0-15%, i.e. the expected bias caused by thin trading is within one standard error of a true beta (which is assumed to be unity). The category 'acceptable' corresponded to securities with expected biases between 15% and 30% (i.e.

between one and two standard errors). 'Poor' corresponded to securities with biases between 30% and 45% (i.e. between two and three standard deviations) and 'extremely poor' corresponded to biases greater than 45%.

The average of the 18 re-categorized transition matrices is given in Table 5.2. It can be seen Table 5.2, for example, that of the 190 securities which were in category 'good' in year p , on average, 178 of them remained in the 'good' category; 11 became more thinly traded and moved to the 'acceptable' category; one became substantially more thinly traded and

Table 5.2

Average transition matrix of thin trading parameters.

Category of estimate in year p	Category of estimate in year $p+1$				Row tot.
	Good	Acceptable	Poor	Extremely poor	
Good ^a	178	11	1	0	190
Acceptable ^b	9	15	4	2	30
Poor ^c	1	4	5	3	13
Extremely poor ^d	0	1	3	9	13
Col. total	188	31	14	14	246

- ^a Securities with a level of thin trading which corresponds to an expected bias in the OLS estimate of beta of less than 15% of the true beta.
^b Securities with an expected bias of between 15 and 30%.
^c Securities with an expected bias of between 30 and 45%.
^d Securities with an expected bias of greater than 45%.

The table was constructed by counting the average number of securities moving from each category in year p to each category in year $p+1$.

moved to the 'poor' category. No securities moved from the 'good' category into the 'extremely poor' category. All the securities off the diagonal of the matrix, i.e. the numbers in the unshaded cells of Table 5.2, are securities which have changed trading frequency enough to change their *theoretical bias* from one category to another. Based on the small average number of securities which do change category each year (39 (16%) of the 246 securities in the sample), the conclusion reached is one of stability in the trading frequencies and it has been noted that stability in trading frequency implies a possibility of spurious stability in the OLS betas.

The next step in evaluating the possibility of distorted evidence about stability of betas is to confirm that the thin trading biases predicted in OLS estimates of beta do in fact occur in actual data.

5.2.3 Empirical evidence of the downward bias in estimates of financial risk caused by thin trading on the JSE

The previous two sections have indicated that if estimators of beta coefficients are biased by thin trading then distorted inferences may be drawn about the stability of betas. In Chapter 3 it was shown analytically how thin trading would affect an OLS estimator of beta. The downward bias in estimates of beta was confirmed in the simulation study of Chapter 4. It is, however, necessary to verify that the theoretical impact of thin trading on estimates of beta are borne out by actual data.

The obvious problem in providing empirical evidence of bias in the estimator is that the true parameter value is unknown. Researchers have generally attempted to avoid the problem of an unknown parameter by reasoning that there is no apparent justification why the more thinly traded shares should have lower (true) betas than the better traded shares.⁷ They then argue that the empirical evidence of the bias is equivalent to a significant relationship existing between the level of thin trading and the estimated betas. More generally, the thin trading bias can be established by observing a *more* significant relationship between the estimated betas and the trading frequencies when OLS is used to estimate the betas than when one of the thin trading correction procedures is used to estimate the betas.

The existence of a relationship between the price ages estimated by q_{jp} (the measure of trading frequency) and the estimates of beta was tested using a similar procedure to that of Dimson and Marsh (1983). The procedure involved regressing the estimates of beta against I_{jp} (the price age of the security estimated using $I_{jp} = 1 - \Psi(1 - \mu_j)$). The model used to perform the test is given by

$$\hat{\beta}_{j,p} = \alpha + \gamma I_{j,p} + u_{j,p} \quad (4)$$

In equation (4) above, $\hat{\beta}_{j,p}$ is the beta of security j in period p estimated by one of the three methods (trade-to-trade, Cohen or OLS); $I_{j,p}$ is the estimated mean price age of security j in period p ; $u_{j,p}$ is a random error term; α is the constant of the regression and γ is the slope coefficient. In order to test for the relationships between the estimated beta coefficients and the

⁷ Dimson and Marsh (1983) do, however, report that in the UK smaller companies tend to have lower betas and suggest that smaller companies are also more likely to be more thinly traded because of their smaller capital issues.

measures of thin trading, OLS, trade-to-trade and Cohen⁸ betas were used. Estimates of the betas were made for the securities for which continuous price information was available in three consecutive non-overlapping five-year⁹ periods. The numbers of securities in each period were thus 289 securities for the regressions in the periods 1982 to 1986 and 1987 to 1991 and 263 securities for 1977 to 1981.

Firstly, the three sets of betas were regressed against the corresponding values of I_{jp} estimated using the trading frequencies based on the proportions of weeks not traded in the periods. Secondly, the sets of betas were regressed against the I_{jp} computed from the proportion of *days* not traded in the period October 1987 until December 1991.¹⁰ The regression was performed even though there was no direct correspondence between the periods in which the betas were estimated and the period in which the thin trading was measured. Performing the regression against the daily data enabled two things to be judged: (a) whether the underestimation of the bias component, which is inevitable when using weekly data,¹¹ reduced the relationship between the OLS estimates of beta in the 1987-1991 period and the measure of thin trading, and (b) for the betas estimated in the other two time periods (1977-1981 and 1982-1986) whether the stability of the thin trading measures was borne out by a consistent relationship between I_{jp} and the estimated betas.

The results of the regressions of the estimated betas on the measures of thin trading are presented in Table 5.3. The key feature of Table 5.3 is that all the slope coefficients (the γ in equation (4)) are negative. In other words, high estimated beta coefficients are related to low levels of thin trading and vice versa. Concentrating on the top panel of Table 5.3, it is apparent that the magnitudes of the slope coefficients are most pronounced for the OLS estimates (on average -2.24) and least so for the Cohen estimator (-1.71). The trade-to-trade estimator has a somewhat intermediate value of -1.85.

⁸ The number of lags for the Cohen estimator was chosen in an adaptive manner. Using the proportion of days as the estimate of thin trading the number of 20-day periods (months) which would be necessary to capture *all* the non-synchronous movements with 95% probability were calculated, i.e. $q_{jp}^{20 \times lag} < 0.05$ was solved for the smallest value of *lag*, assuming there were 20 trading days in a month.

⁹ The conventional 60 months of trading data were used to make the estimates.

¹⁰ The period for which daily data were available.

¹¹ Again, weekly data were available for a longer period of time than daily data. It is therefore instructive to establish the differences which might arise between inferences drawn when using the different frequencies of data. The most important difference when using weekly data is that intra-weekly delays are ignored in measuring the trading frequencies. This will result in a downward bias in the estimates of trading frequency.

Table 5.3

Regressions of estimated betas on the theoretical biases in the OLS estimates

Dep Var	Indep Var		77-81	82-86	87-91	Ave
OLS	Weekly ^a	Const. ^b	.88	.78	.78	.81
		Slope ^c	-2.912	-2.077	-1.732	-2.24
		t-stat	(-6.69)	(-8.41)	(-7.20)	(-7.4)
		R ² (%)	15%	20%	15%	17%
T-T		Const.	.88	.79	.80	.82
		Slope	-2.488	-1.793	-1.263	-1.85
		t-stat	(-5.60)	(-7.35)	(-5.28)	(-6.1)
		R ² (%)	11%	16%	9%	12%
Cohen		Const.	.99	.93	.91	.94
		Slope	-2.307	-1.952	-.858	-1.71
		t-stat	(-4.85)	(-7.64)	(-3.41)	(-5.3)
		R ² (%)	8%	17%	4%	10%
OLS	Daily ^d	Const.	.88	.80	.81	.83
		Slope	-1.816	-1.604	-1.548	-1.66
		t-stat	(-5.30)	(-7.89)	(-8.85)	(-7.3)
		R ² (%)	10%	18%	21%	16%
T-T		Const.	.91	.80	.83	.85
		Slope	-1.920	-1.368	-1.114	-1.47
		t-stat	(-5.66)	(-6.82)	(-6.28)	(-6.3)
		R ² (%)	11%	14%	12%	12%
Cohen		Const.	.96	.94	.93	.94
		Slope	-1.026	-1.468	-.752	-1.08
		t-stat	(-2.74)	(-6.96)	(-3.99)	(-4.6)
		R ² (%)	3%	14%	5%	7%

^a I_{tr} estimated from proportion of weeks not traded^b Alpha in equation (4)^c Gamma in equation (4)^d I_{tr} estimated from proportion of days not traded

The extent of the correlation between the estimates and the trading frequency can be gauged from the R^2 statistics given in Table 5.3. The R^2 of the trade to trade estimator is, on average, 12%, compared with a value of 17% for the OLS estimator. The evidence thus supports the hypothesis that the trade-to-trade and Cohen estimators are less correlated with thin

trading than the OLS estimator. The implication of these findings is that thin trading does indeed cause bias in actual estimates of beta coefficients when OLS is used.¹²

It can also be seen from Table 5.3 that the R^2 from regressing the OLS betas against the I_{jp} calculated from daily data in 1987-1991 is equal to 21%. In contrast it should be noted that the R^2 of the regression between the OLS betas and the measure of thin trading derived from weekly data in the same period is a much lower 15%. The implication of the increased correlation when daily data are used is that the degree of the downward bias in beta is better assessed from the proportion of days on which a share does not trade rather than the proportion of weeks not traded. For example, a share which trades every week may be trading only once every five days, which would lead to bias in the OLS estimates of beta, but which would not be picked up as being thinly traded on a weekly basis.

The second reason for including the daily measures of thin trading in Table 5.3 was to ascertain whether the same relationship between bias and daily data also held when the estimation period for the betas did not correspond to the estimation period for the trading frequencies. From Table 5.3 the R^2 of the relationship between OLS estimates of beta and daily trading frequency for 1977-1981 can be seen to be 10%. This value is lower than the 18% for 1982-1986 which is in turn lower than the R^2 of 21% in 1987-1991. If thin trading were a stable characteristic of securities then the R^2 statistics would be constant over all three periods. The inference that can be drawn from the increasing correlations associated with the convergence of the beta estimation periods and trading frequency estimation periods observed in Table 5.3, is that trading frequencies in adjacent periods are more stable than trading frequencies in more distant periods.

The constant (y-intercept) in the regression can be interpreted as the average estimated beta for securities with no thin trading. OLS estimates should be optimal and unbiased for the

¹² The fact that the trade-to-trade and Cohen estimators are also significantly correlated with thin trading indicates that the true betas and trading frequency are also indirectly associated. In order to test whether the relationship between the OLS betas and the I_{jp} is 'significantly' more correlated than the other estimators, the following testing procedure can be used. The OLS estimates were regressed against the trade-to-trade estimates of beta. The residuals from this regression can be thought of as the components of the OLS betas which are pure bias. In other words, by performing the regression of the OLS betas against the trade-to-trade betas the underlying relationship between the true beta and the thin trading is removed. If the residuals from that regression of the OLS betas against the trade-to-trade betas are still correlated with the trading frequency, then the thin trading bias is being identified directly. The suggested procedure was followed. The regression of the OLS estimates against the trade-to-trade estimates yielded residuals which were regressed against the daily measure of frequency. (The betas from the period 1987-1991 were used because it was for that period that daily trading frequencies were available.) The regression of the residuals yielded an R^2 of 19% and a slope coefficient (t -statistic) of -0.44 (-8.28). The fact that the correlation remains significant (in fact it has hardly changed from the 21% reported in Table 5.3) and that the slope coefficient is still negative, indicate that it is indeed the bias in the OLS estimates which are related to the thin trading measures.

well traded sector of the market if thin trading is the only source of bias. Thus it would be expected that estimators correcting for bias would converge to the OLS estimates for well traded shares, i.e. that the constant would be the same for the estimates of the well traded shares. From Table 5.3 it is clear that the average constant from the trade-to-trade betas (of 0.82) is only marginally higher than the OLS constant (of 0.81). The Cohen estimates yield a value considerably higher than both trade-to-trade and OLS (of 0.94). The higher levels of Cohen beta in the well traded sector of the market suggests that there might be some other cause of friction¹³ in the pricing process (in addition to thin trading) which the Cohen estimator is correcting for, but for which the trade-to-trade estimator is not.

The investigation in this section set out to establish whether the biases theoretically caused by thin trading do actually manifest themselves in empirical estimates of beta. The results presented in Table 5.3 show that thin trading is indeed a source of bias in empirical estimates of beta and furthermore the bias is of the form predicted. In continuing the investigation into the stability of betas, estimators which avoid the biases caused by thin trading are therefore necessary to avoid spurious inference being drawn.

5.2.3.1

Comparison with UK stocks

As with the Dimson and Marsh (1983) study on UK stocks, the slope coefficient remains significantly negative even when estimators which avoid the bias associated with thin trading are used. On average, the slope coefficient for UK securities was reported by Dimson and Marsh to be -1.41 for the OLS estimator and -1.02 for the trade-to-trade estimator (the values for the JSE reported in Table 5.3 are -2.24 and -1.85). The fact that there is still a significant relationship indicates that there might be some other latent factor causing bias in addition to that caused by thin trading. Dimson and Marsh (1983) provide some suggestions as to what this other factor might be: namely, the size of the firms, or delays in the actual transaction prices for small firms. They also show that by adjusting both the trade-to-trade betas and the trading infrequencies to take into account the size of the firm, the negative relationship between the betas and thin trading is reduced considerably.

The Cohen estimator, on the other hand, is designed to correct for all price adjustment delays which might cause biases in estimates of beta coefficients. The evidence from the JSE

¹³ Cohen *et al* (1980, 1983) provide a list of possible other frictions.

given in Table 5.3 is that, although there is still a significant relationship between the Cohen betas and thin trading,¹⁴ the extent of the correlation is less than that of the trade-to-trade estimator. Unfortunately for the sake of a complete comparison with UK stocks, Dimson and Marsh do not actually use the Cohen estimator in their paper on UK securities. However, Dimson (1979) in his aggregated coefficients paper¹⁵ provides average security betas across deciles of securities based on the ranking of thin trading parameters in the UK. Some indirect inference about the extent of any relationship between his aggregated coefficients estimates of beta and thin trading can thus be drawn. In his Table 7 it is apparent that the average beta in deciles 1 to 5 is 1.04, and for deciles 6 to 10 is 0.96. The fact that there is little difference between the estimated betas in the well traded deciles and those in less frequently traded deciles suggests that the slope relating the Cohen estimates of betas in the UK might also be less marked than the slopes observed for the trade-to-trade and OLS estimates.

Before proceeding to the empirical investigation of beta stability on the JSE, it is worth recapping on what the investigations have highlighted thus far. Firstly, the necessity of correcting for any thin trading biases in estimates of beta was demonstrated theoretically. Secondly, an empirical study of the stability of measures of thin trading confirmed that the theoretical distortions of measures of stability would indeed occur *if* trading frequencies and unadjusted OLS estimates of beta *were* correlated as predicted analytically. The findings of the previous section indicated exactly that: thin trading and OLS estimates of beta are significantly correlated. Therefore estimators which correct for the biases caused by thin trading must be used to accurately assess stability. The findings were shown to be similar to those reported by Dimson and Marsh (1983) in the UK.

The next three sections are the central ones of this investigation. In each of them, a time-honoured method of assessing beta stability is applied and the results compared with the UK and US markets.

¹⁴ Possibly due to the firm size effect.

¹⁵ Dimson (1982) argues that his version of the aggregated coefficients methodology gives very similar estimates to the Cohen estimator.

5.2.4 Empirical investigation into the stability of betas using mean squared forecast errors

In the previous sections it has been established that there is indeed a biasing effect on OLS estimates of beta caused by thin trading and that biased estimators provide distorted measures of beta stability. In this and the following two sections, the stability of betas on the JSE is assessed using estimators which avoid the bias caused by thin trading.

The first investigation into stability is based on the well-known mean squared forecast error (MSE) which is used to analyze the predictability of betas from one period to the next. The assessment of stability using MSEs is predominantly a comparative one. The results are contrasted with those recorded on other stock exchanges as well.

The advantage of using the MSE (over the correlation coefficient, for example) for measuring stability, is that it can be divided up into three interpretable components (after Theil (1971)): bias, inefficiency and random error. This decomposition allows researchers to identify the reason *why* one estimator might be a better predictor than the other. If the beta estimated for a security j in a period p is denoted by $\hat{\beta}_{j,p}$ then, following Theil (1971), the MSE can be written as:

$$\begin{aligned}
 MSE &= \sum_{j=1}^n (\hat{\beta}_{j,p+1} - \hat{\beta}_{j,p})^2 / n \\
 &= (\bar{\beta}_{j,p+1} - \bar{\beta}_{j,p})^2 + (1 - \hat{\gamma})^2 \text{Var}[\hat{\beta}_{j,p}] + (1 - \rho^2) \text{Var}[\hat{\beta}_{j,p+1}] \\
 &= \text{bias} + \text{inefficiency} + \text{randomness}
 \end{aligned} \tag{5}$$

where $\hat{\gamma}$ is the slope of the regression of the beta in period $p+1$ on the beta in period p ; ρ is the estimated correlation coefficient between the two estimates of beta; $\bar{\beta}_{j,p}$ is the average beta in period p .

The first term in the decomposition in equation (5) corresponds to 'bias' in the prediction of $\hat{\beta}_{j,p+1}$ by $\hat{\beta}_{j,p}$. It is important to note that the 'bias' referred to in this section is *not* thin trading bias, but rather the extent to which *overall* beta levels have changed from period to period. The bias term will be small if the *average* beta coefficient has not changed from the one period to the next. The second term reflects the 'inefficiency' of the prediction. The inefficiency measures the extent to which the estimates of individual betas have changed. If all

the estimates of the betas are about the same value as in the period before then the $\hat{\gamma}$ will be close to unity and the inefficiency will be close to zero. The third term measures the random error, i.e. the relative order of the securities as ranked by their betas in the two periods.

The expression in equation (5) is used in this section to measure the stability of betas obtained from the OLS, trade-to-trade and Cohen estimators on the JSE from one five-year period to the next. The periods under consideration are January 1977 to December 1981, January 1982 to December 1986, and January 1987 to December 1991.

Another estimator considered in the investigation is the OLS estimator adjusted to correct for the expected bias in the OLS estimator. The adjustment was effected by dividing the OLS estimate of beta by $\Psi(1 - \mu_p)$, which can be recognized as the fraction that the estimated OLS beta is expected to be of the true beta. The adjustment of the OLS estimator makes it an unbiased estimator. However, the adjustment also rescales the variance of the estimate. Thus, for instance, a security beta which is expected to be 50%, say, of the true beta will have its standard error doubled because of the division of the estimate by the fraction, 0.5.

Table 5.4 summarises¹⁶ the results of the MSEs calculated using equation (5). The estimate of the security's beta produced by a particular method made in the first period is used as a predictor for all the other estimates of that security's beta in the second period. This is done so as to avoid prejudging which estimator is best and, further, to give insights into the relative efficiencies of the different unbiased estimators. The estimator which is the most efficient of the unbiased estimators can be expected to yield lower inefficiency components and total MSEs when predicting the estimates for *all* the other unbiased estimators in the second period.

From Table 5.4 it can be seen that the trade-to-trade¹⁷ approach and OLS yields approximately equal total MSEs, 0.162 and 0.165 respectively, when they are the estimators

¹⁶ Because there were sufficient data for three non-overlapping five-year periods, two MSEs for each estimator could be calculated. The results given here are the average MSEs for the two possible forecasts using the data available. The separate results of the two forecasts are presented in Table 5C.1 and Table 5C.2 in Appendix 5.C.

¹⁷ Note that the 'trade-to-trade' betas are based on weekly data, i.e. the price at the time of the last trade in the month is the recorded price at the end of the week in which the last trade occurred. This implies that while 'gross' thin trading will have been corrected for, the effect of the delay until the end of the week will still be present. In the simulation study in Chapter 4 it was shown that, relative to the size of the standard error of estimation, the remaining bias is small.

Table 5.4

MSE of one-period forecasts. The first period estimate of beta is used as a predictor for the second period estimate.

Beta estimator		Mean squared error				
First Period	Second Period	corr ^b	bias ^c	ineff ^d	rnd ^e	total ^f
OLS	OLS	.66	.002	.047	.115	.165
	adj OLS ^a	.57	.002	.054	.130	.187
	Cohen	.50	.025	.085	.147	.256
	T-T	.64	.004	.056	.109	.168
adj OLS	OLS	.60	.005	.068	.132	.206
	adj OLS	.58	.004	.072	.142	.218
	Cohen	.46	.018	.106	.158	.282
	T-T	.60	.004	.066	.123	.203
Cohen	OLS	.57	.030	.076	.142	.248
	adj OLS	.54	.021	.082	.154	.256
	Cohen	.45	.002	.111	.158	.271
	T-T	.56	.020	.086	.132	.237
T-T	OLS	.65	.004	.044	.118	.166
	adj OLS	.61	.002	.050	.132	.186
	Cohen	.50	.041	.079	.146	.244
	T-T	.66	.003	.051	.108	.162

^a OLS adjusted by expected bias.

^b The correlation coefficient between the estimates of beta in successive periods.

^c The squared difference between the average beta coefficients in successive periods.

^d The 'inefficiency' term from the Theil decomposition of the MSE.

^e The 'randomness' term from the Theil decomposition.

^f The sum of the bias, inefficiency and randomness terms. Note that the bias, inefficiency and randomness terms presented in this table may not sum to the total values. This is because the values here have been obtained from the average of the total MSEs from the two possible predictions and may differ from the sum of the individually averaged components due to rounding.

in both the first and second periods.¹⁸ Because of the similarity in total MSEs, one can ascertain that the stability of OLS betas and trade-to-trade betas are approximately the same.

Comparing the total MSEs in the last column of Table 5.4 for the three unbiased estimators (namely the trade-to-trade estimator, the Cohen estimator and the adjusted OLS estimator) it is evident that the trade-to-trade estimator not only has the lowest MSE when predicting trade-to-trade estimates of beta, but also the estimates of beta from the other two methods. Further, by examining the inefficiency and randomness components columns in Table 5.4 for the three unbiased estimators, the fact that the trade-to-trade estimator is the most

¹⁸ The fact that the OLS and trade-to-trade MSEs are about the same reflects the fact that although there is spurious stability in the OLS estimates of beta, the effect is counter-acted by the increased precision with which the trade-to-trade estimator estimates the betas of thinly traded securities.

efficient estimator can be noted empirically. The assessment of superior efficiency of the trade-to-trade estimator arises because inefficient estimation of beta coefficients of thinly traded securities will lead to spuriously increased cross-sectional variation in the estimated betas. The increased cross-sectional variation will cause large inefficiency and randomness components of the MSE, as can be seen from the definitions of the components in (5).

In fact it is clear from Table 5.4 that the Cohen¹⁹ and adjusted OLS methods are substantially less efficient than the trade-to-trade technique, with the Cohen method proving the least efficient of all four methods.

5.2.4.1 Further discussion of MSE results and a comparison with UK findings²⁰

The pattern of the results is very similar to that reported for UK stocks (see Table IV of Dimson and Marsh (1983)). The MSEs on the JSE are also very similar to those reported on the UK market. For example, the total MSEs for the OLS and trade-to-trade estimators on the UK market are reported as being 0.16 and 0.17 respectively which are similar to the values of 0.165 and 0.162 found on the JSE.

For all the estimators the bias (again, note that this is not the thin trading bias being referred to, but rather the difference in overall level of betas in successive periods) components reported in Table 5.4 are relatively small implying that the average estimated betas in successive periods are about the same. It is also clear that the random error component (which depends on the correlation between estimates of beta in successive periods and their cross-sectional variance in the second period) is generally the largest component of the MSE. The proportion of randomness is similar to that of the UK stocks (reported in Dimson and Marsh (1983)) and reflects not only similar correlation coefficients between periods, but also a similar spread of estimates of beta. For example, Dimson and Marsh report a correlation coefficient and randomness component of 0.47 and 0.12, respectively, for the forecast of trade-to-trade betas. The corresponding values for the JSE are fairly similar, being 0.66 and 0.11, respectively.

¹⁹ In this chapter the JSE-Actuaries Overall Index is assumed to be negligibly affected by thin trading since it comprises only those shares which trade relatively often, and only those shares which have a large market capitalization and so no leading coefficients were used.

²⁰ Summary statistics giving the mean beta estimates, their variances and ranges obtained from the four estimators are given in Appendix 5.G.

Because it is fairly difficult to interpret the values of the MSEs on their own, the conclusions that can be drawn are mostly comparative ones. The major results indicate that the *stability* of betas on the JSE (as measured by MSEs) is consistent with that in the UK market. On both markets, the MSEs indicate quite clearly that the trade-to-trade estimator is the most efficient of the unbiased estimators. The superiority of the trade-to-trade estimator is confirmed by comparing the lower forecast errors of the trade-to-trade estimator (when predicting the estimates of the betas produced by the other unbiased estimators) with the higher errors of the other unbiased estimators (when predicting the trade-to-trade betas).

In this section, the stability of betas was assessed using MSEs. The results indicated that the JSE and the UK market are similar with respect to their levels of stability. The assessment of beta stability relative to other stock exchanges is continued in the next section where transition matrices of estimates of beta are considered.

5.2.5 Transition matrices in the assessment of beta stability

While the correlation coefficients given in the tables above reflect a moderate degree of stability of the estimates, transition matrices enable more complete comparisons between the JSE and other markets.

Because of the superiority of the trade-to-trade estimator in terms of efficiency, the remaining investigations will be performed using only trade-to-trade estimates of beta. This decision is similar to that reached by Dimson and Marsh (1983) on the basis of their empirical findings.

The investigation into beta stability described in this section is based on transition matrices. In order to construct the transition matrices, the trade-to-trade betas from a period p (being either 1977-1981 or 1982-1986) were ranked in ascending order. The securities were then categorized into quintiles (with about 58 or 52 shares in each quintile for the periods 1977-1981 and 1982-1986, respectively). The proportion of securities with betas that fell into the corresponding quintiles in the next period, $p+1$, are given in the Table 5.5.

The three panels in Table 5.5 give the results separately for p being 1977-1981 and 1982-1986 as well as the average of the two transition matrices. Looking at the top panel of Table 5.5, for example, it is apparent that 56% of the securities which were in risk class 1 (i.e. the high beta securities) in 1982-1986 remained in that class in 1987-1991; 23% of those securities in class 1 in 1982-1986 changed to risk class 2 in 1987-1991; 12% changed even

Table 5.5

Transition matrices showing the movement of shares from risk classes in period p to risk classes in period $p+1$.

		Percentage in risk class in 1987 - 1991					•	Same or adjacent class
		1	2	3	4	5		
Risk class in the period 1982 to 1986	1	56	23	12	4	5		79
	2	21	28	21	22	9		69
	3	7	26	28	26	14		79
	4	5	17	24	28	26		78
	5	10	7	16	21	47		67
		Chi-square tests of		Association		107.50		
				Symmetry		5.352		
		Percentage in risk class in 1982 to 1986						
Risk class in the period 1977 to 1981	1	81	15	4	0	0		96
	2	19	28	23	11	19		70
	3	0	26	28	21	25		75
	4	0	13	32	36	19		87
	5	0	17	13	33	37		69
		Chi-square tests of		Association		189.02		
				Symmetry		14.774		
		Average percentage of shares in risk class in period $p+1$						
Risk class in period p	1	69	19	8	2	3		88
	2	20	28	22	17	14		70
	3	4	26	28	24	20		77
	4	3	15	28	32	23		83
	5	5	12	15	27	42		68

more and were allocated to the third quintile in 1987-1991, etc. A useful summary figure is given in the last column of Table 5.5, namely the percentage of securities which remained in the same risk class, or which moved only one risk class. (The numbers of securities in the same or adjacent class are highlighted in the transition matrices.) It is thus evident, for example, that 79% of the securities in risk class 1 in 1982-1986 were in quintiles 1 and 2 in 1987-1991. It can be read from the third panel (the average transition matrix) that between 70% and 80% of securities remain in the same or adjacent decile of systematic risk from one five-year period to the next.

Also given in Table 5.5 for each of the top two panels are the chi-squared tests of association between the number of betas in the risk classes in successive periods and a chi-squared statistic testing for a symmetric matrix.²¹ The chi-squared test²² of *association* yields values of 107.5 for the top panel and 189.02 in the second. Thus for both panels the test indicates a significant relationship between the betas in successive periods. The chi-squared test of *symmetry*, on the other hand, yields values of 5.352 and 14.774 which are not significant. The fact that there is no departure from the hypothesized symmetry suggests that there is no evidence of an overall trend in the betas between risk classes from one period to the next.

The results on the JSE show a marked similarity with the UK and US markets. The average proportion of securities on the JSE in the same (*same or adjacent*)²³ quintile is 39.8% (77.2%) compared with figures for the UK of 33% (72%) and for the US of 40% (79%) (reported by Dimson and Marsh (1983) and Sharpe and Cooper (1972), respectively). The corresponding figures for OLS estimates on the JSE are 43% (81%).²⁴

Unlike the results using MSEs in the previous section which showed the trade-to-trade and OLS estimates of betas to be equivalently stable, the transition matrices indicate that the downward bias in OLS betas does indeed increase the apparent stability of betas. It should also be noted that because there is still some small downward bias in the trade-to-trade estimates of beta owing to the fact that weekly (rather than daily) data had to be used, the assessments of stability may be slightly overstated. The results of Chapters 3 and 4 indicate only a marginal downward bias in the trade-to-trade betas when weekly data instead of daily data are used.

The results from the investigation into stability using transition matrices described in this section have confirmed that the stability of beta coefficients on the JSE is similar to that found on overseas markets. However, neither the analysis of transition matrices nor MSEs leads directly to a strict statistical test of significance in estimated beta coefficients. In the next section, therefore, the Chow (1960) test is used to test for stability in the trade-to-trade estimates of beta in successive five-year periods.

²¹ The expected numbers in row k , column l and row l and column k were taken as the average of the two observed numbers in these cells.

²² The 5% critical value for the chi-squared tests is 26.30.

²³ The percentage of securities in the same quintile is given outside the parentheses. Inside the parentheses, the percentage in the same or adjacent quintiles is reported.

²⁴ The OLS figures have been calculated in the same way as the trade-to-trade percentages, but are not given in Table 5.5.

5.2.6 Use of the Chow-Fisher test to test for beta stability

The inferences about the stability of beta coefficients in this chapter have thus far predominantly relied on comparing the measures of stability on the JSE with those from overseas markets. For instance, the use of transition matrices, or examination of MSEs is useful in that the measures describe the movements of beta, but they do not in themselves constitute a 'test' for stability. There have been several procedures developed which specifically test for statistically significant differences in regression coefficients through time. The Chow test (which Franklin Fisher (1970) later expounded and simplified) is one such test.²⁵

The test procedure is simply an application of generalised least squares, and essentially tests whether two separate regression through the two time periods (each of five years) fits the data significantly better than one regression through both time periods (i.e. all ten years) simultaneously. The test statistic itself has an F-distribution²⁶ and is calculated as:

$$F = \frac{(SSR_T - (SSR_1 + SSR_2)) / 2}{(SSR_1 + SSR_2) / (n_t - 4)} \quad (6)$$

where SSR_1 , SSR_2 and SSR_T are the sum of squared residuals from the regressions in the first period, the second period and both periods together, respectively, and n_t is the total number of trade-to-trade returns from both periods.

The Chow test was performed to test for significant differences between the parameters over the three five-year periods described previously, i.e. 1977-1981, 1982-1986 and 1987-1991. Only those shares which had traded every month over the relevant ten-year periods were included in the testing procedure. The numbers of securities for which the hypothesis of a stable beta was rejected are shown in Table 5.6.

From Table 5.6 it is clear that in the period 1982-1991, 10.9% (30) of the 275 securities displayed significant beta changes between the two five-year periods 1982-1986 and 1987-1991. The percentage of securities with significant changes in beta in the 1977-1986

²⁵ Surprisingly the Chow test does not seem to have been used all that often in the literature. Malkamaki's (1992) investigation of the Helsinki Stock Exchange and Carter's (1983) study on the JSE are among the few. Carter used the Chow test in an attempt to select the optimal number of intervals for beta estimation on the JSE rather than testing for beta stability directly.

²⁶ The distribution is of course dependent on the disturbance term in the market model being normally distributed. In a brief simulation study described in Appendix 5.E we show that the levels of non-normality on the JSE are such that the F-distribution is at least a reasonable approximation for the distribution of the test statistic.

Table 5.6

The number of significant changes in beta identified by the Chow test.

Period	Number of shares	Number of significant changes	Percentage significant changes
1982-1991	275	30	10.9%
1977-1986	241	9	3.7%

period is lower at 3.7%.²⁷ It is important to note that, because the tests of stability were all performed at a 5% significance level, it might be expected that in 5% of securities, the hypothesis of a stable beta may be *falsely* rejected. Comparing the expected false rejection rate (5%) with the actual percentages of rejections (the 10.9% and 3.7%) leads to the conclusion that betas are largely stable on the JSE.

However, it should be recognized that the Chow test (as applied in this case) may lead to substantial under-reporting of significant changes in the beta coefficients. Finding an insignificant change in the betas may, however, be attributed to one of three reasons. Firstly, there may not have been any significant changes in the underlying beta coefficients, in other words the Chow test is correctly signalling a stable beta. Secondly, the test may lack power, possibly because of inefficient parameter estimation. Thirdly, the test for stability has not included a search for the most likely change-points of beta in the data and has used rather arbitrary five-year periods.²⁸

The Chow test for detecting instability in regression coefficients was implemented on the JSE and indicated statistically significant instability for only a small percentage of securities. However, the power of the test was brought into question and therefore its results are to be treated with some caution.

²⁷ The higher stability of betas in the 1977-1986 period is also evident in Table 5.5 where the average percentage of securities in the same or adjacent period is 79% compared with 74% for the period 1982-1991.

²⁸ The five-year periods may, for instance, include several changes in systematic risk which cancel each other out.

5.2.7 Conclusions

This section has described an important empirical study of the betas on the JSE. It confirms for the JSE many of the findings about the stability of beta coefficients reported on overseas markets. Using MSE-analysis, transition matrices and the more formal Chow test, the betas on the JSE have been shown to be moderately stable over consecutive five-year periods. Furthermore, the magnitude of the stability appears to be very much in line with the UK and US markets.

An important consideration in testing for beta stability is that the biases in estimates of beta caused by thin trading could be possible sources of distortions in tests of beta stability. It was demonstrated that thin trading may lead to either spurious stability, or instability, in betas unless the estimators of beta were adjusted to correct for the biases caused by thin trading. To demonstrate that thin trading was indeed a problem on the JSE, the relationship between thin trading and unadjusted beta estimates was validated empirically. In addition to providing information on the stability of beta coefficients, the MSE-analysis allowed a comparison of estimators which avoided the biases caused by thin trading. The trade-to-trade estimator was confirmed to be the most efficient of the estimators considered.

In the next major section of this chapter an extension of the tests into beta stability is considered. It deals with a specific form of instability, namely the regression tendency of beta coefficients, i.e. the phenomenon observed on many stock markets where securities with high estimated betas in one period have lower estimated betas in the next period, and vice versa.

5.3 Regression bias and the regression tendencies of betas

5.3.1 Introduction

In section 5.2 the stability of beta coefficients on the JSE was examined in general terms. In this section the focus of the investigation is on a particular form of instability in beta coefficients. Dimson and Marsh (1983) and several other authors (e.g. Blume (1971)) have identified and examined the phenomenon that the high beta coefficients of one period tend to become less extreme in the next. The phenomenon has come to be known as the *regression tendency* of security beta coefficients. The explanations which have been suggested for the observed tendencies are, firstly, the *regression bias* in estimates of betas which are unavoidable

when using the market model and, secondly, a real drift (henceforth referred to as *true regression*) in the true, underlying betas towards some mean value.

This major section (i.e. section 5.3) encompasses an empirical investigation of the JSE designed, firstly, to establish whether the regression tendency of betas manifests itself and, secondly, to try to identify the relative sizes of the *regression bias* and the *true regression*.

In sub-section 5.3.2, an empirical study of betas on the JSE is conducted to ascertain whether there is a regression tendency of estimated betas on the JSE. Following the identification of the regression tendency, the next step is to implement a Bayesian adjustment factor to correct for the regression bias. The Bayesian adjustment is described and discussed in sub-section 5.3.3. In sub-section 5.3.4, the true regression of the betas after the regression bias has been removed is tested for significance.

5.3.2 Empirical evidence of the regression tendency of betas on the JSE

To complete the investigation into the stability of beta estimation and to permit further comparison with the overseas markets the regression tendencies of betas on the JSE are examined in this section. Following the methodology of Dimson and Marsh (1983), securities²⁹ from the JSE were arranged into quintiles on the basis of their ranked betas as computed for period p . The average betas for the quintiles of securities formed in period p were then computed for period p and period $p+1$.

The average betas in each of the quintiles are shown in Table 5.7. From Table 5.7 it is apparent that the regression of the betas towards a mean value is quite marked, i.e. the high beta stocks do, on average, become lower beta stocks, and the lower beta stocks become higher beta stocks. For example, the first quintile of securities in the period 1977-1981 (i.e. the 20% of securities with the highest estimated trade-to-trade beta coefficients in the period 1977-1981) has an average beta of 1.71. The average beta of the same securities in 1982-1986 is 1.41. The fifth quintile of securities from 1977-1981, by contrast, has an average beta coefficient of 0.16 in 1977-1981 which rises to 0.41 in 1982-1986. Very similar results are found when comparing the average betas of the quintiles formed in 1982-1986 with the average betas of those same securities in 1987-1991. For example, the first quintile beta decreases from 1.46 to 1.10, whereas the fifth quintile beta rises from 0.14 to 0.53.

²⁹ The same securities as those selected in section 5.2 were used.

Table 5.7

Average beta estimates in quintiles of estimation period (denoted by asterisk) and in subsequent period				
Quintile	77-81*	82-86	82-86*	87-91
1	1.71	1.41	1.46	1.10
2	.95	.73	.84	.80
3	.58	.50	.59	.68
4	.43	.46	.41	.56
5	.16	.41	.14	.53
Average	.76	.70	.73	.69

* The quintiles were reformed in these periods based on their ranked trade-to-trade betas.

It is evident from the results reported in this section that the regression tendency of portfolio betas exists on the JSE. In the next section a method of correcting for the *regression bias* component of the regression tendency is proposed and implemented.

5.3.3 Correcting for the regression bias using a Bayesian adjustment

It is well known that betas estimated from the market model suffer from *regression bias* (see Blume (1971, 1975)). This *regression bias* can be described as follows: an estimated beta coefficient which is far higher than the 'average' beta of all the other securities is more likely to be an overestimate of the true beta than an underestimate. Similarly, a very low estimated beta is more likely to be an underestimate. The asymmetry in the estimation errors involved with regressions (beta coefficients are invariably estimated using regression analysis) causes the usual trade-to-trade estimates to be inefficient. Thus the estimates of beta obtained from the regression analysis may be sub-optimal for forecasting purposes.

To correct for the regression bias, a 'Bayesian' adjustment of the form suggested by Vasicek (1973) can be implemented. The Vasicek estimator can be interpreted as a weighted average of the trade-to-trade estimate of beta and the average cross-sectional beta. The rationale underpinning the correction is that the extent to which the trade-to-trade beta is adjusted towards the overall mean depends on the relative sizes of the standard error of the estimate of beta and the cross-sectional standard deviation of the betas. For example, if a high trade-to-trade estimate of beta has a high standard error then the beta will be adjusted quite substantially towards the mean. If, on the other hand, the standard error is small relative to the overall

spread of all the betas (where the spread is measured by the cross-sectional standard deviation) then the adjustment will be quite small.

The Vasicek estimator has the form

$$\hat{\beta}_{bayes} = w \beta_{TT} + (1 - w) \bar{\beta}_{TT} \quad (7)$$

where

$$w = \frac{\sigma_{\beta}^2}{(\sigma_{\beta}^2 + \sigma_{TT}^2)} \quad (8)$$

where the following definitions hold

β_{TT} is the trade-to-trade estimate of beta;

$\bar{\beta}_{TT}$ is the average trade-to-trade beta of all the securities in the market;

σ_{TT} is the standard error of the trade-to-trade estimate of beta;

σ_{β} is the cross-sectional standard deviation of all the estimates of beta in the market.

It can readily be seen from equation (7) that the weight (w) assigned to the unadjusted trade-to-trade estimate of the beta will be large if the cross-sectional standard deviation (σ_{β}) is large. In other words, if the spread of the betas across the securities is so large as to make all the values of beta equally likely, then the usual trade-to-trade estimator is optimal. Conversely, if the standard error of the trade-to-trade estimate (σ_{TT}) is large relative to σ_{β} , then w will be small and the Vasicek estimate will be 'shrunk' towards the overall average beta of the securities in the market. The shrinkage arises from the assumed prior knowledge that the spread of betas around a mean value is limited. Therefore an estimate of beta which falls outside the usual spread of betas (and which has a large standard error) is likely to be an overestimate. The estimate must therefore be pulled in towards the mean so that it becomes a more likely estimate given the distribution of betas across the other securities.

Using an assessment based on MSEs, the prediction power of the adjusted trade-to-trade estimator can be evaluated. In an investigation similar to that described in 5.2, betas were estimated using a five-year period and were compared with the estimates in a subsequent period. The forecast MSEs were calculated between the two estimates of beta.

The decompositions (along the lines described in equation (5)) of the forecast MSEs³⁰ are presented in Table 5.8. The first estimator used in Table 5.8 is the overall average beta from the first period. In other words, the average beta from the first period was used as a predictor of the trade-to-trade betas of the second period. The total MSE for the historic average beta estimator is 0.205, which is higher than any of the other MSEs. The high MSE indicates that the historic average beta is a poor predictor. This finding stands in contrast to that of Dimson and Marsh (1983) on the UK market where they found the historic average beta to

Table 5.8

Prediction MSEs for trade-to-trade and Bayesian adjusted trade-to-trade betas

Estimate in period p	Estimate in period $p+1$	Corr. ^d	Mean Squared Error			
			Bias	Ineff	Random	Total
Historic beta ^a	TT	n/a	.003	.000	.202	.205
TT ^b	TT	.66	.003	.051	.108	.162
Adjusted ^c	TT	.65	.002	.023	.107	.132
Adjusted	Adjusted	.67	.002	.039	.062	.104

^a Average trade-to-trade beta in period p used as predictor.

^b Unadjusted trade-to-trade beta.

^c Bayesian adjusted trade-to-trade beta.

^d Correlation coefficient between estimate in period p and period $p+1$.

See Table 5D.1 and Table 5D.2 in Appendix 5.D for separate tables of each forecast period.

be marginally superior to the unadjusted trade-to-trade estimator. The MSE of the unadjusted trade-to-trade estimator is 0.162, which is substantially higher than the total MSE of 0.132 found for the Bayesian adjusted beta estimator. Focusing on the inefficiency column in Table 5.8 of the MSE, it can be seen that it is the improved efficiency of the Bayesian estimator which makes it a superior predictor. The inefficiency of the Bayesian adjusted estimator when predicting the trade-to-trade estimates in the next period is 0.023, which is under half the inefficiency when the unadjusted trade-to-trade estimates are used. The final row of Table 5.8 shows the decomposition of the Bayesian adjusted estimator when predicting the Bayesian adjusted estimates of the second period. The randomness component is seen to

³⁰ The values presented are the average of the values for the two forecasts possible. See Appendix 5.D for separate tables of each forecast.

decrease dramatically to 0.062, compared with 0.107 when the trade-to-trade estimate is used unadjusted in the second period. The decrease is caused by the reduced cross-sectional variance of the betas when the Bayesian adjustment is used.

In this section the vast improvements in efficiency offered by the Bayesian adjustment suggested by Vasicek (1973) have been demonstrated. The improvement is due to the reduced regression bias in the estimates of beta. In the next section, the issue of whether there is any *true regression* (i.e. any regression in the underlying betas after the regression biases in the estimates of beta have been removed) in the betas is examined, along with the possibility that the Vasicek estimator may over-correct for the regression bias.

5.3.4 Identifying true regression after regression bias has been removed

The dilemma as to whether there is true regression in the underlying betas and therefore as to whether or not the Vasicek procedure over- or under-corrects for the regression bias will be investigated in this section using the methodology described by Dimson and Marsh (1983). One way in which *true regression* can be detected is by tracking the changes in average quintile betas over two periods. Another way of detecting true regression is by measuring any regression tendency after the regression bias has been removed using the Bayesian adjustment.

To establish the extent of true regression in the manner discussed above, the securities were assigned to quintiles based on their ranked trade-to-trade betas calculated in the period 1977-1981. The average quintile betas in the two subsequent periods were examined. The average unadjusted trade-to-trade betas in each quintile are shown in the unshaded columns of Table 5.9 for each of the three periods. The average quintile betas display a quite marked regression tendency. The range of average quintile betas from the most risky to the least risky in 1977-1981 (the period in which the quintiles were formed) spans the values 1.71 to 0.16. In period 1982-1986, the range narrows to span the values 1.41 to 0.41. Keeping the quintile constituents the same in the second subsequent period, the range of average betas narrows even further to span 1.07 to 0.66. On its own, the continued regression tendency of the quintile average betas over two periods would indicate that true regression exists.

As indicated earlier, the shaded column of Table 5.9 reports the average quintile betas after the adjustment using the Bayesian procedure suggested by Vasicek (1973). The Bayesian

Table 5.9

Average quintile betas over successive period. Quintiles were formed in 1977-1981 and the average beta of the securities in that quintile were calculated in the two subsequent periods.

Average betas in quintiles based on ranked traded-to-trade betas calculated in 1977-1981				
Quintile	77-81	Bayesian 77-81	82-86	87-91
1	1.71	1.54	1.41	1.07
2	.95	.92	.73	.84
3	.58	.61	.50	.64
4	.43	.46	.46	.57
5	.16	.25	.41	.66

correction does indeed appear to 'pull in' extreme estimates of beta, as predicted.³¹ The range of quintile betas spans 1.54 to 0.25, compared with 1.71 to 0.16 which is the case for the unadjusted averages. If it is assumed that the Bayesian adjustment *correctly* adjusts for the *regression bias*, then the fact that the range of quintile betas in the first subsequent period (1982-1986) narrows more than the Bayesian adjustment (i.e. from 1.54 - 0.25 to 1.41 - 0.41), also suggests that there is a *true regression* in the underlying betas.

However, as pointed out by Dimson and Marsh (1983), the results are also consistent with two alternative hypotheses. The first of these is that the Bayesian adjustment over-corrects for regression bias and the true regression is masked by the 'pulling in' of the Bayesian adjustment. The second hypothesis is that the Bayesian adjustment is under-correcting and there is negligible true regression. In this section the methodology suggested by Dimson and Marsh (1983) is implemented to choose between the hypotheses.

Following Dimson and Marsh (1983), the rate at which betas regress to their mean was calculated from the estimation period (1977-1981) to the first and second subsequent periods by fitting the model:

$$(\beta_{jq} - \bar{\beta}_{jq}) = \lambda (\beta_{jp}^* - \bar{\beta}_{jp}^*) + u_{jp} \quad (9)$$

Both the dependent and independent variables in equation (9) are expressed as differences from their mean values. Dimson and Marsh (1983) argue that using the differences between the betas

³¹ The difference is not as marked as Dimson and Marsh's findings for the UK stocks, where, for instance in their fifth quintile the unadjusted average beta is 0.18, while the adjusted value is 0.60.

and their means removes any confounding caused by shifts in the mean betas between periods. The *rate of regression* for the estimated betas is given by the λ parameter in equation (9). A value of λ close to unity, for example, would indicate very little regression in the estimated betas because the difference between the beta and its mean in period p would be, on average, the same as the difference between the beta and its mean in period q .

The rate of regression (λ) in the estimates of beta between the estimation period (p) and the first and second subsequent periods can be estimated using equation (9) with q set equal to $p+1$ and $p+2$. The relative sizes of the two estimated rates of regression allow the *true rate of regression* in the underlying betas to be computed. If there is only a *regression bias* in the trade-to-trade estimates of beta (and no *true regression* of the underlying betas), then the λ (estimated rate of regression) parameter should be the same for both $q = p+1$ and $q = p+2$, i.e. the rate of regression should be the same for both the first and second subsequent periods.³² In fact, the ratio of the rate of regression of the *estimated* betas in the first period to the rate of regression of the *estimated* betas in the second subsequent period is an estimator of the *true regression*. If the rates of regression in the estimated betas are equal, for instance, the estimated rate of true regression will be unity, i.e. no true regression.

The model in equation (9) was fitted four times in order to help pin point the extent of the true regression. The dependent variable ($\beta_{j,q}$) was set equal to the unadjusted trade-to-trade estimates in the first and second subsequent periods, i.e. q was set equal to $p+1$ and $p+2$. The independent variables ($\beta_{j,p}^*$) were taken for each value of q to be, firstly, the trade-to-trade and, secondly, the Bayesian adjusted trade-to-trade beta in the estimation period p .

The issue as to whether the Bayesian adjustment over- or under-corrects can be gauged by the difference between the rate of regression in the estimated betas after the Bayesian adjustment and estimated *true regression*.

The rates of regression for the estimated betas (slope coefficients from equation (9)) are set out in Table 5.10. From Table 5.10 the slope coefficient in the unadjusted betas from the estimation period to the first subsequent period is 0.677. The value of 0.677 reflects both the *regression bias* and *true regression* of the trade-to-trade betas from the period 1977-1981 to the next five-year period. As argued above, the decrease in slope from 0.677 in the first subsequent period to 0.309 in the second subsequent period (87-91) suggests that there is true

³² The equality of the two rates of regression when there is no true regression stems from the fact that, on average, if there is no movement in the actual betas then both the estimates in the first and second subsequent periods will be unbiased estimates of the true beta.

Table 5.10

Estimation of the rate of regression of betas towards their mean.

Dependent variable (in period q)	Independent variable (betas marked with an asterisk, in period p)	slope ^a	t-statistic	R ²
T-T (82-86)	T-T (77-81)	.677	20.49	62
	Adjusted (77-81)	.819	20.74	62
T-T (87-91)	T-T (77-81)	.309	7.42	17
	Adjusted (77-81)	.372	7.42	17

^a The value of lambda, the rate of regression of the estimates of beta, from equation (9).

regression in the underlying true betas, i.e. the true, underlying portfolio betas are indeed changing. From Table 5.10 it is also evident that the rate of regression for the Bayesian-adjusted estimates of beta is 0.819, which, being less than unity, indicates that there is still regression towards the mean after the adjustment. The rate of regression after the Bayesian adjustment when predicting the beta in the second subsequent period drops to 0.372 from the 0.819, which further confirms the true regression in the beta coefficients.

The information in Table 5.10 also helps in deciding which of the hypotheses about the over- or under-correction of the Bayesian adjustment is most substantiated by the data. As indicated earlier, the ratio of the second rate of regression in the estimated betas to the first rate of regression of the estimated betas is an estimate of the true regression rate. The true regression rate is thus estimated as 0.45 for both the Bayesian estimator (with a ratio of rates of regression equal to $0.372/0.819$) and unadjusted estimators (with a ratio of rates of regression equal to $0.309/0.677$). The evidence from Table 5.10 therefore indicates that the Bayesian adjustment appears to over-adjust for regression bias, i.e. after the adjustment there is a slower *apparent* rate of regression (i.e. 0.819) than there is *true* regression (i.e. 0.45). The over-correction of the Bayesian adjustment means that the predictive power of the Bayesian estimator is improved because it incorporates some of the true movements in the security betas, as well as correcting for the regression bias.

Because only three periods were available (due to data limitations) to investigate the regression of betas only one estimate of the true regression could be obtained as only one 'second subsequent period' was ever available. Nevertheless, two estimates of the rate of regression from one period to the first subsequent period, viz 1977-1981 to 1982-1986 and 1982-1986 to 1987-1991, could be computed. The first estimate of the rate of regression after

Bayesian adjustment is given in Table 5.10 as 0.819. The second estimate based on the rate of regression in the Bayesian betas between the estimation period 1982-1986 and the first subsequent period (1987-1991) is 0.519. The value of 0.519 represents a rate of regression after the correction for regression bias which is much closer to the true regression and suggests that the Bayesian estimator does *not* over-correct significantly when compared with a *true regression* of 0.45.

The conclusion reached in this section is that there is some evidence that the Bayesian adjustment factor does more than just correct for the regression bias. In fact, by over-correcting for the regression bias it also partially adjusts for the *true regression* of beta coefficients.

5.3.5 Conclusion

This section completes the major investigations into stability of beta coefficients on the JSE using thin trading correction procedures. In it, the form of instability which has become known as the *regression tendency* of portfolio betas was investigated. The most significant finding from this section has been the substantial additional efficiency offered by the Bayesian adjustment procedure suggested by Vasicek (1973). The Bayesian adjustment corrects for the inevitable *regression bias* in trade-to-trade estimates of beta and thereby significantly improves its predictive power.

Another interesting result which emerged from the study is that there is evidence of a *true regression* of portfolio betas towards some mean value. The true regression in the underlying betas is an additional effect which compounds the apparent instability of estimated betas caused by the regression bias which stems from the use of regression analysis to obtain beta estimates.

5.4 Some further tests and suggestions for improving the efficiency of beta estimators

In sections 5.2 and 5.3 the stability of beta coefficients was discussed and the trade-to-trade estimator was confirmed as the most efficient, unbiased estimator of systematic risk. In addition, conclusive evidence supporting the use of a Bayesian adjustment was found. In this section two proposals for enhancing the efficiency of the trade-to-trade estimator are investigated. The first method (discussed in 5.4.1) involves weighting the observations so as

to give more influence to those months in which there were large trading *volumes*. The concept of weighting the observations is associated with the hypothesis that trades involving large volumes of shares are more accurately priced³³ than 'marginal' trades. Motivation for such an hypothesis would stem from a view that because the trades associated with large volumes are more likely to be made at prices close to the *true value* of the shares, assigning those observations more weight will improve the efficiency of the estimator.

The second method researched in this section is the use of shorter or longer interval lengths. In section 5.4.2 the studies of the stability and relative efficiencies resulting from different interval lengths are examined in the context of the JSE.

The beta estimators used in both sections include Vasicek's (1973) correction which was discussed in section 5.3 and found to improve efficiency.

5.4.1 Weighting of observations

As mentioned above, some practitioners may hold the view that large volume trades are backed by more careful information analysis and so will be made at a more 'accurate' price, implies in statistical terms that there will be less random error associated with the observation. Equivalently, the effect can be thought of as a correlation between the random error and the volume of shares traded in the security. Because the optimal efficiency property of least squares regression³⁴ estimates depends on *constant* variance in the observations, a significant correlation between volume and variance suggests that the trade-to-trade estimator will be sub-optimal. To avoid the heteroscedasticity the observations must be weighted so as to remove the correlation between variance and volume.

The volumes can be converted into weighting functions in several different ways. The two functions considered in this investigation were (a) the square root of the volume traded in

³³ The more accurate pricing may arise from the increased research which traders in large volumes of stock may put into finding a precise value.

³⁴ The trade-to-trade estimator is a least squares estimator based on trade-to-trade returns which have been adjusted to correct for the changing variance in the observations caused by the varying interval lengths between the last traded days in successive months.

the week and (b) the square root of the empirical distribution function³⁵ of the volume of trade in the week.³⁶

The average standard deviations of the estimates of beta are summarised in Table 5.11. Inspection of Table 5.11 reveals that the weighting schemes appear to *increase* the average standard deviation of the estimate. For example, in the 1977-1981 period, the unweighted estimate of the standard error is 0.2244, whereas the standard errors for the two weighting schemes are 0.2310 (for weighting scheme (b)) and 0.2583 (for weighting scheme (a)). Looking at the final two columns of Table 5.11, in all three periods, the more robust weighting scheme (scheme (b)) outperforms scheme (a) on the basis of small standard errors confirming that the

Table 5.11

Cross-sectional average standard errors of the trade-to-trade beta estimates using different weighting schemes.

Period	Unweighted	Weighted	
		By Distribution Function	By Volume
1977-1981	.2244	.2310	.2583
1982-1986	.2175	.2240	.2356
1987-1991	.1937	.2057	.2292

smaller the effect of weighting scheme, the more efficient the estimator.

The weighting schemes proposed in this section do not appear to improve the precision of beta estimators. In other words, there appears to be very little discrepancy between the accuracy of pricing of smaller trades compared with larger trades. Hence the above-mentioned scheme is not recommended for practical use. It would appear that the view that larger volumes are associated with more accurate prices is invalid in the context of beta estimation.

In the next section, the empirical effect of using longer and shorter interval lengths than the monthly returns used in the previous sections is investigated on the JSE.

³⁵ Calculated by finding the rank of the volume corresponding to the return under consideration relative to the volumes in the rest of the sample.

³⁶ The two proposed functions arose from the following reasoning. Firstly, it was assumed that the precision of the pricing was directly proportional to the volume. On this assumption, the variance of the random term would be proportional to the reciprocal of the volume. Weighting all the observations by the square root of the volumes would therefore eliminate the heteroscedasticity. The second method is just a more robust version of the first. For some securities there are one or two observations with extremely large volumes. Weighting the observations by the square root of these very large values would mean that many of the observations would essentially be ignored. Therefore the *rank* of the volume was used as a robust measure of the actual volume.

5.4.2 Alternative interval lengths in estimating beta

In this section, the effect of using interval lengths other than monthly on the JSE are investigated. From the market model it might appear that more efficient estimates of beta can be obtained by using shorter intervals within a given estimation period. For example, using weekly data over five years would give us approximately four times more observations and a standard error of about a half of that obtained by using monthly observations. Any improvement gained by using weekly return data, for example, is dependent on the relative variability of weekly residuals being about a quarter of monthly variance.

Using shorter return interval lengths (henceforth referred to as 'interval lengths') appears on the surface to be an alternative to lengthening the overall period of estimation and maintaining monthly intervals. One possible advantage of using shorter interval lengths is that assumption of stability in the true betas does not have to be extended over increasingly long periods of time. On the other hand, using shorter interval lengths may lead to other effects. For example, the price adjustment delays other than thin trading may bias even trade-to-trade estimates of beta. Alternatively, the residual variance may not decrease in line with the shorter intervals which would make the use of shorter intervals less efficient.

In this study the effects of varying the interval length are investigated by estimating trade-to-trade betas in the three five-year periods which correspond to those used in the studies in 5.2 and 5.3, i.e. 1977-1981, 1982-1986 and 1987-1991. The mean and average standard deviation of the estimated betas were calculated. Furthermore, the stability of the estimates based on different interval lengths were assessed using MSEs.

A potential drawback in this investigation is that only weekly data were available for the three periods. The problem arises because, as shown in Chapter 3, the bias in trade-to-trade estimates of betas caused by the thin trading *within* a week increases as the interval length is decreased.

Because of the increased biases involved with shorter interval lengths the results must be interpreted with some caution. Table 5.12 gives the average betas for various interval lengths. The results in Table 5.12 demonstrate that the average betas do decrease with shorter interval lengths. For example, in the 1987-1991 period, the average weekly beta is 0.64

Table 5.12

Average estimates and standard errors at varying interval lengths

	Average beta			Average standard error of estimate		
	77-81	82-86	87-91	77-81	82-86	87-91
8-weekly	.88	.72	.71	.281	.235	.305
monthly	.78	.70	.75	.224	.218	.194
biweekly	.75	.62	.69	.184	.157	.161
weekly	.67	.57	.64	.141	.121	.125

whereas the average 8-weekly estimate is 0.71.³⁷ This effect was to some extent anticipated and confirms that for thinly traded securities, the use of interval lengths shorter than monthly is not recommended.

In comparing the average betas in Table 5.12 calculated using monthly data in the three periods (viz. 0.78, 0.70 and 0.75) and those calculated using eight-weekly data (0.88, 0.72 and 0.71) it appears that there is no consistent difference in the average betas. The conclusion drawn is that there is no more bias in monthly estimates of beta than eight-weekly estimates.

Focusing on the right hand side of Table 5.12, it appears as if the average standard errors of the estimates also decreases in almost exactly the expected manner; for example, the average standard error of the eight-weekly estimates in 1991, 0.305, is approximately double the average standard error for the biweekly estimates, 0.161, of which there are about four times as many observations for each estimate.

Based on the summary results given in Table 5.12, it might be expected that shorter interval lengths will appear more stable when measured in terms of MSEs. The increased apparent stability is due not only to improved efficiency in the estimation, but also to the spurious stability caused by increased biases in the estimates of beta. To confirm the expectations about stability, forecast MSEs were examined empirically using similar methodology and data to those measures of stability described in 5.2.

In Table 5.13 the forecast MSEs for the estimates based on varying interval lengths are given. Looking down the far right-hand column of Table 5.13, the total MSEs can be seen to be smaller for shorter intervals in both forecasts. For example, in predicting the estimates of beta in 1987-1991 using the estimation period 1982-1986, the weekly interval lengths yield an

³⁷ The so-called 'intervalling effect' on estimates of beta has been noted in many previous studies on other exchanges where thin trading impacts on the estimation technique; see, for instance, Pogue and Solnik (1974) or Corhay (1992). If OLS estimates of beta were calculated the effect would be even more noticeable.

Table 5.13

MSE for varying interval lengths

	Corr	Bias	Inefficiency	Random	Total
Estimation: 1977-1981 Prediction: 1982-1986					
Weekly	.88	.006	.003	.057	.066
Biweekly	.81	.012	.003	.102	.117
Monthly	.78	.004	.033	.090	.127
8-weekly	.64	.018	.052	.172	.242
Estimation: 1982-1986 Prediction: 1987-1991					
Weekly	.78	.005	.026	.074	.105
Biweekly	.68	.005	.063	.085	.156
Monthly	.51	.002	.068	.126	.196
8-weekly	.28	.000	.164	.177	.341

MSE of 0.105. The monthly MSE is larger at 0.196 and the eight-weekly MSE larger still at 0.341. Again, while it appears as if the increased number of intervals lowers the forecast MSE, it should be remembered that the effects of thin trading are more pronounced at shorter interval lengths and hence also the spurious stability associated with it.

The tentative conclusion of this section is that although shortening the interval length permits improved efficiency of estimation, the bias which remains when using weekly or biweekly data, however, implies that it is still more satisfactory to use monthly estimates.

The major source of bias is undoubtedly thin trading on the JSE. Where markets are well traded (e.g. NYSE), any small remaining bias is caused by other price adjustment delays. Consequently a Cohen estimator may be useful. The last section is somewhat out of place in that it does not necessarily build on any existing development. However, for the sake of completeness and for researchers on other markets, it presents some interesting new insights into the estimation of the Cohen estimators variance.

5.5 Bootstrap estimation of the variance of the Cohen estimator

As discussed in all the previous chapters, thin trading and other price adjustment frictions plague the accurate and efficient estimation of risk statistics on most smaller stock exchanges and even other better traded markets.

The trade-to-trade method of estimating risk statistics removes the effects of infrequent trading (which accounts for most of the bias) but leaves estimates susceptible to a relatively smaller bias from other delays in the price adjustment process (see, e.g., Cohen *et al* (1980, 1983)). On the other hand, Cohen estimators use non-synchronous coefficients to capture *all* relevant correlation between the market and a share. The Cohen estimator is therefore consistent but its small-sample variance is difficult to derive except under very restrictive assumptions (see Dimson (1979)). Methods which have been used to get an idea of the relative size of the sampling error include simulation studies but these offer no practical procedure for estimating *empirical* efficiencies.

A demonstration of how to compute the variance of the Cohen estimator using the computer intensive technique known as *bootstrapping* is given in this section. By using actual data no assumptions about the structure of inefficiencies have to be made. Hence a realistic and practically useful comparison can be made of the bootstrap estimates of the standard error with those obtained using the simplifying assumptions.

In the next sub-section, the key features of the bootstrapping technique are briefly outlined.

5.5.1 Bootstrapping

The technique of bootstrapping (Efron (1982)) has been widely used by statisticians and is a technique suited to econometric applications where the distribution of some sample statistic is unknown. The bootstrap is similar to the much older jackknife method in that it is a resampling plan. More specifically, 'sub-samples' are drawn from the original, single sample which the researcher has collected. From the subsamples repeated measures are made of the statistic of interest. The repeated measures (replications) of the statistic can then be used to infer results about the distribution of the statistic. Its use has not been widespread in the field of risk estimation. One exception is Hui *et al* (1990) who suggested its use in an attempt to

improve beta estimation in thinly traded markets (although their reasoning behind the use of the bootstrap in such a context is not clear).

The risk measure researchers and practitioners in financial economics are most often interested in is the systematic risk or beta of a share. Estimates are usually based on the market model, which is again set out below to facilitate the discussion of the bootstrap methodology. The market model corresponds exactly to a simple linear regression model which can be written as:

$$y_i = \alpha + \beta x_i + \epsilon_i \quad (10)$$

where y_i , x_i and ϵ_i are the return on the share, the return on the market and a disturbance term measured over an interval i , where i runs from 1 to n . Estimates of the regression parameters are then made by minimizing some measure of the distance between the vector of observed share returns, y , and the vector of predictors $\eta(\alpha, \beta)$, which in the case of the market model are $\eta(\alpha, \beta) = [\alpha + \beta x_i]$. The most common measure of distance is the squared difference which results in OLS estimates. The disturbances, ϵ_i , are assumed to come from some common unknown distribution F . The appropriate bootstrap algorithm (which will be referred to as the 'proper' bootstrap to distinguish it from the 'cautious' version mentioned later) is then:

1. Construct $\hat{F}$ by putting a mass of $1/n$ at each estimated residual, $e_i = y_i - \hat{\alpha} - \hat{\beta} x_i$. ($\hat{F}$ represents the empirical distribution function which, it is assumed, forms a good approximation to the true function, F .)
2. Draw a bootstrap data set by selecting at random n estimated residuals (with repeats) (which is denoted by e_i^*) from $\hat{F}$ and then put $y_i^* = \hat{\alpha} + \hat{\beta} x_i + e_i^*$. For the bootstrap data set $(y_i^*; x_i)$ calculate the parameters which minimize the distance between the y_i^* 's and the predictors and denote the parameters thus obtained $\hat{\beta}^*$ and $\hat{\alpha}^*$.
3. Repeat step 2, B times (where B is a large number, say 1000) and hence obtain B replications of the parameters.
4. The estimate of the standard error of $\hat{\beta}$ can be taken as

$$se(\beta) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\beta}^{*b} - \hat{\beta}^{**})^2 \quad (11)$$

where $\hat{\beta}^{*b}$ denotes the b th replication and $\hat{\beta}^{**}$ the average value of the bootstrap betas.

An alternative procedure which is more 'cautious' in its assumption about $\hat{F}$ being a good approximation to F is to draw B random samples of size n from the original data pairs

(x_i, y_i) . In other words, in the more 'cautious' estimator, the n observations in the bootstrap sub-samples are obtained by selecting, at random, n of the original observations (allowing the same original observation to be selected more than once). The 'cautious' bootstrap differs from the 'proper' one in that in the 'proper' version, the same x_i values are used in *all* n sub-samples, whereas in the 'cautious' version, different x_i values are selected for each sub-sample.

Before going on to describe the empirical study conducted on the JSE, it is worth recalling briefly the features of the Cohen estimator which necessitate special modifications in the bootstrapping methodology.

5.5.2 A brief recap of the Cohen estimator

The Cohen estimator is an aggregated coefficients technique designed to correct for the biases in estimates of beta caused by thin trading. The key features of the Cohen estimator are the lagged and leading coefficients which are aggregated to obtain the estimate. If the number of leading intervals to be used in the estimator is denoted by l_1 and the number of lagged coefficients by l_2 , then the estimator can be written as:

$$\hat{\beta} = \frac{\sum_{l=-l_2}^{l_1} \beta^{(l)}}{\sum_{l=-l_2}^{l_1} \rho_m^{(l)}} \quad (12)$$

where

$$\beta^{(l)} = \frac{\text{Cov}[y_i; x_{i+l}]}{\text{Var}[x_{i+l}]} \quad (13)$$

and

$$\rho_m^{(l)} = \frac{\text{Cov}[x_{i+l}; x_i]}{\text{Var}[x_i]} \quad (14)$$

Cohen *et al* originally specified that l_1 and l_2 should be equal. However as the number of lags plus leads increases so the estimator becomes more inefficient. The compromise solution to the dilemma is to use a slightly biased version of the Cohen estimator which will be more efficient. The compromise is effected by selecting different numbers of lagged and leading coefficients. Clearly it is difficult to obtain the small sample distribution of such an estimator.

Under simplifying assumptions (including zero serial correlation in the market returns) it can be argued that the variance of the regression parameter can be calculated as the sum of the individual variances in the $(l_1 + l_2 + 1)$ individual regressions. The estimate of standard error obtained from adding the individual variances is henceforth referred to as the 'naive' estimator of standard error. The 'naive' estimator may well be a conservative measure of the efficiency and therefore lead to the unwarranted rejection of the use of the Cohen estimators in empirical studies. The objective of this study is to gauge how good an estimator of efficiency the 'naive' estimator is.

The use of the first mentioned ('proper') bootstrap procedure involves no further special considerations when applied to the Cohen estimator and the bootstrap algorithm set out earlier can be used. The second (more 'cautious') procedure does, however, necessitate further work. The vital aspect of the Cohen estimator is its use of consecutive returns on the market. Therefore the procedure of selecting random pairs of security and market returns needs to be extended so that the sample comprises sets of *consecutive* pairs. For example, if one lagged coefficient and no leading coefficients are being used, each item in the sub-samples must consist of two consecutive pairs of points.³⁸

In the next section the empirical investigation conducted using securities from the JSE is described. The purpose of the investigation is to compare the 'naive' and bootstrap estimates of the Cohen estimator.

5.5.3 Empirical investigation to compare the simple and bootstrap estimates of the beta inefficiency

In the empirical study to assess the appropriateness of the 'naive' standard error by comparing it with the bootstrap estimator, a set of securities from the JSE were needed. Initially only those shares which had been listed for at least 240 weeks as at 8 July 1992 (529 shares) were included. The restriction was implemented so that estimates of q (the proportion of intervals not traded) could be based on the daily data available to us. The betas and bootstrap variances were then computed as at 31 December 1981, 1986 and 1991. However, the restriction in the share selection meant there was a survivorship bias in the selection of shares.

³⁸ The number of lags used is chosen using the adaptive procedure described in Chapter 4, and in section 5.2.3 of this chapter.

The study was therefore repeated using all the shares available at each of the three dates and q was computed from weekly data.³⁹

Since the results for both 'proper' and 'cautious' bootstrap methodologies and both for the 'unbiased' and 'biased' sets of data were found to be similar, only the figures from the 'proper' bootstrap procedure applied to the unbiased data set are presented in Table 5.14. The results from the other investigations can be found in Appendix 5.H.

Focusing attention of the top two rows of Table 5.14 it is apparent that the 'naive'

Table 5.14

Comparison of bootstrap and 'naive' estimates of the standard errors in the Cohen beta estimator.

	1977-1981	1982-1986	1987-1991	Average ^f
Bootstrap estimate ^a	.311	.268	.429	.363
'Naive' estimate ^b	.346	.301	.450	.390
Average actual Cohen beta	.862	.811	.720	.775
'Bias' ^c	.001	.000	.000	.000
Percentage of shares where bootstrap is lower than 'naive' estimate ^d	95.0	93.4	80.3	86.9
Number of shares in sample ^e	297	320	688	

^a Bootstrap estimate of standard error of Cohen estimate of beta.

^b Square root of the squared standard errors of the β^0 's from equation (4).

^c The difference between the average bootstrap estimates of beta and the actual Cohen estimates.

^d Percentage of the shares in each period for which the bootstrap estimate of standard error of the Cohen estimate of beta is lower than (or equal to) the 'naive' estimate.

^e The number of shares available in the period with sufficient data (at least 30 months) to calculate a Cohen estimate of beta.

^f Weighted average, with weights equal to the number of shares used in the period.

estimator of standard error is a fairly good approximation to the actual standard error (the bootstrap estimate). In each period the average 'naive' estimate is slightly larger (i.e. more conservative) than the bootstrap estimate. In the last two columns of Table 5.14 it can be seen that, on average, the bootstrap estimate of the variance was 0.363 and the 'naive' estimate was 0.390, a difference which represents an average overestimate of about 7% of the standard error.

A further useful statistic reported in Table 5.14 is the percentage of securities for which the bootstrap standard error was smaller than the 'naive' standard error. The average percentage given for the three periods is approximately 87%. It can therefore be inferred that

³⁹ For the earlier dates (1981 and 1986) daily data were not available and hence weekly data had to be substituted with the appropriate adjustments for selecting a suitable number of leads and lags for the Cohen estimator.

the 'naive' estimator is mildly conservative for the majority of the securities, and can be used as a lower bound of the efficiency of the Cohen estimator of betas. From a theoretical standpoint, the appropriateness of the 'naive' standard error depends on assumptions of there being no serial correlation in the monthly market returns and negligible price adjustment frictions in addition to infrequent trading. The empirically observed difference in the 'naive' and bootstrap standard errors indicates that the assumptions are violated to a small extent. Further investigations into the existence of serial autocorrelation in the monthly market returns and the effects that this has on Cohen estimators are given in Chapter 8.

It should be noted that the 'bias' referred to in Table 5.14 is the average (across all shares in the period) difference between the mean estimated beta from the (1000) bootstrap samples and the beta estimated from the original sample. The bias statistic summarises the extent of any bias in the Cohen procedure. It should be close to zero given the construction of the Cohen estimator, and the results from Table 5.14 indicate that this is indeed the case.

5.5.4 Conclusions

The results clearly demonstrate that the 'naive' estimate of the standard error of the Cohen estimator is relatively good, even though it is mildly conservative (i.e. too large). It appears from the results that the added computational burden associated with calculating the bootstrap estimates does not yield a significant improvement in the accuracy of the final estimate of the standard error. A practical consequence of the findings in this section is therefore that if the Cohen estimator is used (for whatever reason) and a measure of the standard deviation is required, then an adjustment downwards of, say, 7% in the 'naive' estimate of the standard error is recommended. However, the fact remains that the Cohen estimator proves unsatisfactory for individual security beta estimation on the JSE because of its inefficiency.

The second important finding in this section is that the bootstrap has shown itself to be a valuable tool in assessing the feasibility of simplifying assumptions about the distributions of econometric variables. In particular, the inefficiency of the Cohen estimator (which has long been evident only through simulation studies and in stability assessments of betas) can be directly assessed by the bootstrap methodology.

5.6 Summary and conclusion of Chapter 5

This chapter comprised four major sections. The results presented in the first section (5.2) contained several important implications for the selection of a ‘best’ beta estimator. First of all, the results verified *empirically* the findings of the *simulation* study in Chapter 4: specifically, that OLS estimates of beta coefficients of beta are biased by thin trading and, secondly, that the *trade-to-trade* estimator is the most efficient of the unbiased estimators of beta. The second part of the investigation in section 5.2 examined the stability of beta coefficients. It was shown the stability of beta coefficients on the JSE was similar to that of the UK and US, once the distortions in measures of stability caused by thin trading were removed.

In the second section (5.3) the regression tendencies of estimated portfolio betas were probed. It was demonstrated that the regression tendency of portfolio betas was evident on the JSE and that the regression tendency could be separated into two components, namely regression bias in the estimates of the betas, and true regression of the underlying betas. In order to correct for the regression bias in the estimates of the betas so that the true regression of the betas could be quantified, a *Bayesian adjustment* was implemented. The Bayesian adjustment did indeed correct for the regression bias and, in fact, markedly improved the efficiency of estimates of beta. Additionally, the application of the Bayesian adjustment revealed that there was a true regression in the portfolio betas, similar to the regression found in the UK.

Thirdly, section 5.4, contained two short empirical studies which investigated two possible modifications of the Bayesian adjusted trade-to-trade estimator. The first modification was based on the view possibly held by some practitioners that trades involving large volumes of shares might be more accurately priced. The validity of the view was empirically assessed by weighting the observations with a function of the trading volumes. However, the results indicated that no improvements in efficiency over and above that of the unweighted estimator were obtained. The second investigation evaluated the impact on estimates of beta when shorter return intervals were used. The results indicated that, in the case of the JSE, biases caused by thin trading in the estimates of beta mitigated the improved efficiency that the shorter intervals provided. Monthly intervals emerged as the best compromise between unbiasedness and efficiency.

The final section, 5.5, demonstrated that the ‘naive’ estimate of the variance of the Cohen estimator (obtained by adding the variances of the leading and lagged coefficients) was

a reasonable estimate of the true standard error. The bootstrap methodology proved itself very useful in helping to assess the quality of financial statistics for which the small sample distributions are not known.

In sum, the major recommendations that emerged from this chapter with respect to beta estimation are the use of the trade-to-trade estimator (to remove the biases caused by thin trading) and the Bayesian adjustment proposed by Vasicek (1973) (which corrects for the regression bias in the estimates of beta).

Appendix 5.A**Relationship between q_{jp} and μ_I and Ψ .**

If q is the probability that the share does not trade in a sub-interval and there are k sub-intervals in the interval used for estimation then:

$$\Psi = (1 - q^k) \quad (15)$$

$$k \mu_I = \frac{q}{1 - q} \frac{1 - q^{k-1}}{1 - q^k} - (k-1) \frac{q^k}{1 - q^k} \quad (16)$$

Appendix 5.B

Average transition matrix of thin trading parameters

Table 5B.1

Average transition matrix for expected bias in OLS betas.

		% bias in OLS estimate in year $p+1$										
		5	10	15	20	25	30	35	40	45	...	85
% bias in OLS est. in year p	5	101	19	2							...	
	10	14	18	8	3	1						
	15	2	6	8	4	2		1				
	20		2	4	4	3	1	1				
	25			2	2	1	1	1				
	30			1	1	1	1	1				
	35				1	1	1	1	1	1		
	40							1				
	45						1	1		1		
	...	...									...	
	85											2

The proportion of shares which fell in the same class was 56%, and 83% fell in the same or adjacent classes. The matrix is relatively symmetric. The contracted matrix discussed in section 5.2.2 of the chapter does not obscure any information in the transition matrix presented here.

Appendix 5.C

MSE tables for the two possible forecasts

(Table 5.4 is calculated as the average of the two tables in this appendix.)

Table 5C.1

MSE of estimators in 1982-1986 for 1987-1991

Beta estimator		Mean squared error				
82-86	87-91	corr	bias	ineff	rnd	total
OLS	OLS	.54	.000	.064	.130	.195
	adj OLS ^a	.49	.003	.075	.142	.221
	Cohen	.33	.041	.121	.158	.320
	T-T	.51	.005	.076	.125	.206
adj OLS	OLS	.51	.000	.079	.137	.216
	adj OLS	.46	.002	.089	.147	.238
	Cohen	.31	.034	.135	.160	.329
	T-T	.49	.003	.089	.129	.221
Cohen	OLS	.47	.019	.086	.145	.251
	adj OLS	.42	.010	.097	.154	.260
	Cohen	.27	.002	.146	.165	.313
	T-T	.43	.008	.101	.139	.247
T-T	OLS	.53	.000	.058	.132	.191
	adj OLS	.49	.001	.067	.143	.211
	Cohen	.32	.032	.112	.159	.304
	T-T	.51	.002	.068	.126	.196

^a OLS beta adjusted by expected bias.

Table 5C.2

MSE of estimators in 1977-1981 for 1982-1986

Beta estimator		Mean squared error				
77-81	82-86	corr	bias	ineff	rnd	total
OLS	OLS	.78	.004	.030	.100	.134
	adj OLS*	.75	.002	.033	.119	.153
	Cohen	.68	.009	.049	.135	.193
	T-T	.78	.002	.035	.093	.129
adj OLS	OLS	.70	.010	.057	.128	.195
	adj OLS	.70	.007	.054	.136	.197
	Cohen	.62	.003	.078	.155	.236
	T-T	.71	.006	.062	.117	.184
Cohen	OLS	.67	.040	.065	.138	.244
	adj OLS	.65	.032	.067	.153	.252
	Cohen	.63	.002	.076	.152	.229
	T-T	.68	.031	.070	.126	.227
T-T	OLS	.77	.007	.031	.103	.142
	adj OLS	.73	.004	.034	.122	.160
	Cohen	.69	.005	.046	.133	.184
	T-T	.78	.004	.033	.090	.127

* OLS beta adjusted by expected bias.

Appendix 5.D

The individual tables of results of using Bayesian adjustment for prediction

(Table 5.8 is calculated as the average of the two tables in this appendix.)

Table 5D.1

MSE for predictions made in 1977-1981 for 1982-1986

Estimate in period 77-81	Estimate in period 82-86	Corr.	Mean Squared Error			
			Bias	Ineff	Random	Total
Historic beta ^a	TT	n/a	.004	.000	.235	.239
TT ^b	TT	.78	.004	.033	.090	.127
Adjusted ^c	TT	.79	.003	.007	.089	.099
Adjusted	Adjusted	.80	.002	.017	.064	.083

^a Average trade-to-trade beta in 1977-1981.

^b Trade-to-trade beta.

^c Bayesian-adjusted beta.

Table 5D.2

MSE for predictions made in 1982-1986 for 1987-1991

Estimate in period 82-86	Estimate in period 87-91	Corr.	Mean Squared Error			
			Bias	Ineff	Random	Total
Historic beta ^a	TT	n/a	.002	.000	.169	.172
TT ^b	TT	.51	.002	.068	.126	.196
Adjusted ^c	TT	.51	.002	.030	.125	.165
Adjusted	Adjusted	.54	.001	.060	.062	.124

^a Average trade-to-trade beta in 1982-1986.

^b Trade-to-trade beta.

^c Bayesian-adjusted beta.

Appendix 5.E**The effect of non-normality on the Chow-Fisher test**

Using the method of Fleishman and the parameters discussed in Chapter 4 the returns on 20 000 shares were simulated for 120 months using a fixed beta of unity over the whole period. The number of 'significant' changes in beta at a 5% level were then calculated.

Of the 20 000 'shares', 934 (4.67%) revealed a spuriously significant change in beta between the two halves of the 120 months. In a similar test using data from a normal distribution 5.25% of shares revealed a 'change' between the two 60 month periods.

It can therefore be concluded that under the null hypothesis of no change, non-normality has a negligible effect.

Appendix 5.F

Comparison between weekly and daily measures of thin trading and stability of weekly measures over consecutive five-year periods

The stability of thin trading measures were also assessed when the q_{jp} were calculated from a period of five years. The ability of the q_{jp} calculated from daily data to predict the proportion of weeks not traded was also measured. The comparison of the predicted proportion of weeks not traded with the proportion actually not traded provides a check on the assumption of randomness of the trading process. If q is the daily estimate of thin trading then q^w , the estimated weekly estimate is given by:

$$q^w = q^5$$

The correlation coefficients were used as indicators of good fit/stability between the daily and weekly estimates.

Table 5F.1

Summary statistics and correlations between different measures of thin trading

Period	Percentage of weeks not traded by shares in the period		Correlation between estimates of q_{jp}		
	Average	Maximum	82-86	87-91	Daily (88-91)
77-81	13.78	95.70	.97	.92	.57
82-86	13.61	90.22		.98	.69
87-91	13.62	87.51			.82
Daily* (88-91)	11.41	89.63			

* Estimate of the number of weeks not traded based on the proportion of days not traded.

The results are presented in Table 5F.1. The stability of the sets of weekly proportions is evident in the very high correlation coefficients: there is a correlation between successive periods of about 0.98 which drops slightly to 0.92 when the prediction is made two periods away.

The estimates of q_{jp} based on the daily data available (October 1987 to December 1991) provide reasonable estimates of the proportion of weeks not traded, particularly in the period

where there is an overlap (1988-1991) in the available data. The correlation between the daily data and the weekly data which do overlap is 0.82, implying that about two-thirds ($0.82^2 = 0.67$) of the variance in the daily figures are explainable by covariance with the weekly figures.

Appendix 5.G**Summary statistics of beta coefficients estimated using the various estimators**

Table 5G.1 gives the summary statistics for the estimated betas of those shares listed over the whole period from 1977 to 1991. The variability of the estimates appears to have decreased over the years, but the range of betas remains much the same. The overall levels of beta were slightly lower in the middle period, but as can be seen from the MSE analysis this

Table 5G.1

Summary statistics of betas estimated.

	Mean	Variance	Minimum	Maximum
<u>1981</u>				
OLS	.74	.317	-.48	2.36
adj OLS*	.76	.315	-.68	2.48
Cohen	.88	.352	-.82	2.85
T-T	.78	.351	-.67	2.96
<u>1986</u>				
OLS	.68	.253	-.34	2.17
adj OLS	.70	.235	-.43	2.18
Cohen	.83	.253	-1.74	2.32
T-T	.70	.265	-.84	2.18
<u>1991</u>				
OLS	.70	.189	-.90	2.42
adj OLS	.75	.174	-.91	2.49
Cohen	.89	.176	-.46	2.35
T-T	.75	.189	-.91	2.64

* OLS estimate adjusted by expected bias.

'bias' is small relative to the variance.

The adjusted OLS and trade-to-trade betas are, as expected, very similar in magnitude. The Cohen betas are on average the largest and the OLS the smallest as anticipated from our earlier findings.

Appendix 5.H**The results using the cautious version of the bootstrap estimate of the Cohen estimator standard error**

The results are very similar to those described in the paper except for the 'cautious' bootstrap in the period 1987-1991. The inconsistency here can be ascribed to the high variance of the market returns over the period including the 1987 market 'crash'. The 'cautious' version of the bootstrap involves random selection from the original data and hence the bootstrap

Table 5H.1

Summary table showing the usual and 'cautious' bootstrap estimates of standard error compared to a 'naive' estimate.

	1977-1981	1982-1986	1987-1991	Average ^a
Bootstrap estimate ^a	.301	.259	.340	.310
'Cautious' bootstrap estimate ^b	.297	.265	.436	.358
'Naive' estimate ^c	.334	.288	.397	.353
Average beta	.876	.831	.785	.819
Bias	-.001	-.001	.001	.000
Bias (cautious method)	.006	.003	-.045	-.020
Percentage of shares for which bootstrap estimate is less than 'naive' estimate	98	93	81	88
As above but using 'cautious' bootstrap method	86	72	41	60
Number of shares ^d	258	276	529	

^a Using the 'proper' method

^b Using the 'cautious' bootstrap method described.

^c The square root of the sum of squared standard errors of the individual coefficients in the Cohen estimator.

^d The number of shares with sufficient data (at least 30 months) for beta estimation chosen from the shares which had been listed for at least 240 weeks on 8 July 1992.

^e Weighted by the number of shares.

estimates involve increased variance due to the high variance of the values in the original data. The 'proper' bootstrap relies on selection from just the estimated residuals. It is apparent that the 'cautious' version is overly conservative in cases where there is high variance in the original variables.

CHAPTER 6

NON-NORMALITY AND ROBUST ESTIMATION OF BETA COEFFICIENTS ON THE JSE

6.1 Introduction

Many of the empirical studies conducted on both the JSE and other stock markets have concluded that the return distributions of securities are significantly non-normal (see, for example, Mandelbrot (1963) and Fama (1965) for early evidence from the NYSE and Schlosberg (1976) and Affleck-Graves (1974) for findings from the JSE). It is well known that the least squares estimators (which can include the trade-to-trade estimator¹) are sensitive to departures from normality (see, e.g., Chow (1983) and that, consequently, their efficiency is seriously undermined. It is therefore essential to investigate and counteract the non-normality of security returns in order to improve the efficiency of beta estimators.

The evidence presented earlier in this thesis in Chapter 4 suggested that the non-normality of return series was more severe on the JSE than on the NYSE. This finding is very much in line with the conjecture of Chan and Lakonishok (1992) who speculated that smaller firms (which are relatively more prevalent on the JSE than on the NYSE) are more likely to have heavier-tailed (i.e. more non-normal) return distributions than the normal.

In this chapter a comprehensive analysis is undertaken, both of the non-normality of security returns on the JSE and of the robust estimators designed to counteract the effect on beta estimation that non-normality has. The analysis is performed empirically and demonstrates that, over and above correcting for thin trading, beta estimators can be significantly improved by alleviating the inefficiency caused by non-normality.

The first objective of this chapter, therefore, is to establish the distributional behaviour of stock market returns on the JSE. The second objective is to extend the work of Chan and Lakonishok (1992) who compared various robust estimators in a simulation study. To this end, a far broader range of robust estimators (some of which have not, to the author's knowledge, been considered in the context of beta estimation) are tested using data from the JSE. The third

¹ The trade-to-trade estimator is named because of application of an estimation technique to the *trade-to-trade returns* rather than the *month-end returns*. In order to emphasize the occasions when the least squares estimation technique is applied to the trade-to-trade returns, the trade-to-trade estimator is referred to as the OLS estimator in this chapter. The term *OLS* used in this chapter should therefore not be confused with the references to the OLS estimator earlier in the thesis (where the phrase was used to indicate the estimation technique which did not correct for thin trading).

aim of the chapter is to use the robust estimators to gain additional insight into beta stability *as well as* into the use and interpretation of the mean squared forecast error (MSE) as a criterion for assessing stability.

The structure of the chapter follows the three objectives outlined above. That is, in section 6.2 the tests of non-normality are introduced and implemented on the data from the JSE. In section 6.3 a brief review of robust estimation is given and the results of an empirical comparison of the estimators are presented. Some of the problems associated with the testing of beta stability are described in Section 6.4 and it is shown how they are alleviated to some extent by the use of robust estimators. Section 6.5 contains a summary and some conclusions.

In order to enhance the readability and flow of the discussion, several results and reviews have been put in appendices at the end of the chapter. Brief outlines of the contents of the appendices are as follows. Appendix 6.A contains a further elaboration of one of the results in the main study. Appendix 6.B is a detailed review and discussion of robust estimators in general, and with particular reference to financial applications. The reader is referred to this appendix for some useful additional insights into which robust estimators are suitable for estimation in the financial area. Appendix 6.C contains, for completeness of discussion, some further technical details about one of the family of robust estimators outlined in Appendix 6.B. Finally, Appendix 6.D elaborates on the measurement of the variance of the robust estimator known as the L_p -estimator. The results discussed in this appendix are useful for researchers wanting to implement the L_p -estimator because they off-set the current difficulty in obtaining a reasonably practical method of estimating the variance.

6.2 Non-normality of residuals

Before beginning the investigation of non-normality, it is worth remembering that it is the distribution of the *residuals* of the market model (rather than that of the actual *returns*) which is critical for efficiency in OLS estimators. Therefore, in this section, the tests for detecting the extent of non-normality are focused on the residuals of securities on the JSE. The outline of the section is as follows: firstly, in sub-section 6.2.1, the measures of non-normality are defined. Then, in sub-section 6.2.2, the measures of non-normality are used to establish the extent of non-normality on the JSE. Finally, in sub-section 6.2.3, the reliability and stability of the measures of non-normality are considered.

6.2.1 Description of tests of non-normality

The focus in this section is on tests for normality using the sample coefficients of skewness and kurtosis of residuals. In the literature other composite tests (e.g. the Studentized range and Shapiro-Francia W' test statistic which were considered, for example, by Affleck-Graves and McDonald (1989)) have also been used to test for normality of the residuals. However, as Affleck-Graves and McDonald (1989) point out, the use of the skewness and kurtosis as indicators of the non-normality have two advantages over composite tests: first, they are at least as sensitive as omnibus tests and, second, they give an indication of the *direction* of any non-normality. In the implementation of robust regression techniques the directional tests provide a further advantage: many of the techniques are adaptive in the sense that the degree to which the regression is 'robustified' depends on the degree to which the residuals are non-normal. The adaptive criterion is usually couched in terms of the kurtosis or skewness of the residuals.

In addition, tests of normality of residuals are plagued by the fact that the actual residuals are unobservable and only the estimated residuals can be tested for normality or otherwise. However, for large enough samples, White and MacDonald (1980) showed that the more powerful tests (such as those based on skewness and kurtosis) are reasonably reliable when applied to estimates of residuals.

Following Affleck-Graves and MacDonald (1989) the coefficients of skewness and kurtosis can be calculated using the formulae:

$$\sqrt{b_1} = m_3 / (m_2^{3/2}) \quad (\text{for skewness});$$

$$b_2 = m_4 / m_2^2 \quad (\text{for kurtosis});$$

where $m_r = \sum (x_i - m'_i)^r$, and $m'_i = \sum x_i / n$.

In order to use the above coefficients as 'tests' for non-normality, their distributions under the null hypothesis of normality have to be known. Unfortunately, the distributions of the sample coefficients are generally difficult to determine precisely and consequently approximate probability integrals are used (see D'Agostino and Stephens, 1986). Nevertheless, Pearson, D'Agostino and Bowman (1977) have provided simulation evidence that tests based

on the coefficients of skewness and kurtosis are reasonably powerful. The tests are based on standardizing the coefficients of skewness and kurtosis so that they have an approximately standard normal distribution.² In the remainder of this chapter the *standardized coefficients* are used to aid interpretation and are henceforth referred to as simply 'skewness' and 'kurtosis'. Using the expressions for skewness and kurtosis above the extent of non-normality in security residuals on the JSE is investigated in the next sub-section.

6.2.2 Extent of non-normality on the JSE

As discussed in the previous section, actual residuals are unobservable and hence their kurtosis and skewness cannot be directly measured. In this investigation into the extent of non-normality in the residuals, the residuals were estimated by fitting the market model (using OLS) to securities on the JSE.

The data in the investigation comprised the monthly return series of all the securities which were listed continuously on the JSE for each of three non-overlapping five-year periods. The periods and numbers of shares in each of the periods were: 286 shares in the first period, i.e. 1978.01-1982.12; 308 shares in 1983.01-1987.12 and 538 shares in 1988.01-1992.12.

In Table 6.1 a frequency table of the resulting standardized coefficients of skewness and kurtosis are shown across various 'significance classes' over the three time periods. Each of the numbers in the table reflects the percentages of shares which have coefficients of skewness or kurtosis which are different from zero at various significance levels. For comparative purposes, the *expected* percentages falling in each 'significance class' (given in the first column) are derived from the standard normal distribution. Table 6.1 thus depicts not only the percentage of shares which have significant departures from normality, but also indicates the direction (either negative or positive) in which the departures lie.

It is evident from a comparison of the *observed numbers* of securities with very large coefficients of skewness and kurtosis with the *expected numbers* in Table 6.1 that there is significant non-normality among the estimated residuals. There are significant departures (from the expected percentages) in the measures of skewness (both negative and positive) and in the measures of kurtosis (only positive) across all three periods. This observation is reflected by

² The formulae for the standardizations of the coefficients of skewness and kurtosis are given in D'Agostino and Stephens (1986) and are based on work by D'Agostino (1970) and Anscombe and Glynn (1983), respectively.

Table 6.1

Extent and distribution of non-normality of residual distributions among securities on the JSE. The figures in the table represent the percentage of securities with a normalised coefficient of skewness/kurtosis in a given range.

Crit. ^a	Exp. ^b	1978-1982		1983-1987		1988-1992	
		Skew ^c	Kurt ^d	Skew	Kurt	Skew	Kurt
< -2.58	0.5	6	0	6	0	11	0
< -1.96	2.0	3	1	3	0	4	0
< -1.64	2.5	2	1	2	0	3	0
< 0	45.0	29	15	26	14	22	9
< 1.64	45.0	36	40	36	46	33	37
< 1.96	2.5	3	8	5	6	5	7
< 2.58	2.0	8	14	8	9	7	17
> 2.58	0.5	13	22	14	25	14	30

The numbers in the table represent the percentage of securities with a normalised coefficient of skewness/kurtosis with a value less than Crit. but greater than Crit. for the row above.

^a Critical values for the normal distribution, i.e. the 0.5%, 2.5%, 5%, 50%, 95%, 97.5% and 99.5% percentile points of the standard normal distribution.

^b The expected percentage of securities with normalised coefficients of skewness/kurtosis less than the critical value, but greater than the critical value in the row above.

^c The normalised coefficient of skewness.

^d The normalised coefficient of kurtosis.

the fact that there are more large (significant) normalised coefficients of skewness and kurtosis than would be expected from a set of securities with normally distributed residuals. For example, consider the data for the five-year period 1978-1982 in Table 6.1. It is apparent from the results that 13% of the securities had standardized coefficients of skewness greater than 2.58, and 6% less than -2.58. If the securities all in fact had normal residual distributions, then it can be seen from the shaded column in Table 6.1 that, on average, only 0.5% of the securities would be expected to have a skewness greater than 2.58, and only 0.5% a skewness less than -2.58. It is clear therefore that many securities do indeed have significant coefficients of skewness. Examining the kurtosis numbers for the 1978-1982 period in Table 6.1, it can be seen that 22% of the securities had coefficients greater than 2.58 compared with the expected percentage of 0.5; clear indication that many securities have heavier tails than the normal. The conclusion regarding heavier tails than the normal distribution among the securities on the JSE is borne out by noting that in Table 6.1 there a *fewer* securities than expected with coefficients less than zero (i.e. with lighter tails than normal). For example, there are no shares with coefficients of kurtosis less than -1.64 in 1983-1987 and 1988-1992, and only two shares in 1978-1982.

Looking more closely at the pattern of non-normality in the securities which is evident in Table 6.1, it is clear that there are slightly more significantly positively skewed results than negatively skewed (viz. 60% had positive skewness in 1978-1982; 63% were positive in 1983-1987 and 59% were positive in 1988-1992). Extreme leptokurtosis is evident in all three periods in the very high proportion of securities (greater than 22% in each time period) with coefficients which are highly significant at the 1% level (two-sided probability, critical magnitude 2.58).

The proportion of shares experiencing significant non-normality appears to be slightly increased over the 1988-1992 period, with 54% of securities having a normalised coefficient of kurtosis greater than 1.64 in contrast with the 40% and 44% for the periods 1983-1987 and 1978-1982, respectively.³

Another statistic which assists in describing the non-normality of securities is the correlation between skewness and kurtosis. However, Bowman and Shenton (1975) argue that the *sample coefficients* of kurtosis and skewness are not independent. Nevertheless, they demonstrate that the sample coefficients *are* actually uncorrelated and can be considered *nearly* independent. Therefore, any correlations between the sample coefficients of skewness and kurtosis can be attributed to a real correlation between the underlying population skewness and kurtosis of the security residuals. The correlations between the kurtosis and absolute skewness for the periods 1978-1982, 1983-1987 and 1988-1992 were found to be fairly high, i.e. 0.79, 0.75 and 0.76, respectively. The high correlations indicate that the shares which have very skewed distributions also predominantly have leptokurtic distributions. The implication of the correlation is that if a security is shown to have a high level of kurtosis then it is also likely to have significant skewness. Conversely, it will also be true that a security with low kurtosis will have low skewness and consequently is likely to have a 'normal' residual distribution.

Although the focus remains on the impact of skewness and kurtosis measures, a composite test of normality involving the joint measures of skewness *and* kurtosis together was implemented. For computational ease the test suggested by D'Agostino and Pearson (1973) was used. Their test statistic is obtained by summing the squares of the standardized coefficients of skewness and kurtosis. Being the sum of two standardized normal variates under the null hypothesis of normality, the test statistic is approximately distributed with a chi-square distribution with two degrees of freedom. It is worth noting that the dependence of the sample

³ The figures were obtained from Table 6.1 by adding all the percentages in the rows marked < 1.96 down to the row headed > 2.58.

coefficients of skewness and kurtosis again obscures the interpretation of the results to some extent. Using the composite test described above, the proportion of significantly non-normally distributed residuals in each of the three periods were found to be 24%, 24% and 32% respectively, when tested at a 0.5% significance level.

It would appear from the above empirical results that at least a quarter of all the securities on the JSE are significantly non-normally distributed. These results are consistent with the evidence reported in Table 6.1: specifically, the finding that the percentage of shares with significant (at a one-sided 0.5% level) kurtosis was reported in Table 6.1 to be at least 22%. This evidence thus confirms that a significant number of securities on the JSE have residuals with non-normal distribution.

Having verified that non-normality is indeed a significant characteristic of the residual components of securities on the JSE, the next important issue to consider is the reliability of the measures of non-normality. In other words the issue is addressed of whether it is reasonable to measure the kurtosis and skewness of a residual series in one period and then to use them to describe the distribution in another period.

6.2.3 Reliability of sample skewness and kurtosis as measures of non-normality

From the previous sub-section it is apparent that the non-normality of share residuals is generally a fairly constant characteristic over the three time periods considered. In this sub-section the extent to which stability of non-normality applies to individual securities is examined.

To the author's knowledge, no detailed empirical study which examines the stability of non-normality for individual securities has been forthcoming on any of the stock exchanges. Interest in the stability of non-normality stems from the argument that a persistent level of non-normality for individual security residuals would indicate that the non-normality is associated with a fundamental characteristic of the security. For example, Chan and Lakonishok (1992) speculated that a characteristic such as the size of the firm may be linked to non-normality.

To investigate the above issues empirically, two sets of data were used. In the first data set all shares which had been listed throughout the ten-year period 1978.01 to 1987.12 (numbering 287) were included. The second data set comprised all those securities listed

continuously over the ten-year period 1988.01 to 1992.12 (numbering 266). The monthly returns⁴ measured over each of the two data sets were separated into four groups:

- (i) the returns measured over the first five years of the period;
- (ii) the returns measured over the second five years of the period;
- (iii) the returns measured every second month starting from the first month of the period, and
- (iv) the returns measured every second month starting from the second month in the period.

The rationale for the split was twofold: first, the *internal reliability* of the measures of non-normality⁵ could be assessed by comparing the measures over the two groups of *alternating* monthly returns, i.e. groups (iii) and (iv); second, the *sequential correlation* in the measures of non-normality could be ascertained by comparing the measures of kurtosis and skewness over *consecutive* (i.e. sequential) five-year periods, i.e. groups (i) and (ii).

The internal and sequential correlations in the measures of kurtosis and of skewness are presented in Table 6.2. The row labelled *internal* gives the correlations between the alternating data sets; i.e. the internal reliability of the measures. In both data sets the coefficients of

Table 6.2

Correlations in the estimates of skewness and of kurtosis in consecutive and alternating data series.

	1978-1987		1983-1992	
	skewness ^a	kurtosis ^b	skewness	kurtosis
internal	-.055	.297	-.035	.334
sequential	.043	.215	.002	.110

^a Correlation in coefficients of skewness and kurtosis measured on alternating monthly returns in the same period and on consecutive five-year periods. (The approximate standard deviation is .059.)

^b Correlation in coefficients of skewness and kurtosis measured on alternating monthly returns in the same period and on consecutive five-year periods. (The approximate standard deviation is .061.)

⁴ To avoid problems relating to thin trading, the monthly returns were computed as trade-to-trade returns. The fact that the 'months' were of varying lengths necessitated adjusting the measured return for this. The adjustment followed common practise by dividing the returns by the square root of the actual length of the interval before performing the regression. The adjustment ensures that heteroscedasticity is avoided (under the plausible assumption - tested in Chapter 4 - that the variance of the residuals grows in proportion to the length of the interval).

⁵ Residuals from a robust technique known as bounded influence $L_{1.5}$ -norm estimation were used in estimating the coefficients of skewness and kurtosis. It is shown in 6.3 that the above estimator is one of the most efficient for securities on the JSE and will therefore ensure that the estimated residuals correspond more closely to the unobservable, true residuals than the estimated residuals from any other estimator.

skewness show negligible internal reliability, i.e. the correlations of -0.055 for 1978-1987 and -0.035 for 1983-1992 are smaller than the standard deviations (which are both 0.059) and hence are 'insignificantly' different from zero. The *internal* correlations for the coefficients of kurtosis reported in Table 6.2 are also small (being 0.297 and 0.334 for the two periods), but are nevertheless significantly larger than zero.

In the row labelled *sequential* in Table 6.2, the sequential correlations in successive five-year periods of the sample coefficients of skewness and kurtosis are reported. The sequential correlations for skewness are negligibly small (i.e. 0.043 and 0.002) for both data sets. The *sequential* correlations for kurtosis are, similarly to the *internal* correlations, significantly positive. The *sequential* correlations of 0.215 and 0.110 are lower than the *internal* correlations of 0.297 and 0.334 for the two data sets, suggesting instability over time of the coefficients of kurtosis. The results from the above analysis indicate that the distributions of securities are not constant over time, and the fact that the distribution of a residual series is normal in a previous period (or even part of the current period) does not imply that it will currently be normal.

α The evidence presented in this chapter so far indicates that there is significant non-normality on the JSE. The next investigation tests whether there is a link between the non-normality characteristic of the JSE and the other prominent feature of the JSE, namely thin trading. A significant relationship between thin trading (which Dimson (1979), for example, has proved to be associated with firm size) and non-normality would offer some confirmation of the hypothesis that small firms and non-normality were mutually correlated.

A correlation analysis using the securities in the 1978-1982, 1983-1987 and 1988-1992 periods resulted in significant negative correlations between the coefficients of kurtosis and the trading frequencies (measured by the proportion of weeks in which the security traded) of -0.38, -0.40 and -0.30, for the above-mentioned periods respectively.

The evidence therefore suggests that non-normality in individual securities can be explained to some extent by infrequent trading, with the better traded securities being relatively less susceptible to non-normality in their residual distributions. Given the severe thin trading prevalent on the JSE, the above finding confirms indirectly that non-normality is a factor which should be taken into account for efficient beta estimation.

Robust estimators are designed to take into account the non-normality of security residuals and consequently to provide efficient estimation of beta coefficients. In the next section, therefore, a variety of robust estimators are discussed.

6.3 A comparison of robust estimators as applied to security returns

This section contains two main themes: first, a theoretical discussion of robust estimators which are suitable for application in beta estimation on the JSE and, second, the results of an empirical study designed to compare and contrast a selection of the robust estimators.

To facilitate the description and discussion of the robust estimators in the context of the market model, the simple linear regression equation must be examined. The simple regression model specifies the relationship between a dependent variable and an explanatory (or independent) variable as follows:

$$y_t = \alpha + \beta x_t + z_t$$

where y_t is the dependent variable, x_t is an independent variable and z_t is a residual term independently and identically distributed around zero.

The use and verification of the model relies on efficient and consistent estimates of the parameters α and β . In the case of the market model where the y and x variables represent the returns on a security and on the market respectively, the β of course corresponds to the systematic risk, or beta of the security.

Estimation of the parameters in the financial and econometric fields has generally been performed using the method of OLS.⁶ While software for performing OLS is widely available and the methodology is computationally quite simple, the main appeal of OLS rests on the Gauss-Markov theorem. The Gauss-Markov theorem states that the OLS estimator is the best (in the sense of minimum variance) of all linear unbiased estimators of β , provided the variance of z is finite. The best linear unbiasedness property is commonly referred to as BLU. While it is not obvious that minimum variance is the only criterion by which to compare unbiased estimators, the use of OLS gains further popularity by corresponding exactly to a maximum likelihood estimator (MLE) provided the z have a normal distribution.

In many instances it is quite plausible and convenient to assume that the normal distribution is appropriate for the residuals, z ; in the case of the market model the z are

⁶ Again, it should be emphasized that the reference to OLS in this chapter includes the trade-to-trade adjustment.

assumed to represent the impact on the security's return of a large number of independent, firm-specific (i.e. diversifiable) influences. If the distributions of the influences on the security's returns have finite variances themselves, then the central limit theorem will apply to the sum of the independent influences and thus the z will indeed have a normal distribution. All of the *usual* probabilistic statements about the estimates of the parameters and the model (e.g. any hypothesis tests about parameter values, or tests of goodness-of-fit) depend on the z having a normal distribution. Therefore it is often convenient to make the assumption of normality.

Easily accessible computer facilities have removed the need to use simple computational methods. The evidence regarding the non-normality of residuals is quite clear and the extent to which the returns of some securities are non-normal suggests that the assumption of finite variances, and hence the minimum variance quality of the OLS estimator, is not always valid. In addition, the maximum likelihood quality of OLS also falls away if the returns are non-normal.

There is a large body of statistical and econometric literature which deals with robust regression. For example, Huber (1981) discusses and has developed much of the theory of robust estimators in general while Koenker (1982) examines the facets of robustness with respect to regression techniques.

More specifically in the econometric literature, Smith and Hall (1972) and Blattberg and Sargent (1971), for example, have shown that MLEs are generally preferable to BLU estimators. Sharpe (1971), Cornell and Dietrich (1978), McDonald and Nelson (1989), Easton (1991) and Chan and Lakonishok (1992) are among the few who have applied robust techniques specifically to financial data. Their findings for NYSE data have indicated that the use of robust estimators offers marginal improvement over the use of OLS. In one of the more recent papers addressing robust estimation in the context of estimating systematic risk, Chan and Lakonishok (1992) assessed several robust estimators in a simulation study. They further investigated the differences in beta estimates provided by trimmed regression quantiles estimators and OLS estimators on data from the NYSE.

The investigation in this chapter extends the work of Chan and Lakonishok by testing a much wider variety of robust estimators on financial data with particular emphasis on the relative *efficiency* of the estimators rather than the differences in the estimates from OLS estimates. The change of emphasis arises because with empirical data the true beta of the security is unknown and so a comparison of the values of the estimates on their own does not provide much information regarding the superiority of the estimators. For example, the OLS

estimator may not be biased but it may be inefficient. In this case simply comparing average estimated betas with those from other estimators will provide little evidence of a change in efficiency.

In this chapter efficiency is measured using a technique common in the psychometric and educational literature whereby the estimator is applied to different parts of the same sample and the different estimates obtained are compared. The comparison of the different estimates provides a measure of the 'internal reliability' of the estimator, which is similar to the test of the stability in the measures of non-normality conducted in section 6.2. By measuring internal reliability focus is placed on the ability of the estimator to produce consistent results when the underlying parameter (i.e. the beta) is assumed to be constant. Once the internal reliability of an estimator is established it can be used to gain inference about changes over time of the underlying parameter.

The rest of section 6.3 has the following structure. In sub-section 6.3.1 which follows, a broad range of robust estimators are briefly outlined. Representative estimators are then selected from the different classes of robustness to be implemented in the empirical study. Sub-section 6.3.2 contains a short discussion of *bounded influence* estimators in which not only outlying residuals are down-weighted, but also outlying market returns. The reason for down-weighting outlying market returns is that they may unduly affect the estimation of the slope coefficient when OLS is used as an estimator. The data used in the empirical investigation are described in sub-section 6.3.3. The main empirical investigation is presented in sub-section 6.3.4. In the investigation, the estimators are initially rated on the basis of their 'internal reliability'. However, cognizance also needs to be taken of the goodness-of-fit of the estimators and so a combined assessment of internal reliability and goodness-of-fit is developed which allows a 'best' estimator to be determined. In sub-section 6.3.5 the reliability of the estimators is compared for securities which have normal and non-normal residual distributions. The reason for the comparison is to establish whether OLS estimators are as reliable as the robust estimators for the securities with normal residuals. For non-normal residuals, the robust estimators are expected to outperform the OLS estimator considerably. The last sub-section, 6.3.6, tests whether robust estimators actually result in significantly different estimates of beta coefficients, in addition to any increased efficiency.

6.3.1 The selection of robust estimators

Robust estimators are often categorized into three main classes (which overlap in specific cases), namely *M*-estimators, *L*-estimators and *R*-estimators. These classes cover nearly all of the robust estimators which can be readily implemented in regression analyses. The purpose of this sub-section is to outline the three different classes of robust estimator and to select from the classes specific estimators which are reasonably easy to implement on stock exchange data. The reader is referred to Appendix 6.B for a more comprehensive discussion of non-normality and robust estimators.

M-estimators encompass a wide range of maximum likelihood type methods. *M*-estimation generally involves maximising a transformation of the likelihood function of the residuals. The likelihood function can thus be considered to be the estimation criterion. Robustness is ensured by selecting an estimation criterion which accords low weights to outlying observations, unlike least squares (which corresponds to the likelihood function of the normal distribution) which accentuates the effects of outliers. Such a criterion is most easily found by specifying a distribution for the residuals which anticipates heavy tails. The *M*-estimators which are used in the comparison of estimators in this chapter are OLS, adaptive L_p -norm estimators, least absolute deviation (LAD) regression and Tukey's biweight estimators. The motivation for selecting the above-mentioned estimators can be summarized as:

OLS	corresponds to MLE for the normal case and is included chiefly for comparative purposes;
L_p -norm	corresponds to the MLE for the Box-Tiao family of distributions which in turn belong to the family of generalised- <i>t</i> distributions suggested by McDonald and Newey (1988) for use with security returns;
LAD	corresponds to the MLE of the Laplace (double exponential) distribution; it is one of the most commonly used and easiest of the robust estimators to implement;
Tukey's biweight	does not correspond to any standard MLE, but is an example of a function chosen to maximize a weighted sum of the residuals, where large, outlying residuals have small weights

Before discussing L -estimators, it is worth elaborating on L_p -norm estimation in order to highlight a practical detail required for its implementation. The L_p -norm estimator involves minimizing an expression of the form:

$$\sum_{i=1}^n |y_i - \alpha - \beta x_i|^p \quad (1)$$

for a selected value of p . If it is noted that the expression between the modulus signs corresponds to the estimated residual of the observation then, for example, it can be noted that by setting p to 2, the L_p -norm estimator is identical to the usual least squares estimator. For values of p less than 2, the estimator becomes more robust as it will accord large residuals less weight than OLS. When p is set to unity, the L_p -norm estimator is identical to the LAD estimator.

L_p -norm estimators can be adaptively implemented. By adaptive it is meant that the p -parameter can be chosen based on preliminary evidence about the distribution of the particular security. Given the instability of residual distributions and the fact that the optimal p value will depend on the extent of non-normality, the adaptive selection of p appears appropriate. Two methods for selecting the p parameter were used: first, the method proposed by Barr (1980) which is based on the empirical relationship between the kurtosis of the residuals and the optimal value of p and, secondly, a method based on minimizing the asymptotic variance of the estimate as derived by Nyquist (1980, 1983). The two variants are referred to as $L_p(k)$ and $L_p(\omega)$, respectively. The computation of the asymptotic variance is not a trivial exercise and can lead to serious practical difficulties. In Appendix 6.D, a novel way of obtaining robust estimates of the asymptotic variance is outlined and its usefulness confirmed against bootstrap estimates of the L_p -norm variances.

The second class of robust estimators which are considered are the L -estimators. L -estimators are based on linear combinations of order statistics. Koenker and Basset (1978), Basset and Koenker (1982) and Koenker and Portnoy (1987) give detailed descriptions of the regression equivalent to order statistics, namely regression quantiles, and demonstrate their application to robust regression. The use of regression quantiles gives rise to the trimmed mean estimators of Koenker and Basset (1978) and Welsh (1987). In order to make the Welsh estimator functional, the ordered residuals from a preliminary regression fit have to be

available. In the empirical study which follows, both OLS and LAD were used to provide these preliminary residuals.

It should be noted that LAD regression can also be formulated as an L -estimator and its joint classification as both L - and M -estimator adds to its intuitive appeal as a robust regression technique applicable in a wide variety of cases. In the empirical study the 15% Koenker-Basset trimmed mean and the 10% Welsh trimmed mean⁷ based on residuals from two preliminary fits (OLS and LAD) are implemented.

The third class of robust estimators, R -estimators, are based on rank statistics. In the study described in sub-section 6.3.4 one version of a family of regression analyses which Olkin and Yitzhaki (1992) call Gini regressions is used. The Gini coefficient and Gini mean difference (which are based on the ranks rather than the actual values of the variables) are fairly well-known in the area of finance, econometrics and economic statistics (see Yitzhaki (1982), Shalit and Yitzhaki (1984), and Carroll, Thistle and Wei (1992), for recent examples).

Scholz (1977) and Sievers (1978) independently derived regression techniques based on the weighted average of the slopes between pairs of observations, which Olkin and Yitzhaki show are special cases of Gini regressions. In this chapter the Gini regression of Sievers and Scholz is computed for comparison with the other estimators.

All the estimators described above are designed to incorporate the non-normality of the residual distributions in order to make the estimation procedure more efficient. However, what they neglect to do is to build in the non-normality of the market returns. The market returns are themselves subject to outlying values and it can be shown that when OLS is used (see Belsey, Kuh and Welsh (1981)) an outlying value of the x -variable in a regression can severely distort the parameter estimation. In the next sub-section a brief account of bounded influence estimation (which avoids these distortions) is presented.

6.3.2 Bounded influence

An important additional concept in the 'robustification' of beta estimation is that of bounding the influence of the explanatory variables. Market returns which are either extremely large or small have large 'leverage' and may dominate the fitting of the regression line (see

⁷ The reason different trimming proportions are used in the Welsh and Koenker-Basset estimators is that the Koenker-Basset estimator actually trims slightly fewer observations than the trimming proportion. De Jongh, De Wet and Welsh (1988) report that the 15% Koenker-Basset actually trimmed about 11% of the observations.

Belsey, Kuh and Welsh (1981)). The 'leverage' effect of an outlying market value in a regression analysis can be seen by considering the simplistic case where all but one of the x -values are bunched together and the outlying value is very large, say 5 standard deviations greater than the average value in the bunch. The application of OLS will, in effect, treat the bunch of points and the single outlying point as just two points through which to fit the line. The single outlying point will thus have a substantial influence on the estimation of the line.

The non-normality of the market index (as discussed in Chapter 4 of this thesis) is such that outlying values are likely to occur in a sample of points. Adjusting for the possible sensitivity of the estimate of beta to potentially outlying market returns is therefore an essential part of obtaining an efficient estimate of beta.

The proposals of Krasker and Welsch (1982) and of Mallows (1973, 1975) for bounding the influence of outlying explanatory variables have gained prominence in the literature. The method of Mallows down-weights observations with outlying market returns irrespective of the size of the residual associated with the observation, whereas the Krasker-Welsch method combines the down-weighting of the residuals and explanatory variables.

In deciding on which of the approaches to adopt the following considerations were taken into account. First, the separation of the two down-weighting processes as in the Mallows scheme results in some efficiency loss when the residuals do follow a normal distribution. Second, however, Giltinan, Carroll and Ruppert (1986) suggest that inferences made using the Mallows approach will be more stable than that of Krasker and Welsch and that in any case neither method is consistently preferable in terms of efficiency. Furthermore, the Mallows approach is much easier to implement, given that it does not involve any preliminary regression analyses. The Mallows approach has therefore been included in the empirical comparison.

Before describing the empirical study, it is worth setting out the full set of estimators considered and their associated abbreviations:

Ordinary Least Squares	OLS
Least Absolute Deviations (L_1 -norm)	LAD
15% Koenker-Basset trimmed mean	KB
10% Welsh trimmed mean (OLS residuals)	W(OLS)
10% Welsh trimmed mean (LAD residuals)	W(LAD)
Tukey's biweight	TBW
L_p -norm (p based on kurtosis)	$L_p(k)$

L_p -norm (p based on asymptotic variance)	$L_p(\omega)$
Gini regression of Sievers (1978)	Gini

Mallows-weights bounded influence versions of OLS, LAD, KB, W(LAD), $L_p(k)$ and $L_p(\omega)$ are also implemented. Their abbreviations are prefixed by *BI*.

The next sub-section then contains a short description of the data used in the empirical study.

6.3.3 The data used in the investigation

The data used in the analysis covered the period from January 1978 to December 1992. Similarly to the investigation into the measures of non-normality discussed in section 6.2, the data were split into two sets of shares: first, all those shares which had been continuously listed from 1978.01 to 1987.12 and, second, all those listed from 1983.01 to 1992.12. The two sets were chosen for compatibility with the stability tests described later in section 6.4. The monthly trade-to-trade returns were calculated for all the shares in the two data sets.

For each of the two sets of shares the monthly returns were further divided into alternate data series in a similar way to that described in Section 6.2, i.e. one series of alternate monthly returns starting with the first month's returns and the other starting with the second month's returns. The two periods (1978-1987 and 1983-1992) contained 287 and 266 shares⁸ respectively and in each there was a maximum of 120 trade-to-trade months.

The next sub-section contains the methodology and results of the comparison of the various beta estimators. As mentioned earlier, the comparison is based on internal reliability. Because there are no benchmark values in empirical studies against which to compare estimates, efficiency can only be judged by assessing the extent to which the estimator produces estimates which are consistent with each other when similar data sets are used. The consistency of results is referred to here as 'internal reliability'.

⁸ Shares for which there were less than 60 trade-to-trade returns in the ten-year period were excluded.

6.3.4 Internal reliability of the various robust estimators

All of the robust estimators were applied to both series of alternating returns for each of the two data sets. The measure of internal reliability was defined to be the mean squared error (MSE) between the betas estimated from the two alternating return series. The MSE for an estimator in a period p is given by

$$MSE_p = \sum_{i=1}^{k_p} (\hat{\beta}_i^1 - \hat{\beta}_i^2)^2 / k_p \quad (2)$$

where $\hat{\beta}_i^m$ is the estimate of beta for share i in the data set p starting in month m (where m is either 1 or 2 in equation (2)) and k_p is the number of shares in the data sets corresponding to data set p .

To obtain further insight into the reliability (as measured by the MSE) the total MSE was decomposed into the terms suggested by Theil (1971), namely bias, inefficiency and randomness. The decomposition establishes the reasons *why* one estimator might be more reliable than another. Following Theil (1971), the decomposition is

$$\begin{aligned} MSE_p &= \sum_{i=1}^{k_p} (\hat{\beta}_i^1 - \hat{\beta}_i^2)^2 / k_p \\ &= (\bar{\beta}_i^1 - \bar{\beta}_i^2)^2 + (1 - \hat{\gamma})^2 \text{Var}[\hat{\beta}_i^2] + (1 - \rho^2) \text{Var}[\hat{\beta}_i^1] \\ &= \text{bias} + \text{inefficiency} + \text{randomness} \end{aligned} \quad (3)$$

where $\hat{\gamma}$ is the slope of the regression of the beta for data set 1 on the beta in period 2; ρ is the estimated correlation coefficient between the two estimates of beta; $\bar{\beta}^m$ is the average beta across all shares in the period p for the data series starting in month m . It is again important to note that the 'bias' referred to above is the level of average change between the estimates of beta between the two alternating series (and is not intended to suggest that the estimators are biased by thin trading, for instance). The second term in equation (3) reflects the inefficiency of the estimator in predicting the beta for the alternate series, and the third term measures the random error, i.e. the degree to which the estimates of beta obtained from the alternate data sets are not correlated.

Before examining the decomposed MSEs, however, the correlations and total MSEs are reported in Table 6.3 for all the beta estimators considered. Also included in the table are the average (over both the alternate data series in each period) inter-quartile range (IQR) for the residuals and the average median absolute residual (MAR). These latter two measures (IQR and MAR) give an indication of the goodness-of-fit of the models. The reason for including the two measures of goodness-of-fit in the comparison is that a reliable (as defined above) estimator which does not fit the data well is clearly not very useful.

Table 6.3

Comparative measures of reliability and goodness-of-fit

Estimator	1978-1987				1983-1992			
	ρ^a	MSE ^b	IQR ^c	MAR ^d	ρ	MSE	IQR	MAR
OLS	.57	.1889	.1184	.0598	.56	.1626	.1192	.0588
LAD	.77	.1226	.1138	.0562	.71	.1202	.1154	.0557
KB	.73	.1192	.1158	.0572	.65	.1238	.1170	.0570
W(OLS)	.75	.1086	.1161	.0579	.71	.0968	.1172	.0572
W(LAD)	.74	.1109	.1160	.0578	.72	.0936	.1172	.0572
TBW	.78	.1001	.1149	.0568	.75	.0885	.1159	.0560
$L_p(k)$	.76	.1066	.1154	.0572	.73	.0962	.1168	.0567
$L_p(\omega)$	.76	.1072	.1157	.0573	.71	.1017	.1168	.0566
Gini	.74	.1180	.1156	.0585	.75	.0903	.1162	.0571
BIOLS	.72	.1308	.1097	.0556	.62	.1365	.1102	.0549
BILAD	.78	.1202	.1137	.0562	.74	.1119	.1156	.0557
BIKB	.79	.0993	.1150	.0572	.78	.0788	.1167	.0567
BIW(LAD)	.79	.1015	.1156	.0576	.77	.0811	.1174	.0573
BIL _p (k)	.79	.1023	.1031	.0509	.77	.0828	.1052	.0512
BIL _p (ω)	.79	.0990	.0990	.0492	.77	.0811	.1004	.0490

- ^a The correlation coefficient between estimates of beta calculated from the two alternating data series in the ten-year period.
^b The mean squared error between the estimates of beta calculated from the two alternating data series in the ten-year period.
^c The average interquartile range of residuals from the model fits to the two alternating data series in the ten-year period.
^d The average median absolute residual from the two model fits to the two alternating data series in the ten-year period.

Examining the columns giving the total MSEs in Table 6.3 it can be seen that the OLS estimator (with total MSEs of 0.1889 and 0.1626 for the two periods 1978-1987 and 1983-1992, respectively) is indeed far less reliable than any of the robust estimators (the maximum MSE across all of the robust estimators is 0.1308 for 1978-1987 and 0.1365 for 1983-1992).

The differences in the MSEs are testimony to the vast improvement in reliability that robust estimators offer over OLS.

Further, it is evident from Table 6.3 that it is the bounded influence estimators (those whose abbreviations are prefixed by *BI*) which generally seem to provide greater reliability than the robust estimators which are not bounded along the x -axis.⁹ For example, focusing on the 1978-1987 data, the MSE for the OLS estimator drops from 0.1889 to 0.1308 when the Mallows bounded influence adjustment is applied. Similarly, the MSE for the Koenker-Basset estimator (KB) is 0.1192 when unbounded and is 0.0993 when it is bounded (i.e. the BIKB estimator).

Adjudged purely on MSEs in Table 6.3, the bounded influence Koenker-Basset (BIKB) estimator appears the most suitable of those investigated in the 1983-1992 period (with an MSE of 0.0788) and one of the bounded influence versions of the L_p -norm estimators, i.e. $BIL_p(\omega)$, appears the best in the 1978-1987 period (with an MSE of 0.0990). The measures of goodness-of-fit in Table 6.3 (i.e. the average IQRs and average MARs) are reasonably constant for all the estimators, although there is some indication that the L_p -norm estimators have slightly better measures of goodness-of-fit. In fact, the MARs and IQRs of the $BIL_p(\omega)$ estimator are the lowest in both data periods and the corresponding measures for the $BIL_p(k)$ are second lowest.

In order to take into account the goodness-of-fit indicators as well as the reliability in a single, joint measure, each figure in the columns MSE and IQR was expressed as a percentage of the OLS figure. Thereafter the average percentage across both periods for both MSE and IQR was computed. This average appears in the last column of Table 6.4 and represents one omnibus criterion to assess the 'best' estimator.

From Table 6.4 it is evident that the bounded influence trimmed mean estimators (BIW(LAD) and BIKB) and L_p -norm estimators show up to be superior when assessed in terms of reliability and goodness-of-fit. The 'score' for the $BIL_p(\omega)$ estimator is the lowest (at 67.5), followed by the $BIL_p(k)$ estimator with a score of 70.1. The third lowest score of 74.0 corresponds to the BIKB estimator, followed closely by BIW(LAD) at 74.9. Of the unbounded estimators, Tukey's biweight performs the best with a 'score' of 75.4. The choice of comparison criterion is somewhat arbitrary as it is clear that a different ranking of the estimators could be achieved by using some other combination of reliability and goodness-of-fit.

⁹ In order to check that the results were not an artifact of taking alternate months, similar estimates of reliability and goodness-of-fit were calculated for data in which alternate 12-months of data were used, i.e. the one series of data comprised the monthly returns of years 1, 3, 5, 7 and 9 while the other comprised the monthly returns of the even years. The comparative figures are given in Appendix 6.A and are consistent with the findings reported in this section.

Table 6.4

Relative percentage of reliability/goodness-of-fit measure to that of value of unadjusted OLS

Estimator	1978-1987		1983-1992		Average
	MSE %	IQR %	MSE %	IQR %	
OLS	100	100	100	100	100
LAD	64.9	96.1	73.9	96.8	82.9
KB	63.1	97.8	76.1	98.2	83.8
W(OLS)	57.5	98.1	59.5	98.3	78.4
W(LAD)	58.7	98.0	57.6	98.3	78.2
TBW	53.0	97.0	54.4	97.2	75.4
$L_p(k)$	56.4	97.5	59.2	98.0	77.8
$L_p(\omega)$	56.7	97.7	62.5	98.0	78.7
Gini	62.5	97.6	55.5	97.5	78.3
BIOLS	69.2	92.7	83.9	92.4	84.6
BILAD	63.6	96.0	68.8	97.0	81.4
BIKB	52.6	97.1	48.5	97.9	74.0
BIW(LAD)	53.7	97.6	49.9	98.5	74.9
BL $_p(k)$	54.2	87.1	50.9	88.3	70.1
BL $_p(\omega)$	52.4	83.6	49.9	84.2	67.5

Nevertheless, the marked difference between the bounded influence robust estimators suggests that they should be used in preference to the usual robust or least squares estimators. Further, the above analysis confirms that the L_p -norm estimators appear to be the best of bounded influence, robust estimators.

In order to establish the source of the improvement in MSE the MSEs were decomposed in the fashion suggested by Theil (1971) and described earlier in this section. The decomposition was effected for the MSEs of the bounded influence versions of the L_p -norm, the Koenker-Basset and biweight estimators, because they had been assessed as superior in the analysis of total MSEs and goodness-of-fit. The unbounded OLS estimator is included in Table 6.5 as a yardstick for comparison purposes.

From Table 6.5 it is apparent that all the robust estimators display a significant reduction in the inefficiency (by nearly two-thirds) and the random (by nearly a half) components of the MSEs. For example, consider the inefficiency components in 1983-1992 given in Table 6.5: the inefficiency of all the robust estimators (other than the BIOLS) are approximately 0.01 compared with 0.0336 for the OLS estimator. All the randomness

Table 6.5

Decomposed mean squared errors measuring reliability of the estimators

1978-1987				
	bias	inefficiency	randomness	total
OLS	.0176	.0426	.1286	.1889
BIOLS	.0080	.0233	.0995	.1308
BIW(LAD)	.0053	.0081	.0893	.1028
BIKB	.0033	.0089	.0898	.1020
TBW	.0041	.0190	.0770	.1001
$BIL_p(k)$	.0046	.0177	.0800	.1023
$BIL_p(\omega)$	.0040	.0184	.0766	.0990
1983-1992				
OLS	.0001	.0336	.1290	.1626
BIOLS	.0009	.0230	.1126	.1365
BIW(LAD)	.0024	.0098	.0689	.0811
BIKB	.0021	.0099	.0680	.0788
TBW	.0013	.0118	.0753	.0885
$BIL_p(k)$	.0013	.0095	.0721	.0828
$BIL_p(\omega)$	.0010	.0102	.0700	.0811

components of the robust estimators in 1983-1992 can be seen to be about 0.07 compared with the OLS randomness component of 0.1290. Examining the results for the bounded influence OLS estimator (BIOLS) it appears as if bounding the influence helps more in reducing the inefficiency (i.e. a drop of 32% from 0.0336 for the unbounded OLS to 0.023) than it does in reducing the randomness (i.e. a drop of only 13% from 0.129 to 0.1126). The reduction of the randomness component can therefore be seen to be a result of down-weighting the outlying residuals values. Apart from the aforementioned comments about the BIOLS, the robust estimators display similar characteristics in terms of the component parts of the total MSEs.

The results of this sub-section have shown that substantial improvements in efficiency can be achieved in beta estimation by using robust, bounded influence regression techniques. In particular, the L_p -norm estimators were assessed to be superior on the joint criteria of internal reliability and goodness-of-fit. As a check on the results a further investigation was performed. Given the Gauss-Markov theorem, OLS estimators should be more reliable and efficient than robust estimators when the security's residuals are normally distributed. In the next sub-section, therefore, the validity of this expectation is tested by comparing the MSEs

of the OLS and robust estimators on securities which have been classified as normal, or non-normal.

6.3.5 Reliability versus level of non-normality

In the previous section the increased reliability of the robust estimators over the more sensitive OLS estimator was demonstrated. In this sub-section the investigation into reliability of robust estimators is aimed specifically at establishing whether the improvement is more significant for the shares which exhibit significant non-normality. To establish whether this is the case, the total MSEs of the estimates were contrasted for those securities which had significant non-normality and those for which no significant non-normality could be detected.

The shares were assigned to four (possibly overlapping) categories based on their residual distributions: (1) significant coefficients of kurtosis (at a 5% level, one-tailed because leptokurtosis was *a priori* expected) in either of the alternate data series; (2) significant skewness (5% level, two-tailed) in either of the alternate data series; (3) significant non-normality determined by the D'Agostino and Pearson (1973) statistic¹⁰ at a 5% (one sided) level; (4) normal shares which were not significant by the criterion in (3).

The total MSEs for the categories of shares are reported in Table 6.6. From Table 6.6 it is evident that all the robust estimators perform consistently across all four categories. For example, consider the $BIL_p(\omega)$ estimator in the period 1978-1987: the MSE for securities which have significant kurtosis is 0.1010; the MSE for securities with significant skewness is similar at 0.1004; for securities which were significantly non-normal based on the D'Agostino and Pearson test, the MSE is 0.1028 and for the normal shares, the MSE drops slightly to 0.0922.

On the other hand, the MSEs in Table 6.6 for the OLS estimator are very high for all the non-normal securities (all above 0.2) but drop to 0.1015 (for 1978-1987) and 0.0671 (for 1983-1992) for the normal shares. The MSEs for the normal shares are very similar to the MSEs for the robust estimators as anticipated. The fact that the OLS estimator does not have a lower MSE than the robust estimators can be explained by the possible mis-categorization of the securities into the normal and non-normal classes. For example, the test of normality may not have enough power to exclude *all* the non-normal shares from the normal category and hence the efficiency of the OLS estimator is compromised. Furthermore, no account has been

¹⁰ A chi-square statistic with two degrees of freedom calculated by summing the squares of the normal deviates for skewness and kurtosis.

Table 6.6

Comparison (using MSEs) of robust and least squares estimators for normal and non-normal distributions

	1978-1987				1983-1992			
	Lepto ^a	Skew ^b	Non-norm ^c	Norm.	Lepto	Skew	Non-norm	Norm
OLS	.2326	.2497	.2377	.1015	.2029	.2223	.2083	.0671
BIW(LAD)	.1039	.1037	.1069	.0918	.0882	.0964	.0925	.0572
BIKB	.1038	.1011	.1067	.0862	.0860	.0931	.0898	.0557
TBW	.1021	.0993	.1032	.0945	.0976	.1069	.1036	.0568
$BIL_p(k)$	.1068	.1024	.1085	.0911	.0867	.0933	.0893	.0693
$BIL_p(\omega)$	.1010	.1004	.1028	.0922	.0873	.0931	.0899	.0627

^a Leptokurtic share residuals. Coefficient of kurtosis significant ($p < 5\%$, one-tailed). (1978-1987: 193 shares; 1983-1992: 188 shares.)

^b Skewed share residuals. Coefficient of skewness significant ($p < 5\%$, two-tailed). (1978-1987: 165 shares; 1983-1992: 150 shares.)

^c Non-normal residual distributions. D'Agostino and Pearson (1973) test ($p < 5\%$, one-sided). (1978-1987: 184 shares; 1983-1992: 180 shares.)

taken of the non-normality of the market returns in the OLS estimator.

The results of this section have confirmed that the superiority of the robust estimators is due to non-normality in the return distributions. Furthermore, it was shown that even when the securities were classified as normal, the robust estimators were as efficient as the OLS estimator. Consequently, robust estimators are recommended for use in estimating the betas of all shares, whether or not they are classified as having normal residual distributions.

In the next sub-section, the actual differences in the estimates of beta coefficients produced by robust estimators and the OLS estimator are examined. The purpose of the investigation is to see whether the non-normality of the residuals leads to any significant differences between the OLS and robust estimates of the betas.

6.3.6 Differences between actual beta estimates with respect to levels of non-normality

The investigation in this section deals with significant differences in robust and OLS estimates of beta. Table 6.7 gives the mean differences between the betas of the robust estimators and the OLS beta. For each robust estimator, eight mean differences from the OLS estimates are reported. Each mean difference corresponds to one of two periods (1978-1987 or 1983-1992) as well as to one of the two alternating series of returns calculated for each share in each period and, finally, to whether the share was classified as having a normal or non-

normal residual distribution. Significant non-normality in their residuals was assessed using the D'Agostino and Pearson (1973) test at a 5% level.

By comparing the average differences between the robust and OLS betas in each of the four 'panels' in Table 6.7, it is clear that there are larger differences in the estimates of beta for shares which have significantly non-normal residuals than for shares with normally

Table 6.7

Mean differences between robust estimates of beta and OLS estimates.

	1983-1992		1978-1987	
	Non-normal	Normal	Non-normal	Normal
First set of alternating returns				
BIOLS	-.033 (-3.01)	.007 (.86)	-.044 (-4.52)	-.012 (-1.47)
BIW(LAD)	-.072 (-3.98)	-.016 (-1.17)	-.111 (-6.88)	-.028 (-2.25)
BIKB	-.081 (-4.14)	-.024 (-1.59)	-.128 (-7.36)	-.034 (-2.43)
TBW	-.096 (-4.89)	-.036 (-2.68)	-.132 (-7.65)	-.032 (-3.36)
$BIL_p(k)$	-.092 (-5.14)	.004 (.32)	-.116 (-7.39)	-.025 (-2.09)
$BIL_p(\omega)$	-.078 (-4.85)	-.006 (-.44)	-.106 (-7.22)	-.030 (-2.49)
Second set of alternating returns				
BIOLS	.000 (.00)	.006 (.48)	.016 (1.13)	.002 (.19)
BIW(LAD)	-.018 (-.95)	-.000 (-.00)	-.002 (-.09)	-.014 (-1.10)
BIKB	-.030 (-1.51)	-.014 (-.81)	-.009 (-.40)	-.021 (-1.52)
TBW	-.062 (-3.50)	-.017 (-1.41)	-.024 (-1.03)	-.034 (-2.79)
$BIL_p(k)$	-.045 (-2.42)	-.007 (-.47)	-.020 (-.89)	-.015 (-1.15)
$BIL_p(\omega)$	-.040 (-2.36)	-.013 (-.87)	-.004 (-.20)	-.018 (-1.40)

t-statistics are given in parentheses.

BIOLS is the bounded influence OLS estimator.

In each period there are two figures for each estimator based on the two alternating series of returns.

distributed residuals. For example, the mean difference in betas estimated using the $BIL_p(\omega)$ estimator and OLS is -0.078 in the 1983-1992 period (using the first set of alternating returns)

for the non-normal shares compared with a mean difference in the betas of -0.006 for the normal shares. The exception to the aforementioned 'rule' is the panel in Table 6.7 corresponding to second set of alternating returns in the 1978-1987 period. In that panel, there is only one significant mean difference (the Tukey Biweight (TBW) estimator with a mean difference of -0.034). In most instances, though, it appears as if the OLS estimator overestimates the betas when the share residuals are non-normally distributed.

The conclusion drawn from the results in Table 6.7 is that not only will robust estimators give more reliable estimates of beta, but also that the actual estimates of beta will differ from the OLS estimates.

In section 6.4 the stability of the betas is assessed using the robust estimators. The assessment is based on a comparison of the 'forecast' MSEs (i.e. the MSEs between estimates in successive periods) and 'internal reliability' MSEs (i.e. the MSEs between the estimates calculated from alternating data sets). It can be appreciated that if the forecast MSEs are higher than the reliability MSEs then there is evidence that the beta of the security has changed.

6.4 Stability of betas on the JSE

The issue of the stability of security betas over time is fundamental to the use of historical estimates of beta. Various tests of beta stability have been performed using OLS estimates of beta on stock exchanges world-wide (see, for example, Dimson and Marsh (1983) on the London Stock Exchange, Blume (1975) and Sharpe and Cooper (1972) on the NYSE, Altman, Jacquillat and Levasseur (1974) on the Paris Bourse, Chapter 5 of this thesis for results from the JSE and Hawawini and Michel (1979) in Belgium) and most have indicated moderate stability. As in the discussion in Chapter 5, it should be noted that securities will only have stable beta coefficients if their fundamental structure and market profile remain constant. The measures of stability therefore indicate how useful historical estimates of beta are in predicting future betas, given not only the sampling error in the estimates, but also that the true, underlying beta may have changed.

An important consideration in testing beta stability is that the true, underlying betas of securities are not observable and hence inference about their stability is only obtainable from estimates of beta. Dimson and Marsh (1983) provide theoretical evidence that biased estimates of beta can lead to spurious inferences about the stability of betas over time. They show that

the correlation between estimates of beta in successive estimation periods is given (under certain assumptions) by:

$$\rho' = \{\rho\sigma_\beta^2 + \rho^*\sigma_\epsilon^2\} / \{\sigma_\beta^2 + \sigma_\epsilon^2\} \quad (4)$$

where ρ' is the correlation between the estimates, ρ is the correlation between the actual betas, σ_β^2 is the cross-sectional variance of betas in the sample of securities, ρ^* is the correlation between the standard errors in the estimates of beta between periods, and σ_ϵ^2 is the variance of the error terms of estimates of beta.

In their article Dimson and Marsh (1983) illustrate how, for example, thin trading biases estimates of beta in successive periods thus making ρ^* larger than zero. Non-zero ρ^* makes ρ' an upwardly biased estimator of ρ . As a consequence they conclude that any study of beta stability must include a thin trading correction procedure in the estimation of beta.

It can be verified from equation (4) that the effect that a non-zero ρ^* has on the accuracy of ρ' as an estimator of ρ is directly proportional to the size of σ_ϵ^2 (which represents the inefficiency of the estimate of β). Consequently, even a slightly biased estimator which is very efficient (and hence has a small σ_ϵ^2) will yield an accurate assessment of the stability of the true betas. It was pointed out in chapters 4 and 5 of this thesis that the trade-to-trade estimator may be slightly biased because it does not take into account any price adjustment delays other than thin trading. However, if the efficiency of the trade-to-trade estimator is improved significantly by using robust estimation techniques then it can still yield useful information about the stability of beta coefficients.

To date there has been no major study on the stability of beta coefficients using robust estimates of beta. It is argued that substantial improvements in the assessments of stability (particularly on small stock markets) can be achieved by using the more efficient estimators discussed in this chapter. In the sub-section that follows, the evidence of stability of beta coefficients using the robust estimators is presented for the JSE.

6.4.1 Empirical assessment of stability of betas

In this sub-section the robust estimates of beta calculated using the bounded influence L_p -norm estimator are used to test for beta stability. The inferences concerning stability are

made on the basis of a comparison between the MSEs of the estimators using the alternate data series and the MSEs which were computed for the estimates of beta calculated over the two non-overlapping five-year periods in each of the data sets (i.e. in each period 1978-1987 and 1983-1992). In other words, the forecast MSE was compared with the reliability MSE. An increase in the forecast MSE over the reliability MSE indicates that there is less stability over consecutive periods than there is in the estimates calculated from the alternating series of data.

The MSEs of the bounded influence L_p -norm estimators are given in Table 6.8 along with the MSEs of the bounded influence OLS estimator as a comparative measure. From the last column of Table 6.8 it can be seen that the MSEs of the estimators are much larger for the consecutive five-year periods than for the alternate data series, particularly for the 1983-1992

Table 6.8

MSEs for betas over consecutive five-year periods

	ρ	bias	1978-1987		
			inefficiency	randomness	total
BIOLS	.68	.0038	.0134	.1236	.1408
$BIL_p(k)$	.75	.0017	.0103	.1059	.1180
$BIL_p(\omega)$	.74	.0022	.0123	.1074	.1219
	ρ	bias	1983-1992		
			inefficiency	randomness	total
BIOLS	.52	.0000	.0257	.1550	.1807
$BIL_p(k)$	.58	.0004	.0165	.1481	.1650
$BIL_p(\omega)$	.59	.0002	.0150	.1452	.1604

MSEs for betas in alternating data series

	ρ	bias	1978-1987		
			inefficiency	randomness	total
BIOLS	.72	.0080	.0233	.0995	.1308
$BIL_p(k)$	.79	.0046	.0177	.0800	.1023
$BIL_p(\omega)$	.79	.0040	.0184	.0766	.0990
	ρ	bias	1983-1992		
			inefficiency	randomness	total
BIOLS	.62	.0009	.0230	.1126	.1365
$BIL_p(k)$	.77	.0013	.0095	.0721	.0828
$BIL_p(\omega)$	.77	.0010	.0102	.0700	.0811

data set. For example, in the 1983-1992 data set, the forecast MSE for the $BIL_p(\omega)$ estimator

is 0.1604 compared with a much smaller reliability MSE of 0.0811. The reliability MSE gives an indication of the expected value of the MSE given that the betas are stable. The increased value of the forecast MSE is therefore due to instability of the betas.

In the column labelled 'bias' in Table 6.8, it is clear that the bias component of the MSE remains very small indicating, as might be expected, that there is no substantial change in the overall average beta estimate for securities in the market over time. The inefficiency component is also approximately the same for the forecast MSEs and reliability MSEs, e.g. for the $BIL_p(\omega)$ estimator, the inefficiency component increases from 0.0102 for the reliability MSE in 1983-1992 to 0.0150 for the forecast MSE, but decreases from a 0.0184 to 0.0123 in 1978-1987. The evidence from Table 6.8 is that the largest absolute increase in MSE arises in the randomness component. The increase is due to that fact that the relative order and spread of the beta estimates among the securities has changed more from one five-year period to the next than from one alternate data series to the other.

The comparison between the two sets of MSEs, namely the reliability and forecast, has permitted greater insight into the stability of betas. On the JSE the evidence suggests that the underlying betas do indeed change over time.

6.5 Concluding remarks and suggestions for beta estimation

This chapter has examined the need for and performance of robust estimators on the JSE. In particular the efficiencies of a wide range of robust estimators along with their goodness-of-fit characteristics were assessed. It was shown that the robustness required for estimating betas involved down-weighting not only the outlying residual values, but also the outlying market returns.

Another aspect of the non-normality of a share's residuals which plays an important role is that the degree of non-normality varies substantially from sample to sample. Furthermore, the changing levels of kurtosis were found to have significant correlations with trading frequency, with the more frequently traded shares generally having smaller deviations from the normal distribution. This finding confirmed the hypothesis that non-normality is likely to be a relatively more severe problem on the smaller markets. The variability in the extent of non-normality from sample to sample also suggested that an adaptive robust procedure would produce more efficient estimates.

From the studies of efficiency (internal reliability) and goodness-of-fit, the *bounded influence* L_p -norm estimator (with adaptive parameter selection) emerged as the most suitable estimator on the JSE. The L_p -norm estimator involves a parameter choice which determines the robustness properties of the estimator. Two procedures for choosing the parameter were considered: the first was an adaptive procedure based on the sample kurtosis of a preliminary fit to the data; the second was based on minimizing the asymptotic variance of the L_p -norm estimator. The empirical results indicated that the second procedure provided marginally more efficient estimates. Two other robust estimators which fared almost as well as the L_p -norm estimator were the *bounded influence* Koenker-Basset and Welsh estimators. It was also found that Tukey's biweight M-estimator performed surprisingly well considering its simplicity and ease of computation; although it was clearly not as effective as the L_p -norm estimators.

It was also clearly demonstrated that OLS estimators do indeed provide far less efficient estimates of beta than their robust counterparts when the securities have non-normal residual distributions. When the securities were categorized as having normal residual distributions, the OLS estimator performed nearly as well as the robust estimators. The fact that the robust estimators (and in particular the L_p -norm estimator) offered superior efficiency not only for the non-normal securities, but also a marginal improvement for the normal securities, highlighted the importance of using robust procedures for estimating betas.

Using the bounded influence L_p -estimator, further evidence of instability in the betas of securities on the JSE was detected by comparing the predictability of betas (measured by their forecast MSEs) over consecutive periods with the reliability MSEs of betas measured using alternate data series over the same period.

Appendix 6.A

Reliability and goodness-of-fit using data from alternate years

Table 6A.1

Reliability and goodness-of-fit measures using data series from alternate years								
	1978-1987				1983-1992			
	ρ	MSE	IQR	MAR	ρ	MSE	IQR	MAR
BIOLS	.74	.1084	.1097	.0545	.56	.1633	.1105	.0555
$BL_p(k)$	.81	.0891	.1031	.0498	.66	.1296	.1057	.0521
$BL_p(\omega)$	.81	.0876	.0993	.0483	.66	.1298	.1013	.0500

Appendix 6.B**Summary of Robust Regression Methods**

In this section a wide variety of robust regression techniques which are appropriate for stock exchange data are described in some detail. It is a technical appendix which provides a theoretical backdrop to the methodologies which are applied to the JSE data in the main body of the chapter. While robustness as a whole and robust regressions in particular are more than adequately described in many sources (see Chan and Lakonishok (1992), Gonin and Money (1989), and Koenker (1982) to name but three) it is useful to have a reasonable overview of the robust regression estimators that were assessed in the empirical analysis of Section 6.4.

Firstly, though, the evidence surrounding the assumption of normality in stock returns on the JSE is examined more closely. The distributions of stock market returns (or log price relatives) have been widely studied, both in the larger stock markets and on the JSE. Bachelier (1900) initiated research in the area with his theory of random walks of security prices. His work concentrated mainly on the distributions of actual price changes rather than log price relatives but the assumptions underpinning his theory made it applicable to simple returns (which are usually very close to log price relatives). Osborne (1959) independently derived Bachelier's model using log price relatives and showed that the distribution function was Gaussian under the assumption of a large number of transactions in the security in the interval in which the log price relative was measured.

Fama (1965) produced considerable empirical evidence from the NYSE that daily stock price returns were leptokurtic or heavy tailed. Osborne (1959), Fama (1965) and Teichmoeller (1971) suggested the use of a simple mixture of normal distributions as appropriate for stock returns but all reported unsatisfactory fits to the data. Mandelbrot (1963), on the other hand, fostered the concept that the security returns were observations from the family of stable Paretian distributions. Consequently Fama (1965), Blume (1970) and Roll (1968) produced some findings suggesting that the stable Paretian family could provide a better fit than the normal distribution.

Praetz (1972) continued the investigation into the distribution of share returns by considering several candidate distributions. In particular Praetz tested the fit of the scaled t distribution, the normal distribution, the compound events distribution (see Press (1967)) and the stable Paretian distribution to returns on share price indices. He showed that the scaled

t distribution showed a superior fit as judged by the Pearson chi-square test. Blattberg and Gonedes (1974) also demonstrated that the scaled t distribution provided a better fit for individual shares than the stable Paretian.

On the JSE, Affleck-Graves (1974) has attempted to fit the stable Paretian to share returns and concludes that it appeared suitable for the small sample of shares used. He also pointed out that in most cases in the literature only small samples of shares have been used in estimating and testing the fits of distributions. Schlosberg (1976) conducted extensive research into the distributions of a small sample of shares on the JSE and concluded that either the scaled t distribution or the compound normal distribution (see Sichel (1973)) was most appropriate. He also noted that most of the research into security returns has concentrated on symmetric distributions or the symmetric forms of the other distributions fitted.

Types of robust regression

Until Huber (1981) and others developed the formal theory of robust estimators, they were regarded as inferior to conventional estimators. Researchers have been loath to depart from the comfort of least squares regression because of the implications of the Gauss-Markov theorem. For reasons of computational simplicity and a belief that BLUness was *the* criterion on which to compare estimators, robust regression has received little attention, particularly in the literature on empirical finance.

Robustness is often seen as an excuse to ignore some of the observations. Early robust estimators did in fact exclude some of the observations from the analysis, but the reasons for doing so were, again, computational simplicity. Robust estimators do not in fact exclude observations but rather build in the expectation that there will be outliers in the data set. Because outliers are expected in robust approaches to regression they are down-weighted *relative* to least squares methods which do not anticipate outliers and hence accord them undue weight. The exclusion of outlying values in early robust estimators corresponded to an approximation of the down-weighting factor.

Robust estimation methods have conventionally been categorised into M-estimators, L-estimators and R-estimators. Bayesian techniques might also be considered 'robust' but the underlying philosophy is quite different.

Broadly speaking M-estimators encompass a wide range of maximum likelihood type methods, ranging from L_p -norm estimation to partially adaptive maximum likelihood. Outliers

are modelled by using distributions which would account for heavier tails than the normal distribution and which hence down-weight their influence.

L-estimators are based on ranked data, for instance winsorized and trimmed means and medians are all L-estimators of location. Koenker and Basset (1978) overcame the apparent problem in applying L-estimators to regression analysis by defining regression quantiles as the solution of a minimization procedure. The quantiles could then be used to define L-type estimators, in particular regression trimmed means.

The estimates known as R-estimates are based on rank tests. For example the Wilcoxon test leads to the Hodges-Lehmann estimate of location, the median of the set $\{\frac{1}{2}(x_i + x_j), \text{ for all } i \text{ and } j\}$.

M-estimators

M-estimators can be further categorized based on their derivations. The categories that are used here are (1) traditional maximum likelihood estimators (MLEs), (2) MLE-type estimators and quasi-MLEs and (3) L_p -norm estimators. Before looking at these in more detail, a brief review the underlying concepts pertinent to M-estimation is given.

Maximum likelihood estimation of a vector of parameters θ using data points (y_i, x_i) is attained by minimizing over θ the expression:

$$\sum_{i=1}^n \rho(y_i, y\{x_i; \theta\}) \quad (5)$$

where the function ρ is the negative logarithm of the probability density and $y\{x_i; \theta\}$ represents the predicted value of y_i based on the vector of covariates x_i and the parameters θ .

In the case of the simple market model, y_i is the return on a security in interval i , x_i are the returns on the market at a time i , n is the number of intervals over which returns are

measured and $\theta = (\alpha, \beta)$. It can be noted that in cases such as this where the function ρ depends on the (appropriately scaled) difference between the measured y_i and the predicted $y\{x_i; \alpha, \beta\}$, i.e. the residuals from the model, and not on the two components separately, a *local* M-estimate of the parameters can be found using the formulation:

$$\text{minimize over } \alpha \text{ and } \beta \quad \sum_{i=1}^n \rho \left(\frac{e_i\{\alpha, \beta\}}{\sigma_i} \right) \quad (6)$$

where ρ is now a function of the single variable, the residual e_i scaled by some weight factor σ_i .

The function, ψ , is defined to be the derivative of ρ with respect to the size of the scaled residual, i.e.

$$\psi(z) = \frac{d\rho(z)}{dz} \quad (7)$$

where $z \equiv (y_i - \alpha - \beta x_i)/\sigma_i$, i.e. the standardized residual. Clearly then the maximum likelihood estimates comes from the solution of the two equations:

$$0 = \sum_{i=1}^n \frac{1}{\sigma_i} \psi \left(\frac{e_i}{\sigma_i} \right) \quad (8)$$

and

$$0 = \sum_{i=1}^n \frac{1}{\sigma_i} \psi \left(\frac{e_i}{\sigma_i} \right) x_i \quad (9)$$

Solutions can be found either by solving the two equations simultaneously or by minimizing the single equation (6).

The influence function (which is proportional to ψ) has been the subject of much investigation which and again examined briefly in the next section of this appendix. It can, for some probability distributions, be regarded as the indicator of sensitivity to outliers. Residuals which produce a large value of $\psi(z)$ will dominate the minimization process (or solution of the two simultaneous equations) and hence have a large influence on the parameter estimates. While this is not necessarily a bad thing it must be recognized that should one data point be 'out of line' with the others and also have a large influence then it will lead to inappropriate

estimates. For example, the normal distribution has $\rho(z) = z^2/2$ and $\psi(z) = z$ and thus the influence of an observation on the estimation procedure increases in direct proportion with the size of the residual. The implication is that extreme outliers which are not typical of the data and which have large influence will cause inefficiency in the estimation process.

The influence function

The major concern in parameter estimation using stock exchange data is the leptokurtosis (heavy tails) of the distribution of residuals and, to a lesser extent, returns. Outliers may therefore be expected in a set of data. M-estimators which have monotonic increasing influence functions, e.g. the normal distribution, will therefore be inefficient when applied to stock market data. If the application of 'pure' M-estimation is desired then a distribution has to be found which fits the residuals and allows for heavy tails by having a bounded or redescending influence function.

The influence function is formally defined (e.g. Hampel, 1974) to be:

$$IC(z) = \frac{\psi(z)}{E[\psi(z)]} \quad (10)$$

Since $E[\psi(z)]$ is constant IC will be directly proportional to ψ , the latter function is usually examined to establish the influence of outliers in a sample on the estimates. In the case of true MLEs ψ is found directly by differentiating the negative logarithm of the probability density which is assumed for the distribution of the residuals. In other cases the shape of the influence curve (or the form of robustness) can be prescribed and can be used to work back to find a suitable distribution for the residuals. For example, it might be desirable to have an influence function which is first of all, close to that of the normal for small residuals (i.e. when the residuals are small, a residual slightly further from zero will have more influence on the estimates than one very close to zero); secondly, which is bounded for large residuals (i.e. after the size of the residuals exceeds a specified value the influence does not increase) and which, thirdly, redescends, i.e. for extreme outliers their influence approaches zero.

It is also entirely feasible to specify the shape of the ψ and hence find a ρ which does not correspond to a well-known distribution density at all. Many authors (see Hampel, 1974) have specified influence functions which are not necessarily derived from specified distributions.

Based on the above theory the three types of M-estimator defined earlier are discussed in more detail.

(1) MLEs

In dealing with stock market data it is well known that the returns differ from variates following a normal distribution chiefly because of the relatively higher number of values far from the mean. There are several common distributions which are suitable for modelling these directly.

Affleck-Graves (1974) provides a thorough review of the Stable distributions. Stable distributions have been widely cited as the most logical family of distributions to consider (they include for example the Cauchy and normal distributions as special cases). They also represent the only possible limiting distributions for sums of independent, identically distributed random variables. The concept of sums of random variables corresponds with the notion that the returns on a share are made up of the sums of returns from one transaction to the next. The family is defined in terms of its characteristic function:

$$\log \phi(t) = i\delta t - \gamma |t|^\alpha (1 + i\beta \frac{t}{|t|} \omega(t, \alpha)) \quad (11)$$

where

$$\omega(t, \alpha) = \begin{cases} \tan \frac{\pi \alpha}{2} & \text{if } \alpha \neq 1 \\ \frac{2}{\pi} \log |t| & \text{if } \alpha = 1 \end{cases} \quad (12)$$

and $i = \sqrt{-1}$.

The parameters in the defining function can rather loosely be regarded as reflecting certain characteristics of the shape of the particular member of the Stable family: δ - location (ranges from $-\infty$ to ∞); γ - scale (positive); β - skewness (ranges from -1 to 1, with symmetric distributions at 0); α - peakedness (characteristic exponent which ranges from strictly greater than 0 to 2, inclusive). Some interesting properties of the Stable family (proofs of which can be found in Levy (1925) or Gnedenko and Kolmogorov (1954)) include:

- (a) for $\alpha = 1$ and $\beta = 0$ the Stable distribution corresponds to the Cauchy (Lorentzian) distribution;

tails than the normal even though they are exponentially decreasing. The functions of interest in this instance are:

$$\rho(z) = |z| \quad (15)$$

and

$$\psi(z) = \text{sign}(z) \quad (16)$$

The influence function in this case is a discontinuous step function, with all residuals being assigned the same weight.

Other distributions which represent some sort of compromise between the sensitive normal and the Cauchy include Student's t -distribution with some low number of degrees of freedom, and in fact all the Stable distributions with values of α between 1 and 2. In the case of the t -distribution it should also be noted that it is equivalent to the Cauchy distribution when the number of degrees of freedom is equal to one.

(2) Quasi-MLEs and MLE-type estimators

By quasi-MLEs it is meant that the parameters which are not of interest (in the case of regression, all the 'nuisance' parameters such as the residual standard deviation, measures of skewness, etc.) are estimated separately from the coefficients. The coefficients of interest are estimated using maximum likelihood conditional on adaptive estimates of the 'nuisance' parameters. In the case of the Stable family the parameters can be estimated by some procedure using the residuals from a preliminary model, based on an arbitrary set of parameters. Once estimates of the parameters have been obtained either the ρ or ψ functions can be used to estimate the regression coefficients.

For some members of the Stable family, McDonald and Newey (1988) have shown that asymptotic efficiency still holds provided that the distributional parameters are chosen to minimize the estimator of the asymptotic covariance matrix of the regression coefficients. McDonald and Newey (1988) and subsequently McDonald and Nelson (1989) promote the use of a partially adaptive generalised t -distribution (GT) in the estimation of regression coefficients. The GT is a very flexible range of distributions which include as special cases the Cauchy, normal, t , double exponential (Laplace) and Box-Tiao distributions. What is further

attractive about this family is that the density can easily be written down in closed form (as opposed to most of the Stable family) and the ρ and ψ functions are easily derivable. The GT family overlaps much of the Stable family, particularly in the range of distributions which might be considered suitable for the distributional form of the market model residuals. McDonald and Newey (1988) also prove asymptotic efficiency of M-estimators based on the GT distributions.

The density of a GT random variate is specified as:

$$f(z; \sigma, p, q) = \frac{p}{2\sigma q^{\frac{1}{p}} B(\frac{1}{p}, q) \left[1 + \frac{|z|^p}{q\sigma^p} \right]^{\frac{p+1}{p}}} \quad -\infty < z < \infty \quad (17)$$

and

$$\psi(z) = \frac{(pq + 1) \text{sign}(z) |z|^{p-1}}{q\sigma^p + |z|^p} \quad (18)$$

It is easily seen that ψ is bounded and redescending and hence fulfils the conditions which might be regarded as ideal for stock market data. The GT distribution encompasses least squares estimation which is equivalent to maximum likelihood using the normal distribution. As q becomes large and $p = 2$ so GT becomes a normal distribution and so GT will yield OLS results. For q large and p equal to unity, GT approaches the double exponential distribution. Maximum likelihood in this case is equivalent to the LAD estimator. In fact for all p fixed and q very large the GT MLEs correspond to L_p estimators which are examined more closely later in this appendix.

McDonald and Newey (1988) proposed a method for providing quasi-MLE using the GT distributions where p , q and σ are estimated from the residuals of a preliminary model. Because of its very general applicability, the method relies heavily on efficient numerical algorithms for the estimation procedures. The details of MacDonald and Newey's proposal are set out in Appendix 6.C. In this chapter the focus has been on a further subset of the GT family, the Box-Tiao family. The M-estimators of this family of distributions correspond to the L_p -norm estimators which are described below.

Other MLE type estimators have been developed by specifying only the ψ function, rather than a specific distribution for the variable. In this chapter Tukey's biweight is considered. It prescribes a ψ function of:

$$\psi(z) = \begin{cases} z(1 - z^2/c^2)^2 & |z| \leq c \\ 0 & |z| > c \end{cases} \quad (19)$$

(3) L_p -norm estimation

The linear L_p -norm estimation problem is defined as finding the parameter vector, i.e. the α and β from the linear model $y_i = \alpha + \beta x_i + z_i$, such that the following function is minimised:

$$\sum_{i=1}^n |z_i|^p \quad (20)$$

The procedure is identical to MLE with $\rho(z) = |z|^p$ (which in turn corresponds to distributions from the GT family when q tends to infinity). The special case $p = 1$ is equivalent to LAD estimation. Apart from OLS, LAD has certainly been the most commonly applied estimation technique for beta estimation on the stock market and in other econometric studies. Blattberg and Sargent (1971), Fama (1965), Mandelbrot (1971), Sims (1971), Zeckhauser and Thompson (1970), Smith and Hall (1972) and Coursey and Nyquist (1983) among others have all applied LAD to econometric data.

Blattberg and Sargent (1971) also considered the symmetric subset of the Stable family allowing α , the characteristic exponent, to vary between 1 and 2. They performed a simulation study, generating data from the Stable family with various characteristic exponents and then compared the efficiency of the LAD, OLS and various other minimum variance estimators (proposed by Wise (1966) and also based on the Stable family). Their conclusions indicated the relatively superiority of the LAD estimator for a wide variety of heavier than normal tails and suggested in particular that LAD be used whenever α is expected to be less than 1.5. They also propose a two step (which could easily be extended into an iterative) procedure for deciding on which estimator to use: OLS (or some other preliminary method) is used to generate residuals for which the most likely characteristic exponent is estimated; thereafter LAD is used if the estimate is close to one, OLS if it is close to two or a minimum dispersion estimator if it is between one and two.

Smith and Hall (1971) investigated the performance of LAD and OLS with residuals generated using a double exponential distribution which implied that although OLS remained the BLU estimator, the LAD was in fact the MLE. Their results indicated that maximum likelihood showed superiority over BLU as a criterion for parameter estimation.

Coursey and Nyquist (1983) extended the seemingly endless stream of acronyms for robust estimators introducing GLAK, GLAU, GLSU and GLSK. These were designed to compare least squares and least absolute errors methods in the face of serial correlation. They also concluded that for fat tailed distributions, the LAD estimator is better than the least squares (and vice versa for more normal distributions); but they pointed out that the LAD estimator has less sensitivity to any true dependency structure in the error terms.

The other special case of L_p -norm estimators which has found widespread appeal in fields other than econometrics is the Chebychev approximation or L_∞ estimator. It is a minimax procedure; the parameters are selected so that the largest residual is minimised. L_∞ estimation is clearly not appropriate to stock market data given that they are subject to genuine outliers which would then dominate the fitting.

For the general case the choice of p has been the subject of wide ranging debate. Most of the research has centred on p in the range between one and two, i.e. allowing for more outliers than the normal but not more than the double exponential distribution. Forsythe (1972), Ekblom (1974), Harter (1977), Money *et al* (1982), Sposito *et al* (1983) and Barr (1980) have all come up with suggestions as to how to select p . The latter three have all proposed adaptive schemes based on the kurtosis of the residuals from preliminary estimates of the regression parameters. For example Barr suggested:

$$p = 1 + \frac{9}{b_2} \quad (21)$$

where b_2 is the coefficient of kurtosis of the residuals. Note that the 9 in equation (21) is the square of the kurtosis of a normal random variate. Nyquist (1980, 1983) proposed an optimal p based on the asymptotic variance of the L_p -norm estimators. In the empirical study described in the main body of the chapter, one-step versions of the two methods have been implemented: initially the residuals were calculated using a $L_{1.5}$ -norm estimator (which was chosen as being reasonably robust). The b_2 of these residuals was then calculated and used in equation (21). The second p was chosen to be that p (from the range 1.05 to 2.50) which minimized the asymptotic

variance. A further discussion of the asymptotic variance of the L_p -norm estimator and a practical means of estimating it are given in Appendix 6.D.

Implementing M-estimation

Except for OLS and some of the maximum likelihood methods, most of the problems involved in quasi-MLE and L_p -norm estimation are computational ones. L_1 -estimation has long been formulated as a linear programming (LP) problem and several advances and modifications (e.g. Barrodale and Roberts, 1973) have resulted in a very efficient algorithm. In the case of L_2 -norm estimation the solution is easily found through Gaussian elimination. As mentioned previously, the least squares solution is linear in the observed dependent variables and is the most efficient of this kind as proved by the Gauss-Markov theorem. For the general p case, a modified Davidon-Fletcher-Powell method as used, for example, by Forsythe (1972) was implemented.

L-estimators

Koenker and Basset (1978), Basset and Koenker (1982) and Koenker and Portnoy (1987) give detailed descriptions of regression quantiles and their application to robust regression. The use of regression quantiles gives rise to the trimmed mean estimators of Koenker and Basset (1978) and Welsh (1987).

The Koenker-Basset 100 q % trimmed regression estimate of beta (which is denoted by $\beta_{KB}(q)$) is obtained as part of the solution to the minimization problem:

$$\min_{\alpha, \beta} \sum_{i=1}^n d_i(q) (y_i - \alpha - \beta x_i)^2 \quad (22)$$

where

$$\begin{aligned} d_i(q) &= 1, & \hat{\beta}_{RQ}(q)x_i \leq y_i - \hat{\alpha}_{RQ}(q) \leq \hat{\beta}_{RQ}(q) \\ &= 0, & \text{otherwise} \end{aligned} \quad (23)$$

and the parameters subscripted RQ are the regression quantiles obtained from the minimization of

$$\min_{\alpha, \beta} \left\{ q \sum_1 z_i + (1 - q) \sum_2 z_i \right\} \quad (24)$$

where $z_i = y_i - \alpha - \beta x_i$ and Σ_1 is taken over i such that $z_i \geq 0$ and Σ_2 is taken over i such that $z_i < 0$.

Welsh (1987) produced an alternative trimmed estimator which uses the residuals from a preliminary fit of the model to the data to 'winsorize' the data set. If the ordered residuals from the preliminary fit are denoted by $\{z_{(i)}\}$ then for a given 'trimming parameter' q between 0 and 1, and $s = [nq]$, the following is firstly defined:

$$\begin{aligned} \zeta_q &= r_{(s)}, & s &= nq \\ &= r_{(s+1)}, & & \text{otherwise} \end{aligned} \quad (25)$$

where $[]$ denotes taking the integer portion of the operand.

The Welsh 100 $q\%$ trimmed mean estimator of the parameter vector $\theta' = (\alpha, \beta)$ is then defined by:

$$\hat{\theta}_{WE} = (X' B_q X)^{-1} X' (B_q Y + h_q(z)) \quad (26)$$

where $B_q = \text{diag}(b_i(q))$ with $b_i(q)$ equal to unity when z_i is between ζ_q and $\zeta_{(1-q)}$ and zero otherwise.

The vector $h_q(z)$ has as its i th component

$$\begin{aligned} h_{q,i} &= \zeta_q - q(\zeta_q + \zeta_{1-q}), & z_i &\leq \zeta_q \\ &= -q(\zeta_q + \zeta_{1-q}), & \zeta_q &\leq z_i \leq \zeta_{1-q} \\ &= \zeta_{1-q} - q(\zeta_q + \zeta_{1-q}), & z_i &> \zeta_{1-q} \end{aligned} \quad (27)$$

Welsh (1987) demonstrated that (theoretically) the Koenker-Basset and his estimators share the same asymptotic properties. The common property binding the above estimators is that they depend on the ordered values of the regression quantiles (for the Koenker-Basset) and the ordered preliminary residuals (for the Welsh).

Other L-estimators which have been used in the literature include the minimum median residual regression and the L_1 (or LAD) regression. In both these cases the estimators can be formulated in terms of regression quantiles (with $q = 0.5$ and 1 , respectively). The fact that LAD can be classified as both an L- and M-estimator adds to its intuitive appeal as a robust regression technique applicable in a wide variety of cases.

R-estimators

The third type of robust regression estimator is based on ranked statistics and is hence classified as an R-estimator. One version of a family of regression analyses which Olkin and Yitzhaki (1992) call Gini regressions is considered in this chapter. The Gini coefficient and Gini mean difference are fairly well-known in finance, econometrics and economic statistics (see Yitzhaki (1982), Shalit and Yitzhaki (1984), or Carroll, Thistle and Wei (1992), for recent examples).

Scholz (1977) and Sievers (1978) independently derived regression techniques based on the weighted average of the slopes between pairs of observations. Olkin and Yitzhaki show that these regression techniques are special cases of Gini regressions. In establishing the power characteristics of his estimator, Sievers (1978) reports that his version of Gini regression has asymptotic efficiency greater than 95% for the normal case and that the efficiency is larger than 86.4% for all continuous distributions. Because of the high degree of non-normality in residual distributions the efficiency of the Gini estimator relative to the OLS estimator (which performs best under normality) will almost certainly be greater than unity.

Sievers (1978) estimator is the weighted median of the slopes between the individual pairs of observations. The weights are given as the absolute difference between the x -values.

Sievers's estimator is given by

$$\hat{\beta}_G = (\hat{\beta}_U + \hat{\beta}_L)/2 \quad (28)$$

where

$\hat{\beta}_U = \sup\{\beta: T_\beta \geq a../2\}$ and $\hat{\beta}_L = \inf\{\beta: T_\beta \leq a../2\}$ with

$$T_\beta = \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_{ij} \phi(z_i, z_j) \quad (29)$$

where $\phi(u, v) = 1$ or 0 according as $u \leq v$ or $u > v$. The a_{ij} represent weights and Sievers suggests $a_{ij} = x_j - x_i$ if $j > i$, on the assumption that the data set is in ascending order of the x 's. The term, $a..$, is the sum of all the a_{ij} 's for $i < j$.

A simpler way of describing the estimator is to consider a probability mass function

$$p(S_{ij}) = a_{ij} / a.. \quad (30)$$

for $i < j$, and S_{ij} equal to the $(y_j - y_i)/(x_j - x_i)$. The proposed estimator is then the median of the mass function. Sievers also demonstrated a method for obtaining confidence intervals and performing hypothesis tests using his estimator.

Appendix 6.C

Distribution of partially adaptive GT estimators

McDonald and Newey (1988) describe two methods for finding the M-estimator based on the GT distribution. The implementation of both these methods rest on a preliminary estimate of the distributional parameters of the GT distribution being available, a process for which McDonald and Newey give two alternatives. The first of these is asymptotically efficient only if the residuals arise from a GT distribution, but the second process whereby the asymptotic variance is minimised, is always at least as efficient as selecting an arbitrary set of distributional parameters.

To describe both the estimation procedures for the distributional parameters and then the regression parameters it is necessary to introduce (in modified form) the notation which McDonald and Newey use:

$$\rho(z) = \ln \left[1 + \frac{1}{q} |z/\sigma|^p \right] \quad (31)$$

$$\rho_z(z) = \partial \rho(z) / \partial z = [p |z/\sigma|^{p-1} \text{sign}(z)] / [\sigma(q + |z/\sigma|^p)] \quad (32)$$

$$\begin{aligned} \rho_{zz}(z) &= \partial^2 \rho(z) / \partial z^2 \\ &= [p(p-1) |z/\sigma|^{p-2} (q + |z/\sigma|^p) - p^2 |z/\sigma|^{2p-2}] / [\sigma(q + |z/\sigma|^p)]^2 \end{aligned} \quad (33)$$

If the functions above are evaluated at the (truncated¹¹) estimated residuals obtained from the model with regression parameters estimated by minimizing $\sum \rho(z_i)$ for some given distributional parameters (p , q and σ) then the asymptotic variance of the estimated slope coefficient ($\hat{\beta}$) is:

$$\hat{V}ar[\hat{\beta}] = \hat{v} \hat{Q}_x^{-1} \quad (34)$$

¹¹ Since ρ_{zz} is unbounded at $z = 0$, MacDonald and Newey replace any zero residuals by a small positive constant which is denoted by δ . In Appendix 6.D an alternative suggestion to replacing the very small or zero residuals is proposed.

where

$$\hat{Q}_x^{-1} = \sum_{i=1}^n x_i^2 / n - \bar{x}^2 \quad (35)$$

and

$$\hat{v} = \frac{\sum_{i=1}^n \hat{\rho}_z^2 / n - \left[\sum_{i=1}^n \hat{\rho}_z / n \right]^2}{\left[\sum_{i=1}^n \hat{\rho}_z / n \right]^2} \quad (36)$$

Consistent estimates of the distributional parameters (which one would need to estimate in a partially adaptive scheme) are derived by minimizing the asymptotic variance evaluated using the residuals from some preliminary regression parameters estimated in a robust way.

McDonald and Newey describe a one-step estimator (providing an estimate $\tilde{\beta}$) which is also consistent and which is computationally much more simple than the minimization of $\sum \rho(z_i)$. The procedure again depends evaluating all the relevant functions using residuals obtained from preliminary estimates of the regression parameters (denoted β^*):

$$\tilde{\beta} = \beta^* + (\nu_2 \hat{Q}_x)^{-1} \sum_{i=1}^n (x_i - \bar{x}) \times \rho_z / n \quad (37)$$

where

$$\nu_2 = \sum_{i=1}^n \rho_z / n \quad (38)$$

Estimation of p , q and σ must be done using a comprehensive minimization routine, although, for technical reasons outlined by McDonald and Newey, p must be bounded and bounded away from 1, q must be at least 1, but bounded and σ must be strictly positive and bounded. The restrictions disallow certain special cases of the GT family (e.g. the normal distribution), but McDonald and Newey conclude that the advantages of applying the restrictions outweigh the disadvantages and note that, although the special cases can never be attained exactly, they can be approached quite nearly.

Appendix 6.D

The Standard Errors of Bounded Influence L_p -norm Estimators

In this appendix, a discussion of the variance of L_p -norm estimators is given. First, the general theoretical results are discussed and then the practical problem of obtaining an estimate of the variance is outlined. A method for estimating the variance reasonably accurately is given and tested using data from the JSE.

The theoretical discussion starts by defining X to be the matrix of explanatory variables and β to be the vector of regression parameters. Nyquist (1980, 1983) showed, on the basis of several plausible assumptions, that if $\mathcal{F}$ is the common distribution function of the residuals then the L_p -norm estimates are asymptotically normally distributed with mean β and variance $\omega_p^2 Q^{-1}/n$, where

$$\omega_p^2 = \begin{cases} \frac{1}{[2\mathcal{F}(0)]^2} & \text{if } p = 1 \\ \frac{E[|z_i|^{2p-2}]}{[(p-1)E[|z_i|^{p-2}]]^2} & \text{if } 1 < p < \infty \end{cases} \quad (39)$$

and $Q = \lim_{n \rightarrow \infty} X'X/n$

Basset and Koenker (1978) considered the case $p = 1$, and derived the following result: the estimate is asymptotically normally distributed with mean β and variance $\lambda^2(X'X)^{-1}/n$ where λ^2/n is the variance of the sample median. For this special case where $p = 1$, an estimate of λ can be derived by one of two methods due to Cox and Hinkley (1974) or McKean and Schrader (1987).

The Cox and Hinkley method depends on a result by Cramer (1946) stating that asymptotically $\lambda = 1 / 2 f(m)$ where $f(m)$ is the ordinate of the error distribution at the median, m . An estimate of $f(m)$ can be obtained from

$$\hat{f}(m)^{-1} = \frac{\hat{z}_{(i)} - \hat{z}_{(j)}}{(i - j) / n} \quad (40)$$

where $\hat{z}_{(i)}$ denotes the ordered residuals. Cox and Hinkley (1974) showed that i and j must be symmetric about the rank of the median residual and the difference between the two must be kept small. Considering only the $\tilde{n}$ residuals which are non-zero

$$i = \left[\frac{\tilde{n} + 1}{2} \right] + \nu \quad \text{and} \quad j = \left[\frac{\tilde{n} + 1}{2} \right] - \nu \quad (41)$$

for $\tilde{n}$ odd and

$$i = \left[\frac{\tilde{n}}{2} \right] + 1 + \nu \quad \text{and} \quad j = \left[\frac{\tilde{n}}{2} \right] - \nu \quad (42)$$

for $\tilde{n}$ even.

ν is a small, positive constant.

On the other hand, McKean and Schrader (1987) suggested an estimate of λ based on a non-parametric confidence interval,

$$\hat{\lambda}_{1-\alpha} = \frac{\sqrt{\tilde{n}} (\hat{z}_{(\tilde{n}-\zeta+1)} - \hat{z}_{(\zeta)})}{2 z_{\alpha/2}^*} \quad (43)$$

where asymptotically

$$\zeta = \frac{\tilde{n} + 1}{2} - z_{\alpha/2}^* \left(\frac{\tilde{n}}{4} \right)^{\frac{1}{2}} \quad (44)$$

and $z_{\alpha/2}^*$ is from the standard normal tables and ζ is rounded to the nearest integer.

In the more general case when p is not equal to one, Gonin and Money (1989) have shown that estimates of ω_p^2 based on sample moments in the numerator and denominator of the expression derived by Nyquist (1980, 1983) are satisfactory for large n and values of p larger than 1.5. However, the denominator of the expression grows without bound for values of p close to one. McDonald and Newey (1988) describe a similar problem in their estimator based on the generalised t distribution.

The sample moment estimate of the asymptotic variance of the L_p -norm estimator is given by

$$\hat{\omega}_p^2 = \frac{m_{2p-2}}{[(p-1)m_{p-2}]^2} \quad (45)$$

with

$$m_r = \frac{1}{n} \sum_{i=1}^n |\hat{z}_i|^r \quad (46)$$

and where $\hat{z}_i$ is the i^{th} residual from the L_p -norm fit.

Note that in (45) small residuals will cause unbounded inflation of the denominator and hence the estimated variance will approach zero. McDonald and Newey (1988) suggested replacing small residuals with a small constant. The downside of their suggestion is that a *very good* fit which produces very small residuals for *all* the observations will be attributed a larger estimate of the variance than is appropriate. Similarly, merely trimming the sample moment may unnecessarily enlarge the variance. As an alternative an *adaptively trimmed, progressive average* method is proposed. The proposed method operates as follows:

The moment in the denominator is calculated progressively, including each residual successively in *descending* order of size. After the number of residuals included is larger than 30 the percentage change in the sample moment caused by the addition of the next residual is compared with a critical value of 50% the previous size of the estimated denominator. If the percentage change is larger than this critical value the sample moment is kept equal to the value before the addition of the residual. If not, the next residual is included and so on. The choice of 30 as a minimum number of residuals to include and 50% as a decision criterion is largely arbitrary and is a source of further research. In most instances in the empirical studies which were performed, the chosen criteria led to fewer than one or two residuals out of the 60 being excluded.

Assessing the quality of the adjusted method of moments estimator

The value of ω_p^2 is unknown for any particular share. To gauge the quality of the estimator of ω_p^2 precisely is therefore impossible. Bootstrapping was therefore used to give a comparative measure of the variance. 200 bootstrap samples for each share were drawn from each of the 1983-1987 and 1988-1992 data sets. The standard errors of the estimates derived from the bootstrap samples were used as estimates of the standard deviations for the individual shares' betas. There were thus two sets of bootstrap variances for each share, corresponding to the two periods. The bootstrap variances were compared with the standard method of moments estimates of variance and with the *adaptively trimmed, progressive average* estimator.

The results of the comparison indicated that the usual measures of ω_p^2 using all the residuals were indeed unreliable for $p < 1.5$. The sample moments were much smaller than the bootstrap estimates, which were taken as benchmarks.

Regressing the adaptively trimmed progressive moments against the bootstrap estimates, however, were encouraging in that they showed a close correlation. For the two periods the constant parameters from the two regressions were found to be small (i.e. not significantly different from zero) and the slope parameters were 0.845 and 0.800 respectively. Under the null hypothesis that the bootstrap and progressive average estimates were equal the constant parameter (intercept) would be expected to be zero and the slope coefficient unity. The regression results indicate that those expectations are nearly met.

The correlation between the two estimates (the bootstrap and the adaptive, progressive average estimates) of the standard error of beta were 0.68 and 0.81 for the two periods. On the basis of the regression findings the conclusion is that the adaptive progressive method yields slightly optimistic estimates of the variance (relative to the bootstrap estimate) but is superior to the usual method of moments.

CHAPTER 7

IMPROVED BETA ESTIMATION USING ADAPTIVE INTERVAL SELECTION

7.1 Introduction

One of the major problems involved in testing financial models is the precision of estimates of financial risk, specifically security beta coefficients. Some researchers have attempted to enhance efficiency by adopting robust estimators for the regression, others by using more sophisticated techniques such as state-space modelling. In this chapter a novel approach to improving beta estimation by modifying the input data is considered. In particular it is demonstrated how the efficiency of beta estimators can be improved by exploiting the fact that there exists a near continuous series of price information for reasonably well traded securities. Even for thinly traded securities, there is generally more information available than just the month-end prices traditionally used in beta estimation.

As noted in the previous chapters of this thesis, the main hurdle in a market such as the JSE for improving efficiency by the use of shorter interval lengths, is the increased bias caused by infrequent trading in the beta estimates. In an attempt to overcome this problem, two proposals are considered in this chapter. The proposals are designed to take advantage of the fact that securities are priced more frequently than at monthly intervals, without prejudicing the unbiasedness of the estimates.

The two proposals have been termed (1) the interval aggregation procedure and (2) the extreme value procedures. The principle underpinning the interval aggregation procedure is the *sequential* selection and aggregation of intervals. The modification of the intervals is made in order to accentuate the linear relationship between the security and market returns (and therefore improve efficiency in the estimation of the slope coefficient, i.e. the beta). The extreme value set of procedures, on the other hand, attempt to improve beta estimation efficiency by picking out the extreme prices in each month so as to emphasize the relative volatility (i.e. beta) of the security.

By virtue of its sequential nature, the interval aggregation procedure suffers from the potential disadvantage of a reduced sample size which will be a counter-active influence on efficiency. In the extreme value procedure, no such loss of sample size occurs. However, there is a built in restriction in the method because one observation *has* to be made each month, even

when a more efficient estimate may have been obtained by aggregating the information in that particular month with the information in the next month.

In this chapter the two proposals for more efficient beta estimation are considered both theoretically and empirically in order to ascertain whether the advantages, or disadvantages, of the methods prevail.

The chapter proceeds as follows: in section 7.2 the interval aggregation procedure is motivated, discussed and tested empirically. In section 7.3 the extreme value set of procedures are investigated and recommendations are made for the efficient estimation of beta coefficients.

7.2 The interval aggregation procedure for improving the efficiency of beta estimation

In this section the issue of improved efficiency of beta estimation is initially considered from a theoretical viewpoint. The fundamental principle that arises from the theoretical development is that estimation efficiency can be improved by spreading the measured market returns as much as possible. By 'stretching' the market returns, any linear relationship between the security and market returns will be accentuated. Consequently more efficient beta estimates can be attained.

The interval aggregation procedure is proposed as a method for sequentially deciding whether or not the returns in two consecutive months (or any other so-called *basic* interval) should be aggregated to increase the spread of the market returns. Throughout the rest of the chapter, the *basic* intervals are defined to be the shortest interval at which data are considered. Thus, for example, a *basic* interval may be a week if weekly data are available instead of monthly data. The development and testing of the interval aggregation procedure proceeds as follows:

In sections 7.2.1 and 7.2.2 the problem of improving the efficiency (i.e. reducing the estimation variance) of an estimator of beta is discussed and the interval aggregation procedure for modifying the input data is motivated. Section 7.2.3 is a more technical section outlining the theoretical effectiveness of the aggregation procedure. In Section 7.2.4 the potential efficacy of the proposal is demonstrated using a specific example. The remaining sub-sections present the results of an empirical study performed on the JSE. In the study the usefulness of the proposal using actual financial data is discussed. In particular, the extent to which the reduced

number of observations (inevitably associated with the interval aggregation procedure) negates the improvements in efficiency is considered.

7.2.1 General regression results

Focusing on the simple linear model relating one-period returns on a security to that of returns on the market, viz.:

$$R_t = \alpha + \beta M_t + u_t \quad (1)$$

several well-known results reveal possible areas where precision in the estimation of the regression parameters (α and β) can be improved. The most important statistic to concentrate on in this preliminary discussion is the variance of the OLS estimator of beta, given by

$$\text{Var}[\hat{\beta}] = \frac{\sigma_u^2}{n \sigma_M^2} \quad (2)$$

where the numerator, σ_u^2 , represents the variance of the residual component and the denominator the sample size, n , multiplied by the variance of the market returns, σ_M^2 . From this expression it can be seen that precision in the estimate of beta (i.e. a reduction in the variance of the estimate) will be achieved by one of three procedures: (1) decreasing the residual variance, (2) increasing the sample size or (3) increasing the variance of the market returns.¹

Robust regression and other more sophisticated estimation techniques are designed to reduce residual variance and are not discussed further in this chapter. The issue of increased sample size is a complex one in the context of security beta estimation: a larger sample size implies either that returns are measured over shorter intervals within a given period of time, or that more intervals are used which implies a longer estimation period. Longer estimation periods are generally associated with less confidence in underlying parameter (beta) stability, and shorter estimation periods are generally associated with increased residual volatility.

¹ Provided, of course, that the other two factors remain constant.

In both sections 7.2 and 7.3 of this chapter the focus is essentially on modifying the market return variance as a method of improving precision.

7.2.2 The interval aggregation procedure for improving precision by increasing the variance of the market returns

Increasing the variance of the market returns is tantamount to spreading the explanatory variable in a regression analysis more widely along the x -axis. The larger the range of values encompassed by the explanatory variable, the clearer any linear relationship between the y and x will be and hence the more efficient any fitting procedure.

One method of achieving the desired increase in market variance is to increase the average magnitude of the measured market returns. If the market returns are all large in magnitude (be they positive or negative) then their variance will also be large. It would appear then that the obvious procedure to adopt would be to use longer interval lengths. However, the use of fixed interval lengths makes attaining a large market variance² nearly impossible. For instance, the market index may go up and then down within the interval leaving a net market return of close to zero. Alternatively, consider a steady bull run which implies that all the market returns in successive fixed intervals will be positive. However, the market variance will be small because of the lack of variability. In fact, it would appear that in this case the aggregation of all the small positive market returns into one large market return would increase the spread of the observed market returns.

Generalizing the above discussion, improved estimation efficiency will be attained by adaptively selecting the interval lengths so that the market variance is maximized (or at least increased). Appendix 7.A contains a mathematical exposition of the optimization procedure required. The *interval aggregation procedure* which is now described is one procedure for increasing the market variance. It is based on the idea of going through the data set sequentially and deciding whether each observation in turn should be aggregated to the previous observation or not.

² By large it is meant large relative to the interval length. It was shown, for example, in Chapter 4 that longer fixed intervals led to an increased variance in both the market returns and the residuals. In fact, the extent of the increases in the variances was nearly exactly in proportion to the interval lengths. What is required is a procedure for increasing the market variance beyond that expected from simply lengthening the intervals. Otherwise the residual variance will increase at a similar rate and cancel the improvement in the efficiency.

More specifically, the procedure starts by calculating the *basic* returns on the market in the first and second fixed, *basic* intervals³. If the sign (either positive or negative) of the second *basic* return is the same as the first *basic* return then the two basic returns are aggregated to form a *modified* return. The third basic return from the third basic interval is then calculated: if it has the same sign as the modified return from intervals one and two then it, too, is aggregated to that modified return. If not, then the third basic return forms the start of a new *modified* return. Successive *basic* returns will in turn be added to the new modified return until the sign of the *basic* return changes.

A *modified* return is thus defined as the sum of *successive basic returns* on the market which have the same sign. The characteristics of the modified series of returns are therefore: (a) the signs of the modified returns alternate between positive and negative and (b) the modified returns are not necessarily measured over equal interval lengths.

It is important to note that modifying the observed returns in the way described above will not increase the residual variance and thereby nullify the improvement in efficiency. The reason for this is that the interval lengths of the modified returns are based on the runs of positive and negative *market* returns. The residuals are assumed to be uncorrelated with the market returns (*however they may be aggregated*). Therefore a similar increase in the residual variance would *not* be expected from the interval aggregation procedure. However, there will obviously be a loss of observations associated with the aggregation of the basic returns into modified returns. Hence, a potential disadvantage of the modified returns from the interval aggregation procedure (in contrast to using the basic returns) is that the reduced sample size may vitiate the gains in efficiency.⁴

7.2.3 A more formal description of the proposal

The above description of the interval aggregation procedure for increasing the variance of measured market returns is rather informal. In this section an analytical description of the proposed procedure is presented by briefly working through some of the mathematics which

³ The basic intervals can be any fixed length, e.g. days, weeks, months, etc.

⁴ The success of the interval aggregation proposal depends on the fundamental assumption that the market model holds over intervals of all lengths. This assumption is fairly common in the context of financial modelling and is also used, for example, in the trade-to-trade estimator of betas.

underpin it. The development starts by examining the variance of the linear regression estimator. It then proceeds by considering how aggregating adjacent observations will affect the variance. Finally, the estimation variance associated with the *interval aggregation procedure* is examined and an inequality derived which establishes under what circumstances efficiency will be improved by the proposal.

The development accordingly starts by considering the general linear model written in matrix notation:

$$y = X\beta + u \quad (3)$$

where y and u are $T \times 1$ vectors, β is a $k \times 1$ vector and X is a $T \times k$ matrix of explanatory variables. The u_i comply with usual assumptions of being independently and identically distributed with zero mean and constant variance.

A is defined to be an $m \times T$ matrix of the form:

$$A = \text{diag} \left[\frac{1}{\sqrt{n_j}} \times \mathbf{1}_{n_j}' \right] \quad (4)$$

where j runs from 1 to m ; $\mathbf{1}_n$ is an n -vector of ones and the sum of the n_j is equal to T .

Premultiplying the linear model above by A yields:

$$Ay = AX\beta + Au \quad (5)$$

Note that the new residual component (Au) in equation (5) retains the properties of serial independence and the constant variance owing to the structure of A . If it is assumed (without

loss of generality) that the residual variance is unity, i.e. $\sigma^2 = 1$, and if equation (5) is now used to estimate β , the variance of the estimate will be given by

$$\begin{aligned} \text{Var}[\hat{\beta}] &= (X'A'AX)^{-1} \\ &= \left[\sum_{j=1}^m 1/n_j X^{(j)'} J_{n_j} X^{(j)} \right]^{-1} \end{aligned} \quad (6)$$

$X^{(j)}$ is defined to be the sub-matrix of X containing the n_j rows (observations) starting at row s_j and ending at row e_j . The sub-matrix thus represents the adjacent observations starting at observation s_j and ending at observation e_j . J_n is defined to be an $n \times n$ matrix of ones.

In the case of the market model where k (the number of explanatory variables) is equal to two, the inverse of the variance-covariance matrix, i.e. $(X'A'AX)$, is equivalent to

$$(X'A'AX) = \sum_{j=1}^m \begin{bmatrix} n_j & \sum_{i=s_j}^{e_j} x_i \\ \sum_{i=s_j}^{e_j} x_i & \frac{1}{n_j} \left(\sum_{i=s_j}^{e_j} x_i \right)^2 \end{bmatrix} \quad (7)$$

where x_i is the i th market return.

If s_j and e_j are randomly selected (i.e. without considering the values of x_i) and the average x_i value is approximately zero then it is relatively easy to demonstrate that the inverse of the variance-covariance matrix tends in limit to:

$$\begin{aligned} (X'A'AX) &\rightarrow \sum_{j=1}^m \begin{bmatrix} n_j & 0 \\ 0 & \sigma_x^2 \end{bmatrix} \\ &= \begin{bmatrix} T & 0 \\ 0 & m \sigma_x^2 \end{bmatrix} \end{aligned} \quad (8)$$

The variance-covariance matrix itself is given by the inverse of the matrix in equation (8) and so will also be a diagonal matrix with components equal to the reciprocal values of the diagonal components in equation (8). Remembering that the systematic risk of a security is given by the second component of β in equation (3), it is clear that the variance of the

estimator equals the reciprocal of $m \sigma_x^2$. The implication of the expression in (8) is then that the variance of the security's estimated beta is minimized when $m \sigma_x^2$ is maximized. This maximum will occur when m is maximized, since σ_x^2 is fixed. The m can be interpreted as the number of modified returns from a random aggregation procedure, i.e. one in which adjacent returns are aggregated without taking into account the signs of the basic market returns. The maximum m is equal to T (i.e. not performing any aggregations) with all the n_j equal to one. Simply stated, the most efficient way of estimating beta is to use all the basic returns, unless the interval aggregation procedure is performed in some other way.

However, if s_j and e_j are selected in the way suggested by the interval aggregation proposal (which results in all the x_i having the same sign when i is between s_j and e_j) then the perspective alters. The primary reason for the change is that, although a randomly selected x_i has zero expectation, the expectation of the market return *conditional on it being positive (negative)* will be non-zero.

In order to derive the limit of equation (7) when the interval aggregation procedure is implemented it is assumed that the distribution of the x_i is symmetric and that the conditional expectation of a (negative) positive x_i value is $(-)\mu_x$. The 'expected' values of the off-diagonal elements of equation (7) are given by $n_j \mu_x$ when the x_j 's are positive and $-n_j \mu_x$ when the x_j 's are negative. Because the distribution of the x_j 's has been assumed symmetric, asymptotically there will be an equal number of $+n_j \mu_x$ terms and $-n_j \mu_x$ terms. Therefore the summation from $j=1$ to $j=m$ makes the limiting values of the off-diagonals zero. The bottom right component of (7) has an 'expectation' of $(n_j^2 \mu_x^2 + n_j \sigma_x^2)/n_j$. The summation over j outside the brackets therefore yields a limit of $m \sigma_x^2 + T \mu_x^2$.

Hence, by virtue of the argument outlined above, it can be shown that the limit of the $X'A'AX$ matrix is

$$(X'A'AX) \rightarrow \begin{bmatrix} T & 0 \\ 0 & (m \sigma_x^2 + T \mu_x^2) \end{bmatrix} \quad (9)$$

Precision of beta estimation will now be improved over that obtained from using the basic intervals whenever

$$\begin{aligned} m \sigma_x^2 + T \mu_x^2 &> T \sigma_x^2 \\ \Rightarrow \frac{m}{T} &> 1 - \frac{\mu_x^2}{\sigma_x^2} \end{aligned} \quad (10)$$

The above finding can be summarized as follows: there will be an increase in precision as long as the proportion of observations lost due to the aggregation is less than the squared ratio of the conditional mean to the standard deviation of the x_i . The relationship in the inequality in (10) also makes some intuitive sense: if the average size of the market returns is small relative to the spread of the market returns then the term on the right hand side of the inequality (10) will be close to unity. Therefore very few observations can be lost before it becomes more beneficial to stay with the *basic* returns. This finding is very much in line with the stated intention earlier of increasing the average magnitude of the market returns so as to make the linear relationship clearer.

Before progressing to the empirical study of the interval aggregation procedure, it is worth considering an hypothetical distribution for the market returns. The point of doing so is to determine theoretically how many observations are needed before the interval aggregation procedure improves the efficiency.

7.2.4 An hypothetical example of the predicted improvements due to the aggregation proposal

In this sub-section a specific example is given of the conditions under which the proposal has merit. In particular the assumption is made that the x_i (the market returns) are generated independently from a contaminated normal distribution. The assumption of a contaminated normal distribution implies that a small proportion (which is denoted by ϵ) of the observations come from a normal distribution with a variance κ times bigger than the usual variance. In other words, an observed market return comes from one of two normal distributions. For the most part, the distribution produces 'good' observations with a particular variance. However, every now and again an outlying or 'bad' observation may occur with a variance which is much larger than that of the 'good' observations.

It is easy enough to show that for a standard normal distribution the conditional expectation of a random variable given that it is positive is

$$E[x_i | x_i > 0] = \sqrt{\frac{2}{\pi}} \quad (11)$$

The conditional mean and variance of the contaminated normal distribution can be simply obtained as the weighted average of the two component normal distributions. Bearing this in mind and the result in (11), equation (10) demonstrates that the maximum proportion of observations which can be lost before it is more efficient to use the *basic* returns instead of the *modified* returns (from the interval aggregation procedure) is

$$\frac{\mu_x^2}{\sigma_x^2} = \frac{\left[(1 - \epsilon) \sqrt{\frac{2}{\pi}} + \epsilon \sqrt{\frac{2}{\pi} \kappa} \right]^2}{1 + \epsilon(\kappa - 1)} \quad (12)$$

Equation (12) represents the theoretical ratio of the squared expectation of the market to the variance of the market, assuming that the market returns are distributed with a contaminated normal distribution. This ratio represents the maximum proportion of *basic* intervals which can be aggregated before it is more efficient not to perform any aggregation at all. From equation (12) it follows that if $\epsilon = 0$, i.e. if the x_i come from an uncontaminated normal distribution with a proportion of outliers equal to zero, then up to $2/\pi$ (i.e. about 64%) of observations can be lost before the proposal is less efficient than using the basic intervals. If ϵ were 0.01 and κ were nine (i.e. 1% of the observations came from a normal distribution with a standard deviation of three times the usual) then the maximum proportion of lost intervals would be 61%. If there were even more 'outliers' (say $\epsilon = 0.1$) then the proportion decreases further to 51%. Although it is not contended that the distribution of market returns is best described by a contaminated normal distribution, some useful insights can nevertheless be drawn from the example. Most importantly it is observed that the heavier-tailed distributions with many outliers have less to gain than the more 'well-behaved' distributions (i.e. the distributions closer to the normal distribution) from the interval aggregation proposal described, because fewer observations can be lost before it is more efficient to use the basic returns.

An empirical study was conducted using securities from the JSE in order to assess the interval aggregation procedure. The methodology and results of the study are discussed in the ensuing sections.

7.2.5 Empirical study to assess the interval aggregation procedure

In order to assess the efficacy of the suggested proposal an empirical study using data from the JSE was conducted. All the securities which had been listed continuously from 1 January 1978 to 18 March 1993 were included. Moreover, the sample of securities was restricted to those which had traded on more than 90% of the weeks in that period so that extremely thinly traded shares were excluded. There were thus a total of 130 securities in the sample.

The measured security returns were regressed against the returns measured on the JSE-Actuaries Overall Index using various *basic* interval lengths, both with and without the aggregation proposal. For each *basic* interval length the return on the security was computed from the log price relative. The returns were then aggregated according the interval aggregation procedure described in section 7.2.2 to form the *modified* returns. In all cases the returns were 'standardized' by dividing them by the square root of the length of the interval over which they were measured.⁵

The results are analyzed by considering the two components of the variance of beta separately: firstly, the ratio of residual to market return variances (σ_u^2/σ_M^2) and, secondly, the number of observations, n , available in the period of observation (see equation (2) for the expression for the estimation variance). As noted earlier, the comparison of the reduced variance with the reduced sample size will indicate whether or not the interval aggregation procedure does indeed reduce the variance of the estimated betas.

The primary results of the study are presented in Table 7.1. For each *basic* interval length the ratio of the average variance of the estimated residuals to the variance of the measured market returns are given, both without and with the aggregated intervals proposal.

The salient feature of Table 7.1 is the reduction in the ratio of residual to market variance for nearly all basic intervals caused by the modification of the return series associated

⁵ The standardization does of course mean that the formula for the standard error of the beta estimate will be slightly different from that in the theoretical development. However the magnitude of the difference is small enough so that it will not change the results or conclusions significantly.

Table 7.1

Effect of aggregating basic returns into modified returns. The percentage decrease in the ratios of residual to market return variances is given alongside the percentage decrease in observations caused by the aggregation procedure. Notice that in each case there is a larger percentage decrease in observations than in the ratios. (All figures in parentheses are negative.)

Basic Interval length (weeks)	Ratio of residual to market variances (no modification)	Ratio of residual to market variances (with modification)	% decrease in ratios	% decrease in observations	No. of basic intervals
1	3.456	1.891	45	58	792
2	2.965	1.736	41	53	396
3	2.601	2.283	12	45	264
4*	2.463	1.984	19	49	198
5	2.329	1.566	33	53	158
6	2.284	1.706	25	52	132
7	1.966	2.066	(5)	50	113
8	1.939	1.600	17	45	99
9	1.973	1.571	20	49	88
10	2.099	1.530	27	48	79
11	1.782	1.548	13	39	72
12	1.879	1.772	6	50	66
13	1.758	2.223	(26)	48	60

* The interval lengths marked in bold (4-week and 5-week basic intervals) are the interval lengths which represent the closest match to monthly intervals which are commonly used in empirical financial investigations.

with the aggregation procedure. In Table 7.1 the reduction can be observed by comparing the two columns headed 'ratio ... no modification' and 'ratio ... with modification'. The ratio of variances using the *basic* intervals (no modification) drops from a value of 3.456 for a *basic* interval of one week, to 1.758 when the *basic* intervals are taken as 13 weeks. The 'standardized' *modified* returns, in comparison, display much lower variance ratios (for example, the ratio is 1.566 for a 5-weekly basic interval compared with 2.329 when the *basic* returns are used). The percentage decrease in the variance ratios is given in the fourth column of Table 7.1. The comparison of the percentage decrease in the variance ratios with the percentage decrease in the observations caused by the aggregation procedure is the empirical counterpart of (10). If the percentage of observations lost is greater than the percentage reduction in the variance ratios, then the *basic* returns yield superior efficiency.

In the second last column of Table 7.1 the percentage decreases in the number of observations caused by the interval aggregation procedure are presented. Looking down the column the first feature to be noted is that the percentage of observations lost is in the region of 50% at most interval lengths. Furthermore, at every interval length the percentage of

observations lost is greater than the percentage reduction in the ratio of the variances. The interval aggregation procedure therefore appears to result in a greater loss in observations than is made up by additional market variance.

7.2.6 Further discussion of results and a check on the theoretical assumptions of the interval aggregation procedure

Before offering some conclusions about the interval aggregation procedure, it is worth examining the empirical evidence for some of the assumptions and speculations made during the theoretical development.

The comparison between the theoretical results and the empirical results yields some important additional insights into the behaviour of stock returns. It is well accepted that the density functions of stock market returns are heavy-tailed (see, for example, Affleck-Graves and MacDonald (1989) for recent evidence from the US market or Chapter 6 of this thesis for South African evidence). As noted in Table 7.1, the improvement in the ratio of residual to market variances caused by the aggregation procedure is relatively small. This finding is therefore in line with the hypothesis that a heavy-tailed distribution of the market returns will be relatively less assisted than a lighter tailed distribution (as indicated in the theoretical section of this section).

As an informal check on one of the other assumptions that was made in the theoretical model, the average *residual* standard deviations were examined. For the model to be strictly appropriate the residual standard deviations of the basic and modified proposals should be equal. The (modified) residual standard deviations from the interval aggregation procedure are given in column two of Table 7.2. Looking down the column it appears as if the longer *basic* intervals are associated with slightly smaller residual standard deviations.⁶ For example, the modified residual standard deviation for a basic interval length of one week is 0.0527 compared with a value of 0.0455 at a basic interval length of 10 weeks. Comparing the modified residual standard deviations with the standard deviations from the usual regression procedure, it is readily apparent that the residual standard deviations from the interval aggregation procedure

⁶ There appears to be an 'elbow' in the pattern of decreases of the standard deviations at a basic interval length of 5 weeks. This finding lends support to the choice of monthly data in most financial empirical studies conducted.

Table 7.2

Some checks on the consistency between the modified and basic models.

Length of basic interval	Modified residual standard deviation	Basic residual standard deviation	Modified average beta	Basic average beta
1	.0527	.0529	.87	.86
2	.0509	.0526	.95	.92
3	.0504	.0512	.97	.95
4	.0493	.0507	.99	.95
5	.0469	.0499	.96	.95
6	.0472	.0497	.98	.96
7	.0473	.0489	.89	.94
8	.0468	.0488	.95	.96
9	.0458	.0481	.98	.98
10	.0455	.0488	1.00	.98
11	.0471	.0485	.98	.97
12	.0481	.0488	.98	.98
13	.0504	.0481	.91	.96

are no bigger than the usual residuals and in fact are slightly reduced.⁷ For example, the modified residual standard deviation at a basic interval length of five weeks is 0.0469 in contrast to 0.0499 for the basic residual standard deviation.

As a final check on the consistency of the two models (the aggregated and basic) the average estimated betas at each interval length were also examined (see the final two columns of Table 7.2). There is marginal evidence that the modified betas are slightly larger on average than the *basic* betas, especially at the shorter interval lengths. The intervalling effect of betas (in which shorter intervals have usually been associated with lower betas⁸) appears to be somewhat mitigated by the aggregation procedure. The reason for the decreased intervalling effect is that the interval aggregation procedure, because of its construction, uses slightly longer intervals than traditional regression using the *basic* intervals.

⁷ The reduction may stem from the standardization procedure implemented in the empirical work: specifically the division of all returns by the square root of the length of the interval over which they were measured. The evidence from Table 7.2 suggests that the variance of the returns may not increase exactly linearly with the length of the interval over which it is measured. Nevertheless, the comparison of modified and basic standard deviations confirms that most of the improvement offered by the aggregation is attributable to the increases spread of the market returns.

⁸ The intervalling effect is discussed in more detail in Chapter 5. The fact that its presence is not as marked in this study compared with others, is that *well traded* securities make up most of the sample.

7.2.7 Concluding remarks about the interval aggregation procedure

In this section of the chapter, an interval aggregation procedure for improving the efficiency of beta estimation using the market model was proposed and tested empirically. The proposed procedure was based on aggregating runs of positive and negative market returns. In developing the procedure, it was noted that it would produce two counteractive effects: firstly, an increase in the average magnitude of the market returns (which would improve efficiency) and, secondly, a reduction in the number of observations (which would decrease efficiency).

It was shown that heavy-tailed distributions of the market returns would tend to decrease the effectiveness of the interval aggregation procedure. An empirical study on the JSE was conducted in order to ascertain which of the effects on efficiency caused by the procedure would dominate. The results of the study indicated that the reduced number of observations outweighed the improvements in the spread of the market returns. The heavy-tailed distributions of returns on the JSE can be cited as a possible reason for the limited effectiveness of the interval aggregation procedure.

The second adaptive procedure considered in this chapter is based on a similar principle, i.e. amplifying the measured returns on the market so as to emphasize the linear relationship between the security and the market. The extreme value procedure described in 7.3 avoids the loss of observations associated with the interval aggregation procedure by taking an observation from every basic interval. However, the observation selected from the basic interval is designed to highlight the relative volatility of the security by making use of the fact that price information is available throughout the interval, rather than just at the end of it.

7.3 The extreme value set of adaptive interval selection procedures

As indicated in section 7.2, aggregation of intervals leads to a loss of observations which is instrumental in reducing estimation efficiency. In this section an alternative method of 'spreading' the market returns is proposed. Specifically, the procedure involves selecting the highest and lowest market index values in each interval as the deciding factor in fixing the days between which returns are measured (as opposed to using the rather arbitrary end of interval dates). By selecting the extreme index values in each interval, the magnitude of the measured market returns will be increased and the relative volatility (beta) of the security emphasized.

The main concern of this section is then to establish whether the adaptive procedure does in fact offer a significant improvement in efficiency over that obtained by using the usual end of interval returns. The investigation proceeds as follows. In section 7.3.1 the extreme value procedures for selecting data intervals over which to measure returns are described. The empirical study detailed in section 7.3.2 provides evidence of the effectiveness of the procedure in increasing market variance. In section 7.3.3 the empirical study is extended to cover the estimation of betas for a sample of shares. In section 7.3.4 a robustification of the procedure and a way to implement the procedure for thinly traded shares are discussed. The conclusions of the study are discussed in section 7.3.5.

7.3.1 Introduction and discussion of the extreme value procedure

The findings of the empirical investigation reported in section 7.2 indicated that the loss of observations caused by the aggregation proposal reduced the effectiveness of the proposal. In this extension of the work another *ad hoc*⁹ procedure (the extreme value procedure) is proposed for increasing the variance of the market returns without the loss of observations associated with the aggregation procedure.

The key concepts underpinning the proposal are (a) the fixing of a constant number of observations from which to estimate beta in a given estimation period and (b) the selection of one index value from each of the fixed number of *basic* intervals in the estimation period. The index values in each *basic* interval are chosen so that the returns measured between the index values have as great a magnitude as possible. The reason for increasing the average magnitude of the returns is exactly the same as that described in 7.2, viz. making the linear relationship between the security and market returns more precise by using a wider range of market returns.

To be more precise, the proposal comprises three main steps. Firstly, a *basic* interval length needs to be selected (in this study calendar months were used). Next, the dates of the highest and lowest market index values within the basic intervals have to be determined. Thirdly, the market returns (and the relevant security returns) have to be calculated. The first market return is measured between the date corresponding to the lowest index value in interval $t-1$, say, and the date of the highest in interval t . The next market return is then measured between the dates of the highest value in interval t and the lowest in $t+1$, etc.

⁹ The procedure in no way guarantees improvements. See the Appendix 7.B for a discussion about the optimality of the procedure.

In the next section a brief description is given of the empirical study undertaken to establish the efficacy of the proposal for increasing the variance of the measured market returns.

7.3.2 The empirical study

The purpose of the empirical study was to show that the extreme value procedure did indeed increase the variance of the market returns above that of the month-end market returns. Furthermore, it needs to be shown that any improvements attained are greater than any other random selection of dates.

In order to perform the empirical study the daily data for the JSE-Actuaries Overall Index between 1 January 1988 (chosen to exclude the volatile period surrounding the 1987 crash) and 30 November 1993 were accessed. It should be noted that the estimation period does include the smaller 1989 crash and the impact of this is alluded to in a later section. *Basic* intervals of one calendar month were used chiefly because of their prevalence in the financial literature, but also based on the results of 7.2 and earlier chapters which indicated that it might be the most suitable.

The first hypothesis to be assessed was that the extreme value procedures did in fact offer an increase in market variance. The data were divided up into the monthly intervals and, for the purposes of investigating the first hypothesis, three sets of data from each interval were selected: the monthly closing value, the high in the month and the low in the month. From the three sets of price data three sets of measured returns were constructed: the returns between the monthly closing prices; the returns between the high and low values of alternate months starting with the low in the first month (the low-high returns) and, thirdly, a set of returns obtained in the same way as described above but starting with the high value in the first month (the high-low returns).¹⁰

The variances of each of the three data sets are presented in the first three rows of Table 7.3. The results in the second panel of Table 7.3 will be discussed later. It is clear from the first three rows that the proposed procedure does indeed increase the standard deviation of the measured market returns substantially in the whole period of investigation and in each of

¹⁰ Each return (log price relative) was divided by the square root of the length of time over which it was measured. This was done in order to standardize them with respect to the varying lengths of interval over which they were measured and thereby to facilitate comparisons of variance.

Table 7.3

The standard deviations of the market returns measured using different dates within each month.

	1988.01 - 1989.12	1990.01 - 1991.12	1992.01 - 1993.11	1988.01 - 1993.11
Closing prices	.011	.011	.011	.011
Alternating (low in first interval)	.017	.017	.014	.016
Alternating (high in first interval)	.023	.018	.015	.019
Average of a 1000 random intervals ^a	.014	.011	.011	.012
Maximum of the 1000 random intervals ^b	.017	.013	.012	.014
Mid-month ^c	.017	.012	.011	.014
Number of intervals	22	22	20	68

^a In each month a random date was selected to be the date at which the value was measured for that month. The process was repeated 1000 times and the average taken.

^b The maximum standard deviation obtained from the 1000 'random' intervals.

^c The value of the index was taken 11 trading days prior to the last trading day in the month.

the three non-overlapping sub-periods. Looking at the last column of Table 7.3 (which gives the results for the whole period of investigation) the standard deviation of the month-end returns is seen to be 0.011.¹¹ The low-high version of the extreme value procedure gives a substantially higher 0.016, and the high-low version an even higher 0.019. When it is remembered that the number of observations remains the same in the extreme value procedure as in the usual month-end version, it would seem that the extreme value procedures offer an increase in efficiency of about 50%. The improvement offered by the proposed procedure is apparent also in all three sub-periods.

An interesting feature of Table 7.3 which emerges when examining the standard deviations in the three sub-periods, is that in the first sub-period (1988.01 - 1989.12) there is a large difference between the low-high and high-low formulations of the procedure; viz. a standard deviation of 0.017 for the low-high version and 0.023 for the high-low. The difference might be due to the 1989 crash in that when a high value is taken in the first month, the crash month happens to coincide with a low value. This single outlying value is responsible for much of the increase in the market standard deviation. In fact, if the proposed procedure for the sub-period 1988.01 - 1989.12 is repeated but with the last two months of data excluded then the

¹¹ The figures are measured *per day*, so 0.011 corresponds to a daily standard deviation of 1.1%.

corresponding values for the standard deviation are 0.017 (low in the first interval) and 0.019 (the high in the first interval).

The results reported in the second panel of Table 7.3 verify the second hypothesis of this section; namely that the improvement offered by the extreme value procedure is superior to other random date selections in each month. In order to demonstrate this assertion, the standard deviations of market returns measured between essentially arbitrary points in each month are reported. The standard deviations reported were obtained from a Monte Carlo-type experiment: in each month a date was selected at random and the returns were measured between the randomly selected dates in successive months. The standard deviations of the return series were then calculated. The random selection was repeated 1000 times and the highest and the average of the 1000 standard deviations were recorded. Furthermore, a 'mid-month' series of returns was obtained by selecting the date eleven trading days from the end of each month as an arbitrary date on which to take the values of the index in each month.

The validity of the second hypothesis can then be assessed by establishing that the following set of relationships between the standard deviations of the various return series hold: firstly, the standard deviation of the extreme value procedure should be the largest; secondly the maximum standard deviation from the set of 1000 return series should be larger than the month-end standard deviation, but not as high as the proposed procedure; thirdly, the average standard deviation of the 1000 random series and the standard deviations of the month-end series and the mid-month series should all be approximately equal.

The results are presented in the second panel of Table 7.3. Firstly, note from Table 7.3 that the for the whole period (the last column) and for the two sub-periods which do not include any major 'crashes' (1990.01 - 1991.12 and 1992.01 - 1993.11) the expectations are quite closely met. From the overall column, it can be seen that the extreme value procedure produced standard deviations of 0.016 and 0.019 which are the highest in the column. In line with the expectations of the second hypothesis, the second highest value is the maximum of the 1000 repetitions, i.e. 0.014. The arbitrariness of the month-end series is confirmed by noting that the average standard deviation of the 1000 randomly selected series is 0.012 which is very similar to the 0.011 from the month-end series. The 'mid-month' series (also theoretically a random series) produced a standard deviation which was surprisingly high, i.e. 0.014, which is as high as the best of the 1000 random series. The same phenomenon of the 'mid-month'

standard deviation being higher than the month-end and average standard deviations occurs in all three sub-periods.¹²

In this section, the increased standard deviation of the market returns associated with the extreme value procedure was confirmed. The next step is to check that the increased spread of market returns translates into the expected increase in efficiency for beta estimates. A further empirical study was therefore conducted and is discussed in the next section.

7.3.3 Beta estimation using the proposed procedure

Although the evidence in the previous section demonstrates the potential for improved precision offered by the proposed procedure, a further empirical study was conducted in which betas were estimated for a sample of securities. The purpose of the second empirical study was to examine the hypothesis that the increased standard deviation of the market returns indeed improved the efficiency of the estimates of beta.

The securities selected for the study were all those which had traded on 95% of the days in the period 5 October 1987 to 7 December 1993. The period chosen corresponded to the available daily data and the selection of securities having the maximum trading frequencies avoided additional problems caused by thin trading biases. There were thus 33 securities in the sample.¹³

The betas and their standard deviations were computed over the same periods discussed in the previous section (1988.01-1989.12, 1990.01-1991.12, 1992.01-1993.11 and the overall period 1988.01-1993.11). The validity of the improved efficiency hypothesis was assessed in a similar way to the testing of increased standard deviation of the market returns described in the previous section. Specifically, the standard errors of the betas computed using the extreme value procedure were compared with the standard errors obtained from the month-end and mid-month return series. The necessity for the Monte Carlo-type study fell away because of the

¹² The chance inclusion of a date near to the crash means that the maximum and average standard deviation of the 1000 random series are higher than that of the closing price series. An explanation for the high mid-month standard deviations is that the mid-month series also happens to coincide with the crash date.

¹³ Many of the securities came from the mining sector which is the more heavily traded on the JSE. The average betas would therefore be expected to be higher than unity given the high systematic risks of the gold and other mining shares.

result reported in the previous section which indicated that the mid-month series promised the same sort of levels of efficiency as the 'best' of the 1000 simulations.

The statistic of primary interest is the *average difference* in the *standard errors* of the beta estimates between those computed from the dates selected according to one of the proposed procedures and those of the usual month-end data. Also of interest is the average difference in the *residual standard deviations*. The invariance of the residual standard deviation with respect to the various data selection procedures is a critical assumption required in the improvement of the precision of the beta estimator. In particular, if the residual standard deviation increases at the same rate as the variance of the market, then any gains in efficiency will be nullified.

A further statistic which was calculated was the mean squared difference (MSD) between the beta estimates of the modified and usual month-end data. The MSD acts as a measure of the variability of the estimates around the month-end estimates. The MSD for the mid-month estimate acts as a benchmark for the MSD of the proposed procedure: provided the MSD of the extreme value proposal is no higher than that of the mid-month MSD¹⁴ then it can be concluded that the selection procedure does not introduce added uncertainty into the estimation.

The statistics referred to above are given in Table 7.4 and are reflected as percentages of the relevant average month-end values, i.e. the MSDs are given as percentages of the average month-end betas, the differences in the standard errors of the betas are given as percentages of the average month-end standard error, etc. The first two rows of Table 7.4 report the statistics for the low-high and high-low versions of the extreme value procedure. The 'significance' of the figures is gauged by comparing them with the 'mid-month' figures in the third row. The final row of the table is different from the others and reports the average beta coefficients, residual standard deviations and beta standard errors. Reading from the fourth row, for example, it can be seen that in the overall period, the mean beta was 1.23, the average residual standard deviation was 0.097 and the average standard error of the beta estimate was 0.227. All the other figures in the table are given as percentages of the appropriate figure in the fourth row.

Focusing on the results in the shaded column of Table 7.4 which give the results for the overall period, it can be ascertained that the mean squared difference between the estimates

¹⁴ The mid-month points are regarded as being a random selection of dates from within the month. The estimates from the mid-month and month-end points should be unbiased with respect to each other and the MSD should therefore measure random influences. The reason is that the dates are based on an essentially exogenous factor, namely a fixed point in a calendar month, as opposed to selecting the dates dependent on the data.

Table 7.4

Percentage differences from month-end figures.

Method	Statistic	88.01-93.11	88.01-89.12	90.01-91.12	92.01-93.11
Alternating high/low (low in first month)	MSD of beta ^a	2.97	7.70	19.86	105.80
	Resid s.d. ^b	.05	.52	.62	1.02
	Std err of beta ^c	-25.4	-29.7	-18.8	+35.2
Alternating High/low (high in first month)	MSD of beta	2.03	10.46	9.52	8.69
	Resid s.d.	.19	-.37	.16	.54
	Std err of beta	-38.6	-55.1	-38.8	-27.7
Mid-month	MSD of beta	3.27	37.84	13.66	12.38
	Resid s.d.	.02	.59	.75	-.14
	Std err of beta	-13.2	-15.4	+21.1	-1.2
Month-end	Average beta	1.23	1.323	1.265	1.754
	Average resid s.d. ^d	.097	.081	.089	.113
	Average std err of beta	.227	.369	.409	.518

^a The mean squared difference (as a percentage of the average month-end beta) between the betas estimated using the modified procedure and those estimated using month-end figures. Use the mid-month figures as a benchmark value for assessing the proposed alternating procedures.

^b The mean difference in residual standard deviation between the procedure and the month-end standard deviation. (Expressed as a percentage of the average month-end standard deviation.)

^c The decrease (-) or increase (+) in the standard errors of the betas relative to the month-end figures. Figures reported as a percentage of the average standard error of beta calculated using month-end figures.

^d The average residual standard error corresponding to a 22-day period (which approximates an average trading month).

of beta obtained from the low-high and high-low procedures were 2.97% and 2.03%, respectively, of the average month-end beta (i.e. 1.23). The MSDs are smaller than the MSD of 3.27% produced when using the mid-month series. The differences between the actual estimates of beta produced by the extreme values procedures and the usual month-end estimates of beta are no bigger than random variation allows for.

The differences between the residual standard deviations of the adaptive procedures and the month-end standard deviations are all less than 1% (i.e. 0.05% for the low-high, 0.19% for the high-low and 0.02% for the mid-month series) of the value of the month-end residual standard deviation. Clearly, the adaptive procedure does not impact substantially on the residual standard deviations.

The most important feature of Table 7.4 is the substantial improvement in the standard errors of the beta estimates offered by the proposed procedure over the whole period and in two of the three sub-periods. Again, looking at the overall figures, the low-high version showed a 25.4% drop in beta standard error and the high-low version a 38.6% drop. The mid-month

procedure (which is essentially an arbitrary selection of dates within each month) offers far less of an improvement (13.2% when the whole period is used).

In summary, the extreme value data selection procedure results in substantial improvements in the efficiency of beta estimation. In fact, the results of the study in this section have indicated that decreases of about 30% can be achieved in the standard errors of the beta estimates by adaptively selecting the input data using the extreme value procedure. As a consequence the adaptive data selection procedure is highly recommended for practical beta estimation.

In the discussion in this section, no mention has been made of the non-normality and thin trading phenomena which characterize small stock markets. The next sub-section therefore contains a brief outline of how the inefficiency and bias associated with non-normality and thin trading, respectively, can be avoided.

7.3.4 Adaptations to the extreme value procedure to incorporate robustness and the trade-to-trade approach

The only exception from the general pattern of results in Table 7.4 outlined above occurs in the sub-period 1992.01-1993.11 when the low-high version of the extreme value procedure is used (i.e. the first row, last column of the table). It can be seen the MSD of the betas is a huge 105.8% and there appears to be an *increase* in the standard error of the estimated betas of about 35%. The figures clearly indicate some major outlying values, which raises an important concept which is discussed first in this sub-section.

In Chapter 6 it was demonstrated that robust regression procedures outperform least squares estimators of beta coefficients for securities on the JSE in terms of efficiency. In particular they showed that one of the most important robustifying procedures was the bounding of the influence of outlying market return values. In other words they showed that the undue influence given to observations associated with market returns which lay far from the rest of the market returns hindered the precise estimation of beta.

It may appear that the proposal (and consequent empirical results) of this chapter contradicts the results of Chapter 6 using, as it does, the *largest* market return values it can find to increase the variance of the market returns. The apparent paradox is cleared up when it is noticed that by picking the largest market returns the average magnitude is being increased of *all* the returns in the observation set used in the estimation period. Therefore the influence

of any *one* large or small market return is being reduced because all the other returns are of the same magnitude. In fact the use of **robust** regression methods continues to be **recommended** practice especially on the modified data set. The robustification will enhance the efficiency beyond the improvement already offered by 'spreading out' **all** the market returns.

Focusing attention on the thin trading characteristic of small markets, it should be noted that although the proposal works well for well traded securities there are additional problems associated with thinly traded securities. The major problem stems from the need to correct for the downward bias in the estimates of beta caused by thin trading. The results from Chapters 4 and 5 have demonstrated that the trade-to-trade estimator is the most efficient of the thin trading correction procedures.

The trade-to-trade estimator selects data from an interval using a slightly different criterion from that of the proposed extreme value procedure: specifically it selects the *last date* on which the *security* traded in the calendar month. The suggested modification to the trade-to-trade estimator would be to select that date in the month on which the index value was at its highest/lowest from the set of all dates on which the security was *actually traded*.

The improvements offered would not be as great as those for the well-traded securities simply because there would be fewer days from which to select an extreme market value. Hence the extreme would inevitably be less marked than an extreme selected from all the days in the month. The decline in efficiency would be relative to the degree of thin trading; nevertheless the improvements that would be observed will still make the extreme value procedure worthwhile.

7.3.5 Conclusion

Based on the partial success of the interval aggregation procedure described in 7.2 for increasing the variance, a so-called 'extreme value procedure' has been designed for increasing the spread of market return values so that the precision of a beta (slope) estimate is enhanced. The selection of dates corresponding to alternately high and low values of the index in each month as the dates between which to measure returns causes a substantial increase in the market variance over that of the (arbitrary) choice of month-end dates. The increase in market variance accentuates the linear relationship between the security and market returns and thus improves the efficiency of beta estimates.

An empirical study was conducted which confirmed that the increased market variance did indeed translate into improved efficiency of beta estimators. Furthermore, it was proved that the increase in efficiency was accomplished without prejudicing unbiasedness and it was shown how the methodology could be extended to include thinly traded securities.

The results of this chapter have demonstrated that the pricing information available *within* each month can be used to increase efficiency without resorting to shorter interval lengths. By doing so, the more significant biasing effects of all price adjustment delays on shorter intervals can therefore be avoided at the same time as efficiency is improved. The efficiency improvements are of important practical consequence and, as a result, the implementation of one or other version of the extreme value procedure is seriously recommended for incorporation into beta estimators.

Appendix 7.A

A description of interval aggregation

In this appendix the daily returns of the market to be denoted by a $T \times 1$ vector, x_d , and it is assumed that there are T days in the investigation. In symbols, the general problem of adaptively selecting the interval to maximize variance can be represented as finding the $N \times T$ matrix A such that the aggregated set of N returns denoted by x and defined by

$$A x_d = x \quad (13)$$

has maximum variance. There are some restrictions on the form of A which are now discussed briefly: a fairly general restriction is that it has the form:

$$A = \begin{bmatrix} 1'_{\Delta_1} & 0 & \dots & 0 \\ 0 & 1'_{\Delta_2} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & 1'_{\Delta_n} \end{bmatrix} \quad (14)$$

where I_k is a k -vector of ones. Note that the Δ_i should sum to T . The restriction on A is more general than it might at first appear. This is because the assumption that the daily returns are serially uncorrelated and homoscedastic implies that the elements of x_d can be permuted. The problem is then fairly easily stated: find the matrix A such that the sample variance of the elements of x is maximized, i.e. so that $(x'x - (I'x)^2)$ is maximized. If T is even fairly large the numerical problems in choosing the $N-1$ positions to split up the elements of x_d for each permutation of the T elements of x_d make the formulation of the problem formidable. A more *ad hoc* manner of finding A is therefore required, namely the interval aggregation procedure.

Appendix 7.B**A brief discussion on the optimality of the proposed procedure**

In the second part of this appendix some optimality issues in the extreme value procedure are discussed.

Using the dates associated with alternate high and then low values in no way guarantees that the return series thus generated will have the highest variance. If the $n-1$ market returns produced as a result of the proposed method in an estimation period comprising n months are denoted by

$$X_{r_j} \quad \text{where } 1 \leq j \leq n-1 \quad (15)$$

and the lengths of the $n-1$ intervals are denoted by δ_j (and it is assumed that the mean market return is close to zero) then the function being maximised can be written as

$$\sum_{j=1}^{n-1} \frac{x_{r_j}^2}{\delta_j} \quad (16)$$

Because the x 's generated by the extreme value procedure are the largest in magnitude that can be constructed (using the fundamental premise of the proposal that there be one observation associated with every basic interval) the proposal would guarantee the largest variance if the lengths of the intervals (the δ 's) were ignored. However, because the measured returns are divided by the square root of the interval lengths the optimality falls away.

In addition the constraint has been imposed that only contiguous periods over which to measure the returns have been considered. The use of disjoint intervals, in which those periods in which there is little or no movement in the market are ignored, might also assist in increasing the variance. The problem associated with this suggestion is explaining why some of the data should be ignored in estimating a model parameter which should hold over all periods.

CHAPTER 8

COVARIANCE ESTIMATION IN THINLY TRADED ENVIRONMENTS

8.1 Introduction

The estimation of covariances is a fundamental step in portfolio analysis and selection as well as multivariate tests of financial models and so its efficient and accurate estimation is essential. As it turns out, the estimators of covariances between securities are biased by thin trading in the same way that the estimators of beta are. In considering possible correction procedures, it is evident that trade-to-trade type approaches are not generally feasible when dealing with covariance estimation in thinly traded environments as it may not always be possible to match returns to days where trades occurred, especially for very thinly traded securities. Hence an aggregated coefficients approach would seem the more suitable approach. The selection of the appropriate number of leads and lags used in the aggregated coefficients estimator is a critical consideration, balancing, as it does, the bias and efficiency of the estimator. In this chapter an adaptive method for selecting a suitable number of lagged and leading coefficient terms is developed.

A novel contribution made in this chapter is the derivation of a covariance estimator under the assumption of autocorrelation in the market return series. It is proved that inefficiency (autocorrelation) in the market will significantly bias estimates of covariance if an aggregated coefficients correction procedure is applied. Therefore an adjustment to remove the bias caused by the autocorrelation in conjunction with the aggregated coefficients estimators is developed in this chapter. It should be noted that the autocorrelation referred to above is the autocorrelation present in the market in addition to that induced by thin trading.

In the next section (8.2) of this chapter the investigation into the extent to which autocorrelation in portfolio returns is caused by thin trading is repeated for the South African market. The results motivate the necessity for examining the biases caused in the aggregated coefficients techniques by autocorrelation. The estimation of covariances in thinly traded environments is directly investigated in section 8.3. An adaptive procedure for selecting the appropriate number of leads and lags is proposed. Additionally, an assumption of market autocorrelation is built into the analysis and consequently a correction factor is derived to

counteract the biases caused by market autocorrelation in covariance estimation. The discussion in section 8.4 extends the results based on the assumption of market autocorrelation into the area of the Cohen *beta* estimator. Although the trade-to-trade estimator is undoubtedly the preferred estimator on markets such as the JSE, it is possible that on other more well traded markets the Cohen estimator may be relatively more efficient. The extension of the results concerning market autocorrelation to beta estimation is thus included not only for completeness, but as a recommended adjustment for practitioners who wish to estimate betas using the Cohen estimator. Finally, in section 8.5, the biases caused by autocorrelation in the very well traded markets are considered, and an adjustment factor derived. The results of this final section are useful for estimating the covariances of securities that trade every day, and where the 'thin trading' is limited to delays in the price adjustment process which are less than a day long. The consideration of market autocorrelation and its implications for covariance estimation form a fairly new area of research and the results of this chapter are therefore valuable, not only for thinly traded markets, but also for the well traded stock markets of the world.

8.2 Non-synchronous security trading and market index autocorrelation on the JSE

In this section the extent to which non-synchronous trading explains observed autocorrelations in the portfolio returns on the JSE is examined. As long ago as 1966 Fisher hypothesized that the autocorrelation observed in portfolio returns could be explained by thin trading. His suggestion helped to explain the paradoxical evidence that *portfolio* returns displayed positive first order autocorrelation although the return series of *individual* securities were found to have little autocorrelation.¹

An important issue which arises from Fisher's hypothesis is establishing the degree to which the autocorrelation in a portfolio's returns is due to thin trading as opposed to other frictions and inefficiencies in the trading process. In their motivation for a study into the magnitude of the autocorrelation caused by thin trading using NYSE data, Atchison, Butler and Simonds (1987) proposed that knowing the source of price adjustment delays² is significant in understanding the price-formation process. The results from this section provide the techniques

¹ See Fama (1965) for empirical results using NYSE securities.

² Cohen *et al* (1980, 1983) class thin trading as one of several price adjustment delays.

needed to estimate the input parameters required in the derivations of covariance and beta estimates described in the remaining essays of this chapter (i.e. 8.2, 8.3 and 8.4). Specifically, the remaining essays demonstrate that the measurement of the level of autocorrelation in the underlying market³ returns is essential for accurate estimation of security covariances (and beta coefficients) when aggregated coefficients methods have to be implemented.

The research of Scholes and Williams (1976, 1977) proves that a model to predict the frequency of trades in securities is required to estimate the extent to which the autocorrelation in portfolio returns is due to thin trading. In their initial working paper, Scholes and Williams (1976) specified a model for non-synchronous trading. The model enabled Atchison *et al* (1987) to use *daily* transaction frequencies to determine (theoretically) the first order autocorrelations caused by thin trading on the NYSE. Specifically, Scholes and Williams postulated that transactions for individual securities arise from independent Poisson processes.⁴ Later, Scholes and Williams (1977) generalized their results to prove that a plausible class of trading processes (including the Poisson process) generated observed returns with the following characteristics: firstly, the observed returns of individual securities will be slightly negatively correlated even if the true return series are uncorrelated. Secondly, cross-correlations between the observed returns of securities will be predominantly positive (even when there is no cross-correlation between the true returns) and hence, thirdly, observed returns of a portfolio of thinly traded shares will be positively autocorrelated.

On the JSE and other smaller stock markets, thin trading is fairly severe relative to the NYSE. In fact, many securities are not traded frequently enough to assume (or estimate the parameters of) a Poisson process for the daily transaction frequencies. In this section it is assumed that all securities which *do* trade on a day do so at the same time, but that for a thinly traded security there is a positive probability that no trade occurs on the day at all. Consequently the focus of the investigation is on measuring the impact that thin trading has on portfolio autocorrelation when the observed returns are measured over intervals of several days, e.g. a month.

In this chapter the suggestion of Scholes and Williams (1976) for the trading process is adapted to be the discrete analogue of the Poisson process, namely a series of independent

³ Because any autocorrelation in the market is usually estimated from a *portfolio* of securities which form a market index, it is useful to be able to detect how much of the estimated autocorrelation is due to thin trading and how much is an estimate of autocorrelation in the true market returns.

⁴ The fundamental assumption of the Poisson process is that trades occur at random during the day.

Bernoulli trials (which is termed a Bernoulli process). A standard result which follows from the assumption of a Bernoulli process for the trades is that the intervals between trades can be modelled by a geometric distribution with parameter $1 - q_j$ (where q_j is defined to be the probability that security j does not trade on a given day).

The structure of this section is as follows. Firstly, in sub-section 8.2.1 a brief theoretical development is given outlining how thin trading causes the autocorrelation in portfolio returns. The discussion here is fairly standard and follows the results discussed in Scholes and Williams (1977) and Atchison *et al* (1987). The additional contributions made are technical ones for deriving the theoretical autocorrelations from the discrete trading process. Sub-section 8.2.2 consists of an empirical study designed to contrast the theoretically expected autocorrelation caused by thin trading with observed autocorrelation in portfolios. Sub-section 8.2.3 contains the conclusions of the study.

8.2.1 Theoretical development

The theoretical development below describes a few of the central conclusions of Scholes and Williams (1977) and Atchison *et al* (1987) in their papers. Following the discussion of the theoretical results derived by Scholes and Williams and Atchison *et al*, the formulae for calculating the theoretical autocorrelations due only to thin trading for the discrete trading process are explained. The initial discussion centres on the expression for portfolio autocorrelation.

If the observed return on a share j is denoted by R_{jt}^T and the return on an equally weighted portfolio by R_{pt}^T then Atchison *et al* show that

$$\text{Corr}[R_{pt}^T, R_{pt-1}^T] = \frac{\frac{1}{n} \text{ave. autocovariance} + \frac{n-1}{n} \text{ave. cross-covariance}}{\frac{1}{n} \text{ave. variance} + \frac{n-1}{n} \text{ave. covariance}} \quad (1)$$

A consequence which can be seen from the above expression is that the observed returns of a portfolio depend not only on the autocovariances of the observed security returns, but also on the cross-covariances of the returns. Therefore, even if the security returns have little or no autocorrelation, but do have positive cross-correlation, then the observed portfolio autocorrelation will be positive. In order to gauge the expected sizes of the components of the

autocorrelation expression, the results established by Scholes and Williams (1977) are invoked. Scholes and Williams demonstrated that the cross-covariances of individual securities can be written as

$$\text{Cov}[R_{it}^T, R_{jt-1}^T] = E[\max(I_i - I_j, 0)] \text{Cov}[R_{it}, R_{jt}] \quad (2)$$

where I_i and I_j are the price ages⁵ of securities i and j , and $\text{Cov}[R_{it}, R_{jt}]$ is the covariance of the *true* (as opposed to observed) returns on shares i and j . The interpretation of the expression in (2) is that if the two securities i and j have the same trading frequencies then there will be zero cross-correlation because I_i and I_j will be equal. On the other hand, if there is positive correlation between the *contemporaneous* true returns of the two securities, then there will be positive cross-correlations in the observed return series.

Scholes and Williams also show that the covariances between the observed security returns are given by

$$\text{Cov}[R_{it}^T, R_{jt}^T] = [1 - \alpha_{q,q_i} - \alpha_{q,q_j}] \text{Cov}[R_{it}, R_{jt}] \quad (3)$$

where the following definition holds:

$$\alpha_{q,q_i} = E[\max(I_i - I_j, 0)] \quad (4)$$

Since the α terms (which can be thought of as the expected differences in the price ages of the securities) are by definition greater than or equal to zero, the covariances between the observed returns will be less than the covariances between the true returns. The importance of the reduced covariances in the observed returns arises because the covariances appear in the *denominator* of the expression for the autocorrelation of the portfolio returns (equation (1)) and result in an apparently increased portfolio autocorrelation.

⁵ A price age is the period of time between the last trade of the security and the end of the interval.

Other results which relate observed security autocovariances with true autocovariances and observed variances with true variances which stem from the Scholes and Williams (1977) model are listed below for later reference:

$$\text{Cov}[R_{it}^T; R_{it-1}^T] = - \left[\frac{\text{Var}[I_i]}{\nu_i^2} \right] \text{Var}[R_{it}] \quad (5)$$

and

$$\text{Var}[R_{it}^T] = \left[1 + 2 \frac{\text{Var}[I_i]}{\nu_i^2} \right] \text{Var}[R_{it}] \quad (6)$$

where

$$\nu_i^2 = \frac{\text{Var}[R_{it}]}{E[R_{it}]^2} \quad (7)$$

The squared coefficient of variation, ν_i^2 , is generally large.⁶ The large squared coefficients of variation have the effect of severely down-weighting the $\text{Var}[I_i]$ terms in equations (5) and (6) and thus it can be seen that the autocovariances and variances between the observed returns are good estimates of the relationships between the true returns.⁷ Consequently, in order to estimate the theoretical portfolio autocorrelations the model for the thin trading process only has to provide estimates of the expected difference in the price ages of the two securities (i.e the α coefficients defined in equation (4)). Although Atchison *et al* give expressions for the α coefficients in the case of a well traded market (which correspond to those markets in which securities trade generally at least once a day), the corresponding expression for thinly traded markets has not yet been established.

⁶ See Scholes and Williams (1977) for evidence from the NYSE. In order to get some feel for the magnitude of the parameter on the JSE, the reader is referred to Table 4.2 in Chapter 4. Using the standard deviation and mean return on the JSE-Actuaries Overall Index reported in the table, the coefficient of variation is calculated to be about 16.7. Because the Overall Index can be thought of as a diversified portfolio of shares and hence will have a smaller variance component than the individual shares, the value 16.7 can be considered a lower bound for the coefficients of variation.

⁷ The small negative autocovariance in individual security returns predicted by the Scholes and Williams (1977) model can be seen to arise from equation (5).

In Appendix 8.A a derivation of the α coefficients is given based on the discrete analogue of the Poisson process which underpins Atchison *et al*'s model. The expression is in terms of the trading infrequencies of the two securities q_j and q_i and the number of days in the return intervals, k . The expression for the α coefficients has the form:

$$\alpha_{q_i q_j} = \frac{1}{1 - q_i q_j} \left[\frac{1 - q_j}{1 - q_j^k} - q_i^k q_j^k \frac{1 - q_i}{1 - q_i^k} \right] / k \quad (8)$$

where k is the number of days in the interval over which the returns are measured, e.g. $k = 20$ for four-weekly returns, and q_i is the probability that share i does not trade in a day.

The basic requirements for the calculation of the theoretical autocorrelation in a portfolio are estimates of q_i , q_j and $Cov[R_{it}, R_{jt}]$. With these values the cross- and autocovariances for shares i and j can be estimated and used in equation (1) to calculate the theoretical autocorrelation of the portfolio.

In the empirical study described in the next section, estimates of the covariance between the true returns of securities i and j were obtained from the market model:

$$Cov[R_{it}; R_{jt}] = \hat{\beta}_i \hat{\beta}_j Var[R_{mt}] \quad (9)$$

where R_{mt} is the true return on the market. The parameters q_i and q_j were obtained from the proportion of days not traded by shares i and j , respectively, using the daily data which were available.

8.2.2 Empirical investigation

The data used were the 506 shares which had continuous trading data for at least 240 weeks ending on 8 July 1992, and which had traded on at least 5% of the days for which daily data were available (October 1987 - July 1992).

Following the methodology of Atchison *et al*, portfolios with 1, 2, 3, ..., 165 securities chosen randomly from the available 506 shares were constructed. For each portfolio the theoretical autocorrelation (from equation (3)) and the observed autocorrelation were calculated. The above step was repeated two hundred times in order to obtain an average theoretical and observed portfolio autocorrelation at each portfolio size. The results are presented in Figure 8.a.

As can be seen from Figure 8.a the theoretical autocorrelation (under the assumption of the cause of the autocorrelation being thin trading) is very small relative to the observed autocorrelation. The observed autocorrelation rises sharply from a small negative value (as predicted by the Scholes and Williams model) and then gradually levels out as the portfolio size increases. Although the theoretical autocorrelation displays the same sharply increasing shape as the observed autocorrelation function it is clearly of a completely different scale. In fact, according to the model, thin trading appears to explain just under 5% of the observed

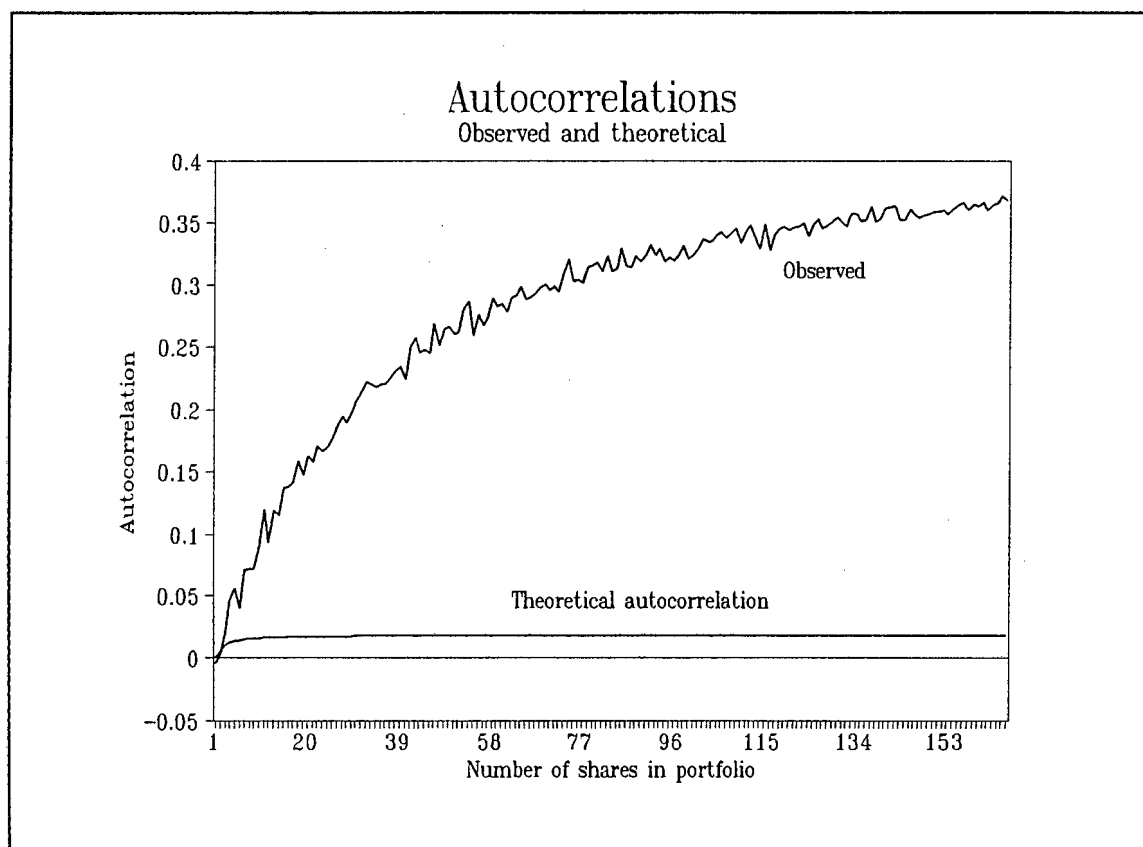


Figure 8.a

autocorrelation. The remainder of the observed autocorrelation can consequently be attributed to other inefficiencies in the pricing process. Using the NYSE securities Atchison *et al* reported a similarly low explanatory power of the thin trading model (15.8%). Although the low explanatory power may arise from either a poor *model* for thin trading or other frictions in the pricing process, Atchison *et al* argue that any biases in the Poisson-type processes will in fact lead to apparently larger theoretical autocorrelations.

8.2.3 Conclusions

The results in this section of the chapter have demonstrated that thin trading explains only a very small component of the observed autocorrelation of portfolio returns. Because stock market indices are generally portfolios of securities, most of the autocorrelation in the *index* returns cannot be explained by thin trading alone.

The average autocorrelations of large portfolios appear to be in the region of 0.4. Further, only about 5% of the autocorrelation in the equally weighted portfolio can be explained by thin trading. The study indicates that the possibility of autocorrelation in the return series of indices (over and above that caused by thin trading) on the JSE cannot be ignored simply because individual security autocorrelations are negligible. Furthermore, results of the model suggest that researchers must take into account the fact that most of the autocorrelation is due to frictions in the pricing process other than thin trading.

The main conclusion therefore is that there is evidence which indicates that significant autocorrelation exists in any large portfolio of securities, e.g. the market indices, even after the effect of thin trading has been considered.

In the next section an adaptive estimator of covariances is developed and the consequences of the autocorrelation in the market index on estimates of covariance are examined.

8.3 On estimating the covariance structure of returns in thinly traded environments

8.3.1 Introduction

The estimation of the covariance matrix of returns has always played an important role in financial empirical work. It is well known that the covariance component between securities is the most important consideration for risk reduction in portfolio design (see for example Brealey and Myers (1991, pp 119-125), and Copeland and Weston (1988, pp 153-155)). Shanken (1987) emphasized the important role that the estimation of the covariance matrix of returns plays in financial empirical work and cited the use of multivariate methods in tests of asset pricing relations as well as multifactor models as examples. For examples of studies that employ the residual covariance matrix of returns in multivariate tests of asset pricing models

see Gibbons (1982), Stambaugh (1982), Jobson and Korkie (1982) and Shanken (1985). Roll and Ross (1980), Reinganum (1981), Chen (1983) and Brown and Weinstein (1983), on the other hand, have employed the covariance matrix of returns for research into multifactor models.

Although Scholes and Williams (1977) have indicated that for non-synchronous data, measured covariances understate (in absolute magnitude) true covariances, most researchers have largely ignored this problem and concentrated instead on beta estimation in the presence of non-synchronous data. Scholes and Williams (1977) and Dimson (1979), for example, have proposed techniques for beta estimation based on aggregating leading and lagged regression coefficients. Cohen *et al* (1983) pointed out a small inconsistency in Dimson's (1979) derivation of the aggregated coefficients method and presented a revised estimator. Further, Dimson and Marsh (1983), among other researchers, have advocated a procedure of matching the market index to dates when securities are traded (known as the trade-to-trade approach). The trade-to-trade approach is only applicable in beta estimation in cases where the market index itself is well traded. The trade-to-trade approach is clearly not suitable for covariance estimation among thinly traded securities as it would not always be possible to find times when the two securities traded simultaneously.

Shanken (1987) has suggested that covariance estimation between individual securities be corrected for the effects of non-synchronous trading using the aggregated coefficients approach. Shanken (1987) pointed out that the derivation of the beta estimator (for non-synchronous data) presented by Cohen *et al* (1983) contains an explicit expression for the estimation of the covariance between two securities. In his study Shanken (1987) selected ten groups of 30 securities from the NYSE and computed (using daily data) the average covariance, both adjusted and unadjusted for non-synchronous trading, for each of the ten subgroups. He found that the adjusted estimates were nearly twice that of the unadjusted estimates and that over 97% of individual covariance estimates increased when the adjusted estimator was used. On the basis of these results Shanken (1987) concluded that price-adjustment delays induce substantial cross-serial correlation between individual observed security returns.

In many of the smaller non-USA markets the extent of thin trading (non-synchronous trading) is severe and has been shown to have a major impact on beta estimation. Dimson (1979) and Dimson and Marsh (1983), for example, have presented evidence on the London Stock Exchange and Chapters 3, 4 and 5 of this thesis give comprehensive evidence from the JSE. Consequently, it appears likely that covariance estimation will also suffer from similar

effects. The major problem in comparing estimators of the covariance is that the parameter that is being estimated, i.e. the 'true' covariance, is unknown. There is, therefore, no yardstick by which to compare empirical estimates of the covariance. Although Shanken (1987) highlighted the impact of non-synchronous trading on covariance estimation of NYSE stocks, for example, he had no knowledge of the 'true' covariance and therefore of whether the correction procedure he had implemented was indeed operating accurately.

In this section an analytical approach will be used to demonstrate how varying levels of thin trading impact on covariance estimation. Furthermore, an adaptive approach to implementing the covariance estimator is proposed to account for the differing levels of thin trading of the securities. The adaptive approach considers a suitable selection procedure for the number of leads and lags to be used depending on the level of thin trading of the securities involved.

An additional aspect of covariance estimation considered in this chapter is the effect on covariance estimation that autocorrelation in the true market returns has. Cohen *et al* (1983) and Atchison *et al* (1987), for example, have shown that autocorrelation exists in market returns in addition to that induced by thin trading. This issue is also investigated here. The autocorrelation in true market returns is shown to lead to upwardly biased estimates of covariance when the Cohen estimator is used. To date, the assumption of zero autocorrelation in the market proxy has not been relaxed in any of the derivations of the thin trading correction models. The assumption of non-zero autocorrelation is made in the ensuing sections of the chapter in order to quantify the extent to which it biases estimates of covariance and beta, and to develop a correction procedure for the bias.

The rest of this section is organized as follows: In sub-section 8.3.2 the covariance estimator is expressed analytically in terms of the trading frequencies of the two securities for two situations: first, for the case where there is no autocorrelation in the true returns of the market index and, second, for the more general case when this autocorrelation is assumed to be non-zero. Sub-section 8.3.3 contains a brief discussion on the methodology used to obtain the expected values for the covariance estimator. The theoretical results are used to derive an adaptive procedure for selecting the appropriate number of leading and lagged coefficients to be used in the Cohen estimator. Sub-section 8.3.4 demonstrates how the adaptive estimator can be adjusted for the case of non-zero autocorrelation in the true market returns. In the final sub-section (i.e. 8.3.5) conclusions and a summary are given.

8.3.2 Theoretical background

The main cause of the bias associated with estimation problems in thinly traded environments has been discussed several times in this thesis and is cited as being the fact that recorded prices are used to represent true, underlying prices. Clearly, estimation problems arise when the series of recorded rather than underlying prices are used in the estimation procedures. In particular, it is evident that the covariance between two recorded securities' returns is likely to be less than the covariance between their underlying returns. The two types of covariance are referred to as the 'observed covariance' and the 'true covariance' between the two securities. The difference between the observed and true covariance is clearly due to the fact that underlying prices reflect movements in the market instantaneously while recorded prices may remain unchanged.

In deriving their corrected beta estimator, Cohen *et al* (1983) showed that the covariance between the true returns of two securities, j and k , comprised the contemporaneous covariance of the observed returns as well as leading and lagging serial cross-covariances. That is

$$\begin{aligned} \text{Cov}(R_{j,t} ; R_{k,t}) = & \text{Cov}(R_{j,t}^o ; R_{k,t}^o) + \sum_{n=1}^L \text{Cov}(R_{j,t}^o ; R_{k,t-n}^o) \\ & + \sum_{n=1}^K \text{Cov}(R_{j,t-n}^o ; R_{k,t}^o) \end{aligned} \quad (10)$$

where

- $R_{j,t}$ is the underlying return on security j at time t ;
- $R_{j,t}^o$ is the *observed (recorded)* return on security j at time t ;
- K is the number of leads;
- L is the number of lags.

In the above formulation the Cohen *et al* (1983) covariance estimator has been extended to allow for unequal numbers of leads and lags. In deciding on the appropriate numbers of leads and lags one must note that as K and L are increased, the model's potential representation of reality is increased, but its statistical efficiency declines as more potential noise is included. On the basis of an exploratory analysis Shanken (1987) used three daily lags and leads for NYSE stocks to adjust for covariance bias. On the basis of a study on the UK stock market

Dimson (1979) recommended that one monthly lead and at least four monthly lagged coefficients should be used in beta estimation.

The Cohen covariance estimator (shown in equation (10)) has two appealing properties. Firstly, Cohen *et al* (1983) have shown it to be a consistent estimator of the true covariance as the interval of differencing is increased; and, secondly, for two well traded securities, the estimator converges to the underlying covariance in efficient markets. The latter can be confirmed by noting that in efficient markets, there will be theoretically no serial correlation and all the non-contemporaneous covariances between two well traded securities will be zero.

Under the assumption of zero market autocorrelation, the expected values of the component terms⁸ of the Cohen covariance estimator (given on the right hand side of equation (10)) can be derived in terms of the trading frequencies of the securities. The derivation begins by noting that a fundamental assumption in Cohen *et al*'s development is that the observed returns on a security can be expressed as:

$$R_{j,t}^o = \sum_{n=0}^{N^*} \gamma_{j,n} R_{j,t-n} \quad (11)$$

where N^* is the maximum number of intervals for which price adjustment delays are non-negligible. The formulation in equation (11) is very general and allows for all price adjustment delays in the trading process. If, however, it is assumed that all trades do occur at the true prices of the security then all price adjustment delays other than infrequent trading can be ignored. (This assumption is very similar to that in Dimson (1979) and has been used previously in this thesis because the thin trading of the JSE dominates all the other delays.) The expected observed return is thus a weighted average of the returns in the current and previous intervals. If it is further assumed that trades occur in an interval at random⁹ and if N^* is allowed to grow to infinity, then the weighting coefficients for the expected observed return on the security are given by

$$\gamma_{j,n} = q_j^n p_j \quad (12)$$

⁸ Strictly speaking, it is not the expected *covariances* which are being calculated, but rather the covariances of the expected observed *returns*. However, to avoid clumsy mathematical expressions and discussions, the phrase 'expected covariance' will be used in this chapter.

⁹ In other words, a Bernoulli process is assumed for the trading process.

where p_j can be interpreted as the probability that security j does trade in any one interval and q_j is defined to be $1 - p_j$. The expected observed return in equation (11) is then equivalent to

$$R_{j,t}^o = \sum_{n=0}^{\infty} p_j q_j^n R_{j,t-n} \quad (13)$$

The reason for making the assumptions outlined above is that the expected performance of the Cohen covariance estimator can then be evaluated analytically. In Appendix 8.B it is shown that the assumed structure of the expected observed returns implies that a typical term from the right hand side of the Cohen covariance estimator in equation (10) has the expectation:

$$E[\text{Cov}(R_{j,t-n}^o ; R_{k,t}^o)] = \text{Cov}(R_{j,t} ; R_{k,t}) \times p_j p_k \frac{q_k^n}{1 - q_j q_k} \quad (14)$$

From the expression in equation (14) above it can be seen that once a value for the true covariance ($\text{Cov}(R_{j,t} ; R_{k,t})$) is fixed, the expected values of the terms of the Cohen covariance estimator can be calculated using only the trading frequencies (the p_k and p_j) of the two securities.

In the investigation described in section 8.3.3, the ability to calculate the expected values of the terms in the Cohen covariance estimator is used to establish an adaptive procedure for selecting the number of leads and lags necessary for the estimator. The decision as to which leads and lags to omit is based on the expected sizes of the terms of the estimator given the trading frequencies of the two securities involved.

8.3.2.1

The non-zero market auto-correlation case

In this section the expected values of the terms of the Cohen covariance estimator are calculated under the additional assumption that there is autocorrelation in the market returns. It is emphasized that the autocorrelation referred to here represents the autocorrelation in the series of true market returns. Consequently, for market indices which do not suffer from the effects of thin trading (i.e. consist of only well traded securities) any market autocorrelation present in the true market returns can be estimated directly from the observed market index return series. For market indices which do suffer from the effects of thinly traded component securities, direct estimates of market index autocorrelation generally reflect two sources: first,

autocorrelation induced by infrequent trading in the component securities, and secondly autocorrelation in the series of true market returns. For a discussion on estimating the magnitude of the two sources of autocorrelation mentioned above, see Atchison, Butler and Simonds (1987) or section 8.2 in this chapter.

In order to gain some idea of the levels of autocorrelation in market indices it is worth noting that empirical estimates of the total autocorrelation for US, Canadian and UK indices were found to be 0.19, 0.26 and 0.25, respectively (see Dimson (1979)). The results in 8.1 and those reported by Atchison *et al* (1987) suggest that very little of the autocorrelation is attributable to thin trading.

Following Cohen *et al* (1983) the true returns for a security j in interval t are assumed to be generated by the market model:

$$R_{j,t} = \alpha_j + \beta_j R_{M,t} + u_{j,t} \quad (15)$$

The usual distributional assumptions about $u_{j,t}$ are made: namely, serial independence and stationarity. However, the novelty in the development arises in the different assumption made about true market returns $R_{M,t}$. In order to allow for autocorrelation in the market returns (and consequently, using equation (15), in the security returns as well) it is assumed that the market returns follow an autoregressive process of order one. The market returns are thus serially correlated with

$$\text{Corr}(R_{M,t} ; R_{M,t-1}) = \rho \quad \text{for all } t.$$

Using the above assumption, a similar expected covariance value to that in equation (14) can be derived, but which now includes the effect of market autocorrelation. In Appendix 8.B the mathematical details of the derivation¹⁰ of the expected value of the Cohen covariance terms is set out for non-zero ρ . The expressions for the expected covariances depend

¹⁰ To date no derivation for the non-zero ρ has been forthcoming in the literature. Thus the development in Appendix 8.B is novel but has been relegated to the appendix as the algebra is fairly tedious.

on the number of lags (or leads) in the particular component being considered. An example of the expected value for one of the Cohen covariance component terms is

$$E[Cov(R_{j,t-n}^o ; R_{k,t}^o)] = Cov(R_{j,t} ; R_{k,t}) p_j p_k \frac{\rho^{n+1}(1-q_j q_k)(1-\rho q_k) + q_k^{n+1}(\rho^2-1)(1-\rho q_j)}{(1-\rho^2)(\rho-q_k)(1-\rho q_j)(1-\rho q_k)(1-q_j q_k)} \quad (16)$$

It can be seen from the above expression that the expected covariance of the component in which security j 's returns are lagged n intervals behind security k 's returns is equal to the true (synchronous) covariance multiplied by a factor which depends on the trading frequencies of the two securities and the autocorrelation coefficient.

In the next section the analytical results are used to design an investigation which is intended to lead to an adaptive manner of selecting the appropriate number of leads and lags for the Cohen covariance estimator.

8.3.3 Methodology

As a starting point to the analysis a cross-section of realistic levels of thin trading which typify smaller markets was required. For the purposes of this study it was assumed (without loss of generalization of the methodology) that weekly data would be used in the covariance

Table 8.1

Average percentages of weeks securities were not traded over the period 1 Jan 1981 to 31 Dec 1986.

Decile	Percentage of weeks in which no trades occurred
1	0.00
2	0.07
3	0.43
4	1.36
5	3.05
6	6.70
7	13.43
8	24.20
9	40.30
10	68.36

calculation. Were other data frequencies used, the methodology for determining the number of leads and lags would have to be repeated using the distribution of trading frequencies based on the, say, monthly, biweekly or daily data. The distribution of probabilities that securities would not trade in a week was extracted from the study on the JSE by Bradfield (1990). Table 8.1 shows this distribution for the population of securities grouped into ten deciles ranked on the basis of their thin trading.

The procedure used to examine the efficacy of the Cohen estimator assumed, initially, zero market autocorrelation (i.e. $\rho = 0$) and an equal number of lags and leads (i.e. $L = K$). For a security, j , from each of deciles one through ten, the expected values of the components of the Cohen covariance estimator (in equation (10)) were calculated for securities, k , in each of the deciles. The expected values were calculated using equation (14).

In the analysis that follows the extent to which lagged and leading coefficients contributed towards unbiased estimation of covariances between securities of all levels of thin trading could be ascertained. Using this investigation as a basis the selection of L and K could be made in an adaptive manner depending on the extent of thin trading.

In the second part of section 8.3.4 the impact of non-zero autocorrelation is examined and a correction procedure derived for any biases in the estimates of covariance.

8.3.4 Results

The expected values of the components of the Cohen covariance estimates, i.e. equation (10) (with $\rho = 0$) are shown in Table 8.2 through to Table 8.5. In Table 8.2 through to Table 8.5 the expected values of the lagged coefficients (i.e. the $E[\text{Cov}(R_{j,t}^o; R_{k,t-n}^o)]$), the unadjusted synchronous covariance estimates (i.e. the $E[\text{Cov}(R_{j,t}^o; R_{k,t}^o)]$) and the leading coefficients (i.e. the $E[\text{Cov}(R_{j,t-n}^o; R_{k,t}^o)]$) are presented. The expected values have been calculated for a 'control security', j 's covariance with returns on securities, k , in deciles one to ten. Tables for j in deciles 1, 4, 7 and 10 have been included for illustrative purposes.¹¹ To facilitate interpretation without loss of generalization, a true covariance of unity has been assumed. Setting the true covariance to unity permits the expected observed covariances to be interpreted as percentages of the true value.

¹¹ In the tables presented the number of leads and lags have been set equal to five ($N = 5$) in all cases, i.e. five lagged and five leading coefficients are included in addition to the synchronous estimate. In arriving at the adaptive procedure for selecting the numbers of leads and lags, the expected component values were examined for a much larger number of leads and lags. The procedure for selecting the number of leads and lags is, however, clearly evident from the reduced set.

Table 8.2

Expected cross-covariance estimates with security from decile 1

		Lagged Coefficients						Leading Coefficients						Total ^a
Decile	Thin trading %	-5	-4	-3	-2	-1	0	1	2	3	4	5		
1	0.00	-	-	-	-	-	1	-	-	-	-	-	1	
2	0.07	-	-	-	-	-	1	-	-	-	-	-	1	
3	0.43	-	-	-	-	-	1	-	-	-	-	-	1	
4	1.36	-	-	-	-	.01	.99	-	-	-	-	-	1	
5	3.05	-	-	-	-	.03	.97	-	-	-	-	-	1	
6	6.70	-	-	-	-	.06	.93	-	-	-	-	-	1	
7	13.43	-	-	-	.02	.12	.87	-	-	-	-	-	1	
8	24.20	-	-	.01	.04	.18	.76	-	-	-	-	-	1	
9	40.30	.01	.02	.04	.10	.24	.60	-	-	-	-	-	1	
10	68.36	.05	.07	.10	.15	.22	.32	-	-	-	-	-	.9	

^a The total is the total expected covariance calculated using five leads and five lags. It is obtained by adding all the figures in the corresponding row.

(Note that - indicates a value which is less than 0.005 and which has been omitted for clarity of presentation.)

Table 8.2 represents the situation where a well traded security (from decile 1) is cross-correlated with a range of securities having various levels of thin trading. Table 8.3 and Table 8.4 give the results for securities with intermediate trading frequencies and Table 8.5 (at the other extreme of trading frequency) represents the situation where a typical, very thinly traded security (from decile 10) is cross-correlated with the same range of securities having various levels of thin trading.

After inspecting all four tables, one of the salient features of all the tables is that the unadjusted covariances (indicated by a lag/lead of zero) decrease as the trading frequency in the k security decreases. In other words, the expected values of the observed covariances are closer to the true covariances for the well traded securities. The difference between the expected and the true covariances increases as the level of thin trading increases. The decrease in observed covariance is consistent for all four 'control securities' confirming that the more frequently either one of the two securities used in the covariance estimation trade, the less downward bias there will be. Consider, for example, the evidence from Table 8.5 (where a hypothetical security from decile 10 is cross-correlated with securities in the other deciles). Table 8.5 shows that the expected unadjusted covariance between a decile 10 and a decile 1 security is 0.32, while the average unadjusted covariance between 2 securities from decile 10 is only 0.19. This indicates that the covariance between two thinly traded securities suffers

Table 8.3

Expected cross-covariance estimates with security from decile 4

Decile	Thin trading %	Lagged Coefficients							Leading Coefficients					Total
		-5	-4	-3	-2	-1	0	1	2	3	4	5		
1	0.00	-	-	-	-	-	.99	.01	-	-	-	-	1	
2	0.07	-	-	-	-	-	.99	.01	-	-	-	-	1	
3	0.43	-	-	-	-	-	.98	.01	-	-	-	-	1	
4	1.36	-	-	-	-	.01	.97	.01	-	-	-	-	1	
5	3.05	-	-	-	-	.03	.96	.01	-	-	-	-	1	
6	6.70	-	-	-	-	.06	.92	.01	-	-	-	-	1	
7	13.43	-	-	-	.02	.11	.86	.01	-	-	-	-	1	
8	24.20	-	-	.01	.04	.18	.75	.01	-	-	-	-	1	
9	40.30	.01	.02	.04	.10	.24	.59	.01	-	-	-	-	1	
10	68.36	.05	.07	.10	.15	.22	.32	-	-	-	-	-	.9	

from more bias than the covariance between a thinly traded and a well traded security.

Another important feature of Table 8.2 through to Table 8.5 is that the covariance biases increase rapidly as the 'control security' becomes more thinly traded (i.e. moving from Table 8.2 through to Table 8.5). For example, in Table 8.2 where the 'control security' is from

Table 8.4

Expected cross-covariance estimates with security from decile 7

Decile	Thin trading %	Lagged Coefficients							Leading Coefficients					Total
		-5	-4	-3	-2	-1	0	1	2	3	4	5		
1	0.00	-	-	-	-	-	.87	.12	.02	-	-	-	1	
2	0.07	-	-	-	-	-	.87	.12	.02	-	-	-	1	
3	0.43	-	-	-	-	-	.86	.12	.02	-	-	-	1	
4	1.36	-	-	-	-	.01	.86	.11	.02	-	-	-	1	
5	3.05	-	-	-	-	.03	.84	.11	.02	-	-	-	1	
6	6.70	-	-	-	-	.05	.82	.11	.01	-	-	-	1	
7	13.43	-	-	-	.01	.10	.76	.10	.01	-	-	-	1	
8	24.20	-	-	.01	.04	.16	.68	.09	.01	-	-	-	1	
9	40.30	.01	.01	.04	.09	.22	.55	.07	.01	-	-	-	1	
10	68.36	.05	.07	.10	.14	.21	.30	.04	.01	-	-	-	.9	

decile 1 (i.e. is a well traded security), the average unadjusted covariance with another security

from decile 1, is 1.00, and it decreases monotonically to an average of 0.32 for decile 10. In Table 8.4, where the 'control security' is from decile 7 (i.e. is not traded in 13.4% of the periods), the corresponding expected values decrease from 0.86 to 0.30. Finally, in Table 8.5 where the 'control security' is from decile 10 (i.e. is not traded in 68.4% of the periods), the corresponding estimation bias is extreme, with the expected covariance estimates decreasing from 0.32 to 0.19.

Turning to the non-synchronous covariance estimates and their role in the adjusted covariance estimator, several features emerge. In all four tables the more thinly traded securities have the largest non-zero lagged non-synchronous covariances. Furthermore, as the

Table 8.5

Expected cross-covariance estimates with security from decile 10

Decile	Thin trading %	Lagged Coefficients							Leading Coefficients				Total
		-5	-4	-3	-2	-1	0	1	2	3	4	5	
1	0.00	-	-	-	-	-	.32	.22	.15	.10	.07	.05	.90
2	0.07	-	-	-	-	-	.32	.22	.15	.10	.07	.05	.90
3	0.43	-	-	-	-	-	.32	.22	.15	.10	.07	.05	.90
4	1.36	-	-	-	-	-	.32	.22	.15	.10	.07	.05	.90
5	3.05	-	-	-	-	.01	.31	.21	.15	.10	.07	.05	.90
6	6.70	-	-	-	-	.02	.31	.21	.14	.10	.07	.05	.90
7	13.43	-	-	-	.01	.04	.30	.21	.14	.10	.07	.05	.90
8	24.20	-	-	-	.02	.07	.29	.20	.13	.09	.06	.04	.91
9	40.30	-	.01	.02	.04	.11	.26	.18	.12	.08	.06	.04	.91
10	68.36	.03	.04	.06	.09	.13	.19	.13	.09	.06	.04	.03	.88

control security becomes more thinly traded, i.e. moving from Table 8.2 through to Table 8.5, the number of leads becomes increasingly important. Consequently in the cases where two thinly traded securities are involved, it is evident that both leads and lags are important.

The features of the tables outlined above indicate that the non-synchronous covariances are indeed capturing the pricing delays induced into the series of observed returns by thin trading. Equation (10) is designed precisely to incorporate the non-synchronous covariances and so remove estimation bias.

The column headed 'Total' in each of the Tables gives the Cohen estimate of the covariance (with $N = 5$). In all cases the Cohen estimate is closer to the true covariance (set equal to unity for ease of interpretation) than the unadjusted synchronous estimate is.

By separating out the individual components of the Cohen estimator additional insights into the number of leads and lags required are gained. For example, if securities j and k are in decile 1, no leads or lags are required, whereas for securities with more extreme thin trading, lagged and leading coefficients are needed. Including leads and lags where they are not necessary for picking up non-synchronous covariance induces inefficiency into the estimator. Therefore, an adaptive estimator is proposed in which the number of leads and lags are selected based on the extent of thin trading of the less traded of securities j and k .

The appropriate number of leads and lags suggested by extensions of the above analysis were used to specify an adaptive procedure. The resulting number of leads and lags are summarized in Table 8.6.

Table 8.6

Leads and lags appropriate for various levels of thin trading

% not traded for most thinly traded security	Decile of most thinly trade security	Corresponding number of leads and lags
0.0 - 0.5	1, 2 or 3	0
0.5 - 10.0	4, 5 or 6	1
10.0 - 20.0	7	2
20.0 - 30.0	8	3
30.0 - 50.0	9	5
50.0 +	10	15

The first column of Table 8.6 represents the range of expected percentages of return intervals having no trades for the more thinly traded of the two securities. The second column gives an indication of the corresponding deciles (for JSE stocks) that are included in this range. The third column suggests the suitable number of leads and lags required by the covariance estimator.

The expected values of the adaptive Cohen covariance estimator for all combinations of securities are shown in Table 8.7. The final estimates of the covariances show a substantial improvement when compared to the unadjusted, synchronous estimates of the covariances shown in Table 8.8. For example, in Table 8.7 where the adaptive procedure is used, the

Table 8.7

Expected adaptive Cohen-based estimates

Decile	Thin trading %	Decile									
		1	2	3	4	5	6	7	8	9	10
1	0.00	1.00									
2	0.07	1.00	1.00								
3	0.43	1.00	1.00	.99							
4	1.36	1.00	1.00	1.00	1.00						
5	3.05	1.00	1.00	1.00	1.00	1.00					
6	6.70	1.00	1.00	1.00	1.00	.99	.99				
7	13.43	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
8	24.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.99		
9	40.30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.99	
10	68.36	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

observed covariance between two decile 10 securities has an expected value of unity which corresponds to the true covariance. In contrast, the expected value of the observed covariance without the adaptive Cohen estimator is given as 0.19 in Table 8.8.

Having established in this section that the adaptive Cohen estimator does indeed provide unbiased estimates for the case of zero autocorrelation, the investigation proceeds by considering the more general case when the assumption of zero autocorrelation is dropped.

Table 8.8

Expected unadjusted covariance estimates

Decile	Thin trading %	Decile									
		1	2	3	4	5	6	7	8	9	10
1	0.00	1.00									
2	0.07	1.00	1.00								
3	0.43	1.00	1.00	.99							
4	1.36	.99	.99	.98	.97						
5	3.05	.97	.97	.97	.96	.94					
6	6.70	.93	.93	.93	.92	.91	.87				
7	13.43	.87	.87	.86	.86	.84	.82	.76			
8	24.20	.76	.76	.76	.75	.74	.72	.68	.61		
9	40.30	.60	.60	.60	.59	.59	.57	.55	.50	.43	
10	68.36	.32	.32	.32	.32	.31	.31	.30	.29	.26	.19

8.3.4.1 The case where $\rho > 0$

The convergence of the Cohen estimator in equation (10) on to the true covariance is conditional on $\rho = 0$. If autocorrelation exists in the market then the Cohen estimator will be biased to the extent that the components calculated in equation (16) differ from those in equation (14).

To demonstrate this assertion two realistic levels of market autocorrelation¹² were selected, specifically $\rho = 0.1$ and $\rho = 0.3$, and the expected values of the components of the Cohen covariance estimator were recomputed using the selection rules in Table 8.6 to give the

Table 8.9

Expected adaptive Cohen-based estimates with $\rho = 0.1$

Decile	Thin trading %	Decile									
		1	2	3	4	5	6	7	8	9	10
1	0.00	1.00									
2	0.07	1.00	1.00								
3	0.43	1.00	1.00	.99							
4	1.36	1.19	1.19	1.19	1.19						
5	3.05	1.19	1.19	1.19	1.19	1.18					
6	6.70	1.18	1.18	1.18	1.18	1.18	1.17				
7	13.43	1.20	1.20	1.20	1.20	1.20	1.20	1.20			
8	24.20	1.21	1.21	1.21	1.21	1.21	1.21	1.21	1.21		
9	40.30	1.21	1.21	1.21	1.21	1.21	1.21	1.21	1.21	1.21	
10	68.36	1.21	1.21	1.21	1.21	1.21	1.21	1.21	1.21	1.21	1.21

number of leads and lags used. The resulting variance-covariance matrices are shown in Table 8.9 (for $\rho = 0.1$) and Table 8.10 (for $\rho = 0.3$).

It is evident from Table 8.9 and Table 8.10 that the prior assumption that $\rho = 0$ is indeed an important one for implementing the Cohen estimator. It is only in the more well traded deciles (i.e. deciles one, two and three) where no leads or lags are used that the estimator (equation (10)) can be seen to produce reasonable estimates. For example, considering Table 8.9, the expected covariance of two securities in decile one corresponds with

¹² The levels were chosen with reference to the autocorrelations in the indices of overseas markets reported in Dimson (1979) and given earlier in this section.

Table 8.10

Expected adaptive Cohen-based estimates with $\rho = 0.3$

Decile	Thin trading %	Decile									
		1	2	3	4	5	6	7	8	9	10
1	0.00	1.00									
2	0.07	1.00	1.00								
3	0.43	1.00	1.00	.99							
4	1.36	1.46	1.46	1.46	1.45						
5	3.05	1.45	1.45	1.45	1.45	1.45					
6	6.70	1.44	1.44	1.44	1.44	1.44	1.43				
7	13.43	1.60	1.60	1.60	1.60	1.60	1.60	1.59			
8	24.20	1.66	1.66	1.66	1.66	1.66	1.66	1.65	1.65		
9	40.30	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	
10	68.36	1.68	1.68	1.68	1.68	1.68	1.68	1.68	1.68	1.68	1.68

the assumed true covariance, namely unity. However, for the more thinly traded securities (e.g. two securities in decile 10) the adaptive Cohen estimator prescribes the use of 15 weekly leads and lags and the expected covariance is biased upwards at a value of 1.21. It can be seen by comparing the corresponding expected covariance in Table 8.10 (i.e. 1.68) that the extent of the upward bias increases with the market autocorrelation.

In the ensuing development a procedure is proposed for adjusting the components of the Cohen estimator so that it will be consistent even in the presence of autocorrelation. The procedure is based on multiplying each of the components of the Cohen covariance estimator (which are on the right hand side of equation (10)) by an adjustment factor denoted in its general form by $f(\rho, q_j, q_k, m, n)$. The adjustment factors are designed to make the expected values of the components of the Cohen covariance estimator equal to the expected values of the components which are based on the assumption that no autocorrelation is present (as given in equation (14)). The adjustment factors depend only on the magnitude of true market autocorrelation (ρ), the extent of thin trading of each of the securities and on the lead or lag being considered and are *not* dependent on the true covariance.

In keeping with the specifications outlined above the proposed adjusted estimator will have the following form:

$$\begin{aligned} \hat{Cov}(R_{j,t} ; R_{k,t}) &= Cov(R_{j,t}^o ; R_{k,t}^o) \times f(\rho, q_j, q_k, 0, 0) \\ &+ \sum_{n=1}^N Cov(R_{j,t}^o ; R_{k,t-n}^o) \times f(\rho, q_j, q_k, 0, n) \\ &+ \sum_{n=1}^N Cov(R_{j,t-n}^o ; R_{k,t}^o) \times f(\rho, q_j, q_k, n, 0) \end{aligned} \quad (17)$$

The above estimator has very much the same form as the original Cohen covariance estimator (given in equation (10)) in which zero autocorrelation is assumed except that in equation (17) above, each of the components on the right hand side is multiplied by a factor depending on the trading frequencies of the securities, the autocorrelation in the market and the lead or lag being considered.

For the estimator given above in equation (17) to be unbiased and to converge onto the true covariance ($Cov(R_{j,t} ; R_{k,t})$), the expected values of the components on the right hand side of equation (17) (which have a general form $Cov(R_{j,t-n}^o ; R_{k,t-m}^o) \times f(\rho, q_j, q_k, m, n)$) must be equal to the expected covariance between the observed returns when $\rho = 0$ (i.e. the same condition under which Cohen *et al* (1983) proved the consistency of equation (10)). This requirement can be expressed symbolically as:

$$E[Cov(R_{j,t-n}^o ; R_{k,t-m}^o) | \rho = 0] = E[Cov(R_{j,t-n}^o ; R_{k,t-m}^o) | \rho \neq 0] \times f(\rho, q_j, q_k, m, n) \quad (18)$$

where $E[Cov(R_{j,t-n}^o ; R_{k,t-m}^o) | \rho = 0]$ is the expected value of the component of the covariance estimator when there is no autocorrelation and $E[Cov(R_{j,t-n}^o ; R_{k,t-m}^o) | \rho \neq 0]$ is the expected value of the component of the covariance estimator in the cases when there might be autocorrelation.

From the relationship in equation (18) it is evident that a typical adjustment factors can be derived as follows:

$$\begin{aligned}
 f(\rho, q_j, q_k, n, 0) &= \frac{E[\text{Cov}(R_{j,t-n}^o; R_{k,t}^o) | \rho = 0]}{E[\text{Cov}(R_{j,t-n}^o; R_{k,t}^o) | \rho \neq 0]} \\
 &= \frac{(1-\rho^2)(1-\rho q_j)(1-\rho q_k)(\rho - q_k)}{(\rho/q_k)^n \rho (1-q_j q_k)(1-\rho q_k) + q_k (\rho^2 - 1)(1-\rho q_j)}
 \end{aligned} \tag{19}$$

where the numerator is obtained from equation (14) and the denominator from equation (16). The fact that the adjustment factors are derived as ratios of two expressions both of which are a multiple of the true covariance term means that the true covariance term will cancel out in the ratios and thus the adjustment factors will not depend on the true covariance. Therefore, given ρ , the factors can be calculated using only q_j , q_k , m and n .

Using the adjustment factors to nullify the bias caused by autocorrelation, the Cohen estimator can still be used in the case of non-zero autocorrelation. The adaptive Cohen estimates of the covariances adjusted (using the proposal described above) for the cases $\rho = 0.1$ and $\rho = 0.3$ are identical to those given in Table 8.7. That is, the bias will be fully countered by the adjustment factors.

8.3.5 Conclusion and summary

Although not emphasized to a great extent in the literature, it is known that the usual covariance estimators are seriously biased in thinly traded environments. In this section, the impact on covariance estimation has been demonstrated analytically and was shown to be significant even for fairly well traded securities. On the basis of Shanken's (1987) suggestion a covariance correction procedure was implemented and it was shown (using an analytical approach) that this correction procedure yields substantial improvements in covariance estimation. Further, some adaptations concerning the number of leads and lags were considered for different levels of thin trading in order to improve the efficiency of the estimator.

The consideration takes account of the trade-off between using too many leads or lags (which would result in a decrease in efficiency), and too few leads or lags (which would not correct for all the thin trading bias). Results of the analysis suggest an adaptive type of covariance estimator be used for covariance estimation in thinly traded environments.

Finally, a novel proposal was developed to adjust the covariance estimator where autocorrelation (in addition to that induced by thin trading) exists in the market index. The effectiveness of the proposal was also demonstrated analytically. On the basis of the results it can be recommended that the proposals considered here be used for covariance estimation in thinly traded environments.

In the next section the concept of correcting for biases caused by market autocorrelation in aggregated coefficients estimators is extended into beta estimation.

8.4 Beta estimation using the Cohen *et al* estimator in the presence of market inefficiency

In section 8.3 an adjustment procedure for removing the bias in the Cohen covariance estimators caused by market inefficiency was derived. The correction procedure was based on the assumption that thin trading was the only cause of friction in the pricing process. Explicitly the assumption was that a security was either traded or not in any period and that the probability of a trade remained the same over time.

In this section the focus is shifted from covariance estimation to beta estimation in thinly traded environments. Although beta estimation is not the specific focus of this chapter, the developments necessary for covariance estimation proposed in the previous section can also be extended to beta estimation when the aggregated coefficients estimator is used. Even though the conclusions of Chapters 4 and 5 suggested that the trade-to-trade estimator is a more suitable beta estimator for the JSE (because of the extreme thin trading), the aggregated coefficients approach may well be more suitable on *other* markets with less severe thin trading. In this section it is demonstrated how the Cohen estimator (derived by Cohen *et al* (1983)) of the beta coefficient would also be biased under the hypothesis of first order autocorrelation in the market returns. Approximate expressions for correcting for the bias are derived and illustrated.

8.4.1 Theoretical development

Briefly, the correction procedure for the Cohen estimator of beta in the face of market autocorrelation proceeds as follows. Each of the components of the Cohen estimator is adjusted using a factor designed to make its expected values correspond to the expected value in the hypothetical case where there is no market autocorrelation. Each adjustment factor is therefore the ratio of the expected value of the component in the hypothetical case where there is no autocorrelation to the expected value of the component in the case where there is autocorrelation.

If the return on a security j in interval t is denoted by $R_{j,t}$ and the return on the market is denoted by $R_{M,t}$ then the usual Cohen estimator is given by

$$\hat{\beta}_j = \frac{b_j^o + \sum_{n=1}^N b_{j(+n)}^o + \sum_{n=1}^N b_{j(-n)}^o}{1 + \sum_{n=1}^N b_{M(+n)}^o + \sum_{n=1}^N b_{M(-n)}^o} \quad (20)$$

where

$$\begin{aligned} b_{j(+n)}^o &= \text{Cov}(R_{j,t+n}; R_{M,t}) \\ b_{M(+n)}^o &= \text{Cov}(R_{M,t+n}; R_{M,t}) \end{aligned} \quad (21)$$

To illustrate the technique used to calculate the adjustment factors for market autocorrelation, it is instructive to examine more closely one of the components for which the b^o is an estimator in equation (20). Specifically one of the lagged beta coefficients is examined:

$$\beta_{j(+n)} = \frac{\text{Cov}(R_{j,t+n}^o; R_{M,t}^o)}{\text{Var}(R_{M,t}^o)} = \frac{\text{Cov}(R_{j,t}^o; R_{M,t-n}^o)}{\text{Var}(R_{M,t}^o)} \quad (22)$$

Both the numerator and denominator of the expression will be affected by market autocorrelation. Two factors are therefore required to adjust the b^o estimates in (20). If the market capitalization weighting of the k th security in the market proxy is denoted by x_k then the numerator of the component in (22) can be expanded as follows:

$$\begin{aligned}
Cov(R_{j,t+n}^o ; R_{M,t}^o) &= Cov(R_{j,t+n}^o ; \sum_k x_k R_{k,t}^o) \\
&= \sum_k x_k Cov(R_{j,t+n}^o ; R_{k,t}^o)
\end{aligned} \tag{23}$$

It can clearly be seen from the above expansion that the lagged covariance of the security with the market is a weighted sum of the covariances between the security and all the other component securities in the market index. The factors to correct for the market autocorrelation are introduced by first noting that under stationarity assumptions

$$E[Cov(R_{j,t+n}^o ; R_{k,t}^o)] = E[Cov(R_{j,t}^o ; R_{k,t-n}^o)] \tag{24}$$

i.e. the expected lagged and leading covariance terms are equal. Secondly, using more explicit notation for the conditions regarding the autocorrelation in the market being considered, the adjustment factors for market autocorrelation are derived by considering the following equality which must hold for the component of the Cohen estimator to be unbiased:

$$E[Cov(R_{j,t+n}^o ; R_{M,t}^o) | \rho = 0] = \sum_k x_k E[Cov(R_{j,t+n}^o ; R_{k,t}^o) | \rho \neq 0] \times f(\rho, q_j, q_k, 0, n) \tag{25}$$

It should be noted from the above formulation that the factor f can be derived similarly to that in section 8.3.

Of course the above equation can be used to obtain an unbiased estimator of the component term of the Cohen estimator directly. However, it is somewhat clumsy and requires full information regarding the x_k , i.e. the weights of the securities in the market index. Some additional assumptions about the market and market proxy will enable a more tractable adjustment factor to be designed. The assumptions centre on a well traded market proxy, or the combination of a fairly well traded index with low autocorrelation.

To establish the simplifications, the full expression for the factor f is written out and thus the right hand side of equation (25) becomes

$$\sum_k x_k E[Cov(R_{j,t+n}^o ; R_{k,t}^o) | \rho \neq 0] \times \frac{(1-\rho^2)(1-\rho q_j)(1-\rho q_k)(\rho - q_j)}{(\rho/q_j)^n \rho (1-q_j q_k)(1-\rho q_j) + q_j (\rho^2 - 1)(1-\rho q_k)} \tag{26}$$

Under one of two fairly plausible conditions the expression in equation (26) can be simplified. The two conditions are: (1) x_k is nearly zero for q_k larger than zero (in other words only well traded securities constitute the index), or (2) ρq_k and q_k^2 are very small when x_k is larger than zero (which is equivalent to fairly well traded securities being in the index, but that the market autocorrelation is small so that the product of the autocorrelation coefficient and the measure of thin trading is nearly zero).

Assuming that one of the above two conditions holds, the numerator of equation (22), which is given in equation (26), can be simplified down to:

$$\begin{aligned} E[\text{Cov}(R_{j,t+n}^o ; R_{M,t}^o) | \rho = 0] &\approx (1-\rho^2) \left[\frac{(1-\rho q_j)(\rho - q_j)}{(\rho/q_j)^n \rho (1-\rho q_j) + q_j (\rho^2 - 1)} \right] E[\text{Cov}(R_{j,t+n}^o ; R_{M,t}^o) | \rho \neq 0] \\ &= (1-\rho^2) \psi(n, q_j, \rho) \times E[\text{Cov}(R_{j,t+n}^o ; R_{M,t}^o) | \rho \neq 0] \end{aligned} \quad (27)$$

From the above expression it is apparent that the expected covariance between a security and the market when there is no autocorrelation is approximately equal to the expected covariance when there is autocorrelation multiplied by a factor which depends on the lag, the trading frequency of the security and the autocorrelation coefficient. For convenience and to emphasize that the factor depends only on the measure of trading frequency of the security, the 'factor' for a lag of n is henceforth written $\psi(n, q_j, \rho)$ (for a security j which has a probability of not trading of q_j and a market which has autocorrelated returns with coefficient ρ). As would be expected for the special case in which a well traded security is considered (i.e. where $q_j = 0$ and there is no delay in information being priced into the security), the factor equals zero, for n non-zero.

Having derived a simplified expression for the numerator in the discussion above, the denominator of the component term of the Cohen estimator in equation (22) is now considered, i.e.:

$$\text{Var}(R_{M,t}^o) = \sum_k \sum_j x_k x_j \text{Cov}(R_{k,t}^o ; R_{j,t}^o) \quad (28)$$

The denominator above is adjusted to correspond to the case where there is no autocorrelation as follows:

$$\begin{aligned}
 E[\text{Var}(R_{M,t}^o) \mid \rho = 0] &= \sum_k \sum_j x_k x_j E[\text{Cov}(R_{k,t}^o ; R_{j,t}^o) \mid \rho \neq 0] \times f(\rho, q_k, q_j, 0, 0) \\
 &= \sum_k \sum_j x_k x_j E[\text{Cov}(R_{k,t}^o ; R_{j,t}^o) \mid \rho \neq 0] \times \frac{(1-\rho^2)(1-\rho q_k)(1-\rho q_j)}{(1-\rho^2 q_k q_j)} \quad (29) \\
 &\approx (1-\rho^2) E[\text{Var}(R_{M,t}^o) \mid \rho \neq 0]
 \end{aligned}$$

where the final (approximate) equality holds under the same conditions discussed when obtaining the simplified expression given in equation (26) for the numerator of the component term. The above result demonstrates that autocorrelation in the index will lead to underestimates of the market variance.

Combining the results from equations (27) and (29) for the numerator and denominator, respectively, an adjusted lagged component for the Cohen estimator can be calculated as follows:

$$E[\hat{\beta}_{j(n)} \mid \rho = 0] = \psi(n, q_j, \rho) \times E[\hat{\beta}_{j(n)} \mid \rho \neq 0] \quad (30)$$

The above expression indicates that the expected value of the estimator of the component given in equation (22) (where there is zero autocorrelation assumed) is equal to an adjustment factor, $\psi(0, q_j, \rho)$, multiplied by the expected value of the estimator when there is autocorrelation in the market returns. In the case where $n = 0$, i.e. the synchronous component β_j^o , has an adjustment factor of $\psi(0, q_j, \rho) = (1 - \rho q_j)$.

In this section, the adjustment factor required to make one of the lagged components of the Cohen estimator of beta unbiased when there is market autocorrelation was derived. Before considering the other components of the Cohen estimator and adjustment factors derived it is interesting to examine a few examples of the factors at various levels of thin trading and autocorrelation.

8.4.1.1 Examples of factor sizes for various levels of thin trading and autocorrelation

In Table 8.11 values for the factors to be applied to the lagged coefficients for various levels of thin trading are calculated for two levels of market inefficiency. In the first panel of

the table, the adjustment factors are given for an autocorrelation in the market of 0.1 and three levels of thin trading (measured as the proportion of intervals not traded) for the security, namely 0.11, 0.31, 0.51. It is readily apparent that factors adjust less for the more thinly traded shares. For example, at a lag of one, the adjustment factor is 0.99 for $q = 0.11$ compared with 0.95 when the thin trading is measured as 0.51. By comparing the adjustment factors in the panel with autocorrelation of 0.1 with the adjustment factors in the panel with autocorrelation of 0.3, it is also apparent that the factors will adjust more for higher levels of autocorrelation, as might be intuitively expected. For example, at a lag of five and trading infrequency of 0.11, the factor for $\rho = 0.1$ is 0.21 compared with a factor of zero for $\rho = 0.3$. That is, when the autocorrelation is very high, any observed covariances between the security and the market at a lag of five intervals is spurious, and will have already have been incorporated into the Cohen estimator. However, if the autocorrelation is less severe, then a small component of the covariance is real and has not yet been incorporated into the estimator. It is important to note when interpreting this table that it is the *factors* which have been presented, and not the *expected covariances* at the various lags.

Another interesting aspect that can be determined from the tabled factors is that they decrease to zero only in the case when the level of autocorrelation is greater than the extent of thin trading, i.e. when $\rho = 0.3$ and $q = 0.11$.

In the following sub-section, the adjustments required to the other components of the Cohen covariance estimator to make them unbiased are discussed.

8.4.1.2 Other terms in the Cohen estimator

In the discussion above one of the lagged component terms of the Cohen estimator was examined in some detail and simplifications considered. In this section adjustments to the other components in the Cohen estimator are described somewhat more briefly because much of the reasoning follows the same tack as for the lagged components. First, consider the numerator of the leading beta estimator

$$E[\text{Cov}(R_{j,t-n}^o ; R_{M,t}^o) \mid \rho = 0] = \sum_k x_k E[\text{Cov}(R_{j,t-n}^o ; R_{k,t}^o) \mid \rho \neq 0] \quad (31)$$

$$\times \frac{(1-\rho^2)(1-\rho q_j)(1-\rho q_k)(\rho - q_k)}{(\rho/q_k)^n \rho (1-q_j q_k)(1-\rho q_k) + q_k (\rho^2 - 1)(1-\rho q_j)}$$

Table 8.11

Evaluation of factor values for lagged coefficients of Cohen estimator, i.e. $\psi(n, q, \rho)$.

ρ	0.10			0.30		
	0.11	0.31	0.51	0.11	0.31	0.51
q_0						
$n = 0$	.99	.97	.95	.97	.91	.85
1	.52	.74	.80	.27	.48	.57
2	.36	.69	.78	.09	.33	.47
3	.29	.67	.77	.03	.26	.43
4	.24	.67	.77	.01	.21	.41
5	.21	.66	.77	.00	.18	.40
6	.19	.66	.77	.00	.15	.39
7	.17	.66	.77	.00	.14	.39
8	.16	.66	.77	.00	.12	.39
9	.15	.66	.77	.00	.11	.39
10	.14	.66	.77	.00	.11	.38
11	.13	.66	.77	.00	.10	.38
12	.13	.66	.77	.00	.09	.38
13	.12	.66	.77	.00	.09	.38
14	.12	.66	.77	.00	.08	.38
15	.12	.66	.77	.00	.08	.38
16	.11	.66	.77	.00	.07	.38
17	.11	.66	.77	.00	.07	.38
18	.11	.66	.77	.00	.07	.38
19	.11	.66	.77	.00	.07	.38
20	.10	.66	.77	.00	.06	.38

Noting that $q_k^2 \approx 0$ and $\rho q_k \approx 0$ under both assumptions (the very well traded index, or the fairly well traded index and the mild autocorrelation assumptions), the adjustment factor for the estimator of the leading beta coefficient (i.e. the last term in equation (31)) can be seen to tend to zero.

The remaining two types of component term in the Cohen estimator that have to be considered are:

$$\text{Cov}(R_{M,t+n}^o ; R_{M,t}^o) = \text{Cov}(R_{M,t}^o ; R_{M,t-n}^o), \text{ and}$$

$$\text{Cov}(R_{M,t-n}^o ; R_{M,t}^o).$$

Using very similar reasoning to the derivations given above and in the previous section, it does not take much further work to see that under either of the simplifying assumptions about the trading characteristics of the market index constituents, the adjustment factors are zero for

$n > 0$ and are equal to unity in the synchronous case. In other words, based on the assumptions the expected autocovariance in the market is zero, and the estimated variance using the observed returns is unbiased.

8.4.1.3 The adjusted Cohen estimator

Combining the results in the previous section permits the adjusted Cohen estimator to be set out under either of the two simplifying assumptions about the market trading frequencies, viz:

$$\hat{\beta}_j = (1 - \rho q_j) b_j^o + \sum_{n=1}^N \psi(n, q_j, \rho) b_{j(n)}^o \quad (32)$$

The above formulation thus specifies that the adjusted Cohen estimator is given by a weighted sum of the synchronous and lagged coefficients, with the weights equal to the adjustment factors.

In summary the results above show that the components of the Cohen estimator have to be adjusted to provide consistent estimates in the case where there is autocorrelation in the true, unobservable market returns. Correction factors have been derived under two reasonably plausible simplifying assumptions. The major adjustment factors are applied to the synchronous and lagged coefficients in the numerator of equation (20). These depend only on the level of thin trading in the security and the size of the autocorrelation coefficient of the market.

In this section an extension of the methodology developed in 8.3 was considered for adjusting the Cohen beta estimator. In the next section the concept of adjusting for market autocorrelation when aggregated coefficients estimators are used is extended to a continuous-time framework. It should be noted that although the concept of correcting for the bias in the Cohen estimator caused by market autocorrelation is the same as in section 8.3, the actual mathematics and assumptions vary considerably and are indeed far more technical.

8.5 Covariance estimation in the presence of thin trading and market autocorrelation: a continuous-time framework

8.5.1 Introduction

This section comprises a novel extension of the earlier work on covariance estimation in thinly traded environments described in section 8.3 of this chapter. In this section attention is concentrated on covariance estimation in markets where thin trading is limited to delays which are less than one day in length. In other words the focus is on covariance estimation in markets where securities are traded at least once per day,¹³ but where the actual times of the trades are not recorded. Because the JSE is a very thinly traded stock market (where price adjustment delays of a few hours are negligible when compared with the more common delays of several days or weeks), the results in this section of the chapter are not directly applicable in South Africa. The reason for including the analysis in the thesis is to show how the research which emerged for a thinly traded market can be extended to the well traded exchanges, particularly because covariance estimation has been relatively ignored in the literature compared with systematic risk estimation on well traded exchanges.

In addition, the findings indicate that it is imperative to implement the suggested adjustment procedures in estimating covariances using daily data on well traded stock markets such as the NYSE. The importance attached to trading frictions and price adjustment delays of at most a few hours, stems from the common use of daily data with which to estimate financial risk statistics on the NYSE.¹⁴ The delays of a few hours are clearly of relatively more significance daily returns rather than, say, monthly returns.

The contents of this section are fairly technical because of the difficulties that exist in dealing with continuous-time time series and because the use of continuous-time time series analysis in estimating risk statistics in financial economics is not very common.¹⁵ However, the central concepts and general approach are very similar to section 8.3: the way in which

¹³ The results can very easily be generalized to any case where observations are available at longer intervals than the intervals at which trading activity occurs, i.e. cases where the model timing intervals are shorter than the observation intervals. (See Harvey (1989, pp309 *et seq*) for a fuller discussion of aggregation of these cases.)

¹⁴ See Scholes and Williams (1977).

¹⁵ The obvious exception is the work done on continuous-time versions of the CAPM, see, e.g., Merton (1973).

autocorrelation in the market causes biases in covariance estimates is detailed and then a way for correcting for the bias is suggested.

The rest of the section is organized as follows: In sub-section 8.5.2 the model on which the analysis is based is presented. In sub-section 8.5.3 the model is used to derive the expected covariance between two hypothetical securities. Some sample numerical values are presented in sub-section 8.5.4 along with a suggested procedure for adjusting empirical estimates of covariances to account for the bias caused by market autocorrelation. A summary and some conclusions are offered in sub-section 8.5.5.

8.5.2 The model for security returns

As pointed out in section 8.3, Shanken (1987) noted that the derivation of the beta estimator of Cohen *et al* (1983) contains an explicit expression for the covariance between the returns of two securities. Under the assumption that securities are traded at least once per day, the estimator for the covariance between two securities j and k based on Cohen *et al* (1983) can be written as:

$$\hat{C}(j,k) = \text{Cov}(R_{j,t}^o ; R_{k,t}^o) + \text{Cov}(R_{j,t}^o ; R_{k,t-1}^o) + \text{Cov}(R_{j,t}^o ; R_{k,t+1}^o) \quad (33)$$

where $R_{j,t}^o$ is the observed return on security j in day t . Cohen *et al* (1983) provide an argument which demonstrates that the estimator given in equation (33) will be consistent and unbiased when there is no market autocorrelation. In order to establish the effect of market autocorrelation on the estimator given in (33), it is necessary to re-examine the Market Model in a continuous-time time series framework.

The set of *observed* daily returns based on daily closing prices (the $R_{j,t}^o$) also constitutes a set of *true*, underlying returns. The additional point to notice is that these true returns are measured over intervals of unknown length¹⁶ (because the times of the trades are not recorded). With a slight change in notation, the series of *observed* returns can also be denoted by $R_{j,\tau}$ ($\tau = 1, 2, 3, \dots, n$). The return on the market in day τ (which is measured over the same period of time as the observed return on security j) is defined to be $M_{j,\tau}$. The reason for the change of notation for the observed set of returns is that in the ensuing development the

¹⁶ Because it has been assumed that the security is traded every day, the length of each interval will take on a value between zero and two days.

emphasis falls on the fact that the observed returns are also a set of true returns measured over unknown intervals ending in day τ .

A further piece of notation needed to facilitate later discussion is the representation of the (unknown) interval length over which $R_{j,\tau}$ and $M_{j,\tau}$ are measured by $\delta_\tau^j = t_\tau^j - t_{\tau-1}^j$, where t_τ^j is the time of the last trade of security j in day τ . The development that follows considers how autocorrelation in the market can be incorporated into the model. The results will provide insights as to how the covariances between the observed returns will be affected by the autocorrelation.

Using the above notation and using the idea that the observed returns correspond to true returns measured over unknown intervals, the Market Model can then be written as:

$$R_{j,\tau} = \alpha_j \delta_\tau^j + \beta_j M_{j,\tau} + u_{j,\tau} \quad (34)$$

Using the new notation, the Cohen *et al* (1983) estimator given in equation (33) can be rewritten as follows:

$$\hat{C}(j,k) = \text{Cov}(R_{j,\tau} ; R_{k,\tau-1} + R_{k,\tau} + R_{k,\tau+1}) \quad (35)$$

which is consistent, provided, among other things, that the market returns are not autocorrelated. If it is assumed that the beta coefficients of securities j and k are equal to unity (without loss of generalization), then the estimator in (35) is easily shown to be equal to

$$\hat{C}(j,k) = \text{Cov}(M_{j,\tau} ; \sum_{s=-1}^1 M_{k,\tau+s}) \quad (36)$$

The above expression demonstrates that the covariance between the two securities will be equal to the covariance of one market return measured between the last trade in day $\tau-1$ of security j and the last trade in day τ of security j and a second market return measured between the last trade in day $\tau-2$ of security k and the last trade of security k in day $\tau+1$.

In order to facilitate the inclusion of autocorrelation into the estimator of covariance given in (36) a 'market process' is defined. The market process can be thought of as the return

on the market over an infinitely small interval. The process is denoted by $\tilde{M}(t)$. The ‘hypothetically observed’¹⁷ market returns, $M_{j,\tau}$, can then be written as

$$M_{j,\tau} = \int_{t_{\tau-1}}^{t_{\tau}} \tilde{M}(t) dt \quad (37)$$

Atchison, Butler and Simonds (1987) have shown in an empirical study on the NYSE that even when the effects of thin trading are removed there remains autocorrelation in the market returns. Although this ‘residual’ autocorrelation may admittedly arise in part from other sources of friction (which could be corrected for in a more complete model of trading frictions) the evidence does seem to support the existence of autocorrelation in the market returns.¹⁸ The autoregressive nature of the market returns is embedded in the model by means of a first-order continuous-time autoregressive (CAR(1)) process for the market. The rationale for using a CAR(1) process is that the return on the market over a given interval comprises the sum of all the instantaneous returns in the interval. By allowing the instantaneous returns (the market process) to be autoregressive in nature, autocorrelation is automatically embedded in the observed daily market returns.

A CAR(1) process can be written in the form of a stochastic differential equation (see Harvey (1989, p481) or Jones (1986)):

$$d\tilde{M}(t) = \gamma + \alpha \tilde{M}(t) + dW(t) \quad (38)$$

where $W(t)$ is a Wiener process (the integral of a white-noise process) with parameter σ^2 . Note that a technical constraint is that α must be less than zero for the process to be stationary. The case where α is zero corresponds to a continuous-time version of the random walk with drift (of γ). The case where α tends to negative infinity corresponds to an uncorrelated series of observed market returns.

¹⁷ The returns on the market are not actually observed because the times of the trades on the securities are unknown.

¹⁸ Of course, if there exists autocorrelation in the market returns and the Market Model holds, then the returns of the securities will also be autocorrelated to some extent.

The solution to equation (38) (again, see Harvey (1989) or Jones (1986)) is given by

$$\tilde{M}(t) = -\gamma / \alpha + \int_{-\infty}^t e^{\alpha(t-r)} dW(r) \quad (39)$$

The key to establishing the effect of autocorrelation in the market process on the covariance estimates, rests on substituting the market process defined in equations (39) and (37). Once the substitution has been performed, several important results emerge which can then be used to find the bias in covariance estimation. In order to facilitate understanding the results which emerge, it is helpful to consider the usual, zero mean, discrete time autoregressive process, i.e.

$$M_t = \rho M_{t-1} + \epsilon_{M,t} \quad (40)$$

The elements of the above equation are the market return in interval t , M_t , the autocorrelation coefficient, ρ , and the random input, $\epsilon_{M,t}$, which is assumed to have a variance of σ_m^2 and zero mean. Because the elements in the discrete time case (as set out in (40)) are more intuitively associated, each of the features of the continuous-time case which are discussed below will (where necessary and possible) be associated with the corresponding feature of the discrete time case.

The results which are important in establishing the bias in covariance estimates are firstly, the mean, variance and autocovariance of the market returns. The mean and variance of $M_{j,\tau}$ are easily derived by substituting the expression for the market process given in equation (39) into the definition of the market returns given in (37) (see Harvey, 1989, p482):

$$E[M_{j,\tau}] = -(\gamma/\alpha) \delta_\tau^j \quad (41)$$

$$Var[M_{j,\tau}] = -(\sigma^2/2\alpha) \delta_\tau^j \quad (42)$$

Note that if the ‘drift’ parameter (λ) is assumed to be zero, then the mean market return (in equation (41)) over the interval from $t_{\tau-1}$ to t_τ is zero. The autocovariance of the

market *returns* is also relatively straightforward to derive if the general result for the autocovariance of the market *process* given in Jones (1986) is invoked, viz:

$$\text{Cov}(\tilde{M}(t); \tilde{M}(s)) = e^{\alpha|t-s|}(-\sigma^2/2\alpha) \quad (43)$$

Using the above expression for the autocovariance of the market *process*, the autocovariance of the market *returns* can be obtained as

$$\begin{aligned} \text{Cov}(M_{j,\tau} ; M_{j,\tau-1}) &= \int_0^{\delta_t} \int_0^{\delta_{t-1}} \text{Cov}(\tilde{M}(t_{\tau-1} + r) ; \tilde{M}(t_{\tau-1} - s)) dr ds \\ &= -\frac{\sigma^2}{2\alpha^3} (e^{\alpha\delta_t} - 1) (e^{\alpha\delta_{t-1}} - 1) \end{aligned} \quad (44)$$

If the above expressions are examined more closely some important implications can be ascertained. Firstly, as α tends to zero (i.e. as the process converges on a random walk) in equations (41) and (42), so the mean and variance become ill-defined (which is as expected for a random walk). Secondly, when α is negative (i.e. when the market process is a CAR(1) process), there will be positive autocorrelation in the market returns because the autocovariance term in equation (44) will be positive.

If attention is restricted to fixed unit intervals (i.e. the case where the security is very well traded), then an expression for the first order autocorrelation coefficient (which is given by ρ in the discrete time case) is obtained by dividing the autocovariance function in (44) by the variance function in equation (42) resulting in

$$\rho = \frac{(e^{\alpha} - 1)^2}{\alpha^2} \quad (45)$$

Another result which is important in the discussion below is the variance of the random input of the market process integrated over a unit interval, or the variance of $\epsilon_{M,t}$ in the discrete time case. It can quite easily be shown (see, e.g. Harvey, 1989) that the variance can be evaluated using

$$\sigma_m^2 = \sigma^2 \frac{(e^{2\alpha} - 1)}{2\alpha} \quad (46)$$

In this section a few of the most important characteristics of the model for the market returns have been presented. In the next section the expected value of the covariance of the two securities is derived using the characteristics of the market process set out above.

8.5.3 An expression for the covariance

In this section the model specified in the previous section for the market process is used to find the expected value of the covariance based on the Cohen *et al* procedure. The expression for the expected value is shown to depend on the parameters α and σ^2 , i.e. the autocovariance and innovation variance of the market, and on measures of the trading frequency of each of the two securities.¹⁹

Combining the expressions for the market returns and the covariance between the securities given in equations (37) and (36), respectively, the covariance estimator can be rewritten using the market process as:

$$\hat{C}(j,k) = \text{Cov}[M_{j,\tau} ; \int_{t_{\tau-2}}^{t_{\tau-1}} \tilde{M}(s) ds] \quad (47)$$

¹⁹ As mentioned earlier, the assumption has been made that the beta coefficients are equal to unity in order to clarify the exposition. In the actual estimation process the expression for the covariance will also depend on the values of the two security beta coefficients.

The above expression gives the covariance estimator in terms of the market process. The derivation proceeds by noting that the right hand side of equation (47) can be divided up into three parts:

$$\begin{aligned} \text{Cov}[M_{j,r} ; \int_{t_{r-2}}^{t_{r+1}} \tilde{M}(s) ds] &= \text{Cov}[M_{j,r} ; \int_{t_{r-2}}^{t_{r-1}} \tilde{M}(s) ds] + \text{Cov}[M_{j,r} ; \int_{t_{r-1}}^{t_r} \tilde{M}(s) ds] \\ &\quad + \text{Cov}[M_{j,r} ; \int_{t_r}^{t_{r+1}} \tilde{M}(s) ds] \end{aligned} \quad (48)$$

Based on the above decomposition and recalling the definition in equation (37), the following clearly holds:

$$\begin{aligned} \text{Cov}[M_{j,r} ; \int_{t_{r-2}}^{t_{r+1}} \tilde{M}(s) ds] &= \text{Var}[M_{j,r}] \\ &\quad + \text{Cov}[M_{j,r} ; \int_{t_{r-2}}^{t_{r-1}} \tilde{M}(s) ds] \\ &\quad + \text{Cov}[M_{j,r} ; \int_{t_r}^{t_{r+1}} \tilde{M}(s) ds] \end{aligned} \quad (49)$$

For convenience in the discussion that follows, the interval lengths which appear in the expressions above are given more succinct notation, i.e.:

$$\begin{aligned} \delta_1^* &= t_r^j - t_{r-1}^j \\ \delta_0^* &= t_{r-1}^j - t_{r-2}^k \\ \delta_2^* &= t_{r+1}^k - t_r^j \end{aligned} \quad (50)$$

Using the expressions for the market variance and covariance terms in equations (42) and (44), an expression for the expected value of the estimated covariance can now be presented in terms of the α coefficient (which defines the extent of the autocorrelation in the market process) and the interval lengths defined in equation (50):

$$\begin{aligned}\hat{C}(j,k) = & -(\sigma^2/2\alpha) \delta_i^* \\ & - (\sigma^2/2\alpha^3) (e^{\delta_i^* \alpha} - 1) (e^{\delta_i^* \alpha} - 1) \\ & - (\sigma^2/2\alpha^3) (e^{\delta_i^* \alpha} - 1) (e^{\delta_i^* \alpha} - 1)\end{aligned}\quad (51)$$

The expected covariances given in equation (51) could be easily evaluated were the δ_i^* known and constant over time. For well traded securities, for example, the δ_i^* would all be equal to unity. In the case of thinly traded securities, on the other hand, the intervals will not only be of unknown length but will vary over time.

Along with models proposed by several other researchers (e.g. Atchison *et al* (1987) and Scholes and Williams (1977)) the model in this chapter incorporates the additional variance caused by unknown trading times by assuming the trading process to be a curtailed Poisson process.²⁰ Using standard statistical results, the market variance can be split up into two components as follows

$$Var[M_{j,\tau}] = E[Var[M_{j,\tau} | \delta_i^*]] + Var[E[M_{j,\tau} | \delta_i^*]] \quad (52)$$

If, for simplicity, it is assumed that γ is small, i.e. that the drift parameter in the market process is small, then the last term on the right hand side can be ignored (see equation (41)). The covariance terms in equation (51) can be broken up into two parts in a similar way to the variance term in equation (52). The discussion now proceeds by finding the expectations of the three terms on the right hand side of equation (51) with respect to the trading frequencies of the two securities.

²⁰ By a *curtailed* Poisson process it is meant a process that forces there to be at least one trade in the interval. See Atchison *et al* (1987) for a discussion on the reasonableness of the Poisson process as model for trading.

If the parameters for the Poisson processes are represented by λ_j and λ_k for the two securities respectively,²¹ then tedious, but relatively straightforward algebra results in the following expressions for the expected values of the covariance terms with respect to the trading frequencies (λ_j and λ_k):

$$E[(e^{\delta_j^* \alpha} - 1)(e^{\delta_k^*} - 1)] = e^\alpha \frac{\lambda_j}{\lambda_j + \alpha} \frac{1 - e^{-(\lambda_j + \alpha)}}{1 - e^{-\lambda_j}} \times \left[(e^\alpha - 1) \frac{\lambda_k}{\lambda_k - \alpha} \frac{1 - e^{-(\lambda_k - \alpha)}}{1 - e^{-\lambda_k}} - \frac{\lambda_j}{\lambda_j - \alpha} \frac{1 - e^{-(\lambda_j - \alpha)}}{1 - e^{-\lambda_j}} \right] + 1 \quad (53)$$

and

$$E[(e^{\delta_k^* \alpha} - 1)(e^{\delta_j^*} - 1)] = e^\alpha \frac{\lambda_j}{\lambda_j - \alpha} \frac{1 - e^{-(\lambda_j - \alpha)}}{1 - e^{-\lambda_j}} \times \left[(e^\alpha - 1) \frac{\lambda_k}{\lambda_k + \alpha} \frac{1 - e^{-(\lambda_k + \alpha)}}{1 - e^{-\lambda_k}} - \frac{\lambda_j}{\lambda_j + \alpha} \frac{1 - e^{-(\lambda_j + \alpha)}}{1 - e^{-\lambda_j}} \right] + 1 \quad (54)$$

Substituting the terms for the expected values of the covariances in equations (53) and (54) into the appropriate places in equation (51) (and further noting that $E[\delta_i^*] = 1$) yields the expression for the covariance depending on only the thin trading parameters and the autocovariance and variance.

To recap on what has been shown in this section and the preceding one: (a) the covariance of the *security* returns was expressed in terms of covariances between *market* returns; (b) the market *returns* were expressed in terms of a market *process* which incorporated an autoregressive parameter (α); (c) the covariance in the market returns could thus be expressed in terms of the interval lengths between the times of the trades of the securities and α ; (d) the expected value of the covariance was obtained in terms of α and the trading frequencies of the two securities by assuming a distribution based on the Poisson process for the interval lengths.

In the next section the theoretical development outlined above is moulded into a practical procedure for correcting for the biases caused by market autocorrelation.

²¹ The parameters can be loosely thought of as being the average number of trades in the securities per day.

8.5.4 A procedure for correcting covariance estimates in the presence of market autocorrelation

In this section a procedure is outlined for implementing the results derived in the previous sections to correct for the bias in covariance estimation caused by market autocorrelation. It is assumed that a value for ρ , the autocorrelation in the daily market returns is available. The procedure runs as follows:

- [1] Solve equation (45) to obtain a value for α given a value of ρ , the market autocorrelation for daily returns;
- [2] Using the value of α obtained from step [1], obtain a value for σ^2 (the innovation's variance), by setting σ_m^2 in equation (46) equal to unity (for reasons outlined in [3] below);
- [3] Using estimates of λ_j and λ_k calculated from empirical data,²² the expected value of the Cohen *et al* (1983) estimator of the covariance can be calculated using the expressions derived in Section 8.5.3. Because the expected covariance thus calculated is based on $\sigma_m^2 = 1$, it represents the ratio of the expected value of the covariance to the expected covariance that would have been obtained were $\rho = 0$, i.e. if the market process were uncorrelated. An unbiased estimator of the covariance is therefore the Cohen *et al* (1983) estimator divided by the ratio.

A numerical example is given in Table 8.12. The results indicate the expected covariances between securities, j and k where security j is assumed to have a fixed level of thin trading ($\lambda_j = 40.66$) and the second security (k) is assumed to take on three different levels of thin trading. The levels of trading frequency (the λ 's) are taken from the Atchison, Butler and Simonds (1987) paper and correspond to the minimum, average and maximum parameter values for shares on the NYSE. The first column in the table gives the level of daily autocorrelation which it is desired to model. The second column gives the value of α implied by the level of autocorrelation and the third column indicates the value of σ^2 that is required to have a unit variance for the market. The fourth, fifth and sixth columns give the expected covariances between the two securities at the given level of thin trading and market autocorrelation. For example, if it can be ascertained from the table that if the daily market autocorrelation is 0.1, then the α parameter will be -3.006 and the σ^2 parameter will be 6.027. Then, noting that the

²² The λ parameters can be loosely thought of as the average number of trades per day in the securities. A method-of-moments estimator is the solution to $\bar{f} = \lambda / (1 - \exp(-\lambda))$, where $\bar{f}$ is the average number of trades per day obtained from a sample of days.

Table 8.12

Expected covariances between securities j and k at different levels of market autocorrelation and trading frequency.

λ_j 40.66	ρ	α	σ^2	λ_k		
				1.86	40.66	286.08
	0.990	-0.010	1.010	149.433	149.521	149.521
	0.500	-0.738	1.914	2.532	2.591	2.591
	0.300	-1.355	2.903	1.674	1.714	1.714
	0.100	-3.006	6.027	1.191	1.203	1.203
	0.050	-4.419	8.839	1.096	1.100	1.100

true covariance has been set to unity, it can be seen that the covariances will be biased upwards by 19.1%, 20.3% or 20.3% depending on whether the trading frequency of security k is low, average or high.

In general, the expected covariances reflected in Table 8.12 differ little across levels of thin trading, but change markedly with levels of market autocorrelation. Even quite moderate autocorrelation (at $\rho = 0.05$) produces an increase of about 10% in the estimated covariance. The autocorrelation in large portfolios (which includes market proxies) not explained by thin trading on the NYSE has values in the range from about 0.10 to 0.30 (see Atchison *et al* (1987)) and the estimated covariances will thus have a severe upward bias. Using a similar argument to that applied in 8.3, it can also be seen that there will be some upward bias in *beta* estimates using the Cohen estimator, although that investigation is not performed here.

Although there is apparent stability in the bias across levels of thin trading, it should be noted that there is slightly less bias in the covariance with the most thinly traded security ($\lambda_k = 1.86$). The reduction is in line with expectations: thin trading causes a downward bias on the covariance estimates which mitigates (to a very small extent) the effect of the upward bias caused by the autocorrelation. The effect would be more readily seen if both securities were very thinly traded.

8.5.5 Conclusion and summary

In this section it has been demonstrated how market inefficiency (evidenced by autocorrelation in the hypothetical market process) will cause severe bias in the covariance estimates of series of returns of individual securities. A model was developed which enabled

the quantification of the biases. An immediate application of the model is that of correcting for covariance estimation bias when using aggregated coefficients techniques. Although the concept of adjusting for the bias caused by market autocorrelation in estimates of covariance is very much the same as that in section 8.3 the mathematical approach required to cope with the 'thin trading' on markets such as the NYSE are very different.

8.6 Chapter summary and conclusions

The main focus of this chapter was covariance estimation on thinly traded stock markets. In particular, it was argued that trade-to-trade estimators were generally impracticable in thinly traded markets because of the scarcity of days on which both (infrequently traded) securities were traded. Hence it was contended that aggregated coefficients estimators are generally more appropriate. Use of an aggregated coefficients estimator necessitated consideration of two further problems: namely, the selection of the number of leads and lags and, secondly, the presence of autocorrelation in the security returns.

In section 8.3 an adaptive procedure was developed for selecting the appropriate number of leading and lagged coefficients in an aggregated coefficients covariance estimator. The adaptive procedure was based on the trading frequencies of the securities, in particular the levels of thin trading which are typical of securities on the JSE. Fundamental to the unbiasedness and consistency of the aggregated coefficients estimator is the assumption of zero market autocorrelation. However, evidence presented in section 8.2 (as well as elsewhere in the literature) have indicated that such an assumption is invalid. Therefore, the additional complication of market autocorrelation was considered.

It was proved analytically that autocorrelation in the market resulted in substantial upward biases in aggregated coefficients estimates of covariances. In order to combat the biasing effect of the autocorrelation, a correction procedure was consequently proposed for adjusting the components of the aggregated coefficients estimator to make them consistent. One of the key elements required in the adjustment factors turned out to be an estimate of the market autocorrelation. However, estimates of market autocorrelation based on a market *index* introduces the additional problem of spurious autocorrelation which results from thin trading in the constituent securities of the index.

It was therefore argued that estimating market autocorrelation using a market index involved separating out the proportion of the autocorrelation caused by thin trading. It was

shown how the separation of the two parts of the observed autocorrelation could be achieved using a modification of the methodology devised by Atchison *et al* (1987). In fact, in section 8.2, an empirical study revealed that, on the JSE, only a small fraction of observed portfolio autocorrelation can be attributed to thin trading.

In section 8.4 it was shown how the adjustment procedure for correcting covariance estimates for market autocorrelation could also be extended to counteract the bias which is consequently caused in beta estimation. Although beta estimation using the Cohen estimator has been assessed to be relatively inefficient in earlier chapters of this thesis, it was suggested that the results may be useful on more frequently traded exchanges where the Cohen estimator is more appropriate.

Finally, in section 8.5, the concept of correcting for the bias caused by market autocorrelation in aggregated coefficients estimators was extended to incorporate the levels of thin trading and other price adjustment delays which are prevalent on the very well traded exchanges (such as the NYSE). Using a continuous-time time series framework, a reasonably practical algorithm was developed for predicting the bias in covariance estimates at a given level of daily market autocorrelation. Once the *expected* bias is known, it was shown how the actual estimators can easily be adjusted to make them unbiased. The results of the development showed that the biases caused by market autocorrelation are indeed severe, and the implementation of either the correction procedure given in 8.3 or that in 8.5 (depending on the levels of thin trading in the market) is highly recommended.

Appendix 8.A

In this appendix the derivation of

$$\alpha_{q_i, q_j} = E[\max(I_i - I_j, 0)] \quad (55)$$

is given.

More convenient notation for the price ages of two hypothetical securities, i and j , are defined as $X_i = I_i \times k$ and $X_j = I_j \times k$. The X_i and X_j have geometric distributions curtailed at k days.

The transformations $Y_1 = X_i + X_j$ and $Y_2 = X_i - X_j$ imply that $X_i = (Y_1 + Y_2)/2$ and $X_j = (Y_1 - Y_2)/2$ provided both values are integers. On the assumption that the two trading processes are independent the joint density of Y_1 and Y_2 is given by the product of the marginal distributions:

$$f_{Y_1, Y_2} = \Psi p_i p_j q_i^{\frac{n+Y_2}{2}} q_j^{\frac{n-Y_2}{2}} \quad (56)$$

where $\Psi = 1/(1 - q_i^k)(1 - q_j^k)$ and provided the following conditions hold:

1. $|y_2| \leq y_1 \leq 2(k-1) - |y_2|$;
2. $-(k-1) \leq y_2 \leq (k-1)$;
3. $y_1 + y_2 = 2m_1$ for some integer m_1 ;
4. $y_1 - y_2 = 2m_2$ for some integer m_2 ,

or is otherwise zero.

To obtain the marginal distribution of Y_2 (the random variable of interest) a summation over the domain of Y_1 is required:

$$\begin{aligned} f_{Y_2} &= \Psi p_i p_j \sum_{y_1 = |y_2|}^{2(k-1) - |y_2|} f_{y_1, y_2} \\ &= \Psi p_i p_j \sum (q_i q_j)^{\frac{n}{2}} \left(\frac{q_i}{q_j} \right)^{\frac{Y_2}{2}}, \text{ provided conditions 3 and 4 hold} \\ &= \Psi p_i p_j \sum_{m=0}^{(k-1) - |y_2|} (q_i q_j)^{\frac{n+Y_2}{2}} \left(\frac{q_i}{q_j} \right)^{\frac{Y_2}{2}} \\ &= \Psi p_i p_j (q_i q_j)^{\frac{n+Y_2}{2}} \left(\frac{q_i}{q_j} \right)^{\frac{Y_2}{2}} \frac{1 - (q_i q_j)^{k-|y_2|}}{1 - (q_i q_j)} \end{aligned}$$

$$= \frac{\Psi p_i p_j}{1 - (q_i q_j)} \begin{cases} q_j^{-y_2} - (q_i q_j)^k q_i^{y_2}; & y_2 < 0 \\ q_i^{y_2} - (q_i q_j)^k q_j^{-y_2}; & y_2 \geq 0 \end{cases}$$

The required expectation is then easily obtained by summing over the suitable domain:

$$\begin{aligned} k \alpha_{q_i q_j} &= \sum_{y_2=0}^{(k-1)} y_2 f_{Y_2}(y_2) \\ &= \frac{\Psi p_i p_j}{1 - (q_i q_j)} \sum_{y_2=0}^{(k-1)} (q_i^{y_2} - (q_i q_j)^k q_j^{-y_2}) \\ &= \frac{\Psi p_i p_j}{1 - (q_i q_j)} \left[\frac{1 - q_i^k}{1 - q_i} - q_i^k q_j^k \frac{1 - q_j^k}{1 - q_j} \right] \end{aligned}$$

Appendix 8.B

Derivation of components of the covariance estimator

$$\text{Let } C_{n0} = \text{Cov}(R_{j,t-n}^o ; R_{k,t}^o) \\ C = \text{Cov}(R_{j,t} ; R_{k,t})$$

Then it follows that

$$C_{n0} = \text{Cov} \left[\sum_{l=0}^{N^*} q_j^l p_j R_{j,t-n-l} ; \sum_{m=0}^{N^*} q_k^m p_k R_{k,t-m} \right] \\ = C \sum_{l=0}^{N^*} \sum_{m=0}^{N^*} q_j^l p_j q_k^m p_k \rho^{|l+n-m|}$$

If $\rho = 0$ then $\text{Cov}(R_{j,t-n-l} ; R_{k,t-m}) = 0$ unless $m = n + l$, i.e.

$$C_{n0} = C \sum_{l=0}^{N^*-n} q_j^l p_j q_k^{n+l} p_k \\ \rightarrow C p_j p_k q_k^n \frac{1}{1 - q_j q_k} \quad \text{as } N^* \rightarrow \infty$$

If $\rho \neq 0$ then

$$C_{n0} = C p_j p_k \sum_{l=0}^{N^*} q_j^l \left\{ \sum_{m=0}^{n+l} \rho^{n+l-m} (q_k/\rho)^m + \sum_{m=n+l+1}^{N^*} \rho^{m-(n+l)} q_k^m \right\} \\ = C p_j p_k \sum_{l=0}^{N^*} q_j^l \left\{ \frac{\rho^{n+l} - (q_k/\rho)^{n+l+1}}{1 - (q_k/\rho)} + q_k^{n+l+1} \rho \frac{1 - (\rho q_k)^{N^*-(n+l)}}{1 - \rho q_k} \right\} \\ = C p_j p_k \sum_{l=0}^{N^*} \left\{ \frac{\rho^n}{1 - (q_k/\rho)} (\rho q_j)^l + \frac{(\rho^2 - 1)}{(1 - \rho q_k)(\rho - q_k)} (q_j q_k)^l - \frac{\rho^{N^*+1-n} q_k^{N^*+1}}{1 - \rho q_k} (q_j/\rho)^l \right\} \\ = C p_j p_k \left\{ \frac{\rho^n}{1 - q_k/\rho} \frac{1 - (\rho q_j)^{N^*+1}}{1 - \rho q_j} + \frac{q_k^{n+1}(\rho^2 - 1)}{(1 - \rho q_k)(\rho - q_k)} \frac{1 - (q_j q_k)^{N^*+1}}{1 - q_j q_k} - \frac{\rho^{-n} q_k^{N^*+1}}{1 - \rho q_k} \frac{\rho^{N^*+1} - q_j^{N^*+1}}{1 - q_j/\rho} \right\} \\ \rightarrow C p_j p_k \frac{\rho^{n+1}(1 - q_j q_k)(1 - \rho q_k) + q_k^{n+1}(\rho^2 - 1)(1 - \rho q_j)}{(\rho - q_k)(1 - \rho q_j)(1 - \rho q_k)(1 - q_j q_k)} \quad \text{as } N^* \rightarrow \infty$$

The above derivation assumes that $\rho \neq q_j$ and $\rho \neq q_k$. The algebra is easily changed to accommodate these cases. One should also note that in the case of a first order autocorrelation process the variance of the market and hence C is inflated by a factor of $1/(1 - \rho^2)$.

Appendix 8.C**Correction factors**

The correction factor is obtained as the ratio of the expression for the case where ρ is zero to the expression for the case where ρ may be non-zero.

The ratio simplifies down to:

$$\frac{(1-\rho^2)(1-\rho q_j)(1-\rho q_k)(\rho-q_k)}{(\rho/q_k)^n \rho(1-q_j q_k)(1-\rho q_k) + q_k(\rho^2-1)(1-\rho q_j)}$$

when q_k is non-zero and is identically zero when q_k is zero.

When n is zero, i.e. for the synchronous component of the estimator, the factor reduces to $(1-\rho^2)(1-\rho q_j)(1-\rho q_k) / (1-\rho^2 q_j q_k)$ which is symmetric in q_j and q_k as would be expected.

CHAPTER 9

A MULTIVARIATE STATE-SPACE APPROACH TO ESTIMATING TIME-VARYING BETAS

9.1 Introduction

In this chapter an extension to the estimation of time-varying betas is proposed using state-space modelling and the Kalman filter. The proposal is based on the hypothesis that the movements in beta coefficients of the securities are correlated with each other and therefore that modelling the co-movements has important implications for the management of portfolios with selected levels of systematic risk.

Most of the effort in dynamic beta estimation in the past has centred on a univariate approach, with betas being allowed to vary independently of each other. Blume (1971, 1975) pioneered research into beta stability using fairly *ad hoc* methods. Later Bos and Newbold (1984), Simonds *et al* (1986), Chang and Weiss (1991), Hays and Upton (1986) and many others conducted research into the stability of betas on the NYSE and all found the betas to be unstable. The research into beta stability has extended into other markets where similar findings have been reported (see, for example, Faff *et al* (1992) and Brooks *et al* (1992) for evidence from Australia; de Jong, Kemna and Kloek (1992) for evidence from the Dutch market; Bradfield *et al* (1982) and Chapters 5 and 6 in this thesis for evidence from the JSE and Dimson and Marsh (1983) for evidence from the London Stock Exchange).

The appropriate structural model for betas that vary over time remains a source of contention. The random coefficients model, random walk model and a variant of the AR(1) model have been the most popular alternatives to the constant beta hypothesis. Evidence has been presented by Bos and Newbold (1984) and Brooks *et al* (1992) that the random coefficients model offers superior fits to the data. These researchers suggested that the evidence in support of the random coefficients model is not completely conclusive because of the low power of the tests applied to discriminate between the models.

This chapter differs somewhat from the previous chapters in the thesis in that the conclusions presented are not aimed as directly at practitioners, but rather at researchers in the field of systematic risk estimation. As discussed above, the statistical testing procedures for finding the most suitable model for time-varying betas are not sufficiently powerful to be

rigorous. In addition, the models for the time-varying beta coefficients have been based on considerations which are predominantly statistical rather than based in financial theory. Although it is possible to speculate about firms' risk management strategies and thereby arrive at a statistical model for time-varying betas, it is apparent that assuming a strategy which is generally applicable to *all* shares is impracticable.

The focus in this chapter is therefore on providing further insights into the movements of beta coefficients rather than on testing any one particular model. It is argued that knowledge of the covariance of systematic risk of sectors on the stock market will contribute towards understanding the *causes* of the moves. The style adopted in discussing the issue of covarying betas is expository although some interesting conclusions are reached from an empirical study on the JSE. It is hoped that the methodology and the results may provide a springboard for further research in this area.

In section 9.2 the multivariate state-space approach to modelling beta coefficients is described and presented mathematically. Although the Kalman filter has been used before in the state-space modelling of systematic risk (see, for example, De Jong *et al* (1992) or Malkalmaki (1992)), the exposition in this chapter concentrates on the *multivariate* features which can be incorporated in a state-space approach in a similar vein to Hall, Miles and Taylor (1989). In section 9.3 a description of the Kalman filter as a method of estimating the parameters of the state-space model of systematic risk is given. The estimates which are of particular interest in this chapter are the components of the covariance matrix of the beta coefficients. It is asserted that information about the sectors which have positively or negatively correlated betas will shed light on the *reasons* for the changes in the betas. The discussion in section 9.4 centres on adapting the Kalman filter to accommodate non-normality in the distributions of the returns series. The robustness adaptation is necessary because non-normality presents a fairly serious problem for filtering techniques as discussed in the review of the literature in Chapter 2.

Both the usual Kalman filter and the robust version are implemented in an empirical study of beta coefficients from the JSE. Some important practical considerations for conducting such an empirical study are discussed in section 9.5 and the methodology and results of the study are described in section 9.6. The covariance matrices of the beta coefficients form the centre of attention in the section and in fact give rise to interesting observations about how the management of a portfolio's systematic risk can be enhanced. Further, a statistical test for the significance of the covariance components in improving the fit of the multivariate state-space

model is conducted. Finally, in section 9.7, some implications of the results are deliberated and conclusions offered.

9.2 Multivariate state-space modelling of beta coefficients

In this section a multivariate state-space model of beta coefficients is described. The detailed description of the state-space model for systematic risk is included in the chapter in order to provide a fairly comprehensive reference for future research using state-space models. It is initially based on a version of the Ohlson and Rosenberg (1982) ARMA(1,1) model. This is a *very* general model and involves a large number of parameters which require estimation in order for the model to be implemented and estimates of beta obtained. Nevertheless, it offers a suitable starting point from which to justify and develop a more simplified model. For the purpose of clarifying the extension to the multivariate model, the exposition begins by focusing on the univariate model for time-varying beta coefficients.

The univariate state-space model for beta coefficients comprises two sets of equations: the *observation equation* and the *state equations*. The *observation equation* appears very similar to the usual market model, but allows the parameters (α and β) to vary with time. The *state equations*, on the other hand, provide a model for the time-varying parameters from the *observation equation*. In the univariate case, the *observation equation* is:

$$y_{i,t} = \alpha_{i,t} + \beta_{i,t} r_{m,t} + \epsilon_{i,t} \quad (1)$$

where the return on the market in interval t is denoted by $r_{m,t}$; the return on security i in interval t is denoted by $y_{i,t}$ and the residual component in the return is denoted by $\epsilon_{i,t}$.

The *state equations* for β are given in (2) below and specify that the β parameters vary according to an ARMA(1,1) process:

$$\begin{aligned} (\beta_{i,t} - \bar{\beta}_{i,t}) &= \phi_i^\beta (\beta_{i,t-1} - \bar{\beta}_{i,t-1}) + \eta_{i,t}^\beta \\ \bar{\beta}_{i,t} &= \bar{\beta}_i + u_{i,t}^\beta \end{aligned} \quad (2)$$

The first of the state equations indicates that the time series of differences between the security betas and their mean values (in intervals t and $t-1$ the differences are denoted by $\beta_{i,t} - \bar{\beta}_{i,t}$ and $\beta_{i,t-1} - \bar{\beta}_{i,t-1}$, respectively) follow an autoregressive process of order one. In other words, the

model posits that the difference between the current beta and its mean is correlated with the difference between the previous beta and its mean. The closer the ϕ parameter is to unity, the more highly correlated the differences are. If the ϕ parameter were near zero then the difference between the beta and the mean value would be random and therefore unrelated to the difference between the previous beta and its mean. The second equation allows further variation in the series of beta coefficients by specifying that the mean value of the beta in each interval ($\bar{\beta}_{i,t}$) is *itself* a random variable which varies around an *overall* mean beta value, $\bar{\beta}_i$. (The $\eta_{i,t}^\beta$ and $u_{i,t}^\beta$ are independent, zero mean disturbance terms.)

If matrix representation is considered, it can readily be seen that the univariate model which applied to a given security on its own can be written more compactly in an observation equation and a *single* state equation, viz:

$$\begin{aligned} y_{i,t} &= z_t' \gamma_{i,t} + \epsilon_{i,t} \\ \gamma_{i,t} &= T_i \gamma_{i,t-1} + \nu_{i,t} \end{aligned} \quad (3)$$

for $t = 1, 2, 3, \dots, T$ and for $i = 1, 2, 3, \dots, N$

and where

$$z_t' = [1 \ r_{m,t} \ 1 \ r_{m,t} \ 1 \ r_{m,t}] \quad (4)$$

$$T_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi_i^\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & \phi_i^\beta \end{bmatrix} \quad (5)$$

$$\gamma_{i,t} = \begin{bmatrix} \bar{\alpha}_i \\ \bar{\beta}_i \\ \bar{\alpha}_{i,t} - \bar{\alpha}_i \\ \bar{\beta}_{i,t} - \bar{\beta}_i \\ \alpha_{i,t} - \bar{\alpha}_{i,t} \\ \beta_{i,t} - \bar{\beta}_{i,t} \end{bmatrix} \quad (6)$$

and

$$\nu'_{i,t} = [0 \ 0 \ u_{i,t}^\alpha \ u_{i,t}^\beta \ \eta_{i,t}^\alpha \ \eta_{i,t}^\beta]$$

Nothing in expressions (4), (5), (6) and (7) above is new; they are intended as definitions which permit the components of the state-space model to be referred to more concisely. Using the definitions in equations (4) to (7) above (and introducing a few fairly obvious extensions to them), the multivariate state-space model for the several securities can then be written as

$$\begin{aligned} y_t &= Z_t \gamma_t + \epsilon_t \\ \gamma_t &= T \gamma_{t-1} + \nu_t \end{aligned} \quad (8)$$

where Z_t is block diagonal with z_t blocks; T is also block diagonal with T_t blocks; γ_t is obtained by stacking all the $\gamma_{i,t}$ vectors from each security on top of each other and ν_t is similarly obtained by stacking all the $\nu_{i,t}$ vectors.

In this section, the mathematical formulation of a very general, multivariate state-space model for security returns has been presented. In the next sub-section some simplifications of the model which make it easier to use are considered.

9.2.1 Simplified versions of the general model

It should be noted that there are many simplifications that can be considered in the general model presented in the previous section which will greatly facilitate any empirical investigations and which may in fact be as useful as the general model.¹ Firstly, one can consider the random coefficients case which corresponds to the general case with the ϕ parameters along with the values of the u random influences being set to zero. Several studies (e.g. Brooks *et al* (1992) and Bos and Newbold(1984)) have shown empirically that the random coefficients model is a more appropriate model than the full autoregressive model. In the random coefficients model, the current beta fluctuates randomly around a stable, overall mean value.

A second simplification that one could consider is setting constant autoregressive parameters (the ϕ) across all securities. The effect of setting a constant ϕ is to specify an equal tendency for the betas of all the securities to drift towards their mean values. If it were considered reasonable to assume that all firms with a given level of systematic risk would adopt a similar risk strategy in selecting the new projects to be taken on, then the same autocorrelation in the betas for all shares would be justified.

Thirdly, one could consider (as many authors have), a pure AR(1) process which involves setting the u random values equal to zero. The pure AR(1) model removes the additional source of variation in the current beta by omitting the so-called *measurement error* in the mean values (the $\bar{\beta}_{i,t}$).

A fourth, well-known model is the random walk, which would be obtained by setting the autoregressive parameters to unity and the u values to zero. A random walk model supposes that the systematic risk of the firm has no average, long-term value, but varies randomly from interval to interval. In practical terms the firm is assumed to select new projects without reference to a desired long-term level of systematic risk.

Clearly other variations on the general model are possible, e.g. a random walk in the α 's and a random coefficient model in the β 's, and the Kalman filter offers a flexible tool for fitting the models. In this chapter the chosen model corresponds with the case where α 's are

¹ Discussion of the types of structural models which have been proposed in the literature for time-varying betas can be found in Chapter 2.

constant and β 's vary with a random coefficients pattern. The reasons for choosing this model are, firstly, the support that the random coefficients model has received from the literature and, secondly, its simplicity. An interpretation of the model might be that a firm's current systematic risk deviates from its long-term systematic risk because of its current portfolio of projects. The assumption underlying the model is then that the firm always chooses to invest in projects which hold approximately the same average systematic risk, but which vary randomly around that average. Symbolically the model is obtained by setting

$$T_i = \text{diag}(1, 1, 0)$$

$$z_t' = [1 \ r_{m,t} \ r_{m,t}] \quad (9)$$

$$\gamma_{i,t} = \begin{bmatrix} \alpha_i \\ \bar{\beta}_i \\ \beta_{i,t} - \bar{\beta}_i \end{bmatrix} \quad (10)$$

and

$$v_{i,t}' = [0 \ 0 \ \eta_{i,t}] \quad (11)$$

In order to use the conventional Kalman recursions (see Harvey (1989) or Meinhold and Singpurwalla (1983)) the residual and random components are assumed to be Gaussian. The issue of non-normality is dealt with more comprehensively in section 9.4.

To be precise the following assumptions are made in the state-space model. The ϵ_t (the residual returns from the *observation equation* in (8)) are assumed to be both serially and cross-sectionally independently distributed with mean vector $\mathbf{0}$ and variance-covariance matrix $H = \sigma_\epsilon^2 \text{diag}(1/\lambda_i)$ for $i = 1, 2, 3, \dots, N$. The random beta movements, η_t , are distributed with zero means and variance covariance matrix Q_η . The variance-covariance matrix (denoted Q) of the random components of the state equations (i.e. the v vector) thus has a block diagonal structure with matrices $\mathbf{0}$, $\mathbf{0}$ and Q_η on the diagonal, where the $\mathbf{0}$ are $N \times N$ matrices. The Q_η matrix is the matrix of interest as it reflects the covariance structure of the sector betas.

The discussion in this section has centred on finding a simplification of the general state-space model which is appropriate for beta estimation. In particular, the random coefficients model which has emerged from the literature as the leading univariate contender for modelling time-varying betas was formulated in a *multivariate* state-space framework. The next section describes how the structural model set out above can be *estimated* using the Kalman filter. The description follows closely that given in Harvey (1989) and is presented here for completeness and to facilitate the discussion later in the chapter.

9.3 Estimation using the multivariate Kalman filter

The Kalman filtering process for estimating state-space model parameters is described in this section. Briefly, the filter comprises two sets of equations; the *prediction* and *updating* equations. The *prediction* equations specify the mean and variance of an estimator for the state equation parameters at time t based on all the observations up until time $t-1$. The *updating* equations update the mean and the variance of the estimator with the new observation at time t . The two sets of equations build recursively on each other with the one being dependent on the other at a lag of one observation.

The estimation error matrix P_{t-1} is defined to be the $3N \times 3N$ matrix of the optimal estimator of γ_{t-1} (i.e. the true alpha and beta parameters), a_{t-1} :

$$P_{t-1} = E[(\gamma_{t-1} - a_{t-1})(\gamma_{t-1} - a_{t-1})'] \quad (12)$$

Given a_{t-1} and P_{t-1} (the estimates of the parameters and the error matrix at time $t-1$), the optimal estimator of γ_t , $a_{t|t-1}$, is

$$a_{t|t-1} = Ta_{t-1} \quad (13)$$

The covariance matrix of the estimation error, $P_{t|t-1}$, is given in

$$P_{t|t-1} = TP_{t-1}T' + Q \quad (14)$$

The above set of relationships (equations (13) and (14)) are referred to as the *prediction equations*. Once the observation vector (i.e. the returns on all the securities) at time t becomes known, the predictions of the parameters are updated. The *updating equations* are then

$$\begin{aligned} a_t &= a_{t|t-1} + P_{t|t-1}Z_t'F_t^{-1}(y_t - Z_t a_{t|t-1}) \\ P_t &= P_{t|t-1} - P_{t|t-1}Z_t'F_t^{-1}Z_t P_{t|t-1} \end{aligned} \quad (15)$$

with the matrix F_t defined as

$$F_t = Z_t P_{t|t-1} Z_t' + H \quad (16)$$

which can be thought of as the variance-covariance matrix of the prediction errors, $y_t - Z_t a_{t|t-1}$.

The Kalman filter is made functional by suitable choices of a_0 and P_0 . Harvey (1989, p121), for example, suggests $a_0 = 0$, $P_0 = \text{block diag}(Q, \kappa I)$ where κ is a very large value used to approximate infinity. The elements of the variance-covariance matrices (Q and H) are estimated using maximum likelihood, the details of which are described in section 9.3.1.

It turns out that if multivariate normality is assumed in the distributions of the residual terms then there are more efficient algorithms for the Kalman filter. Two of these alternative algorithms are outlined in Appendix 9.A. The second of them (the generalised least squares approach) is used in estimating the parameters in the empirical investigation described in section 9.6.

The Kalman filter as briefly outlined in this section was shown to be the appropriate methodology for actually estimating the α 's and β 's of the securities in the state-space framework. As discussed above, the Kalman filter operates by considering each observation in turn and updating the estimates of the α 's and β 's depending on how divergent the actual observation and the prediction using all the data up to the previous observation are. In the next section, the estimation of the covariance matrices needed for the functioning of the Kalman filter is discussed.

9.3.1 Maximum likelihood estimation of the variance-covariance matrices

The parameters in the matrices $\mathbf{Q}$ (the beta variance-covariance matrix) and $\mathbf{H}$ (the covariances of the residuals in the observation equation) will not, in general, be known and yet they are essential for the operation of the Kalman filter. While the properties of the Kalman filter strictly apply only when these ‘hyperparameters’ are known and fixed, it is usually reasonable to estimate them from the same data set. Additionally, the estimated ‘hyperparameters’ are in this case the parameters of particular interest, because $\mathbf{Q}$ corresponds to the covariances of the beta coefficients from the different sectors.

The method of maximum likelihood is used to estimate the hyperparameters. The maximum likelihood estimates are obtained by maximizing the (logarithm of the) prediction error density, i.e. the multivariate likelihood function of the $y_t - \mathbf{Z}_t \mathbf{a}_{t|t-1}$. In symbols this entails maximizing

$$\sum_{t=1}^T \log(p(y_t | Y_{t-1})) \quad (17)$$

with respect to the components of the variance-covariance matrices.

In the usual case where the observation equation residuals are Gaussian, the likelihood function² can be written as

$$\log L = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T |\mathbf{F}_t| - \frac{1}{2} \sum_{t=1}^T \nu_t \mathbf{F}_t^{-1} \nu_t \quad (18)$$

where $\nu_t = y_t - \mathbf{Z}_t \mathbf{a}_{t|t-1}$. The first term on the right hand side of equation (18) is constant over time for a given sample and is not involved in the maximization process. The second term in equation (18) can be thought of as requiring the variance of the prediction errors to be small in order for the log-likelihood to be maximized. The final term in the log-likelihood represents the sum of the squared, standardized residuals from the prediction equations. In order to maximize the likelihood this term must be minimized as far as possible.

² See Appendix 9.A for details on how the likelihood is modified when the generalised least squares approach is used.

In this sub-section, it has been reasoned that estimates of the variance-covariance matrices can be obtained using the method of maximum likelihood on the prediction errors, i.e. the differences between the predicted returns and the observed security returns.

Because the prediction residuals are *squared*, it is the final term of the likelihood which might make maximum likelihood estimates of the hyperparameters sensitive to outlying observations. In addition, the actual beta estimates (as well as the estimates of the hyperparameters) of the Kalman filter will be sensitive to departures from normality (see Meinhold and Singpurwalla (1989)). Therefore, even if the maximum-likelihood estimates of the hyperparameters are similar for the normal and robust cases, the estimates of the time-varying parameters might not be. Hence a modification to state-space filtering which does not assume normality is considered in the next section.

9.4 Robustification of the Kalman filter

Above it was noted that both the estimation of the alpha and beta parameters as well as the variance-covariance matrices were sensitive to outliers which might arise if security returns were non-normally distributed. In fact, it is well known that the distributions of security returns are indeed non-Gaussian (see, for example, Affleck-Graves and MacDonald (1989) for recent evidence from the NYSE, or Chapter 6 in this thesis for evidence from the JSE). The simple recursive nature of the Kalman filter is based on the property of the normal distribution which allows it to be defined by only the expectation and variance-covariance matrix. When the residual components cannot be assumed normal, the Kalman filter loses some of its optimality properties but still produces minimum mean square *linear* estimators of the state equation parameters (see Harvey, 1989, p111). Carlin, Polson and Stoffer (1992) have commented, however, that a *best linear* predictor might be substantially different from an optimal forecast of the beta coefficients.

Recursive routines similar to that of the Kalman filter can be developed which have as their aim the derivation of the *entire distribution* functions of the state equation parameters (the alphas and betas) conditional on the observations up until the current time period. Unfortunately the derivation of the distribution is generally fraught with numerical problems as it is found by convolving two densities. Kitagawa (1987) has suggested using a 'brute force' numerical integration procedure, but, having programmed it for the univariate case, the author has found (along with many others) that it is too time-consuming to be of practical use in investigating

time-varying beta coefficients. Carlin, Polson and Stoffer (1992) tackled the problem using a Monte Carlo approach based on a relatively recent statistical technique known as the Gibbs sampler;³ again, this approach involved quite considerable computational time. Meinhold and Singpurwalla (1989) proposed a robustification of the Kalman filter based on an approximation involving 'poly-t' distributions. However, their approach is really best suited to univariate situations as they point out themselves. They also concentrated more on developing the whole distribution of the estimate of beta rather than finding a point estimator.

The problem with considering the whole conditional distribution is that it does not aid the practitioner who is, after all, generally interested in obtaining only one value - the 'best' estimate of beta. In many cases it is difficult enough to get the user to consider at least the variance component.

The work of Masreliez (1975) and subsequent developments by Martin (1979, 1981) do yield a reasonably attractive way of surmounting the problem. The approach is based on what are known as *approximate maximum-likelihood* type estimates. The basic assumptions of Masreliez's approximate filter are that the innovations (random changes) in the betas conditional on the observations are Gaussian, but that the residuals of the observations equation need not be. In what follows, the theorem proved by Masreliez is described and a few of the implications for beta estimation are noted. The main difference between Masreliez's filter and the standard Kalman filter lies in the updating equations (equation (15)) and centres on reducing the impact that an outlying observation has on the updated beta estimate.

The updating equations are modified on the basis of the assumptions so that they become

$$\begin{aligned} a_t &= a_{t|t-1} + P_{t|t-1} Z_t' g_t(y_t) \\ P_t &= P_{t|t-1} - P_{t|t-1} Z_t' G_t(y_t) Z_t P_{t|t-1} \end{aligned} \quad (19)$$

³ Carlin *et al* (1992) discuss the implementation of the Gibbs sampler in some detail and note that many of the properties of the Gibbs sampler remain under investigation due to its recent appearance in the statistical literature.

where $\mathbf{g}_t(\mathbf{y}_t)$ is a column vector with components

$$\{\mathbf{g}_t(\mathbf{y}_t)\}_i = - \left[\frac{\partial p(\mathbf{y}_t | \mathbf{Y}_{t-1})}{\partial (\mathbf{y}_t)_i} \right] \cdot [p(\mathbf{y}_t | \mathbf{Y}_{t-1})]^{-1} \quad (20)$$

and $\mathbf{G}_t(\mathbf{y}_t)$ is a matrix with elements

$$\{\mathbf{G}_t(\mathbf{y}_t)\}_{ij} = \frac{\partial \{\mathbf{g}_t(\mathbf{y}_t)\}_i}{\partial (\mathbf{y}_t)_j} \quad (21)$$

Note that $p(\mathbf{y}_t | \mathbf{Y}_{t-1})$ is the density function of the current observation conditional on all past observations. The function $\mathbf{g}_t$ can therefore be recognized as the score function of the conditional density. It can be seen from the first of the updating equations in (19) that the larger the value of the score function at a particular time, the greater the influence that observation has on the updated estimates of the time-varying parameters. By comparing equations (19) and (15) it can be seen that in the usual Kalman filter $\mathbf{g}_t$ is a vector of the standardized errors from the prediction equation. A few points which arise from the new updating equations are:

- (a) When the density of the observation equation residuals is Gaussian then the approximate filter and the usual Kalman filter are equivalent.
- (b) Although it is generally not easy to derive the score function, $\mathbf{g}_t$, because of the convolution required, it can be chosen directly to have suitable properties (this aspect is discussed in more detail in the next section).
- (c) The filter is now *approximate* (the conditional distribution of the state-equation parameters are approximated by a Gaussian distribution and this assumption will rarely hold *exactly*).
- (d) The estimates of the hyperparameters will no longer be maximum-likelihood estimates, but will rather be *approximate* maximum-likelihood estimates owing to the assumption referred to in (c).

Masreliez's theorem can therefore be used as a basis for performing a Kalman filter-type procedure which will yield parameter and hyperparameter estimates which are not sensitive to the one or two outlying values inevitably found in security returns. The next sub-section

contains a short discussion of how to select the g_t function which is necessary for the execution of the Masreliez filter.

9.4.1 The influence of an observation

When the density of the observation residuals is Gaussian the score function (g_t) is linear. The linearity can be identified by remembering that in the usual Kalman filter (which assumed normality) the g_t vector is just the vector of standardized residuals. Hence a large residual value will directly cause a large score function value. As mentioned in the discussion on Masreliez's filter, a large score function has great *influence* on the updated estimates of the time-varying parameters. The implication of linearity in the score function is therefore that observations which are far from the 'average' will have greater *influence* on the estimation procedure than those closer to the mean. In the situation where the observation equation is subject to outliers the linearity of g_t is undesirable: observations which are subject to extreme *random* disturbances will exert undue influence on the estimates of the betas. In other words, one uncharacteristically large or small security return will severely distort the estimates of the betas and the covariance matrices. A more satisfactory situation is to ensure that the score function down-weights extreme observations while still allowing small and moderately large observations to influence the estimation in the usual Gaussian fashion.

There are two ways of making g_t comply with the ideals set out in the previous paragraph. Firstly, the density function of the residuals can be chosen to be a heavy-tailed one. Because the score function of a heavy-tailed distribution at outlying values is smaller than value itself, the down-weighting is achieved automatically. This method has its problems because $p(y_t | Y_{t-1})$ has to be obtained by convolving two densities. Except in a few cases, the convolution operation is non-trivial. A second method is to select the g_t function directly without reference to any particular distribution. This method has the disadvantage that it may imply a rather unconventional density function for the residuals (which may be difficult to justify in practice).

The numerical convolution referred to in the first method can be performed quite easily if the residual distribution is assumed to be a contaminated normal distribution. In 9.4.2 the details of implementing the approximate filter with the assumption of a contaminated normal for the residual distribution are described. In 9.4.3 an example of score function which satisfies the down-weighting criterion, but which does not correspond to a particular distribution of the residual, is also suggested.

9.4.2 Contaminated normal distribution

In the previous section it was concluded that one way of robustifying the Kalman filter was to specify a non-normal distribution for the return residuals. The particular distribution considered in this section is the contaminated normal, which was also used in Chapter 7. As described in Masreliez (1975) and Martin (1979, 1981) the contaminated normal distribution is a simple linear mixture of (usually two) normal distributions. The one distribution can be thought of as describing the bulk of the observations (the 'good' observations), whereas the second has a larger variance and describes the distribution of the 'bad' or outlying observations. The outlying observations are assumed to appear with a frequency of ζ , say. In practical terms, the contaminated normal distribution might be considered reasonable for the residuals if it were assumed that information arriving in the market was either 'usual' information (which would lead to 'good' observations) or 'unusual' information which *could* lead to extreme reactions in the security prices (i.e. 'bad' observations).

The formulae for the score vector and its derivative matrix are given in Appendix 9.B, but a sketch of the univariate equivalent of the score function serves to demonstrate a few points relevant to the discussion of robustness. In the following discussions the score function is referred to as the *influence* function in order to better describe its operation on the observations.

Figure 9.a depicts the influence function based on a contaminated normal distribution for the residuals with a relative frequency of $\zeta = 10\%$ for the outlying observations and a multiplier of 9.8 for the variance. The relative frequency of the outliers and the variance of the outliers have been selected to correspond with those used in the empirical study described in section 9.6. The interpretation of the parameters of the contaminated normal distribution selected is that it is assumed that the 10% of the observations which might be outliers have a standard deviation of just over three times that of a 'normal' observation. The observations with the higher standard deviation can be thought of as arising from information about the firm which *might* cause an extreme reaction in the share price. Figure 9.a illustrates how the influence function is linear in the vicinity of the mean,⁴ but starts to decrease for larger observations. Hence large observations are down-weighted in the updating equations so that they will influence the estimate of the betas only marginally. The one drawback of the

⁴ The mean of prediction errors is zero.

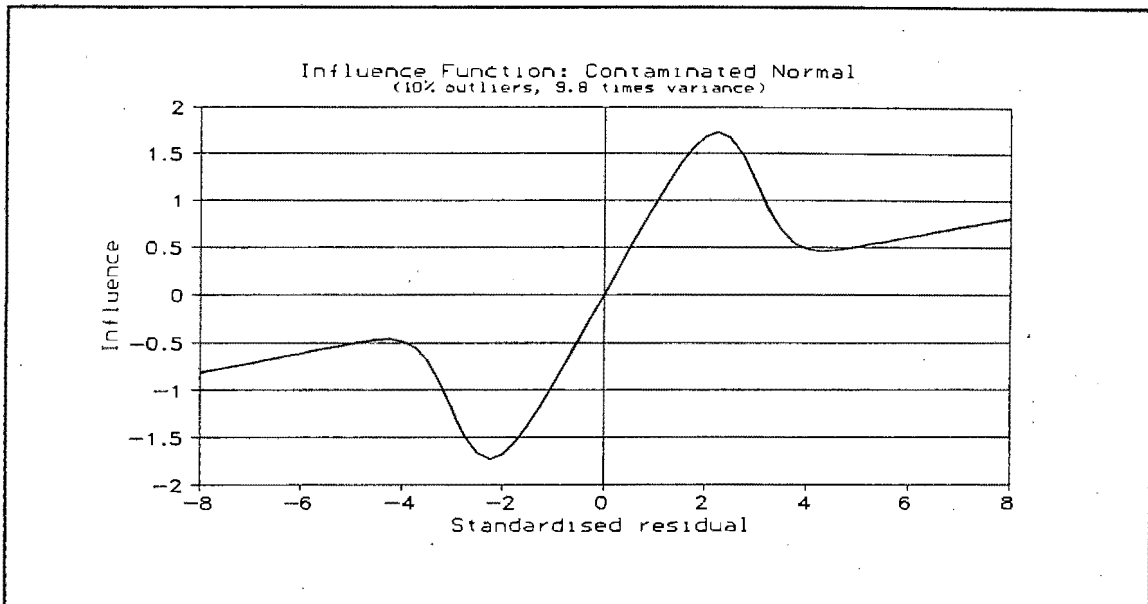


Figure 9.a

contaminated normal assumption is that extremely large observations will once again have a significant impact on the estimation process. In Figure 9.a the drawback manifests itself in that the influence function starts to increase again for extreme values.

By manipulating the value of the multiplier or the ζ , the properties of the influence function can be chosen to suit the particular data set. For example, increasing ζ to 20% will cause the influence function to tend back to zero ('redescend') for smaller prediction errors than the case where ζ is 10%, i.e. more observations will be categorized as outliers and hence the influence of more of the observations will be down-weighted. On the other hand decreasing the multiplier to, say, four, will cause a less marked redescension of the influence curve back to zero and the down-weighting of outliers will not be as severe. It is apparent from the discussion in this section that the contaminated normal distribution offers a fairly easily interpretable distribution for the residuals as well as offering a flexible range of influence functions to robustify the filter.

In the next section an influence function which does not have a readily interpretable distribution associated with it is described briefly. Although the Huber function is not used in the empirical study described in section 9.6, it is mentioned here because it is a well-known function in the literature on robust estimation. It also demonstrates how the Masreliez filter offers other methods for robustifying the Kalman filter.

9.4.3 A Huber influence function

Huber (1964) proposed a monotone influence function which has a univariate form:

$$g(v) = \begin{cases} v, & |v| \leq k \\ \text{sign}(v) \times k, & |v| > k \end{cases} \quad (22)$$

The concept in the Huber influence function is to give all prediction errors above a certain limit (k) a fixed influence. In contrast with the influence function described for the contaminated normal in the previous section, very large prediction errors are assigned the *same influence* as moderately large errors rather than *smaller influence*.

Huber's function is specified in univariate terms for a scalar valued residual. In the multivariate situation the prediction errors are vectors of errors corresponding to the errors of each security in the sample. The error vector is denoted v_t . A possible solution is to downweight all observations with a norm greater than some fixed value k so that the adjusted error vector has a norm set to the fixed value, k . Thus

$$g_t(v_t) = \begin{cases} v_t & \|v_t\|^{0.5} \leq k \\ v_t \times \frac{k}{\|v_t\|^{0.5}} & \|v_t\|^{0.5} > k \end{cases} \quad (23)$$

The influence function retains the relative sizes of the prediction errors at each observation, but down-weights the whole observation if the 'magnitude' of the errors is large.

Before describing the empirical investigation which was undertaken to obtain information about the correlations between the betas of various sectors, some practical considerations with regard to implementing the filters are discussed.

9.5 Some practical considerations in testing the filters

Thus far in the chapter much of the discussion has been theoretical and has concentrated on selecting the ideal methodology for estimating betas and the covariances. In

this section the ideas and techniques outlined previously are re-assessed in the light of empirical limitations. Ideally one should apply Masreliez's version of the Kalman filter to all the securities on the stock exchange using as much time series data as possible. There are, however, some important practical considerations which might make such a complete empirical investigation infeasible. Firstly, as discussed in Chapter 3, severe thin trading on the JSE causes non-synchronicity in the measurement of returns. The non-synchronicity implies that it will be impossible to measure the returns of a set of firms all at the same time because not all the shares will have been traded at that time. Secondly, new listings, delistings and suspensions of securities will generate missing observations in the return time series. A third point is that the large number of hyperparameters associated with a large number of securities require estimation. The resulting loss in degrees of freedom will reduce the power of any tests of lack of fit or of the significance of parameters considerably.

Although the first two issues could be dealt with using slightly more complex filter algorithms based in a continuous-time framework (see Harvey, 1989), it is felt that the additional complications would not necessarily generate more useful results, particularly at this early stage in assessing the usefulness of the multivariate approach. Another reason for not using the more complex models is that the power of the statistical tests would be decreased because of the extra parameters requiring estimation in the continuous time framework.

With regard to combatting the problem of non-synchronous trading in securities in a practical way, Cohen *et al* (1983) have suggested that the use of long intervals over which to measure returns will reduce the impact of any frictions in the pricing process. In the empirical study which is discussed in the next section, quarterly intervals have been used.

In addition to using long intervals over which to measure returns, the problem of thin trading was also avoided by using the JSE-Actuaries Sector Indices as the 'securities' of which the betas were estimated. The sector indices are weighted portfolios of generally well traded securities. Apart from helping to avoid the non-synchronicity problem and the problem of missing data resulting from delistings, new listings and suspensions, the use of *sector* betas permits interpretations of the correlations of the betas which are not obscured by *firm specific* characteristics.

The details and results of the empirical tests performed using data from the JSE are described in the next section.

9.6 Empirical investigation

As motivated in the previous section, the JSE-Actuaries sector indices were used as inputs in the multivariate state-space model. The major sector indices were used, namely Gold, Other Metals and Minerals, Mining Houses, Mining Holdings, Financial and, finally, Industrial. The period of investigation ran from October 1979 until the beginning of October 1993 and comprised 60 quarters. The results are reported in two parts: firstly, the results using the usual Kalman filter and the Gaussian assumption and, secondly, using the Masreliez filter with the contaminated normal distribution. The focus of the results is on the correlation matrix of the betas from the different sectors. Other results reported to check on the plausibility of the model are the alpha coefficients and the overall, mean values of the beta coefficients referred to in section 9.2 (which described the random coefficients model used in the empirical investigation). The mean beta values can be interpreted as the best forecasts of the betas for the sectors given that any future variations are modelled as random, unpredictable fluctuations around these values.

Table 9.1 gives the values of the mean betas and the alphas. From Table 9.1 it is evident that the values of the beta coefficients are very much in line with estimates based on the traditional static beta assumptions, as might be expected. The plausibility of the model is

Table 9.1

The base values of the sector betas, the alphas and observation error standard deviations. (Quarterly data, 1979.10 to 1993.10)

Sector	Mining Houses	Mining Holdings	Industrial	Financial	Gold	Other metals and minerals
Beta	1.216	1.143	.635	.537	1.383	1.137
Alpha	-.005	-.022	.028	.028	-.037	-.001
S.D.	.047	.060	.035	.054	.090	.085

confirmed, for example, by noting that the mining sectors have very much higher systematic risk components than the industrial and financial sectors (e.g. the All Gold sector has a beta of 1.383 and the Industrial sector has a beta of 0.635).⁵ In the third row of Table 9.1 are the

⁵ Estimates of the sector betas resulting from the static beta model can be found in Bradfield and Bowie, *The Financial Risk Service*.

maximum likelihood estimates of the standard deviations of the observation error terms, i.e. the diagonal elements of H . Again, it is reassuring that the estimates from the state-space model are of the same order of magnitude as those from the static beta model. For example, the fact that the All Gold sector index has a residual standard error of 0.090 compared with 0.054 for the Industrial index is in line with the conventional wisdom among investors on the JSE that gold shares not only have higher systematic risk components, but also have higher firm specific (residual) variances.

Table 9.2 displays one of the main results of the empirical study, namely the correlation matrix of the stochastic elements of the betas. The standard deviations of the elements are given on the diagonal. From Table 9.2 it is apparent that there is strong positive

Table 9.2

Correlation matrix of stochastic elements of betas. Standard deviations of the stochastic elements are given on the diagonal. (Quarterly data, 1979.10 to 1993.10.)

Mining Houses	Mining Holdings	Industrial	Financial	Gold	Other Metals and Minerals	
.113	.78	-.70	-.60	.74	-.04	Mining Houses
	.651	-.98	-.92	1.00	.44	Mining Holdings
		.620	.98	-1.00	-.60	Industrial
			.538	-.96	-.74	Financial
				.902	.52	Gold
					.325	Other Metals and Minerals

correlation between the industrial and financial sectors, i.e. 0.98. There are also pairwise positive correlations between the Mining Houses, Mining Holdings, Gold and Other Metals and Minerals sectors with the exception being a virtually zero correlation between the Mining Houses and Other Metals and Minerals sectors (i.e. -0.04). Further, there are negative correlations between the movements in the betas of the non-mining sectors (the Financial and the Industrial) and all the mining sectors, e.g. the correlation between the All Gold and Financial sectors is -0.96.

The interpretation of the correlation matrix of betas is that any information absorbed by the market appears to hold divergent consequences for the *systematic risk* (as well as the

returns) of the shares on the two main boards of the JSE (the *mining* board⁶ and the *financial and industrial* board⁷). The betas of the sectors on the mining board nearly all⁸ have strong positive correlations with each other. Similarly the beta of the Financial Index is strongly positively correlated (the correlation coefficient is 0.98) with the Industrial Index. In contrast, the correlation between the betas of each sector of the financial and industrial board and each sector of the mining board is strongly negative. For example, the betas of the All Gold and Industrial sectors have a nearly perfect negative correlation (the estimate of the correlation coefficient reported in the table has been rounded to -1.00).

An implication of the split between the two boards is that to manage their risk properly, investors must consider securities from both boards in their portfolio. Should investors invest only in the firms of one board then they will be exposed not only to the increased firm specific risk caused by limited diversification, but also to systematic risk which will vary with time. On the other hand, should the investor hold shares in both boards then, for example, an upward swing in the systematic risk of the mining shares may well be compensated for by a downward swing in the systematic risk of the financial and industrial shares. The result of the compensating changes in systematic risk of the shares from the two boards is that the investor remains exposed to the *same* selected systematic risk over time.

9.6.1 Empirical results using contaminated normal approximate filter

In this section the results obtained using the robust, approximate filter⁹ described in section 9.4 are reported. The results from both the usual Kalman filter and the robust¹⁰ version are presented in this chapter for comparative purposes.

⁶ The mining board consists of the All Gold, the Other Metals and Minerals, the Mining Houses and the Mining Holdings sectors.

⁷ The financial and industrial board comprises the Financial and the Industrial sectors.

⁸ The only exception is the very small negative correlation of the Other Metals and Minerals sector and the Mining Houses sector.

⁹ Details of the score function and its derivative are given in Appendix 9.B.

¹⁰ The contaminated normal distribution was used with an assumed 10% probability for an observation from the outlier distribution. The optimization routine used to maximize the likelihood function for estimates of the hyperparameters was also used to select the variance inflation factor. On convergence of the routine the inflation factor had a value of 9.80 (i.e. the distribution of the potentially outlying values had a standard deviation approximately three times that of the usual observations).

In Table 9.3 the correlation matrix of the time varying components of the betas is presented. The results are largely the same as those found without the robust filter, i.e. there is positive correlation between the Financial and Industrial sectors (of 0.97), but negative

Table 9.3

Correlation matrix of stochastic elements of betas. Standard deviations of the stochastic elements are given on the diagonal. (Quarterly data, 1979.10 to 1993.10., robust filter)

Mining Houses	Mining Holdings	Industrial	Financial	Gold	Other Metals and Minerals	
.219	.48	-.55	-.52	.44	.20	Mining Houses
	.737	-.96	-.87	.99	.34	Mining Holdings
		.834	.97	-.95	-.56	Industrial
			.671	-.88	-.69	Financial
				1.25	.31	Gold
					.721	Other Metals and Minerals

correlations between the *financial and industrial* board and the *mining* board. The fact that there is little difference between the results here and the results described in Table 9.2 does not imply that the additional complications caused by allowing for non-normality are unnecessary. Although the maximum likelihood estimates of the *hyperparameters* are similar, the actual estimates of the *beta coefficients* will be different. Further, the return distributions of sector indices may well be closer to the normal distribution than those of individual securities.

One difference between the correlation coefficients of the betas in the normal case compared with the betas in the contaminated normal case which is apparent, is the correlation between the beta of the Other Metals and Minerals sector and the Mining Houses sector which has increased from -0.04 to 0.20. The positive correlation between the two sectors is now in line with all the other positive correlations among the sectors on the mining board.

In Table 9.4 the standard deviations of the (non-outlying) return residuals and the final beta and alpha estimates are presented. It is evident from the table that the estimates are very similar to those obtained from the normal Kalman filter. For example, the largest percentage difference in the estimates of the mean betas is just over 9% (the increase in the beta of the Other Metals and Minerals from 1.040 in the normal filter to 1.137 in the robust filter). The

Table 9.4

The base values of the sector betas, the alphas and observation error standard deviations. (Quarterly data, 1979.10 to 1993.10, robust filter)

Sector	Mining Houses	Mining Holdings	Industrial	Financial	Gold	Other metals and minerals
Beta	1.213	1.127	.644	.563	1.446	1.040
Alpha	-.005	-.028	.030	.023	-.057	.014
S.D.	.031	.046	.017	.049	.050	.051

residual standard deviations of the observation equation are all smaller than in the normal filter. The reason for the reduction is that the standard deviations in the contaminated normal case represent the standard deviations of the 'good' observations. The standard deviations of the outlying observations are approximately three times the value of these. In the normal filter, the standard deviations are a measure of the average spread of the 'good' and the outlying observations. Therefore the values are somewhat larger than the standard deviations of the 'good' observations taken on their own.

Although the multivariate state-space model appears to offer improved insights into the management of systematic risk of portfolios, it is important to test statistically the improvement offered by the additional parameters in the model. If the improvement is not sufficient to justify the increased number of parameters then many of the interpretations of the enlarged model will be spurious. In the next section, therefore, the improvement in the fit offered by the multivariate model is tested for significance.

9.6.2 Statistical tests of the model

In the previous section the multivariate state-space framework was implemented using the main sector indices on the JSE. Although the results were plausible and added insight into the co-movements of beta coefficients, it is necessary to test statistically whether the additional estimation requirements are warranted. There are several tests for the significance or otherwise of the additional hyperparameters which describe the non-stationarity of the betas as specified

in the model. The likelihood ratio (LR) test appears to be the easiest of these to perform. It is based on testing the validity of a set of linear restrictions on the vector of hyperparameters, Ψ , viz.:

$$R\Psi = r \quad (24)$$

The likelihood ratio test stems from the Neyman-Pearson lemma and is based on the fact that

$$LR = -2 \log \lambda \quad (25)$$

is asymptotically distributed as χ_m^2 (m is the number of parameters restricted under the null hypothesis and λ is the ratio of the maximised likelihood when $Q_\eta = 0$ to the unrestricted maximized likelihood).

The results of the LR test are given in the middle column of Table 9.5. Two sets of results are presented which test for the significance of various components of the covariance matrix. The first result considers the test for the significance of the whole variance-covariance matrix of the betas. The likelihood ratio reported in Table 9.5 for the significance of the whole Q matrix is a highly significant 48.19. Second, the ratio corresponding to a test for the significance of the *off-diagonal* components of Q is reported. By testing for the significance of the off-diagonal elements it can be ascertained whether the multivariate approach adds significantly to the model, beyond simply allowing for time-varying betas. From Table 9.5 it is apparent that the likelihood ratio for the off-diagonal elements in addition to the variance components is 47.89. The likelihood ratio test gives highly significant results in both cases, i.e. the full model explains significantly more than either of the reduced models. (The critical χ_m^2 values are 38.93 (1% level) for the test against a zero matrix and 30.58 (also a 1% level) for the test against zero off-diagonals¹¹). The very small (and insignificant) decrease in the likelihood ratio (λ) when going from the zero matrix model to the zero off-diagonal model helps to amplify the apparent importance that the correlation structure in the movements in beta plays.

¹¹ The degrees of freedom (the m) are 26-5=21 for the first test and 26-11=15 for the second.

Also given in Table 9.5 are the Akaike Information Criterion (AIC) values (calculated as twice the number of parameters estimated less twice the maximum log-likelihood) for each model. The AIC values offer a quick way of comparing models with different numbers of parameters, with the lower the AIC, the better the model. The AIC of the full model (i.e. -189)

Table 9.5

Likelihood ratio tests and AICs for testing the significance of the variance-covariance matrix of the state equation parameters.

Model	Likelihood ratio (relative to unrestricted model)	Akaike Information Criterion
$Q = 0$	48.19	-183
The off-diagonals of Q set equal to zero	47.89	-171
Unrestricted model	n/a	-189

is less than that of either of the other two models (which are -183 for a zero Q matrix and -171 for zero off-diagonal elements in the Q matrix).

The results of both the likelihood ratio test and the AIC therefore prove that the effort required to estimate the additional parameters in the multivariate model is justified by a significantly better fit of the model.

9.7 Conclusions and summary

This chapter has presented a new way of modelling and estimating time-varying beta coefficients. The multivariate state-space approach and Kalman filter offer a means of estimating time-varying beta coefficients from a variety of models. It was shown how models ranging from the very elaborate to the far more simple could be accommodated in the same framework and estimated using the same procedures. The multivariate nature of the approach further provides insights into the covariances of beta coefficients between securities. These insights into the extent to which the betas of securities co-vary are extremely useful in monitoring and adjusting portfolio systematic risk components.

A serious complication arises in implementing the Kalman filter on stock market data: namely, the non-normality of security returns. Non-normality is usually associated with the existence of potential outliers in the data set which might distort beta estimates and estimates of the beta covariance matrix. The necessity of using a robust version of the Kalman filter for

estimating the parameters and hyperparameters of the state-space model because of the non-normality of security returns was therefore discussed. A filtering procedure based on a theorem of Masreliez (1975) was proposed for performing the robustification. The details of the filtering techniques needed for state-space estimation were spelled out in some detail to provide a complete reference for further research in this area.

The key practical result which emerged from the empirical study performed using sectors from the JSE is that in order to maintain a desired level of systematic risk in a portfolio of shares, the investor must identify the sectors that have betas which are negatively correlated and then invest in shares from both sectors. On the JSE, the two broad sectors are the *mining* and the *financial and industrial* boards.

The empirical results showed that not only does the multivariate state-space framework offer a more *sophisticated* modelling procedure for returns, but that statistical testing of the fit of the model on the JSE indicated that the additional complexity of the model provided *significant improvements* in the fit of the model. In conclusion it is hoped that the encouraging results presented in this chapter pave the way for further research in multivariate estimation of systematic risk.

Appendix 9.A

Rosenberg's algorithm

In the first part of the appendix two algorithms are described for estimating the non-time varying parameters of the state equation. The first of these is known as Rosenberg's algorithm and the second is based on generalised least squares (GLS).

Although the security alphas and the mean betas at time 0 could be estimated using maximum likelihood along with the variance-covariance matrices, the numerical optimisation is greatly simplified by concentrating these values out of the likelihood. One way of concentrating them out is to use an algorithm devised by Rosenberg (1973).

The idea behind the algorithm is to consider γ_0 as fixed rather than stochastic. The algorithm is started by calculating a set of prediction errors vectors ν_t^* based on $a_0 = 0$ and $P_0 = 0$. A set of recursions for G_t is set up as:

$$G_t = (T - K_t Z_t) G_{t-1} \quad (26)$$

with $G_0 = T$ and where the Kalman gain matrix, K_t , is defined by

$$K_t = T P_{t|t-1} Z_t' F_t^{-1} \quad (27)$$

A maximum likelihood estimator (conditional on the other parameters in the model) for γ_0 is then given by

$$\tilde{\gamma}_0 = \left[\sum_{t=1}^T G_{t-1}' Z_t' F_t^{-1} Z_t G_{t-1} \right]^{-1} \sum_{t=1}^T G_{t-1}' Z_t' F_t^{-1} \nu_t^* \quad (28)$$

The log likelihood function for the filter is then given by

$$\log L = -\frac{1}{2} \sum_{t=1}^T \log |F_t| - \frac{1}{2} \sum_{t=1}^T \mathbf{v}_t^*{}' F_t^{-1} \mathbf{v}_t^* + \frac{1}{2} \tilde{\gamma}_0 \sum_{t=1}^T \mathbf{G}_{t-1}' \mathbf{Z}_t' F_t^{-1} \mathbf{v}_t^* \quad (29)$$

and this expression should be maximised with respect to the unknown ‘hyperparameters’ (which correspond to the unknown components of the variance-covariance matrices H and Q).

Generalised Least Squares approach

However a method based on a generalised least squares approach is more convenient to use given the models proposed. What follows is a brief description of the GLS method in as applied to the model proposed in this chapter. The motivation behind the procedure can be read in Harvey (1989, p130 *et seq*).

The observation equation is first rewritten to make explicit the distinction between those parameters which vary over time (which we will denote γ_t) and those which are time invariant (which we denote by δ). The observation and state equations are then written as

$$\begin{aligned} \mathbf{y}_t &= \mathbf{Z}_t \gamma_t + \mathbf{X}_t \delta + \epsilon_t \\ \gamma_t &= \eta_t \end{aligned} \quad (30)$$

where now the definitions of some of the matrices have changed slightly. In particular, $\mathbf{z}_t' = \mathbf{r}_{m,t}$, $\mathbf{X}_t$ is an $2N \times N$ block diagonal matrix with $[I \ \mathbf{r}_{m,t}]$ on the diagonal, γ_t contains just the time-varying elements of the parameter vector, and δ is obtained by stacking the security alphas and mean betas of all the securities. Also, Q is now just equivalent to Q_η .

Conditional on the hyperparameters (the variance-covariance matrices) the GLS estimator of δ is given by

$$\tilde{\delta} = \left[\sum_{t=1}^T \mathbf{X}_t^*{}' F_t^{-1} \mathbf{X}_t^* \right]^{-1} \sum_{t=1}^T \mathbf{X}_t^*{}' F_t^{-1} \mathbf{y}_t^* \quad (31)$$

where the variables superscripted with asterisks are the ‘innovations’ from applying the Kalman filter to $\mathbf{y}_t$ and each column of $\mathbf{X}_t$. Note that the same Kalman filter is applied each time and

so the recursion for $\mathbf{P}_{t|t-1}$ is run only once, with only the values of $\mathbf{a}_t$ changing each time the filter is applied.

The residuals ν_t are calculated from

$$\nu_t = y_t^* - \mathbf{X}_t^* \tilde{\delta} \quad (32)$$

and are used to construct the concentrated likelihood function. Further simplification can be achieved by concentrating out a variance component from $\mathbf{Q}_\eta$ which is denoted σ_η^2 . $\mathbf{Q}_\eta$ still denotes the variance-covariance matrix although it must now be understood that it is a matrix of relative covariances. The elements of the $\mathbf{H}$ matrix are also specified relative to unity with the upper left component fixed at that value to provide a reference point. The maximum likelihood estimate of σ_η^2 is given by

$$\tilde{\sigma}_*^2 = \frac{1}{TN} \sum_{t=1}^T \nu_t' \mathbf{F}_t^{-1} \nu_t \quad (33)$$

and the fully concentrated likelihood is

$$\begin{aligned} \log L &= -\frac{TN}{2} \log \tilde{\sigma}_*^2 - \frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t| - \frac{1}{2\tilde{\sigma}_*^2} \sum_{t=1}^T \nu_t' \mathbf{F}_t^{-1} \nu_t \\ &= -\frac{TN}{2} \left[1 + \log \sum_{t=1}^T \nu_t' \mathbf{F}_t^{-1} \nu_t \right] - \frac{1}{2} \sum_{t=1}^T \log |\mathbf{F}_t| \end{aligned} \quad (34)$$

The algorithm for fitting the model is

- [1] Find suitable starting values for the hyperparameters $\mathbf{H}$ and $\mathbf{Q}_\eta$, i.e. the error covariance matrices for the observed returns and the stochastic elements of the betas, respectively;
- [2] Set $\mathbf{a}_0 = \mathbf{0}$ and $\mathbf{P}_0 = \mathbf{Q}_\eta$, as suggested by Harvey (1989, p121);
- [3] Run the Kalman filter on the observations and columns of the $\mathbf{X}_t$ matrix to obtain the asterisked values;
- [4] Estimate δ (the base values of the parameters) using GLS;
- [5] Estimate σ_η^2 (the standard deviation of the first security) using ML;
- [6] Calculate the likelihood function;
- [7] Adjust the values of the hyperparameters and repeat steps [2] to [6] until the likelihood is maximized.

Appendix 9.B

Influence function for the contaminated normal assumption

In this appendix the influence function of the contaminated normal distribution and its derivative matrix are described. The approximate maximum likelihood function is also given. The derivation of the results is fairly straight-forward given the definition of the score function and fact that the contaminated normal density function is equivalent to the weighted sum of the density functions of two normal distributions. The reader is referred to Martin (1979, 1981) for equivalent results in the univariate case.

The influence function is given by

$$g_i(\nu_i) = (1 - b^{-1}(\nu_i)) F_{1,i}^{-1} \nu_i + b^{-1}(\nu_i) F_{2,i}^{-1} \nu_i \quad (35)$$

where

$$F_{\kappa,i} = Z_i P_{i|i-1} Z_i' + \sigma_\kappa^2 H \quad (36)$$

and

$$b(\nu_i) = 1 + \frac{1-\zeta}{\zeta} \left\{ \frac{|F_{2,i}|}{|F_{1,i}|} \right\}^{\frac{1}{2}} \times \exp\left\{-\frac{1}{2} \nu_i' (F_{1,i}^{-1} - F_{2,i}^{-1}) \nu_i\right\} \quad (37)$$

(In equation (36), $\kappa = 1$ refers to the distribution of the 'good' observations; $\kappa = 2$ refers to the distribution of the outlying observations.)

The derivative matrix is obtainable as

$$G_i(\nu_i) = (1 - b^{-1}(\nu_i)) F_{1,i}^{-1} + b^{-1}(\nu_i) F_{2,i}^{-1} + \frac{1 - b(\nu_i)}{b^2(\nu_i)} (F_{1,i}^{-1} - F_{2,i}^{-1}) \nu_i \nu_i' (F_{1,i}^{-1} - F_{2,i}^{-1}) \quad (38)$$

The function maximized to give the approximate MLEs of Q and H is given by

$$L = -\frac{NT}{2} \log 2\pi + \log \zeta - \frac{1}{2} \sum \nu_i' F_{2,i}^{-1} \nu_i - \frac{1}{2} \sum |F_{2,i}| + \sum \log b(\nu_i) \quad (39)$$

CHAPTER 10

SUMMARY AND CONCLUSIONS

The estimation of systematic risk and other financial risk parameters has long been the subject of empirical research on the large and relatively well traded stock markets of the world. The motivation for this thesis arose from the realization that beta estimation problems had not been adequately solved in the small market context. The aim of the thesis has therefore been to address the problems which characterize small markets so that concrete recommendations can be made to improve the estimates of systematic risk in these markets.

In particular, one of the practical issues which caught the attention of researchers in the '70s, was the problem that price adjustment delays could lead to biased estimates of systematic risk. As a consequence, several estimators which corrected for the biases emerged and were tested on the major international markets such as the New York and London Stock Exchanges. It has been argued in the thesis, however, that the 'thin trading' of the NYSE, for instance, is of a completely different magnitude to that of the smaller, less well traded markets of the world, e.g. the Johannesburg Stock Exchange. It has been essential, therefore, to examine here the adaptations required to the correction procedures (which were developed on the larger, more well traded exchanges) in the light of the vastly different levels of thin trading.¹

Another characteristic of stock market data which has been identified in this thesis as an important issue is the non-normality of the security return distributions. The assumption of normality has long been questioned on the well traded markets, but it has been shown in the literature (and indeed in this thesis) that it is the smaller, less well traded securities on the exchanges which exhibit the most pronounced deviations from normality. As it is also well known that the mis-specification of the residual and return distributions leads to inefficient parameter estimation when traditional least squares methods are used, it is clear that the effects of non-normality need to be countered.

This thesis has therefore tackled the two problems associated with estimating systematic risk on small stock markets, i.e. thin trading and non-normality. The investigations undertaken in the thesis have provided several practical conclusions for estimating betas and covariances.

¹ In the concluding remarks of his doctoral dissertation, Bradfield (1989) in fact suggested that 'the improvement of beta estimation and covariance estimation would be an important contribution to analysts and investors who are active in thinly traded environments'.

More specifically, the simulation and empirical results of Chapters 4 and 5 have indicated that the *trade-to-trade* approach to avoiding thin trading biases is far more efficient than the aggregated coefficients procedures. Furthermore this superior efficiency persists even when weekly data have to be substituted for daily data (because of possible data limitations) .

The modifications to the trade-to-trade estimator proposed in Chapter 7 (i.e. the adaptive selection of the days in the month so as to accentuate the linear relationship between the security and the market returns) indicated that the near continuous set of pricing information *within* the calendar months should be utilized to make substantial gains in efficiency. Based on the empirical findings the adaptive selection of return interval in conjunction with the trade-to-trade estimator is strongly recommended.

A further modification to the trade-to-trade estimator which proved itself in the empirical assessments of efficiency described in Chapter 5, was the Bayesian adjustment suggested by Vasicek (1973). The Bayesian adjustment is designed to counteract the so-called regression bias which is inevitable when using the market model (or any other linear model for that matter) to estimate systematic risk.

Chapter 6 in this thesis contained a major study of robust regression methods. The least squares criterion is the most commonly applied in estimating the parameters of the market model. However, as indicated earlier, least squares can lead to very inefficient estimates of beta coefficients because the returns on the securities *and* the market index are non-normal. A broad cross-section of estimators were thoroughly tested and a *bounded influence* version of the L_p -norm estimators was shown to be the most efficient, as well as providing the best 'fit' to the data. It was shown that 'bounding the influence' of outlying market returns was as important as down-weighting the effect of outlying residual returns. It is therefore recommended that the bounded influence L_p -norm regression technique be used as the estimation methodology in conjunction with the adaptively selected, trade-to-trade return data.

It is apparent from the above summary that the final recommendations for practical beta estimation on small stock markets are therefore a hierarchy of modifications to the trade-to-trade approach, with each additional step contributing significantly to improved precision and efficiency in the resultant beta estimate.

Unlike the estimation of systematic risk, the estimation of covariances has not been as intensively researched. As demonstrated in Chapter 8 of the thesis, however, thin trading biases the estimation of covariances in the same way as it does the estimation of betas. It has been argued that the trade-to-trade approach is not readily applied to covariance estimation because

both securities may be very thinly traded and so there will be very few days on which both securities are traded. Accordingly it appears more apt to use an aggregated coefficients method for covariance estimation in thinly traded markets. The first part of an investigation into the estimation of covariances (described in Chapter 8 of this thesis) was aimed at finding a simple adaptive procedure for selecting the appropriate number of leading and lagged coefficients for a Cohen-type covariance estimator, based on the trading frequencies of the two securities.

The other major contribution of Chapter 8 was the examination of the biases caused in covariance estimation by return autocorrelation. In particular, the case where the market returns (and hence also the individual security returns via the market model) were serially correlated was considered. It was shown that any market autocorrelation (in addition to that induced by thin trading in the constituent components of the index) would result in severe upward biases in estimates of covariance (and also beta estimates, if aggregated coefficients were used). In order to correct for the upward bias, a series of adjustment factors were derived. The results were extended using a different theoretical framework (namely, continuous-time time series) so that similar adjustment factors could be applied to estimates from the more well traded exchanges (where thin trading is generally limited to delays of less than one day in length).

The important issue of the *stability* of beta coefficients on the JSE was assessed in Chapters 5 and 6. The empirical findings from the JSE were shown to correspond with those from the overseas markets, i.e. it was found that betas were only moderately stable. In order to incorporate the movements of the betas into the estimation process, the estimation techniques appropriate for time-varying parameters were investigated. Consequently, in Chapter 9, a robust multivariate state-space approach to estimating security betas was developed and the important new insights which could be gained from such a conceptual model were highlighted. As a first step towards measuring the effectiveness of the multivariate model, the correlations between the movements in the betas of the major sectors of the JSE were estimated. The magnitudes of the correlations were shown to be statistically highly significant and to hold important implications for portfolio management: for example, if a stable portfolio beta is desired then investments have to be spread over the sectors of the market which have negatively correlated beta coefficients. This new dimension to portfolio management will, it is hoped, form the basis for fresh research in the area.

The primary objective of this thesis was to offer firm recommendations for practitioners wanting to estimate security beta coefficients. In keeping with this objective, the following

section details the main findings and steps required to estimate systematic risk efficiently and accurately on small stock markets.

Recommendations for beta estimation

The main recommendations emerging from the thesis for improving estimates of beta coefficients on a small stock exchange are:

1. the use of trade-to-trade returns;
2. the use of a robust estimation procedure;
3. the use of the extreme value adaptive interval selection procedure;
4. the use of the Bayesian adjustment procedure.

The evidence presented in this thesis has indicated that the above four recommendations should be implemented jointly in the routine estimation of systematic risk on small stock markets for maximum improvement. However, it is recognized that the appropriate technology for implementing all the recommendations may not always be easily available to practitioners. It should, therefore, be noted that in most cases the implementation of any subset of the procedures will improve the resulting beta estimates substantially over naive OLS estimates. The ensuing discussion gives some insight into the choice of the recommended procedures.

In any study involving a cross-section of beta estimates from a thinly traded stock exchange it is imperative that a thin trading correction technique be used. In particular the use of **trade-to-trade** returns described in section 3.2.1 was demonstrated to be the most efficient of the approaches to correcting the biases, removing the downward biases of up to 60% in OLS estimates of beta without any loss in efficiency. Data requirements for the trade-to-trade approach are the price and date of the last trade of the security in each month, as well as the value of the market index on those dates.

The empirical analysis in Chapter 6 indicated that **robust estimation** techniques (briefly described in section 6.3.1) improved the reliability and efficiency of the estimation process by nearly 50%. The *bounded influence*, L_p -norm estimator, in particular, showed substantial gains in efficiency, especially for securities which had non-normally distributed return series. It should be emphasized that in cases where the securities are thinly traded, it is essential to apply the robust estimator to *trade-to-trade* returns in order to avoid the downward biases caused by infrequent trading. Of the four main recommendations, the implementation of the robust

estimation procedure is possibly the most difficult, requiring, as it does, sophisticated software and hardware.²

The **extreme value adaptive interval selection** procedure detailed in section 7.3.1 is an easy adaptation of the trade-to-trade procedure. The empirical tests in Chapter 7 demonstrated that improvements in the standard errors of the estimates of beta of around 30% could be attained. The additional data requirements for the extreme value procedure (over and above the data requirements for the trade-to-trade approach) are given in section 7.3.4.

The **Bayesian adjustment** to estimates of beta is outlined in section 5.3.3. Empirical testing of the Bayesian adjustment in Chapter 5 showed it to remove much of the regression bias in estimates of beta and to improve forecasting power by nearly 20%. It is very easy to incorporate into routine beta estimation, requiring only an estimate of the average beta in the market and the standard deviation of the estimated betas.

The above discussion has dealt with each of the recommended steps individually, each of which would provide some improvements to estimates of beta on their own. What follows is an algorithmic description of the 'ideal' procedure, which incorporates all four recommendations for routine beta estimation on a small stock market.

The first major step involves measuring the return series of the security and market. In order to select the intervals over which to measure the return series in such a way as to avoid biases arising from thin trading and to enhance the efficiency of the estimation process by 'spreading' the market returns, the trade-to-trade and extreme value approaches have to be combined in the manner described in Chapter 7. Specifically, in the first phase of the interval selection process the dates on which shares in the security were traded in each month of a five-year period need to be identified. The second phase in the data selection process involves picking out from the traded dates those dates which are consistent with the extreme value approach. The ensuing procedure briefly describes a way of performing this second phase: from the traded dates identified in the first month, the date corresponding to the lowest market index value is selected; from the traded dates in the second month, the date matching the highest market index value is selected; in the third month, the traded date corresponding to the

² Although a particular bounded influence, robust estimator has been recommended here, other robust estimators also provide significant improvements in efficiency over OLS. Practitioners finding it difficult to implement the bounded influence L_p -norm estimator recommended here are strongly advised to try one of the other robust estimators, preferably a bounded influence one. The results in the final column of Table 6.4 may help in assessing the performance of other robust estimators which the practitioner might find easier to implement.

lowest market index value is selected, etc. The foregoing procedure will result in a series of intervals over which the returns on the security and market should be measured.

The next major step involves regressing the set of security returns obtained from the above procedure against the corresponding market returns in order to estimate the beta coefficient. The empirical evidence presented in Chapter 6 indicated that a robust, bounded influence estimator should be implemented. In particular, the results in Chapter 6 indicated that the L_p -norm estimator is the most efficient estimator in light of the non-normality of the return distributions of most securities.

The third and final recommended step in the estimation of beta coefficients is to apply the Bayesian adjustment discussed in Chapter 5. The adjustment corrects for the regression bias inherent in estimates of beta and thereby substantially improves the efficiency of the estimation.

It should be noted that the resulting estimate of beta is of the *historical* beta of the security in the five years prior to the estimation date. The evidence from the JSE and other markets indicates that securities generally have stable beta coefficients. Therefore the historical estimates will be good indicators of future beta coefficients in most cases. On the other hand it should be emphasized that if the firm undergoes major operational or financial restructuring (a process which is usually announced to interested parties) then the historical beta will not necessarily be a good guide to the future beta coefficient.

REFERENCES

- Affleck-Graves, J.F., 1974, Portfolio Selection on the Johannesburg Stock Exchange, *MSc Thesis*, University of Cape Town
- Affleck-Graves, J.F. and B. McDonald, 1989, Nonnormalities and Tests of Asset Pricing Theories, *Journal of Finance*, Vol 44, No 4, 889-908
- Affleck-Graves, J.F. and D.J. Bradfield, 1992, An Examination of the Power of Univariate Tests of the CAPM: A Simulation Approach, *Journal of Economics and Business* (in press)
- Alexander, G. and G. Benson, 1982, More on Beta as a Random Coefficient, *Journal of Financial and Quantitative Analysis*, Vol 17, 27-36
- Altman, E.I., B. Jacquillat and M. Levasseur, 1974, Comparative Analysis of Risk Measures: France and the United States, *Journal of Finance*, Vol 29, 1495-1511
- Anscombe, F.J. and W.J. Glynn, 1983, Distribution of the Kurtosis Statistic b_2 for Normal Samples, *Biometrika*, Vol 70, 227-234
- Atchison, M.D., Butler, K.C. and R.R. Simonds, 1987, Nonsynchronous Security Trading and Market Index Autocorrelation, *Journal of Finance*, Vol 42, 111-118
- Bachelier, L., 1900, *Theorie de la Speculation*, Paris: Gauthier-Villars, Reprinted in *The Random Character of Stock Market Prices*, (ed. P.H. Cootner), 1964, MIT Press, Cambridge
- Ball, R., 1977, A Note on Errors in Variables and Estimates of Systematic Risk, *Australian Journal of Management*, Vol 2, 79-84
- Barr, G.D.I, 1980, A Contribution to Adaptive Robust Estimation, *PhD Thesis*, University of Cape Town
- Barrodale, I. and F.D.K. Roberts, 1973, An Improved Algorithm for Discrete L_1 Linear Approximation, *SIAM Journal of Numerical Analysis*, Vol 10, 839-848
- Basset, G.W. and R.W. Koenker, 1982, An Empirical Quantile Function for Linear Models with i.i.d. Errors, *Journal of the American Statistical Association*, Vol 77, 407-415
- Beaver, W., P. Kettler and M. Scholes, 1970, The Association between Market Determined and Accounting Determined Risk Measures, *Accounting Review*, Vol 45, 654-682
- Belsey, D.A., Kuh, E. and R.E. Welsch, 1981, *Regression Diagnostics*, Wiley
- Berglund, T., Liljeblom, E. and A. Löflund, 1989, Estimating Betas on Daily Data for a Small Stock Market, *Journal of Banking and Finance*, Vol 13, 41-64
- Black, F., 1972, Capital Market Equilibrium with Restricted Borrowing, *Journal of Business*, Vol 45, 444-455
- Black, F., Jensen, M.C. and M. Scholes, 1972, The Capital Asset Pricing Model: Some Empirical Tests, in *Studies in the Theory of Capital Markets* (ed. M.C. Jensen), Praeger Publishers Inc., New York
- Blattberg, R.C. and N.J. Gonedes, 1974, A Comparison of the Stable and Students Distributions as Statistical Models for Stock Prices, *Journal of Business*, Vol 47, 244-280
- Blattberg, R.C. and T. Sargent, 1971, Regression with Non-Gaussian Disturbances: Some Sampling Results, *Econometrica*, Vol 39, 501-510
- Blume, M.E., 1970, Portfolio Theory: A Step Towards its Practical Application, *Journal of Business*, Vol 43, 152-173
- Blume, M.E., 1971, On the Assessment of Risk, *Journal of Finance*, Vol 26, 1-10
- Blume, M.E., 1975, Betas and their Regression Tendencies, *Journal of Finance*, Vol 30, 785-795
- Blume, M.E. and I. Friend, 1973, A New Look at the Capital Pricing Model, *Journal of Finance*, Vol 28, 19-33
- Booth, G.G., Hatem, J., Virtanen, I. and P. Yli-Olli, 1992, Stochastic Modelling of Security Returns: Evidence from the Helsinki Stock Exchange, *European Journal of Operational Research*, Vol 56, 98-106
- Bos, T., T.A. Fetherston, T. Martikainen and J. Perttunen, 1993, The International Co-Movements of Finnish Stocks, *Working Paper*, University of Vaasa, Finland
- Bos, T. and P. Newbold, 1984, An Empirical Investigation of the Possibility of Stochastic Systematic Risk in the Market Model, *Journal of Business*, Vol 57, 35-41
- Bowie, D.C. and D.J. Bradfield, 1993a, Improved beta estimation on the JSE: A Simulation study, *South African Journal of Business Management*, Vol 24, 118-123

- Bowie, D.C. and D.J. Bradfield, 1993b, A Review of Systematic Risk Estimation on the JSE, *De Ratione*, Vol 7, 6-22
- Bowman, K.O. and B.R. Shenton, 1975, Omnibus Test Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2 , *Biometrika*, Vol 62, 243-250
- Box, G.E.P. and G.M. Jenkins, 1976, *Time Series Analysis: Forecasting and Control*, revised edn., San Francisco: Holden-Day
- Box, G.E.P. and G.C. Tiao, 1973, *Bayesian Inferences in Statistical Analysis*, Addison-Wesley Publishing, Reading, MA
- Bradfield, D.J., 1989, Applications and Extensions of Financial Models to Small Markets: The South African Case, *PhD Thesis*, University of Cape Town
- Bradfield, D.J., 1990, A Note on the Estimation Problems Caused by Thin Trading on the JSE, *De Ratione*, Vol 3, 22-25
- Bradfield, D.J. and J.F. Affleck-Graves, 1991, Multivariate Tests of the CAPM: South African Evidence, *South African Statistical Journal*, Vol 25, 19-44
- Bradfield, D.J., J.F. Affleck-Graves and G.D.I. Barr, 1982, Stability Tests for Alphas and Betas over Bull and Bear Market Conditions on the JSE, Technical Report, University of Cape Town
- Bradfield, D.J. and D.C. Bowie, The Financial Risk Service (a quarterly subscription service), University of Cape Town
- Bradfield, D.J. and D.C. Bowie, 1991, On Estimating the Risk that Shareholders Bear During Hostile Merger Activity, *South African Journal of Business Management*, Vol 22, 83-86
- Bradfield, D.J. and G.D.I. Barr, 1989, Risk Estimation in the Thinly-Traded JSE Environment, *South African Journal of Business Management*, Vol 20, 169-173
- Brealey, R. and S. Myers, 1991, *Principles of Corporate Finance*, McGraw-Hill, New York.
- Breusch, T.S. and A.R. Pagan, 1979, A Simple Test for Heteroscedasticity and Random Coefficient Variation, *Econometrica*, Vol 47, 1287-1294
- Brooks, R.D., 1991b, Hildreth-Houck or Return to Normalcy Random Coefficients Regression Coefficients: Which is the Most Appropriate Testing Alternative, *mimeo*, Monash University
- Brooks, R.D. and M.L. King, 1991, Testing Hildreth-Houck Against Return to Normalcy Random Regression Coefficients, *mimeo*, Monash University
- Brooks, R.D., Faff, R.W. and J.H.H. Lee, 1992, The Form of Time Variation of Systematic Risk: Some Australian Evidence, *Applied Financial Economics*, Vol 2, 191-198
- Brooks, R.D., 1991a, Alternative Point Optimal Tests for Regression Coefficient Stability, *Working Paper 10/91*, Department of Econometrics, Monash University
- Brown, R.L., Durbin, J. and J.M. Evans, 1975, Techniques for Testing the Constancy of Regression Relationships over Time, *Journal of the Royal Statistical Society (series B)*, Vol 37, 149-192
- Brown, S. and M. Weinstein, 1983, A new approach to testing asset pricing models: The bilinear paradigm. *Journal of Finance*, 38, 711-743.
- Carlin, B.P., Polson, N.G. and D.S. Stoffer, 1992, A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modelling, *Journal of the American Statistical Association*, Vol 87, 493-500
- Carroll, C., Thistle, P.D. and K.C.J. Wei, 1992, The Robustness of Risk-Return Nonlinearities to the Normality Assumption, *Journal of Financial and Quantitative Analysis*, Vol 27, 419-435
- Carter, K.J., 1983, The Estimation of Security Beta Coefficients on the Johannesburg Stock Exchange, *PhD Thesis*, University of Cape Town
- Chan, K.C., 1985, Market Value Changes and Time-Varying Risk, *Dissertation* submitted to the Faculty of the Graduate School of Business, University of Chicago
- Chan, L.K.C. and J. Lakonishok, 1992, Robust Measurement of Beta Risk, *Journal of Financial and Quantitative Analysis*, vol 27, 265-282
- Chang, W.-C. and D.E. Weiss, 1991, An Examination of the Time Series Properties of Beta in the Market Model, *Journal of the American Statistical Association*, Vol 86, 883-890
- Chen, N., 1983, Some empirical tests of the theory of Arbitrage Pricing, *Journal of Finance*, 38, 1393-1414.
- Chow, G.C., 1960, Test of Equality between Subsets of Coefficients in Two Linear Regressions, *Econometrica*, Vol 28, 591-605
- Chow, G.C., 1983, *Econometrics*, Wiley, New York

- Cohen, K.J., Hawawini, G.A., Maier, S.F., Schwartz, R.A. and D.K. Whitcomb, 1980, Implications of Microstructure Theory for Empirical Research on Stock Price Behaviour, *Journal of Finance*, Vol 35, 249-257
- Cohen, K.J., G.A. Hawawini, S.F. Maier, R.A. Schwartz and D.K. Whitcomb, 1983, Friction in the Trading Process and the Estimation of Systematic Risk, *Journal of Financial Economics*, Vol 12, 263-278
- Collins, D.W., Ledolter, J. and J. Rayburn, 1987, Some Further Evidence on the Stochastic Properties of Systematic Risk, *Journal of Business*, Vol 60, 425-448
- Copeland, T. and F. Weston, 1988, *Financial Theory and Corporate Policy*, Addison-Wesley Publishing Company, Massachusetts
- Corhay, A., 1992, The Intervalling Effect Bias in Beta: A Note, *Journal of Banking and Finance*, Vol 16, 61-73
- Cornell, B. and J.K. Dietrich, 1978, Mean-Absolute-Deviation Versus Least-Squares Regression Estimation of Beta Coefficients, *Journal of Financial and Quantitative Analysis*, Vol 13, 123-131
- Coursey, D. and H. Nyquist, 1983, On Least Absolute Error Estimation with Linear Regression Models with Dependent Stable Residuals, *Review of Economics and Statistics*, Vol 65, 687-692
- Cox, D.R. and D.V. Hinkley, 1974, *Theoretical Statistics*, Chapman and Hall, London
- Cramer, H., 1946, *Mathematical Methods of Statistics*, Princeton University press, Princeton
- D'Agostino, R.B., 1970, Transformations to Normality of the Null Distribution of g_1 , *Biometrika*, Vol 57, 679-681
- D'Agostino, R.B. and M.A. Stephens (editors), 1986, *Goodness-of-fit Techniques*, New York and Basel: Marcel Dekker
- D'Agostino, R.B. and E.S. Pearson, 1973, Testing for Departures from Normality I. Fuller Empirical Results for the Distribution of b_2 and $\sqrt{b_1}$, *Biometrika*, Vol 60, 613-622
- De Jong, F., Kemna, A. and T. Kloek, 1992, A Contribution to Event Study Methodology with an Application to the Dutch Stock Market, *Journal of Banking and Finance*, Vol 16, 11-36
- De Jongh, P.J., De Wet, T. and A.H. Welsh, 1988, Mallows-Type Bounded Influence Regression Trimmed Means, *Journal of the American Statistical Association*, Vol 83, 805-810
- Dimson, E., 1979, Risk Measurement when Shares are Subject to Infrequent Trading, *Journal of Financial Economics*, Vol 7, 197-206
- Dimson, E., 1982, Friction in the Trading Process and Risk Measurement: Response, *Working Paper*, London Business School
- Dimson, E. and P.R. Marsh, 1983, The Stability of UK Risk Measures and The Problem of Thin Trading, *Journal of Finance*, Vol 38, 753-783
- Easton, G.S., 1991, Compromise Maximum Likelihood Estimators of Location, *Journal of the American Statistical Association*, Vol 86, 1051-1064
- Efron, B., 1982, *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM, Philadelphia
- Ekblom, H., 1974, L_p -Methods for Robust Regression, *BIT*, Vol 14, 22-32
- Eubank, A.A. and J.K. Zumwalt, 1979, An Analysis of the Forecast Error Impact of Alternative Beta Adjustment Techniques and Risk Classes, *Journal of Finance*, Vol 34, No 3, 761-776
- Fabozzi, F.J. and J.C. Francis, 1977, Stability Tests for Alphas and Betas over Bull and Bear Market Conditions, *Journal of Finance*, Vol 32, 1093-1099
- Fabozzi, F.J. and J.C. Francis, 1978, Beta as a Random Coefficient, *Journal of Financial and Quantitative Analysis*, Vol 13, 101-116
- Faff, R.W., Lee, J.H.H. and T.R.L. Fry, 1992, Time Stationarity of Systematic Risk: Some Australian Evidence, *Journal of Business Finance and Accounting*, Vol 19, 253-270
- Fama, E.F., 1965, The Behaviour of Stock-Market Prices, *Journal of Business*, Vol 38, 34-105
- Fama, E.F., 1976, *Foundations of Finance*, Basic Books Inc., New York
- Fama, E.F. and J.D. MacBeth, 1973, Risk, Return and Equilibrium: Empirical Tests, *Journal of Political Economy*, Vol 81, 607-636
- Fama, E.F. and K.R. French, 1992, The Cross-section of Expected Stock Returns, *Journal of Finance*, Vol 67, 427-465
- Fisher, F.M., 1970, Tests of Equality between Sets of Coefficients in Two Linear Regressions: An Expository Note, *Econometrics*, Vol 38, 361-366
- Fisher, L., 1966, Some New Stock Market Indexes, *Journal of Business*, Vol 39, 191-225

- Fleishman, A.I., 1978, A Method for Simulating Non-normal Distributions, *Psychometrika*, Vol 43, 521-532
- Forsythe, A.B., 1972, Robust Estimation of Straight Line Regression Coefficients by Minimizing p th Power Deviations, *Technometrics*, Vol 14, 181-186
- Fowler, D.J. and C.H. Rorke, 1983, Risk Measurement when Shares are Subject to Infrequent Trading: Comment, *Journal of Financial Economics*, Vol 12, 279-283
- Francis, J.C. and F.J. Fabozzi, 1979, The Effects of Changing Macro-economic Conditions on the Parameters of the Single Index Model, *Journal of Financial and Quantitative Analysis*, Vol 14, 351-360
- Franks, J.R., J.E. Broyles and M.J. Hecht, 1977, An Industry Study of the Profitability of Mergers in the United Kingdom, *Journal of Finance*, Vol 32, 1513-1525
- Gibbons, M., 1982, Multivariate Tests of Financial Models: A New Approach. *Journal of Financial Economics*, 10, 3-27.
- Giltinan, D.M., Carroll, R.J. and D. Ruppert, 1986, Some New Estimation Methods for Weighted Regression When There are Possible Outliers, *Technometrics*, Vol 28, 219-230
- Gnedenko, B.V. and A.N. Kolmogorov, 1954, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Massachusetts
- Gonin, R. and A.H. Money, 1989, *Nonlinear L_p -norm Estimation*, Marcel Dekker Inc., New York and Basel
- Hall, S., Miles, D. and M.P. Taylor, 1989, Modelling Asset Prices with Time-varying Betas. Some Evidence from the London Stock Exchange, *Manchester School of Economics and Social Studies*, Vol 2, 225-240
- Hampel, F.R., 1974, The Influence Curve and its Role in Robust Estimation, *Journal of the American Statistical Association*, Vol 69, 383-393
- Handa, P., Kothari, S.P. and C.E. Wasley, 1989, The Relation between Return Interval and Betas: Implications for the Size Effect, *Journal of Financial Economics*, Vol 23, 79-100
- Handa, P., Kothari, S.P. and C.E. Wasley, 1993, Sensitivity of Multivariate Tests of the Capital Asset Pricing Model to the Return Measurement Interval, *Journal of Finance*, Vol 48, 1543-1551
- Harrison, P.J. and C.F. Stevens, 1976, Bayesian Forecasting, *Journal of the Royal Statistical Society (series B)*, Vol 38, 205-247
- Harter, H.L., 1977, The Nonuniqueness of Absolute Value Regression, *Communications in Statistics - Simulation and Computation*, Vol 6, 829-838
- Harvey, A.C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge
- Hawawini, G.A. and P.A. Michel, 1979, An Assessment of Risk in Thinner Markets: The Belgian Case, *Journal of Economics and Business*, Vol 31, 196-201
- Hays, P.A. and D.E. Upton, 1986, A Shifting Regimes Approach to the Stationarity of the Market Model Parameters of Individual Securities, *Journal of Financial and Quantitative Analysis*, Vol 21, 307-321
- Hildreth, C. and J.P. Houck, 1968, Some Estimators for a Linear Model with Random Coefficients, *Journal of the American Statistical Association*, Vol 63, 584-595
- Huber, P.J., 1981, *Robust Statistics*, John Wiley, New York
- Hui, T-K., Kwan, K-C. and K-L. Lim, 1990, Beta Estimation for Thinly Traded Shares: A Bootstrap Approach, *OMEGA International Journal of Management Science*, Vol 18, 329-333
- Ibbotson, F.G., 1975, Price Performance of Common Stock New Issues, *Journal of Financial Economics*, Vol 3, 235-272
- Jacob, N.L. and R.R. Pettit, 1984, *Investments*, Richard D. Irwin Inc., Homewood, Illinois
- Jobson, J.D. and R. Korkie, 1982, Potential Performance and Tests of Portfolio Efficiency, *Journal of Financial Economics*, 10, 433-466.
- Jones, R.H., 1986, Time Series Regression with Unequally Spaced Data (in *Essays in Time Series and Allied Processes*, Gani, J. and M.B. Priestly (eds)), The Universities Press, Belfast
- Kalman, R.E., 1960, A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering*, Vol 82, 269-294
- King, M.L., 1987, An Alternative Test for Regression Coefficient Stability, *Review of Economics and Statistics*, Vol 69, 379-381
- Kitagawa, G., 1987, Non-Gaussian State-Space Modelling of Nonstationary Time Series (with discussion), *Journal of the American Statistical Association*, Vol 82, 1032-1063

- Klemkosky, R.C. and J.D. Martin, 1975, The Adjustment of Beta Forecasts, *Journal of Finance*, Vol 30, 1123-1128
- Knif, J., 1989, Parameter Variability in the Single Factor Market Model, An Empirical Comparison of Tests and Estimation Procedures Using Data from the Helsinki Stock Exchange, *Commentationes Scientiarum Socialium* Vol 40, Societas Scientiarum Fennica
- Koenker, R.W., 1982, Robust Methods in Econometrics, *Econometric Reviews*, Vol 1, 213-255
- Koenker, R.W. and G.W. Basset, 1978, Regression Quantiles, *Econometrica*, Vol 46, 33-50
- Koenker, R.W. and S. Portnoy, 1987, L-Estimation for Linear Models, *Journal of the American Statistical Association*, Vol 82, 851-857
- Kon, S.J., 1984, Models of Stock Returns - A Comparison, *Journal of Finance*, Vol 39, 147-165
- Kon, S.J. and F.C. Jen, 1978, Estimation of Time Varying Systematic Risk Statistics, *Journal of Business Research*, Vol 8, 263-275
- Kotz, S. and N.L. Johnson (eds), *Encyclopedia of Statistical Sciences*, Wiley-Interscience, New York
- Krasker, W.S. and R.E. Welsch, 1982, Efficient Bounded Influence Regression Estimation, *Journal of the American Statistical Association*, Vol 77, 595-604
- Lakonishok, J. and A.C. Shapiro, 1986, Systematic Risk, Total Risk and Size as Determinants of Stock Market Returns, *Journal of Banking and Finance*, Vol 10, 115-132
- Levy, P., 1925, *Calcul des Probabilités*, Gauthier-Villars, Paris
- Lintner, J., 1965, The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, *Review of Economics and Statistics*, Vol 47, 13-37
- Luoma, M., T. Martikainen and J. Perttunen, 1993a, Adjusting for the Interval Effect Bias in Beta Coefficients on a Thin Security Market: Application of a Lag Distribution Model, *International Journal of Systems Science*, Vol 24, 2391-2398
- Luoma, M., T. Martikainen and J. Perttunen, 1993b, Thin Trading and Estimation of Systematic Risk: An Application of an Error-correction Model, *Annals of Operations Research*, Vol 45, 297-305
- Luoma, M., T. Martikainen and J. Perttunen, 1993c, A Pseudo-Criterion for Security Betas in the Finnish Stock Market, *Working Paper*, University of Vaasa, Finland
- Malkamäki, M., 1992, Essays on Conditional Pricing of Finnish Stocks, *PhD Thesis*, Faculty of Accounting and Industrial Management, University of Vaasa, Finland
- Mallows, C.L., 1975, On Some Topics in Robustness, *unpublished memorandum*, Bell Telephone Laboratories, Murray Hill, New Jersey
- Mallows, C.L., 1973, Influence Functions, *unpublished paper presented at a conference on robust regression held at Cambridge, Massachusetts (sponsored by National Bureau for Economic Research)*
- Mandelbrot, B., 1963, The Variation of Certain Speculative Prices, *Journal of Business*, Vol 36, 394-419
- Mandelbrot, B., 1967, The Variation of Some Other Speculative Prices, *Journal Business*, Vol 40, 393-413
- Markowitz, H.M., 1984, The 'Two Beta' Trap, *Journal of Portfolio Management*, Vol 11, 12-20
- Markowitz, H.M., 1952, Portfolio Selection, *Journal of Finance*, Vol 7, 77-91
- Markowitz, H.M., 1959, *Portfolio Selection: Efficient Diversification of Investments*, Wiley and Sons, New York
- Marsh, P.R., 1979, Equity Rights and the Efficiency of the UK Stock Market, *Journal of Finance*, Vol 34, 839-863
- Martin, R.D., 1979, Approximate Conditional-Mean Type Smoothers and Interpolators, (in *Smoothing Techniques for Curve Estimation*, T. Gasser and M. Rosenblatt (eds)), Berlin: Springer-Verlag, 117-143
- Martin, R.D., 1981, Robust Methods for Time Series, (in *Applied Time Series Analysis*, Findley, D.F. (ed)), Academic Press, New York
- Masreliez, C.J., 1975, Approximate Non-Gaussian Filtering with Linear State and Observation Relations, *IEEE Transactions on Automatic Control*, AC-20, 107-110
- McDonald, J.B. and R.D. Nelson, 1989, Alternative Beta Estimation for the Market Model using Partially Adaptive Techniques, *Communications in Statistics - Theory and Methodology*, Vol 18, No 11, 4039-4058
- McDonald, J.B. and W.K. Newey, 1988, Partially Adaptive Estimation of Regression Models Via the Generalized T Distribution, *Econometric Theory*, Vol 4, 428-457

- McKean, J.W. and R.M. Schrader, 1987, Least Absolute Errors Analysis of Variance, (in *Statistical Data Analysis based on the L_1 -norm and Related Methods*, Dodge, Y. (ed)), North Holland, Amsterdam, 297-305
- Meinhold, R.J. and N.D. Singpurwalla, 1983, Understanding the Kalman Filter, *The American Statistician*, Vol 37, 123-127
- Meinhold, R.J. and N.D. Singpurwalla, 1989, Robustification of Kalman Filter Models, *Journal of the American Statistical Association*, Vol 84, 479-486
- Merton, R.C., 1973, An Intertemporal Capital Asset Pricing Model, *Econometrica*, vol 41, 867-887
- Money, A.H., Affleck-Graves, J.F., Hart, M.L. and G.D.I. Barr, 1982, The Linear Regression Model: L_p -norm Estimation and the Choice of p , *Communications in Statistics - Simulation and Computation*, Vol 11, 89-109
- Murray, M., 1992, A Multiprocess state space model for modelling the rate of return on ordinary stock, *South African Statistical Journal*, Vol 26, 95-120
- Ng, L., 1991, Tests of the CAPM with Time-Varying Covariances: A Multivariate GARCH Approach, *Journal of Finance*, 46, 1507-1521
- Nyquist, H., 1980, Recent Studies on L_p -norm Estimation, *PhD Thesis*, University of Umea, Sweden
- Nyquist, H., 1983, The Optimal L_p -norm Estimator in Linear Regression Models, *Communications in Statistics - Theory and Methods*, Vol 12, 2511-2524
- Ohlson, J. and B. Rosenberg, 1982, Systematic Risk of the CRSP Equal-Weighted Common Stock Index: A History Estimated by Stochastic Parameter Regression, *Journal of Business*, Vol 55, 121-145
- Olkin, I. and S. Yitzhaki, 1992, Gini Regression Analysis, *International Statistical Review*, Vol 60, No 2, 185-196
- Osborne, M.F.M., 1959, Brownian Motion in the Stock Market, *Operations Research*, Vol 7, 145-173
- Pearson, E.S., D'Agostino, R.B. and K.O. Bowman, 1977, Tests for Departure from Normality. Comparison of Powers, *Biometrika*, Vol 64, 231-246
- Pogue, G. and B. Solnik, 1974, The Market Model Applied to European Common Stocks: Some Empirical Results, *Journal of Financial and Quantitative Analysis*, Vol 9, 917-944
- Praetz, P.D., 1969, Australian Share Prices and the Random Walk Hypothesis, *Australian Journal of Statistics*, Vol 11, 123-139
- Praetz, P.D., 1972, The Distribution of Share Price Changes, *Journal of Business*, Vol 45, 49-55
- Press, S.J., 1967, A Compound Events Model for Security Prices, *Journal of Business*, Vol 40, 317-335
- Quandt, R.E., 1958, The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes, *Journal of the American Statistical Association*, Vol 53, 873-880
- Quandt, R.E., 1960, Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes, *Journal of the American Statistical Association*, Vol 55, 324-330
- Reinganum, M.R., 1981, Misspecification and Capital Asset Pricing: Empirical anomalies based on earnings yields and Market Values, *Journal of Financial Economics*, 9, 19-46
- Reinganum, M.R., 1981, A New Empirical Perspective on the CAPM, *Journal of Financial and Quantitative Analysis*, Vol 16, 439-462
- Roll, R., 1968, The Efficient Market Model Applied to US Treasury Bill Rates, *PhD Thesis*, GSB University of Chicago
- Roll, R., 1977, A Critique of the Asset Pricing Theory's Tests, *Journal of Financial Economics*, Vol 4, 129-176
- Roll, R. and S. Ross, 1980, An Empirical Investigation of the Arbitrage Pricing Theory, *Journal of Finance*, 35, 1073-1103
- Roll, R. and S. Ross, 1994, On the Cross-sectional Relation between Expected Returns and Betas, *Journal of Finance*, Vol 44, 101-121
- Rosenberg, B. and J. Guy, 1976a, Beta and Investment Fundamentals, *Financial Analysts Journal*, Vol 32, 60-72
- Rosenberg, B. and J. Guy, 1976b, Beta and Investment Fundamentals II, *Financial Analysts Journal*, Vol 32, 62-70
- Rosenberg, B. and V. Marathe, 1975, The Prediction of Investment Risk: Systematic Risk and Residual Risk, in *Proceedings: Seminar on the Analysis of Security Prices*, Centre for Research into Security Prices, Graduate School of Business, The University of Chicago
- Rosenberg, B., 1973, The Analysis of a Cross-section of Time Series by Stochastically Convergent Parameter Regression, *Annals of Economic and Social Measurement*, Vol 2, 367-386

- Rosenberg, B. and V. Marathe, 1976, Common Factors in Security Returns: Microeconomic Determinants and Macroeconomic Correlates, in *Proceedings: Seminar on the Analysis of Security Prices*, Centre for Research into Security Prices, Graduate School of Business, The University of Chicago
- Schlosberg, H.H., 1976, Distribution Models for Daily Log-Price Relatives, *MBA Research Report*, University of the Witwatersrand
- Scholes, M. and J. Williams, 1976, Estimating Betas from Daily Data, *Working Paper*, University of Chicago
- Scholes, M. and J. Williams, 1977, Estimating Beta from Non-Synchronous Data, *Journal of Financial Economics*, Vol 5, 309-327
- Scholz, F.W., 1977, Weighted Median Regression Estimates, *Institute of Mathematical Statistics Bulletin*, Vol 6, 44
- Schwert, G.W., 1977, Stock Exchange Seats as Capital Assets, *Journal of Financial Economics*, Vol 5, 51-78
- Scott, E. and S. Brown, 1980, Biased Estimators and Unstable Betas, *Journal of Finance*, Vol 35, 49-55
- Shalit, H. and S. Yitzhaki, 1984, Mean-Gini, Portfolio Theory and the Pricing of Risky Assets, *Journal of Finance*, Vol 39, 1449-1468
- Shanken, J., 1985, Multivariate tests of the zero-beta CAPM, *Journal of Financial Economics*, Vol 14, 327-348.
- Shanken, J., 1987, Nonsynchronous data and the covariance-factor structure of returns. *Journal of Finance*, Vol 42, 221-231.
- Sharpe, W.F., 1963, A Simplified Model for Portfolio Selection, *Management Science*, Vol 9, 277-293
- Sharpe, W.F., 1964, Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *Journal of Finance*, Vol 19, 425-442
- Sharpe, W.F., 1971, Mean-Absolute-Deviation Characteristic Lines for Securities and Portfolios, *Management Science*, Vol 18, 1-13
- Sharpe, W.F. and G.M. Cooper, 1972, Risk-Return Classes of New York Stock Exchange Common Stocks, 1931-1967, *Financial Analysts Journal*, Vol 28, 46-54
- Shively, T.S., 1988, An Analysis of Tests for Regression Coefficient Stability, *Journal of Econometrics*, Vol 39, 367-386
- Sichel, H.S., 1973, Statistical Valuation of Diamondiferous Deposits, *Journal of South African Institute of Mining and Metallurgy*, Vol 73, 235-243
- Sievers, G.L., 1978, Weighted Rank Statistics for Simple Linear Regression, *Journal of the American Statistical Association*, Vol 73, 628-631
- Simonds, R.R., LaMotte, L.R. and A. McWhorter, 1986, Testing for Nonstationarity of Market Risk: An Exact Test and Power Considerations, *Journal of Financial and Quantitative Analysis*, Vol 21, 209-220
- Sims, C.A., 1971, A Comparison of Maximum Likelihood Versus BLUE Estimators, *Review of Economics and Statistics*, Vol 53, 204-205
- Smith, A.F.M. and M. West, 1983, Monitoring Renal Transplants: An Application of the Multiprocess Kalman Filter, *Biometrics*, Vol 39, 267-271
- Smith, V.K. and T.W. Hall, 1972, A Comparison of Maximum Likelihood Versus BLUE Estimators, *Review of Economics and Statistics*, Vol 54, 186-190
- Sposito, V.A., Hand, M.L. and B. Skarpness, 1983, On the Efficiency of using the Sample Kurtosis in Selecting Optimal L_p Estimators, *Communications in Statistics - Theory and Methods*, Vol 12, 265-272
- Stambaugh, R., 1982, On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis, *Journal of Financial Economics*, Vol 10, 237-268.
- Stapleton, R.C. and M.G. Subrahmanyam, 1983, The Market Model and Capital Asset Pricing Theory: A Note, *Journal of Finance*, Vol 38, 1637-1642
- Sunder, S., 1980, Stationarity of Market Risk: Random Coefficients Tests for Individual Stocks, *Journal of Finance*, Vol 35, 883-896
- Teichmoeller, J., 1971, A Note on the Distribution of Stock Price Changes, *Journal of the American Statistical Association*, Vol 66, 282-284
- Theil, H., 1971, *Principles of Econometrics*, Wiley and Sons, New York
- Vasicek, O.A., 1973, A Note on Using Cross-Sectional Information in Bayesian Estimation of Security Betas, *Journal of Finance*, Vol 28, 1233-1239

- Watson, M.W. and R.F. Engle, 1985, Testing for Regression Coefficient Stability with a Stationary AR(1) Alternative, *Review of Economics and Statistics*, Vol 67, 341-346
- Welsh, A.H., 1987, The Trimmed Mean in the Linear Model, *Annals of Statistics*, Vol 15, 20-45
- West, M., Harrison, P.J. and H.S. Migon, 1985, Dynamic Generalized Linear Models and Bayesian Forecasting, *Journal of the American Statistical Association*, Vol 80, 73-97
- White, H. and G.M. MacDonald, 1980, Some Large-sample Tests for Non-normality in the Regression Model, *Journal of the American Statistical Association*, Vol 75, 16-28
- Wise, J., 1966, Linear Estimators for Linear Regression Systems Having Infinite Residual Variances, *manuscript referred to in Blattberg and Sargent (1971)*
- Yitzhaki, S., 1982, Stochastic Dominance, Mean Variance and Gini's Mean Difference, *American Economic Review*, Vol 72, 178-185
- Zeckhauser, R. and M. Thompson, 1970, Linear Regression with Non-Normal Error Terms, *Review of Economics and Statistics*, Vol 52, 280-286