Information and Optimisation in Investment and Risk Measurement

by

Nato Kemkhadze

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Industrial and Business Studies

The research was conducted under the supervision of
Prof Stewart D. Hodges & Dr Victor Podinovski

University of Warwick, Warwick Business School
April 2004

Table of Contents

List of Figures

List of Tables

Abstract

The thesis explores applications of optimisation in investment management and risk measurement. In investment management the information issues are largely concerned with generating optimal forecasts. It is difficult to get inputs that have the properties they are supposed to have. Thus optimisation is prone to *'Garbage In, Garbage Out'*, that leads to substantial biases in portfolio selection, unless forecasts are adjusted suitably for estimation error. We consider three case studies where we investigate the impact of forecast error on portfolio performance and examine ways of adjusting for resulting bias.

Treynor and Black (1973) first tried to make the best possible use of the information provided by security analysis based on Markovitz (1952) portfolio selection. They established a relationship between the correlation of forecasts, the number of independent securities available and the Sharpe ratio which can be obtained. Their analysis was based on the assumption that the correlation between the forecasts and outcomes is known precisely. In practice, given the low levels of correlation possible, an investor may believe himself to have a different degree of correlation from what he actually has. Using two different metrics we explore how the portfolio performance depends on both the anticipated and realised correlation when these differ. One measure, the Sharpe ratio, captures the efficiency loss, attributed to the change in reward for risk. The other measure, the Generalised Sharpe Ratio (GSR), introduced by Hodges (1997), quantifies the reduction in the welfare of a particular investor due to adopting an inappropriate risk profile. We show that these two metrics, the Sharpe ratio and GSR, complement each other and in combination provide a fair ranking of existing investment opportunities.

Using Bayesian adjustment is a popular way of dealing with estimation error in portfolio selection. In a Bayesian implementation, we study how to use non-sample information to infer optimal scaling of unknown forecasts of asset returns in the presence of uncertainty about the quality of our information, and how the efficient use of information affects portfolio decision. Optimal portfolios, derived under full use of information, differ strikingly from those derived from the sample information only; the latter, unlike the former, are highly affected by estimation error and favour several (up to ten) times larger holdings.

The impact of estimation error in a dynamic setting is particularly severe because of the complexity of the setting in which it is necessary to have time varying forecasts. We take Brennan, Schwartz and Lagnado's structure (1997) as a specific illustration of a generic problem and investigate the bias in long-term portfolio selection models that comes from optimisation with (unadjusted) parameters estimated from historical data. Using a Monte Carlo simulation analysis, we quantify the degree of bias in the optimisation approach of Brennan, Schwartz and Lagnado. We find that estimated parameters make an investor believe in investment opportunities five times larger than they actually are. Also a mild real time-variation in opportunities inflates wildly when measured with estimated parameters.

In the latter part of the thesis we look at slightly less straightforward optimisation applications in risk measurement, which arise in reporting risk. We ask, what is the most efficient way of complying with the rules? In other words, we investigate how to report the smallest exposure within a rule. For this purpose we develop two optimal efficient algorithms that calculate the minimal amount of the *position risk required*, to cover a firm's open positions and obligations, as required by respective rules in the FSA (Financial Securities Association) Handbook. Both algorithms lead to interesting generalisations.

Chapter 1

Introduction

1.1. Overview

The work contained within this thesis explores some of the applications of optimisation both in investment management and risk measurement.

Modern portfolio management typically employs many financial theoretical concepts and advanced academic techniques. The use of formal optimisation models in finance goes back to the roots of modern finance with the world of Markovitz (1952). Today the technology of portfolio selection has advanced considerably from the static framework (e.g. Black and Scholes (1973)). Efficiency of optimisation models in investment management rests entirely upon forecasts and information. It is difficult to get inputs that have the properties they are supposed to have as the historical data do not constitute a good representation for deriving reliable estimates for model parameters (see, for example, Campbell, Lo and MacKinlay (1997)). Thus optimisation is prone to the problem of *'Garbage In, Garbage Out'*, that reduces model efficiency and leads to substantial biases in portfolio selection, unless models / forecasts are adjusted suitably for estimation error. In the investment part of the thesis we consider three case studies where we investigate the impact of forecast error on portfolio efficiency from various perspectives and examine different ways of adjusting for bias.

In risk management we look at slightly less straightforward applications which arise in reporting risk. We consider two risk measurement rules from the *FSA* (Financial Securities Authority) *Handbook* (2001) and develop original optimal algorithms that calculate the smallest risk exposures in settings of respective rules. One of the algorithms generalises to the Theorem on Convex Optimisation, formulated and proved in the thesis.

This introductory chapter surveys the relevant literature. To embed our work within existing research, we will give pointers to how these ideas are developed in the body of the thesis.

The rest of this chapter is structured as follows. We start with a general discussion of information issues in investment (Section 1.2) and move on to the problem of estimation error in mean-variance optimisation (Section 1.3). In Section 1.4 we analyse different implementations of the Bayesian approach to adjusting for estimation bias in forecasts in a mean-variance framework. Section 1.5 explores the portfolio selection model of Treynor and Black (1973) and subsequent generalisations to their model. We explain how we extend the Treynor-Black analysis to account for forecast errors in Chapter 2 of the thesis. Section 1.6 describes briefly how we implement our Bayesian adjustment to infer optimal scaling of forecasts of expected returns under incomplete information on asset returns. This problem is addressed in Chapter 3. Section 1.7 describes other (non-Bayesian) approaches to adjusting for estimation error in inputs of optimisation models. In Section 1.8 we discuss the inference issues in long-term portfolio selection in presence of stochastic and predictable investment opportunities. Dynamic optimisation with parameters estimated from limited historical data introduces estimation bias in the model. We describe how we quantify the extent of such bias which appears to be substantial indeed (for rigorous treatment see Chapter 4). Section 1.9 discusses the risk measurement and capital adequacy issues, in the context of requirements by financial regulators. We discuss the optimal algorithms we developed to calculate minimal capital risk required within two FSA rules. These algorithms and their generalisations form Chapters 5 & 6 of the thesis. Section 1.10 concludes with the outline of the thesis.

1.2. Information in Portfolio Theory

As P. Bernstein (1992, p.75) recalls in his "Capital Ideas", the question, "Do you beat the market?" used to be insulting to portfolio managers about four decades ago, as it was taken for granted that all managers outperformed the market. "The only room for argument was over which professional manager's returns were furthest above average." The equilibrium model well known as the capital asset pricing model (CAPM) of Sharpe (1964), Treynor (1965), Lintner (1965), and Mossin (1966), proved them wrong. CAPM concludes that the stock market itself is the optimal portfolio[1] and holding the market is the optimal strategy for an average investor. Empirically, Treynor (1965), Sharpe (1966) and Jensen (1968) were the earliest papers to analyse the performance of professional investment managers and suggest that they have not been very successful. In a more recent work Gruber (1996) and Carhart (1997) make a similar point. So, who beats the market?

[1] It is optimal in the sense that no other portfolio can offer a larger return for same risk, or less risk for same expected return.

Today it is well understood that tracking the market portfolio is already good enough and only a relatively small number of managers with superior investment skills (e.g. asset allocation skills, stock picking skills[2]), and the ability to optimally use these skills can actually beat the market (Cochrane, 1999). The finance community is still in search of intelligent answers on how to deliver above-average returns. In the light of limited information provided by financial markets, the importance of analysing this information in the most efficient way can never be overestimated.

The literature on information analysis started with work of Treynor and Black (1973), Hodges and Brealey (1973), and Ambachtsheer (1974). Other references, using a similar approach and Bayesian statistics, include Blume (1971, 1975), and Vasicek (1973). Ideas suggested by these papers remain intuitively appealing today and have largely inspired this thesis. The information issues in long- and short-term portfolio optimisation differ. Long-term, or *strategic asset allocation* decisions relate to relative amounts invested in different asset classes over the long term while short term portfolio decisions are associated with *tactical asset allocation*[3]. These two therefore complement each other — strategic asset allocation defines a broad picture of investments, subject to long-term market forecasts, and tactical asset allocation 'tilts' holdings between active and passive (market) portfolios, depending on short-term forecasts of market movements and, at the same time, complying with risk/reward constraints defined at the strategic allocation levels. We start with the analysis of forecast errors in a single-period MV portfolio selection framework and then move onto more complex setting of continuous-time long-term portfolio management. However this separation is rather conditional. The approaches to adjusting for estimation error (e.g.

[2]Asset allocation skills tend to play a more important role in investment management, reducing the possibility of obtaining substantial gain through stock picking alone (see Brown and Harlow (1990), Brinson, Hood, and Beebower (1986), and Brinson, Singer, and Beebower (1991)).

[3]According to Haugen (2001), the time horizon for TAA is limited to one year.

Bayesian adjustment, robust optimisation) surveyed in the context of MV optimisation represent essential means of resolving uncertainty in more complicated settings too. Similarly, many of the models referenced in the beginning are continuous-time (e.g. Cvitanic, Lazrak, Martellini and Zapatero (2003)) and few are not on mean-variance optimisation at all (e.g. Maenhout (2004)). We hope such presentation will not cause any confusion.

1.3. Estimation Error in MV Approach

Markovitz's (1952) mean-variance (MV) portfolio optimisation model is the most quoted quantitative model in the investment literature. Given estimates of expected return, standard deviation or variance, and correlation of returns for a set of assets, MV efficiency provides the investor with an exact prescription for optimal allocation of capital. MV optimisation problem is given by

$$\max_{w} \quad \mu' w - \frac{1}{2}\lambda w' \Sigma w \tag{1.1}$$

where w is the $N \times 1$ vector of portfolio weights, μ is the $N \times 1$ vector of expected returns, Σ is the $N \times N$ covariance matrix of returns, and λ denotes risk aversion. In each period the investor chooses his portfolio w^* $\left(w^* = \frac{1}{2}\lambda\Sigma^{-1}\mu\right)$ to maximise the value of the objective function, trading expected portfolio return, $\mu' w$, against portfolio variance, $w'\Sigma w$.

There is considerable literature on the strengths and limitations of MV analysis (e.g. Markovitz (1987)). Markovitz's mean-variance optimisation model is simple and intuitive but is widely criticised for its sensitivity to the inputs: the first two moments of the distribution of future asset returns, and the matrix of expected future correlations of all returns. In the classical implementation of (1.1), μ and Σ are replaced by their estimates $\hat{\mu}$ and $\hat{\Sigma}$ respectively, changing (1.1) to the following:

$$\max_{w} \ \hat{\mu}'w - \frac{1}{2}\lambda w'\hat{\Sigma}w$$

Not surprisingly, estimation error is known to have a huge impact on MV optimised portfolios. They are unbalanced and fluctuate widely over time. The extreme sensitivity of a mean-variance portfolio to replacing unknown parameters by estimated sample values was first reported by Hodges and Brealey (1973). Other references include Michaud (1989), Best and Grauer (1991), Chopra and Ziemba (1993), Chopra (1993), Ziemba and Mulvey (1998), and Litterman (2003) who present some empirical and theoretical results on the sensitivity of optimal portfolios to changes in means, variances and covariances.

Solution to (1.1) overweights those assets that have large estimated expected returns, low estimated variances and low estimated correlations to other assets. According to Chopra and Ziemba (1993), out of all the parameters, estimates of expected means have greatest impact on a MV portfolio. They find that errors in means are about ten times as important as errors in variances, and errors in variances are about twice as important as errors in covariances. Best and Grauer (1991) note that "a surprisingly small increase in the mean of just one asset drives half the securities from the portfolio." The difficulty of statistical estimation of mean returns (see Merton (1980)) means that the model often allocates the highest portion to the asset class with the largest estimation error.

There are several approaches to dealing with estimation error in portfolio optimisation. One approach is to correct for errors in the inputs, using Bayesian inference. Under different implementations of Bayesian adjustment, the sample estimates are shrunk to some prior values in order to reduce the effect of sampling error. Another popular approach is robust optimisation. To account for parameter uncertainty, an estimation robust investor considers a set of plausible parameters and chooses the worst case (see Section 1.7.1). Other less popular approaches are discussed in Section 1.7.2.

1.4. Bayesian Adjustment

As argued above, using the information available within a sample alone is not sufficient to generate reliable portfolio rules. Examples of outside of sample information include equilibrium models, the expertise and views of investors or financial analysts, and results of unrelated experiments. External information can be formally modelled using Bayesian analysis.

Zellner (1971) and Berger (1985) give a general introduction to Bayesian analysis and Bawa, Brown, and Klein (1979) to its application to portfolio theory. Using Bayesian estimates is the most popular way of dealing with estimation error in portfolio management. The Bayesian approach may be implemented in several ways.

In one implementation, a number of papers, including Barry (1974), Klein and Bawa (1976), Bawa, Brown, and Klein (1979), and Brown (1979), use either a non-informative diffuse prior or a predictive distribution obtained by integrating over the unknown parameter rather than using the historical estimate of the parameter value. The advocates of a diffuse prior argue against imposing an informative prior due to difficulties in justifying a particular informative prior.

To deal with estimation risk, another stream of papers, such as Korkie and Ratti (1979), Jobson and Korkie (1980), Jorion (1985, 1986), Frost and Savarino (1986), Dumas and Jacquillat (1990), and Chopra, Hensel, and Turner (1993), uses empirical Bayes estimators, which are equivalent to weak informative priors. For example, Jorion (1985, 1986) develops the Bayes/Stein estimator that a priori assumes equal expected returns for all assets and hence shrinks the MV optimised portfolio towards the minimum-variance-portfolio (MVP).

MVP excludes the sample information on expected returns and uses only the covariance matrix of returns.

In a third implementation of the Bayesian approach, authors use the equilibrium implications of an asset pricing model to establish a prior. Important papers on this subject include Black and Litterman (1990, 1991, 1992), Jorion (1991), Pastor (2000), and Pastor and Stambaugh (2000). We will give detailed analysis of the Black and Litterman model later in the chapter, in Section 1.5.2.

In the next section we discuss how we account for parameter uncertainty in the model of Treynor and Black (1973), and how we adjust for forecast error.

1.5. Forecast Errors in Portfolio Selection: Treynor and Black (1973) Perspective

Treynor and Black (1973) first tried to make the best possible use of the information provided by security analysis based on Markovitz portfolio selection. In this section we provide a brief description of their methodology and findings. We discuss their results in the light of estimation error and quantify the extent of bias in optimised portfolio, caused by the estimation error. We provide an alternative analysis that better accounts for available information as well as unknown parameters, and gives a fair picture of existing investment opportunities (see **Chapter 2**). The last section of this paragraph discusses the literature related to the work of Treynor and Black.

1.5.1. Asset Allocation by Treynor and Black (1973)

Sharpe (1966) proposed a measure of fund performance that incorporated the ratio of excess return to the standard deviation of return μ/σ. The higher the Sharpe ratio, the more risk-adjusted return the investor expects. Treynor-Black accepted that definition and first stressed the analysts' role in portfolio management (Treynor and Black, 1973). They showed that the optimal holding of an individual stock should not depend on the investment manager's expectations regarding the general market. An investment manager is assumed to optimise his portfolio by combining n independent assets (active portfolio) and a market (passive) portfolio. Two managers with radically different expectations regarding the general market will select the active portfolios with the same relative proportions. For the optimal holdings $h_i^* = \frac{1}{\lambda}\frac{\mu_i}{\sigma_i^2}$ the portfolio manager obtains the highest Sharpe Ratio μ_p/σ_p. Its square $\left(\mu_p/\sigma_p\right)^2$ equals

$$\frac{\mu_p^2}{\sigma_p^2} = M^2 + \sum \frac{\mu_i^2}{\sigma_i^2}, \tag{1.2}$$

where μ_p, μ_i's and σ_p, σ_i's are expected excess returns and standard deviations of the portfolio and assets respectively, M is the market Sharpe ratio. Therefore the positions in securities are taken purely on the basis of expected independent returns and variance. Unless the analyst is able to anticipate all the events affecting the price, hence the return, some portion of the independent return variance remains unexplained by his forecasts. Let us assume that the correlation between the portfolio manager's predictions of excess return and the realisations is ρ. Then the squared Sharpe ratio becomes:

$$\frac{\mu_p^2}{\sigma_p^2} = M^2 + \frac{n\rho^2}{1-\rho^2}$$

Scalings of μ_i and w_i depend on what ρ is. As in reality an investor works with his anticipated correlation which may differ from the actual one, the interesting question is, how vulnerable is the Sharpe ratio to changes in investor's prior beliefs about the market? Particularly, the investor thinks that the correlations between his forecasts and the realised outcomes are R, but actually those correlations are ρ's.

In our analysis, we extend the Treynor-Black classical setting to incorporate such a discrepancy between the anticipated and actual correlations. We quantify how much the investor's actual Sharpe ratio differs from the one he anticipates and what the cost of the forecast error is. Interestingly, the Sharpe ratio we derive, that accounts for the difference in correlations, asymmetrically treats under and overestimation errors in forecasts — generally overestimation of the forecasts gives a larger Sharpe ratio rather than underestimation. We explain this paradox of success of overconfidence by the inability of the Sharpe ratio alone to take a full account of portfolio efficiency loss which, except for the wrongly assumed reward for risk, is attributed to the utility loss that comes from adopting an inappropriate risk profile. The following figure, taken from **Chapter 2**, provides an insight into the portfolio performance reduction due to forecast error:

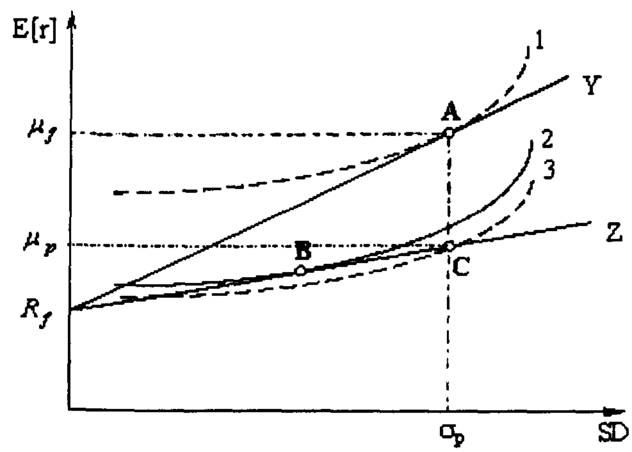

Figure 1.1

The investor thinks the opportunity set (optimal Sharpe ratio) is $R_f - Y$ and chooses (what he thinks is optimal) portfolio A on indifference curve 1 (see Figure 1.1). The actual opportunity set is $R_f - Z$ and if he knew that he would choose B on indifference curve 2. Instead he takes the risks of A, but gets a lower $E[R]$ as at C, which is only on indifference curve 3.

In light of the above argument, we extend our analysis to the utility based portfolio performance measure, the Generalised Sharpe Ratio (GSR), introduced by Hodges (1997), which reflects the reduction in the portfolio efficiency due to utility loss. We show that these two measures, the Sharpe ratio and GSR, in combination give a better account of portfolio quality.

1.5.2. Accounting for Uncertainty in Active Portfolio Management

Treynor-Black findings, essential to active portfolio management, have been unjustly neglected for many years. They have not lost their significance today, over three decades after publication, and recently started enjoying revived interest in the finance community. The simplicity of Treynor-Black ideas makes them highly influential among practitioners (see Taggart (1996)) as well as academics — recently published leading investment textbooks provide their detailed analysis (see Grinold and Kahn (1999), Bodie, Kane and Marcus (2001)).

Cvitanic, Lazrak, Martellini, and Zapatero (2002) extended the work of Treynor and Black to a dynamic setting, taking into account the hedging demands of an investor and learning during the investment horizon. While trying to optimise a dynamic portfolio, they rely heavily on estimated parameters that bring extra sources of bias into the model.

Treynor-Black's idea was refined by Black and Litterman (1991, 1992) by introducing uncertainty about the model in a Bayesian framework. Black and Litterman's asset allocation model combines market equilibrium with subjective views of investors about market opportunities. Instead of a single vector of expected excess returns as in the classical MV optimisation, a Black-Litterman investor is asked to specify as many views as he wishes, where each is an expectation of the return to a portfolio of his choosing. Each portfolio is referred to as a "view portfolio" (see Litterman (2003)). Besides view portfolios, the investor also specifies a degree of confidence as a standard deviation around the expectation. After such reformulation of the investor's problem, in an unconstrained optimisation Black-Litterman's optimal portfolio is a weighted combination of the equilibrium portfolio and the view portfolios (see He and Litterman (1999)). The optimal portfolio would tilt towards view portfolios with projected higher magnitude and confidence, and away from view portfolios with projected lower magnitude and confidence. The implementation of the model, especially the translation of uncertainty around the investor's view into the covariance matrix, is far from straightforward and has been studied by many authors. Most important papers among those are: Bevan and Winkelmann (1998), He and Litterman (1999), Satchell and Scowcroft (2000), Lee (2000).

These papers are different from ours. We identify a specific source of uncertainty and adjust the Treynor-Black formula for the inherent estimation error, based on almost as little information as the original formula had. At the same time, we explain the shortcomings of the original measure of portfolio performance and complement it with the utility-based measure.

1.6. Inferring Optimal Scaling of Forecasts of Expected Returns

This section presents some of the findings of **Chapter 3** where we study the problem of inferring optimal scaling of forecasts of asset returns in the presence of uncertainty about the quality of our information. In **Chapter 3** we look at the cross-section of asset returns where assets possess predictability but up to an unknown scale function. This is a corollary to the Treynor and Black model in an idealised world.

Consider a single period investor who faces financial market with n independent risky assets. Assume that the asset returns x_i come from the following model:

$$\begin{aligned} x_i &= s\delta_i + \sigma\varepsilon_i \\ i &= 1, ..., n \end{aligned} \tag{1.3}$$

where $\varepsilon_i \sim N(0,1)$ is a random noise, scaled by the known volatility σ that is constant across assets. Unobservable $\delta_i \sim N(0,1)$ represent the prior information on expected returns, but only up to an unknown scale function s. We are unsure what scale function should be applied. Assuming stability in the return generating process already given, we investigate how best to use historical data to infer the optimal scaling of our expected return forecasts. Then we solve the investor's portfolio problem in a mean-variance framework.

Based on the sample information only, the scale function is computed using a maximum likelihood estimation (MLE) which, due to finite sample limitations, is affected tremendously by the estimation error and cannot offer a satisfactory estimate of scaling. To help forecast expected returns we turn to Bayesian inference and use the prior that comes from outcomes of unrelated experiments. Our prior is the historical forecasting skill of an investor, modelled by a distribution reflecting the investor's historical correlation between forecasts and returns. The distributions are chosen so that they project into either normal $s \sim N(m, v)$ or uniform $s \sim [\underline{s}, \overline{s}]$ prior of the scaling. Additional information that updates prior scaling is the sample of observed returns.

After having estimated the scaling, we solve the investor's portfolio problem in a mean-variance framework. We find that, to fully exploit available information, an investor needs only the first two moments of the posterior distribution on his forecasting skills. Portfolio holdings, derived under efficient use of information, strikingly differ from the holdings derived from the sample information only (i.e. based on classical MLE). Ignoring the prior and optimising on observed returns only results in either several (up to 10) times larger holdings, or no investments at all. The analysis is extended to a multi-manager portfolio, where each manager works on a particular class of assets (each class of assets has its own unknown scale function). Besides being a more plausible setting, this framework offers new insights into investment under uncertainty.

With a modified assumption of observable systematic returns δ_i, our model is a generalisation to Ambachtsheer (1977) which provides a "IC ("Information" correlation) adjustment" that converts ex-ante alpha α_i into ex-post A_i, scaled input for optimisation. His scaling d equals the following:

$$d = IC\sigma\left(A\right)/\sigma\left(\alpha\right)$$

where

IC = cross-sectional "information" correlation of ex-ante and ex-post alpha,

$\sigma\left(A\right)$ = cross-sectional standard deviation of ex-post alpha,

$\sigma\left(\alpha\right)$ = cross-sectional standard deviation of ex-ante alpha.

$d\alpha_i$ afterwards can be interpreted as the excess return associated with forecast alpha, α_i. In our model we introduce uncertainty around the correlation (IC) and show analytically its impact on optimal MV portfolio.

Our model bears some similarities to the Black and Litterman approach (see Section 1.5.2). For a prior, they use the equilibrium returns and update them with an investor's views. Our prior instead is a measure of an investor's long-term investment performance (IC), which is transformed into his asset-specific forecasting skill using observed returns.

The optimal portfolios, analytically expressed for both models, are highly diversified and less sensitive to estimation error. We scale unobservable alpha, while Black and Litterman compute alpha based on equilibrium implications and use it as a neutral prior for expected returns. Our adjustment applies directly to the scaling of expected-return forecasts, i.e. to the model parameter, rather than to the expected-return forecasts the model produces.

Another paper that also refines the model parameters instead of forecasts, is Connor (1997). In his empirical study, Connor examines the impact of Bayesian adjustment on portfolio weights in a multi-period model where r_t is a time t return on the asset and x_{t-1} a variable observable at time $(t-1)$ that the investor believes can be used linearly to predict the time t return r_t:

$$r_t = bx_{t-1} + \varepsilon_t$$

The unexplained return is distributed normally with known variance $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. To adjust for the OLS estimate of b, Connor too applies a Bayesian adjustment. In his implementation, Connor sets the prior expected value of b to zero ($b \sim N(0, \sigma_b^2)$) which "is equivalent to weak-form market efficiency because it implies that the forecasting model has no ability to predict returns"[4] (see Connor (1997), p.44). He claims this is similar to the assumptions of Black and Litterman who "imposed their prior on expected returns based on the efficient markets hypothesis." Unlike Connor, our prior distributions on scaling have non-negative means since we expect a skilled investor to trade profitably on the difference between his expectations and those of the market. In the meantime, we do not exclude Connor's weak-form market efficiency since the posterior estimate of scaling can be zero. There is a difference in formulation between our models: Connor's predictable variable x_t is

[4] This analysis is similar to the β-adjustments of Blume (1971, 1975) and Vasicek (1973), where βs are shrunk towards the market beta of 1.

observable, unlike our unobservable δ_i. We also note that our implementation of Bayesian adjustment differs from that of Connor. The Bayesian estimation in Connor's work is based on the solution to a standard Bayesian regression problem (see Chow (1983)). We solve explicitly for optimal holdings while Connor presents numerical solutions only. Despite these differences, Connor too finds that the portfolio weights based on the unadjusted forecasting model are extremely aggressive.

1.7. Non-Bayesian Approaches

Recent growth of non-Bayesian approaches has been so striking that there is a sufficient reason for us to provide a brief introduction to the corresponding literature even though none of our work is of this type.

1.7.1. Robust Portfolio Choice

Robust optimisation is another approach that offers vehicles to incorporate estimation risk into the decision-making process in portfolio selection. In robust optimisation an investor is assumed to be aware that an estimated model is only approximately true. To account for the model uncertainty, he considers a set of all plausible models. Unlike a Bayesian investor, a robust investor has too little information to assign probabilities to alternative models and instead considers the least favourable model. The size of the set depends on the degree of required robustness — the larger the set the poorer the worst case. Extreme robustness in portfolio selection results in highly conservative portfolio strategies. For example, Maenhout (2003) finds that introduction of robustness drastically reduces the demand for a risky asset.

In the presence of uncertainty about model parameters, investors may consider multiple priors about the mean and the variance of asset returns instead of a single prior. Anderson, Hansen and Sargent (1999) is a key reference that develops a model of decision making that allows for multiple priors and where the decision maker is not neutral to uncertainty[5] (see also Chen and Epstein (2002), and Uppal and Wang (2003)). Such a framework is consistent with the ambiguity-aversion known to be experienced towards having multiple priors (see Ellsberg (1961)).

A few important papers that study robust portfolio choice besides those already listed, include Gilboa and Schmeidler (1989), Hansen and Sargent (2003), Rustem, Becker and Marty (2000). Robust portfolio rules in the context of input parameter uncertainty is examined by Goldfarb and Iyengar (2003), Lutgens and Schotman (2004), Tutuncu and Koenig (2003), Halldorsson and Tutuncu (2003), Garlappi, Uppal, Wang (2004).

Goldfarb and Iyengar (2003) proposed a robust portfolio selection model, based on Ben-Tal and Nemirovski (1998, 1999)[6]. They assume that the unknown market parameters lie in a known and bounded uncertainty set, and the robust portfolio is computed by solving a *max-min* mean-variance problem assuming worst case behaviour of parameter values within the set. The uncertainty sets here can be interpreted as confidence regions around the point estimates of the parameters. Depending on the degree of confidence there is a probabilistic guarantee on the performance of the robust portfolio.

The following is an example of robust optimisation in the face of input/model uncertainty. If a robust investor obtains advice from J experts about the inputs of MV model (μ_j, Σ_j) , $j \in J$, and believes that one of them is right, he will consider the worst case by

[5] In contrast, a Bayesian investor is neutral to uncertainty.
[6] Ben-Tal and Nemirovski (1998, 1999) represent most important references to the optimisation techniques.

reformulating the MV optimisation problem:

$$\max_{w} \min_{j \in J} \ \mu_j' w - \frac{1}{2}\lambda w' \Sigma_j w$$

This corresponds to the model with the rival return and rival risk scenarios of Rustem, Becker and Marty (2000). In Lutgens and Schotman (2004), each expert supplies a set of uncertainty U_j around his estimates of the mean and/or variance. If investors consider uncertainty in the estimator of the expected return only, the estimation and model robust portfolio is found by

$$\max_{w} \min_{j} \min_{\mu \in U_j} \ \mu' w - \frac{1}{2}\lambda w' \Sigma_j w$$

Expert j believes the uncertainty set U_j to contain all the plausible parameter values for μ_j. A critical aspect of the analysis is specifying the sets of uncertainty. In general multi-prior model takes the following form:

$$\max_{w} \min_{\mu} \ \mu' w - \frac{1}{2}\lambda w' \Sigma w$$

subject to

$$f(\mu, \hat{\mu}, \Sigma) \leq \varepsilon$$

Robust optimisation is a relatively new and fast-growing area of optimisation and is being applied successfully to various areas of finance. The main characteristic of robust optimisation is that it takes a pessimistic view, based on the assumption of ambiguity aversion, related to Ellsberg-style experiments (see Ellsberg (1961)). Some believe this assumption to be non-rational.

1.7.2. Other Approaches

To smooth the estimation error, papers like Grauer and Shen (2000), Frost and Savarino (1988) advocate imposing arbitrary portfolio constraints (see also Jaganathan and Ma (2003)).

Michaud (1998) proposes a method of resampled efficiency, by drawing repeatedly from the return distribution based on the original optimisation inputs and finds efficient frontier portfolios based on these resampled returns. His method, however, is rather arbitrary as he provides no statistical justification for choosing an interval for resampling. Although his method generally obtains a better answer than raw MV optimisation, we have no grounds to believe this method is at least 'locally optimal'. Applying meaningful Bayesian prior instead would be more reasonable. Scherer (2002) discusses pros and cons of Michaud's theory of resampled efficiency.

Scenario-based stochastic programming models have also been proposed for handling the uncertainty in parameters (Ziemba and Mulvey (1998) give a survey of this research). This approach becomes very inefficient as the number of assets grows.

1.8. Long-Term Portfolio Choice

1.8.1. Predictability in Long-Term Asset Returns

The problem of estimation error is not limited to the mean-variance framework although, for its simplicity and analytical tractability, the MV framework best illustrates the extent of inefficiencies that an optimised portfolio may suffer from. Since MV utility, concerned with only the mean and the variance of investor's final wealth, is not rich enough to provide a satisfactory match to market participants' priorities in multiperiod and continuous-time models, more complex utility functions (e.g. power utility, Epstein-Zin utility) are favoured for long-term portfolio management models.

The investor's time horizon has a significant effect on the composition of the optimal portfolio. The process of portfolio management in long-term continuous-time models becomes dynamic where an investor seeks to dynamically maximise expected utility function over a horizon and will choose a portfolio, which is optimal for this objective. This involves both static optimisation in each period and dynamic allocation over time. It consists of continuously readjusting portfolio portions so as to take the evolution of the market into account. Now optimisation depends on several parameters that are continuously reevaluated and fed back into the optimisation model. We know that even in a simple setting the limited size of historical data does not allow for unbiased estimates of parameters that drive the asset returns process. The continuous time setting inevitably complicates the problem of accounting for bias in optimisation with a long horizon. Here, in the presence of several interrelated and independent sources of bias, the optimality of portfolio is closely tied to how efficiently the available information is being analysed and processed.

Based on seminal work by Merton (1969, 1971) and Samuelson (1969), it has been understood that intertemporal portfolio problems would not reduce to a straightforward sequence of single-period problems unless

- Investors have no labour income and investment opportunities are constant (or uncorrelated with asset returns) over time[7];

  and / or

[7]The opportunity set summarises investor's information on the distribution of asset returns over the remaining investment horizon. This assumption eliminates either uncertainty regarding his opportunity set, which the investor may wish to hedge against or speculate on, or the investor's ability to profit from hedging/speculation.

- Investors have logarithmic utility[8].

That means, if either assumption given above holds, it is optimal for an investor with a 20-year-horizon to hold the same portfolio as for an investor with a 1-year-horizon (*ceteris paribus*). The conventional wisdom suggests that long-term investors should hold more of the risky asset than their short-horizon equivalents. If this is true and both investors have power utilities, the assumption of constant investment opportunities cannot be realistic. On the theoretical side, it has been known since Merton (1973) that variation in expected returns over time can potentially introduce horizon effects.

There is now a consensus in empirical finance that the asset class returns are to some extent predictable. Papers that found significant evidence of predictability in the long-term asset return (i.e. systematic component of return), include Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988), Fama and French (1989), Ferson and Harvey (1991), and Bakaert and Hodrick (1992). The explanatory variables as well as their significance vary across studies but all agree that the explanatory power of detected predictive relationships is statistically weak but economically significant.

Predictable time-variation in expected asset returns makes it possible to time the market. While theoretical formulae in the dynamic case are available in general contexts, their implementation under realistic assumptions gives rise to complex terms that do not have explicit forms and are difficult to evaluate numerically. Under incomplete information on expected asset distribution, to what extent does the evidence of predictability affect optimal portfolio rules?

Kim and Omberg (1996) analytically examined the long-term portfolio decision in the presence of predictability in the market, for HARA utilities. In their model, the investor

[8] With logarithmic utility the optimal portfolio strategy is myopic even in the presence of time-dependent market opportunities.

trades two assets, a riskless bond and a risky asset, in a complete market. The riskless rate r is constant and the price P_t of the risky asset follows a diffusion process

$$\frac{dP_t}{P_t} = \mu_t dt + \sigma_t dz$$

where $\mu_t = r + X_t\sigma_t$. The state variable X_t is the risk premium on the risky asset and follows the Ornstein-Uhlenbeck process

$$dX_t = -\lambda \left(X_t - \overline{X}\right) dt - \sigma_X dz_X$$

where $\lambda, \overline{X}$ and σ_X are positive constants. The correlation between the asset-return and risk-premium processes is given by

$$E\left[dz dz_X\right] = \rho dt$$

Kim and Omberg explicitly solve the non-myopic portfolio problem under different assumptions on interrelations between the parameters. They demonstrate how diverse are the expectations that market participants can have, depending on their beliefs about parameters. They find that such model specification permits for various biases in the process that lead to overprediction of returns. For example, given constant σ_t, the sign of correlation ρ results in a bias towards higher expected returns after price rises and vice versa. The extreme sensitivity of expected returns to the model inputs demonstrates that the process Kim and Omberg use for modelling long-term expected returns is inadequate and some adjustments to the model and/or parameters are required to get plausible results.

Brennan, Schwartz and Lagnado (1997) made an empirical attempt to solve the optimal portfolio problem based on a drift-driven model when the world dynamics was governed by three state variables: short-term risk rate r_t, console bond yield l_t, and the dividend yield δ_t. The asset choice consisted of stock, bonds and cash, and expected asset returns were predictable. They used the following joint stochastic process to model the market evolution:

$$
\begin{aligned}
\frac{dS_t}{S_t} &= (a_{11} + a_{12}r_t + a_{13}l_t + a_{14}\delta_t)\,dt + \sigma_1 dz_1 \\
dr_t &= (a_{21} + a_{22}r_t + a_{23}l_t + a_{24}\delta_t)\,dt + r_t\sigma_2 dz_2 \\
dl_t &= l_t\,(a_{31} + a_{32}r_t + a_{33}l_t + a_{34}\delta_t)\,dt + l_t\sigma_3 dz_3 \\
d\delta_t &= (a_{41} + a_{42}r_t + a_{43}l_t + a_{44}\delta_t)\,dt + \delta_t\sigma_4 dz_4
\end{aligned}
\tag{1.4}
$$

where S_t is the stock price, the variance covariance matrix is constant as are all a_{ij} coefficients in drifts. The model parameters were estimated from fitting regressions to 20-year-data on corresponding market observations. The investor in the model has power utility defined over terminal wealth with a risk aversion of (-5). Based on these estimated parameters, Brennan, Schwartz and Lagnado solved the portfolio choice problem in a dynamic setting.

In their numerical solution to strategic asset allocation Brennan, Schwartz and Lagnado found extraordinarily large predictable investment opportunities, accompanied by extreme sensitivity of holdings to parameter estimates. Merton (1980) pointed out that an observer of a continuous price path can estimate a constant volatility with arbitrary precision over an arbitrary short period of calendar time, provided he has access to arbitrarily high-frequency data. The estimate of expected returns, by contrast, depends on the length of the sample period rather than the frequency of the data available. 20-year-data can provide a set of satisfactory estimates for volatilities and correlations, but it is too short to reliably estimate expected returns, even under the assumption that the historical data constitute a good representation of forthcoming periods[9].

Part of the investment opportunities, predicted by Brennan, Schwartz and Lagnado, comes from estimation error but to what degree, has not been quantified so far. The subsequent literature has advanced considerably from the assumption that coefficients are estimated with certainty. However it has not yet realised the dangers associated with the portfolio selection approach pioneered by Brennan, Schwartz and Lagnado — long-term optimisation based on parameters estimated from short samples. In subsections 1.8.2 - 1.8.3 we discuss the related literature that studies the long-term portfolio problem in the presence of parameter uncertainty (learning and uncertainty about predictability) but has not addressed the kind of bias we investigate in this thesis.

1.8.2. Inference Issues in Presence of Learning about Parameters

The optimal portfolio solution changes when we allow for uncertainty about model parameters. Researchers, investigating the impact of learning on portfolio allocation decision, assume that as investors observe asset returns over time they will learn about the true parameter values; and the anticipation of learning has an impact on their earlier investment decision. It implies that the portfolio choice problem can be solved in two steps: parameters first estimated, and then portfolios are chosen conditional on these parameter estimates. This separation of the estimation and optimisation steps is optimal only when the property of certainty equivalence applies[10]. Early contributions to the topic of portfolio selection with learning include Williams (1977), Gennotte (1986), Detemple (1986), and Dothan and

[9] The assumption of constant covariance matrix over 20 years remains questionable.

[10] In particular, certainty equivalence states that unknown state variables can be replaced in the optimisation problem by their least squares estimators. It applies when the objective function is quadratic and the process is a linear function of unobservable state variables (See Lucas and Sargent (1981)).

Feldman (1986). Under the Gaussian-Markov structure, they demonstrate that a separation principle holds. Namely, agents can solve the inference problem to form their expectations, and then solve their optimisation problem based on this set of expectations which is equivalent to their original optimisation problem. Their findings enable an investor to derive optimal estimators for the unobservable expected instantaneous returns using observations of past realised returns and use those for utility maximisation. For example, Gennotte (1986) establishes a separation theorem that extends the discrete-time (quadratic-linear) certainty equivalence principle to continuous time without restrictions on utility function[11]. Based on the separation result, Gennotte (1986) analyses the impact of estimation error on portfolio selection. He finds that in continuous time models with diffusion the effects of parameter uncertainty are different from those found in single-period models. Gennotte shows that for non-logarithmic preferences, the level of investments and the term structure depend upon future variations of the derived opportunity set, which arise from randomness in both the true opportunity set and in the estimation error[12]. As a result, a risk-averse investor (with non-logarithmic utility) reduces his investment in a risky asset when the uncertainty around its expected return increases, which is consistent with the literature (see, for example, Black and Litterman (1992)).

In more recent work, Brennan's (1998) learning model assumes constant investment opportunities and his investor has to learn only the long-term mean of the risky asset.

In a simple illustration we scratch the surface of how the learning works. Consider an investor with the power utility (γ risk aversion), who maximises his terminal wealth. Investment opportunities are fixed and there is a riskless and a risky asset in the market. Instantaneous returns on the riskless asset follows

[11] This result is based on Liptser and Shiryayev (1977).

[12] An investor with logarithmic utility optimally ignores stochastic variation in the future investment opportunity set.

$$\frac{dB_t}{B_t} = rdt$$

and on the risky asset —

$$\frac{dP_t}{P_t} = \mu dt + \sigma dz$$

All parameters r, μ and σ are assumed to be constant. We know, that if they are known, the investor would put portion x^* of his wealth in risky asset, $x^* = (\mu - r)/\gamma\sigma^2$. Following the argument in 1.8.1, an investor is not likely to know the true value of μ. Therefore he makes his investment decision based on his current beliefs about μ and in the light of future learning about this parameter. Standard assumption for μ at time 0 is a normal prior, updated in a Bayesian way. It is notable, that if underlying parameters are constant, the investor eventually $(t \longrightarrow \infty)$ learns true μ (see Gennotte (1986)). The described process suggests that learning creates a positive correlation between realised returns and revisions in return forecasts.

Brennan and Xia (1998) is another interesting paper in this stream which, like other papers in this section, studies the impact of learning on portfolio choice without predictability.

The learning considered here does not solve the estimation problem we posed earlier. We discuss more complex learning papers in the following subsection which examines effects of learning when returns are to some degree predictable.

1.8.3. Predictability in the Presence of Estimation Error

A pioneering work by Kandel and Stambaugh (1996) studied the impact of predictability in the portfolio choice problem in the presence of estimation error. They consider a multi-period setting but their investor is assumed to have a single-period horizon. They find that predictability remains economically significant even if it is uncertain — the empirically observed predictability in the market still allows the investor to make superior returns even if the market is timed correctly only 1 out of 100 times.

Barberis (2000) extends the analysis of Kandel and Stambaugh (1996) and investigates the impact of parameter uncertainty in a dynamic asset allocation framework. Relaxing Brennan, Schwartz and Lagnado (1997)'s assumption, that investors know the parameters of the stochastic process generating asset returns with certainty, he introduces parameter uncertainty in the portfolio choice. Barberis (2000) assumes uncertainty surrounding the predictive relation of returns and predictive variables and derives a dynamic strategy for a Bayesian investor in a discrete time setting, considering estimation error. Barberis finds, that even in the presence of parameter uncertainty, a long-term investor allocates to risky assets more than does a myopic investor but significantly less than an investor using Brennan, Schwartz and Lagnado's strategy.

Xia (2001) examines the combined impact of learning and predictability on portfolio decision, drawing on the earlier work on learning (e.g. Gennotte (1986))[13] as well as predictability in the presence of parameter uncertainty (e.g. Barberis (2000)). In her model the investor knows the true long-term mean stock return, but does not know the short-term dynamics of expected return. As a result, revisions in the estimated parameter are no longer

[13] The separation theorem established in Gennotte (1986) is not affected by the predictability (see Xia (2001) for more details).

perfectly positively correlated with innovation in stock returns (which is the case with learning without predictability; see Section 1.8.2.). Now the correlation can be either negative or positive, depending on other factors. This implies another intertemporal hedging term in the demand for stocks which offsets already existing positive intertemporal hedge term.[14]

Another important paper on parameter uncertainty that a long-term investor faces, is Avramov (2002) who considers a model with many risky assets. Other contributions on the optimal decision rules in the presence of returns with differing degrees of predictability include Campbell and Viceira (1999), and Brandt (1999).

The optimisation approaches in these papers, investigating the impact of parameter uncertainty and learning that started with Barberis (2000) and Xia (2001), are still liable to create bias in the spirit of Brennan, Schwartz and Lagnado (1997). They optimise portfolios with the parameter estimates derived from fitting the historical data to regressions. Given the small explanatory power of these regressions, and a limited time-span for learning, estimated (and updated) parameters cannot be relied upon. In **Chapter 4** we conduct a Monte Carlo simulation analysis where we take Brennan, Schwartz and Lagnado's model and show how misleading the whole procedure followed by these authors can be. The problem of bias is so severe, that even allowing for uncertainty in predictability, or learning cannot resolve it, given that the methodology of optimisation remains the same. Prior literature on parameter uncertainty has missed the point we address here and the analysis presented in this thesis is, to our knowledge, the first attempt to highlight the degree of bias in long-term optimisation with parameters estimated from historical data.

[14]Barberis (2000), which ignores learning about predictability, finds only positive intertemporal hedge term.

1.8.4. Bias in Long-Term Portfolio Optimisation Models with Estimated Parameters

In this section we describe our Monte Carlo analysis and show that the way in which Brennan, Schwartz and Lagnado estimated their model parameters leads to massively biased inputs for the optimisation model.

We assume that the true return generating process is (1.4), the one used in Brennan, Schwartz and Lagnado. To make the model realistic, the parameters are calibrated so that an instantaneous Sharpe ratio, measuring the market opportunities, is within its real-life range with the average of 0.5 at each instant over 20 years. With this model, we simulate a 20-year returns sample similar to that used by Brennan, Schwartz and Lagnado as their dataset. We assume that an investor knows the true model as well as true volatilities and the correlation matrix of the return/predictor variables, but has to estimate drifts of the true model from the simulated data using regression estimates, as was done in Brennan, Schwartz and Lagnado. Since historical datasets as long as this are usually used for inferring model parameters for long-term portfolio decisions, it is essential for us to know to what degree true and inferred parameters match each other. For reasons of accuracy, we conduct 10,000 independent simulations and produced a distribution of 10,000 estimates for each parameter that had to be estimated. The actual parameter estimates we use for the estimated model are averages of the corresponding distributions. For a new model with estimated parameters we calculate expected investment opportunities measured by an instantaneous Sharpe ratio, given that the real world dynamic follows the true model. We find that the estimated model promises outstanding investment opportunities corresponding, on average, to the Sharpe ratio of 2.2, as opposed to the true Sharpe ratio of 0.5! The difference between the two is attributed solely to the estimation error. Moreover, a scarcely noticeable actual time-variation in investment opportunities inflate significantly once measured with least squares estimates, derived through regressions — the average Sharpe ratio varies between 1.85 and 2.45 during twenty years.

Our Monte Carlo simulation analysis showed that the distributions of parameter estimators are rather large. Each individual estimate in this distribution is a least squares estimate for the corresponding path which is believed to be the best possible historical prior for that parameter. A Bayesian investor will attempt to infer these parameters over time but the investment horizon (say, 20 years) is not sufficient for the prior estimate to reach the true parameter value. In our experiment, it took 10,000 years for all estimates to converge to the true values of their parameters respectively. Our parameters were assumed constant, while in real-life continued shocks and structural shifts may prevent investors from ever resolving the uncertainty.

We argue that our results, striking as they are, are bound to be less biased compared to those of Brennan, Schwartz and Lagnado. Firstly, in our simulated world we eliminate a few sources of bias which exist in the Brennan, Schwartz and Lagnado model (e.g. model misspecification). Also knowing the true dynamics of the world, we fairly 'discount' inflated expected opportunities. In the optimiser's view, placing his faith in the estimated model, these opportunities are several times larger. Another reason we believe our picture is more moderate than that of Brennan, Schwartz and Lagnado, is that our reported results (all individual coefficient estimates as well as the Sharpe ratio) are averages over 10,000 simulations, while in real-life we are limited to one path of realisations only.

1.8.5. How Can We Adjust for Bias in Dynamic Portfolio Management? No Answers...

After having overviewed existing optimisation models, we conclude that there is a need for radical and meaningful adjustment in the models of long-term portfolio returns. Unfortunately, there is no straightforward way of doing this. Grinold and Kahn (1999) suggest Bayesian refinement to the parameter estimates, but there is no obvious way of implementing them in a complex environment of continuous-time portfolio selection.

The solution to the problem may lie in making adjustments so that the forecasts, generated by a return-generating model, are consistent. The role forecast evolution in time may play in portfolio selection has not been fully understood so far. By the definition of forecasts, the model efficiency is strongly tied to forecast optimality which is especially apparent when modelling long-term returns.

We regard this problem of adjusting for bias as a generalisation of β-adjustment problem, examined by Blume (1971, 1975), Vasicek (1973). With β however, there is prior information on cross-sectional distribution of βs and in their setting implementation of the Bayesian approach is straightforward. In continuous-time asset returns models we have a class of assets which is not even a homogeneous group and we cannot consider their mean and dispersion as we did for β.

Alternatively, incorporating an asset-pricing model into the analysis, as done by Black and Litterman (1990, 1992), Pastor and Stambaugh (2000), and combining the results of optimisation with the implications of an asset-pricing model by using Bayesian inference, can give a reasonable treatment to the problem.

More research is required to come up with sensible ways of adjusting model parameters for estimation error. At present we have limited ourselves to only general suggestions around the topic.

1.9. Optimisation in Risk Measurement

1.9.1. Risk Measurement and Capital Adequacy: Regulatory Issues

Risk measurement forms an important part of the complex and multi-stage process of risk management (see Jorion (2003), Gallai (2003), Marrison (2002), Rustem and Howe (2002)).

Sound risk control is essential to the prudent operation of a financial institution and to promoting the stability of the financial system as a whole. The efficient functioning of markets requires participants to have confidence in each other's stability and ability to transact business. Capital rules help foster this confidence because they require each member of the financial community to have, among other things, adequate capital. This capital must be sufficient to protect a financial organisation's depositors and counterparties from the institution's on- and off-balance sheet risks. Top of the list are credit and market risks; not surprisingly, firms are required to set aside capital to cover these two main risks.

Capital standards are designed by regulators, like the Financial Securities Authority (FSA), in order to allow a firm to absorb its losses, and in the worst case, to allow a firm to wind down its business without loss to customers, counterparties and without disrupting the orderly functioning of financial markets. Minimum capital standards are thus a vital tool to reducing systemic risk. They also play a central role in how regulators supervise financial institutions.

Capital requirements have been criticised for being simple mechanical rules rather than applications of sophisticated risk-adjusted models (see, for example, Marrison (2002)). That fairly raises issues about the optimality of current rules but we do not aim to study that. Instead, we investigate how we can report the smallest exposure within a rule in the most cost-effective and capital-efficient way; this is a legitimate interest of banks (see Gallai (2003)). To cope with moderations in rules, the ways of calculating minimal exposures must not only be optimal but also flexible to accommodate future changes.

We have picked two independent rules in the FSA Handbook and developed the optimisation algorithms that calculate minimal capital requirements in their settings. To do so, we found optimal ways to offset a large number of individual positions or obligations. The next section describes them in more detail.

1.9.2. Description of Algorithms

Both algorithms we developed are sufficiently general to account for reasonable modifications to rules and remain optimal and efficient. **Chapter 5** describes them in detail. All algorithms and theorems in the thesis are original. To derive them, we used a wide range of optimisation/mathematical programming techniques (see, for example, Ahuja, Magnanti and Orlin (1993), Rockafellar (1996)). Rustem and Howe (2002) provides algorithms for computing the best decision in view of the worst-case scenario. The main tool they use is *minimax*, which ensures robust policies with guaranteed optimal performance that will improve further if the worst case is not realised. Our work however is not on *minimax* optimisation and our algorithms are not related to those described in Rustem and Howe. In fact, our risk capital minimisation is actually minimising the sums of absolute values of exposures, so it is in the $L1$ norm, not the $L\infty$ norm.

Risk Offsetting Algorithm

The Risk Offsetting Algorithm minimises the required capital held by a firm to cover its positions in commodities with different maturities. Offsetting opposite positions of different maturities incurs additional cost. The smallest risk exposure results from netting all opposite positions in the most cost-efficient way.

Using combinatorial optimisation and backward induction, we develop an efficient algorithm that yields all optimal solutions to the problem in closed form and is easy to implement with a standard optimisation package. The optimality of the algorithm follows from the Theorem on Convex Optimisation that we formulate and prove in **Chapter 6**. This theorem delivers all solutions for a specific combinatorial optimisation problem (including the risk offsetting algorithm in question). More precisely, if $x_1(t), x_2(t), ..., x_n(t)$ are continuous convex functions on $[a, b] \subset R$, the theorem finds all sets of points $\{t_1, t_2, ..., t_n\}$ from $[a, b]$, arranged in the ascending order, so that $\sum_{i=1}^{n} x_i(t_i)$ were minimal. In notation, it solves

$$\begin{aligned} & Min \quad \sum_{i=1}^{n} x_i(t_i) \\ S.t.\ a &= t_0 \leq t_1 \leq t_2 \leq ... \leq t_n \leq b \end{aligned}$$

where the optimisation is with respect to $t_1, t_2, ..., t_n$ variables.

The main contribution of this theorem is that it transforms an n-variable convex optimisation problem into a one-variable convex optimisation problem.

The theorem is proved using three lemmas, also given and proved in **Chapter 6**.

Optimal Grouping Algorithm

The other rule we consider here requires calculating the minimal interest rate risk by finding the optimal grouping of bonds with different maturities. We develop the Optimal Grouping Algorithm that finds the minimal interest rate risk within the rule, using a linear programming approach.

Next we consider a modified setting of the above problem, which is appealing for its economic insight, and solve this explicitly using dynamic programming techniques. The problem is as follows:

$\Psi(.)$ is a risk-measuring function of one-variable, which maps a (possibly negative) exposure to a non-negative risk measure. C_i's $(j = 1, ..., n)$ are given exposures (cashflows) that may be partitioned into K (or more) groups of length not greater than L for "offset" purposes. Find the grouping which minimises reported risk as

$$Min \sum_{j}^{K} \Psi(g_j)$$

$$s.t. \quad \sum_{i=l_{j-1}+1}^{l_j} C_i = g_j; \; j = 1, ..., K$$

$$1 \leq l_j - l_{j-1} \leq L; \; j = 1, ..., K$$

$$\begin{aligned} l_o &= 0, ..., l_K = N \\ C_i &\in R \end{aligned}$$

In other words, we find the best way to position the indices into the sets $G_1, ..., G_k$, so that each set contains no more than L entries. $k \geq K$ to ensure at least K groups. Similarly, the constraint $l_j - l_{j-1} \leq L$ guarantees at most L elements in each group.

The grouping problem is solved using the dynamic programming approach, based on the principle of optimality. It is notable that no restrictions are placed on the risk function $\Psi(.)$. This algorithm with complexity of at most $\left(N\frac{N-1}{2}\right)$ efficiently identifies the optimal solution among a finite set of $\left(2^{N-1}\right)$ alternatives.

1.10. Structure of the Thesis

This thesis analyses information and optimisation issues in portfolio selection and risk measurement.

The impact of estimation error on portfolio selection and suboptimality of the 'optimised' portfolio has long been recognised in the finance literature. Such failures are often driven by inefficient use of available information. The associated problems vary depending on the optimisation horizon an investor faces, as well as on the underlying model. The thesis starts with analysing forecast errors in a mean-variance framework. In Chapter 2 we re-examine earlier work by Treynor and Black (1973) who give a portfolio selection prescription in the presence of relatively little information. Retaining the simplicity of their analysis, we adjust their formula for forecast error and give a clearer picture of how to rank investment opportunities. Chapter 3 investigates how to use information from the outside of the model, to help infer the optimal scaling of forecasts of expected returns. The estimate of scaling derived using the sample information only, gives an inadequate portfolio which responds to unobservable estimation error with either overly aggressive allocations, or no risky holdings at all. In Chapter 4 we move from the inference issues in a static setting to continuous-time long-term portfolio selection models. Many authors model long-term asset returns with drift-driven diffusion processes where parameters are estimated from the historical data and, based on such models, attempt to solve the long-term dynamic portfolio management problem. Due to estimation error, this way of optimisation is bound to lead to substantial forecast error and, therefore, suboptimal portfolio rules. We quantify the degree of bias in a particular setting and find that it is indeed overwhelming. In Chapter 5 we consider two risk measurement problems from the *FSA Handbook* and develop original algorithms to calculate minimal position risk required under these rules. Both algorithms have interesting generalisation. One of them leads to the Theorem on Convex Optimisation, formulated and proved in Chapter 6; the other to a risk management problem, solved using dynamic programming techniques. Final remarks as well as suggestions for future work appear in the concluding Chapter 7.

Chapter 2

Overconfident Forecasts and Active Portfolio Performance

2.1. Introduction

This chapter examines the problem of portfolio selection based on imperfect forecasts in a mean-variance framework. Early work by Treynor and Black (1973) established a relationship between the correlation of forecasts, the number of independent securities available and the Sharpe Ratio which can be obtained.

Under the assumption of the Sharpe's diagonal model of security covariances (Sharpe, 1963) with the added assumption that direct investment in the market index is possible, they showed that the square of the Sharpe ratio (SR) provided by the optimal portfolio is equal to:

$$\frac{\mu_p^2}{\sigma_p^2} = M^2 + \sum \frac{\mu_i^2}{\sigma_i^2} \tag{2.1}$$

where μ_p and σ_p^2 respectively are the portfolio expected excess return and its variance, M is the market (Sharpe) ratio, μ_i is the expected abnormal return on security i (i.e. expected deviation from CAPM) and σ_i^2 is the residual variance for security i.

Assuming the correlation between the investor's forecasts on expected abnormal return of the particular securities and subsequent returns is known to be ρ, the squared Sharpe ratio could be rewritten as[1]

$$\frac{\mu_p^2}{\sigma_p^2} = M^2 + \frac{n\rho^2}{1-\rho^2}. \tag{2.2}$$

It seems that (2.2) allows an investor to rank his investment opportunities assuming that he has some control over the parameters in (2.2). Given the near efficiency of security markets, we know that the correlation ρ will be low. For example, for $M = 1/2$ and $n = 75$ a value of ρ as low as 0.10 (so $\rho^2 = 0.01$) would be sufficient to increase the portfolio Sharpe ratio from a passive (market) figure of 0.50 to 1.00 (which would delight most portfolio managers and their clients).

Correlations as low as this are difficult to make inferences about. Rather than assume that the investor knows his correlation coefficient, we assume that he believes it to be R when in fact it is ρ.

This will enable us to model the cost of not knowing exactly what the correlation is. The issue is examined in context of the Sharpe ratio and a measure of expected utility.

Acting according to an erroneous assumption about forecast correlation has two disad-

[1] We will derive equations (2.1) and (2.2) later on.

vantages. First, since the relative holdings of active stocks and the market are no longer optimal, the Sharpe ratio is reduced. Second, since the investor's estimate of expected return is now biased, he is unable to take the optimal risk exposure given this modified opportunity set.

The investor thinks the opportunity set (optimal SR) is $R_f - Y$ and chooses portfolio A on indifference curve 1 (see Figure 2.1).

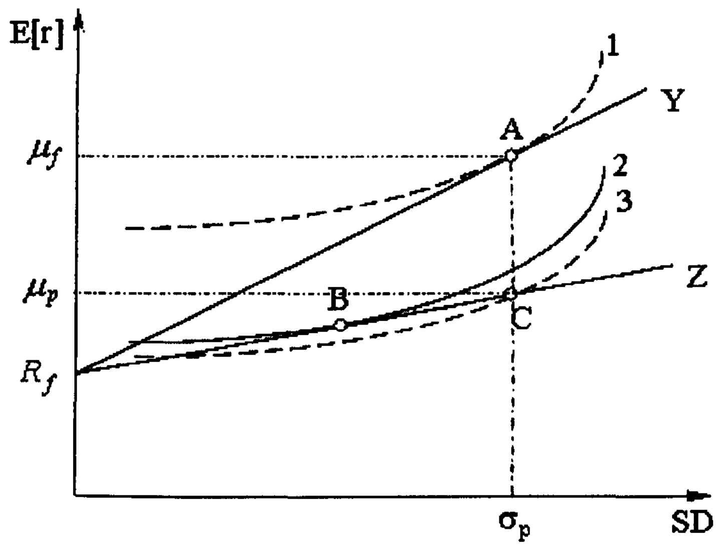

Figure 2.1

The actual opportunity set is $R_f - Z$ and if he knew that he would choose B on indifference curve 2. Instead he takes the risks of A, but gets a lower $E[R]$ as at C, which is only on indifference curve 3.

It is therefore useful to describe the effects of the Sharpe ratio reduction and the utility reduction separately.

We measure the expected utility obtained in the following way: We assume the investor has utility function $-e^{-\lambda W}$. Given that we are working with normal distributions, expected utility is $-e^{-\lambda\left[\mu_p-\frac{\lambda}{2}\sigma_p^2\right]}$ with μ_p differing from his expectation μ_f. This figure is then reinterpreted as an equivalent Sharpe ratio — that of the known opportunity which would provide the same level of expected utility. This is the same principle as in Hodges (1997), introducing the Generalised Sharpe Ratio (GSR). The generalisation is based on the expected utility to investors with constant absolute risk aversion (CARA). Given an investor with CARA, the GSR measures the extent of market opportunities irrespective of his level of risk aversion.

Here is a preview of the results of this chapter. We explore the properties of the Sharpe ratio and GSR and draw a number of conclusions for these two measures of portfolio performance. As expected, both of them are optimised when the investor knows precisely where his forecasting skills lie. For overconfident forecasters the Sharpe ratio, surprisingly, stays close to the optimum, while the underconfident investor significantly underperforms the overconfident one. On the other hand, the GSR is almost symmetric to the degree of over/underestimation and the smaller the deviation is between R and ρ, the better value it assigns to the forecast. Using these two different metrics we explore how portfolio performance depends on both the anticipated and realised correlations when these differ. The Sharpe ratio captures the efficiency loss caused by the change in reward for risk, while GSR reflects how the welfare of a particular investor is affected by adopting an inappropriate risk profile. Therefore, the Sharpe ratio and GSR complement each other and, in combination, provide a better measure of portfolio performance. This result is emphasised by examining the comparative statics with respect to the market price of risk and a number of securities in the portfolio. We also examine the worst-case scenario portfolio performance under the Sharpe ratio and GSR and conclude that a significant uncertainty around relatively high assumed correlation yields much smaller level of the 'guaranteed' utility, than a lower correlation with smaller uncertainty.

The rest of the paper is organised as follows. Sections 2.2-2.5 explore how the Sharpe Ratio and GSR depend on ρ and R, derive their properties and explain the 'paradox' of successful overestimation. Section 2.6 provides comparative statics with respect to the market price of risk and the number of securities in the portfolio. In Section 2.7 we compute the worst-case SR/GSR given an investor's certainty about his correlation. In Conclusion the results are summarised and a few suggestions for further work made.

Technical details appear in Appendices A1-A4 at the end of the paper.

2.2. How the Sharpe Ratio & GSR Depend on ρ and R

We provide a generalisation to the Treynor-Black's analysis of mean-variance portfolio selection. The procedure, used by Treynor-Black to derive (2.1) and (2.2), can be outlined as follows. They optimise the mean-variance utility with respect to h_i portfolio holdings of n independent risky assets and a market index, which stands for $(n+1)$st asset:

$$Max \ \sum_{i=1}^{n+1} h_i\mu_i - \frac{\lambda}{2}\sum_{i=1}^{n+1}\sigma_i^2 h_i^2$$

where expected abnormal returns on every ith asset (and the market index) are distributed normally with $N(\mu_i, \sigma_i^2)$, λ is a risk aversion coefficient. The portfolio is optimised for h_i^*, and abnormal return μ_p and standard deviation σ_p of the corresponding portfolio are

$$h_i^* = \frac{1}{\lambda}\frac{\mu_i}{\sigma_i^2}; \quad \mu_p = \lambda\sum_{i=1}^{n+1}\frac{\mu_i^2}{\sigma_i^2}; \quad \sigma_p = \lambda\sqrt{\sum_{i=1}^{n+1}\frac{\mu_i^2}{\sigma_i^2}}$$

It is straightforward that the squared Sharpe ratio of the portfolio is

$$\frac{\mu_p^2}{\sigma_p^2} = \sum_{i=1}^{n+1} \frac{\mu_i^2}{\sigma_i^2},$$

which is the same as (2.1) where M denotes the market Sharpe ratio. Assuming that an analyst's degree of forecasting ability, correlation ρ, is constant across n assets, (2.1) is easily rewritten as (2.2).[2]

Following our assumptions, the investor, whose actual correlation ρ is different from his anticipated correlation R, wrongly thinks that he has obtained the Sharpe Ratio of

$$\frac{\tilde{\mu}_p}{\tilde{\sigma}_p} = \sqrt{M^2 + \frac{nR^2}{1-R^2}}. \tag{2.3}$$

We derived that the Sharpe ratio actually obtained by this investor is the following (see Appendix A1):

$$SR = \frac{\mu_p}{\sigma_p} = \frac{M^2 + \frac{n\rho R}{1-R^2}}{\sqrt{M^2 + \frac{nR^2(1-\rho^2)}{(1-R^2)^2}}}. \tag{2.4}$$

We perform similar analysis for a measure of market opportunities, the Generalised Sharpe Ratio (GSR). It was introduced by Hodges (1997) as

$$GSR = \sqrt{-2\ln(-U^*)},$$

where U^* is the optimal utility, for an investor who maximizes $E\left[U(\widetilde{W})\right]$ with

$$U = -e^{-\lambda W}$$

If the investor's forward investment opportunity set (for a myopic investor) has future

[2]See Appendix A1 for details.

outcomes distributed as $N(\mu, \sigma^2)$, then

$$U = -\exp\left\{-\lambda\left(\mu h - \frac{1}{2}\lambda\sigma^2 h^2\right)\right\}$$

When $h = \frac{\mu}{\lambda\sigma^2}$, optimal U^* is

$$U^* = -\exp\left\{-\frac{1}{2}\frac{\mu^2}{\sigma^2}\right\} = -\exp\left\{-\frac{1}{2}GSR^2\right\}$$

The GSR, a measure of expected utility, is expressed through ρ and R (see Appendix A2) as

$$GSR = \sqrt{M^2 + \left(2\frac{n\rho R}{1-R^2} - \frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)}. \tag{2.5}$$

It is notable that the risk aversion measure does not participate in either expression (2.4) or (2.5). Neither measure will be affected by the introduction of cash in the model.

We will use these formulae to estimate the effect of over and underconfidence errors on the portfolio performance.

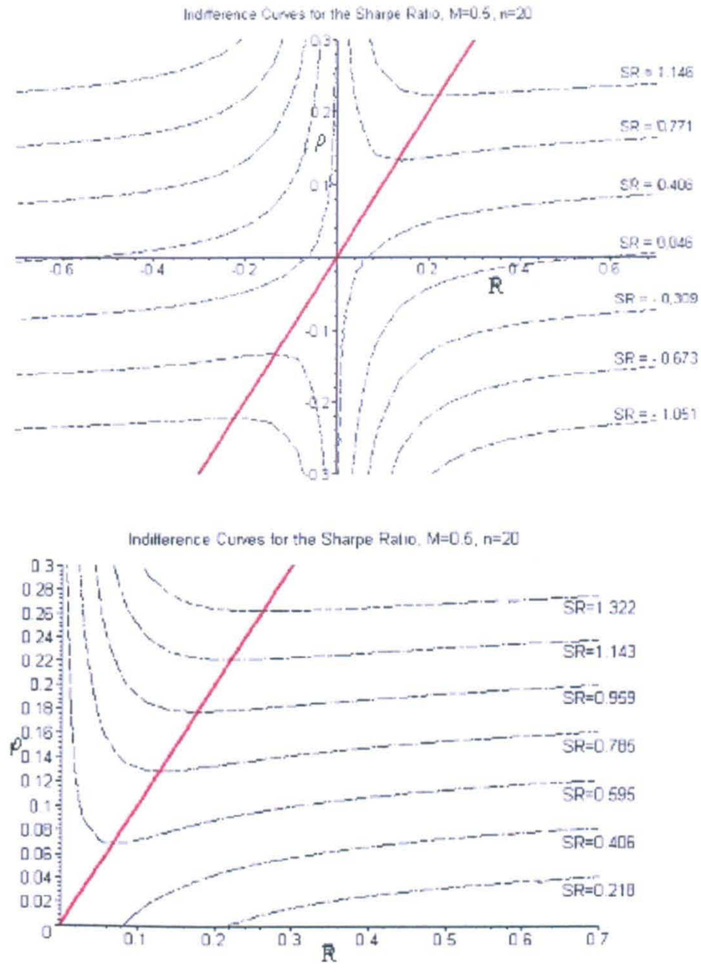

Figure 2.2

2.3. Properties of the Sharpe Ratio

In Figure 2.2 the indifference curves of the Sharpe ratio (for $M = 0.5$ and $n = 20$)[3] are plotted, where each curve represents the Sharpe ratio of a certain level for corresponding combinations of ρ and R. Inclusion of negative correlation quadrants presents a clearer picture of the world as well as the functional form of SR, and is of some interest when investigating the effect of negative ρ correlation, but negativity of R is unlikely to be true.

As anticipated, for given ρ, the Sharpe ratio is optimal when $R = \rho$. The optimal Sharpe

[3] Unless specified otherwise, we assume that M=0.5 and n=20.

ratios appear in red in Figure 2. To see the degree of bias coming from misestimation of skills, we investigate the Sharpe ratio's dependence on R. In Figure 2.3 the Sharpe ratio is plotted against R, for several realisations of ρ. The concavity of the curves that correspond to most likely values of parameters ρ&R, implies that overconfidence beats underconfidence, except when the true correlation ρ is very close to zero. In such a case, overestimation becomes a losing strategy.

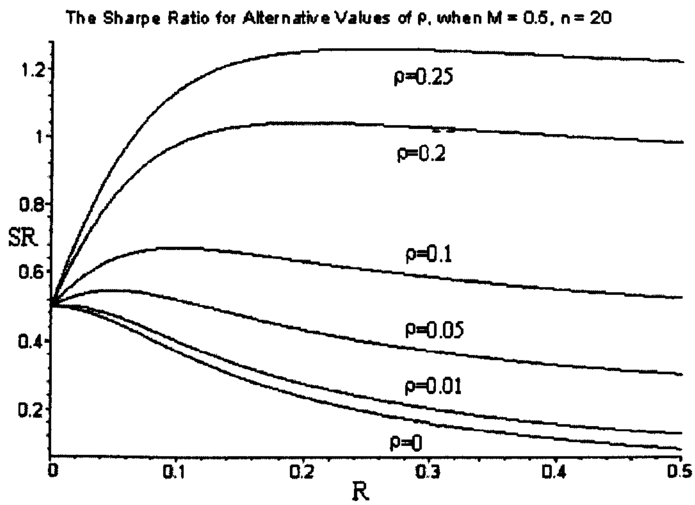

Figure 2.3

The shaded area of Figure 2.4 presents the combinations of (ρ, R) that generate a market-beating Sharpe ratio. The level curves of the Sharpe ratio move further away from the axes as skills improve. It means that investors with some forecasting skills (non-zero ρ) may be better off by holding an active portfolio, as far as they have a fair knowledge or even overestimate their forecasting skills. Meanwhile, even exceptionally good forecasting abilities (large ρ) may lead to a miserable Sharpe ratio if these abilities are substantially underestimated (small R). Solid curves in Figure 2.4 separate not only the market beating parameters (ρ, R) from those underperforming the market (ρ, R), but also act as a boundary between the pairs (ρ, R) which achieve a larger Sharpe ratio with aggressive R, and those that give a better Sharpe ratio with moderate R. In particular, in the shaded area, overconfident forecasts are superior to underconfident ones, in the nonshaded area — the relationship reverses. Solid curves represent the locus of (ρ, R) whose Sharpe ratio equals the market Sharpe ratio M.

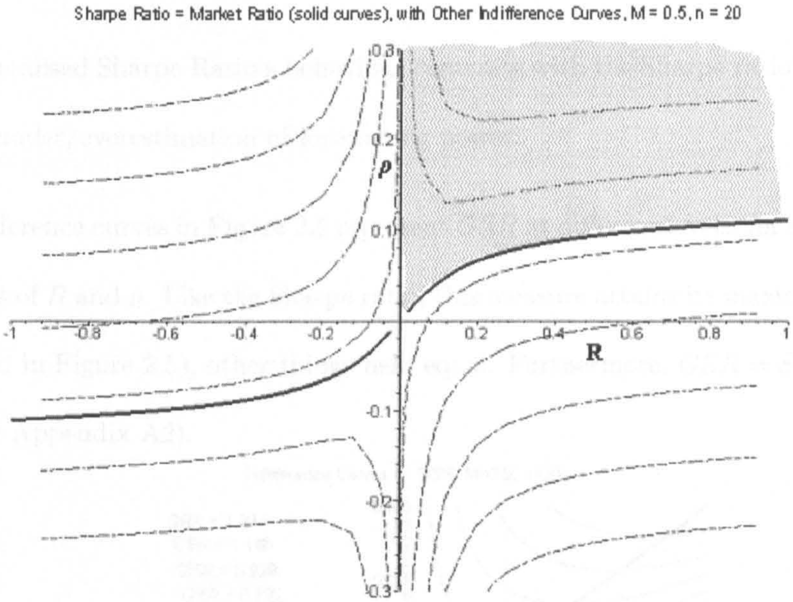

Figure 2.4

Based on (2.4) , we can compute what true correlation ρ corresponds to a target Sharpe ratio SR_{tgt}, for a given R :

$$\rho\left(SR_{tgt}, R\right) = \frac{M^2 n\left(R^2-1\right) + SR_{tgt}\sqrt{n}\sqrt{\left[\left(n+SR_{tgt}^2\right)\left[M^2\left(1-R^2\right)^2+nR^2\right] - M^4\left(1-R^2\right)^2\right]}}{\left(n+SR_{tgt}^2\right)nR}.$$

SR_{tgt} may represent any level curve in Figure 2.4 which separates (ρ, R) giving a Sharpe ratio better than SR_{tgt}, from those giving a Sharpe ratio worse than SR_{tgt}. A higher target shrinks the set of (ρ, R) that leads to a better than the target Sharpe ratio.

2.4. Properties of the Generalised Sharpe Ratio

The Generalised Sharpe Ratio's behaviour contrasts with the Sharpe ratio's asymmetric response to under/overestimation of forecasting power.

The indifference curves in Figure 2.5 represent GSR at different levels, for corresponding combinations of R and ρ. Like the Sharpe ratio, this measure attains its maximum at $R = \rho$ (drawn in red in Figure 2.5), other things held equal. Furthermore, $GSR = SR$ if and only if $R = \rho$ (see Appendix A2).

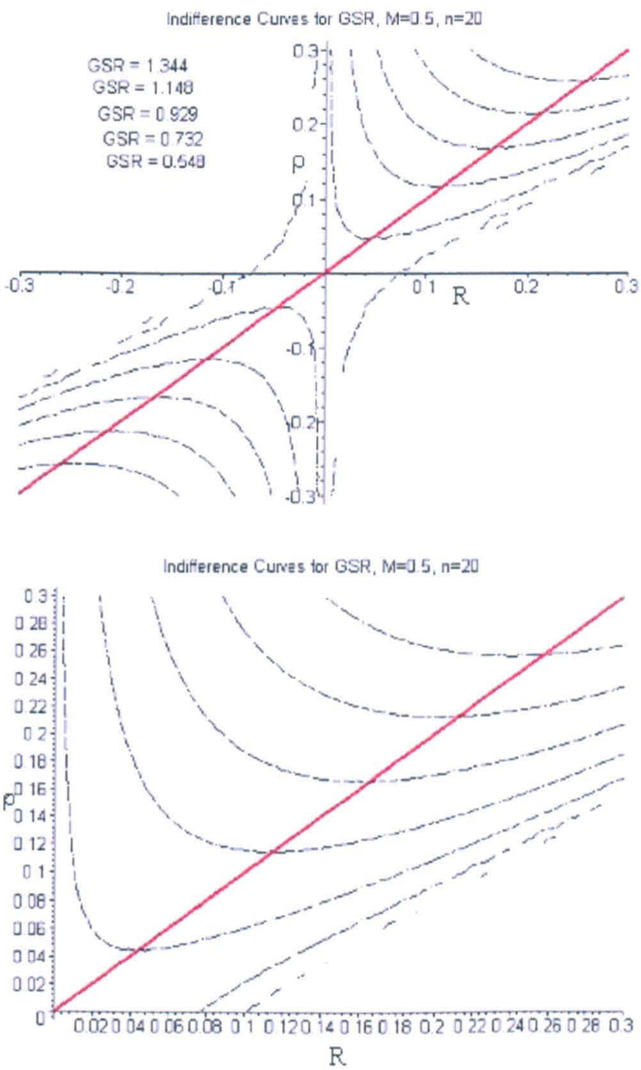

Figure 2.5

There is, however, a difference between how GSR and SR respond to forecast errors. Figure 2.6, where GSR is plotted against R for alternative realisations of ρ, demonstrates this difference. The GSR curves are almost equally sensitive to both under and overestimation errors in forecasts. In other words, unlike SR, the penalty for the overestimation with GSR is as significant as for the underestimation. It is worth noting that GSR is not even defined for a big gap between R and ρ, hence the sparsely populated quadrant of positive correlation pairs (ρ, R).

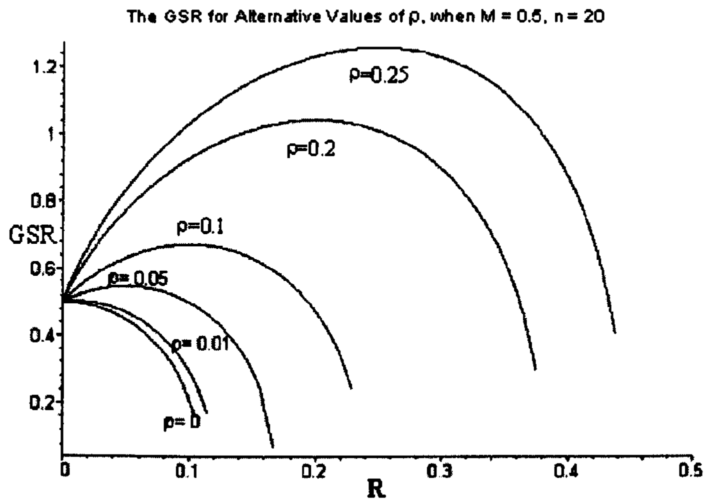

Figure 2.6

For given R, the correlation ρ corresponding to a target value GSR_{tgt}, is solved from (2.5) as follows:

$$\rho\left(GSR_{tgt}, R\right) = \frac{n\left(R^2-1\right)+\sqrt{n\left[\left(R^2-1\right)^2\left(n-M^2+K^2\right)+R^2\right]}}{R}.$$

Figure 2.7 highlights (ρ, R)'s, when $GSR_{tgt} = M$. The interpretation is similar to that of Figure 2.4.

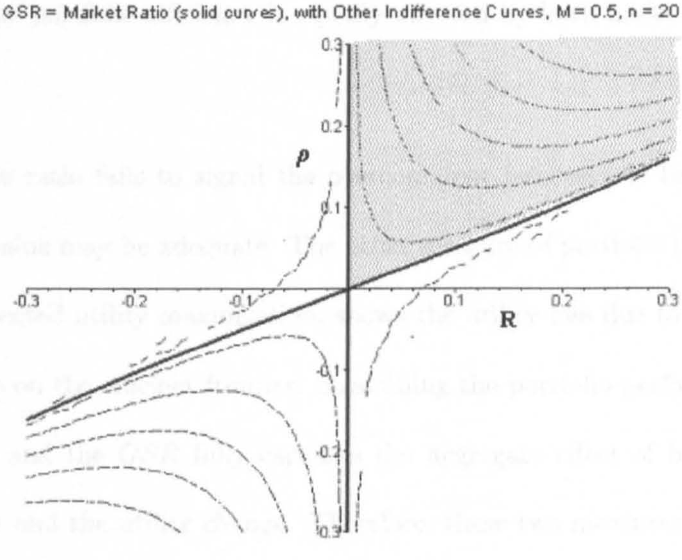

Figure 2.7

2.5. Paradox of Success of Overconfidence

Although the expected utility measure and the Sharpe ratio are optimal and equal when $\rho = R$, they have different properties with respect to misspecified forecasts. The Sharpe ratio penalises for the underconfidence substantially but forgives the overconfidence almost fully. On the contrary, GSR is symmetric in penalising for both over and underconfidence. This section explains this difference between the two measures of portfolio performance, and its implications.

An investor, whose $R \neq \rho$, loses not only in the Sharpe ratio that measures his reward for risk, but also in his utility for adopting an inappropriate risk profile. Therefore the portfolio efficiency loss is attributed to both misspecified expected reward for risk and taking a wrong risk profile. If $R > \rho$, the Sharpe ratio is slightly reduced while the corresponding utility decreases significantly. On the other hand, for $R < \rho$, the Sharpe ratio goes down fast, but not the respective utility. These two effects offset each other in either $R > \rho$ or $R < \rho$ case and, therefore, the portfolio efficiency is equally affected by both under and overconfident forecast errors.

As the Sharpe ratio fails to signal the overconfident forecasts, it becomes misleading even though its value may be adequate. The other measure of portfolio performance GSR, based on the expected utility maximisition, shows the utility loss due to choosing a wrong indifference curve on the efficient frontier. Examining the portfolio performance with both the Sharpe ratio and the GSR fully captures the aggregate effect of both the reduction in reward for risk and the utility change. Therefore, these two measures complement each other and in combination provide a better ranking of investment opportunities.

2.6. Comparative Statics on the Size of the Market Risk Premium and the Number of Securities

The Sharpe ratio and GSR are differently affected by the variation in the market risk premium. It is interesting to explore this issue as the specialists cannot agree what the actual size of the market risk premium is[4].

The Sharpe ratio's discriminating sensitivity to overconfident forecast error is magnified by a small market risk premium and a large number of securities in portfolio (see Figure 2.8) while GSR shows robustness to these changes (Figure 2.9). Furthermore, for small M and large n, the Sharpe ratio becomes insensitive to changes in R. The sensitivity of SR to the anticipated correlation increases with a bigger marker risk premium and/or fewer securities in the portfolio. Simultaneous changes in M and n emphasise/offset each other's effects.

[4] See, for example, Dimson, Marsh and Staunton (2002).

The number of independent assets n is an important parameter in the portfolio's performance. In general, the smaller the ρ, the easier it becomes to find more assets so that the correlation between the expected and realised returns equals ρ. According to expression of the Sharpe ratio, investing in many securities even in the presence of low forecast correlation, can be highly rewarding — the Sharpe ratio becomes unbounded for large n, irrespective of the value of R; such an investment is a luxury, that we cannot expect in reality. Unlike the Sharpe ratio, no n can eliminate R from the formula of GSR. Even with a large number n of stocks and positive ρ, utility can easily diminish with n, if $R > 2\rho$.[5]

This section confirms our intuition about the significance of the investor knowing what his forecasting skills are, and shows the importance of estimating portfolio performance with both the GSR and SR.

2.7. The Worst-Case Scenario Sharpe Ratio and GSR

In the previous sections we examined how the measures of portfolio performance, the Sharpe ratio and GSR, change in the presence of forecast errors. Their modified expressions are of little practical importance to an investor who at no point can be certain about his true correlation coefficient, unless he is able to estimate the downside potential of his portfolio. In this section we give an estimate of the worst-case portfolio performance under the assumption of discrepancy between the actual and expected SR/GSR. To proceed, we assume that a rational investor can provide an interval estimate of his true ρ, around his anticipated R. Suppose the investor is certain that his ρ is bounded by $R - \varepsilon$ from below, for some positive ε. The worst-case Sharpe ratio and GSR are the minimal Sharpe ratio and GSR

[5] See details in Appendix A3.

that an investor may obtain over the specified interval of uncertainty around his expected R. $\min_{\rho \geq R-\varepsilon} SR$ and $\min_{\rho \geq R-\varepsilon} GSR$ denote the minimal Sharpe ratio and GSR respectively.

Both measures, the Sharpe ratio and GSR, represent increasing functions of ρ and achieve their minimums over the given interval at $(R-\varepsilon)$. The further R is from ρ, the worse the performance. Thus, ε is a measure of investor's certainty about his skills which (ε) influences the worst-case portfolio performance. Assuming that $\varepsilon \leq \frac{R}{2}$ and considering that the GSR and the Sharpe ratio are equal when $R = \rho$, we get the following (see Appendix A4):

$$\begin{aligned} \min_{\rho \geq R-\varepsilon} SR &\approx \min_{\rho \geq R-\varepsilon} GSR \\ &\approx \sqrt{M^2 + \frac{n\left(R-\varepsilon\right)^2}{1-R^2}} \end{aligned} \tag{2.6}$$

As formula (2.6) suggests, great uncertainty around relatively high expected correlation yields much smaller level of the 'guaranteed' utility, than a lower correlation with smaller uncertainty. It confirms that the practical value of either portfolio performance measure is closely tied to adequate knowledge of forecasting abilities.

2.8. Conclusion

This chapter examined the problem of portfolio selection based on over/underconfident forecasts in a mean-variance framework. Early work by Treynor and Black (1973) established a relationship between the correlation of forecasts, the number of independent securities available and the Sharpe ratio which can be obtained. Their analysis was based on the assumption that the correlation between the forecasts and outcomes is known precisely. In practice, given the low levels of correlation possible, an investor may believe himself to have a different degree of correlation from what he actually has.

The current paper therefore described how the portfolio performance depends on both the anticipated (R) and realised correlation (ρ) when these differ. The portfolio performance was assessed according to the Sharpe ratio and a measure of expected utility obtained from investing on the basis of the forecasts, Generalised Sharpe Ratio (GSR). The Sharpe ratio measures the efficiency loss, attributed to misetimation in reward for risk. The reduction in the utility, the other source of the portfolio efficiency loss, is measured with the GSR which quantifies the utility obtained under this information set (i.e. different R and ρ) in terms of the Sharpe Ratio that would give the same utility.

The portfolio performance is assessed according to these two metrics — the Sharpe ratio and GSR. We found that the investor's degree of self-confidence plays an important role in the performance obtained. Provided that his actual forecasting skills are not too bad, an overconfident forecaster ($R > \rho$) attains almost as good Sharpe ratio as one who knows precisely where his forecasting skills lie ($R = \rho$), while an underconfident strategy ($R < \rho$) results in a significantly lower Sharpe ratio. Therefore, the Sharpe ratio alone is unable to capture all the efficiency loss which, except for the change in reward for risk, is attributed to the loss of utility due to chasing an inappropriate portfolio. Meanwhile, the Generalised Sharpe Ratio, the measure of expected utility reflects the reduction in the welfare of a particular investor, due to adopting a wrong risk profile. Interestingly, GSR stays almost symmetric to both over and underestimation errors. These two measures balance each other and, in combination, fully explain portfolio efficiency loss and provide a fair ranking of investment opportunities.

We explored the comparative statics of these two metrics with respect to the market risk premium and the number of securities in the portfolio; it confirmed the importance of evaluating portfolio performance jointly with the Sharpe ratio and *GSR*. We also showed that the level of uncertainty about the anticipated correlation coefficient plays a crucial role in estimating the worst-case scenario portfolio performance.

In future work the model may be extended to a multi-period setting. Another suggestion is to classify assets in several groups according to their degree of predictability, instead of considering them in a single homogeneous group with constant correlation ρ between forecasts and realisations. Then alternative ρ's would have different π's and a Bayesian analysis would be applied.

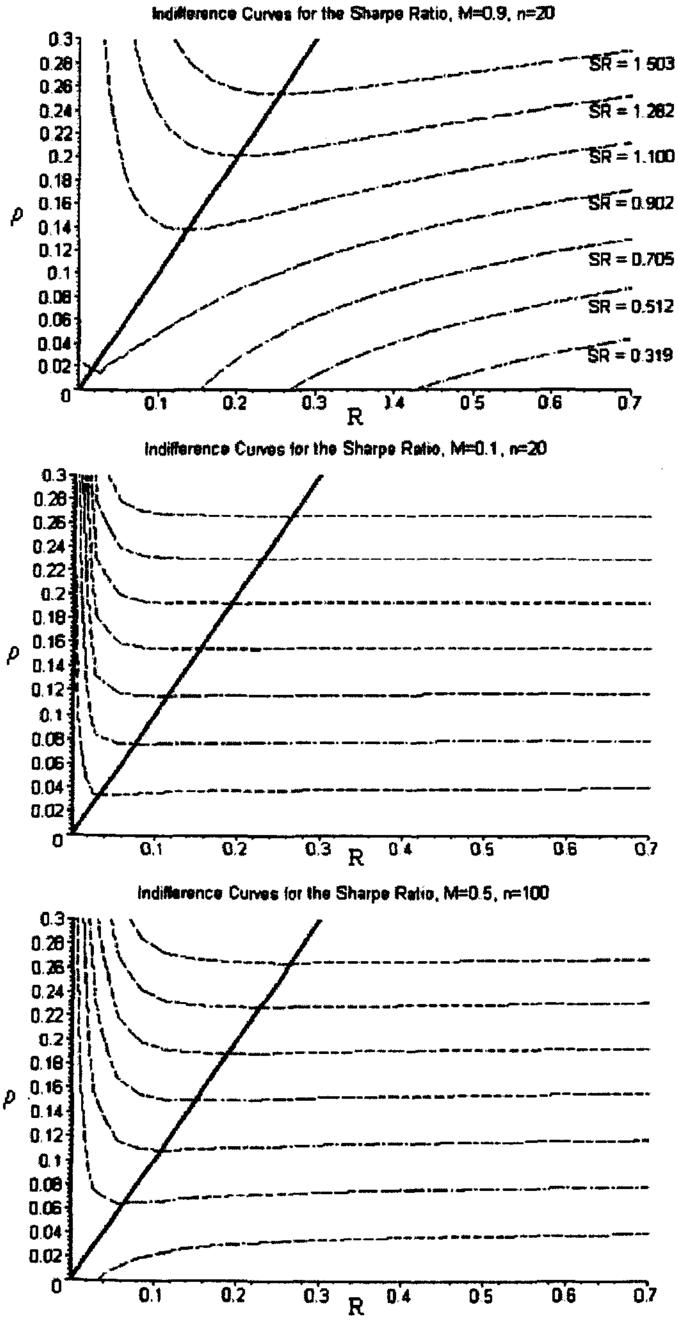
Indifference Curves for the Sharpe Ratio, M=0.9, n=20
SR = 1.503
SR = 1.282
SR = 1.100
SR = 0.902
SR = 0.705
SR = 0.512
SR = 0.319
ρ
R
Indifference Curves for the Sharpe Ratio, M=0.1, n=20
ρ
R
Indifference Curves for the Sharpe Ratio, M=0.5, n=100
ρ
R

Figure 2.8

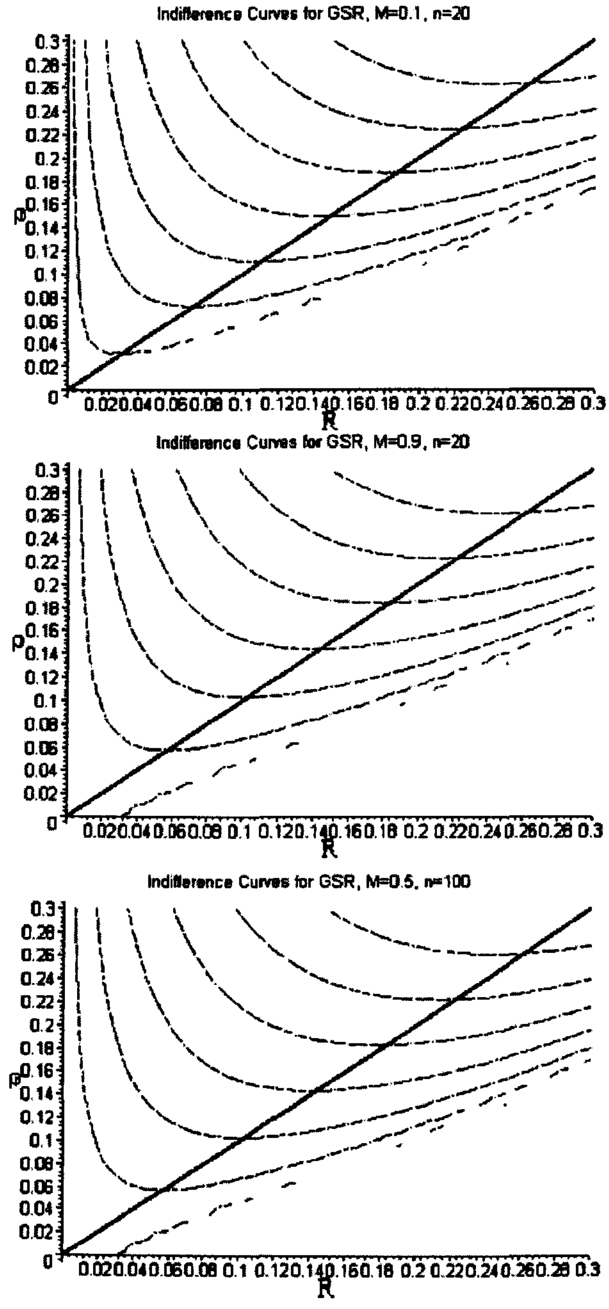
Indifference Curves for GSR, M=0.1, n=20
ρ
R
Indifference Curves for GSR, M=0.9, n=20
ρ
R
Indifference Curves for GSR, M=0.5, n=100
ρ
R

Figure 2.9

Appendix A

A1. Derivation of the Sharpe Ratio

We followed Treynor-Black's analysis of mean-variance portfolio and adopted their notation to derive formulae for the Sharpe ratio. Symbols are defined as:

β_i is market sensitivity of the ith security;

μ_i is the expected abnormal return on the ith security (i.e. expected deviation from CAPM) and σ_i^2 is the residual variance for security i;

(μ_m, σ_m^2) and (μ_p, σ_p^2) stand for the expected return and variance of the market portfolio and the investor's portfolio respectively.

Treynor-Black derived that for the optimal holdings h_i $(i = 1, ..., n)$ of n securities and the market portfolio h_m

$$\begin{aligned} h_i &= \frac{\sigma_p^2}{\mu_p}\frac{\mu_i}{\sigma_i^2}, \ (i = 1, ..., n) \\ h_m &= \frac{\sigma_p^2}{\mu_p}\left(\frac{\mu_m}{\sigma_m^2} - \sum_{i=1}^{n}\frac{\beta_i\mu_i}{\sigma_i^2}\right) \end{aligned}$$

the Sharpe Ratio achieves its maximum:

$$\frac{\mu_p^2}{\sigma_p^2} = \frac{\mu_m^2}{\sigma_m^2} + \sum_{i=1}^{n}\frac{\mu_i^2}{\sigma_i^2} = M^2 + \sum_{i=1}^{n}\frac{\mu_i^2}{\sigma_i^2} \tag{2A.1}$$

$\lambda = \frac{\sigma_p^2}{\mu_p}$ (a risk-aversion measure) does not participate in (2A.1) which means that for the optimal portfolio the proportions of positions are constant for all investors and do not depend on the risk aversion.

Investor takes positions in the securities according to his personal (subjective) expectations of abnormal asset returns $(\widehat{\mu}_i)$. Suppose that the correlation between the investor's forecasts on these particular securities and actual returns is ρ. The relevant measure of risk is the fraction of the variance σ_i^2 in abnormal returns not anticipated by the investor. Therefore we have

$$\widehat{\sigma}_i^2 = (1-\rho^2)\sigma_i; \quad h_i = \lambda \frac{\widehat{\mu}_i}{(1-\rho^2)\sigma_i^2} \tag{2A.2}$$

Note that investors, following the Treynor-Black model, cross-sectionally scale their forecasts $(\widehat{\mu}_i)$ so that

$$\begin{aligned} E\left[\widehat{\mu}_i\right] &= 0, \quad i = 1, ..., n; \qquad \sum_{i=1}^{n} \widehat{\mu}_i = 0 \\ \widehat{\mu}_i^{\,2} &= \rho^2 \sigma_i^2, \quad i = 1, ..., n; \end{aligned} \tag{2A.3}$$

Substituting expressions (2A.2) into (2A.1) returns:

$$\frac{\mu_p^2}{\sigma_p^2} = \frac{\mu_m^2}{\sigma_m^2} + \frac{n\rho^2}{1-\rho^2}$$

Suppose that an investor expects the correlations between his forecasts $\widetilde{\mu}_i$ and realised abnormal returns to be R's whilst in reality they are ρ's. Therefore, he wrongly thinks he has attained

$$\frac{\mu_p^2}{\sigma_p^2} = \frac{\mu_m^2}{\sigma_m^2} + \frac{nR^2}{1-R^2}$$

having taken the positions

$$h_i = \lambda \frac{\widetilde{\mu}_i}{(1-R^2)\sigma_i^2}, \ i = 1, ..., n \tag{2A.4}$$

as implied by (2A.3) when the correlation is R. Instead, given his prediction skills, he should be expecting

$$\widehat{\mu}_i = \frac{\rho}{R}\widetilde{\mu}_i; \qquad \widehat{\sigma}_i^2 = (1-\rho^2)\sigma_i^2$$

Choosing the $\widetilde{\mu}_i$'s instead of the $\widehat{\mu}_i$'s changes expected returns on the portfolios. The returns on the active and passive portfolios respectively become:

$$\begin{aligned}
\mu_{active} &= \sum_{i=1}^{n} h_i \widehat{\mu}_i = \sum_{i=1}^{n} \lambda \frac{\widetilde{\mu}_i}{(1-R^2)\sigma_i^2} \frac{\rho}{R} \widetilde{\mu}_i \\
&= \lambda \frac{\rho}{R} \frac{nR^2\sigma_i^2}{(1-R^2)\sigma_i^2} = \lambda \frac{n\rho R}{1-R^2} \\
\mu_{passive} &= \lambda \frac{\mu_m^2}{\sigma_m^2}
\end{aligned}$$

This makes the portfolio return equal to

$$\mu_p = \mu_{passive} + \mu_{active} = \lambda \left(\frac{\mu_m^2}{\sigma_m^2} + \frac{n\rho R}{1-R^2} \right)$$

The variances of active portfolio and the whole portfolio are:

$$\begin{aligned}
\sigma_{active}^2 &= \sum_{i=1}^{n} h_i^2 \widehat{\sigma}_i^2 = \lambda^2 \sum_{i=1}^{n} \frac{\widetilde{\mu}_i^2}{(1-R^2)^2\sigma_i^4}(1-\rho^2)\sigma_i^2 \\
&= \lambda^2 \sum_{i=1}^{n} \frac{R^2\sigma_i^2}{(1-R^2)^2\sigma_i^4}(1-\rho^2)\sigma_i^2 = \lambda^2 \frac{nR^2(1-\rho^2)}{(1-R^2)^2} \\
\sigma_p^2 &= \sigma_{passive}^2 + \sigma_{active}^2 = \lambda^2 \left(\frac{\mu_m^2}{\sigma_m^2} + \frac{nR^2(1-\rho^2)}{(1-R^2)^2} \right)
\end{aligned}$$

The squared Sharpe Ratio becomes:

$$SR^2 = \frac{\mu_p^2}{\sigma_p^2} = \frac{\left(\frac{\mu_m^2}{\sigma_m^2} + \frac{n\rho R}{1-R^2}\right)^2}{\frac{\mu_m^2}{\sigma_m^2} + \frac{nR^2(1-\rho^2)}{(1-R^2)^2}} = \frac{\left(M^2 + \frac{n\rho R}{1-R^2}\right)^2}{M^2 + \frac{nR^2(1-\rho^2)}{(1-R^2)^2}}$$

The Sharpe Ratio is a square root of the above, as follows:

$$SR = \frac{\left(M^2 + \frac{n\rho R}{1-R^2}\right)}{\sqrt{M^2 + \frac{nR^2(1-\rho^2)}{(1-R^2)^2}}}$$

A2. Derivation of the Generalised Sharpe Ratio (GSR)

Hodges (1997) introduced the Generalised Sharpe Ratio GSR as a measure of market opportunities as

$$GSR = \sqrt{-2\ln(-U^*)}$$

where is the optimal utility, for an investor who maximizes $E\left[U(\widetilde{W})\right]$ with

$$U = -e^{-\lambda W}$$

If the investor's forward investment opportunity set (for a myopic investor) has future outcomes distributed as $N(\mu, \sigma^2)$, then

$$U = -\exp\left\{-\lambda\left(\mu h - \frac{1}{2}\lambda\sigma^2 h^2\right)\right\}$$

When , $h = \frac{\mu}{\lambda\sigma^2}$, optimal U^* is

$$U^* = -\exp\left\{-\frac{1}{2}\frac{\mu^2}{\sigma^2}\right\} = -\exp\left\{-\frac{1}{2}GSR^2\right\}$$

and $GSR = SR$.

Considering the difference between the forecast and actually obtained portfolio, we will assume the distribution of portfolio's returns to be $N(\mu_F, \sigma_F^2)$, as expected by the investor, and by $N(\mu_p, \sigma_p^2)$, as it is in reality. According to Appendix A1, they are as follows:

$$\begin{aligned}
\mu_p &= \lambda\left(\frac{\mu_m^2}{\sigma_m^2} + \frac{n\rho R}{1-R^2}\right) \\
\sigma_p^2 &= \lambda^2\left(\frac{\mu_m^2}{\sigma_m^2} + \frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right) \\
\mu_F &= \lambda\left(\frac{\mu_m^2}{\sigma_m^2} + \frac{nR^2}{1-R^2}\right) \\
\sigma_F^2 &= \lambda^2\left(\frac{\mu_m^2}{\sigma_m^2} + \frac{nR^2}{1-R^2}\right)
\end{aligned}$$

As expected, the investor optimises his expected utility with $h = \frac{\mu_F}{\lambda\sigma_F^2}$ and, instead of expected $U^* = -\exp\left\{-\frac{1}{2}\frac{\mu^2}{\sigma^2}\right\}$, he obtains

$$\begin{aligned}
U^* &= -\exp\left\{-\mu_p\frac{\mu_F}{\sigma_F^2} + \frac{1}{2}\sigma_P^2\frac{\mu_F^2}{\sigma_F^4}\right\} = -\exp\left\{-\frac{\mu_F}{\sigma_F^2}\left(\mu_p - \frac{1}{2}\sigma_P^2\frac{\mu_F}{\sigma_F^2}\right)\right\} \\
&= -\exp\left\{-\frac{1}{2}GSR^2\right\}
\end{aligned}$$

$$\therefore\ GSR = \sqrt{\frac{2\mu_F}{\sigma_F^2}\left(\mu_p - \frac{1}{2}\frac{\mu_F}{\sigma_F^2}\sigma_P^2\right)}$$

In terms of ρ and R, GSR is expressed as

$$GSR = \sqrt{M^2 + \left(2\frac{n\rho R}{1-R^2} - \frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)}$$

GSR and SR are equal if and only if $R = \rho$:

$$
\begin{aligned}
M^2 + \left(2\frac{n\rho R}{1-R^2} - \frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right) &= \frac{\left(M^2 + \frac{n\rho R}{1-R^2}\right)^2}{M^2 + \frac{nR^2(1-\rho^2)}{(1-R^2)^2}} \\
M^4 + 2\left(\frac{n\rho R}{1-R^2}\right)^2 \frac{nR^2(1-\rho^2)}{(1-R^2)^2} - \left(\frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)^2 &= M^4 + \left(\frac{n\rho R}{1-R^2}\right)^2 \\
\left(\frac{n\rho R}{1-R^2} - \frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)^2 &= 0
\end{aligned}
$$

A3. The Sharpe Ratio and GSR for Large n

When $n \to \infty$, the Sharpe Ratio becomes unlimited if ρ is not too small:

$$
\begin{aligned}
\lim_{n\to\infty} SR^2 &= \lim_{n\to\infty} \frac{\mu_p^2}{\sigma_p^2} = \lim_{n\to\infty} \frac{\left(M^2 + \frac{n\rho R}{1-R^2}\right)^2}{M^2 + \frac{nR^2(1-\rho^2)}{(1-R^2)^2}} = \lim_{n\to\infty} \frac{M^4 + 2M^2\frac{n\rho R}{1-R^2} + \left(\frac{n\rho R}{1-R^2}\right)^2}{M^2 + \frac{nR^2(1-\rho^2)}{(1-R^2)^2}} \\
&= \lim_{n\to\infty} \frac{\left(\frac{n\rho R}{1-R^2}\right)^2}{\frac{nR^2(1-\rho^2)}{(1-R^2)^2}} = \lim_{n\to\infty} \frac{\frac{n^2\rho^2 R^2}{(1-R^2)^2}}{\frac{nR^2(1-\rho^2)}{(1-R^2)^2}} = \lim_{n\to\infty} \frac{n\rho^2}{(1-\rho^2)}
\end{aligned}
$$

Similar conclusion could not be reached for the GSR. As shown below, it increases in n provided that, approximately, $\rho > R/2$:

$$
\begin{aligned}
2\frac{n\rho R}{1-R^2} - \frac{nR^2(1-\rho^2)}{(1-R^2)^2} &> 0 \\
2\rho &> \frac{R(1-\rho^2)}{1-R^2}
\end{aligned}
$$

A4. Estimating the Worst-Case Sharpe Ratio/GSR

In this appendix we express the Sharpe Ratio and GSR at $\rho = R - \varepsilon$ as the Taylor series expansion.

$$SR^2(R-\varepsilon) = SR^2(R) + \frac{dSR^2}{d\rho}(R) * (-\varepsilon) + \frac{1}{2}\frac{d^2SR^2}{d\rho^2}(R) * \varepsilon^2 + \frac{1}{6}\frac{d^3SR^2}{d\rho^3}(R) * (-\varepsilon)^3 + ...$$

The first differential of the squared Sharpe Ratio is

$$\frac{d\left(SR^2\right)}{d\rho}(\rho) = 2\frac{\left(M^2+\frac{n\rho R}{1-R^2}\right)nR}{\left(M^2+\frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)(1-R^2)}\left(1+\frac{\left(M^2+\frac{n\rho R}{1-R^2}\right)R\rho}{\left(M^2+\frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)(1-R^2)}\right)$$

Its value at $\rho = R$ is the following:

$$\frac{d\left(SR^2\right)}{d\rho}(R) = 2\frac{nR}{(1-R^2)}\left(1+\frac{R^2}{(1-R^2)}\right)$$

Similarly, the second order derivative with respect to ρ and its value at $\rho = R$ are respectively

$$\begin{aligned}\frac{1}{2}\frac{d^2SR^2}{d\rho^2}(\rho) \quad = \quad & \frac{nR^2}{(1-R^2)^2\left(M^2+\frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)}\left[n+4\frac{\left(M^2+\frac{n\rho R}{1-R^2}\right)nR\rho}{\left(M^2+\frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)(1-R^2)}\right.\\ & \left.+4\frac{\left(M^2+\frac{n\rho R}{1-R^2}\right)^2 nR^2\rho^2}{\left(M^2+\frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)^2(1-R^2)^2}+\frac{\left(M^2+\frac{n\rho R}{1-R^2}\right)^2}{\left(M^2+\frac{nR^2(1-\rho^2)}{(1-R^2)^2}\right)}\right]\end{aligned}$$

and

$$\frac{1}{2}\frac{d^2SR^2}{d\rho^2}(R) = \frac{n}{1-R^2}\left(1+5\frac{R^2}{1-R^2}+4\frac{R^4}{(1-R^2)^2}\right)$$

We can assume that $\varepsilon \in [0, R/2]$, given that R itself is relatively small. With ε and R being of the same order of magnitude we are able to ignore the 3rd and higher order terms as insignificant. It simplifies the expression for $SR(R-\varepsilon)$ to the following:

$$SR^2(R-\varepsilon) \approx SR^2(R) + \frac{dSR^2}{dr}(R) * (-\varepsilon) + \frac{1}{2}\frac{d^2SR^2}{dr^2}(R) * \varepsilon^2$$

$$= M^2 + \frac{nR^2}{1-R^2} - 2\frac{nR}{(1-R^2)}\left(1 + \frac{R^2}{(1-R^2)}\right) * \varepsilon + \frac{n}{1-R^2}\left(1 + 5\frac{R^2}{1-R^2} + 4\frac{R^4}{(1-R^2)^2}\right) * \varepsilon^2$$

$$\begin{aligned} &\approx M^2 + \frac{nR^2}{1-R^2} - 2\frac{nR}{1-R^2} * \varepsilon + \frac{n}{1-R^2} * \varepsilon^2 \\ &= M^2 + \frac{n\,(R-\varepsilon)^2}{1-R^2} \end{aligned}$$

Similar result obtains for $GSR^2(R-\varepsilon)$ too. I.e.:

$$SR^2(R-\varepsilon) \approx GSR^2(R-\varepsilon) \approx M^2 + \frac{n\,(R-\varepsilon)^2}{1-R^2}$$

Chapter 3

Inferring Optimal Scaling of Forecasts of Expected Returns in Presence of Uncertainty About the Quality of Your Information

3.1. Introduction

The work within this chapter is a corollary to the Treynor-Black model in an idealised world.

One of the central issues in portfolio optimisation is, how to deal with the error in estimating expected asset returns. The effects of estimation error in expected returns are especially evident in a mean-variance portfolio optimisation, causing substantial fluctuations in optimal weights (for the literature review see Section 1.3).

In this chapter we investigate the ways of reducing the impact of estimation error on portfolio selection in a particular MV setting. We consider the cross-section of asset returns where assets possess predictability but up to an unknown scale function. Assuming stability in the return generating process, how can we best use historical data to infer the optimal scaling of our expected return forecasts? Based on the sample information only, the scale function is computed using a maximum likelihood estimation (MLE) method. Although it is an unbiased ordinary least squares estimate, in a small sample like ours the ML estimator cannot scale the forecasts of expected returns in a consistent way.

To filter the unknown scaling of unobservable predictable components from the sample, it is desirable to use whatever information is available, which can be formally modelled using Bayesian analysis. A review of different implementations of the Bayesian approach in the portfolio theory are given in Sections 1.4 - 1.6.

In our implementation of the Bayesian procedure, to help forecast expected returns we use the prior that comes from outcomes of unrelated experiments. Our prior is the historical forecasting skill of an investor, modelled by a distribution reflecting an investor's historical correlation between his forecasts of returns and the corresponding realisations[1]. The distributions are chosen so that they project into either normal or uniform prior of the scaling. Bayesian learning updates a prior, by incorporating observed returns into it, and turns it (the prior) into the posterior distribution of the scaling which describes the model-specific forecasting ability of the investor. Afterwards we discuss the ways of inferring the optimal scaling from the posterior distribution, and constructing the optimal portfolio.

We express the optimal amount of money put at risk as a complicated function of a prior

[1] The distribution is centered at the correlation coefficient between forecasts and realisations. Its standard deviation may be derived from the standard error of the correlation coefficient. Later in the chapter we show that it is the *existence* of uncertainty about the correlation that matters the most in portfolio decision, rather than its precise level.

on investor's forecasting skills and observed returns. We find that the optimal portfolio holdings of investors, who fully account for available sample and prior information, are based on the first two moments of the posterior distribution of correlation. When investors only partially exploit the sample and prior information, their optimal holdings are based on the mode of the ex-post distribution of correlation. Portfolios optimised without a prior depend on conventional MLE of the scaling. We regard mean-based estimates *efficient* and mode-based (with or without a prior) ones *naive*.

We examine the differences in portfolio performance arising from estimation with/without prior, uniform/normal prior, and naive/efficient use of information. The benchmark for comparisons is the MV optimal portfolio without prior, based on conventional maximum likelihood estimate (MLE). Optimal investment decisions, corresponding to the efficient use of prior, dramatically differ from those based on the MLE. Optimisation in the efficient way with prior recommends concentrated portfolio holdings. Since concentrated holdings contribute to portfolio risk, fund managers can afford not to be conservative on small holdings but should treat bigger holdings with caution. Ignoring the prior and optimising on observed returns only result in either several (up to 10) times larger holdings, or no investments at all.

Under efficient use of information, an investor in our framework is less sensitive to the luck of the draw of error. In this respect, the strategies using available information efficiently sharply contrast with the strategies based on conventional MLE.

We find that more able forecasters take larger holdings compared to less able ones. An investor, whose average prior expected value of the correlation between forecasts and returns is 15%, takes 50% larger holdings than the investor with the correlation of 10%. Despite such an increase, the more able forecaster's holdings remain up to eight times smaller than those found under the MLE-based strategy. We find that efficient portfolios under both normal and uniform prior give almost identical performance. These conclusions make it easy to test whether a fund manager applies adjustments to his forecasts or not.

The analysis is extended to a richer setting of multi-manager portfolio, where each manager works on a particular class of homogeneously predictable assets. Under the same mean-variance framework and *iid* assets, we solve the problem of optimal portfolio. We find that the optimal portfolio is a combination of optimal sub-portfolios of homogeneously scaled assets.

The papers from previous literature closely related to our work include Black and Litterman (1990, 1991, 1992), Connor (1997), and Ambachtsheer (1977). For detailed discussion of similarities between these papers and our analysis see Section 1.6.

We proceed as follows. In Section 3.2 we present the returns model and demonstrate how the conventional MLE fails to deliver a consistent estimate of unknown scaling. In Section 3.3 we introduce a prior on investor's forecasting skills and discuss how naively and efficiently derived scalings of forecasts compare to each other, and to the scaling derived using conventional MLE. The section also explains shortcomings of individual estimates. Section 3.4 solves MV optimisation in the framework of our model, expressing optimal holdings as a function of prior (if any) and observed returns. In Section 3.5 we examine the portfolio performance under different estimation strategies and draw our conclusion in favour of methods making efficient use of prior. The portfolio performance is analysed in the context of Sharpe ratio. Section 3.6 extends the analysis to a multi-manager case of non-homogeneously predictable asset classes. Section 3.7 concludes with the summary of findings and suggestions for future research. Technical details appear in Appendices B1-B8 in the end.

3.2. Model Description

Consider a single period investor who faces financial market with n independent risky assets. Assume that the asset returns x_i come from the following model:

$$\begin{aligned} x_i &= s\delta_i + \sigma\varepsilon_i \\ i &= 1, ..., n \end{aligned} \tag{3.1}$$

where $\varepsilon_i \sim N(0,1)$ is a random noise, scaled by the known volatility σ that is constant across assets. Unobservable[2] $\delta_i \sim N(0,1)$ represent the prior information on expected returns, but only up to an unknown scale function s. We are unsure what scale function should be applied.

Assuming stability in the return generating process already given, how best we use historical data to infer the optimal scaling of our expected return forecasts?

δ and ε are assumed orthogonal $\delta \perp \varepsilon$.

3.2.1. Failure of Classical MLE

Conventional maximum likelihood estimation (MLE) offers a simple formula to estimate unknown s^2 as

$$\hat{s}^2 = \frac{1}{n} \sum x_i^2 - \sigma^2 \tag{3.2}$$

Appendix B1 gives the details of derivation.

[2]The assumption of observable systematic returns δ_i would change the optimisation problem to the scaling problem as in Ambachtsheer (1977). See Section 1.6.

The unbiased ordinary least squares estimate in (3.2) however becomes biased, when conditioned on a finite sample. Due to small samples, return realisations may be such that $\frac{1}{n}\sum x_i^2 < \sigma^2$ which, according to (3.2), implies negative $\hat{s}^2$. Thereafter, the scaling is not even defined for small return variation (i.e. when $\frac{1}{n}\sum x_i^2 < \sigma^2$). Figure 3.1 illustrates our point:

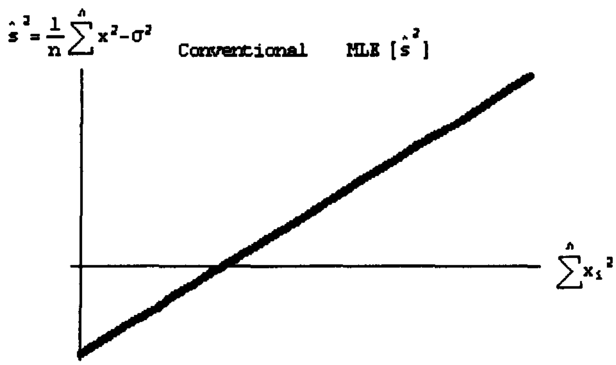

Figure 3.1

Replacing negative values of $\hat{s}^2$ by zero resolves a mathematical shortcoming of (3.2), but does not make the estimate economically meaningful. It is hard to believe that attractive investment opportunities disappear ($\hat{s} = 0$) and reappear ($\hat{s} > 0$) as soon as the average variation $\frac{1}{n}\sum x_i^2$ hits σ^2. Given a statistical similarity between δ (predictable part of returns) and ε (random noise), it is possible that small sample variation, characterising a significant portion of possible returns[3], signals an unlucky draw of error rather than no forecastable variation. Note that none of these problems would be there if either the sample size (n), or the observation interval of returns were infinite to satisfy the assumptions of MLE. For a real-life sample of returns observed over a reasonably short period of time the conventional MLE is not able to infer an optimal scaling of expected return forecasts in a consistent way. Therefore we explore other ways of estimating the scaling to obtain better forecasts.

[3]According to the tables of $\aleph^2$ distribution, with our model parameters such realisations may account for about 40% of all outcomes (see Subsection 3.3.1).

3.3. Estimating the Scaling Using a Bayesian Procedure

Subsection 3.3.1 explains how to use a prior distribution on investor's forecasting skill to predict the correlation between forecasts and returns in our framework. It also relates the scaling to the correlation and derives posterior distribution of the scaling. Subsections 3.3.2 and 3.3.3 analyse the procedure for inferring the optimal scaling from the posterior distribution in the naive and efficient ways respectively. The ex-post correlation is discussed in subsection 3.3.4.

3.3.1. Introducing a Prior on Forecasting Skills

Expected returns $s\delta_i$ are optimal forecasts of the next period's returns. The size of s therefore suggests how much we know about expected returns, or, equivalently, what the correlation between the forecasts $s\delta_i$ and realisations x_i is. The unknown correlation coefficient ρ measures explained variation in returns as given by the following expression relating ρ to the model parameters:

$$\rho = \frac{s}{\sqrt{s^2 + \sigma^2}} = Corr\,(x, \delta) \tag{3.3}$$

See Appendix B2 for rigorous derivation.

Equation (3.3) would give us s if the investor's correlation ρ were known. To help forecast the model-specific ρ, we use the prior information that reflects an investor's general forecasting potential, and update it with the returns information. The historical distribution of the correlation between the investor's forecasts and returns comes as a natural choice for the prior, since a rational investor is able to accurately specify the range of his historical forecasting skills. Prior distributions of correlation are chosen so that they map into either normal or uniform prior distribution on s via (3.3). Bayesian procedure updates a prior on s with the observed returns x_i and turns it into a posterior on the sample-specific scaling for that investor. Correctly anticipated posterior can predict what ex-post scaling *and* correlation a particular forecaster should expect for given realisations x_i.

The Normal prior $s \sim N(m, v)$ implies that the manager centers his skills at m, with a standard deviation of $\sqrt{v}$. Uniform prior $s \sim [\underline{s}, \overline{s}]$ specifies $\underline{s}$ and $\overline{s}$ as boundaries of his confidence level. Without loss of generality, $\underline{s} \geq 0$ and $m - 2\sqrt{v} \geq 0$ [4]. Both prior distributions permit for x_i being the random noise by including zero in the distribution. By choosing normal and uniform distributions as priors, we are able to examine the implications of both unbounded (normal) and bounded (uniform) priors and at the same time take advantage of the simplicity and intuitiveness of these distributions[5].

For the purposes of graphical illustration and numerical procedures we provide numeric values for the model parameters. The residual volatility σ is set at 40%. Prior scaling of forecasts of expected returns is centered at 4%, in order to get the real-life average correlation ρ between forecasts and returns of approximately 10% (see Ambachtsheer (1974), Grinold and Kahn (2000)), via $\rho = \frac{s}{\sqrt{s^2+\sigma^2}}$. With a standard deviation of 2%, the Normal prior on $s \sim N(0.04, 0.02^2)$ provides an ex-ante distribution of scaling s. Under the uniform prior, $s \sim [\underline{s}, \overline{s}]$. The choice of $\underline{s} = 0$, $\overline{s} = 0.08$ guarantees parameter consistency under uniform and normal priors.

The number of assets n is set at 100.

Next we look at the range of potential return variation for the calibrated parameters.

[4] Negative prior on s would reverse the optimal holdings found in Section 3.4.

[5] Note that the uniform prior on s is equivalent to the β-distribution prior on s^2 (for some parameters that define β-distribution).

According to the model, returns $x_i \sim N(0, s^2+\sigma^2)$ and $\frac{1}{s^2+\sigma^2}\sum x_i^2 \sim \chi_n^2$. It means that with our parameters ($\sigma = 40\%$, $n = 100$) the expected range of $\sum x_i^2$ varies roughly between 10 and 50 (in accordance with statistical tables).

Analysing how likely we are to get the returns we actually observed if our historical prior correctly anticipated the scaling s, delivers the posterior on s by the conditional Bayesian analysis[6]:

$$pdf\,(s|x) = \frac{pdf\,(x|s)\,pdf\,(s)}{\int pdf\,(x|s)\,pdf\,(s)\,ds}.$$

Posterior distribution on s gives the investor's updated beliefs on what scaling to expect. Figures 3.2 and 3.3 give the plots of posterior distributions for several levels of variation in realised returns.

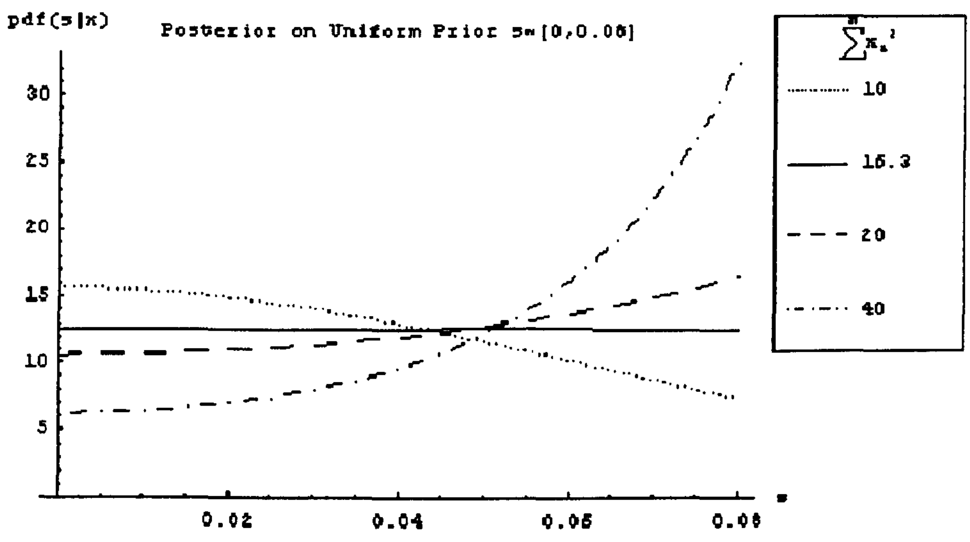

Figure 3.2

[6] x is a vector of observed returns x_i.

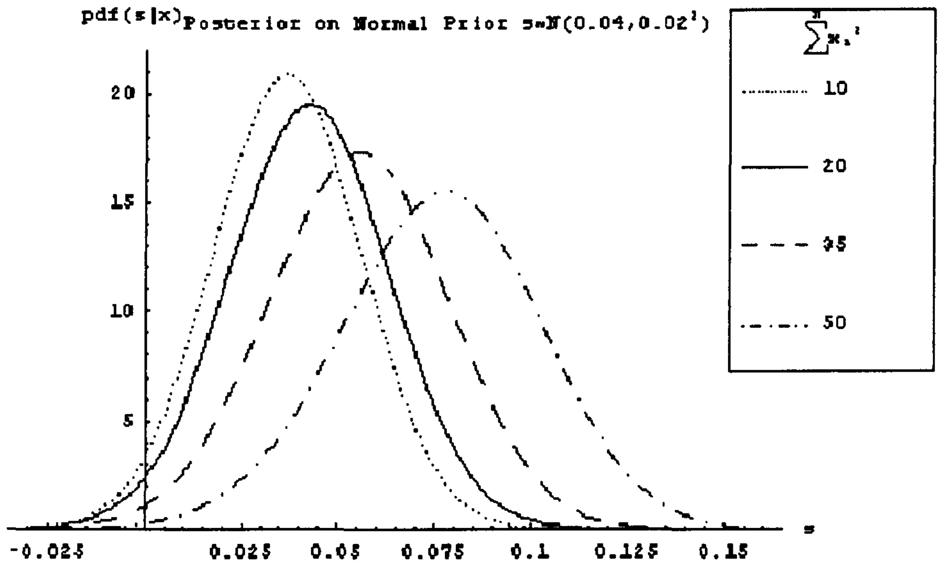

Figure 3.3

3.3.2. Estimating the Scaling with Partial Use of Information

One estimate of s, that comes from the posterior distribution on scaling, is the mode of s's posterior distribution. We call this function of the prior scaling and observed returns *the modified maximum likelihood estimate with corresponding prior* and label it $MMLE[s|Prior]$ as applicable.

Both modified maximum likelihood estimates $MMLE\,[s|Unif]$ and $MMLE\,[s|Norm]$[7] are presented in Figure 3.4, next to $MLE\,[s]$:[8]

[7]See Appendix B4 for details.

[8]The benchmark $MLE[s]$ is $\sqrt{MLE[s^2]}$, which was derived earlier in Subsection 3.2.1. Negative values of $MLE[s^2]$ are replaced by zero.

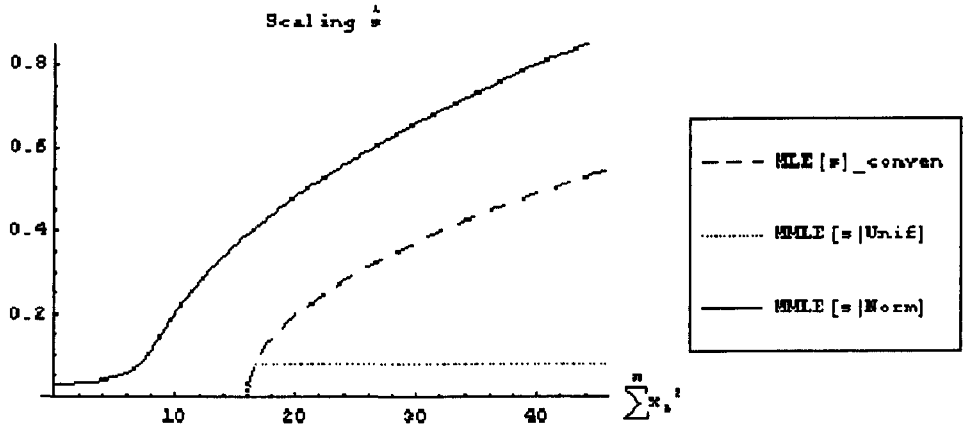

Figure 3.4

$MMLE\left[s|Unif\right]$, given by the formula

$$MMLE\left[s|Unif\right] = \min\left(\sqrt{\max\left(\frac{1}{n}\sum x_i^2 - \sigma^2, \underline{s}\right)}, \overline{s}\right),$$

equals zero for $\frac{1}{n}\sum x_i^2 < \sigma^2$ (conventional MLE of s was not even defined there). With the increase in return variation, it soon reaches the upper boundary of the manager's confidence level $\overline{s}$ to which $MMLE\left[s|Unif\right]$'s upside potential is limited. The behaviour of $MMLE\left[s|Unif\right]$ is explained by s's posterior distribution, shown in Figure 3.2. Here the mode is placed on the boundaries of the prior for almost all return realisations. It would be foolish to believe that the optimal estimate gains in value as much as $\left[\overline{s} - \underline{s}\right]$ within a unit growth in the realised return variation around σ^2. Therefore we have sufficient grounds to question the credibility of $MMLE\left[s|Unif\right]$.

Another estimate of s, the posterior MLE under normal prior denoted $MMLE\left[s|Norm\right]$, is excessively optimistic (see Figure 3.4). Notably, it keeps the scaling of ex-post returns positive even for the smallest return variation of zero. The plots of the posterior in Figure 3.3 confirm that $MMLE\left[s|Norm\right]$ is overly sensitive to growth in return variation. Explaining this much variation in the financial market (the scaling of 50% corresponds to the correlation of 90% when $\sigma = 40\%$) is beyond the rational expectations of any investor. Hence we deem $MMLE\left[s|Norm\right]$ too inadequate.

The inability of the estimates considered above to deliver reasonable performance shows the inadequacy of their estimation method in our framework of a finite sample. Despite using Bayesian learning, $MMLE\left[s|Norm\right]$ and $MMLE\left[s|Unif\right]$ remain essentially *naive* or single-point estimates (like the conventional maximum likelihood estimate $MLE[s]$) and are inconsistent when conditioned on small samples. We argue, these methods fail because they make only partial use of information. In the rest of the chapter, we refer to the $MMLE\left[.|Prior\right]$'s as the *naive* estimates.

3.3.3. Estimating the Scaling with Full Use of Information

It has been understood that a mean of a distribution is a more informative statistic than its mode.

Here we look at an alternative estimate of ex-post scaling, the mean of the posterior on s, calculated as

$$E\left[s|x, Prior\right] = \int s * pdf\left(s|x\right) ds. \tag{3.4}$$

Unlike the mode estimates considered before, posterior means account for the entire posterior distribution $pdf\left(s|x\right)$ and are less affected by individual outcomes in the posterior. Depending on the prior they use, we denote posterior mean estimates $E\left[s|Unif\right]$ and

$E[s|Norm]$[9] and express them as follows:

$$E[s|Unif] = \frac{\int_{\underline{s}}^{\overline{s}} s \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] ds}{\int_{\underline{s}}^{\overline{s}} \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] ds}$$

$$E[s|Norm] = \frac{\int_{-\infty}^{\infty} s \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] e^{-\frac{(s-m)^2}{2v}} ds}{\int_{-\infty}^{\infty} \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] e^{-\frac{(s-m)^2}{2v}} ds}$$

For detailed derivation, see Appendix B5.

Figure 3.5 below plots the mean estimates of scaling next to the benchmark MLE-based scaling:

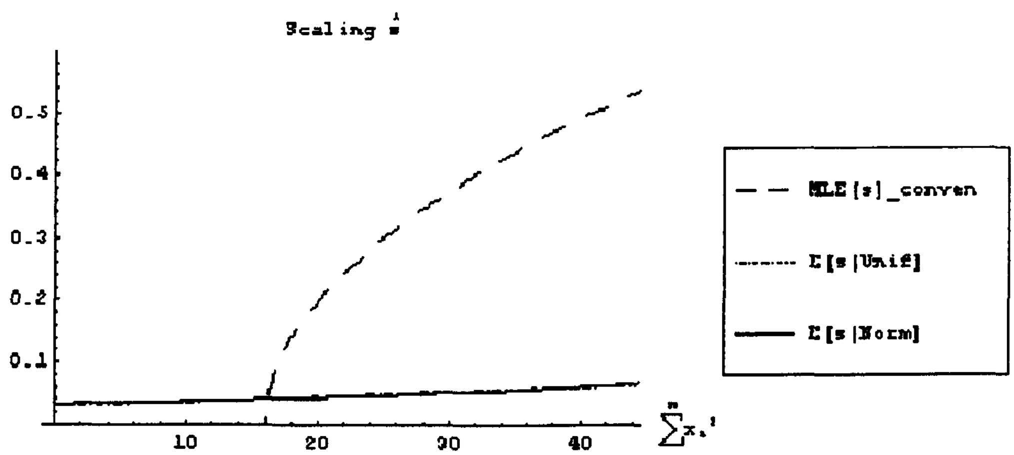

Figure 3.5

Figure 3.5 suggests that both posterior-mean estimates efficiently solve the problem of an unlucky draw of error, implying the presence of predictable returns even for small $\sum x_i^2$. Moreover, the mean estimates remain moderate throughout different return realisations and smoothly and slowly increase with growing return variation.

[9]For simplicity, we leave out *conditioning on x returns* in the notation of $E[s|Unif]$ and $E[s|Norm]$, as well as $MMLE[s|Norm]$ and $MMLE[s|Unif]$.

More consistent performance of the mean-based ex-post scaling (i.e. $E[s|Prior]$) compared to the mode-based (i.e. $MMLE[s|Prior]$ and $MLE[s]$) scaling is explained by the fact that the mean summarises all the available information, while the mode only partially exploits it. We regard the posterior mean-based estimates as *efficient.*

3.3.4. Ex-Post Correlation between Forecasts and Returns

To estimate what ex-post correlation an investor should expect between his forecasts and returns in the framework of our model, we use the distribution of ex-post scaling and formula (3.3), derived earlier, that establishes a relationship between scaling and correlation $\rho = \frac{s}{\sqrt{s^2+\sigma^2}} = Corr(x, \delta)$.

Correlation estimates $\hat{\rho}$, a function of scaling $\hat{s}$, differ across different methods of estimation but due to (3.3), the shapes of individual $\hat{\rho}$'s are similar to the shape of their corresponding $\hat{s}$, as confirmed by Figure 3.6. For consistency, in the rest of the chapter we use similar notation for estimates $\hat{s}$ and $\hat{\rho}$ that correspond to each other. For instance, $E[\rho|Unif]$ denotes the correlation corresponding to $E[s|Unif]$.

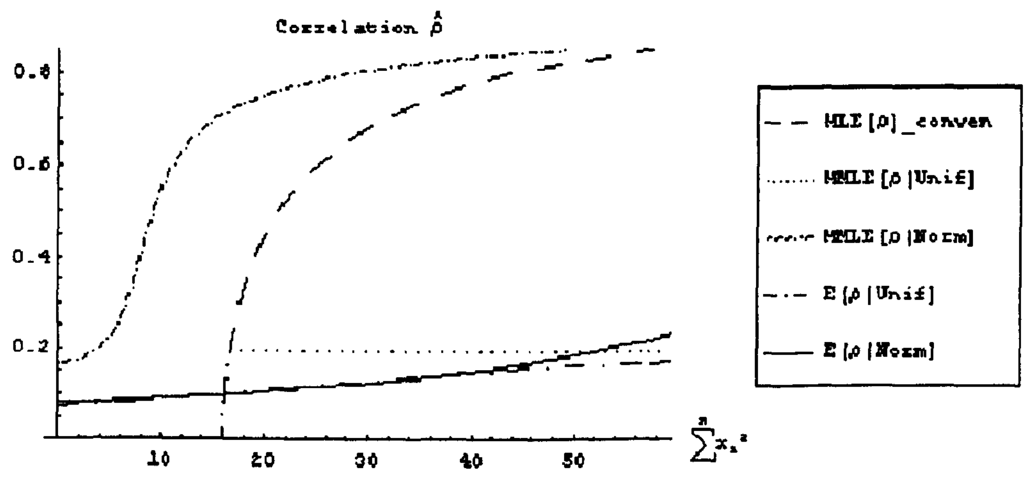

Figure 3.6

As expected, conventional $MLE[\rho]$ and $MMLE[\rho|Normal]$ are both explosive and promise unbelievably high correlations. The uniform prior-based $MMLE[\rho|Unif]$ too is inadequate — zero correlation is almost instantly followed by the largest correlation allowed under this scheme, which stays constant for subsequent growth in $\sum x_i^2$.

On the other hand, $E[\rho|Unif]$ and $E[\rho|Norm]$ both start off at a reasonable level and slowly increase with rising $\sum x_i^2$ — a likely signal of improving investment opportunities. However their paths diverge at some point — the uniform prior-based correlation $E[\rho|Unif]$ becomes practically insensitive to further growth in the return variation as it asymptotically approaches $MMLE[\rho|Unif]$, while the average ex-post correlation under normal prior $E[\rho|Norm]$ keeps increasing. This shortcoming of $E[\rho|Unif]$ implied by uniform prior is not likely to affect its real-life performance. Given limited sample variation in returns, $\sum x_i^2$ is not expected to grow so much that the difference between uniform and normal prior-based estimates becomes significant. It is worth noting that the efficiently estimated correlation under normal prior $E[\rho|Norm]$ is free from the kind of deficiencies described above.

3.4. The Optimal Portfolio Solution

How does the optimal portfolio depend on the posterior distribution of s? It is unlikely to use just the mode or the mean that we have considered so far. In Section 3.3 we discussed different ways in which an investor may estimate his ex-post correlation between forecasts and returns. Below, for each estimate we present the portfolio weights that maximise the expected MV utility of the portfolio.

Given $\rho = \frac{s}{\sqrt{s^2+\sigma^2}}$, the parameters of classical mean-variance optimisation are expressed as

$$
\begin{aligned}
E\left[R_p|x\right] &= \int w'x\frac{s}{\sqrt{s^2+\sigma^2}} * pdf\left(s|x\right)ds \\
&= \int w'x\rho * pdf\left(s|x\right)ds = E\left[\rho|Prior\right]\omega'x \\
Var\left[R_p|x\right] &= Var\left[\rho|Prior\right]\left(xx'\right)\left(w'w\right) + \sigma^2\left(w'w\right)
\end{aligned}
$$

where w is a vector of optimal holdings, x — a vector of realised returns.

The expressions given above lead to the following maximisation problem, under corresponding prior:

$$
\max_{w} \; E\left[\rho|Prior\right]w'x - \frac{\gamma}{2}\left[Var\left[\rho|Prior\right]\left(xx'\right)\left(w'w\right) + \sigma^2\left(w'w\right)\right] \tag{3.5}
$$

where γ is a coefficient of risk aversion (see detailed derivation of (3.5) in Appendix B6). The optimal stock allocation does not depend on the value of wealth (a feature of CARA utility functions). Optimal holding of the ith asset under efficient use of information is denoted $w\_efficient_i$ and equals the following function of corresponding prior and observed returns[10]:

$$
w\_efficient_i = \frac{E\left[\rho|Prior\right]}{Var\left[\rho|Prior\right]\sum x_i^2 + \sigma^2}\frac{x_i}{\gamma} \tag{3.6}
$$

where

$$
\begin{aligned}
E\left[\rho|Prior\right] &= \frac{\int_{-\infty}^{\infty}\frac{s}{\sqrt{s^2+\sigma^2}}f\left(s\right)}{\int_{-\infty}^{\infty}f\left(s\right)ds}ds \\
Var\left[\rho|Prior\right] &= \frac{\int_{-\infty}^{\infty}\frac{s^2}{s^2+\sigma^2}f\left(s\right)}{\int_{-\infty}^{\infty}f\left(s\right)ds}ds - \left(\frac{\int_{-\infty}^{\infty}\frac{s}{\sqrt{s^2+\sigma^2}}f\left(s\right)}{\int_{-\infty}^{\infty}f\left(s\right)ds}ds\right)^2
\end{aligned}
$$

[10]Details appear in Appendix B6.

where $f(s)$ under the normal prior $s \sim N(m,v)$ and uniform prior $s \sim [\underline{s}, \overline{s}]$ is respectively

$$\begin{aligned} \text{normal:} \quad & f(s) = \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] e^{-\frac{(s-m)^2}{2v}} \\ \text{uniform:} \quad & f(s) = \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \end{aligned}$$

An alternative set of solutions is offered for an investor, who uses information in the naive way (i.e. uses the mode-based estimates $MLE[\rho]$, $MMLE[\rho|Unif]$, $MMLE[\rho|Norm]$). He ignores the estimation error and a point-estimate $\hat{\rho}$ is treated as the true value. As a result, instead of (3.5) he maximises

$$\max_w \ \hat{\rho} w'x - \frac{\gamma}{2}\sigma^2 (w'w)$$

and chooses the following holdings (labeled w_naive_i to keep consistency with other naively derived estimates)[11]:

$$w\_naive_i = \frac{\widehat{\rho}}{\sigma^2}\frac{x_i}{\gamma} \tag{3.7}$$

which in the case of conventional MLE-based scaling becomes

$$w\_conven_i = \frac{x_i}{\gamma\sigma^2}\frac{MLE[s]}{\sqrt{MLE[s]^2+\sigma^2}}$$

It is clear that (3.7) is a special case within (3.6). Equation (3.6), the expression for $w\_efficient_i$, is the main finding of this section which states: To fully use the available information, we only need the mean and the variance of the posterior distribution on scaling.

[11]See details in Appendix B6.

3.5. Implications of the Optimal Portfolio

3.5.1. Comparing Different Portfolio Strategies

Figure 3.7, given next, presents the optimal weights under different parameter estimation strategies, when x_i varies in proportion with the average variation in sample, i.e. $x_i = \frac{1}{n}\sqrt{\sum x_i^2}$.

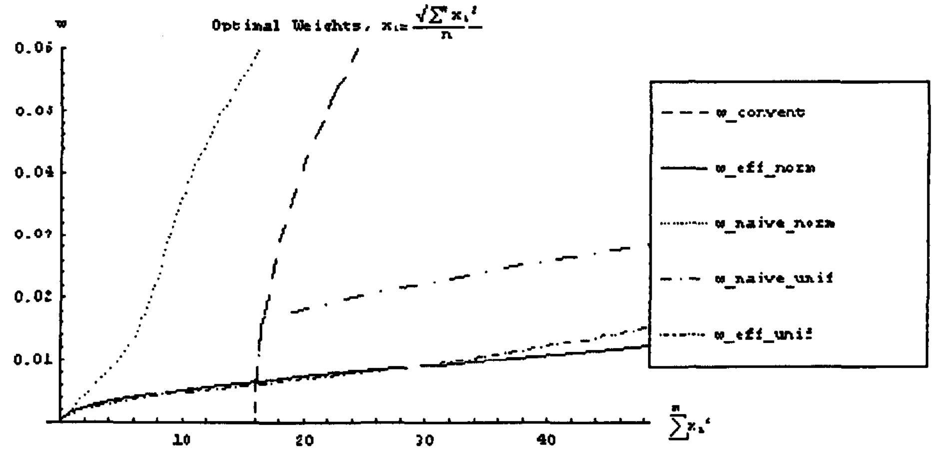

Figure 3.7

Weights based on full Bayesian posterior are strikingly different to those from point estimates (with or without prior). All three point-estimate-based weights vary largely subject to return variation, while the weights based on full posterior remain moderate throughout return realisations.

In particular, benchmark MLE-based w_conven, like w_naive_unif, is zero until the average variation $\frac{1}{n}\sum x_i^2$ reaches σ^2. Immediately afterwards, the conventional strategy recommends large-size investments in risky assets. The holdings corresponding to naive estimator with normal prior are overly optimistic. Weights as fluctuating as these in the presence of moderate prior on skills, provide evidence of the inadequacy of the corresponding estimation methodology.

Meanwhile, within the range of our interest, portfolio holdings under efficient strategies with normal and uniform prior closely track each other and gradually but steadily incorporate signals on better returns. Optimal holdings under efficient use of information are several times smaller than those recommended by the benchmark MLE-based strategy, as confirmed by Figure 3.8.

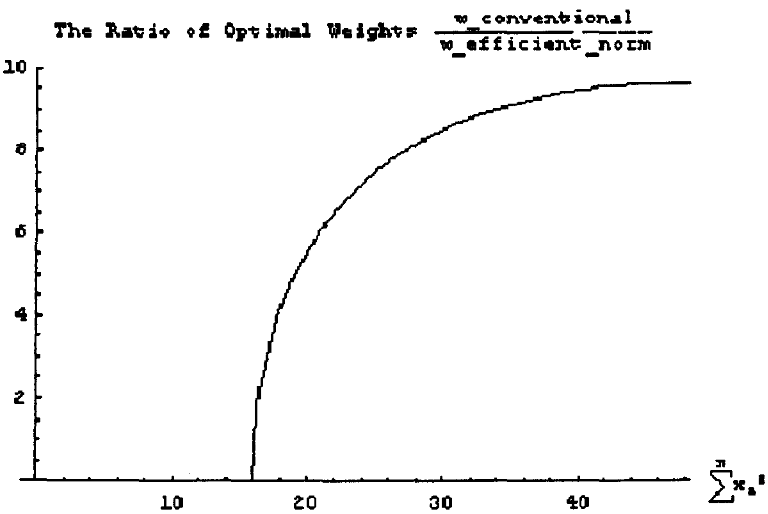

Figure 3.8

Moderate holdings of the efficient strategies guard us against sample error. This way a manager adjusts against underestimating investment opportunities when $\sum x_i^2$ is small (possibly due to the unlucky draw of error), as well as against overestimating investment opportunities when a large $\sum x_i^2$ may not be a signal of superior returns but merely the sample error. When the information is used efficiently in our framework, investors choose concentrated holdings. As these contribute to the portfolio risk, fund managers can afford not to be conservative on small holdings but should treat larger holdings with caution. Ignoring the prior and optimising on observed returns only, can result in either several times (up to ten) as extreme investments as when optimising with efficient use of prior, or no investments at all.

The flip-side of our findings regarding efficient money allocation strategy is using actual investment levels as a performance criteria to select fund managers. Regardless of the manager's confidence level, risk aversion kept equal, it is easy to distinguish between the naive and efficient strategies he may follow — the efficient way of optimisation always results in conservative holdings.

3.5.2. Analysing Portfolio Performance Using the Sharpe Ratio

We have now seen how the way in which information is used affects the portfolio weights. In this section we will look at how it affects the portfolio performance, as measured by the Sharpe ratio.

The Sharpe ratio of portfolios selected under efficient use of information, is expressed as follows:

$$SR_{efficient} = \frac{E\left[\rho|Prior\right]}{\sqrt{Var\left[\rho|Prior\right]\sum x_i^2 + \sigma^2}}\sqrt{\sum x_i^2} \tag{3.8}$$

With point estimates (i.e. $MLE[.]/MMLE[.|.]$-based estimates) of ρ, the above formula changes to

$$SR_{conven\ /\ naive} = \frac{\widehat{\rho}}{\sigma}\sqrt{\sum x_i^2} \tag{3.9}$$

where $\widehat{\rho}$ is one of $MLE[.]$ or $MMLE[.|Prior]$ with corresponding prior. See details in Appendix B6.

Formulae (3.8) and (3.9) measuring the portfolio performance are similar to the optimal holdings formulae (3.6) and (3.7), hence the similarity between Figure 3.9, displaying the Sharpe ratios, and Figure 3.7, displaying the optimal portfolio weights.

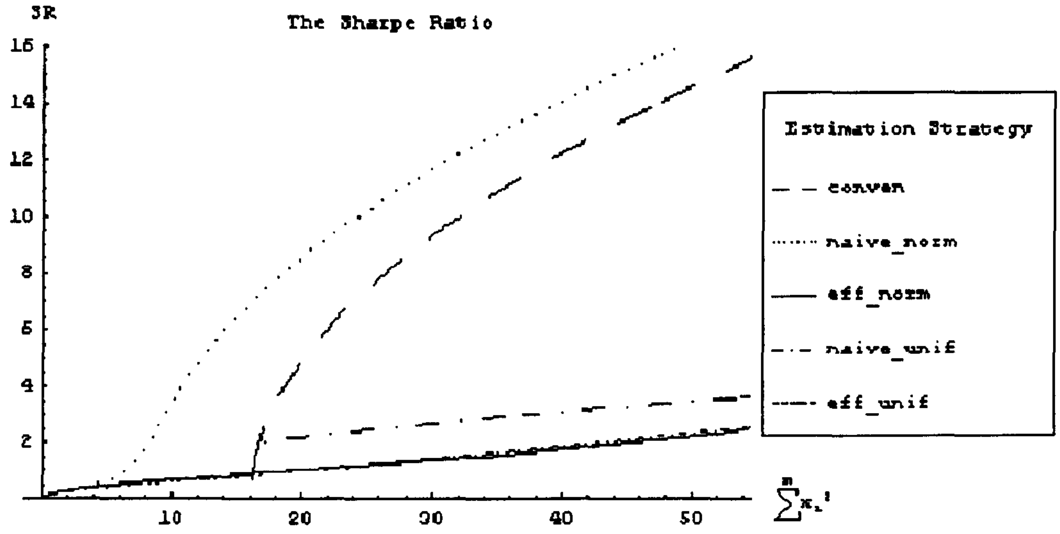

Figure 3.9

Because the Sharpe ratio measures expected portfolio mean with respect to associated volatility, portfolios with proportional mean-volatility attain equal Sharpe ratios. That is why this measure of portfolio performance ranks naive and conventional MLE-based strategies higher than the efficient ones and is not able to spot the efficiency loss attributed to exaggerated holdings in risky assets. However, the sizes of the Sharpe ratio achieved by partial use of information are mostly suspiciously high. A rational forecaster cannot expect to get such a high Sharpe ratio (as in Figure 3.9), and therefore cannot trust these strategies. On the other hand, portfolios derived via accounting for prior information in the efficient way, offer a much lower but reasonable-size Sharpe ratio that grows adequately with the variation in the data.

The true Sharpe ratio, meanwhile, remains indifferent to estimation methods and equals $SR_{true} = \frac{s}{\sigma}\frac{\sum x_i \delta_i}{\sqrt{\sum x_i^2}}$, due to homogeneous predictability of all assets assumed in the model. The assumption will be relaxed in Section 3.6.

3.5.3. Implications of the Assumed Distribution: Uniform Versus Normal Prior

This section examines marginal differences in holdings implied by normal Vs uniform prior, as well as the impact of changing the parameters of assumed distributions. We consider optimal weights derived under full use of information. As noted earlier, the optimal weights as well as the Sharpe ratio are almost identical under both priors. We vary the parameters as follows. In the already analysed uniform distribution $s \sim [\underline{s} = 0, \overline{s} = 0.08]$, we shift $\overline{s}$ by ± 0.02; $\underline{s}$ stays fixed at zero to allow for no predictability in the sample. In the normal $s \sim N(m = 0.04, v = 0.02^2)$ the mean changes to 0.03 and 0.05.[12]

As illustrated in Figure 3.10, the efficient weights, corresponding to different levels of $\overline{s}$ in the uniform prior, move very close to each other at the start but diverge widely afterwards.

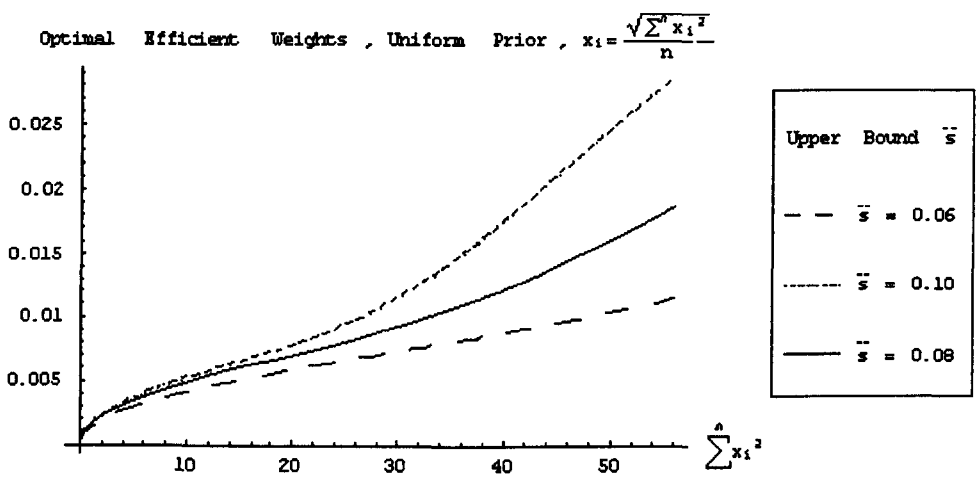

Figure 3.10

Efficient weights for normal priors with different means, on the other hand, diverge from the very start and this divergence between them increases proportionally with the return

[12]Altering v does not give any extra insight into the behaviour of asset holdings. Therefore we keep $\sqrt{v}$ fixed at 0.02.

variation. (See Figure 3.11.) Overall, a smaller confidence level (lower m and $\underline{s}$) results in a flatter curve of optimal weights for both priors.

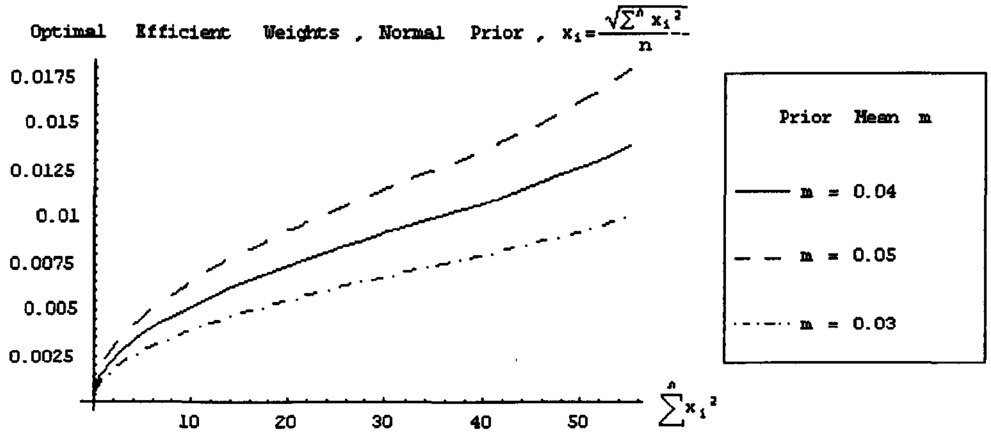

Figure 3.11

Such behaviour agrees with our expectations regarding normal and uniform priors. Someone, well aware of his forecasting skill, projects its average level into the mean of a normal prior distribution, with due uncertainty. This way the marginal difference in holdings between more and less able forecasters is proportional to their holdings at all times. Uniform prior, on the other hand, is less informative compared to the normal and equally favours all of its possible realisations within a specified range. Thus, under uniform prior the performance of less confident forecasters is similar to that of more confident forecasters, unless the return variation is quite large. During large $\sum x_i^2$ better forecasting means more bullish investments. It is another confirmation of our early conclusion that larger positions (implied by large $\sum x_i^2$) should be treated with caution.

Figure 3.12 plots the ratios of efficiently derived weights under normal and uniform priors for the three pairs of priors considered above. Throughout the uncertainty $\sqrt{v}$ remains constant at 2% and $\underline{s} = 0$.

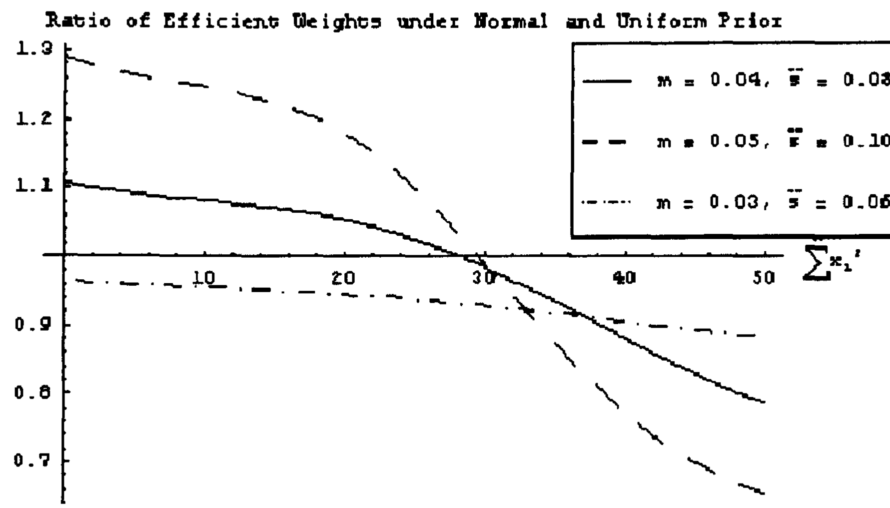

Figure 3.12

The ratio corresponding to our default parameters ($m = 0,04$, $\overline{s} = 0.08$) shows the most balanced performance of the weights derived efficiently under different priors. The balance changes with varying parameters of two distributions. Depending on the size of return variation, either normal or uniform prior gives slightly more aggressive holdings.

Because the scaling and correlation are so small, we are in no position to favour either prior. Nor have we aimed to provide their ranking. We have just pointed out several differences between portfolio decisions arising due to a prior choice, which an investor should be aware of when making distributional assumptions. However, regardless of the minor inconsistencies, efficiently derived optimal holdings with either prior with a reasonable degree of confidence stay several (about eight) times smaller than holdings under the conventional MLE. Thereafter, in terms of reducing the impact of the estimation error, the superiority of the efficient use of information with prior over conventional ways of optimisation is beyond any doubt, for either normal or uniform priors.

3.5.4. Effects of Uncertainty on Weights and the Sharpe Ratio

Earlier in the chapter we derived that naive/conventional investors are more optimistic about their portfolios than those efficiently using available information. The following partly explains such performance. Investors, using MLE/MMLE-based methods, assume they know their skills (the correlation ρ) to a single estimate although their expectation may well be misleading. Efficient optimisation, to the contrary, allows for parameter uncertainty (uncertainty in ρ, in this case) which lowers corresponding ex-post correlation and, consequently, portfolio holdings and expected utility to reasonable levels.

Measuring the uncertainty around forecasting skills may be problematic and an investor may wish to ignore it. Then the formula of optimal weights changes from (3.6) to the following ($w\_effic\_approx_i$ denotes the efficient portfolio weight without uncertainty about ρ):

$$w\_efficient_i \approx w\_effic\_approx_i = \frac{E\left[\rho|Prior\right]}{\sigma^2}\frac{x_i}{\gamma} \tag{3.10}$$

Note that (3.10) is similar to naive portfolio weights (3.7) but here the maximum likelihood estimate is replaced by the posterior mean.

Although portfolio holdings increase once uncertainty is ignored as in (3.10), these holdings remain substantially smaller than their counterparts based on the MLE. It is straightforward that the ratio of optimal weights (as in (3.6)) to approximately optimal weights (as in (3.10)) is slightly less than 1:

$$\frac{w\_efficient_i}{w\_effic\_approx_i} = \frac{\sigma^2}{Var\left[\rho|Prior\right]\sum x_i^2 + \sigma^2} \approx 1$$

The implication of the above is that, if an investor has to choose a point estimate for correlation ρ, he would be better off using a posterior mean estimate $E[\rho|Prior]$ instead of the maximum likelihood estimates, i.e. $MLE[\rho]$ and $MMLE[\rho|Prior]$.

3.6. Generalisation to Multi-Manager Portfolio

Up to now we have been studying the problem of inferring the optimal scaling in a setting of homogeneously predictable assets. Next we extend the setting to several asset classes of likewise predictable assets where each class possesses its own degree of predictability but again up to an unknown scaler (scale functions differ from one asset class to another). This is equivalent to combining efforts of several, say m, managers with the jth manager making a portfolio decision for the jth class of n_j assets (x_{ij}, $i = 1, ..., n_j$, $j = 1, ..., m$) for which he possesses some forecasting ability ρ_j, $j = 1, ..., m$. The overall setting remains mean-variance as before, with no short-sales restriction and all assets being *iid*. In notation,

$$\begin{aligned} x_{ij} &= s_j \delta_{ij} + \sigma_j \varepsilon_{ij} \\ i &= 1, ..., n_j;\ j = 1, ..., m \end{aligned} \tag{3.11}$$

where $\varepsilon_{ij} \sim N(0,1)$ is random noise, scaled by the known volatility σ_j that is constant within the jth class of assets. As before, unobservable $\delta_{ij} \sim N(0,1)$ represent the prior information on expected returns, but only up to an unknown scale function s_j which is constant for each class.

Under both naive and efficient ways of using the information, optimal asset positions, chosen by individual managers for their own class of assets, will simply be combined to form an optimal portfolio of all assets considered. The risk aversion γ is assumed constant for all managers. The following formulae of asset positions in the optimal portfolio, $w\_efficient\_gen_{ij}$ and $w\_naive\_gen_{ij}$ under efficient and naive use of information respectively, are straightforward generalisations of the similar formulae for a single-class portfolio[13]:

$$w\_efficient\_gen_{ij} = \frac{E\left[\rho_j|Prior\right]}{Var\left[\rho_j|Prior\right]\sum_i^{n_j} x_{ij}^2 + \sigma_j^2}\frac{x_{ij}}{\gamma}$$

$$\begin{aligned} w\_naive\_gen_{ij} &= \frac{1}{\gamma\sigma_j^2}\widehat{\rho}_j x_{ij} \\ i &= 1,...,n_j;\ j = 1,...,m \end{aligned}$$

The following are the Sharpe ratios of the combined portfolio as expected by managers:

$$\begin{aligned} SR_{efficient\_gen} &= \sqrt{\sum_j^m \frac{E^2\left[\rho_j|Prior\right]}{Var\left[\rho_j|Prior\right]\sum_i^{n_j} x_{ij}^2 + \sigma_j^2}\sum_i^{n_j} x_{ij}^2} \\ SR_{naive\_gen} &= \sqrt{\sum_j^m \frac{\widehat{\rho}_j^2}{\sigma_j^2}\sum_i^{n_j} x_{ij}^2} \end{aligned}$$

Meanwhile, the true Sharpe ratios in the case of several managers with different forecasting skills become

$$SR_{efficient\_gen\_true} = \frac{\sum_j^m \frac{E\left[\rho_j|Prior\right]}{Var\left[\rho_j|Prior\right]\sum_i^{n_j} x_{ij}^2+\sigma_j^2} s_j \sum_i^{n_j} x_{ij}\delta_{ij}}{\sqrt{\sum_j^m \frac{E^2\left[\rho_j|Prior\right]}{\left(Var\left[\rho_j|Prior\right]\sum_i^{n_j} x_{ij}^2+\sigma_j^2\right)^2}\sigma_j^2 \sum_i^{n_j} x_{ij}^2}} \quad (3.12)$$

$$SR_{naive\_gen\_true} = \frac{\sum_j^m \frac{\widehat{\rho}_j}{\sigma_j^2} s_j \sum_i^{n_j} x_{ij}\delta_{ij}}{\sqrt{\sum_j^m \frac{\widehat{\rho}_j^2}{\sigma_j^2}\sum_i^{n_j} x_{ij}^2}} \quad (3.13)$$

[13]All formulas in this section are derived in Appendix B8.

which are apparently distinct from each other, unlike the true Sharpe ratio for homogeneously predictable assets which was constant at $\frac{s}{\sigma}\frac{\sum x_i\delta_i}{\sqrt{\sum x_i^2}}$ for all analysed ways of estimation.

3.7. Conclusion

In this chapter we examine how to infer the optimal scaling of forecasts of expected returns in presence of uncertainty about the quality of our information.

We consider a model of returns where returns possess predictability but up to an unknown scale function. Predictable parts of returns are unobservable. Assuming stability in the return generating process, we investigate how best we can use historical data to infer the optimal scaling of the expected return forecasts.

The conventional maximum likelihood estimation (MLE) is one way of inferring the scaling but in small samples like ours this method is largely affected by the estimation error and fails to deliver an adequate scaling. To help forecast the scaling, we turn to outside-of-sample information to get Bayesian prior. For the prior we use an investor's forecasting skills, modelled as a distribution of his historical correlation between return forecasts and realisations. Bayesian learning updates the prior with observed returns. The posterior distribution on correlation helps us solve the mean-variance optimisation problems of an investor. We consider two alternative ways of estimating true correlation: one is via partial use of information — based on the MLE of the posterior (called naive estimates with uniform/normal prior); the other is via efficient use of information — based on the mean of the posterior distribution (respectively called efficient estimates with uniform/normal prior). Naive estimates behave inadequately and are largely sensitive to return variation. The efficient estimates, on the other hand, give reasonable and balanced performance. These diversities are explained by the ability of a mean to summarise the entire distribution in one number much better than is done by a mode.

We solve the MV portfolio selection problem and express the optimal amount of money put at risk as a complicated function of a prior and observed returns.

Investment decisions, corresponding to efficient strategies, dramatically differ from those based on the MLE method. MLE-based portfolios are overly optimistic and their holdings are several (up to ten) times larger than the holdings under efficient strategies. Efficient strategies favour concentrated holdings. Since concentrated holdings contribute to the portfolio risk, an investor can afford not to be conservative on small holdings but should treat larger holdings with caution.

Furthermore, during the times of low realised returns the efficient portfolio strategy still recommends investment in risky assets allowing for the possibility that the small variation may be due to the unlucky draw of errors, which overshadows the predictive variation of returns. This is contrary to the conventional MLE-based advice, which is not even defined for small return variation. Meanwhile, so called 'small variation' constitutes a large portion of all return realisations.

Investors with better forecasting skills have larger holdings and therefore higher Sharpe ratio. Despite the increase in a holding arising due to better forecasting skills these holdings remain several times smaller when compared to the holdings suggested by conventional MLE strategies. Therefore, regardless of assumed forecasting skills, the size of efficiently derived portfolio holdings should remain moderate. This finding can help rank fund managers' performance, and select them. Differences in portfolio holdings, arising due to assuming normal or uniform prior on scaling, are negligible. Although we highlight a number of deficiencies associated with the uniform prior, in our framework of limited return variation and moderate scaling both give almost identical results.

The findings are generalised to a non-homogeneous setting when informative parts of assets are not being scaled by the same risk factor. We solve the mean-variance portfolio selection problem in the presence of several categories of assets when assets in each category are equally predictable. This is equivalent to considering a multi-manager scenario, where each manager works on one class on assets and their efforts are combined to form optimal portfolio. Provided that all assets are *iid*, optimal weights are the same as they would be for 'sub-portfolios' made up of assets within individual categories. The expressions of the Sharpe ratio however change and the true Sharpe ratio becomes dependant on how efficiently the information is being used. This result differs from the case of homogeneously scaled assets, where the true Sharpe ratio was constant for all ways of optimisation because all assets were scaled equally.

Throughout the analysis, we restricted ourselves to positive correlation coefficients between forecasts and returns. Allowing for negative correlation would simply reverse the holdings. Also we assumed that systematic returns δ_i were unobservable. If they were observable, the problem would change to finding their optimal scaling following the same Bayesian approach. This kind of posterior scaling would be an improvement over the existing work on scaling the forecasts by Ambachtsheer (1977) who suggests a point estimate to scale known alpha (See Section 1.6 for more details).

Our model is similar to the Black and Litterman model (1990, 1991, 1992) but the implementation of our model is much more straightforward. Connor (1997) takes an approach similar to ours in refining model parameters rather than the forecasts produced by this model. Section 1.6 compares our work with theirs in detail.

An interesting question arising from this work is, to what extent can we use external information to predict model-specific-parameters? We assumed that investors have the distributional knowledge of their forecasting skills. It will be interesting to investigate the utility loss due to misestimation of forecasting skills. Modelling of utility loss caused by misestimation is left for future research.

Appendix B

B1. Classical MLE

Consider the classical MLE estimate of the parameter s. In (3.1) the density will be:

$$\begin{aligned} L &= \frac{1}{\left(\sqrt{\sigma^2+s^2}\sqrt{2\pi}\right)^n}\exp[-\frac{\sum x_i^2}{2\left(\sigma^2+s^2\right)}] \\ \ln L &= -\frac{n}{2}\ln\left(\sigma^2+s^2\right)-\sum\frac{x_i^2}{2\left(\sigma^2+s^2\right)} \end{aligned}$$

FOC:

$$\frac{\partial \ln L}{\partial s} = -\frac{ns}{\sigma^2+s^2}+\frac{s}{\left(\sigma^2+s^2\right)^2}\sum x_i^2 \equiv 0$$

which results in (3.2) :

$$\hat{s}^2 = \frac{1}{n}\sum x_i^2 - \sigma^2$$

Note that (3.2) is an unbiased OLS estimate. However due to sampling error, $\hat{s}$ may go negative if the realisations are such that

$$\frac{1}{n}\sum x_i^2 < \sigma^2$$

It is obvious that such a possibility exists in a finite sample of asset returns.

B2. Relating Forecasts and Returns

Assume that $g(.,.)$ is a function that takes observations x_i and returns optimal forecasts for the next period's return $s\delta_i$. In other words, $g(s, x_i)$ is an estimate $\widehat{(s\delta_i)}$:

$$\widehat{(s\delta_i)} = g(s, x_i), \ i = 1, ..., n. \tag{3B.1}$$

This also means, that $g(s, x_i)$ extracts means from sample observations.

We denote the next period's return for the ith asset that is to be estimated y_i, next period's error realisation – u_j. (We bring in new variables in order to avoid adding time subscripts to those already in use.)

$$y_i = g(s, x_i) + \sigma u_i$$

Using the Taylor series expansion,

$$\begin{aligned} \widehat{(s\delta_i)} &= g(s, x_i) = g(s, s\delta_i + \sigma u_i) \\ &\approx g(s, s\delta_i) + \sigma u_i g'(s, s\delta_i) + \frac{\sigma^2 u_i^2}{2} g''(s, s\delta_i) \\ i &= 1, ..., n. \end{aligned} \tag{3B.2}$$

Assume $g(.,.)$ is a linear function of its second argument. That means,

$$g(s, s\delta_i) = a_0 + a_1(s\delta_i), \ i = 1, ..., n. \tag{3B.3}$$

In equality $(3B.2)$ sample variations of two sides are equal. Below we calculate them, considering that $g(s, x_i)$ is linear:

$$\begin{aligned} Var[s\delta] &= s^2; \\ Var[g(s, s\delta_i)] &= a_1^2 s^2 \\ Var[\sigma u_i g'(s, s\delta_i)] &= a_1^2 \sigma^2 \end{aligned}$$

Next we equalise the sample variations of two sides in $(3B.2)$:

$$\begin{aligned} s^2 &= a_1^2 \left(s^2 + \sigma^2\right) \\ a_1 &= \frac{s}{\sqrt{s^2 + \sigma^2}} \end{aligned} \tag{3B.4}$$

It is straightforward that $g(s, x_i) = \frac{s}{\sqrt{s^2+\sigma^2}} x_i$.

Fraction $\frac{s}{\sqrt{s^2+\sigma^2}}$ is the correlation between observed return x and extracted means δ and summarises the predictive power of an investor. We denote it ρ :

$$\rho = \frac{s}{\sqrt{s^2 + \sigma^2}} = Corr(x, \delta)$$

The next period's returns become:

$$y_i = \frac{s}{\sqrt{s^2 + \sigma^2}} x_i + \sigma u_i = \rho x_i + \sigma u_i \tag{3B.5}$$

B3. Bayesian Approach: Posterior PDF of s Given Data

The following is the Bayes rule for probability density functions:

$$pdf\,(s|x)\,pdf\,(x) = pdf\,(x|s)\,pdf\,(s) \tag{3B.6}$$

Let $pdf\,(s)$ be a prior on s. Then $pdf\,(x|s)$ is

$$pdf\,(x|s) = \frac{1}{(2\pi)^{\frac{n}{2}}\left(\sigma^2+s^2\right)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2\left(\sigma^2+s^2\right)}\right]$$

i.e. $x \sim N\left(0, \left(\sigma^2+s^2\right)\varepsilon\right)$, where ε is a vector of 1's.

$pdf\,(x)$, unconditional distribution of n-dimensional x vector, is expressed as

$$pdf\,(x) = \int\limits_s pdf\,(x|s)\,pdf\,(s)\,ds \tag{3B.7}$$

Then, using (3B.6),

$$\begin{aligned} pdf\,(s|x) &= \frac{pdf\,(x|s)\,pdf\,(s)}{pdf\,(x)} \\ &= \frac{\frac{1}{(2\pi)^{\frac{n}{2}}\left(\sigma^2+s^2\right)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2\left(\sigma^2+s^2\right)}\right]pdf\,(s)}{\int\limits_{\{s\}}\frac{1}{(2\pi)^{\frac{n}{2}}\left(\sigma^2+s^2\right)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2\left(\sigma^2+s^2\right)}\right]pdf\,(s)\,ds} \end{aligned} \tag{3B.8}$$

$\{s\}$ represents the support of s. Let us make sure that (3B.7) is a pdf indeed. Consider its integral

$$\underbrace{\int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{\infty}}_{n} pdf\,(x_1)\ldots pdf\,(x_n)\,dx_1\ldots dx_n$$

(Here we change the multi-integral $\underbrace{\int\limits_{-\infty}^{\infty}...\int\limits_{-\infty}^{\infty}}_{n}$ to a single $\int\limits_{-\infty}^{\infty}$, and $dx_1...dx_n$ to dx which is an abuse of notation but simplifies the formulae.)

$$
\begin{aligned}
&= \int\limits_{-\infty}^{\infty}\int\limits_{\{s\}} pdf\,(x|s)\,pdf\,(s)\,dsdx \\
&= \int\limits_{\{s\}}\left[\int\limits_{-\infty}^{\infty} pdf\,(x|s^2)\,dx\right] pdf\,(s)\,ds \\
&= \int\limits_{\{s\}} 1 * pdf\,(s)\,ds = 1
\end{aligned}
\tag{3B.9}
$$

(3B.9) confirms that (3B.7) defines a *pdf* correctly as claimed.

B4. Modified Maximum Likelihood Estimates

In this appendix we derive the Modified Maximum Likelihood Estimate $MMLE\,(s|.)$ for the uniform and normal priors on s.

When the prior on s is uniform $[\underline{s},\overline{s}]$, $MMLE\,(s|Unif)$ is derived as follows:

Assume that s is distributed uniformly[14] over $[\underline{s},\overline{s}]$:

$$pdf\,(s) = \frac{1}{\overline{s}-\underline{s}}$$

[14]Note that uniform prior of $pdf\,(s)$ implies Beta distribution prior for $pdf\,(s^2)$.

Rewrite (3B.8) :

$$
\begin{aligned}
pdf\,(s|x) &= \frac{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]\frac{1}{\bar{s}-\underline{s}}}{\int_{\underline{s}}^{\bar{s}}\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]\frac{1}{\bar{s}-\underline{s}}ds} \\
&= \frac{\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]}{\int_{\underline{s}}^{\bar{s}}\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds} \qquad (3B.10)
\end{aligned}
$$

Note that the denominator $\int_{\underline{s}}^{\bar{s}}\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds$ is not a function of s, it is just a scaler for $pdf\,(s|x)$. Therefore

$pdf\,(s|x) \propto \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]$ and maximum likelihood estimate of s depends on the numerator only.

Maximum likelihood estimate of s solves the following:

$$\partial_s\left(pdf\,(s|x)\right) = 0 \qquad (3B.11)$$

Derivative of $pdf\,(s|x)$ is proportional to

$$
\begin{aligned}
\partial_s\left(pdf\,(s|x)\right) &\propto -\left(\frac{s\left(-\sum x_i^2+n\left(s^2+\sigma^2\right)\right)}{\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]\sqrt{(2\pi)^n}\left(s^2+\sigma^2\right)^2\sqrt{(s^2+\sigma^2)^n}}\right) \\
&\propto -s\left(-\sum x_i^2+n\left(s^2+\sigma^2\right)\right)
\end{aligned}
$$

which implies that (3B.11) has three solutions, if the expression under the root is positive:

$$s = 0;\ s = -\sqrt{\frac{\sum x_i^2}{n}-\sigma^2};\ s = \sqrt{\frac{\sum x_i^2}{n}-\sigma^2}$$

Analysis of the second derivative shows that the pdf is maximised either for $s = 0$

(when $\frac{1}{n}\sum x_i^2 < \sigma^2$), or for $s = \sqrt{\frac{1}{n}\sum x_i^2 - \sigma^2}$ (when $\frac{1}{n}\sum x_i^2 > \sigma^2$). Therefore, we estimated s to be $\sqrt{\max\left(\frac{1}{n}\sum x_i^2 - \sigma^2, 0\right)}$. It is worth noting that although the choice of $[\underline{s}, \overline{s}]$ parameters does not affect the maximum likelihood estimate through (3B.11) they must be considered nevertheless, because $\sqrt{\max\left(\frac{1}{n}\sum x_i^2 - \sigma^2, 0\right)}$ should be within $[\underline{s}, \overline{s}]$, or $\sqrt{\max\left(\frac{1}{n}\sum x_i^2 - \sigma^2, 0\right)} < \overline{s}$. In fact, for growing $\sum x_i^2$ upper bound $\overline{s}$ is the only restriction that is supposed to keep s estimate under control. Thereafter, we present the following modified maximum likelihood estimate $MMLE[s]$:

$$MMLE\,[s|Unif] = \min\left(\sqrt{\max\left(\frac{1}{n}\sum x_i^2 - \sigma^2, 0\right)}, \overline{s}\right) \tag{3B.12}$$

Next we derive $MMLE\,(s|Norm)$, a Modified Maximum Likelihood Estimate of s given the normal prior $N(m,v)$ on s

Assume that a priori s is distributed normally with $N(m,v)$. Then $pdf\,(s) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}}$. Plug it into (3B.8) :

$$pdf\,(s|x) = \frac{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}}}{\int\limits_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds} \tag{3B.13}$$

Furthermore, $\int\limits_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds$ is independent of s and

$$pdf\,(s|x) \propto \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} \tag{3B.14}$$

To find the maximum we solve $\partial_s\,(pdf\,(s|x)) = 0$.

A derivative of *pdf* $(s|x)$ is proportional to the following:

$$\partial_s\,(pdf\,(s|x)) \propto \frac{\exp\left[-\frac{(m-s)^2}{2v} - \frac{x^2}{2(s^2+\sigma^2)}\right]\left(m\,\left(s^2+\sigma^2\right)^2 - s\left(s^4 - v\,x^2 + n\,v\,\sigma^2 + \sigma^4 + s^2\left(n\,v + 2\,\sigma^2\right)\right)\right)}{\sqrt{(2\,\pi)^{1+n}}\,v^{\frac{3}{2}}\,\left(s^2+\sigma^2\right)^2\,\sqrt{\left(s^2+\sigma^2\right)^n}}$$

Or,

$$\partial_s\,(pdf\,(s|x)) \propto \left(m\,\left(s^2+\sigma^2\right)^2 - s\left(s^4 - v\,x^2 + n\,v\,\sigma^2 + \sigma^4 + s^2\left(n\,v + 2\,\sigma^2\right)\right)\right)$$

To find a new estimate of s, based on modified maximum likelihood method, we maximise its posterior *pdf* $(s|x)$, or find a maximum of real solutions to $\partial_s\,(pdf\,(s|x)) = 0$. We denote it $MMLE\,(s|Norm)$.

It is straightforward that to provide a reasonable prior for s, m should be positive. However, closed-form solutions to the 5^{th} order polynomial $\partial_s\,(pdf\,(s|x)) = 0$ are not readily available for non-zero m. Therefore, we have to restrict ourselves to numerical solutions and graphical output for non-zero mean m of the Normal prior distribution.

Note that for positive m, $MMLE\,(s|Norm) > 0$, for any $\sum x_i^2$.

B5. Posterior Means of s and ρ

We start with calculating posterior moments of s and ρ, given the normal prior $N(m, v)$

For $s \sim N(m,v)$, posterior on s satisfies (3B.13) and $E[s|x]$ in (3.4) becomes:

$$\begin{aligned} E[s|x] &= \int_{-\infty}^{\infty} s * pdf(s|x)\, ds \\ &= \int_{-\infty}^{\infty} s \frac{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}}}{\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds} ds \end{aligned} \tag{3B.15}$$

Here we calculate $E[\rho|x]$ and $Var[\rho|x]$ under assumption of normal prior on $s \sim N(m,v)$. $E[\rho|x]$ is as follows:

$$\begin{aligned} E[\rho|x] &= E\left[\frac{s}{\sqrt{s^2+\sigma^2}}|x\right] = \int \frac{s}{\sqrt{s^2+\sigma^2}} * pdf(s|x)\, ds \\ &= \frac{\int_{-\infty}^{\infty} \frac{s}{\sqrt{s^2+\sigma^2}} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds}{\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds} \end{aligned} \tag{3B.16}$$

Next is the expression of $Var[\rho|x]$:

$$\begin{aligned} Var[\rho|x] &= E\left[\rho^2|x\right] - E[\rho|x]^2 \qquad (3B.17) \\ &= \frac{\int_{-\infty}^{\infty} \left(\frac{s}{\sqrt{s^2+\sigma^2}}\right)^2 \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds}{\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds} \\ &\quad - \left(\frac{\int_{-\infty}^{\infty} \frac{s}{\sqrt{s^2+\sigma^2}} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds}{\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds}\right)^2 \end{aligned}$$

Afterwards we calculate the posterior moments of s and ρ, given the uniform prior $[\underline{s},\overline{s}]$

Rewrite (3.4) for $s \sim [\underline{s}, \overline{s}]$ using expression (3B.10) for $pdf\,(s|x)$:

$$
\begin{aligned}
E[s|x] &= \int_{\underline{s}}^{\overline{s}} s * pdf\,(s|x)\,ds = \int_{\underline{s}}^{\overline{s}} s * \frac{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]}{\int_{\underline{s}}^{\overline{s}}\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds} ds \\
&= \frac{\int_{\underline{s}}^{\overline{s}} s\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds}{\int_{\underline{s}}^{\overline{s}}\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds} \qquad (3B.18)
\end{aligned}
$$

Next we calculate $E[\rho|x]$ and $Var[\rho|x]$, given the uniform prior $[\underline{s}, \overline{s}]$:

$$
\begin{aligned}
E[\rho|x] &= E\left[\frac{s}{\sqrt{s^2+\sigma^2}}|x\right] = \int \frac{s}{\sqrt{s^2+\sigma^2}} * pdf\,(s|x)\,ds \\
&= \int_{\underline{s}}^{\overline{s}} \frac{s}{\sqrt{s^2+\sigma^2}} * \frac{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]}{\int_{\underline{s}}^{\overline{s}}\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds} ds \\
&= \frac{\int_{\underline{s}}^{\overline{s}} \frac{s}{\sqrt{s^2+\sigma^2}}\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds}{\int_{\underline{s}}^{\overline{s}}\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds} \qquad (3B.19)
\end{aligned}
$$

$$
\begin{aligned}
Var[\rho|x] &= E\left[\rho^2|x\right] - E[\rho|x]^2 \qquad (3B.20) \\
&= \frac{\int_{\underline{s}}^{\overline{s}} \left(\frac{s}{\sqrt{s^2+\sigma^2}}\right)^2 \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds}{\int_{\underline{s}}^{\overline{s}}\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds} \\
&\quad - \left(\frac{\int_{\underline{s}}^{\overline{s}} \frac{s}{\sqrt{s^2+\sigma^2}}\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds}{\int_{\underline{s}}^{\overline{s}}\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}}\exp\left[-\frac{\sum x_i^2}{2(\sigma^2+s^2)}\right]ds}\right)^2
\end{aligned}
$$

B6. MV Optimisation: Optimal Weights and Corresponding Sharpe Ratios

Given that the asset returns follow (3.1), the mean-variance optimisation is expressed as follows:

$$\max_{w} E\left[R_p|x\right] - \frac{\gamma}{2} Var\left[R_p|x\right] \tag{3B.21}$$

where $E\left[R_p|x\right]$ is the expected portfolio return, $Var\left[R_p|x\right]$ the portfolio variance and w the vector of portfolio weights.

As shown in (3B.5), the next period's returns are

$$y_i = \rho x_i + \sigma u_i \tag{3B.22}$$

The expectation $E\left[R_p|x\right]$ equals

$$\begin{aligned} E\left[R_p|x\right] &= \int w'x\rho * pdf\left(s|x\right) ds \\ &= w'x \int \rho * pdf\left(s|x\right) ds \\ &= E\left[\rho|x\right] \omega' x \end{aligned} \tag{3B.23}$$

The variance $Var\left[R_p|x\right]$ is

$$Var\left[R_p|x\right] = E\left[R_p^2|x\right] - E\left[R_p|x\right]^2 \tag{3B.24}$$

where

$$\begin{aligned}
E\left[R_p^2|x\right] &= \int (w'x\rho)^2 pdf\left(s|x\right)ds + \sigma^2 \int (w'u)^2 Ndist\left(u\right)du \\
&= (w'x)^2 \int \rho^2 * pdf\left(s|x\right)ds + \sigma^2 \omega' E\left[uu'\right]\omega \\
&= E\left[\rho^2|x\right](x'x)(w'w) + \sigma^2 (w'w)
\end{aligned}$$

Back to the variance in $(3B.24)$:

$$\begin{aligned}
Var\left[R_p|x\right] &= E\left[\rho^2|x\right] w' (x'x) w + \sigma^2 (w'w) - E\left[\rho|x\right]^2 w' (x'x) w \\
&= Var\left[\rho|x\right](x'x)(w'w) + \sigma^2 (w'w) \qquad (3B.25)
\end{aligned}$$

Plugging $(3B.23)$ and $(3B.25)$ into $(3B.21)$ yields:

$$\begin{aligned}
&\max_w \quad E\left[R_p|x\right] - \frac{\gamma}{2} Var\left[R_p|x\right] \qquad (3B.26) \\
= \; &\max_w E\left[\rho|x\right] w'x - \frac{\gamma}{2}\left[Var\left[\rho|x\right](x'x)(w'w) + \sigma^2 (w'w)\right]
\end{aligned}$$

The above is optimised when portfolio weights are the following:

$$\begin{aligned}
w_efficient_i &= \frac{E\left[\rho|x\right] x_i}{Var\left[\rho|x\right](xx') + \sigma^2}\frac{1}{\gamma} \\
&= \frac{E\left[\frac{s}{\sqrt{s^2+\sigma^2}}|x\right] x_i}{Var\left[\frac{s}{\sqrt{s^2+\sigma^2}}|x\right]\sum x_i^2 + \sigma^2}\frac{1}{\gamma} \\
&= \frac{x_i \int \frac{s}{\sqrt{s^2+\sigma^2}} pdf\left(s|x\right)ds}{\left(\int \frac{s^2}{s^2+\sigma^2} pdf\left(s|x\right)ds - \left(\int \frac{s}{\sqrt{s^2+\sigma^2}} pdf\left(s|x\right)ds\right)^2\right)\sum x_i^2 + \sigma^2}\frac{1}{\gamma}
\end{aligned}$$

The expected Sharpe ratio under efficient use of information is:

$$
\begin{aligned}
SR_{efficient} &= \frac{E\left[R_{Portf}|efficient\right]}{SD\left[R_{Portf}|efficient\right]} \\
&= \frac{\sum \frac{E[\rho|x]x_i}{Var[\rho|x]\sum x_i^2+\sigma^2} E\left[\rho|x\right]x_i}{\sqrt{\sum\left(\frac{E[\rho|x]x_i}{Var[\rho|x]\sum x_i^2+\sigma^2}\right)^2\left(\sigma^2+\sum x_i^2 Var\left[\rho|x\right]\right)}} \\
&= \frac{\frac{E[\rho|x]\sum x_i^2}{Var[\rho|x]\sum x_i^2+\sigma^2}}{\sqrt{\frac{\sum x_i^2}{\left(Var[\rho|x]\sum x_i^2+\sigma^2\right)^2}\left(\sigma^2+\sum x_i^2 Var\left[\rho|x\right]\right)}} = \frac{E\left[\rho|x\right]\sqrt{\sum x_i^2}}{\sqrt{Var\left[\rho|x\right]\sum x_i^2+\sigma^2}}
\end{aligned}
$$

Therefore, we get (3.8) :

$$
SR_{efficient} = \frac{E\left[\rho|x\right]\sqrt{\sum x_i^2}}{\sqrt{Var\left[\rho|x\right]\sum x_i^2+\sigma^2}}
$$

With point-estimates, the correlation is assumed to be constant. Ignoring uncertainty around it simplifies MV optimisation problem $(3B.26)$ to

$$
\max_{w} \ \widehat{\rho} w'x - \frac{\gamma}{2}\sigma^2 \left(w'w\right)
$$

which is solved by:

$$
w\_naive_i = \frac{1}{\gamma\sigma^2}\widehat{\rho}x_i
$$

With point estimates the expected Sharpe ratio is:

$$
\begin{aligned}
SR_{naive} &= \frac{\frac{1}{\gamma\sigma^2}\sum\left(\widehat{\rho}x_i\right)^2}{\sqrt{\sum\sigma^2\left(\frac{1}{\gamma\sigma^2}\widehat{\rho}x_i\right)^2}} = \frac{\frac{1}{\sigma^2}\sum\left(\widehat{\rho}x_i\right)^2}{\sqrt{\sigma^2\left(\frac{1}{\sigma^2}\right)^2\sum\left(\widehat{\rho}x_i\right)^2}} \\
&= \frac{\frac{1}{\sigma^2}\sum\left(\widehat{\rho}x_i\right)^2}{\sigma\frac{1}{\sigma^2}\sqrt{\sum\left(\widehat{\rho}x_i\right)^2}} = \frac{\sum\left(\widehat{\rho}x_i\right)^2}{\sigma\sqrt{\sum\left(\widehat{\rho}x_i\right)^2}} = \frac{\widehat{\rho}}{\sigma}\sqrt{\sum x_i^2}
\end{aligned}
$$

Therefore,

$$SR_{naive} = \frac{\widehat{\rho}}{\sigma}\sqrt{\sum x_i^2}$$

B7. Formulas for Portfolio Weights and Associated Sharpe Ratios

This appendix gives explicit expressions of portfolio weights and associated Sharpe ratios under different methods of estimating the scaling.

Weights/The Sharpe Ratio are as follows under the efficient use of information with normal prior on s :

Explicitly, the optimal portfolio weights with normal prior and efficient use of information are derived after substituting out $E\left[\rho|x\right]$ and $Var\left[\rho|x\right]$ in (3.6) by the expressions (3B.16) and (3B.17) respectively.

$$w\_efficient\_norm_i$$

$$= \frac{x_i \int\limits_{-\infty}^{\infty} \frac{s}{\sqrt{s^2+\sigma^2}} \frac{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)} \ \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}}}{\int\limits_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)} \ \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds} ds}{\left(\begin{array}{c} \int\limits_{-\infty}^{\infty} \frac{s^2}{s^2+\sigma^2} \frac{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)} \ \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}}}{\int\limits_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)} \ \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds} ds \\ - \left(\int\limits_{-\infty}^{\infty} \frac{s}{\sqrt{s^2+\sigma^2}} \frac{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)} \ \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}}}{\int\limits_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)} \ \frac{1}{\sqrt{2\pi v}} e^{-\frac{(s-m)^2}{2v}} ds} ds \right)^2 \end{array} \right) \sum x_i^2 + \sigma^2} \frac{1}{\gamma} \tag{3B.27}$$

Plugging the same $E\left[\rho|x\right]$ and $Var\left[\rho|x\right]$ into (3.8) , gives us the corresponding Sharpe ratio $SR_{efficient\_normal}$.

Weights/the Sharpe ratio are as follows under the efficient use of information with uniform prior on s :

Similar to $w\_efficient\_norm_i$, the formula $w\_efficient\_unif_i$ is as follows:

$$w\_efficient\_unif_i = \frac{x \int_{\underline{s}}^{\overline{s}} \frac{s}{\sqrt{s^2+\sigma^2}} \frac{\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)}}{\int_{\underline{s}}^{\overline{s}} \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)} ds} ds}{\left(\int_{\underline{s}}^{\overline{s}} \frac{s^2}{s^2+\sigma^2} \frac{\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)}}{\int_{\underline{s}}^{\overline{s}} \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)} ds} ds - \left(\int_{\underline{s}}^{\overline{s}} \frac{s}{\sqrt{s^2+\sigma^2}} \frac{\frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)}}{\int_{\underline{s}}^{\overline{s}} \frac{1}{(\sigma^2+s^2)^{\frac{n}{2}}} \exp -\frac{\sum x_i^2}{2(\sigma^2+s^2)} ds} ds \right)^2 \right) \sum x_i^2 + \sigma^2} \frac{1}{\gamma} \tag{3B.28}$$

The corresponding Sharpe ratio $SR_{efficient\_uniform}$ is obtained in a similar way.

Weights/the Sharpe ratio under naive use of information are given below:

In the formula of the naive Sharpe ratio (3.9), $\widehat{\rho}$ is based on $MMLE[s|Prior]$, given the corresponding prior:

$$\widehat{\rho} = \frac{MMLE[s|Prior]}{\sqrt{MMLE[s|Prior]^2 + \sigma^2}} \tag{3B.29}$$

Explicitly, optimal weights are as follows:

Under naive use of information with uniform prior,

$$w\_naive\_unif_i = \frac{x_i}{\gamma\sigma^2} \frac{MMLE[s|Unif]}{\sqrt{MMLE[s|Unif]^2 + \sigma^2}};$$

Under naive use of information with normal prior,

$$w\_naive\_norm_i = \frac{x_i}{\gamma\sigma^2} \frac{MMLE\left[s|Norm\right]}{\sqrt{MMLE\left[s|Norm\right]^2 + \sigma^2}}.$$

Depending on Normal and Uniform prior, expressions for the Sharpe ratio respectively change.

Under the conventional MLE-based approach, optimal weights satisfy (3.7) where the correlation is as follows:

$$\widehat{\rho} = \frac{MLE\left[s\right]}{\sqrt{MLE\left[s\right]^2 + \sigma^2}}$$

The Sharpe ratio in (3.9) changes respectively, depending on $\widehat{\rho}$.

B8. Formulas for Non-Homogeneously Predictable Asset Returns

The solution to the portfolio optimisation problem with several asset classes remains looks similar to the case with a single asset class:

$$w\_efficient\_gen_{ij} = \frac{E\left[\rho_j|Prior\right]}{Var\left[\rho_j|Prior\right]\sum_i^{n_j} x_{ij}^2 + \sigma_j^2} \frac{x_{ij}}{\gamma}$$

$$\begin{aligned} w\_naive\_gen_{ij} &= \frac{1}{\gamma\sigma_j^2}\widehat{\rho}_j x_{ij} \\ i &= 1, ..., n_j;\ j = 1, ..., m \end{aligned}$$

The expected Sharpe ratio under efficient use of information becomes the following:

$$SR_{efficient\_gen} = \frac{\sum_j^m \frac{E[\rho_j|Prior]}{Var[\rho_j|Prior]\sum_i^{n_j} x_{ij}^2+\sigma_j^2} E[\rho_j|Prior] \sum_i^{n_j} x_{ij}^2}{\sqrt{\sum_j^m \frac{E^2[\rho_j|Prior]}{\left(Var[\rho_j|Prior]\sum_i^{n_j} x_{ij}^2+\sigma_j^2\right)^2} \sum_i^{n_j} x_{ij}^2 \left(Var[\rho_j|Prior]\sum_i^{n_j} x_{ij}^2+\sigma_j^2\right)}}$$

$$= \sqrt{\sum_j^m \frac{E^2[\rho_j|Prior]}{Var[\rho_j|Prior]\sum_i^{n_j} x_{ij}^2+\sigma_j^2} \sum_i^{n_j} x_{ij}^2}$$

The true Sharpe ratio (denoted $SR_{efficient\_gen\_true}$) however is different provided that the actual returns and the variance of the ijth asset are $s_j\delta_{ij}$ and σ_j^2 respectively:

$$SR_{efficient\_gen\_true} = \frac{\sum_j^m \sum_i^{n_j} w\_efficient\_gen_{ij} * s_j\delta_{ij}}{\sqrt{\sum_j^m \sum_i^{n_j} (w\_efficient\_gen_{ij} * \sigma_j)^2}}$$

$$= \frac{\sum_j^m \frac{E[\rho_j|Prior]}{Var[\rho_j|Prior]\sum_i^{n_j} x_{ij}^2+\sigma_j^2} s_j \sum_i^{n_j} x_{ij}\delta_{ij}}{\sqrt{\sum_j^m \frac{E^2[\rho_j|Prior]}{\left(Var[\rho_j|Prior]\sum_i^{n_j} x_{ij}^2+\sigma_j^2\right)^2} \sigma_j^2 \sum_i^{n_j} x_{ij}^2}}$$

The expected Sharpe ratio under naive use of information is expressed in the following way:

$$SR_{naive\_gen} = \frac{\sum_j^m \sum_i^{n_j} \frac{\widehat{\rho}_j x_{ij}}{\sigma_j^2} * \widehat{\rho}_j x_{ij}}{\sqrt{\sum_j^m \sum_i^{n_j} \frac{\widehat{\rho}_j^2 x_{ij}^2}{\sigma_j^4} \sigma_j^2}}$$

$$= \sqrt{\sum_j^m \frac{\widehat{\rho}_j^2}{\sigma_j^2} \sum_i^{n_j} x_{ij}^2}$$

The true Sharpe ratios in this case becomes

$$SR_{naive\_gen\_true} = \frac{\sum_j^m \frac{\widehat{\rho}_j}{\sigma_j^2} s_j \sum_i^{n_j} x_{ij}\delta_{ij}}{\sqrt{\sum_j^m \frac{\widehat{\rho}_j^2}{\sigma_j^2} \sum_i^{n_j} x_{ij}^2}}$$

Chapter 4

Bias in Dynamic Asset Allocation Models

4.1. Introduction

In previous chapters we investigated estimation error in the context of a single-period portfolio selection. Now we look at it in the context of dynamic portfolio strategies. The impact of estimation error in a dynamic setting is particularly severe because of the complexity of the setting in which it is necessary to have time varying forecasts. We take Brennan, Schwartz and Lagnado's structure (1997) as a specific illustration of a generic problem and investigate the bias in long-term portfolio selection models that comes from estimating parameters from short-sample historical data.

Brennan, Schwartz and Lagnado (1997) proposed a solution to the *Strategic Asset Al-*

location[1] problem when the asset choice consisted of stock, bonds and cash, and expected asset returns were predictable. They found extraordinarily high levels of predictable investment opportunities. It has been understood that part of these investment opportunities comes from estimation error, but to what degree has not been quantified so far. We investigate how much bias could come from the portfolio selection procedure they followed — estimating parameters of the stochastic return process from the historical data and, based on the estimated parameters, optimising their utility.

We conduct a Monte Carlo simulation analysis in the setting of Brennan, Schwartz and Lagnado (1997) to quantify the degree of bias created by their kind of portfolio allocation model. We adopt a return generating stochastic process from Brennan, Schwartz and Lagnado and assume that it is the true return generating process. The parameters in the process are calibrated so that over 20 years an instantaneous Sharpe ratio[2], measuring the market opportunities, is within its real-life range with the average of 0.5. With this model, we simulate a 20-year returns sample similar to the one used by Brennan, Schwartz and Lagnado as their dataset. We assume that an investor knows true (constant) volatilities as well as the correlation matrix of the innovations in return/predictor variables, but has to estimate constant drifts of the true model from the simulated data using regressions, as it was done in Brennan, Schwartz and Lagnado. To guarantee a high level of accuracy in estimated parameters, we simulate 10,000 independent datasets, and for each dataset estimate the set of unknown parameters. Each estimated parameter of the new model represents the mean of the distribution of 10,000 estimates where each is derived from one simulated dataset using regression analysis. Given that the real world dynamic follows the true model,

[1]The jargon *strategic asset allocation* in the context of a long-horizon investor was proposed by Brennan, Schwartz and Lagnado (1997), as opposed to the tactical asset allocation of a short-horizon investor.

[2]Note that in this chapter by *the Sharpe ratio* we always refer to an instantaneous Sharpe ratio which is not affected by investment horizon.

we find that the investment opportunities expected by a Brennan, Schwartz and Lagnado's investor with his estimated drift terms are outrageously large — the average cross-sectional Sharpe ratio varies between 1.85 and 2.45 over twenty years.

We argue that the degree of bias detected in our model is more moderate than that in Brennan, Schwartz and Lagnado because we assumed the true world dynamic was known and we had to estimate *eight* parameters only. Instead Brennan, Schwartz and Lagnado had to estimate *sixteen* from data, covering a similar period of time, and end up with a single set of parameter estimates. Our estimates were averaged over 10,000 simulations. Histograms of 10,000 estimates of individual parameters, estimated from regressions across each of 10,000 simulation paths, show a wide variation in estimates across simulations. What is even worse, state variables with wild estimated parameters produce even wilder expected investment opportunities when conditioned on slightly changing real investment opportunities. A hardly visible upward trend in the true Sharpe[3] ratio over 20 years changes to a large difference in expected market opportunities between the start (the Sharpe ratio of 1.85) and the end (the Sharpe ratio of 2.45) of the 20-year-period, when measured with estimated parameters.

The bias in optimisation with parameters estimated from noisy data is a common problem in the dynamic portfolio selection literature and, in general terms, any application where estimated parameters are used for optimisation, unless an explicit way is found to adjust for bias. There is considerable literature on dealing with the estimation error in portfolio selection, but the kind of bias we address here has not been acknowledged by other finance researchers so far. However prior literature has tackled related problems, associated with long-term portfolio management models[4].

[3]The Sharpe ratio here is the cross-sectional average of instantaneous Sharpe ratios calculated at each instant for 20 years.

[4]For a thorough review, see Section 1.8.

Kandel and Stambaugh (1996), Barberis (2000), Xia (2001), and Avramov (2002) are some of the key papers on learning/parameter uncertainty when returns are predictable. These papers adopt a Bayesian setting to incorporate the information revealed through the investment horizon. Although they report spectacular reduction in risky asset holdings in the optimal portfolio due to introduction of uncertainty and learning about parameters in the model and offer a valuable insight into how the long-term investor should behave, they do not solve the problem of bias we tackle in this chapter. Their priors on parameters come from historical data and neither the length of their investment horizon nor the sample giving the prior is sufficiently large to provide reliable estimates. Given the complex dynamics of their models, the priors they use cannot guarantee consistent evolution of expected returns through time when resolution of uncertainty is expected in such a short period of time. Neither is the learning these Bayesian investors undertake sufficient to reveal the true parameters of the return generating process[5]. In other words, their models too create the kind of bias we address here although to a lesser extent than the models without uncertainty like Brennan, Schwartz and Lagnado (1997).[6]

There is a need for more radical adjustments to the estimated parameters to avoid the bias associated with optimisation based on estimated parameters, leading to inflated expectations. However we argue there is no straightforward way of adjusting for bias in continuous time portfolio selection models and it is not obvious even what prior should be applied for successful Bayesian adjustment.

We proceed as follows: The next section reviews the long-term portfolio optimisation

[5]In a much simpler setting of constant but unknown investment opportunities, Gennotte (1986) shows that learning eventually resolves parameter uncertainty as $t \longrightarrow \infty$. See Section 1.8.

[6]For detailed discussion of papers on long-term portfolio optimisation in the face of learning and uncertainty about parameters, see Section 1.8.

problem of Brennan, Schwartz and Lagnado; then we describe our Monte Carlo simulation framework and report our findings, followed by the discussion of the statistics of the estimated parameters and their reliability; following section examines the difficulties associated with adjusting the model parameters for bias; Section 4.5 concludes. Technical details on the Monte Carlo simulation analysis are given in Appendix C.

4.2. A Review of Brennan, Schwartz and Lagnado's Model

Investment opportunities in Brennan, Schwartz and Lagnado (BSL) are governed by three state variables, the short-term interest rate r_t, the rate on long-term bonds l_t, and the dividend yield on a stock portfolio δ_t, which are all assumed to follow a joint Markov process. An investor can invest in stock, bonds and cash.

The stock return $\frac{dS_t}{S_t}$ is modelled as

$$\frac{dS_t}{S_t} = (a_{11} + a_{12}r_t + a_{13}l_t + a_{14}\delta_t)\, dt + \sigma_1 dz_1 \tag{4.1}$$

The dynamic of the state variables is governed by the following system:

$$dr_t = (a_{21} + a_{22}r_t + a_{23}l_t + a_{24}\delta_t)\, dt + r_t\sigma_2 dz_2 \tag{4.2}$$

$$dl_t = l_t\,(a_{31} + a_{32}r_t + a_{33}l_t + a_{34}\delta_t)\, dt + l_t\sigma_3 dz_3 \tag{4.3}$$

$$d\delta_t = (a_{41} + a_{42}r_t + a_{43}l_t + a_{44}\delta_t)\, dt + \delta_t\sigma_4 dz_4 \tag{4.4}$$

The consol bond return $\left(\frac{dB_t}{B_t} + l_t dt\right)$ is modelled as

$$\frac{dB_t}{B_t} + l_t dt = \left[l_t - (a_{31} + a_{32}r_t + a_{33}l_t + a_{34}\delta_t) + \sigma_3^2\right] dt - \sigma_3 dz_3 \tag{4.5}$$

which simplifies to

$$= \left[\left(\sigma_3^2 - a_{31}\right) - a_{32}r_t + (1 - a_{33})\, l_t - a_{34}\delta_t\right] dt - \sigma_3 dz_3$$

BSL consider an investor with initial wealth W who is interested in maximising the expected utility of wealth at the end of a twenty year horizon. His utility function is assumed to be of the iso-elastic family (with the risk aversion γ of (-5)):

$$U(W) = \frac{1}{\gamma} W^{\gamma}$$

They define x as the proportion of the investment portfolio that is invested in stock, y the proportion that is invested in the consol bond, and $V(r, l, W, \tau)$ the expected utility under the optimal policy when there are τ periods to the horizon. The Bellman equation is:

$$\underset{x,y}{Max}\; E\left[dV\right] = 0 \tag{4.6}$$

The first-order conditions imply that the optimal controls $x^* \equiv x^*(r, l, \delta, \tau)$ and $y^* \equiv y^*(r, l, \delta, \tau)$ are given as functions of the parameters of the stochastic processes (4.1-4.5) for the state variables.

BSL assume the volatilities $\sigma_1, \sigma_2, \sigma_3$ and σ_4, the correlation matrix $\left(\rho_{ij}\right)$, and the drift coefficients a_{ij} are constant and estimate them from the historical data. According to them, "The joint stochastic process was estimated by using a discrete approximation to the continuous process, and using monthly data for the period January 1972 to December 1991. The stock return was taken as the rate of return on CRSP value weighted market index.

The short rate was taken as the yield on a one month Treasury Bill which was taken from the CRSP Government Bond File. The long rate was taken as the yield to maturity on the longest maturity taxable, non-callable government bond, excluding flower bonds; bond yield data were from the CRSP Government Bond File. The dividend yield was defined as the sum of the past 12 months' dividends on the CRSP value weighted index, divided by the current value of the index" (BSL, pp. 1387-1388). They estimated the system of equations by regressions, based on the data.

In solving the control problem BSL note that it is not possible to evaluate formally the stability of the stochastic differential equation system on account of the non-linearity entering through the equation for l. Instead they "followed the empirical procedure of starting the system at points corresponding to historical joint realizations of the state variables, and then simulating the system forward while setting the innovations equal to zero: in all cases the system converged."

Once the parameters in (4.1-4.5) have been estimated, BSL substitute $x^*(r, l, \delta, \tau)$ and $y^*(r, l, \delta, \tau)$ (which depend on estimated parameters) in the Bellman equation (4.6) and solve numerically the resulting non-linear partial differential equation for the value function. They conclude that an investor with a twenty year horizon should invest very aggressively in a risky asset. Overall, they report extraordinarily high predictable time-variation in returns.

4.3. The Monte Carlo Simulation Framework

The financial markets would not clear if the results of BSL were plausible. As empirical findings suggest otherwise, the investment opportunities found by BSL must be exaggerated. We want to quantify to what degree their results are misleading. We adopt their setting (4.1-

4.5), and assume that these are true processes that generate asset returns and state variables. We calibrate parameters in (4.1-4.4) so that the investment opportunities offered by this model are within reasonable bounds. The size of investment opportunities is measured with an instantaneous Sharpe ratio[7]. The calibrated drift coefficients (a_{ij}) are reported in Table 4.1, the volatilities σ_i — in Table 4.2 and the correlations (ρ_{ij}) — in Table 4.3:[8]

a_{ij} j / i	1	2	3	4
1	−0.028	0.5	0	1.7
2	0.01	−0.2	0	0
3	0.018	0.24	−0.3	−0.3
4	0.02	0	0	−0.5

Table 4.1

σ_1	0.085
σ_2	0.2
σ_3	0.037
σ_4	0.1

Table 4.2

ρ_{ij} j / i	1	2	3	4
1	1	−0.037	−0.33	−0.995
2	−0.037	1	0.33	0.032
3	−0.33	0.33	1	0.298
4	−0.995	0.032	0.298	1

Table 4.3

[7] See Appendix C for the expression of an instantaneous Sharpe ratio.
[8] We took the variance-covariance matrix (i.e. Tables 4.2 & 4.3) from the BSL model.

In particular, we simulated 10,000 datasets using model (4.1-4.4) with true coefficients from Tables 4.1, 4.2 & 4.3. For every simulation path at each point in time the instantaneous Sharpe ratio was evaluated and then averaged across the path (i.e. over 20 years). Figure 4.1 presents the histogram of 10,000 average Sharpe ratios over 20 years, that measures average investment opportunities offered by model (4.1-4.4) with true coefficients during this period:

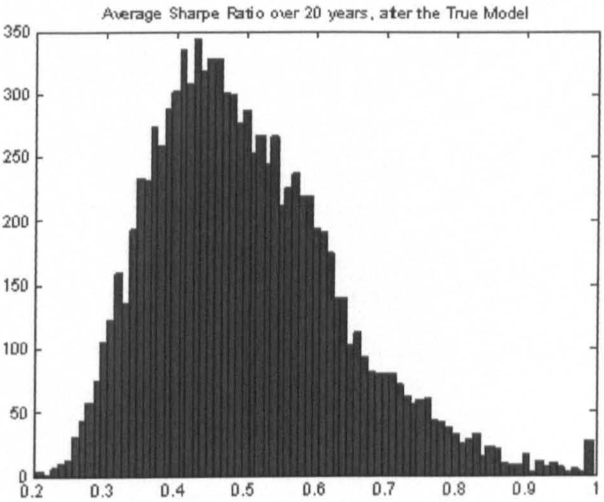

Figure 4.1

Figures 4.2 & 4.3 plot the histogram of the average Sharpe ratios over the first 10 and second 10 years respectively, and detect a scarcely visible time-variation in expected investment opportunities.

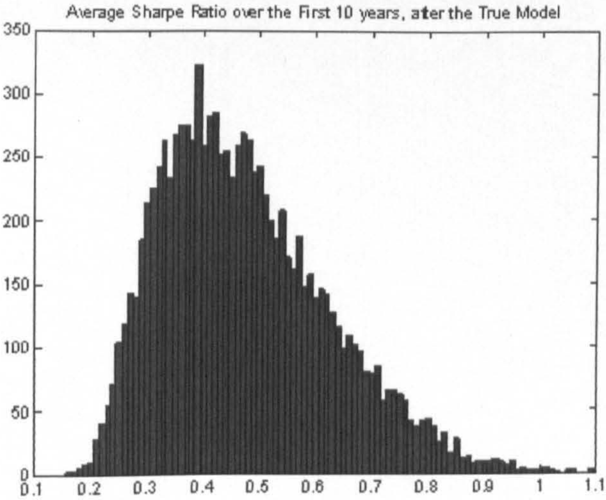

Figure 4.2

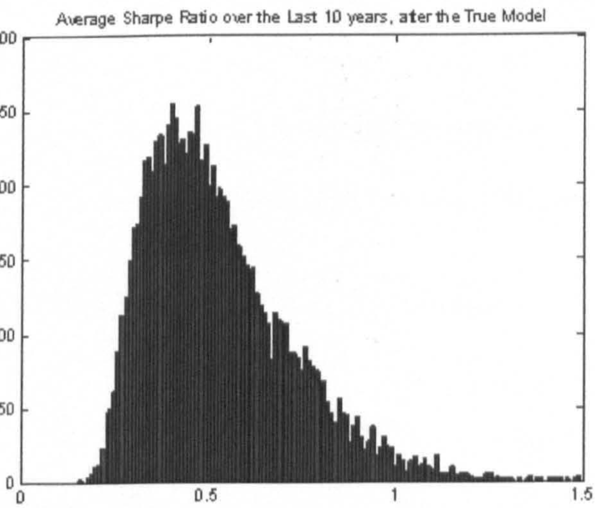

Figure 4.3

Assuming that (4.1-4.5) with calibrated parameters (given in Tables 4.1 - 4.3) is the true return generating model, we simulate (10,000 times) a twenty-year-long dataset on returns and state variables, which is similar to the data used by the BSL[9]. Each dataset will be used as the historical data for inferring the unknown parameters of the stochastic system by regressions (as it was done by BSL). If there were no estimation error, the

[9]Actually, these are datasets for which the average Sharpe ratio was evaluated before (see Figures 4.1, 4.2 & 4.3).

parameters estimated from the simulated twenty-year-long dataset would coincide with true parameters from Tables 4.1 - 4.3 that were used for generating the dataset. Without loss of generality, the volatilities $\sigma_1, \sigma_2, \sigma_3$ and σ_4, and the correlation matrix (ρ_{ij}) are known to an investor with certainty. He also knows the true return generating process (we ignore model uncertainty) and uses our simulated dataset to estimate unknown parameters $(\hat{a}_{ij})$, by fitting respective regressions to the data. His regressions predict increments in state variables and stock returns as functions of the three state variables r_t, l_t, and δ_t, over a twenty-year-horizon. So, the BSL investor estimates 16 parameters $(\hat{a}_{ij})$ from the data and measures investment opportunities based on these estimated parameters. We, on the other hand, know that the actual world dynamic follows (4.1-4.5) with true coefficients from Tables 4.1 - 4.3. This implies that the market opportunities evolve as they should in the real world, but at each instant the BSL investor measures them with his $(\hat{a}_{ij})$ estimates. For our purpose of measuring what the BSL investor's expectations of investment opportunities are, we need to estimate parameters for two regressions only in the above system, those for $\frac{d\ln S_t}{dt}$ and $\frac{d\ln l_t}{dt}$. Therefore, our task is simpler than that of the BSL investor as we estimate eight parameters only in the following equations:

$$\frac{d\ln S_t}{dt} = \left(a_{11} - \frac{\sigma_1^2}{2}\right) + a_{12}r_t + a_{13}l_t + a_{14}\delta_t \tag{4.7}$$

$$\frac{d\ln l_t}{dt} = a_{31} - \frac{1}{2}\sigma_3^2 + a_{32}r_t + a_{33}l_t + a_{34}\delta_t \tag{4.8}$$

Taken separately, the two equations have the stochastic structure of the classical linear model; which is to say that the disturbances are independently and identically distributed with an expected value of zero and a common variance. The two contemporaneous disturbances in the vector of residuals have nonzero covariances for all t. It transpires that the efficient system-wide estimator amounts to nothing more than the repeated application of the ordinary least-squares procedure to generate the regression estimates $(\hat{a}_{ij})$.

Here is a brief summary of what we have done so far: First, we generated 10,000 datasets where each is the data on the evolution of the state variables over twenty years; second, for each dataset we estimated unknown sets of $(\hat{a}_{ij})$ parameters using regression analysis. As a result, we derived 10,000 estimates for every $(\hat{a}_{ij})$ in (4.7-4.8). Table 4.4 reports the mean values of estimated parameters averaged over 10,000 simulations:

$\hat{a}_{ij}$ j / i	1	2	3	4
1	-0.51	0.3	2.47	10.16
3	0.37	0.31	-6.27	-0.45

Table 4.4

A BSL investor thinks that the world evolves after (4.1-4.5) with estimated $(\hat{a}_{ij})$ parameters from Table 4.4 in drifts, and true volatilities and correlations. Hence he measures future investment opportunities based on his estimated model. The discrepancy between estimated $(\hat{a}_{ij})$ in Table 4.4 and true (a_{ij}) in Table 4.1 leads to the bias in expected market opportunities. We measure the opportunities he expects with an instantaneous Sharpe ratio, given that the actual world dynamic evolves according to (4.1-4.5) with true coefficients from Tables 4.1-4.3 (with the Sharpe ratio of 0.5, on average). In other words, at each instant the BSL investor measures true opportunities with his $(\hat{a}_{ij})$ estimates instead of the actual (a_{ij}). The distribution of 10,000 average Sharpe ratios (each is the average of instantaneous Sharpe ratios over 20 years across a particular simulation path), as expected by the BSL investor is given in Figure 4.4:

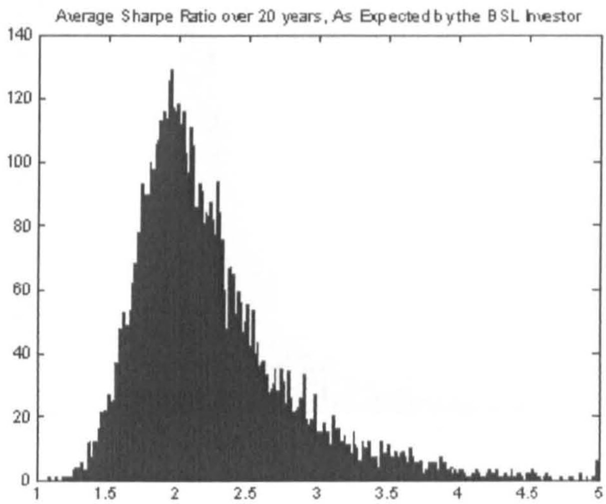

Figure 4.4

The difference between the true and expected market opportunities which corresponds to the BSL way of optimisation, is alarming (compare Figure 4.4 to Figure 4.1). We plot the average market opportunities expected by the BSL investor in the first and second halves of the estimation period, and find a substantial time bias there:

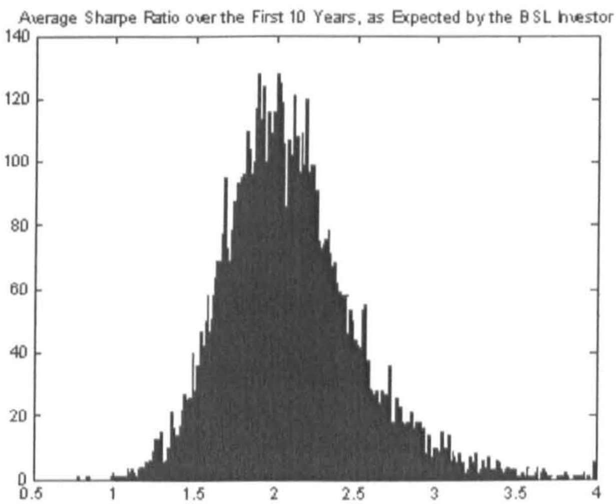

Figure 4.5

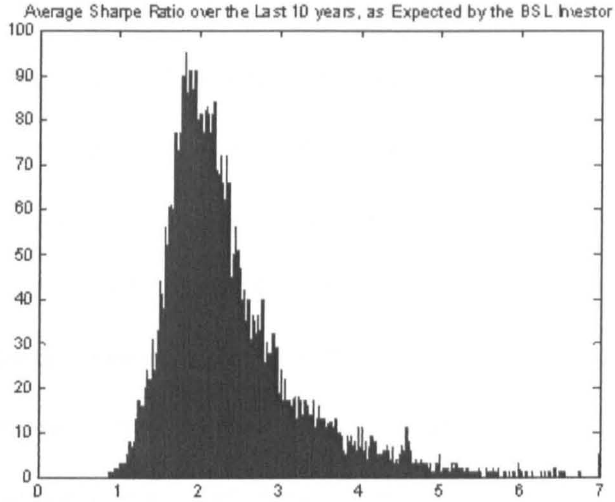

Figure 4.6

It is also interesting to compare the cross-sectional mean SR's over a 20-year-horizon with true coefficients to those with estimated coefficients, as plotted in Figure 4.7 (at each instant both SR, true and expected under BSL, are averaged across 10,000 simulations):

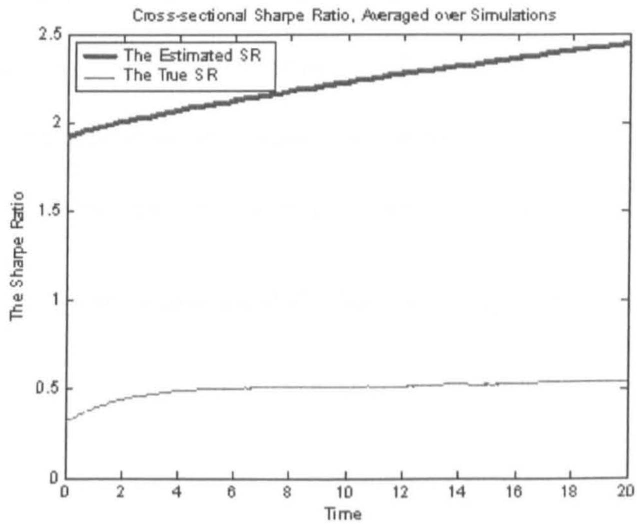

Figure 4.7

The two plots in Figure 4.7 differ not only for their Sharpe ratio, but for their shape as well. The true Sharpe ratio has a tiny upward trend (better seen in Figures 4.2 & 4.3, displaying the distributions of a true Sharpe ratio in the first 10 and last 10 years respectively). The trend of the average Sharpe ratio, expected by the BSL investor, is sharper and linearly increases from less than 2 at the start to about 2.5 in the end. For a given model the wilder the state variables we put in the model, the greater the opportunities are. Furthermore, if these are conditioned on slightly time-dependent opportunities, the bias in expected Sharpe ratio is a compound effect of two things: the regression estimation and the actual time-variation in the Sharpe ratio.

4.4. Analysis of Results

Let $Var\left(\overline{SR}_k\right)$ be the variation of the distribution of 20-year-averages of the Sharpe ratio, given in Figures 4.1 & 4.4. $Var\left(\overline{SR}_k\right)$ is found as the variation of the distribution of $\frac{1}{T}\sum_{t=1}^{T} SR_{kt}$, where T is the number of instances in 20 years, i.e. $T = 250 \times 20 = 5,000$. Let V_k be the variance of the Sharpe ratio across each simulation path; $\frac{1}{10,000}\sum_{k=1}^{10,000} V_k$ will be the average variation of the Sharpe ratio, associated with simulations.

Table 4.5 summarises the statistics of the average Sharpe ratio under both sets of parameters:

	With True (a_{ij})	With Estimated $(\hat{a}_{ij})$
The Sharpe Ratio		
20-year-average	0.501	2.218
first 10-year-average	0.474	2.093
second 10-year-average	0.528	2.343
The Var of the Sharpe Ratio over 20 years		
$Var\left(\overline{SR}_k\right)$	0.017	0.272
$\frac{1}{10{,}000}\sum_{k=1}^{10{,}000} V_k$	0.065	0.333
The Ratio of Var's		
$Var\left(\overline{SR}_k\right)/\frac{1}{10{,}000}\sum_{k=1}^{10{,}000} V_k$	0.271	0.815

Table 4.5

The ratio $Var\left(\overline{SR}_k\right)/\frac{1}{10{,}000}\sum_{k=1}^{10{,}000} V_k$ for the true parameters is 27.1%. This means there is a lot of cross-sectional variation in the Sharpe ratio, but investment opportunities are mean reverting in time. According to Table 4.5, both $Var\left(\overline{SR}_k\right)$ and $\frac{1}{10{,}000}\sum_{k=1}^{10{,}000} V_k$ as well as their ratio substantially increase for the model with estimated parameters. Moreover, their ratio triples to 81.54% which indicates that investment opportunities are largely affected by some sets of parameters giving a larger Sharpe ratio than others; this kind of variation dominates the variation in estimated investment opportunities.

Table 4.6 reports the parameter statistics. The standard error is measured in two alternative ways: one based on the standard deviation of the distribution of 10,000 independent $(\hat{a}_{ij})_k$ $(k = 1, ..., 10,000)$, where each $(\hat{a}_{ij})_k$ is estimated in the respective regression across a simulation path; the other as a standard error of a given coefficient in respective regression. There are eight different standard errors of the first kind, and $8 \times 10,000$ different standard errors of the second kind (i.e. one for each coefficient in each simulation). The t-value we compute measures the significance of the deviation of estimated coefficient from real one. For every coefficient we compute two different t-values — the first is based on the standard deviation of the distribution of estimates $\frac{\hat{a}_{ij}-a_{ij}}{\sqrt{\frac{1}{10,000} s.d.(\hat{a}_{ij})}}$, the second is the overall t-value, depending on 10,000 standard errors of the given coefficient $\frac{\hat{a}_{ij}-a_{ij}}{\sqrt{\frac{1}{10,000}\sum_{k=1}^{10,000} s.e.^2(\hat{a}_{ij})_k}}$.

	Parameter Population Value	Estimated Value $\hat{a}_{ij}$ (mean of 10,000 *ols* estimates)	St.Dev. $s.d.\ (\hat{a}_{ij})$ (s.d. of 10,000 *ols* estimates)	First t-value as $\frac{\hat{a}_{ij}-a_{ij}}{\sqrt{\frac{1}{10,000} s.d.(\hat{a}_{ij})}}$	Second t-value as $\frac{\hat{a}_{ij}-a_{ij}}{\sqrt{\frac{1}{10,000}\sum_{k=1}^{10,000} s.e.^2(\hat{a}_{ij})_k}}$
a_{11}	−0.028	−0.51	0.61	−79.88	−2.84
a_{12}	0.5	0.30	3.00	−6.60	−0.52
a_{13}	0	2.47	10.21	24.25	3.63
a_{14}	1.7	10.16	8.18	103.32	13.17
a_{31}	0.018	0.37	0.29	120.95	3.19
a_{32}	0.24	0.31	1.30	5.68	0.26
a_{33}	−0.3	−6.27	4.97	−120.07	−13.29
a_{34}	−0.3	−0.45	3.52	−4.18	−0.35

Table 4.6

These overwhelmingly high t-values indicate the degree of bias associated with this estimation procedure which is adopted by most people.

It is worth noting that the bias we show here is much more moderate compared to the one arrived at using the BSL method of estimation. Unlike us, they estimate 16 parameters instead of 8. They cannot apply standard *ols* regressions as we do which increases sources of bias in their estimates. Depending on realisations, their coefficients of state variables may end up anywhere in the distribution of $(\hat{a}_{ij})$ (for distributions of $(\hat{a}_{ij})$ considered here, see Figures $4.8 - 4.11$ for the estimates of parameters in the regression of $\frac{d \ln S_t}{dt}$, and Figures $4.12 - 4.15$ for the estimates of parameters in the regression of $\frac{d \ln l_t}{dt}$). What is even worse, they think the world evolves according to their model with wild estimates, which further increases the expected Sharpe ratio to extraordinary heights. It takes $10,000$ years for all parameter estimates $(\hat{a}_{ij})$ to get reasonably close to their respective true parameter values (a_{ij}).

4.5. Adjustment for Bias: Issues

Our Monte Carlo simulation analysis highlights the dangers associated with optimisation that uses a model with parameters estimated from noisy data. Even though we ignored the model uncertainty and kept parameters of predictable state variables constant, estimating drift terms from a limited-size sample had produced market opportunities up to five times larger than those implied by the true model.

It is straightforward that the problem of bias is not limited to estimating parameters using linear regressions. A non-linear regression specification would further magnify the impact of estimation error and produce even wilder state variables and, therefore, investment opportunities. Papers, studying long-term portfolio selection with regression-estimated parameters but incorporating the Bayesian learning of uncertain parameters and/or model, report reduced investment opportunities compared to BSL (see, for example, Xia (2001)). However, because their prior on parameters comes from a similar-size dataset and Bayesian adjustments are based on market realisations over a limited investment horizon, their models too are prone to creating substantial bias.

There is a need for a more radical adjustment for bias but what adjustment can we make?

The role forecast evolution in time may play in portfolio selection has not been fully understood so far. By the nature of model construction and development, its efficiency is strongly tied to the corresponding forecast optimality. Adjusting for estimation error in a model of long-term asset returns is inseparable from making consistent end-of-date return forecasts with this model at different times. Previous chapters discuss different ways of dealing with estimation error. Grinold and Kahn (1999) suggest Bayesian refinement to the parameter estimates, but there is no obvious way of implementing them in a complex environment of continuous-time portfolio selection, as single-period forecast errors inflate uncontrollably over a long horizon.

We regard this bias adjustment problem as a generalisation of β-adjustment problem, examined by Blume (1971, 1975), and Vasicek (1973). There is however prior information on cross-sectional distribution of βs and in their setting implementation of the Bayesian approach is straightforward. In continuous-time asset returns models we have a class of assets which is not even a homogeneous group and we cannot talk about their mean and dispersion as we did for β.

A main kind of prior we can use is some sort of distribution of plausible Sharpe ratio; so that the Sharpe ratio is a controllable number when we make an adjustment to the parameters. Meanwhile there is no obvious projection to making this adjustment. A possible adjustment would be to work out what kind of scaling would take us back to lower Sharpe ratio. Then we could scale the forecast drifts either separately or altogether back towards the zero risk premium.

This kind of adjustment can be misleading though. Suppose we have a true return generating structure with a weak dividend effect. Then, if a new model has a significant dividend effect and we scale back, we bet on spurious signals. The more we keep the multiple dimension and complicated regression terms, the more that can go wrong with the new model. Overall, this way of separating what's spurious from what's actually there can lead to a new kind of bias.

4.6. Conclusion

This chapter has addressed the issue of bias in dynamic portfolio selection models arising from the estimation of parameters from a limited-size historical data. We took Brennan, Schwartz and Lagnado (1997)'s long-horizon portfolio optimisation model as a specific illustration of a generic problem in long-term portfolio optimisation. Based on Brennan, Schwartz and Lagnado's strategic asset allocation model, using a Monte Carlo simulation analysis we quantified the degree of bias inherent in applications that include optimisation with estimated parameters.

We started off with a return generating process giving at each instant investment opportunities at a size of 0.5 of the instantaneous Sharpe ratio (on average). The stochastic process as well as state variables were adopted from Brennan, Schwartz and Lagnado. With this model we simulated the data covering 20 years of returns/state variables and, based on the simulated sample, estimated back the 'unknown' parameters of the true return generating process that produced this sample. Since historical datasets as long as this are often used for inferring model parameters for long-term portfolio decisions, it is essential for us to know to what degree true and inferred parameters match each other. For reasons of accuracy, we conducted 10,000 independent simulations and produced a distribution of 10,000 estimates for each parameter that had to be estimated. The actual parameter estimates we used for the estimated model were averages of the corresponding distributions. For a new model with estimated parameters we calculated expected investment opportunities measured with an instantaneous Sharpe ratio. We found that the estimated model promises outstanding investment opportunities corresponding, on average, to the Sharpe ratio of 2.2, as opposed to the true Sharpe ratio of 0.5! The difference between the two is attributed solely to the estimation error.

We pointed out that papers investigating the impact of learning and predictability on portfolio decision are also liable to create the kind of bias addressed in this chapter, although to a lesser extent. We argue there is a need for consistent adjustment for bias in estimated parameters however there is no straightforward way of making such adjustment.

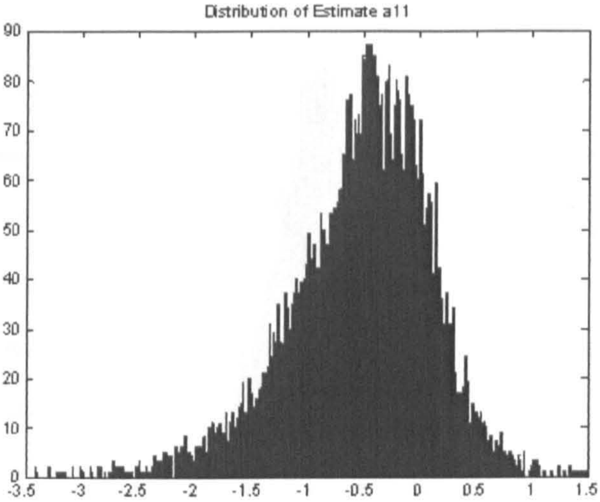

Figure 4.8

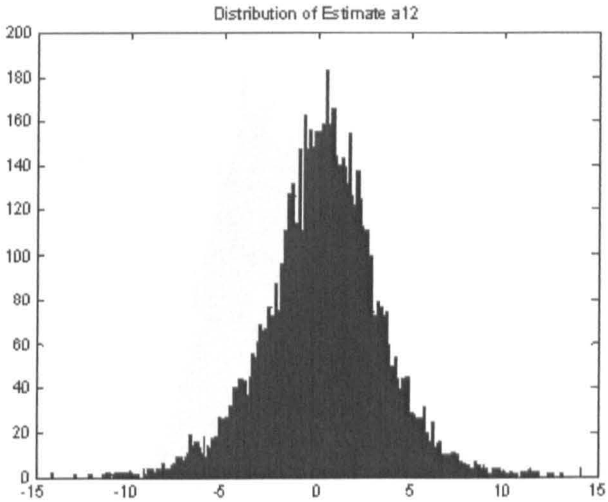

Figure 4.9

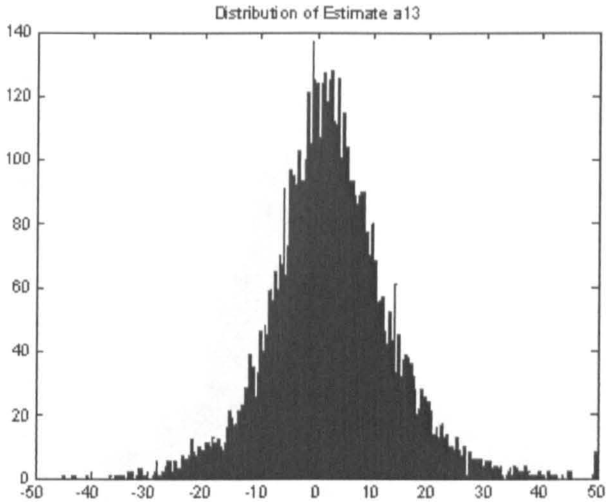

Figure 4.10

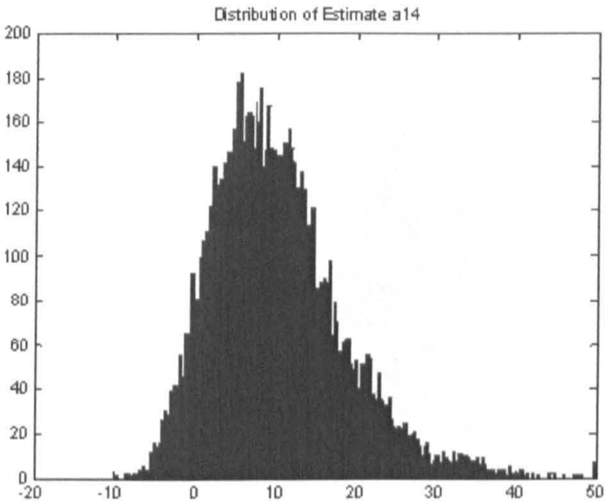

Figure 4.11

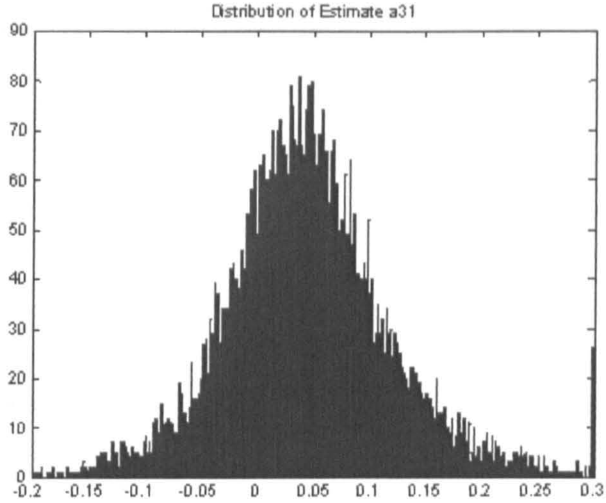

Figure 4.12

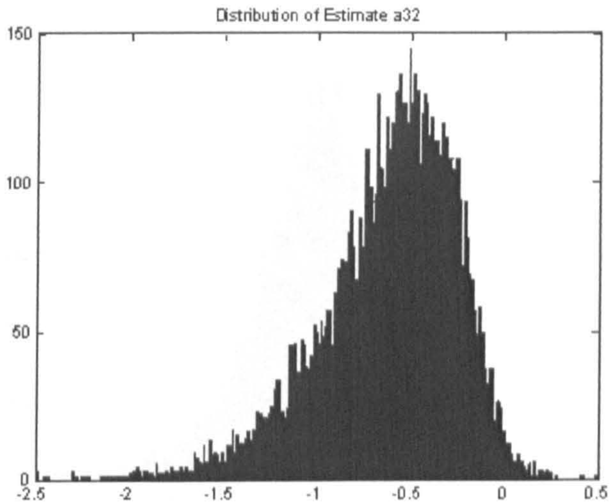

Figure 4.13

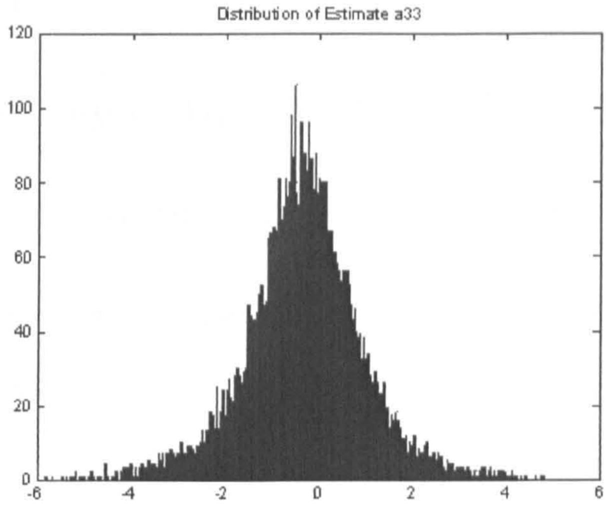

Figure 4.14

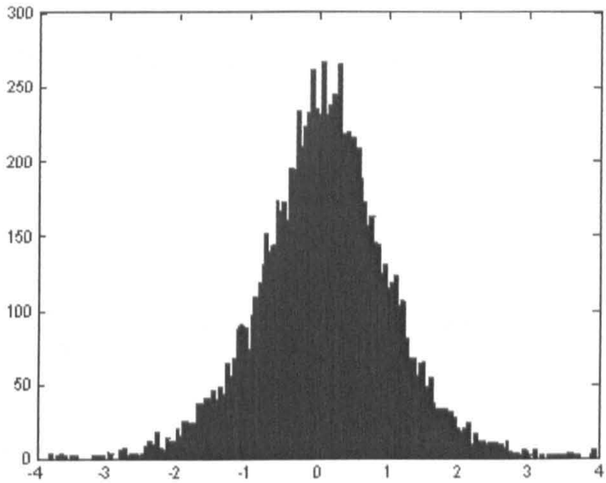

Figure 4.15

Appendix C

C1. The Setting of the Monte Carlo Simulation

We use a variance-stabilising transformation of a state variable.

BSL uses the following processes for state variables (the riskfree rate r_t, the consol bond rate l_t and dividend δ_t):

$$\begin{aligned}
dr_t &= (a_{21} + a_{22}r_t + a_{23}l_t + a_{24}\delta_t)\,dt + r_t\sigma_2 dz_2 \\
dl_t &= l_t\,(a_{31} + a_{32}r_t + a_{33}l_t + a_{34}\delta_t)\,dt + l_t\sigma_3 dz_3 \\
d\delta_t &= (a_{41} + a_{42}r_t + a_{43}l_t + a_{44}\delta_t)\,dt + \delta_t\sigma_4 dz_4
\end{aligned} \tag{4C.1}$$

the stock return

$$\frac{dS_t}{S_t} = (a_{11} + a_{12}r_t + a_{13}l_t + a_{14}\delta_t)\,dt + \sigma_1 dz_1 \tag{4C.2}$$

and, the consol bond return:

$$\begin{aligned}
\frac{dB_t}{B_t} + l_t dt &= \left[l_t - (a_{31} + a_{32}r_t + a_{33}l_t + a_{34}\delta_t) + \sigma_3^2\right] dt - \sigma_3 dz_3 \\
&= \left[\left(\sigma_3^2 - a_{31}\right) - a_{32}r_t + (1 - a_{33})\,l_t - a_{34}\delta_t\right] dt - \sigma_3 dz_3
\end{aligned}$$

Using the Ito calculus we transform $d\delta_t$ as follows:

$$\begin{aligned} d\ln\delta_t &= \frac{1}{\delta_t}d\delta_t - \frac{1}{2}\frac{1}{\delta_t^2}d\delta_t^2 \\ &= \frac{1}{\delta_t}\left[(a_{41} + a_{44}\delta_t)\,dt + \delta_t\sigma_4 dz_4\right] \\ &\quad -\frac{1}{2}\frac{1}{\delta_t^2}\delta_t^2\sigma_\delta^2 dt \\ &= \frac{1}{\delta_t}\left(a_{41} + a_{44}\delta_t - \frac{1}{2}\sigma_\delta^2\delta_t\right)dt - \sigma_\delta dz_1 \end{aligned}$$

Similar transformation applies to dr_t.

Next we rewrite (4C.1) for log's. It is a variance stabilising transformation procedure of the state variables that changes stochastic volatility terms to constants. As a result, it improves the reliability of simulation-based inferences and decreases standard errors. This procedure has been used in prior literature (see, for instance, Detemple, Garcia and Rindisbacher (2003)).

$$\begin{aligned} d\ln r_t &= \frac{a_{21} + \left(a_{22} - \frac{1}{2}\sigma_2^2\right)r_t + a_{23}l_t + a_{24}\delta_t}{r_t}dt + \sigma_2 dz_2 \\ d\ln l_t &= \left(a_{31} - \frac{1}{2}\sigma_3^2\right)dt + \left(a_{32}r_t + a_{33}l_t + a_{34}\delta_t\right)dt + \sigma_3 dz_3 \\ d\ln\delta_t &= \frac{a_{41} + a_{42}r_t + a_{43}l_t + \left(a_{44} - \frac{1}{2}\sigma_4^2\right)\delta_t}{\delta_t}dt + \sigma_4 dz_4 \end{aligned} \tag{4C.3}$$

$(4C.3) - (4C.2)$ are used to simulate the log state variables and the log stock process.

An instantaneous Sharpe ratio SR_t at time t is calculated as

$$SR_t = \sqrt{e_t' V^{-1} e_t}$$

where expected excess return e_t is

$$
\begin{aligned}
e_t &= E[R_t] - r_t 1, \text{ where} \\
E[R_t] &= \begin{bmatrix} (a_{11} + a_{12} r_t + a_{13} l_t + a_{14} \delta_t) \\ [(\sigma_3^2 - a_{31}) - a_{32} r_t + (1 - a_{33}) l_t - a_{34} \delta_t] \end{bmatrix}
\end{aligned}
$$

and the variance-covariance matrix V is

$$
V = \begin{bmatrix} \sigma_1^2 & -\rho_{13}\sigma_1\sigma_3 \\ -\rho_{13}\sigma_1\sigma_3 & \sigma_3^2 \end{bmatrix}
$$

Note, that V is constant here.

The Monte Carlo simulation follows the Euler scheme. For, say, $d \ln r_t$ it evolves in the following way:

$$
\begin{aligned}
(d \ln r)_t^{(i)} &= \left(\frac{a_{21} + \left(a_{22} - \frac{1}{2}\sigma_2^2\right) r_t + a_{23} l_t + a_{24} \delta_t}{r_t} dt \right)^{(i)} + \sigma_2 (dz_2)^{(i)} \\
\ln r_{t+1}^{(i)} &= \ln r_t^{(i)} + (d \ln r)_t^{(i)}
\end{aligned}
$$

The actual value of the interest rate is found as

$$
r_{t+1}^{(i)} = \exp\left(\ln r_{t+1}^{(i)}\right)
$$

Chapter 5

Optimisation in Risk Measurement

5.1. Introduction

Banks and other financial institutions need to meet regulatory requirements for risk measurement and capital. The regulators, like the Financial Services Authority (FSA), want to be sure a firm's/bank's potential for catastrophic net worth loss is accurately measured and that their capital is sufficient to survive such a loss.

Regulators establish levels of risk based capital requirements which each institution must maintain. The optimal allocation of funds raises the question, how small can the regulatory capital required be given that it passes the regulations? We consider two particular rules from the *FSA Handbook* and develop efficient algorithms to calculate corresponding minimal capital required. The algorithms give optimal solutions in closed form and are easy to implement.

Frequent reporting of risk exposures in a dynamic market environment calls for efficient algorithms in risk measurement that give reliable estimates of future capital requirement under different scenarios. The algorithms developed here fall in this category. They can easily deal with changes in rules, which is an important aspect of the risk management with time dimension. Both algorithms lead to interesting generalisations. The following paragraphs give a detailed description.

The "Risk Offsetting Algorithm" minimises the required capital held by a firm to cover the risk arising from its open positions in commodities with different maturities. The algorithm gives all optimal solutions to the problem and is easy to implement with a standard optimisation package. The algorithm motivated our intuition to formulate and prove the Theorem on Convex Optimisation (see Chapter 6), which delivers all optimal solutions in closed-form for a specific combinatorial optimisation problem. The optimality of the solution given by this algorithm is accepted as a special case of the Theorem on Convex Optimisation.

The "Optimal Grouping Algorithm" calculates the minimal interest rate risk by finding the optimal grouping of bonds with different maturities. We formulate the problem and solve it using a linear programming approach. Next we consider its modified version that is interesting for its economic insight, and solve this explicitly using dynamic programming techniques.

These two problems, as well as their methods of solution, are not interrelated and are analysed separately. The chapter is organised as follows: In part one we present the risk offsetting problem, in part two — a problem of optimal grouping of risk. Conclusion gives the summary of the chapter. Technical details appear in Appendices D1-D2 in the end.

PART 1

5.2. The Risk Offsetting Problem

In this part we address the problem of efficient calculation of the Position Risk Required (*PRR*) in commodity portfolios based on the documentation of the *FSA Handbook*[1]. According to the regulations, a firm must calculate the *PRR* on all positions in commodities following one of its four approaches. We developed an algorithm that calculates the minimal *PRR* within the FSA rule.

The algorithm delivers all the optimal solutions to the problem in closed form. It can also cope with a few interesting modifications to the rules and remain optimal.

The outline of Part 1 is the following. The next section describes the risk offsetting problem from the *FSA Handbook* and discusses approaches to calculating the required capital. Sections 5.3 & 5.4 formulate the problem mathematically and solve it using the linear programming approach. In Section 5.5 we develop the risk offsetting algorithm that delivers the minimal *PRR*, and express all solutions analytically. The discussion of possible extensions to the model concludes Part 1.

5.2.1. Summary of the Corresponding FSA Rule

The FSA sets a Position Risk Required (*PRR*) on positions in the firm's commodities. According to the FSA regulations, a firm must calculate the *PRR* on all positions in the commodities, following one of four approaches:

[1] See FSA Handbook (2001)

- A Simplified Approach
- A Maturity Ladder Approach
- An Extended Maturity Ladder Approach
- The Firm's Own Internal Model.

Internal models are subject to a number of qualitative and quantitative requirements and are subject to the FSA approval individually. The other three are standard external approaches. The Simplified Approach should be abandoned as unreasonably expensive in favour of Maturity/Extended Maturity Ladder Approaches that allow a firm to minimise the PRR by offsetting contracts in the same commodity against each other, subject to certain conditions. The next section provides a description of these methods.

5.2.2. Maturity/Extended Maturity Ladder Approach

A commodity is defined as a physical product which is or can be traded on the secondary market. Commodities include precious metals (except gold, which is to be treated as a foreign currency), agricultural products, minerals and base metals, oil and other energy products.

As explained in the guidelines, a firm must calculate the PRR for each commodity separately, except that

(a) different sub-categories of the same commodity that are deliverable against each other may be treated together; and

(b) with the FSA's prior written permission, commodities which are close substitutes for each other, and whose price movements over a minimum period of one year can be shown by the firm to exhibit a stable and reliable correlation of at least 0.9, may be treated together.

All positions in each commodity or commodity derivatives must be expressed in terms of the standard unit of measurement for that commodity (such as tonnes, barrels or kilos). A firm must allocate net positions on any given day to the appropriate maturity band in Table 5.1 below. Physical stock must be assigned to the first band.

MaturityBandsforMaturityLadderApproach
0-1 month
1-3 months
3-6 months
6-12 months
1-2 years
2-3 years
over 3 years

Table 5.1

The objective of the firm is to offset long and short positions within and between maturity bands, in accordance with the following guidelines:

(a) for each maturity band, the firm must sum all the open long positions, and sum all the open short positions. The firm may then subtract the shorts from the longs to form the overall net position. The amount subtracted is the "matched amount". The firm must multiply twice the matched amount by the spread rate of 1.5%, and then by the spot price for the commodity to arrive at the spread risk charge.

(b) the firm may then carry backwards or forwards all or part of the overall net position within a band to an adjacent maturity band for further netting allowances. Where this is the case, the firm must calculate:

(i) a carry charge, by multiplying the amount carried by the carry rate of 0.6%, and

(ii) a spread charge, in accordance with (b) above, where the carried position is matched against a position in an adjacent maturity band.

The firm may repeat the procedure for carrying positions through to other maturity bands as appropriate. An additional carry charge and spread charge must be calculated at each stage of the process.

(c) The firm must multiply any positions remaining after the permitted offsetting by the outright rate of 15%, and then by the spot price of the commodity, to arrive at the outright charge.

(d) The total *PRR* for each commodity is the sum of the spread risk charge, the carry charge, and the outright charge converted to the firm's reporting currency at current spot rates.

An Extended Maturity Ladder Approach is the same as a Maturity Ladder Approach, except that it can assign different, more relaxed charges to different commodities such as precious metals, base metals, soft commodities and other commodities.

We propose an algorithm that minimises the regulatory capital required for both approaches.

5.3. Problem Description and Mathematical Formulation

We start by introducing some notation.

As explained in the rule, offsetting positions within a group is free. Hence, assume the open position of a single band is represented by one real number; i.e. let $a_1, a_2, ..., a_n$ be open positions in the corresponding n bands. Without loss of generality, assume that $\sum_{i=1}^{n} a_i \geq 0$.

The PRR (risk capital) is defined as the sum of the *outright charge*, the *carry charge* and the *spread risk (matched amount) charge* with the rates of OC, CC and SC per unit respectively. Here, $OC = 15\%$, $CC = 0.6\%$ and $SC = 1.5\%$. These charges should be converted into currency by multiplying them by the corresponding spot price. For simplicity, we assume that $price = 1$ *unit* in the firm's reporting currency and omit it from the subsequent analysis. As noted above, the *spread risk charge* and the *carry charge* are insignificant compared to the *outright charge*, which makes it optimal to offset opposite positions across all bands under the realistic assumption that the potential length of the ladder $n < 45$ (see Appendix D1). This assumption may be relaxed depending on an individual case. In notation the condition is $n < \left[\frac{2(OC-SC)}{CC} + 1\right]$. Offsetting opposite positions is accomplished by carrying the desired portion of the open position to the desired band. Following the argument of optimality of complete offsetting, we conclude that all open positions that remain after offsetting, must have the same sign as the overall sum of a_i positions. Following our assumption $\sum_{i=1}^{n} a_i \geq 0$, all remaining positions will be non-negative (long).

Next we discuss how to model the flows of positions between bands. According to regulations, the order of offsetting has no effect on the PRR calculation. The PRR is affected by the number of units moved from a band to a band, and the distance between these bands[2]. The firm pays for both the distance and the quantity moved. For instance, moving 5 units from band 1 to band 3 costs $5\times(3\text{-}1)\times$(the cost of a unit move to the next band). This flow can be broken down into two flows: 5 units from band 1 to band 2, and 5 units from band 2 to band 3. The charge for the two submoves equals exactly the charge for the original move. Without loss of generality, in the rest of the chapter we will consider only the flows between adjacent bands.

[2]The distance is measured as a difference between the ordinal positions of the corresponding two bands. For example, the distance between band 3 and band 1 is $2 = 3 - 1$.

Before proceeding further, we introduce more notation. Let the final (optimal) disposition be denoted $b_1, b_2, ..., b_n$ where, as expected, $\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} a_i \geq 0$. All the b_i's are non-negative, with zero b_i's corresponding to the initial non-positive positions of the a_i's (see Appendix D1). Also, $0 \leq b_i \leq |a_i|$. The connection between the a_i's and the b_i's is illustrated in Figure 5.1. Here each a_i is broken down into two terms, b_i and f_i, where the latter represents the optimal flow (amount transferred) from the ith band to the $(i+1)$st band.

$$
\begin{array}{ccccccc}
a_1 \longrightarrow & a_2 + f_1 \longrightarrow & a_3 + f_2 \longrightarrow & \cdots \longrightarrow & a_n + f_{n-1} \\
\downarrow & \downarrow & \downarrow & & \downarrow \\
b_1 & b_2 & b_3 & \cdots & b_n .
\end{array}
$$

Figure 5.1

$f_1 = a_1 - b_1$, $f_2 = a_2 + f_1 - b_2$, $f_3 = a_3 + f_2 - b_3$, ... , $f_n = a_n + f_{n-1} - b_n$. As explained earlier, flows are defined for the neighbouring bands only and are directed. Hence, negative f_i means positive flow from the $(i+1)$st band to the ith.

The objective is to calculate the minimal risk capital (PRR) so that the capital adequacy standards are satisfied. The risk capital minimisation problem is as follows:

$$
\underset{f_i, b_i}{Min}\ PRR \tag{5.1}
$$
$$
s.t.\ \{a_1, a_2, ..., a_n \in R\}
$$

For non-negative $b_1, b_2, ..., b_n$, the PRR consists of the following: the *outright charge* of $(\sum b_i) * OC$, the *carry charge* of $(\sum |f_i|) * CC$, and the *spread risk (matched amount) charge* of $(\sum |a_i - b_i|) * SC$. The outright charge and the spread risk charge are uniquely defined for the optimal disposition of the b_i's and cannot be minimised. On the other hand, the cost associated with carrying the amount between the bands is controllable and can be minimised for an optimal allocation of the b_i's. Expressing it mathematically, optimal $PRR = Const_1$[3]$+Const_2 * \sum_{i=1}^{n} |f_i|$.

As the term by term breakdown of PRR shows, its minimisation is equivalent to the minimisation of its carry charge $\sum_{i=1}^{n} |f_i|$. Subsequently, the risk minimisation problem (5.1) may be rewritten as

$$\begin{aligned} & \underset{f_i,b_i}{Min} \sum_{i=1}^{n} |f_i| \\ s.t.\ f_i &= \sum_{j=1}^{i} a_j - \sum_{j=1}^{i} b_j,\ i = 1, ..., n \\ b_1, b_2, ..., b_n &\geq 0 \\ & f_i \text{ sign free} \end{aligned}$$

Notice that b_n is solved from the constraint $\sum_{i=1}^{n} b_i = \sum_{i=1}^{n} a_i$ (i.e. once $b_1, b_2, ..., b_{n-1}$ are found, then $b_n = \sum_{i=1}^{n} a_i - \sum_{i=1}^{n-1} b_i$). This makes b_n a redundant decision variable. Following this argument, the above problem is equivalent to the following

[3] $Const_1$ is the fixed cost associated with both the *outright charge* and the *spread amount charge*.

$$
\begin{aligned}
& \underset{f_i, b_i}{Min} \sum_{i=1}^{n-1} |f_i| && (5.2) \\
s.t. \ f_i &= \sum_{j=1}^{i} a_j - \sum_{j=1}^{i} b_j, \ i = 1, ..., n-1 \\
b_n &= \sum_{i=1}^{n} a_i - \sum_{i=1}^{n-1} b_i \\
b_1, b_2, ..., b_{n-1} &\geq 0 \\
& f_i \text{ sign free}
\end{aligned}
$$

5.4. Linear Programming Approach

This optimisation problem can be formulated as a linear programming problem. To remove the module of f_i, we present it as a sum of 'in' and 'out' flows to/from the ith to the $(i+1)$st band. An 'in' flow from the ith to the $(i+1)$st band is an 'out' flow from the $(i+1)$st to the ith. Let x_i be the amount flowing from the ith to the $(i+1)$st band and z_i — from the $(i+1)$st to the ith. Then the net flow from the ith to the $(i+1)$st band will be $f_i = x_i - z_i$ (see Figure 5.2).

$$
a_1 \underset{-z_1}{\overset{x_1}{\rightrightarrows}} a_2 \underset{-z_2}{\overset{x_2}{\rightrightarrows}} a_3 \underset{-z_3}{\overset{x_3}{\rightrightarrows}} \cdots \underset{-z_{n-1}}{\overset{x_{n-1}}{\rightrightarrows}} a_n
$$

Figure 5.2

After rearranging the terms, the new disposition will be a sequence of b_i's, as shown in Figure 5.3:

$$b_1 \text{ —— } b_2 \text{ —— } b_3 \text{ —— } \cdots \text{ —— } b_n$$

$$b_1 = a_1 - x_1 + z_1$$

$$b_2 = a_2 + x_1 - z_1 - x_2 + z_2$$

$$b_3 = a_3 + x_2 - z_2 - x_3 + z_3$$

$$b_n = a_n + x_{n-1} - z_{n-1}$$

Figure 5.3

The objective function to be minimised will be an algebraic sum of the absolute values of all positive and negative flows $\sum_{i=1}^{n}(x_i + z_i)$. The non-negativity requirement for the b_i's, and for the combination of x_i's and z_i's will form the constraints. This gives

$$\begin{aligned}
\underset{x_i, z_i}{Min}\; w &= \sum_{i=1}^{n}(x_i + z_i) \\
s.t. \quad -x_1 + z_1 &\geq -a_1 \\
x_{i-1} - z_{i-1} - x_i + z_i &\geq -a_i, \quad i = 2, ..., n-2 \\
x_{n-1} - z_{n-1} &\geq -a_n \\
x_i, z_i &\geq 0, \ 1 \leq i \leq n
\end{aligned}$$

This approach can also take into account different carry charges from one band to another.

5.5. Algorithm for the Risk Minimisation Problem

In this section we reformulate (5.2) and describe the algorithm that gives all optimal solutions to the problem in closed form. Furthermore, it has all the benefits of the linear programming solution and is easy to implement as a regulatory requirement constraint to the portfolio risk management problem. The algorithm will be generalised to the Theorem on Convex Optimisation (see Chapter 6).

5.5.1. Transformation of the Problem

In this section we transform (5.2) to make combinatorial manipulations easier.

Define $s_i = \sum_{j=1}^{i} a_j$ and $t_i = \sum_{j=1}^{i} b_j$. After setting $t_0 = 0$, (5.2) can be rewritten as

$$\begin{aligned} & \underset{t_i}{Min} \sum_{i=1}^{n-1} |s_i - t_i| \\ s.t. \quad 0 &= t_0 \le t_1 \le t_2 \le ... \le t_{n-1} \le s_n \end{aligned}$$

Define S_i as $S_i = \sum_{j=1}^{i} |s_j - t_j|$. It is obvious that $S_i = S_i(t_1, t_2, ..., t_i)$. For simplicity, let S denote S_{n-1}. Then the optimisation problem becomes:

$$\begin{aligned} & \underset{t_i}{Min} S(t_1, t_2, ..., t_{n-1}) \qquad (5.3) \\ s.t. \quad 0 &= t_0 \le t_1 \le t_2 \le ... \le t_{n-1} \le s_n \end{aligned}$$

The functions S_i play a key role in optimisation. The following section analyses them in detail.

5.5.2. Properties of the Functions S_i

S_i as a Function of One Variable

Before describing the optimisation algorithm, we investigate properties of the functions S_i and provide an insight into optimisation of such functions.

Consider real numbers $(c_1, c_2, ..., c_{n-1})$ such that $0 \le c_1 \le c_2 \le ... \le c_{n-1} \le C$. We claim that the following inequality holds[4]:

$$\underset{\substack{t_1,t_2,...,t_{n-1} \\ s.t.\ 0\le t_1\le t_2\le ...\le t_{n-1}\le C}}{Min} S(t_1, t_2, ..., t_{n-1}) \le \underset{\substack{t_2,...,t_{n-1} \\ s.t.\ 0\le c_1\le t_2\le ...\le t_{n-1}\le C}}{Min} S(c_1, t_2, ..., t_{n-1})$$

The above is obvious because setting the variable t_1 to the constant $t_1 = c_1$ in $\underset{0\le c_1\le t_2\le\cdots\le t_{n-1}\le C}{Min} S(c_1, t_2, ..., t_{n-1})$, restricts t_1 from assuming any value different from c_1 that may yield a smaller value for S. Meanwhile, $\underset{0\le t_1\le t_2\le\cdots\le t_{n-1}\le C}{Min} S(t_1, t_2, ..., t_{n-1})$ is free of similar restrictions and its t_1 variable may take on any value, including c_1. In short, the difference between the LHS and the RHS in the above inequality is the difference between the solutions to unconstrained and constrained optimisation problems.

For simplicity, we denote $\underset{\substack{t_1 \\ s.t.\ 0\le t_1\le s_n}}{Min} S_i(t_1, ..., t_1)$ by $\underset{0\le t_1\le s_n}{Min} S_1(t_1, ..., t_1)$, and apply similar notation in the rest of the chapter.

The issue of one-variable $S_i(t_j, t_j, ..., t_j)$ Vs i-variable $S_i(t_1, t_2, ..., t_i)$ needs to be dealt with delicately as there is a tendency of mixing the functions. Also, it is important which variable is the argument of a one-variable $S_i(t_j, t_j, ..., t_j)$, but introducing new notation in a consistent way may have brought in the model a new set of S tildes and S hats. Thereafter, to avoid confusion we keep using the same notations.

The Shape of S_i

Since all $|s_j - t|$'s have similar shapes, all $S_i(t, ..., t)$'s are continuous convex piecewise linear functions[5]. Therefore S_i's are almost everywhere differentiable and the differentials are

[4] We sacrifice notational precision for simplicity of expression.

[5] Here $t = t_j$, $j = 1, ..., n-1$. The S_i's are one-variable functions and do not depend on the t_j's sequential order, hence notation $S_i(t, ..., t)$.

constants across the segments. The differentials are not defined at the kinks and endpoints, i.e. for $t \in \{s_1, s_2, ..., s_{i-1}, s_i; 0, s_{n-1}\}$. According to the optimisation rules, these are critical points for minimisation along with the solutions to $\frac{dS_i(t,...,t)}{dt} = 0$, if any.

Note that if the s_j's are different from one another, a solution to the FOC exists for even i's only, and it is a whole segment excluding its endpoints[6]. This is made clear in Figure 5.4, which shows the sums of modules, or S_i's for $i = 1, 2, 3, 4$. Derivatives $\frac{dS_i(t,...,t)}{dt}$ are defined and constant across each segment excluding its endpoints. They are zeros only across mid-segments of S_i's (with even subscripts i), as these segments have zero slopes.

5.5.3. Describing the Algorithm

Solving the minimisation problem means finding values of the variables $t_1 = t_1^*, t_2 = t_2^*, ..., t_{n-1} = t_{n-1}^*$ which will minimise (5.3), subject to $0 \leq t_1^* \leq t_2^* \leq ... \leq t_{n-1}^* \leq s_n$. We find the elements of the optimal sequence $\{t_1^*, t_2^*, ..., t_{n-1}^*\}$ one by one starting from t_1^* and finishing with t_{n-1}^*.

Step 1: Optimisation with respect to t_1[7]

Through finding t_1^*, this section explains the intuition behind the solution. Subsequent choices are analogous and derived via induction.

To find the optimal value of t_1, we will consider $S_1(t_1)$, $S_2(t_1, t_1)$, $S_3(t_1, t_1, t_1)$, ... , $S_{n-1}(t_1, t_1, ..., t_1)$.

Choice of t_1^* is crucial for subsequent values of $t_2, t_3, ..., t_{n-1}$ as they cannot go below the limit fixed by t_1^*. The choice of t_1^* should not cease the decreasing potential of any

[6]To express these endpoint analytically, later we introduce the concepts of the *Upper Median* and the *Lower Median*.

[7]For the purposes of illustration only, we assume that all s_i's are different for these solutions at these starting stages of induction only. For a general case we drop this assumption.

$t_2, t_3, ..., t_{n-1}$-dependent S_i function, that eventually affects the objective S function. Not ceasing the S_i's minimisation potential by choosing t_1^* means avoiding the values of t_1 where $S_i(t_1^*, ..., t_1^*)$ increases. If there is such S_i that increases when its argument equals t_1^*, it has already skipped its minimum and can never go back to it for any values of remaining controls $t_2, t_3, ..., t_{n-1}$. In other words, such t_1^* has already incurred irrecoverable loss on the objective function which contains S_i as an addend. Therefore we must restrict t_1 variable to the region where none of the S_i functions is increasing, $i = 1, ..., n-1$. Choosing t_1^* under such constraints will optimise the objective function (the total sum) but not necessarily the individual sums (it may well happen that no individual sum is optimised when the argument is set to t_1^*). But the given scheme guarantees that the sums will be optimised for remaining control variables if it is going to benefit the objective function.

According to the above, we minimise each of S_i $(i = 1, ..., n-1)$ with respect to t_1 independently, and set the minimum of these solutions to t_1^* as follows:

$$t_1^* = min \left\{ \arg \underset{\substack{t_1 \\ s.t.\ 0 \le t_1 \le s_n}}{Min} S_i(t_1, ..., t_1), \quad i = 1, ..., n-1; s_n \right\}$$

In detail,

$$t_1^* = min \left\{ \begin{array}{c} \arg \underset{0 \le t_1 \le s_n}{Min} S_1(t_1, ..., t_1) \\ \arg \underset{0 \le t_1 \le s_n}{Min} S_2(t_1, ..., t_1) \\ \cdots \\ \arg \underset{0 \le t_1 \le s_n}{Min} S_{n-1}(t_1, ..., t_1) \\ s_n \end{array} \right\}$$

Figure 5.4 provides a graphical illustration of $S_1(t_1)$, $S_2(t_1, t_1,)$, $S_3(t_1, t_1, t_1)$ and $S_4(t_1, t_1, t_1, t_1)$ functions of t_1 for a special case, when $s_i \neq s_j, \ \ i, j = 1, ..., 4$.

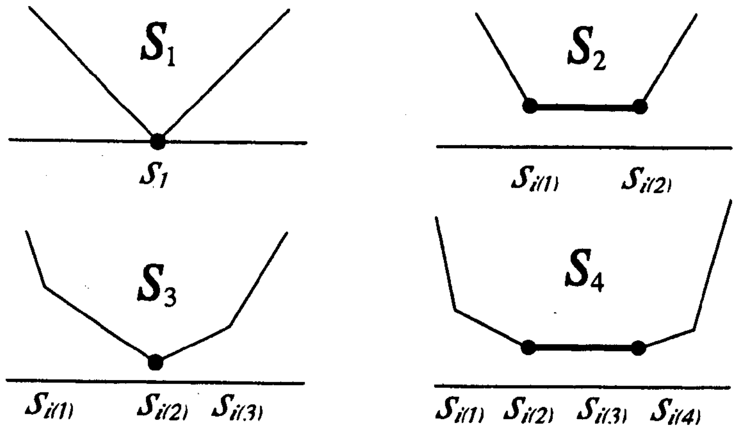

Figure 5.4

First, $S_1 = |s_1 - t_1|$ is minimised. If s_1 is positive, S_1 will be optimal for $t_1 = s_1$. If s_1 is non-positive, S_1 will be minimal for $t_1 = 0$ (since every $t_i \geq 0$). If $t_1(S_j) = \underset{0 \leq t_1 \leq s_n}{Min}\ S_j(t_1, ..., t_1)$, $j = 1, ..., n-1$ denotes the optimal t_1 for S_j, then $t_1(S_1) = \max\{0, s_1\}$.

Second, we choose $t_1(S_2)$. The sum of two modules S_2 is constant and, therefore optimal (minimal) along $[s_1, s_2]$. For $t_1 \geq 0$, S_2 will be optimal for all non-negative t_1's drawn from $[s_{i(1)},\ s_{i(2)}]$ segment[8], or zero, whichever is greater. Thus $t_1(S_2) \in [0, \infty] \cap [s_{i(1)},\ s_{i(2)}]$, or $t_1(S_2) = 0$, if $[0, \infty] \cap [s_{i(1)},\ s_{i(2)}] = \varnothing$

Next we find $t_1(S_3)$. As a sum of three modules, S_3 is optimal at one of the s_i's ($i = 1, 2, 3$) located between the other two (at, say, $s_{i(2)}$ in the ascending sequence of s_i's), or zero. Consequently, $t_1(S_3) = \max\{0, s_{i(2)}\}$.

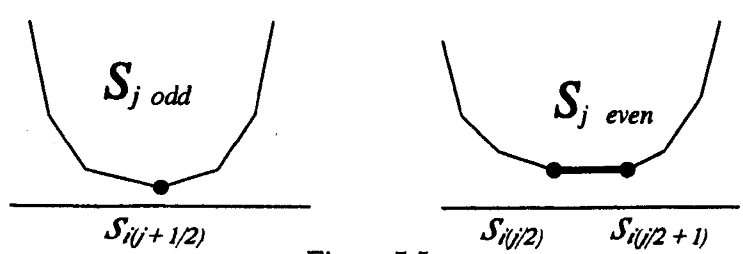

Figure 5.5

Similarly, the other S_j's are optimised with respect to t_1 and for odd j's the optimum is the maximum between zero and the median of the sequence $\{s_{i(1)}, s_{i(2)}, ..., s_{i(j)}\}$, when

[8] $i(.)$ is a permutation of $\{1, 2, ..., n-1\}$ that arranges $\{s_1, s_2, ..., s_{n-1}\}$ in ascending order.

the latter is arranged in ascending order; in other words, for even j's the optimum is the non-negative part of the segment $\left[s_{i(j/2)}, s_{i((j/2)+1)}\right]$ where $s_{i(j/2)}$ and $s_{i((j/2)+1)}$ are $j/2$th and $((j/2)+1)$st points respectively[9] from $\left\{s_{i(1)}, s_{i(2)}, ..., s_{i(j)}\right\}$ sequence arranged in ascending order. For odd j the optimum $t_1(S_j) = \max\left\{0, s_{i((j+1)/2)}\right\}$; For even j, $t_1(S_j) \in [0, \infty] \cap \left[s_{i(j/2)}, s_{i((j/2)+1)}\right]$, or $t_1(S_j) = 0$, if $[0, \infty] \cap \left[s_{i(j/2)}, s_{i((j/2)+1)}\right] = \varnothing$.

As shown above, the optimal t_1 will be the minimum of the values of t_1 that optimise S_i's, and s_n, (the upper limit): $t_1^* = \min\left\{t_1(S_1), t_1(S_2), ..., t_1(S_{n-1}), s_n\right\}$.

This completes the first step of optimisation.

Remark 1 *Note that the process of replacing all t_i's by t_1 will neither affect the subsequent choice of t_i's, nor restrict the functions S_i, because $t_1 \leq t_2 \leq ... \leq t_{n-1}$ and $MinS_i(t_1, t_1, ..., t_1)$ exceeds $S_i(t_1, t_2, ..., t_i)$ for any values of variables.*

Step 2: Optimisation with respect to t_2

Note that when $t_1 = t_1^*$, the sum $S_1(t_1^*) = |s_1 - t_1^*|$ becomes constant, like all first addends of the other sums $|s_i - t_1^*|$, $i = 1, ..., n-1$, and the ith sums $S_i(t_1^*, t_2, ..., t_i)$ become the functions of the $(i-1)$ controls.

Step 2 is the same as Step 1 apart from changing the control variable from t_1 to t_2 and the partial sums $S_1(t_1)$, $S_2(t_1, t_1)$, $S_3(t_1, t_1, t_1)$, ... , $S_{n-1}(t_1, t_1, ..., t_1)$ to $S_2(t_1^*, t_2)$, $S_3(t_1^*, t_2, t_2)$, ... , $S_{n-1}(t_1^*, t_2, ..., t_2)$. As $S_1(t_1^*)$ has been fixed by the previous step, at this stage there are only $(n-2)$ functions S_i's to optimise, $i = 2, ..., n-1$.

[9] As $j/2$ and $(j/2)+1$ are the indices of two points in the middle, later we will introduce the concepts of the Upper Median and the Lower Median to refer to such medians and distinguish between them.

Keeping the same assumption as above regarding distinct kinks, and similar notation $t_2(S_j) = \underset{0\leq t_1^*\leq t_2\leq s_n}{Min} S_j(t_1^*, t_2, ..., t_2)$, $j = 2, ..., n-1$, we note that $\underset{0\leq ...\leq s_n}{Min} S_i$'s with even indices have solutions at one point, and $\underset{0\leq ...\leq s_n}{Min} S_i$'s with odd indices may have solutions across entire segments[10]. Having optima at one point or across a segment depends on the number of remaining control variables in S_i function at the time of optimisation. At the even (e.g. 2nd) stage of optimisation even i-indexed functions give point solutions to the FOCs and, similarly, at the odd (e.g. 1st) stage of optimisation the odd i-indexed functions give point solutions to the FOCs. A similar rule of thumb applies to the segment solutions to FOCs, with apparent modifications. The rest is as in Step 1.

Optimal $t_2^* = \min\{t_2(S_2), ..., t_2(S_{n-1}), s_n\}$, where

$t_2(S_2) = \max\{t_1^*, s_2\}$

$t_2(S_3) \in [t_1^*, \infty] \cap [s_{i(1)},\ s_{i(2)}]$, or $t_2(S_3) = t_1^*$, if $[t_1^*, \infty] \cap [s_{i(1)},\ s_{i(2)}] = \varnothing$

$t_2(S_4) = \max\{t_1^*, s_{i(2)}\}$

...

For even j, $t_2(S_j) = \max\{t_1^*, s_{i(j/2+1)}\}$; For odd j, $t_2(S_j) \in [t_1^*, \infty] \cap [s_{i(j+1/2)}, s_{i(j+1/2+1)}]$, or $t_2(S_j) = t_1^*$, if $[t_1^*, \infty] \cap [s_{i(j+1/2)}, s_{i(j+1/2+1)}] = \varnothing$.

Step k: Optimisation with respect to t_k $(1 \leq k \leq n-1)$[11]

At the kth step we calculate the solution t_k^* as a minimum of

$\{t_k(S_k), t_k(S_{k+1}), ..., t_k(S_{n-1}); s_n\}$.

[10]Here i is a permutation of $\{2, ..., n-2\}$ that arranges $\{s_2, s_3, ..., s_{n-1}\}$ in the ascending order.
[11]As this is a general step, here we allow for equal kinks $s_k = s_j$, $1 \leq k, j \leq n-1$.

$t_k(S_j)$ is defined as:

$$t_k(S_j) = \underset{0\le t_1^*\le\ldots\le t_{k-1}^*\le t_k\le\ldots\le t_j\le\ldots\le s_n}{Min} S_j\left(t_1^*, \ldots, t_{k-1}^*, t_k, \ldots, t_k\right), j = k, \ldots, n-1$$

Figure 5.6 offers another graphical insight into the algorithm. We assume the following ordered sequence of real numbers $\{s_k, s_{k+1}, \ldots, s_n\}$. In Figure 5.6, the solid horizontal line corresponds to $t = t_{k-1}^*$. Points s_j's, $j = k, \ldots, n$ are placed in a sequential order above/below $t = t_{k-1}^*$ line[12].

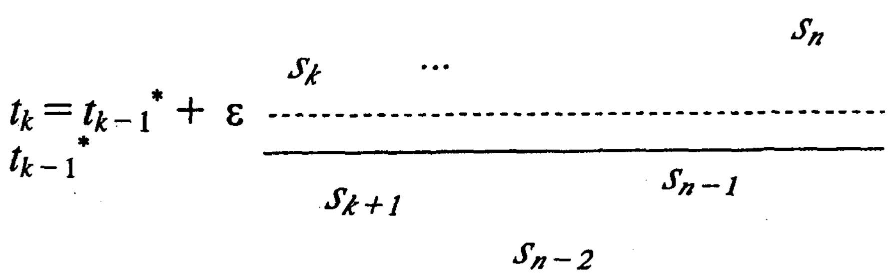

Figure 5.6

t_k is defined as $t_k = t_{k-1}^* + \varepsilon$, $\varepsilon \ge 0$. The dashed horizontal line in Figure 5.6 corresponds to $t = t_k$. If $s_k > t_{k-1}^*$, t_k will approach s_k from left until one of S_j's starts increasing $(j \ge k)$. It is equivalent to the requirement that for any $j \ge k$, the number of $s_k, \ldots, s_j$ located below the line $t = t_k$ is not greater than the number of $s_k, \ldots, s_j$ located above the $t = t_k$ border line. t_k chosen in this way will be a solution $t_k^* = \min\{t_k(S_k), t_k(S_{k+1}), \ldots, t_k(S_{n-1}), s_n\}$.

Once all t_j^*s are found, the corresponding value of the function $S = S_{n-1}(t_1^*, t_2^*, \ldots, t_{i-2}^*, t_{n-1}^*)$ is the optimal solution to the risk minimisation problem (5.2). In the next section we express the solutions t_k^*s in closed form.

[12] s_j's are drawn wrt imaginary vertical axis. The horizontal axis only guarantees that the sequential order $\{k, k+1, \ldots, n\}$ is maintained.

5.5.4. Expressing Solutions Mathematically

Upper and Lower Medians

The following definitions will help us express the optimal values of t_is and, subsequently the optimal flows from one group to another, through rigorous mathematical notation.

Definition 1 *Let $\{x_i, i = 1, ..., n\} \subset R$. Without loss of generality, assume $x_i s$ are arranged in ascending order $\{x_1 \leq x_2 \leq ... \leq x_n\} \subset R$ (if not, re-index them).*

The Upper Median UM of the set $\{x_i, i = 1, ..., n\}$ is defined as

$$UM\{x_i, i = 1, ..., n\} = \begin{cases} x_{k+1}, \ if \ n = 2k \\ x_{k+1}, \ if \ n = 2k+1 \end{cases}$$

Definition 2 *Let $\{x_i, i = 1, ..., n\} \subset R$. Without loss of generality, assume $x_i s$ are arranged in ascending order $\{x_1 \leq x_2 \leq ... \leq x_n\} \subset R$ (if not, re-index them).*

The Lower Median LM of the set $\{x_i, i = 1, ..., n\}$ is defined as

$$LM\{x_i, i = 1, ..., n\} = \begin{cases} x_k, \ if \ n = 2k \\ x_{k+1}, \ if \ n = 2k+1 \end{cases}$$

As it is clear from the definitions, there are at most the same number of points above the Upper Median as below it, and the other way around for the Lower Median. For odd n, $LM = UM$ and both are conventional medians.

Explicit Solution to the Minimisation Problem

Theorem 1 *The full set of solutions to the optimisation problem*

$$\begin{aligned} & \underset{t_1,t_2,\ldots,t_{n-1}}{Min} \sum_{i=1}^{n} |s_i - t_i| \\ s.t.\ 0 &= t_0 \le t_1 \le t_2 \le \ldots \le t_n \le C \end{aligned} \tag{3'}$$

is given by any ascending sequence of t_i^*s $\{(t_1^* \le t_2^* \le \cdots \le t_n^*),\ t_i^* \in D_i\}$ from D,

where

$$D = D_1 \times D_2 \times \ldots \times D_n \subset R^n$$

$$D_i = \begin{bmatrix} Max\{t_{i-1}, Min\{LM_k^1, LM_k^2, \ldots, LM_k^{n-1}, C\}\}, \\ \\ Max\{t_{i-1}, Min\{UM_k^1, UM_k^2, \ldots, UM_k^{n-1}, C\}\} \end{bmatrix} \subset R$$

$$UM_i^j = UM\{s_i, s_{i+1}, \ldots, s_{j-1}, s_j\},\ 1 \le i \le j \le n$$

$$LM_i^j = LM\{s_i, s_{i+1}, \ldots, s_{j-1}, s_j\},\ 1 \le i \le j \le n.$$

Proof. The theorem is a special case of the Theorem on Convex Optimisation formulated and is proved in Chapter 6. ■

The following are the explicit expressions for one set of the optimal solution:

$$t_1^* = Max\{0, Min\{UM_1^1, UM_1^2, \ldots, UM_1^{n-1}, C\}\},$$

$$\ldots$$

$$t_k^* = Max\{0, Min\{UM_k^1, UM_k^2, \ldots, UM_k^{n-1}, C\}\},\ k = 1, \ldots, n-1.$$

5.5.5. Finding the Optimal PRR

Let the f_i^*s denote the optimal flows. After finding all t_i^*s, we solve the optimal flows f_i^*s from the following equations:

$$f_i^* = s_i - t_i^*, i = 1, ..., n-1.$$

Once f_i^*s are found, $b_1, b_2, ..., b_n$ are calculated from the following relationships: $b_1 = a_1 - f_1^*$, $b_2 = a_2 + f_1^* - f_2^*$, $b_3 = a_3 + f_2^* - f_3^*, \cdots$, $b_n = a_n + f_{n-1}^* - f_n^*$. The optimal PRR^* is found after substituting the optimal f_i^*s and b_is in the expression of the PRR:

$$PRR^* = \left(\left(\sum_{i=1}^{n} a_i\right) * OC + \left(\sum_{i=1}^{n} |a_i - b_i|\right) * 2 * SC\right) + \left(\sum_{i=1}^{n} |f_i^*|\right) * CC$$

where, according to the regulations, $OC = 15\%$, $SC = 0.15\%$ and $CC = 0.6\%$.

A numerical example on optimisation will be given in Chapter 6, after having proved the Theorem on Convex Optimisation.

5.5.6. Extensions to the Model

The algorithm is sufficiently general to take into account a carry charge which varies from a band to a band, by assigning different weights to the corresponding flows. This extension is not required under current FSA regulations.

PART 2

5.6. Optimal Grouping of Risk

The Optimal Grouping of the Risk Problem, similar to the Risk Offsetting Problem, comes from the *FSA Handbook* and addresses the problem of optimal grouping in the context of calculating minimal interest rate risk. We formulate the problem and solve it using a linear programming approach. Next we consider its modified version and solve it using dynamic programming techniques. Enhancements and applications are discussed in the end.

5.6.1. Summary of the Corresponding FSA Rule

Here is a brief summary of the FSA rule, from the *FSA Handbook* (2001).

In measuring its positions, a bank may net, by value, long and short positions in the same debt instrument to generate the individual net position in that instrument.

Instruments are considered to be the same where the issuer is the same, they have the equivalent ranking in a liquidation, and the currency, the coupon, and the maturity are the same.

A bank may net by value a long or short position in one tranche of a debt instrument against another tranche of the same instrument where the relevant tranches:

(a) rank pari passu in all respects, and

(b) become fungible within 180 days and thereafter the debt instruments of one tranche can be delivered in settlement of the other tranche.

Where a bank does not have a relevant recognised pre-processing model,trading book positions in derivatives (other than options), and all positions in repos, reverse repos and similar products should be decomposed into their components within each time band prior to the calculation of individual net positions for general market risk.

Where a bank does not have a relevant recognised model, it may nevertheless choose to net notional positions in government bonds (i.e. notional bond legs) that arise from the decomposition of foreign currency forwards, deposit futures, FRAs, swaps and other derivatives.

For the netting of notional government bond positions to be recognised,

(a) the positions should be in the same currency;

(b) their coupons, if any, should be within 15 basis points; and

(c) the next interest fixing date, or residual maturity, should correspond with the limits given in Table 5.2:

Maturity/Interest fixing date	**Permissible mismatch**
Less than one month hence	Same day
One month to one year hence	Within seven days
Over one year hence	Within thirty days

Table 5.2

Interest rate exposures arising from cash borrowing and lending and from cash legs of repo/reverse repo may also be netted against one another (on the same basis as above), but they should not be netted against positions in notional government bonds.

5.7. Problem Description and Mathematical Formulation

Let a_i $(i = 1, ..., n)$ be the open positions where i stands for the maturity/interest fixing date. Without loss of generality, let an open position a_i stand for one particular date i. If $n \leq 30$, the risk of holding corresponding a_i's is the algebraic sum $\sum_{i=1}^{30} a_i$. For $30 < n$ the risk of holding positions a_is could be minimised by netting long and short positions within permissible periods – one week for $n \leq 365$, and one month for $n > 365$.

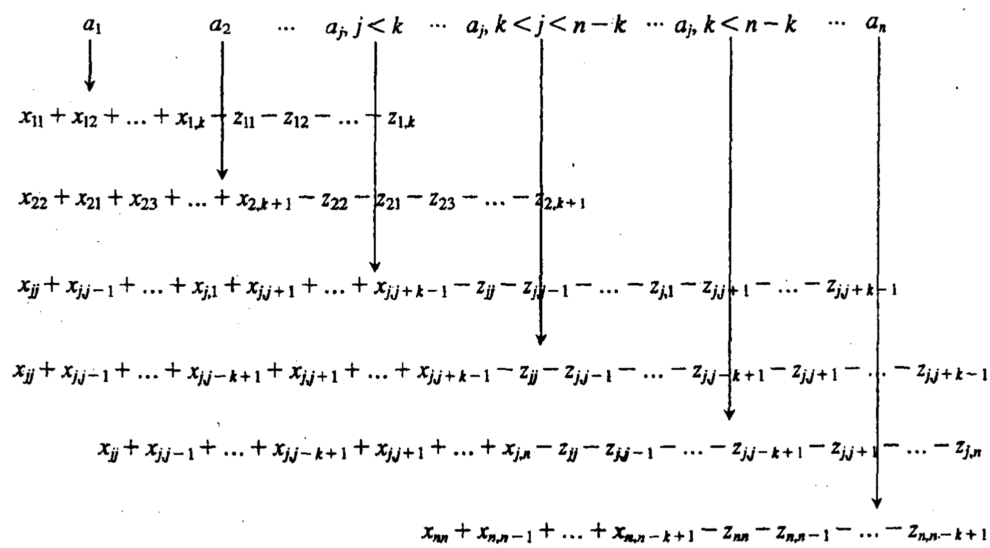

Figure 5.7

Without loss of generality, let the subscripts of $a_1, a_2, ..., a_n$ stand for the maturity/interest fixing date and k be the number of days, within which the permissible mismatch is allowed. In order to determine the optimal netting strategy, we partition a_is into positive x_{ij} and negative $-z_{ij}$ portions, whichever applicable. x_{ij} and z_{ij} represent respectively the absolute values of positive and negative parts of a_i, given that $x_{i,i-j}$ is the positive portion of a_i, that is offset by the the negative $-z_{i-j,i}$ portion of a_{i-j}, $x_{i,i+j}$ is the positive portion of a_i, that is offset by the negative portion $-z_{i+j,i}$ of a_{i+j}. As follows, $-z_{i,i+j}$ is a negative portion of a_i, that is offset by the $x_{i+j,i}$ — positive portion of a_{i+j}, $-z_{i,i-j}$ is a negative portion of a_i, that is offset by the $x_{i+j,i}$ — positive portion of a_{i-j}. Clearly, $j < k$. According to the formulation, the mismatched position at the ith maturity/interest fixing date is either x_{ii} or z_{ii}. In case of $k \geq i$ or $j + i > n$ the process of partitioning into $x_{i,i-j}, x_{i,i+j}$ and $(-z_{i,i-j}), (-z_{i,i+j})$ terminates when $i - j = 1$ (for $x_{i,i-j}$ and $z_{i,i-j}$) and $j = n - i$ (for $x_{i,i+j}$ and $z_{i,i+j}$).

The linear programming model will minimise the objective function which is the sum of absolute values of all positive and negative mismatched positions x_{ii} and z_{ii}. The n constraints ensure that the sum of x_{ij} and $-z_{ij}$ is indeed a_i. Also, a set of constraints forces x_{ij}s to be equal to z_{ji}s. To make sure that a_i is not partitioned into positive and negative parts which cancel each other but may exceed a_i in absolute value, we require that sum of positive/negative parts of a_i be smaller than the module of a_i.

The risk minimisation problem is formulated as follows:

$$\underset{x_{ii}, z_{ii}}{Min} \sum_{i=1}^{n} (x_{ii} + z_{ii}) \tag{5.4}$$

$$\begin{aligned}
s.t.\ x_{ij} - z_{ji} &= 0,\ 1 \leq i,j \leq n,\ i \neq j, |i-j| < k \\
\sum_{j=\max(i-k+1,1)}^{\min(i+k-1,n)} (x_{i,j} - z_{i,j}) &= a_i, \quad i = 1, ..., n \\
\sum_{j=\max(i-k+1,1)}^{\min(i+k-1,n)} x_{ij} &\leq |a_i|, \quad i = 1, ..., n \\
\sum_{j=\max(i-k+1,1)}^{\min(i+k-1,n)} z_{ij} &\leq |a_i|, \quad i = 1, ..., n \\
x_{ij}, z_{ji} &\geq 0,\ 1 \leq i,j \leq n
\end{aligned}$$

For a given set of positions, problem (5.4) is easily solved in any linear programming software.

5.8. Dynamic Programming Approach for a Modified Version

In this section we consider a modified version of the above problem, which will be interesting from the risk management perspective. The problem is solved using the dynamic programming approach.

Let $\Psi(.)$ be a risk-measuring function of one-variable, which maps a (possibly negative) exposure to a non-negative risk measure. C_i's $(j = 1, ..., n)$ are given exposures (cashflows) that may be partitioned into K (or more) groups of length not greater than L for "offset" purposes. Find the grouping which minimises reported risk as

$$Min \sum_{j}^{K} \Psi(g_j) \tag{5.5}$$

$$\begin{aligned} s.t. \quad \sum_{i=l_{j-1}+1}^{l_j} C_i &= g_j;\ j = 1, ..., K \\ 1 &\leq l_j - l_{j-1} \leq L;\ \ j = 1, ..., K \\ l_o &= 0, ..., l_K = N \\ C_i &\in R \end{aligned}$$

In other words, we want to find the best way to position the indices into the sets $G_1, ..., G_k$, so that each set contains no more than L entries. $k \geq K$ to ensure at least K groups. Similarly, the constraint $l_j - l_{j-1} \leq L$ guarantees at most L elements in each group:

$$
\begin{array}{ccccc}
0 & 1 & & \cdots & N \\
l_o & \underbrace{l_0+1 \ldots l_1}_{G_1} & \underbrace{l_1+1 \ldots l_2}_{G_2} & \cdots & \underbrace{l_{k-1}+1 \ldots l_k}_{G_k}
\end{array}
$$

Note that no restrictions are placed on the one-variable function $\Psi(.)$.

The grouping problem is solved using the dynamic programming approach. The dynamic recursive relationship is defined in the following way:

$f(n,k)$ — the optimal risk (as in (5.5)) for exposures $(C_n, ..., C_N)$, when they are arranged in k groups;

$[f(n,k)]$ — an optimal decision associated with $f(n,k)$. Note that there may be more than one optimal decision yielding the same $f(n,k)$.

The dynamics of the process are given by the following equations:

$$f(n, N+1-n) = \sum_{i=1}^{N} \Psi(C_i) \tag{5.6}$$

$$f(n,0) = 0$$

$$f(n,k) = \underset{j=1,\ldots,L;\; n+j \le N}{Min} \left\{ \Psi\left(\sum_{i=n}^{n+j-1} C_i \right) + f(n+j, k-1) \right\},\ n+k \le N \tag{5.7}$$

The optimum of (5.5) is found as

$$\underset{K \le k \le N}{Min} \{f(1,k)\} \tag{5.8}$$

To make the problem easier to visualise, we organise optimal solutions in a NxN truncated matrix, which has a triangular form. The (n, k)th entry of the matrix is $f(n, k)$, with the corresponding $[f(n, k)]$. This helps us find optimal values at every following stage as well as allows us to trace back the chains of optimal groupings.

Following the system dynamics given by (5.6-5.8), we start filling the decision matrix from the last row, and at each subsequent stage of optimisation we will use the optimal decisions that have been made at earlier stages (i.e. the entries from the lower rows).

The Nth row considers only the Nth exposure, hence delivers the trivial solution $f(N, 1) = \Psi(C_N)$ and $[f(N, 1)] = [C_N]$. In the $(N-1)$st row the optimal grouping of Nth and $(N-1)$st exposures is found from alternative ways of grouping: when grouped together (in the $(N-1, 1)$th entry) and separately (in the $(N-1, 2)$th entry). In general, the $(N-p)$th row gives the optimal groupings of $\{N-p, N-p+1, ..., N\}$th exposures, subject to the smallest number of groups varying from 1 to $(p+1)$. The first row gives optimal groupings of all elements, for permitted numbers of groups and the elements within the group.

$$
\begin{array}{cc}
K & \\
\begin{bmatrix}
\begin{matrix} f(1,1), \\ [f(1,1)] \end{matrix} & \cdots & \begin{matrix} f(1,K-1), \\ [f(1,K-1)] \end{matrix} & \begin{matrix} \mathbf{f(1,K)}, \\ [\mathbf{f(1,K)}] \end{matrix} & \cdots & \begin{matrix} \mathbf{f(1,N)}, \\ [\mathbf{f(1,N)}] \end{matrix} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \\
\begin{matrix} f(n,1), \\ [f(n,1)] \end{matrix} & \cdots & \begin{matrix} f(n,K-1), \\ [f(n,K-1)] \end{matrix} & \begin{matrix} \mathbf{f(n,K)}, \\ [\mathbf{f(n,K)}] \end{matrix} & & \\
\cdots & \cdots & \cdots & & & \\
\begin{matrix} f(N,1), \\ [f(N,1)] \end{matrix} & & & & &
\end{bmatrix} & n
\end{array}
$$

Notice that for each (n, k)th entry, k denotes not only the column index of that entry but also the smallest number of groups considered for the optimal decision at this stage.

The final decision is based on the entries where the minimal number of groups is K, i.e. the decision is based on the first row entries that come from the k-indexed columns, where $K \leq k \leq N$ (shown in bold in the matrix).

The algorithm gives all optimal groupings, based on the same decision matrix, if the current constraint "At least K groups" is replaced by either "At most K groups" and "Exactly K groups". The optimal solution for "At most K groups" will become $\underset{1 \leq k \leq K}{Min} \{f(1,k)\}$, and for "Exactly K groups" — $f(1,K)$.

Note that the elements on the main diagonal are described by (5.6), and the first column $(k=1)$ corresponds to (5.7), when $f(n+j,k-1)=0$.

The complexity of the algorithm is $\Omega\left(N^{\frac{N-1}{2}}\right)$.

5.8.1. Numerical Illustration with a Quadratic Risk Function

Here we illustrate on a simple function that the dynamic principle of optimality holds for the algorithm, described above. Let $\sum_j^K \Psi(g_j)$ from (5.6) be $\sum_j^K g_j^2$ — the sum of squared sums g_j^2 of exposures within individual buckets.

Let $N=10$ and the cashflows C_i, $i=1,...,10$ be the following sequence of long and short exposures:

$$5,\ -6,\ -8,\ 5,\ 3,\ 2,\ 5,\ 4,\ -5,\ 2$$

Let the minimal number of buckets be $K=3$ and the number of cashflows within a bucket be restricted to $L=10$.

We rewrite (5.5) as

$$Min\sum_{j}^{K} g_j^2$$

$$\begin{aligned} s.t. \sum_{i=l_{j-1}+1}^{l_j} C_i &= g_j;\ j = 1,...,3 \\ 1 &\leq l_j - l_{j-1} \leq 4;\ j = 1,...,3 \\ l_o &= 0,...,l_K = 10 \\ C_1 &= 5,\ C_2 = -6,\ C_3 = -8,\ C_4 = 5,\ C_5 = 3, \\ C_6 &= 2,\ C_7 = 5,\ C_8 = 4,\ C_9 = -5,\ C_{10} = 2 \end{aligned}$$

To construct the solution matrix for this problem, we calculate $f(n,k)$ functions for the problem, with corresponding decisions $[f(n,k)]$.

$f(n,1)$ entries in the first column are $\left(\sum_{i=n}^{n+j-1} C_n\right)^2$ for corresponding n, $n+j \leq N$, $n = 1,...,N$. These are the *one-group* optimal solutions.

$f(10,1) = \underset{j=1,...,L;\ n+j\leq N}{Min} \left\{\Psi\left(\sum_{i=n}^{n+j-1} C_i\right) + f(n+j, k-1)\right\}$, hence

$f(10,1) = (C_{10})^2 = 2^2 = 4$, decision $[f(10,1)] = [10]$;

$f(9,1) = (C_9 + C_{10})^2 = (-5+2)^2 = 9$, decision $[f(9,1)] = [9,10]$;

...

$f(1,1) = (C_1 + ... + C_{10})^2 = (5 + ... + 2)^2 = 49$, decision $[f(1,1)] = [1,10]$

The remaining entries are found in a similar way. The decision matrix is presented below:

f(1,1) =49 [1 − 10]	**f(1,2)** =29, [1][$f(2,1)$] [1 − 9][$f(10,1)$]	**f(1,3)** =21, [1 − 6][$f(7,2)$]	**f(1,4)** =25, [1,2][$f(3,3)$] [1 − 5][$f(6,3)$]	**f(1,5)** =25, [1,2][$f(3,4)$]	**f(1,6)** =35, [1,2][$f(3,5)$]	**f(1,7)** =53, [1,2][$f(3,6)$]	**f(1,8)** =93 [1,2][$f(3,7)$]	**f(1,9)** =153 [1][$f(2,8)$]	**f(1,10)** =233 [1] [$f(2,9)$]
f(2,1) =4 [2 − 10]	**f(2,2)** =2, [2 − 7][$f(8,1)$]	**f(2,3)** =6, [2 − 7][$f(8,2)$]	**f(2,4)** =46, [2 − 6][$f(7,3)$] [2 − 7][$f(8,3)$]	**f(2,5)** =60, [2][$f(3,4)$]	**f(2,6)** =70, [2][$f(3,5)$]	**f(2,7)** =88, [2][$f(3,6)$]	**f(2,8)** =128 [2][$f(3,7)$]	**f(2,9)** =208 [2] [$f(3,8)$]	
f(3,1) =64 [3 − 10]	**f(3,2)** =40, [3 − 6][$f(7,1)$] [3 − 9][$f(10,1)$]	**f(3,3)** =24, [3 − 6][$f(7,2)$]	**f(3,4)** =24, [3 − 5][$f(6,3)$]	**f(3,5)** =34, [3 − 5][$f(6,4)$]	**f(3,6)** =52, [3,4][$f(5,5)$]	**f(3,7)** =92, [3,4][$f(5,6)$]	**f(3,8)** =172 [3] [$f(4,7)$]		
f(4,1) =256 [4 − 10]	**f(4,2)** =128, [4,5][$f(6,1)$]	**f(4,3)** =86, [4][$f(5,2)$]	**f(4,4)** =70, [4][$f(5,3)$]	**f(4,5)** =58, [4][$f(5,4)$]	**f(4,6)** =68, [4][$f(5,5)$]	**f(4,7)** =108, [4] [$f(5,6)$]			
f(5,1) =121 [5 − 10]	**f(5,2)** =61, [5,6][$f(7,1)$]	**f(5,3)** =45, [5,6][$f(7,2)$]	**f(5,4)** =33, [5][$f(6,3)$]	**f(5,5)** =43, [5][$f(6,4)$]	**f(5,6)** =83, [5][$f(6,5)$]				
f(6,1) =64 [6 − 10]	**f(6,2)** =40, [6][$f(7,1)$] [6 − 9][$f(10,1)$]	**f(6,3)** =24, [6][$f(7,2)$]	**f(6,4)** =34, [6][$f(7,3)$]	**f(6,5)** =74, [6][$f(7,4)$]					
f(7,1) =36 [7 − 10]	**f(7,2)** =20 [7 − 9][$f(10,1)$]	**f(7,3)** =30 [7][$f(8,2)$]	**f(7,4)** =70 [7][$f(8,3)$]						
f(8,1) =1 [8 − 10]	**f(8,2)** =5 [8 − 9][$f(10,1)$]	**f(8,3)** =45 [8][$f(9,2)$]							
f(9,1) =9. [9 − 10]	**f(9,2)** =29, [9][$f(10,1)$]								
f(10,1) =4. [10]									

The optimal solution is the minimum of the first raw entries that come from columns $3, 4, ..., 10$ — $\underset{3 \le k \le 10}{Min} \{f(1,k)\}$. This is $f(1,3) = 21$. The corresponding optimal grouping is $[f(1,3)] = [1-6]\,[f(7,2)]$, where $[f(7,2)]$ was identified at $(7,2)$th stage of optimisation as $[f(7,2)] = [7-9][f(10,1)]$. Similarly, $[f(10,1)] = [10]$.

Therefore, for $K = 3, L = 10$ the optimal grouping of exposures is given by [1-6], [7-9], [10].

If $K = 4, L = 10$, the optimal risk would be given by either $\mathbf{f(1,4)}$ =25 or $\mathbf{f(1,5)}$ =25, and the number of optimal groupings would increase to 3, where 2 solutions contain 4 buckets and 1 contains 5 buckets:

$$\begin{aligned}
[f(1,4)] &= \begin{cases} [1,2],[3-6],[7-9],[10] \\ [1-5],[6],[7-9],[10] \end{cases} \\
[f(1,5)] &= [1,2],[3-5],[6],[7-9],[10]
\end{aligned}$$

If the value of L, the maximum number of exposures within a bucket, changes, a few entries in the decision matrix should be modified in order to find the optimal solution to the same risk measurement problem.

5.8.2. Applications of the Dynamic Programming Algorithm

The proposed algorithm with complexity $\Omega\left(N\frac{N-1}{2}\right)$ identifies efficiently the optimal solution among a finite set of $\left(2^{N-1}\right)$ alternatives.

As mentioned above, the algorithm may lend itself to different applications in optimisation, including those in the area of financial risk management.

The algorithm is important in an environment where there is a need for maintaining the sequential order of elements, that have some risk structure. For example, it can be used for asset/liability management.

5.9. Conclusion

In this chapter we have considered two rules from the *FSA Handbook (2001)* and developed optimal algorithms that calculate minimal capital requirements within these rules. The first algorithm leads to the Theorem on Convex Optimisation, formulated and proved in Chapter 6. A modified version of the second rule is solved using a dynamic programming algorithm with complexity $\Omega\left(N^{\frac{N-1}{2}}\right)$, which efficiently identifies all optimal solutions among a finite set of $\left(2^{N-1}\right)$ alternatives. This algorithm may have a number of applications in the financial risk management. The optimisation methods considered in this chapter are computationally cheap and can effectively deal with modifications in the corresponding rules.

Appendix D

D1. Properties of Optimal $b_1, b_2, ..., b_n$

With given rates of charge, non-zero positions in the final disposition incur the highest outright charge. The flows from one band to the next band cost only 0.6% per unit carried over which makes it optimal to offset all opposite positions (i.e. short versus long).

The following is an example of an extreme allocation and illustrates our point. Assume we are in the setting of Section 5.3. Let $b_1, b_2, ..., b_n$ be the optimal allocation, where b_1 is negative, and the subsequent b_is are zeros except for the last b_n which is positive and greater than the absolute value of b_1 — $b_n \geq |b_1|$ (following the assumption $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i \geq 0$). The outright charge for the disposition $b_1, b_2, ..., b_n$ is $P_1 = |b_1| * OC + b_n * OC$. If b_1 was to be moved to the nth band, the outright charge P_1 would be replaced by $P_2 = (|b_1| * 2 * SC + (b_1 + b_n) * OC + |b_1| * (n-1) * CC)$[13]. Let us compare P_1 to P_2 :

$$P_1 \quad \text{V} \quad P_2$$

$$|b_1| * OC + b_n * OC \quad \text{V} \quad |b_1| * 2 * SC + (b_1 + b_n) * OC + |b_1| * (n-1) * CC$$

$$2 * OC * |b_1| \quad \text{V} \quad |b_1| \{2 * SC + (n-1) * CC\}$$

$$\frac{2*(OC-SC)}{CC} + 1 \quad \text{V} \quad n$$

As the last comparison shows, for any value of negative b_1, P_1 is not worse than P_2 if $n < \frac{2*(OC-SC)}{CC} + 1$. Given that $OC = 15\%$, $CC = 0.6\%$ $SC = 3\%$, this condition is satisfied if the length n does not exceed 45, or $n < 45$. Taking into account the potential length of the maturity ladder, also the prices of a unit of the commodities, we conclude that a realistic

[13] As before, we assume that the gross flow corresponding to the move of a one-unit-position from ith to $(i+j)$th band is j.

n does actually fall within the specified range[14], $P_2 < P_1$ and the disposition $b_1, b_2, ..., b_n$ in not optimal in the sense of minimal PRR. Therefore, the assumption of optimality of $b_1, b_2, ..., b_n$ with sign-varying entries has led us to the contradiction which proves that all opposite positions should be offset in order to optimise the PRR. As a result, all optimal non-zero b_i's should be positive or, in general, have the same sign as $\sum_{i=1}^{n} a_i$. Also note that zero b_i's will correspond to the non-positive a_i's since carrying the positive amount of r to the ith band will not improve anything but only increase the carry charge by at least $r * 0.006$. For the same reason, all $b_i \leq |a_i|$.

D2. Analysis of the PRR

Following our notation, the optimal *outright charge* is $(\sum |b_i|)*0.15$, the *carry charge* — $(\sum |f_i|)*0.006$, and the *spread risk (matched amount) charge* — $(\sum ||a_i| - |b_i||)*2*0.015$. Therefore, the risk minimisation problem is

$$Min\ PRR$$

$$s.t.\ a_1, a_2, ..., a_n$$

$$PRR = \left(\sum_{i=1}^{n} |b_i|\right) * 0.15 + \left(\sum_{i=1}^{n} ||a_i| - |b_i||\right) * 2 * 0.015 + \left(\sum_{i=1}^{n} |f_i|\right) * 0.006$$

[14]Note that the case considered here is extreme and in particular cases (when the opposite positions are not that distant) the restrictions on n can be relaxed.

As $b_1, b_2, ..., b_n$ are positive, the outright charge component of the optimal PRR is uniquely defined for a given $a_1, a_2, ..., a_n$ and is equal to the sum of the a_i's times a constant: $\left(\sum_{i=1}^{n} |b_i|\right) * 0.15 = \left(\sum_{i=1}^{n} b_i\right) * 0.15 = \left(\sum_{i=1}^{n} a_i\right) * 0.15$. Another term in the PRR, the matched amount represents the absolute value of the gross negative positions expressed as

$$\sum_{i=1}^{n} ||a_i| - |b_i|| = \sum_{i=1}^{n} ||a_i| - b_i| = \sum_{i=1}^{n} |a_i - b_i|$$

$$= \sum_{a_i<0} |a_i - 0| + \sum_{a_i \geq 0} (a_i - b_i) = \sum_{a_i<0} |a_i| + \sum_{a_i>0} a_i - \sum_{a_i \geq 0} b_i$$

$$= \sum_{a_i<0} |a_i| + \sum_{a_i \geq 0} |a_i| - \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} |a_i| - \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} |a_i| - \sum_{i=1}^{n} a_i$$

which, similar to the outright charge, is fixed and cannot be minimised. In contrast to it, the carry charge varies depending on the total f_i flows between the maturity bands.

$$PRR = \left(\sum_{i=1}^{n} a_i\right) * 0.15 + \left(\sum_{i=1}^{n} |a_i - b_i|\right) * 2 * 0.015 + \left(\sum_{i=1}^{n} |f_i|\right) * 0.006$$

The above minimisation problem becomes equivalent to the minimisation of its carry charge component. After ignoring the constants in PRR, the problem is rewritten as

$$\begin{aligned} & Min \sum_{i=1}^{n} |f_i| \\ s.t.\ f_i &= \sum_{j=1}^{i} a_j - \sum_{j=1}^{i} b_j,\ i = 1, ..., n \\ b_1, b_2, ..., b_n &\geq 0 \\ & f_i \text{ sign free} \end{aligned}$$

As argued in the main body, $f_n = \sum_{j=1}^{n} a_j - \sum_{j=1}^{n} b_j = 0$, and $\sum_{i=1}^{n} |f_i| = \sum_{i=1}^{n-1} |f_i|$. This delivers the following problem:

$$
\begin{aligned}
& Min \sum_{i=1}^{n-1} |f_i| \\
s.t.\;\; f_i &= \sum_{j=1}^{i} a_j - \sum_{j=1}^{i} b_j,\; i = 1, ..., n-1 \\
b_n &= \sum_{i=1}^{n} a_i - \sum_{i=1}^{n-1} b_i \\
b_1, b_2, ..., b_{n-1} &\geq 0 \\
& f_i \text{ sign free}
\end{aligned}
$$

Chapter 6

Theorem on Convex Optimisation

6.1. Introduction

Chapter 6, the Theorem on Convex Optimisation, was inspired by the risk offsetting algorithm in Chapter 5. The theorem delivers a closed-form solution for a specific combinatorial optimisation problem. It is remarkable, that this optimisation problem imposes no restriction on the sequence of continuous convex functions whose sum is being minimised.

The theorem is accompanied ty the three lemmas that ensure all technicalities in the proof are explicitly dealt with. These lemmas also help us to understand the bigger picture of the optimisation problem.

It is straightforward, that the result can be rewritten as a theorem on concave optimisation (namely, maximisation) in a similar setting.

Overall, the theorem has potential for interesting applications in the area of the finance and operational research.

The chapter is structured in the following way. We start by providing technical preliminaries regarding convex functions. Next the Theorem on Convex Optimisation is formulated, followed by Lemma 1 that develops the intuition behind the Theorem. The result is illustrated with a simple numerical example in Section 3. Discussions of further extensions conclude. The proof is given in Appendix A with all supporting Lemmas.

6.2. Preliminaries

Let $x(t)$ be a convex function, defined on $[a,b]$. We can classify the points $t \in [a,b]$ according to the behaviour of $x(.)$ in some neighbourhood ε_t of the point t.

To proceed, we give the definition of a neighbourhood of $t \in R$ (see, for example, Rockafellar, 1996). The set $\{\tau \in R : |t-\tau| < \rho\}$ is called the open ball, $B(t;\rho)$, of radius ρ about the point t. A set $\varepsilon_t \subset R$ is called a neighbourhood of $t \in \varepsilon_t$ if $B(t;\rho) \subset \varepsilon_t$ for some $\rho > 0$. The definition extends to the definitions of left and right neighbourhoods (ε_t^- and ε_t^+ respectively) in a straightforward way.

$C_1 = \{t \in [a,b] : x \text{ is strictly decreasing in } \varepsilon_t \text{ for some } \rho > 0\}$

$C_2 = \{t \in [a,b] : x \text{ is strictly increasing in } \varepsilon_t \text{ for some } \rho > 0\}$

$C_3 = \{t \in [a,b] : x \text{ is constant in } \varepsilon_t \text{ for some } \rho > 0\}$

$C_4 = \{t \in [a,b] : x \text{ is strictly decreasing in } \varepsilon_t^- \text{ and } x \text{ is constant in } \varepsilon_t^+ \text{ for some } \rho > 0\}$

$C_5 = \{t \in [a,b] : x \text{ is constant in } \varepsilon_t^- \text{ and } x \text{ is strictly increasing in } \varepsilon_t^+ \text{ for some } \rho > 0\}$

$C_6 = \{t \in [a,b] : x$ is strictly decreasing in ε_t^- and x is strictly increasing in ε_t^+ for some

$\rho > 0\}$

Since C_is are convex, they obey the following arrangement rules:

$$C_1\{\begin{matrix} C_4C_3C_5 \\ \\ C_6 \end{matrix}\}C_2,$$ possibly with some empty C_is.

Based on the definition of convexity, below we give inequalities (6.1) and (6.2), which are used in the proof of the theorem. If $x(\tau)$ is convex and $\tau_1 < \tau_2 < \tau_3$, then

$$x(\tau_3) - x(\tau_2) \geq \frac{\tau_3 - \tau_2}{\tau_2 - \tau_1} * [x(\tau_2) - x(\tau_1)] \tag{6.1}$$

$$x(\tau_2) - x(\tau_3) \leq \frac{\tau_3 - \tau_2}{\tau_2 - \tau_1} * [x(\tau_1) - x(\tau_2)] \tag{6.2}$$

6.3. Theorem on Convex Optimisation

Before formulating the theorem, we introduce some notation.

Note, that if $x_1(t), x_2(t), ..., x_n(t)$ are convex, then all sums $\sum_{i=j}^{k} x_i(t)$ are also convex, for $1 \leq j \leq k \leq n$. Denote the set where the convex function $\sum_{i=j}^{k} x_i(t)$ strictly decreases by D_j^k; the set where the convex function $\sum_{i=j}^{k} x_i(t)$ strictly increases by I_j^k.

Theorem 1 (The Theorem on Convex Optimisation) *Let $x_1(t), x_2(t), ..., x_n(t)$ be continuous convex functions on $[a, b] \subset R$.*

Consider the following optimisation problem:

$$Min \sum_{i=1}^{n} x_i(t_i)$$

$$S.t. \;\; a \leq t_1 \leq t_2 \leq ... \leq t_n \leq b$$

where the optimisation is with respect to $t_1, t_2, ..., t_n$ variables.

The full solution to this minimisation problem is given by

$$t_l^* \in \begin{cases} \{b\}, \text{ if } b \in \bigcap_{k=l}^{n} D_l^k \\ \{t_{l-1}^*\}, \text{ if } t_{l-1}^* \in \bigcup_{k=l}^{n} I_l^k \\ [t_{l-1}^*, b] \backslash \left[\left(\bigcap_{k=l}^{n} D_l^k \right) \cup \left(\bigcup_{k=l}^{n} I_l^k \right) \right] \text{ otherwise.} \end{cases} \tag{6.3}$$

where $t_0 = a$ and $l = 1, ..., n$.

Proof. For full proof see Appendix E1. ■

Here we provide the intuition behind the proof of the Theorem. First we will state Lemma 1, which will help us to explain the Theorem on Convex Optimisation.

Lemma 1 *Assume that we are in the setting of the Theorem on Convex Optimisation. Let $x_1(t), x_2(t), ..., x_n(t)$ be any n convex functions, $t_1^* \le t_2^* \le ... \le t_n^*$ be given by (6.3), and $t_{n+1}^* = b$. Consider the above sequence of functions extended by a constant function, i.e. $x_1(t), x_2(t), ..., x_n(t), x_{n+1}(t) = 0$. For any $0 < l \le n+1$, let $t_l', t_{l+1}', ..., t_n', t_{n+1}'$ be any ordered sequence of $(n+2-l)$ points from $[a, b]$ such that $a \le t_l' \le t_{l+1}' \le \cdots \le t_n' \le t_{n+1}' = b$.*

Denote $g(l) = \min(l, n)$ and consider the following (i)-(ii)-(iii) conditions:

(i) $b \in \bigcap_{k=g(l)}^{n} D_{g(l)}^k$;

(ii) $t_l' \ge t_l^*$;

(iii) $t_l' < t_l^*$ *and* $t_l^* \in [t_{l-1}^*, b] \backslash \left[\left(\bigcap_{k=g(l)}^{n} D_{g(l)}^k \right) \cup \left(\bigcup_{k=g(l)}^{n} I_{g(l)}^k \right) \right]$;

If one of the above listed (i)-(ii)-(iii) conditions holds, then the following inequality is satisfied:

$$\sum_{i=l}^{n+1} x_i(t_i^*) - \sum_{i=l}^{n+1} x_i(t_i') \leq 0$$

or, equivalently,

$$\sum_{i=l}^{n+1} x_i(t_i^*) \leq \sum_{i=l}^{n+1} x_i(t_i') \tag{6.4}$$

Moreover, given that at least one of the points $\{t_l', t_{l+1}', ..., t_n'\}$ *is not given by* (6.3)*, then the inequality* (6.4) *is strict.*

Proof. See Appendix E1. ∎

Lemma 1 tells us that the *tail* of the sequence given by (6.3) is optimal in the sense of the Theorem on Convex Opimisation (i.e. it minimises the corresponding *tail sum* of the convex functions). We have had to resort to the concept of best tail in order to apply backwards induction. The best tail of m (m-tail) means that there is no other sequence of $(t_i)_{i\geq m}^n$ points from $[a, b]$ region that gives a 'tail' solution to $Min \sum_{i=n-m}^{n} x_i(t_i)$ superior to the tail of m of the solution obtained by (6.3) (For the sake of simplicity, here we talk about n elements instead of $(n+1)$ as in Lemma 1. However it should not cause confusion as the full proof in Appendix E1 specifies correctly the length of every sequence).

The induction method imposes no bounds on the length of the 'tail', thus allows us to consider the whole sequence of points chosen through (6.3) as the largest 1-tail which is optimal. After having proved Lemma 1, we accept the optimality of m-tail, for $m = 1, ..., n$. Thus we accept the superiority of the 1-tail which, indeed, is the full sequence of the solution given by (6.3).

Note that, for technical reasons, we compliment the set of n arbitrary convex functions $x_i(t)$ with a constant function on $[a, b]$. Without loss of generality, assume this strictly decreasing function is 0 identically. This extra function helps us to anchor the sequence of $x_i(t)$ functions and conduct the first step of backward induction when $m = n + 1$. In Appendix E1 we will show that introducing this extra function in the original set of $x_i(t)$s has no influence over optimality of the solution given by (6.3).

Also note that Lemma 1 proves the optimality of m-tail when (i)-(ii)-(iii) conditions listed above are satisfied. Lemma 2 in Appendix E1 states that one and only one of (i)-(ii)-(iii) conditions can be satisfied for a full sequence. Possible overlaps for smaller 'tails' cause no confusion due to the structure of the proof of Lemma 1. Indeed, in the proof we discuss three mutually exclusive cases that closely relate to (i)-(ii)-(iii) conditions and all possible overlaps are successfully dealt with. Lemma 3 shows the equivalence of two lots of optimal solutions: on one hand, the solution set $\left\{t_l^*, t_{l+1}^*, ..., t_n^*, t_{n+1}^*\right\}$ to the problem $Min \sum\limits_{i=l}^{n+1} x_i(t_i^*)$ and, on the other hand, the solution set $\left\{t_l^*, t_{l+1}^*, ..., t_n^*\right\}$ to the problem $Min \sum\limits_{i=l}^{n} x_i(t_i^*)$.

6.4. Numerical Example

A simple example of the convex optimisation problem given here illustrates how the Theorem on Convex Optimisation works in practice.

Let $t \in [a, b] = [0, 3]$

$x_1(t) = (t-1)^2$

$x_2(t) = (t-2)^2$

$x_3(t) = t + 4$

We have to find $\{t_1, t_2, t_3\}, 0 \le t_1 \le t_2 \le t_3 \le 3$ that minimise the following sum:

$$\begin{aligned}
&\underset{t_1,t_2,t_3}{Min} \sum_{i=1}^{3} x_i(t_i) \\
S.t.\ 0 &= t_0 \leq t_1 \leq t_2 \leq t_3 \leq 3
\end{aligned}$$

According to the definitions of D_i^j and I_i^j,

$$\begin{array}{ll}
D_1^1 = (0,1) & I_1^1 = (1,3) \\
D_1^2 = (0,1.5) & I_1^2 = (1.5,3) \\
D_1^3 = (0,1.25) & I_1^3 = (1.25,3)
\end{array}$$

$$\begin{array}{ll}
D_2^2 = (0,2) & I_2^2 = (2,3) \\
D_2^3 = (0,1.5) & I_2^3 = (1.5,3)
\end{array}$$

$$\begin{array}{ll}
D_3^3 = \varnothing & I_3^3 = (0,3)
\end{array}$$

It is straightforward, that $t_0^* = a = 0$. Also,

$$\begin{aligned}
\bigcap_{k=1}^{3} D_1^k &= (0,1)\,; \bigcap_{k=2}^{3} D_2^k = (0,1.5)\,; D_3^3 = \varnothing \\
\bigcup_{k=1}^{3} I_1^k &= (1,3)\,; \bigcup_{k=2}^{3} I_2^k = (1.5,3)\,; I_3^3 = (0,3)
\end{aligned}$$

The theorem suggests that

$$\begin{aligned}
t_0^* &= a = 0 \notin \bigcup_{k=1}^{3} I_1^k = (1,3)\,; \\
b &= 3 \notin \bigcap_{k=1}^{3} D_1^k = (0,1)\,;
\end{aligned}$$

So,

$$
\begin{aligned}
t_1^* &\in [a,b] \backslash \left[\left(\bigcap_{k=1}^{3} D_1^k \right) \cup \left(\bigcup_{k=1}^{3} I_1^k \right) \right] \\
t_1^* &\in [0,3] \backslash [(0,1) \cup (1,3)] = \{1\}
\end{aligned}
$$

Similarly for t_2^* :

$$
\begin{aligned}
t_1^* &= 1 \notin \bigcup_{k=2}^{3} I_2^k = (1.5, 3)\,; \\
b &= 3 \notin \bigcap_{k=2}^{3} D_2^k = (0, 1.5)\,;
\end{aligned}
$$

So,

$$
\begin{aligned}
t_2^* &\in [a,b] \backslash \left[\left(\bigcap_{k=2}^{3} D_2^k \right) \cup \left(\bigcup_{k=2}^{3} I_2^k \right) \right] \\
t_2^* &\in [0,3] \backslash [(0,1.5) \cup (1.5,3)] = \{1.5\}
\end{aligned}
$$

For t_3^* :

$$
\begin{aligned}
t_2^* &= 1.5 \in I_3^3 = (1,3)\,; \\
&\Rightarrow t_3^* = t_2^* = 1.5
\end{aligned}
$$

We verify that the conditions, given in (11), are mutually exclusive. The following shows that the conditions do not overlap indeed:

$$
b = 3 \notin D_3^3 = \varnothing
$$

6.5. Conclusion

In this chapter we proved the Theorem on Convex Optimisation that gives all optimal solutions in closed form for a particular minimisation problem.

It is straightforward, that the Theorem easily extends to the Theorem on Concave Optimisation (namely, maximisation). It can also deal with the reversed $\{t_1, t_2, ...t_n\}$ sequence (i.e. the sequence put in descending order $b = t_0 \geq t_1 \geq t_2 \geq ... \geq t_n \geq a$).

Appendix E

E1. The Proof of the Theorem

The proof of the Theorem on Convex Optimisation will follow from proving the three lemmas given here. We start with Lemma 1.

Lemma 1 *(formulated in the main body, section 6.3)*

Remark 1 *In the lengthy proof of the lemma, we highlight the 'plot' of the proof in bold font which will make it easier to follow.*

Proof. The Lemma is proved by backwards induction.

At first, we prove that the Lemma 1 is true for $l = (n+1)$ (Step 1). Next assume the correctness of the result for all $l > m$ (so that $1 \leq l \leq n$) and, based on that, prove the result for $l = m$ (Step 2). l is the induction variable.

Step 1:

Let $l = (n+1)$. For any $t'_{n+1} \leq b$ we have $x_{n+1}(t^*_{n+1}) = x_{n+1}(t'_{n+1})$. Therefore $x_{n+1}(t^*_{n+1}) \geq x_{n+1}(t'_{n+1})$ and (6.4) holds.

Step 2:

Consider the following:

m is any fixed number between 1 and n:[1] $1 \leq m \leq n$;

$\{t'_l, t'_{l+1}, ..., t'_n\}$ is an arbitrary sequence where $a \leq t'_l \leq t'_{l+1} \leq ... \leq t'_n \leq t'_{n+1} \leq b$;

and one of the conditions (i)-(iii) is satisfied.

[1]Remember that here the induction variable is l.

The induction assumption is that (6.4) holds for any $l > m$, under the above listed conditions. Then we have to prove that (6.4) holds for any $l = m$.[2] We consider three mutually exclusive cases[3].

Case 1.

Case 1 assumes $b \in \bigcap_{k=m}^{n} D_m^k$ and proves (6.4) for this assumption.

Let $b \in \bigcap_{k=m}^{n} D_m^k \implies t_i^* = b, \; m \le i \le n.$

According to the definition of set $\bigcap_{k=m}^{n} D_m^k$, all sums $\sum_{i=m}^{j} x_i(.)$ are strictly decreasing. This means, for any $\{\tau_1, \tau_2\} \in [a, b]$, $\tau_1 < \tau_2$ and any $m \le j \le n$ we have

$$\sum_{i=m}^{j} x_i(\tau_2) < \sum_{i=m}^{j} x_i(\tau_1).$$

For any $a \le t'_m \le t'_{m+1} \le \dots \le t'_{n+1} \le b$ consider

$$\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{n+1} x_i(t'_i) = \sum_{i=m}^{n+1} x_i(b) - \sum_{i=m}^{n+1} x_i(t'_i).$$

Note, that $t'_i \le t'_{i+1} \implies$

$$\sum_{i=m}^{j} x_i(t'_j) \ge \sum_{i=m}^{j} x_i(t'_{j+1}) \quad \text{for all} \quad 1 \le j \le n \tag{6E.1}$$

$t'_{n+1} \le b \implies$

$$\sum_{i=m}^{n+1} x_i(t'_{n+1}) \ge \sum_{i=m}^{n+1} x_i(b) \tag{6E.2}$$

[2]Since $m < n + 1$ at this stage of the proof, $g(m) = \min(m, n) = m$.

[3]Here we prove the lemma for three cases suggested by conditions (i)-(ii)-(iii). To avoid any confusion regarding any overlap between the sets in conditions (i)-(ii)-(iii), we make the *Cases* mutually exclusive states of the world that span the whole world.

In the inequalities $(6E.1)$ the right hand side of the i-th inequality differs from the left hand side of the $(i+1)$-th inequality only by the term $x_{i+1}(t'_{i+1})$.

Therefore, by adding up the LHSs and RHSs of all inequalities from $(6E.1)$ respectively, we get

$$\sum_{i=m}^{n} x_i(t'_j) \geq \sum_{i=m}^{n} x_i(t'_{j+1}) \tag{6E.3}$$

Combining $(6E.3)$ and $(6E.2)$ returns:

$$\sum_{i=m}^{n+1} x_i(t'_i) \geq \sum_{i=m}^{n+1} x_i(b) \tag{6E.4}$$

or $\sum\limits_{i=m}^{n+1} x_i(b) - \sum\limits_{i=m}^{n+1} x_i(t'_i) \leq 0$, which is the inequality (6.4) for the case $b \in \bigcap\limits_{k=m}^{n} D^k_m$.

To complete the proof for Case 1, we need to show the strict inequality in $(6E.4)$ if at least one of t'_j is not given by (6.3). If we assume t'_j is not given by (6.3) for some j, then $t'_j < b$ for that j, i.e. either $(6E.3)$ or one of $(6E.2)$s is strict. Thus $(6E.4)$ too becomes strict.

The lemma is proved for Case 1.

For the rest of the proof we assume $b \notin \bigcap\limits_{k=m}^{n} D^k_m$.

Case 2.

In Case 2 we let $t'_m \geq t^*_m$.

First consider the subcase when $t'_m = t^*_m$. Then $x_m(t^*_m) = x_m(t'_m)$. The LHS of the inequality (6.4) simplifies to the LHS of inequality (6.4) for $l = m + 1$.

$$\begin{aligned}
&\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{n+1} x_i(t_i') \\
= \;& x_m(t_m^*) + \sum_{i=m+1}^{n+1} x_i(t_i^*) - x_m(t_m') - \sum_{i=m+1}^{n+1} x_i(t_i') \\
= \;& \sum_{i=m+1}^{n+1} x_i(t_i^*) - \sum_{i=m+1}^{n+1} x_i(t_i'),
\end{aligned}$$

According to the induction assumption, the last expression is non-negative/strictly positive. **This proves (6.4) for this particular sub-case.**

For the rest of this case assume $t_m' > t_m^*$. Since $t_m^* \notin \bigcap_{k=m}^{n} D_m^k$, there exists q $(m \leq q \leq n+1)$, such that $t_m^* \in \bigcap_{k=m}^{q-1} D_m^k$ and $t_m^* \notin D_m^q$; If $m = q$, then assume $D_m^{m-1} = \varnothing$, and $t_m^* \notin D_m^m$.

Next we introduce an integer r for which $(t_m^* = t_{m+1}^* = \ldots = t_{r-1}^* \leq t_r^*)$ is true. This expression will be used later in this case.

Denote by r the smallest integer $r > m$ such that[4]

$$t_r^* \in [t_{r-1}^*, b] \backslash \left[\left(\bigcap_{k=g(r)}^{n} D_{g(r)}^k \right) \cup \left(\bigcup_{k=g(r)}^{n} I_{g(r)}^k \right) \right]$$

It is straightforward that

$$t_m^* = t_{m+1}^* = \ldots = t_{r-1}^* \leq t_r^*$$

We can show that $r \geq q + 1$.

[4]Its existence is guaranteed, unless all sums are strictly increasing. If they are, then the rest of the proof becomes trivial.

Indeed, ($t^*_m \leq t^*_r$ and $t^*_m \notin D^q_m$) $\implies$ $t^*_r \notin D^q_m$. But $t^*_r \notin I^q_r \implies t^*_m \notin I^q_r$.

Thus $t^*_m \notin D^{r-1}_m \implies r-1 \geq q$. In particular, $r \geq 2$.[5]

The assumption that $t'_m > t^*_m$ implies, there exists $p:\ m \leq p \leq q$ such that $\sum_{i=m}^{j} x_i(t^*_m) \geq \sum_{i=m}^{j} x_i(t'_m)$ for all $m \leq j < p$ but $\sum_{i=m}^{p} x_i(t^*_m) < \sum_{i=m}^{p} x_i(t'_m)$.

We want to show (6E.9).

Rearranging above given two inequalities, we get

$$\sum_{i=m}^{j}[x_i(t^*_m) - x_i(t'_m)] \geq 0 \quad \text{for all } m \leq\ j < p \tag{6E.5}$$

and

$$x_p(t'_m) - x_p(t^*_m) > \sum_{i=m}^{p-1}[x_i(t^*_m) - x_i(t'_m)] \geq 0 \tag{6E.6}$$

From (6.1)

$$x_p(t'_p) - x_p(t'_m) \geq \alpha_p * [x_p(t'_m) - x_p(t^*_m)] \tag{6E.7}$$

From (6.2)

$$x_i(t'_m) - x_i(t'_i) \leq \alpha_i * [x_i(t^*_m) - x_i(t'_m)], \tag{6E.8}$$

Where $\alpha_i = \frac{t'_i - t'_m}{t'_m - t^*_m}$. Notice, that $0 \leq \alpha_m \leq \alpha_{m+1} \leq \ldots \leq \alpha_p$.

To proceed, we consider and relate the following relationships:

$$x_p(t'_p) - x_p(t'_m)$$

applying $(6E.7)$

[5]Also note that for $r = 2$ either condition (ii) or condition (iii) holds, and for $r > 2$ only condition (ii) holds.

$$\geq \alpha_p * [x_p(t'_m) - x_p(t^*_m)]$$

applying $(6E.6)$

$$\begin{aligned}
&\geq \alpha_p * \sum_{i=m}^{p-1}[x_i(t^*_m) - x_i(t'_m)] \\
&= \alpha_{p-1} * \sum_{i=m}^{p-1}[x_i(t^*_m) - x_i(t'_m)] + (\alpha_p - \alpha_{p-1}) * \sum_{i=m}^{p-1}[x_i(t^*_m) - x_i(t'_m)]
\end{aligned}$$

applying $(6E.5)$

$$\begin{aligned}
&\geq \alpha_{p-1} * \sum_{i=m}^{p-1}[x_i(t^*_m) - x_i(t'_m)] \\
&= \alpha_{p-1} * [x_{p-1}(t^*_m) - x_{p-1}(t'_m)] + \alpha_{p-1} * \sum_{i=m}^{p-2}[x_i(t^*_m) - x_i(t'_m)]
\end{aligned}$$

applying $(6E.8)$

$$\begin{aligned}
&\geq x_{p-1}(t'_m) - x_{p-1}(t'_{p-1}) + \alpha_{p-1} * \sum_{i=m}^{p-2}[x_i(t^*_m) - x_i(t'_m)] \\
&= \sum_{i=p-1}^{p-1}[x_i(t'_m) - x_i(t'_i)] + \alpha_{p-1} * \sum_{i=m}^{p-2}[x_i(t^*_m) - x_i(t'_m)] \ldots
\end{aligned}$$

continue these iterations,

$$\begin{aligned}
\ldots \geq\ & \sum_{i=j+1}^{p-1} [x_i(t'_m) - x_i(t'_i)] + \alpha_{j+1} * \sum_{i=m}^{j} [x_i(t^*_m) - x_i(t'_m)] \\
=\ & \sum_{i=j+1}^{p-1} [x_i(t'_m) - x_i(t'_i)] + \alpha_j * \sum_{i=m}^{j} [x_i(t^*_m) - x_i(t'_m)] \\
& + (\alpha_{j+1} - \alpha_j) * \sum_{i=m}^{j} [x_i(t^*_m) - x_i(t'_m)]
\end{aligned}$$

applying $(6E.5)$

$$\begin{aligned}
\geq\ & \sum_{i=j+1}^{p-1} [x_i(t'_m) - x_i(t'_i)] + \alpha_j * \sum_{i=m}^{j} [x_i(t^*_m) - x_i(t'_m)] \\
=\ & \sum_{i=j+1}^{p-1} [x_i(t'_m) - x_i(t'_i)] + \alpha_j * \sum_{i=m}^{j-1} [x_i(t^*_m) - x_i(t'_m)] \\
& + \alpha_j * \ [x_j(t^*_m) - x_j(t'_m)]
\end{aligned}$$

applying $(6E.8)$

$$\begin{aligned}
\geq\ & \sum_{i=j+1}^{p-1} [x_i(t'_m) - x_i(t'_i)] + \alpha_j * \sum_{i=m}^{j-1} [x_i(t^*_m) - x_i(t'_m)] \\
& + [x_j(t'_m) - x_i(t'_j)] \\
=\ & \sum_{i=j}^{p-1} [x_i(t'_m) - x_i(t'_i)] + \alpha_j * \sum_{i=m}^{j-1} [x_i(t^*_m) - x_i(t'_m)] \ldots \\
\ldots \geq\ & \sum_{i=m}^{p-1} [x_i(t'_m) - x_i(t'_i)].
\end{aligned}$$

Relating the last expression to the first, we have

$$x_p(t'_p) - x_p(t'_m) \geq \sum_{i=m}^{p-1} [x_i(t'_m) - x_i(t'_i)].$$

Summing this inequality and $(6E.6)$ delivers

$$x_p(t'_p) - x_p(t^*_m) > \sum_{i=m}^{p-1}[x_i(t^*_m) - x_i(t'_i)].$$

After grouping all terms in the right hand side of the above inequality, we get

$$\begin{aligned}
[x_p(t'_p) - x_p(t^*_m)] - \sum_{i=m}^{p-1}[x_i(t^*_m) - x_i(t'_i)] &> 0 \qquad \Longrightarrow \\
[x_p(t'_p) - x_p(t^*_m)] + \sum_{i=m}^{p-1}[x_i(t'_i) - x_i(t^*_m)] &> 0 \qquad \Longrightarrow
\end{aligned}$$

$$\sum_{i=m}^{p}[x_i(t'_i) - x_i(t^*_m)] > 0 \tag{6E.9}$$

We have just shown the correctness of (6E.9).

Now consider $\left(\sum_{i=m}^{n+1} x_i(t^*_i) - \sum_{i=m}^{n+1} x_i(t'_i)\right)$ **and show it is non-negative, based on** $(6E.9)$ **and properties of** r.

$$\begin{aligned}
& \sum_{i=m}^{n+1} x_i(t^*_i) - \sum_{i=m}^{n+1} x_i(t'_i) \\
= & \sum_{i=m}^{n+1} x_i(t^*_i) - \sum_{i=m}^{p} x_i(t'_i) - \sum_{i=p+1}^{n+1} x_i(t'_i) \\
= & \sum_{i=m}^{n+1} x_i(t^*_i) - \left\{\sum_{i=m}^{p} x_i(t^*_m) + \sum_{i=m}^{p} [x_i(t'_i) - x_i(t^*_m)]\right\} - \sum_{i=p+1}^{n+1} x_i(t'_i) \\
= & \sum_{i=m}^{n+1} x_i(t^*_i) - \sum_{i=m}^{p} x_i(t^*_m) - \sum_{i=m}^{p} [x_i(t'_i) - x_i(t^*_m)] - \sum_{i=p+1}^{n+1} x_i(t'_i)
\end{aligned}$$

applying (6E.9)

$$
\begin{aligned}
&< \sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=p+1}^{n+1} x_i(t_i') \\
&= \sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) + \sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=p+1}^{n+1} x_i(t_i') \\
&= \sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=p+1}^{n+1} x_i(t_i')
\end{aligned}
$$

Since $r \geq p+1$ and $(t_m^* = t_{m+1}^* = ... = t_{r-1}^* \leq t_r^*)$ holds, the above expression becomes:

$$
\begin{aligned}
&\sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=p+1}^{n+1} x_i(t_i') \\
&= \sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=p+1}^{n+1} x_i(t_i') \\
&= \sum_{i=m}^{p} x_i(t_i^*) + \sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=p+1}^{n+1} x_i(t_i') \\
&= \sum_{i=m}^{p} x_i(t_m^*) + \sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=p+1}^{n+1} x_i(t_i') \\
&= \sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=p+1}^{n+1} x_i(t_i')
\end{aligned}
$$

The rest follows from the induction assumptions.

If $r > p+1$, then $t_{p+1}^* = t_m^* < t_{p+1}'$, hence condition (ii) holds for $l = p+1$ The last expression $\sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=p+1}^{n+1} x_i(t_i')$ is non-positive because of the induction assumption for (6.4). Going back to where we started these iterations, yields

$$\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{n+1} x_i(t_i') \geq 0,$$

which was to be proved.

If $r = p+1$, then condition (iii) holds for $l = p+1$. Thus, we can use the induction assumption, i.e. the correctness of (6.4), and claim

$$\sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=p+1}^{n+1} x_i(t_i') \leq 0$$

Again, using the induction assumption of strict inequality for $l = p+1$ immediately delivers a strict version of the inequality (6.4).

The above proves Lemma 1 for Case 2 fully.

Case 3.

In Case 3 let $t_m^* > t_m'$, $t_m^* \in [t_{m-1}^*, b] \backslash \left[\left(\bigcap_{k=m}^{n} D_m^k \right) \cup \left(\bigcup_{k=m}^{n} I_m^k \right) \right]$.

$t_m^* > t_m'$. There may exist $p : \; m \leq p \leq n$ such that $t_j' \leq t_m^*$ for all $m \leq j \leq p$ and $t_{p+1}' > t_m^*$. If not, assume $p = n+1$.

These two relationships $t_m^* \notin \bigcup_{k=m}^{n} I_m^k$, on one hand, and on the other hand, $t_j' \leq t_m^*$ for all $m \leq j \leq p$, imply $t_j', t_{j+1}' \notin \bigcup_{k=m}^{n} I_m^k$. The latter, combined with $t_j' \leq t_{j+1}'$, delivers:

$$\sum_{i=m}^{j} x_i(t_j') \geq \sum_{i=m}^{j} x_i(t_{j+1}') \;\text{ for all }\; m \leq j \leq p-1 \tag{6E.10}$$

$t_p' \leq t_m^* \implies$

$$\sum_{i=m}^{p} x_i(t_p') \geq \sum_{i=m}^{p} x_i(t_m^*) \tag{6E.11}$$

In $(6E.10)$, the right hand side of the j-th inequality differs from the left hand side of the $(j+1)$st inequality only by $x_{j+1}(t_{j+1}')$ term. Therefore, by adding up all inequalities from $(6E.10)$ we derive

$$\sum_{i=m}^{p-1} x_i(t_i') \geq \sum_{i=m}^{p-1} x_i(t_p') \tag{6E.12}$$

Remark. If any of t_j' is not given by (6.3) $\implies t_j' \in \bigcap_{k=l}^{n} D_l^k$ for that j, i.e. either $(6E.12)$ or one of $(6E.11)$s is strict.

Adding $(6E.11)$ and $(6E.12)$, we get

$$\sum_{i=m}^{p} x_i(t_i') \geq \sum_{i=m}^{p} x_i(t_m^*) \tag{6E.13}$$

or

$$\sum_{i=m}^{p} [x_i(t_i') - x_i(t_m^*)] \geq 0 \tag{6E.14}$$

Consider the LHS of (6.4):

$$\begin{aligned}
&\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{n+1} x_i(t_i') \\
= &\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_i') - \sum_{i=p+1}^{n+1} x_i(t_i') \\
= &\sum_{i=m}^{n+1} x_i(t_i^*) - \left\{ \sum_{i=m}^{p} x_i(t_m^*) + \sum_{i=m}^{p} [x_i(t_i') - x_i(t_m^*)] \right\} - \sum_{i=p+1}^{n+1} x_i(t_i') \\
= &\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=m}^{p} [x_i(t_i') - x_i(t_m^*)] - \sum_{i=p+m}^{n+1} x_i(t_i')
\end{aligned}$$

Applying $(6E.14)$ gives:

$$\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{n+1} x_i(t_i') \leq \sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=m}^{p+1} x_i(t_i') \tag{6E.15}$$

Proving the RHS of $(6E.15)$ is non-positive is equivalent to proving (6.4).

Introduce r:

Denote by r the smallest integer $r > m$ such that

$$t_r^* \in [t_{r-1}^*, b] \backslash \left[\left(\bigcap_{k=g(r)}^{n} D_{g(r)}^k \right) \cup \left(\bigcup_{k=g(r)}^{n} I_{g(r)}^k \right) \right]$$

Below we show that the RHS of (6E.15) is non-positive for both $r \geq p+1$ and $r < p+1$ that, once shown, will prove (6.4) for Case 3.

If $r \geq p+1$, then $(6E.15)$ simplifies to

$$\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=p+1}^{n+1} x_i(t_i')$$
$$= \sum_{i=m}^{p} x_i(t_i^*) + \sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=p+1}^{n+1} x_i(t_i')$$

because all t_r^*, for the r's from m to p are equal,

$$= \sum_{i=m}^{p} x_i(t_m^*) + \sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=p+1}^{n+1} x_i(t_i')$$
$$= \sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=p+1}^{n+1} x_i(t_i')$$

For $r > p+1$, $t_{p+1}^* = t_m^* < t_{p+1}'$, hence condition (ii) holds for $l = p+1$.

For $r = p+1$, either condition (ii) or condition (iii) holds for $l = p+1$.

Thus, when $r \geq p+1$ we can use the assumption of the induction (i.e.(6.4)) and claim

$$\sum_{i=p+1}^{n+1} x_i(t_i^*) - \sum_{i=p+1}^{n+1} x_i(t_i') \leq 0$$

which immediately delivers inequality (6.4):

$$\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{n+1} x_i(t_i') \leq 0$$

Next we prove the same when $r < p + 1$.

If $r < p + 1$, then RHS of ($6E$.15) becomes:

$$\begin{aligned}
&\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{p} x_i(t_m^*) - \sum_{i=p+1}^{n+1} x_i(t_i') \\
= \; &\sum_{i=m}^{r-1} x_i(t_i^*) + \sum_{i=r}^{n+1} x_i(t_i^*) - \sum_{i=m}^{r-1} x_i(t_m^*) - \sum_{i=r}^{n+1} x_i(t_i')
\end{aligned}$$

following the same argument as above,

$$\begin{aligned}
&= \sum_{i=m}^{r-1} x_i(t_m^*) + \sum_{i=r}^{n+1} x_i(t_i^*) - \sum_{i=m}^{r-1} x_i(t_m^*) - \sum_{i=r}^{n+1} x_i(t_i') \\
&= x_i(t_m^*) + \sum_{i=m+1}^{r-1} x_i(t_m^*) + \sum_{i=r}^{n+1} x_i(t_i^*) - x_i(t_m^*) - \sum_{i=m+1}^{r-1} x_i(t_m^*) - \sum_{i=r}^{n+1} x_i(t_i') \\
&= \sum_{i=m+1}^{r-1} x_i(t_m^*) + \sum_{i=r}^{n+1} x_i(t_i^*) - \sum_{i=m+1}^{r-1} x_i(t_m^*) - \sum_{i=r}^{n+1} x_i(t_i') \\
&= \sum_{i=m+1}^{r-1} x_i(t_i^*) + \sum_{i=r}^{n+1} x_i(t_i^*) - \sum_{i=m+1}^{r-1} x_i(t_m^*) - \sum_{i=r}^{n+1} x_i(t_i') \\
&= \sum_{i=m+1}^{n+1} x_i(t_i^*) - \left[\sum_{i=m+1}^{r-1} x_i(t_m^*) + \sum_{i=r}^{n+1} x_i(t_i') \right] \qquad (23)
\end{aligned}$$

We change notations as $t_i'' = t_m^*$ if $m + 1 \le i \le p$ and $t_i'' = t_i'$ if $p < i \le n + 1$. The new sequence $\{t_i'', m + 1 \le i \le n + 1\}$ is ordered as well. The change of notation yields:

$$\left[\sum_{i=m+1}^{r-1} x_i(t_m^*) + \sum_{i=r}^{n+1} x_i(t_i') \right] = \sum_{i=m+1}^{n+1} x_i(t_i'')$$

($6E$.16) is equal to $\sum\limits_{i=m+1}^{n+1} x_i(t_i^*) - \sum\limits_{i=m+1}^{n+1} x_i(t_i'')$, which is the RHS of inequality (6.4) for $l = m + 1$. According to the induction assumption,

$$\sum_{i=m}^{n+1} x_i(t_i^*) - \sum_{i=m}^{n+1} x_i(t_i') \le 0$$

From the derivation it is obvious that if either $(6E.13)$ or $(6E.14)$ are strict, then so is (6.4).

If t'_m is not given by (6.3), then $(6E.14)$ is strict $\Longrightarrow$ (6.4) is strict too.

If any of t'_i $(m < i < n+1)$ is not given by (6.3), then Case 1 yields strict inequality $\Longrightarrow$ (6.4) is strict .

Lemma is proved for Case 3.

Hence the lemma is proved. ■

Remark: Lemma 1 is valid for any m, satisfying $0 \leq m < l \leq n$. Hence it is valid when $m = 1$, meaning

$$\sum_{i=1}^{n+1} x_i(t_i^*) \leq \sum_{i=1}^{n+1} x_i(t_i')$$

The latter implies the optimality of (6.3) under conditions (i), (ii) and (iii). Next we show that conditions (i), (ii) and (iii) cover all possible states of the world.

Lemma 2 *Conditions (i), (ii) and (iii) given in Lemma 1 cover all possibilities.*

Proof. Consider $l = 1$. It is straightforward to see that conditions (i), (ii) and (iii) cover all possibilities and they are mutually exclusive. ■

Thus, solution to (6.3) delivers the validity of (6.4) for any other ordered sequence $\{t'_l, t'_{l+1}, ..., t'_n\}$.

It remains to be proved that removing the last function in Lemma 1 does not affect the optimality of the solution to (6.3).

Lemma 3 *Assume we are in the setting of Lemma 1. Consider the set of functions given in Lemma 1:*

$$\{x_1(t), x_2(t), ..., x_n(t), x_{n+1}(t) = 0\}$$

Let $\left\{t_1^*, t_2^*, ..., t_n^*, t_{n+1}^* = b\right\}$ *be a problem solution given by* (6.3) *and extended by* $t_{n+1}^* = b$ *as in Lemma 1.*

Next consider

$$\{x_1(t), x_2(t), ..., x_n(t)\}$$

and $\left\{t_1^*, t_2^*, ..., t_n^*\right\}$, *which is a problem solution given by* (6.3).

We claim that, for any arbitrary sequence $\left\{t_1', t_2', ..., t_n', t_{n+1}'\right\}$ *from Lemma 1, the following inequalities are equivalent to each other:*

$$\sum_{i=l}^{n} x_i(t_i^*) \leq \sum_{i=l}^{n} x_i(t_i') \tag{6E.16}$$

and

$$\sum_{i=l}^{n+1} x_i(t_i^*) \leq \sum_{i=l}^{n+1} x_i(t_i') \tag{6E.17}$$

Proof. The equivalence follows from the fact that the last function $x_{n+1}(b) = (b-b)^2 = x_{n+1}(t_{n+1}^*) = x_{n+1}(t_{n+1}')$ and the fact that x_{n+1} does not affect conditions (i), (ii) and (iii).

Lemma 3 is proved. ■

The validity of the Theorem on Convex Optimisation directly follows from Lemmas 1, 2, 3.

Chapter 7

Conclusion

7.1. General Remarks

Many studies have illustrated the sensitivity of the mean-variance optimisation method of Markovitz (1952) to the data and how this method can emphasise data errors and lead to an incoherent allocation (e.g. Hodges and Brealey, 1973). The sensitivity of the optimal proportions to the mean return values are magnified by the mean returns being difficult to estimate statistically (see, for example, Merton (1980)). Today it is still problematic to get the inputs that have the properties they are supposed to have for optimally translating the historical data into efficient forecasts.

Treynor and Black (1973) were first to raise issues about using a little available information efficiently. In investment management the information issues are largely concerned with generating optimal forecasts. We investigate, what the cost of forecast error is and how we can adjust for it. When exploring forecast errors in both simple (one-shot) and complex (continuous-time) framework, we draw attention to several problems related to estimation bias. The thesis starts with extending the Treynor-Black model to the case of an unknown correlation coefficient; the following chapter studies how to retrieve parameters from the noisy data using Bayesian techniques; then we move on to the input problems in long-term portfolio management where, in a multi-variate continuous-time environment, mathematical techniques, all meaningful on their own right, may in combination lead to biased results and new puzzles, unless treated with caution.

In risk measurement the optimisation applications are less straightforward. We focus on minimising reported risk within the requirements of regulatory bodies. The rest of the section gives a summary of the work contained within the thesis.

The introductory Chapter 1 provides a critical review of the relevant literature and embeds our work within existing research.

Chapter 2 examines the problem of portfolio selection based on over/underconfident forecasts in a mean-variance framework. Early work by Treynor and Black (1973) established a relationship between the correlation of forecasts, the number of independent securities available and the Sharpe ratio which can be obtained. Their analysis was based on the assumption that the correlation between the forecasts and outcomes is known precisely. In practice, given the low levels of correlation possible, an investor may believe himself to have a different degree of correlation from that which he actually has. Using two different metrics, we explore how portfolio performance depends on both the anticipated and realised correlation when these differ. One measure, the Sharpe ratio, captures the efficiency loss, attributed to the change in reward for risk. The other one, the Generalised Sharpe Ratio

(GSR) introduced by Hodges (1997), measures how the welfare of a particular investor is affected by adopting an inappropriate risk profile. More precisely, GSR quantifies the utility obtained under this information set (i.e. different correlation coefficients) in terms of the Sharpe ratio that would give the same utility. We show that these two metrics, the Sharpe ratio and GSR, complement each other and, in combination, provide a fair ranking of existing investment opportunities.

Chapter 3 explores different ways of inferring optimal scaling of forecasts of unobservable expected returns when the quality of our information is uncertain. In a small sample of observed returns, the maximum likelihood estimate of the unknown scaling produces inconsistent and unreliable return forecasts, heavily influenced by the estimation error. Using a Bayesian implementation, we combine the sample information with the investor's forecasting skills to infer the scaling and derive portfolio holdings for the assumptions of uniform and normal prior distributions on the scaling factor. We find that under full use of information, optimal portfolio holdings depend only on the mean and the variance of the posterior distribution of investor's forecasting skills. These holdings remain conservative through changing cross-sectional variation in returns, in which (the variation) the extent of error is unknown. Such a portfolio decision dramatically differs to the one based on the sample information alone. The latter favours either overly aggressive holdings (up to 10 times larger than the holdings under efficient use of information), or no investments at all; it fluctuates largely, depending on unobservable error terms. Our results in this chapter are consistent with others' work who, using different ways of Bayesian refinement, also reduce asset holdings after implementing Bayesian adjustment.

In Chapter 4 we move on to the analysis of forecast errors in long-term continuous-time portfolio optimisation problems. The impact of estimation error in a dynamic setting is particularly severe because of the complexity of the setting in which it is necessary to have time varying forecasts. We take Brennan, Schwartz and Lagnado's structure (1997) as a specific illustration of a generic problem and investigate the bias in long-term portfolio selection models that comes from optimisation with (unadjusted) parameters estimated from historical data. Based on a Monte Carlo simulation analysis, we quantify the degree of bias associated with the optimisation approach in the spirit of Brennan, Schwartz and Lagnado. We find that in their setting the estimation bias may make an investor believe in five times larger investment opportunities compared to reality. According to our results, minor time-variation in investment opportunities inflates substantially when measured with estimated parameters. This kind of bias has not been recognised by the literature and, to our knowledge, our work is the first attempt to raise issues about the extreme degree of bias inherent in this kind of optimisation approach.

Chapter 5 explores optimisation issues in risk measurement. Regulatory bodies set minimum capital standards to protect a financial organisation's depositors and counterparties from the institution's on- and off-balance sheet risks. Chapter 5 investigates how to report the smallest exposure within a rule. The question is answered for two independent rules from the *FSA Handbook* (2001). Using a wide range of mathematical programming techniques, we develop optimal and efficient algorithms that calculate the minimal required risk capital. Both algorithms have interesting generalisations. One leads to the Theorem on Convex Optimisation (see Chapter 6), the other to the elegant risk minimisation problem, which is solved using dynamic programming techniques.

Chapter 6 formulates and proves the Theorem on Convex Optimisation. The theorem applies to n ordered convex continuous functions $x_i(t)$ whose sum $\sum_{i=1}^{n} x_i(t_i)$ is being optimised with respect to the t_i's. The t_i's are drawn from a given interval and are arranged in ascending order. The theorem finds all optimal solutions in closed form. In the course of the proof, we formulate three original lemmas and prove them. The optimality of one of the algorithms, developed in Chapter 5, follows from the Theorem on Convex Optimisation, as its special case. The theorem is interesting from the optimisation perspective and can be used when dealing with uncertainty in a non-stochastic environment. It is easily modified to a theorem on concave maximisation for n ordered concave continuous functions under similar conditions.

7.2. Suggestions for Future Work

This thesis demonstrates that there is much work still to be done in assessing the impact of information processing and optimisation approaches on portfolio selection as well as risk management. The statistical resolution of the problem of bias, coming from optimisation with noisy estimates, would require hundreds of years of market returns data; this inevitably raises the inference issues and requires further research on how to make the best use of the little available information. In this thesis we were able to identify several estimation-related inefficiencies in the area that had been overlooked before. Until many more similar puzzles are uncovered and resolved, we will keep maximising the estimation error without even suspecting we are doing so.

During the course of this research a number of ideas about further work came to light that were not pursued. We outline some of these below.

The Treynor-Black Model

We extended the Treynor and Black (1973) framework to account for a discrepancy between anticipated and true correlation coefficients between forecasts and realised returns.

However, we kept their original assumption that all assets were equally predictable (hence, one correlation coefficient ρ). Introducing non-homogeneity in assets would result in alternative ρ's having different π's. This framework would require a Bayesian analysis. Extending the model to a multi-period setting, with our moderations as in Chapter 2, is another suggestion for future work.

Inferring the Optimal Scaling of Forecasts of Expected Returns in Presence of Ucertainty about the Quality of Our Information

In the framework of our returns model in Chapter 3, several modifications may be made that would enrich the model and make a valuable contribution to the literature on Bayesian learning. Extending the analysis to the discrete-time framework is the most straightforward generalisation of the model.

Adjustment for Bias in Long-Term Portfolio Management

We showed that long-term portfolio optimisation with noisy parameter estimates, as in Brennan, Schwartz and Lagnado (1997), results in substantial bias. None of the current approaches in the literature seems to provide a suitable answer to this puzzle. As we argued earlier, there is no obvious way to adjust for estimation error and further research is required to shed a light on this problem.

7.3. Final Comments

In extending the literature on information analysis and optimisation in the context of portfolio selection and risk measurement, we have revealed some of the problems the forecasts have both in a single-period and long-term framework, and provided results that are consistent with prior literature such as demonstrating that unadjusted forecasts are overly aggressive. Adjusting the portfolio performance measure for forecast errors in the Treynor-Black model, explains why expected portfolio performance differs from the realised one. Bayesian refinement of an unknown risk factor in a one-period returns model dramatically changes the portfolio decision made under the sample information only. The latter largely fluctuates with error realisation, while the former stays irrelevant to errors and favours conservative holdings throughout.

In long-term portfolio selection with continuous-time asset returns models, the literature widely uses the optimisation method, pioneered by Brennan, Schwartz and Lagnado (1997). This thesis provides the first quantification of the degree of estimation bias, inherent in such approach, optimisation with parameters estimated from short-term sample, and aims to persuade other researchers to change this approach to long-term portfolio optimisation.

In application of optimisation to risk measurement, we found the smallest reported risk capital within two rules from the *FSA Handbook*. The solution led to interesting generalisations (e.g. the Theorem on Convex Optimisation).

References

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin, 1993, Network Flows: Theory, Algorithms, and Applications, Prentice Hall.

Ambachtsheer, K., 1974, "Profit Potential In An 'Almost Efficient' Market," Journal of Portfolio Management, 1, 84-87.

Ambachtsheer, K., 1977, "Where Are the Custormers' Alphas?" Journal of Portfolio Management, 4, 52-56.

Anderson, E., L. Hansen, and T. Sargent, 1999, "Robustness Detection and the Price of Risk," Working Paper, University of Chicago.

Avramov, D., 2002, "Stock Return Predictability and Model Uncertainty," Journal of Financial Economics, 64, 423–458.

Barberis, N., 2000, "Investing for the Long Run When Returns Are Predictable," Journal of Finance, 55, 225–64.

Barry, C. B., 1974, "Portfolio Analysis under Uncertain Means, Variances, and Covariances," Journal of Finance, 29, 515–22.

Bawa, V. S., S. Brown, and R. Klein, 1979, Estimation Risk and Optimal Portfolio Choice, North Holland, Amsterdam.

Bekaert, G., and R. Hodrick, 1992, "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets," Journal of Finance, 47, 467-509.

Ben-Tal, A., and A. Nemirovski, 1998, "Robust Convex Optimization," Mathematica of Operations Research, 23, 769-805.

Ben-Tal, A., and A. Nemirovski, 1999, "Robust Solutions of Uncertain Linear Programs," Operations Research Letters, 25, 1-13.

Bernstein P. L., 1992, Capital Ideas: Improbable Origins of Modern Wall Street, The Free Press.

Best, M. J., and R. R. Grauer, 1991, "On the Sensitivity of Mean-Variance-E cient Portfolios to Changes in Asset Means: Some Analytical and Computational Results," Review of Financial Studies, 4, 315–42.

Best, M. J., and R. R. Grauer, 1992, "Positively Weighted Minimum-Variance Portfolios and the Structure of Asset Expected Returns," Journal of Financial and Quantitative Analysis, 27, 513–37.

Bevan, A., and Winkelmann, K. 1998, "Using the Black-Litterman Global Asset Allocation Model: Three Years of Practical Experience," Fixed Income Research, Goldman, Sachs & Co.

Black, F., and R. Litterman, 1990, "Asset Allocation: Combining Investor Views with Market Equilibrium," Discussion paper, Goldman, Sachs & Co.

Black, F., and R. Litterman, 1991, "Global Asset Allocation with Equities, Bonds and Currencies," Fixed-Income Research, Goldman, Sachs & Co.

Black, F., and R. Litterman, 1992, "Global Portfolio Optimization," Financial Analysts Journal, 48, 28–43.

Black, F. and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," Journal of Policital Economy, 81, 637-659.

Blume, M. E., 1971, "On Assesment of Risk," Journal of Finance, 26, 1-10.

Blume, M. E., 1975, "Betas and their Regression Tendencies", The Journal of Finance, 30, 785-795.

Bodie, Z., A. Kane, and A. Marcus, 2001, Investments, 5th edition, Irwin-McGraw-Hill.

Brandt, M., 1999, "Estimating Portfolio and Consumption Choice: a Conditional Euler Equations Approach," Journal of Finance, 54, 1609-1645.

Brennan, M. J., E. S. Schwartz, and R. Lagnado, 1997, "Strategic Asset Allocation," Journal of Economic Dynamics and Control, 21, 1377-1403.

Brennan, M., 1998, "The Role of Learning in Dynamic Portfolio Decisions", European Economic Review, 1, 295-306.

Brennan, M., and Y. Xia, 2001, "Assessing Asset Pricing Anomalies," Review of Financial Studies, 14, 905-945.

Brinson, G. P., L. R. Hood, and G. L. Beebower,1986, "Determinants of Portfolio Performance," Financial Analysts Journal, July-August, 39-44.

Brinson, G. P., B. D. Singer, and G. L. Beebower, 1991, "Determinants of Performance II: An Update," Financial Analysts Journal, May-June, 40-48.

Brown, S., 1979, "The Effect of Estimation Risk on Capital Market Equilibrium," Journal of Financial and Quantitative Analysis, 14, 215–220.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, The Econometrics of Financial Markets, Princeton University Press.

Campbell, J. Y., and R. Shiller, 1988, "Stock Prices, Earnings and Expected Dividends," Journal of Finance, 43, 661-676.

Campbell, J. Y., and L. Viceira, 1999, "Consumptions and Portfolio Decisions when Expected Returns are Time Varying," Quarterly Journal of Economics, 114, 433-495.

Campbell, J. Y., and L. Viceira, 2002, Strategic Asset Allocation: Portfolio Choice for Long-Term Investors, Oxford University Press.

Carhart, M. M., 1997, "On Persistence in Mutual Fund Performance," Journal of Finance, 52, 57-82.

Chen, Z., and L. Epstein, 2002, "Ambiguity, Risk, and Asset Returns in Continuous Time," Econometrica, 70, 1403-43.

Chopra, V. K., 1993, "Improving Optimization," Journal of Investing, 8, 51–59.

Chopra, V. K., and W. T. Ziemba, 1993, "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice," Journal of Portfolio Management, 19, 6–11.

Chopra, V. K., C. R. Hensel, and A. L. Turner, 1993, "Massaging Mean-Variance Inputs: Returns from Alternative Global Investment Strategies in the 1980s," Management Science, 39, 845–855.

Chow, G. C., 1983, Econometrics, New-York: McGraw-Hill.

Cochrane, J., 1999, "New Facts in Finance", Economic Perspectives (Federal Reserve Bank of Chicago), XXIII (3).

Connor, G., 1997, "Sensible Return Forecasting for Portfolio Management," Financial Analysts Journal, 53, 45–51.

Cvitanic, J., A. Lazrak, L. Martellini and F. Zapatero, 2002, "Revisiting Treynor and Black (1973): an Intertemporal Model of Active Portfolio Management," Working Paper, USC.

Daniel, K., and S. Titman, 1997, "Evidence on the Characteristics of Cross Sectional Variation in Stock Returns," Journal of Finance, 52, 1–33.

Detemple, J., R. Garcia, and M. Rindisbacher, 2003, "A Monte Carlo method for optimal portfolios," Journal of Finance, 58, 401-446.

Detemple, J. B., 1986, "Asset Pricing in a Production Economy with Incomplete Information," Journal of Finance, 41, 383-391.

Dimson E., P. Marsh, and M. Staunton, 2002, Triumph of the Optimists: 101 Years of Global Investment Returns', NJ: Princeton University Press.

Dothan M., and D. Feldman, 1986, "Equilibrium Interest Rates and Multiperiod Bonds in a Partially Observable Economy," Journal of Finance, 41, 369-382.

Dumas, B., and B. Jacquillat, 1990, "Performance of Currency Portfolios Chosen by a Bayesian Technique: 1967-1985," Journal of Banking and Finance, 14, 539–58.

Ellsberg, D., 1961, "Risk, Ambiguity and the Savage Axioms," Quarterly Journal of Economics, 75, 643-669.

Epstein L., and S. Zin, 1989, "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," Econometrica, 57, 937-969.

Epstein L., and S. Zin,1991, "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," Journal of Political Economy, 99, 263-286.

Fama, E. F., and K. R. French, 1989, "Business Conditions and Expected Returns on Stocks and Bonds," Journal of Finance, 48, 663-680.

Ferson, W., and C. Harvey, 1991, "Sources of Predictability in Portfolio Returns," Financial Analysts Journal, 47, 49-56.

Frost, P. A., and J. E. Savarino, 1986, "An Empirical Bayes Approach to E cient Portfolio Selection," Journal of Financial and Quantitative Analysis, 21, 293–305.

Frost, P. A., and J. E. Savarino, 1988, "For Better Performance Constrain Portfolio Weights," Journal of Portfolio Management, 15, 29–34.

FSA Handbook, 2001, Consultation Paper 54a, The Investment Business Interim Prudential Sourcebook, Volume 1 of 2, Financial Services Authority, London.

Gallai, R., 2003, Risk Management and Capital Adequacy, Mc Graw Hill.

Garlappi, L., R. Uppal, and T. Wang, 2004. "Portfolio Selection with Parameter and Model Uncertainty: a Multi-prior Approach," Working paper, University of British Columbia.

Gennotte, G., 1986, "Optimal portfolio choice under incomplete information," Journal of Finance, 41, 733-746.

Gilboa, I., and D. Schmeidler, 1989, "Maxmin Expected Utility Theory with Non-Unique Prior," Journal of Mathematical Economics, 18, 141–153.

Goldfarb, D., and G. Iyengar, 2003, "Robust Portfolio Selection Problems," Mathematica of Operations Research, 28, 1-38.

Grauer, R., and F. Shen, 2000, "Do Constraints Improve Portfolio Performance?" Journal of Banking and Finance, 24, 1253-1274.

Grinold, R., and R. Kahn, 1999, Active Portfolio Management: a Quantitative Approach to Providing Superior Returns and Controlling Risk, McGraw Hill, 2nd edn.

Gruber, M. J., 1996, "Another Puzzle: The Growth in Actively Managed Mutual Funds," Journal of Finance, 51, 783-810.

Halldorsson, B., and R. Tutuncu, 2003, "An Interior-Point Method for a Class of Saddle Point Problems," Journal of Optimization Theory and Applications, 116, 559-590.

Hansen, L. P., and T. J. Sargent, 2003, Robust Control and Model Uncertainty in Macroeconomics, MIT-press.

Hansen L. P., and T. J. Sargent, 2001, "Robust control and model uncertainty," American Economic Review, 91, 60-66.

Haugen R., 2001, Modern Investment Theory, Prentice Hall, 5th edn.

He, G., and R. Litterman, 1999, "The Intuition Behind Black-Litterman Model Portfolios," Investment Management Research, Goldman, Sachs & Co.

Hodges, S. D., 1997, "A Generalization of the Sharpe Ratio and Its Applications to Valuation and Risk Measures," Financial Options Research Centre, University of Warwick, preprint 1998/88.

Hodges, S. D., and R. A. Brealey, 1978, "Portfolio Selection in a Dynamic and Uncertin World," in James H. Lorie, and R. A. Brealey, Modern Developments in Investment Management, Dryden Press, Hinsdale, Illionois, 2nd edn.

Hodges, S. D., and R. A. Brealey, 1973, "Portfolio Selection in a Dynamic and Uncertain World," Financial Analysts Journal, 29(2), 50-65.

Ingersoll J. J., 1987, Theory of Financial Decision Making, Rowman & Littlefield.

Jagannathan, R., and T. Ma, 2003, "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," Journal of Finance, 58, 1651–1684.

Jensen, M. C., 1968, "The Performance of Mutual Funds in the Period 1945-1964," The Journal of Finance 23, 389-416.

Jobson, J. D., and R. Korkie, 1980, "Estimation for Markowitz E cient Portfolios," Journal of the American Statistical Association, 75, 544–554.

Jobson, J. D., R. Korkie, and V. Ratti, 1979, "Improved Estimation for Markowitz Portfolios Using James-Stein Type Estimators," Proceedings of the American Statistical Association, 41, 279–292.

Johnson, R., and D. W. Wichern, 1992, Applied Multivariate Statistical Analysis, Prentice Hall.

Jorion, P., 1985, "International Portfolio Diversification with Estimation Risk," Journal of Business, 58, 259–278.

Jorion, P., 1986, "Bayes-Stein Estimation for Portfolio Analysis," Journal of Financial and Quantitative Analysis, 21, 279–92.

Jorion, P., 1991, "Bayesian and CAPM Estimators of the Means: Implications for Portfolio Selection," Journal of Banking and Finance, 15, 717–27.

Jorion, P., 2003, Financial Risk Manager Handbook, Hoboken, NJ: Wiley, 2nd edn.

Kandel, S., and R. F. Stambaugh, 1996, "On the Predictability of Stock Returns: An Asset-Allocation Perspective," Journal of Finance, 51, 385–424.

Karatzas, I. and S. E. Shreve, 1991, Brownian Motions and Stochastic Calculus, Springer-Verlag.

Keim, D., and R. Stambaugh, 1986, Predicting Returns in the Stock and Bond Markets, Journal of Financial Economics, 17, 357-390.

Kim, T. S., and E. Omberg, 1996, "Dynamic Nonmyopic Portfolio Behavior," The Review of Financial Studies, 9, 141-161.

Klein, R. W., and V. S. Bawa, 1976, "The Effect of Estimation Risk on Optimal Portfolio Choice," Journal of Financial Economics, 3, 215–31.

Knight, F., 1921, Risk, Uncertainty and Profit, Houghton Mifflin, Boston.

Lee, W., 2000, Advanced Theory and Methodology of Tactical Asset Allocation. New York: John Wiley & Sons.

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, 47, 13-37.

Litterman, R. and the Quantitative Resources Group, 2003, Goldman Sachs Asset Management, Modern Investment Management: An Equilibrium Approach. New Jersey: John Wiley & Sons.

Liptser, R. S., and Shiryayev, A. N., 1977., Statistics of Random Processes I: General Theory, Springer Verlag.

Liptser, R. S., and Shiryayev, A. N., 1978, Statistics of Random Processes II: Applications, Springer Verlag.

Lutgens, F., and P. Schotman, 2004, "Robust Portfolio Optimization with Multiple Experts," Working Paper, Maastricht University.

Maenhout, P., 2004, "Robust Portfolio Rules and Asset Pricing," Review of Financial Studies, 17, 951-983.

Markowitz, H. M., 1952, "Mean-Variance Analysis in Portfolio Choice and Capital Markets," Journal of Finance, 7, 77–91.

Markovitz, H. M, 1987, Mean Variance Analysis in Portfolio Choice and Capital Markets, Blackwell

Marrison, C., 2002, The Fundamentals of Risk Measurement, McGraw-Hill.

Merton, R. C., 1969, "Lifetime Portfolio Selection under Uncertainty: the Continuous-time Case," Review of Economics and Statistics, 51, 247-257.

Merton, R. C., 1971, "Optimum consumption and portfolio rules in a continuous-time model," Journal of Economic Theory, 3, 373-413.

Merton, R. O., 1980, "On estimating the Expected Return on the Market: an Exploratory Investigation," Journal of Financial Economics, 8, 323-362.

Michaud, R. O., 1989, "The Markowitz Optimization Enigma: Is Optimized Optimal," Financial Analysts Journal, 45, 31–42.

Michaud, R. O., 1998, Effcient Asset Management, Harvard Business School Press, Boston.

Mossin, J., 1966, "Equilibrium in a Capital Asset Market," Econometrica, 35, 768-783.

Mulvey J. M., 1994, "An Asset-Liability Investment System," Interfaces, 24, 22-33.

Mulvey, J. M., and A. Ruszczynski, 1995, "A New Scenario Decomposition Method for Large-scale Stochastic Optimization," Operations Research, 43, 477-490.

Mulvey, J. M., and Vladimirou, H., 1992, "Stochastic Network Programming for Financial Planning Problems," Management Science 38, 1642-1664.

Pastor, L., 2000, "Portfolio Selection and Asset Pricing Models," Journal of Finance, 55, 179–223.

Pastor, L., and R. F. Stambaugh, 2000, "Comparing Asset Pricing Models: An Investment Perspective," Journal of Financial Economics, 56, 335–81.

Rockafellar, R. T., 1996, Convex Analysis, Princeton University Press.

Rustem, B., R.G. Becker and W. Marty, 2000, "Robust Min-max Portfolio Strategies for Rival Forecast and Risk Scenarios," Journal of Economic Dynamics and Control, 24, 1591–1621.

Rustem, B. & M. A. Howe, 2002, Algorithms for Worst-Case Design and Applications to Risk Management, Princeton University Press.

Samuelson, P., 1969, "Lifetime portfolio selection by dynamic stochastic programming," Review of Economics and Statistics, 51, 239-246.

Satchell, S. and A. Scowcroft, 2000, "A Demystification of the Black-Litterman Model: Managing Quantitative and Traditional Construction," Journal of Asset Management, 1, 144-161.

Scherer, B., 2002, "Portfolio Resampling: Review and Critique," Financial Analysts Journal, 58, 98–109.

Sharpe, W. F., 1963, "A Simplified Model of Portfolio Analysis," Management Science, 9, 277-293.

Sharpe, W. F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," Journal of Finance, 19, 425-442.

Sharpe, W. F., 1966, "Mutual Fund Performance," The Journal of Business 39, pp. 119-138.

Sharpe, W. F., 1970, Portfolio Theory and Capital Markets, McGraw-Hill, New York.

Sharpe, W. F., 1994, "The Sharpe Ratio", Journal of Portfolio Management, 21, 49-59.

Sharpe, W. F., G. J. Alexander, J. V. Bailey, 1995, Investments, Prentice Hall, 6th edn.

Stein, C. 1955, "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," Proceeding of the 3rd Berkely Symposium on Probability and Statistics 1, Berkely. University of California Press

Taggart, R. A. Jr., 1996, Quantitative Analysis for Investment Management, Prentice Hall.

Treynor, J., 1965, "How to Rate Management of Investment Funds," Harvard Business Review 43, 63-75.

Treynor, J., 1974, "Analyst's Role in Portfolio Management", Financial Analyst's Handbook, 45, 1343-1352.

Tutuncu, H., and M. Koenig, 2004, "Robust Asset Allocation," Annals of Operational Research, 132, 157-187.

Treynor, J., F. Black, 1973, "How to Use Security Analysis to Improve Portfolio Selection," The Journal of Business, 1, 66-86.

Uppal, R., and T. Wang, 2003, "Model Misspecification and Underdiversification," Journal of Finance, 58, 2465-2486.

Vasicek, O. A., 1973, "A Note on Using Cross-section Information in Bayesian Estimation of Security Betas," Journal of Finance, 28, 1233-39.

Williams, J. T., 1977, Capital Asset Prices with Heterogeneous Beliefs, Journal of Financial Economics, 5, 219-240.

Xia, Y., 2001, "Learning about Predictability: the Effect on Parameter Uncertainty on Dynamic Asset Allocation," Journal of Finance, 56, 205-246.

Ziemba, W. T., and J. M. Mulvey, 1998, Worldwide Asset and Liability Modeling, Cambridge University Press, Cambridge, U.K.

Zellner, A., 1971, An Introduction to Bayesian Inference in Econometrics, New York: Wiley.