QUANTITATIVE PERFORMANCE EVALUATION OF BENCHMARKED ACTIVE FUNDS

By

Ashraf Chaudhry

December 2005

This thesis is submitted for the degree of Doctor of Philosophy of The Australian National University.
Statement of Originality

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University and, to my best knowledge and belief, this material contains no material published or written by another person, except where due reference is made in the thesis.

Ashraf Chaudhry
December 2005
Acknowledgements

I would like to gratefully acknowledge my two supervisors, Professor Tom Smith and Dr Borek Puza, for their assistance and guidance. I would also like to thank Dr Michael Martin and Professor Terry O’Neill for their continued support throughout my undergraduate and postgraduate degrees. To the administrators, Jennifer Hunt and Tracy Skinner, thanks for your friendship and help over the years. To my friends and colleagues, Steven Roberts, Tim Higgins, Shumi Akhtar and Emma Welch, it has been a long road and your encouragement throughout has been invaluable. To my family, a special thank you for everything. Finally, I would like to thank my love, Helen Johnson, for her intellectual and emotional support and for her tireless efforts on numerous drafts of this thesis.
Abstract

This thesis seeks to determine the most appropriate measures of performance for active fund managers who have clearly identifiable passive investment portfolios as benchmarks.

A range of measures is examined including the Sharpe Selection ratio (SSR), Downside Performance Metric (DPM), Student’s t-test, Fama and French regression model, as well as a method that estimates the decay in the probability of underperforming relative to benchmark through time. In this work, the SSR and DPM are found to perform better than the other measures.

A simulation study is then used to assess the power and bias of the SSR and DPM based on variations in the size of a notional dataset and changes in the mean performance of two simulated funds. The DPM is found to be the superior performance measure since it exhibits similar power and bias under the assumption of symmetry in the distribution of excess returns and more power and less bias than the SSR when the distribution of excess returns are skewed.

An alternative performance evaluation technique is then applied based upon Bayesian methods whereby a flexible range of prior beliefs is placed on the skill set of fund managers in conjunction with the observed data in order to obtain a posterior distribution of the DPM and SSR. In the absence of a closed form solution, these posterior performance metrics are estimated using Markov Chain Monte Carlo (MCMC) methods – more specifically, a Gibbs sampling algorithm. When sceptical dogmatic prior beliefs on manager skill are assumed, the Bayesian methodology can be viewed as a technique for separating the lucky managers from the skilled managers, as well as assessing the overall desirability of investment in active management. Those managers with higher returns, longer return histories and less variation in excess returns provide more evidence of skill and will be less influenced by the prior beliefs defined
under the Bayesian model. When *a priori* ignorance is assumed regarding the skills set parameters, the model estimates the parameters based on the performance of all active managers and uses these results to estimate posterior performance measures for each individual manager.

Under this framework, 99% of managers were found to have skill, however, only 27% of managers had enough skill to outperform benchmark net of the fees and transaction costs associated with active managements as well as any losses through trading with other skilled managers. These data-based estimates of the skill set parameters were then used to obtain posterior performance measures for each individual manager as an alternative to traditional performance evaluation techniques.
# Table of Contents

Statement of Originality ........................................................................................................... i  
Acknowledgements ................................................................................................................... ii  
Abstract .................................................................................................................................... iii  
Table of Contents ........................................................................................................................ v  
List of Tables ................................................................................................................................ vii  
List of Figures ................................................................................................................................ viii  
Chapter 1 ....................................................................................................................................... ix  
  Introduction .................................................................................................................................. 1  
    1.1 Introduction .......................................................................................................................... 1  
    1.2 Motivation and contribution ............................................................................................... 5  
Chapter 2 ....................................................................................................................................... 6  
  Background ................................................................................................................................... 6  
    2.1 Introduction .......................................................................................................................... 6  
Chapter 3 ....................................................................................................................................... 13  
  Classical performance evaluation ............................................................................................... 13  
    3.1 Introduction .......................................................................................................................... 13  
    3.2 Development of single-factor models ............................................................................... 14  
    3.3 Development of multi-factor models ............................................................................... 19  
    3.4 Choice of performance measures ....................................................................................... 21  
    3.5 Data ..................................................................................................................................... 22  
    3.6 Empirical results ................................................................................................................... 24  
    3.7 Size, power and bias ............................................................................................................ 30  
    3.8 Simulation results ................................................................................................................ 33  
      3.8.1 Power ................................................................................................................................ 34  
      3.8.2 Bias ................................................................................................................................... 39  
    3.9 Empirical distributional properties of alpha ...................................................................... 44  
    3.10 Discussion .......................................................................................................................... 45  
Chapter 4 ....................................................................................................................................... 47  
  Bayesian methodology and model ............................................................................................ 47  
    4.1 Introduction ......................................................................................................................... 47  
    4.2 Bayes’ theorem and the posterior distribution .................................................................. 50  
    4.3 Markov Chain Monte Carlo methods ............................................................................... 51  
      4.3.1 The Gibbs sampling algorithm .................................................................................... 51  
      4.3.2 Posterior inference and asymptotic results ................................................................. 53  
    4.4 The Bayesian DPM and SSR ............................................................................................... 54  
    4.5 The Bayesian model ............................................................................................................ 56  
      4.5.1 The likelihoood or data .................................................................................................. 56  
      4.5.2 The prior distributions of $\mu$ and $z$ ....................................................................... 57  
      4.5.3 The remaining prior specification ............................................................................... 59  
      4.5.4 The posterior distribution for the general case ............................................................ 62  
      4.5.5 The priors on $p$ and $q$ ............................................................................................... 63  
      4.5.6 Calculation of $\hat{\mu}$ and $\sigma_a$ ............................................................................. 64  
    4.6 Discussion ............................................................................................................................ 65  
Chapter 5 ....................................................................................................................................... 67  
  Bayesian Results ......................................................................................................................... 67  
    5.1 Introduction .......................................................................................................................... 67  
    5.2 Posterior distribution of performance measures when $p$ and $q$ are constant............. 68
5.2.1 Simulation results.......................................................................................69
5.3 Posterior distribution of parameters of interest when p and q are random....80
5.4 Discussion..................................................................................................83
Chapter 6.....................................................................................................88
Summary and discussion..............................................................................88
Appendix A....................................................................................................91
Formulation of mixture distribution..........................................................91
Appendix B....................................................................................................95
  Calculation of the posterior DPM .............................................................95
Appendix C....................................................................................................98
  Calculation of $\mu$ and $\sigma_a$ .................................................................98
Appendix D....................................................................................................102
  Details of the Gibbs Sampler ....................................................................102
Bibliography...............................................................................................108
List of Tables

Table 3.1: Summary statistics of 287 US value and large capitalisation funds over the period January 1998 through to June 2004. Alpha is the monthly fund return minus benchmark.........................................................24

Table 3.2: Top 50 US value and large capitalisation funds ranked by DPM over the period January 1998 through to June 2004. Relative rankings are also provided using the SSR, t-statistic, decay rate and Fama and French three-factor regression performance measures.................................................................28

Table 3.3: Correlations of performance measures on US value and large capitalisation funds over the period January 1998 through to June 2004..........................................................31

Table 3.4: Percentage of times Fund B outranks Fund A where Fund A’s alphas are drawn from a N(0.001, 0.015) and Fund B from a normal distribution with standard deviation equal to 0.015 and mean varying as shown in the table........................................35

Table 3.5: Percentage of times Fund B outranks Fund A where Fund A’s alphas are drawn from a positively skewed mixture distribution with mean 0.001 and standard deviation 0.015 and Fund B from a similar mixture distribution with the same standard deviation, however, with varying means as shown in the table..................................................39

Table 3.6: Percentage of times Fund B outranks Fund A where Fund A’s alphas are drawn from a N(0.001, 0.015) and Fund B’s alphas are drawn from a positively skewed mixture distribution with mean $\mu$ and standard deviation 0.015, with the mean $\mu$ varying as shown in the table..................................................40

Table 3.7: Empirical confidence intervals on managers exhibiting skewness from a bootstrap of the 158 managers with returns history exceeding thirty observations........45

Table 5.1: Values for $\mu$, $\sigma$ and $U$ for different values of $p$ and $q$........................................69

Table 5.2: Comparisons of Bayes DPM performance measure relative to other Bayes measures ($p=0.05$, $q=0.04$)..........................................................71

Table 5.3: Comparisons of Bayes DPM performance measure relative to other Bayes measures ($p=0.05$, $q=0.01$)..........................................................72

Table 5.4: Comparisons of Bayes DPM performance measure relative to other Bayes measures ($p=0.01$, $q=0.008$)..................................................74

Table 5.5: Comparisons of Bayes DPM performance measure relative to other Bayes measures ($p=0.01$, $q=0.002$)..................................................75

Table 5.6: Correlations between Bayesian parameters of interest and traditional performance measures (DPM and SSR) under varying prior assumptions relating to $p$ and $q$..................................................76
Table 5.7: Manager rank using Bayesian DPM and SSR performance measures where \( q \) and \( p \) are \textit{a priori} unknown.

Table 5.8: Comparisons of manager rank using DPM and SSR performance measures.
List of Figures

Figure 3.1: Distributional properties of the average alphas for 287 large value funds from January 1998 to June 2004. Alpha is measured as fund return minus benchmark (Wilshire Large Cap Value Index). .................................................. 25

Figure 3.2: Cumulative absolute performance of the Wilshire Large Capitalisation Value Index against the average performance of the 287 US value and large capitalisation funds over the period January 1998 through to June 2004. ................. 26

Figure 3.3: A comparison of a normal distribution and a positively skewed distribution both with mean equal to 0.1% and standard deviation equal to 1.5%. ........................................ 33

Figure 3.4: Proportion of times that Fund A is ranked above Fund B where both funds alphas are drawn from a N(0.001,0.015). .................................................................................................. 36

Figure 3.5: Proportion of times that Fund A is ranked above Fund B. Fund A is drawn from a N(0.001,0.015) and Fund B from a N(0.002,0.015). ................................................................. 37

Figure 3.6: Proportion of times that Fund B is ranked above Fund A. Fund A is drawn from a N(0.001,0.015) and Fund B from a positively skewed mixture distribution with same mean and variance. ............................................................ 41

Figure 3.7: Percentage bias of SSR and DPM. Alpha simulated from N(0.001,0.015). 42

Figure 3.8: Percentage bias of SSR and DPM. Alpha simulated from a positively skewed distribution with mean equal to 0.001 and standard deviation equal to 0.015... 42

Figure 4.1: Plot of the prior distributions on µ ......................................................... 60

Figure 5.1: Observed and posterior distributions of alpha for two US large capitalisation fund managers over the period January 1998 through to June 2004. Manager 133 had the highest rank based upon classical performance measures and Manager 92 had the highest rank under Bayesian performance measures. ......................... 78

Figure 5.2: Posterior distributions of the DPM and SSR for two US large capitalisation fund managers over the period January 1998 through to June 2004. Manager 133 had the highest rank based upon classical performance measures and Manager 92 had the highest rank under Bayesian performance measures. ....................... 79

Figure 5.3: Posterior distributions of p and q. ................................................................ 81

Figure 5.4: Distribution of average alpha under Bayesian model where the parameters q and p are estimated from the Gibbs sampling algorithm. ................................................ 82

Figure A1: Lognormal distribution with parameters µ = 0 and σ = 0.5 .................. 93
Chapter 1

Introduction

1.1 Introduction

Numerous studies have addressed the investment decision between active and passive funds management. Trillions of dollars\(^1\) are invested in active managers from a vast array of investors – from superannuation funds and life insurance companies to less sophisticated personal investors. The role of active management should be to provide a superior return relative to passive investment in a relevant diversified portfolio of securities. Therefore, the decision to invest in active funds management must be based on the ability of the active fund manager to consistently provide performance above passive investment. This thesis seeks to determine the most appropriate measures of performance for active fund managers who have clearly identifiable passive investment portfolios as benchmarks. Under these conditions, the use of traditional performance evaluation methods may be inappropriate as they do not explicitly account for the benchmarks that the managers are mandated to outperform.

1.2 Motivation and contribution

While the performance of investment fund managers has been subject to a significant body of research, this study differs in many respects. Firstly, this study will investigate the suitability of existing performance measures under the assumption of a clearly

defined benchmark. Secondly, a Bayesian approach will be employed to evaluate performance.

In recent years there has been a trend towards mandating wholesale active fund managers to outperform specific passive investment portfolios whilst maintaining an acceptable level of risk, where risk generally represents significant deviations in performance relative to the passive benchmark. One of the rationales behind these mandates is to provide quality control and risk management in protecting the funds invested with the active manager. Other motivations stem from the beliefs of the investor and include portfolio tilts and/or diversification requirements as advocated under portfolio theory. Under these circumstances, fund manager remuneration is linked directly to performance and risk relative to the mandated benchmark and therefore, careful analysis of the appropriate measures of performance is necessary. The use of conventional multi-factor regression models that are promoted in the extant literature may be inappropriate as they do not explicitly account for performance relative to benchmark. Instead, performance is measured relative to a number of factors including (but not limited to) the excess market return above the risk-free rate of return, value versus growth portfolios, large versus small capitalisation portfolios and bond yield spreads.

This study will investigate the properties of performance measures that explicitly identify the benchmark that the active manager is mandated to outperform. Initially these performance measures will include the Student’s t-statistic and the Foster and Stutzer (2002) Decay Rate. However, the main focus of this thesis will be a comparison between the Sharpe Selection Ratio (SSR) and a Downside Performance Measure (DPM) that is based upon the lower partial moments approach of Bawa (1967). The SSR is estimated as the average excess returns divided by the standard deviation of the excess returns relative to benchmark. The DPM is defined as the average excess return over the square root of the sum of the squared negative excess returns divided by the number of observations.
A comparison of the above measures is conducted by ranking a subset of fund managers within the US market that have clearly identifiable benchmarks. A simulation study is then used to assess the power and bias of the SSR and DPM, based on variations in the size of the notional dataset, and changes in the mean performance of two simulated funds where it is assumed that the excess returns (alphas) of the funds above benchmark are normally distributed. The simulation is then re-evaluated under the assumption of skewness in the underlying distributions of alpha. The results suggest similar power and bias in the SSR and DPM when the underlying distribution of alpha is normal. However, less bias and more power is present in the DPM when the distribution of alpha is skewed. A bootstrap on the US fund manager dataset provides evidence of positive skewness in the empirical distributions of alpha and consequently the DPM is advocated as the choice performance measure.

A Bayesian approach analogous to the work of Baks, Metrick and Wachter (2001) is also employed to evaluate performance. Prior beliefs about the skill set of fund managers are used in conjunction with the observed data to obtain posterior estimates of performance measures. While Baks et al. (2001) obtain a closed form solution using a posterior expected value of alpha from a multi-factor regression as their measure of performance, this study will adopt the DPM and the SSR as the underlying measures to be used in calculating the posterior performance metrics. As outlined earlier, these measures are believed to be more appropriate under the assumption of a clearly defined benchmark. In the absence of a closed form solution, Markov Chain Monte Carlo (MCMC) methods – more specifically a Gibbs sampling algorithm – will be used to estimate the posterior distributions of the DPM and SSR.

Baks et al. (2001) advocate the use of Bayesian methods to assess the desirability of investment in active fund management. Unskilled managers are assumed to underperform benchmarks by the fees and transaction costs associated with active management as well as any losses from trading with skilled managers. The distribution
of the skilled managers is assumed to be half normal above the performance of the unskilled managers. Baks et al. (2001) use this Bayesian model to assess whether investment in active management may be desirable under the most sceptical of prior beliefs on the probability of active management skill.

This Bayesian methodology can be viewed as a technique for separating the lucky managers from the skilled managers when no specific benchmark can be identified. Managers with high returns, longer return histories and less variation in excess returns provide more evidence of skill and will be less influenced by the prior beliefs defined under the Bayesian model. The skill parameters can be adjusted according to the efficiency of the market the manager invests within.

Under the assumption of a clearly identifiable benchmark, the Baks et al. (2001) use of multi-factor regression models is argued to be inappropriate. Therefore, in this study a Bayesian DPM and SSR are estimated for each manager in the US dataset. The probability that a manager has skill as well as the probability that a manager has enough skill to exceed the benchmark are allowed to vary, and the posterior estimates of the DPM and SSR are estimated for each manager. This analysis provides an alternative technique for ranking the managers as well as a tool to test whether investment in active management is desirable based upon prior beliefs on market efficiency.

Through the use of MCMC methods this study will also contribute by allowing the prior beliefs on the skills of active managers to vary. Unlike Baks et al. (2001), who place a point estimate on the priors, the Bayesian model in this study allows the prior probabilities of having skill and enough skill to exceed benchmark to have a probability density distribution. This added complexity allows for a richer set of outcomes, depending on the purpose of the analysis and the confidence relating to the prior beliefs. The initial model described above can be viewed as one extreme of the general model where dogmatic prior beliefs are placed on the prior parameters.
At the other extreme, *a priori* ignorance can be posited on the probability that a manager has skill and the probability that a manager has enough skill to exceed the benchmark return. These parameters can then be estimated based on the data within the Bayesian model and used to obtain a posterior DPM and SSR for each manager. This version of the model also has intuitive appeal as it can be viewed as obtaining the priors relating to skill directly from the observed data of all fund managers within the dataset and then using these priors to estimate the posterior performance measures relating to each specific manager.

Furthermore, for mathematical simplicity Baks et al. (2001) assume the likelihood within the Bayesian model is multivariate normal in order to obtain a closed form solution for their measure of risk-adjusted performance. In this study, such restrictions on the likelihood are not necessary, and any distribution can be assumed for alpha which can then be used within the MCMC simulation algorithm. If there is strong empirical evidence of skewness in alpha, the model can be easily adapted by assuming a skewed distribution to represent the likelihood within the Bayesian model.

1.3 **Organisation of the thesis**

The structure of this thesis will be as follows. Chapter 2 will provide a background on each of the relevant issues that are related to performance evaluation. In Chapter 3, the classical techniques and performance measures will be discussed and calculated through the use of both data and a simulation study. In Chapter 4, the Bayesian methodology and model that will be employed to estimate the posterior distribution of the DPM and SSR will be described. In Chapter 5, the results obtained through the methods introduced in Chapter 4 will be discussed. In Chapter 6, the research that has been carried out in this thesis will be summarised. The technical steps involved in this research are presented in the Appendices.
Chapter 2

Background

2.1 Introduction

Investment managers who are engaged in active selection assume that they can use historical information regarding securities to achieve a superior risk-adjusted return. However, the Efficient Market Hypothesis (EMH) states that in a liquid market that is characterised by a large number of rational, expected utility-maximising participants making unbiased forecasts of the future, stocks will be appropriately priced and will reflect all available information. If the stock market is efficient in an informational sense, then no information or analysis can be expected to result in fund managers earning superior risk adjusted returns or economic rents.

Tests relating to the EMH are generally split into three informational taxonomies as advocated by Roberts (1967). The first is weak-form efficiency where the information set only includes historic prices. The second is semi-strong form efficiency where the information set includes all information known to all market participants (all public information). Finally, under strong-form efficiency the information set includes all information known to any market participant (including private information). Empirical testing of the efficiency of markets conditional on each of the above information sets generally provides support for semi-strong efficiency within developed equity markets.
The semi-strong theory of market efficiency suggests that market prices should fully reflect all public information available at that time. As new information enters the market, prices react simultaneously and adjust to a new “fair” price. This implies that active fund managers should not be able to earn superior returns above benchmark in the long term and should underperform by the transaction costs and fees that are associated with active management. Under this theory, fluctuations in performance around the expected loss will be based solely on chance.

Grossman and Stiglitz (1980) provide an alternative to the EMH in which they argue that prices reflect the information of informed individuals but only partially. Therefore, those managers who expend resources to obtain information do ultimately receive compensation and, as a result, uninformed investors can expect to be on the unprofitable side of a trade relative to the informed investors who have collected information. Thus, informed investors can earn higher risk-adjusted expected returns. However they also incur more expenses. In equilibrium, informed and uninformed investors can expect to earn the same return net of expenses.

Recent empirical evidence suggests that the average active fund manager may actually be able to add value gross of expenses yet underperform net of expenses. From the 1960s and onwards, there have been numerous studies that have found evidence that risk-adjusted outperformance (defined as alpha) are negative but above the fees and transaction costs involved in active management. Initially, performance was evaluated on an absolute basis and managers were ranked on their average returns over the period of investment. Whilst this is an intuitive and obvious measure of performance (supported by Hallahan and Faff, 2001), it fails to adjust for the risk that is associated with the choice of securities that the manager invests in as well as compare performance to the returns forgone in selecting the active fund manager. Consequently, performance measures have evolved to assess performance relative to passive investment in securities of similar characteristics.
The first series of risk-adjusted performance measures was proposed by Treynor (1965), Sharpe (1966) and Jensen (1968). Each of these studies proposed a single factor model based on the ability of the fund manager to outperform a benchmark once an adjustment has been made for risk. The benchmark for the Treynor Index is the risk free rate of return and the measure of risk which stems directly from the Capital Asset Pricing Model (CAPM) is beta. The Jensen measure evaluates performance relative to a risk-adjusted market return. Both the Treynor Index and Jensen’s Alpha rely on the validity of the CAPM in the estimation of their respective performance metrics. Jensen (1972), Admati and Ross (1985) and Dybvig and Ross (1985) found evidence that Jensen’s Alpha can result in spurious negative abnormal returns for managers possessing market timing. Furthermore, the CAPM theory requires the market portfolio to be mean-variance efficient. Given that the market portfolio is unobservable, Roll (1978, 1979) argues that performance metrics based on a market portfolio could be biased due to the inefficiency of the proxy.

During the 1990s, the intercept from a time series regression on a number of relevant factors was used to evaluate outperformance. These multi-factor models suggest that there are multiple factors that drive performance and that outperformance should be measured once the effect of these factors has been removed. A number of different multi-factor models have been proposed in the literature and these will be discussed in more detail in Chapter 3.

Overall, these models provide little support for active fund managers being able to outperform a passive benchmark of relevant factors by an amount up to the transaction costs and fees associated with active management. This suggests that the ability of fund managers to outperform lies somewhere between the semi-strong form of market efficiency proposed by Roberts (1967) and Grossman and Stiglitz’s (1980) alternative EMH. The performance of fund managers may be questioned further due to the effect of survivorship bias. This bias is associated with the removal of dead funds that no longer exist as a consequence of their underperformance. If these dead funds are not
included in performance evaluation then overall performance of the fund managers would be positively biased\(^2\). Given this evidence regarding the underperformance of the average active fund, the desirability of investment in active fund management must be questioned. One possible reason why investors still choose to invest actively may be based upon their belief that they have the ability to choose a top tier fund manager who actually does outperform benchmark. Given this assumption, investors must use some selection criteria in their choice of fund. One obvious criterion for fund selection would be based upon selecting funds that have done well in the past. However, past research has shown little evidence in general of performance persistence, especially among the top performing active funds\(^3\). This lack of performance persistence suggests there is evidence that active fund manager returns have a large component of luck associated with them. Therefore, a sophisticated investor would want some mechanism for differentiating the lucky from the skilled. Some would argue that the passage of time will disaggregate the lucky from the skilled, however, also over time, the people who drive the investment process of the fund may change and the process itself may have been altered.

All of these issues suggest that an investor should be very cautious in basing their evaluation of a manager purely on past performance. Manager selection should be based on qualitative as well as quantitative factors, and being over enthusiastic about a manager who obtains superior returns relative to benchmark under any classical statistical inference should be discouraged. Nevertheless, there may be some justification for past performance evaluation in manager selection. The question then becomes how performance should be evaluated in a more conservative way that accounts for the possibility of luck in manager performance.

---


The research on appropriate performance measures conducted in this thesis will focus on managers who are mandated to a specific benchmark. In the equity market, this would involve a comparison of performance relative to an index with the same characteristics as the active fund manager. Institutional investors typically mandate active managers to outperform a designated benchmark with clauses restricting the fund managers' levels of allowable risk, which are usually measured in terms of ex-ante standard deviation of excess returns (tracking error) around the designated benchmark. The use of CAPM or multi-factor models under the assumption of a clearly defined benchmark may be inappropriate as the fund managers are not directly compared to the performance of the benchmark which they have been mandated to outperform. Instead managers are assessed on their performance relative to the factors and, even if they outperform benchmark, their alpha from such models may be negative.

Given the increased focus on a fund manager's performance relative to a specific benchmark, Sharpe (1994) proposed another performance measure know as the Sharpe Selection Ratio. This measure estimates the average excess return of the fund over a benchmark per unit of benchmark risk where benchmark risk is defined as the standard deviation of excess returns around the benchmark. Whilst this performance measure is regarded as superior to the previously mentioned measures under the assumption of a clearly defined benchmark, the suitability of the SSR is questionable when applied to non-symmetric distributions of alpha. For example, if the distribution of alpha is positively skewed, then the manager will be penalised for achieving high alphas, since this outcome will invariably increase the standard deviation of alpha. At the same time, the risk of significantly underperforming the benchmark is lower relative to a symmetric or negatively skewed distribution of alpha. Therefore, if risk is viewed as underperformance relative to benchmark, the use of a Downside Performance Metric may be more appropriate. The DPM is measured as the average alpha per unit of downside side risk, where downside risk is the square root of the squared sum of alphas below benchmark divided by the number of observations, or equivalently is the square
root of the second order lower partial moment (LPM) of alpha. The use of the LPM as a measure of risk was initially proposed by Bawa (1967), and a comparison between the DPM and the SSR is conducted in Chapter 3.

Another single factor model that is similar in spirit to the downside performance metric has been proposed in a working paper by Foster and Stutzer (2002). In this work, fund managers are ranked on the basis of decay in probability of underperforming the benchmark through time. A manager whose underperformance probability decays to zero at a faster rate will be ranked higher than a manager with a lower decay rate. Under the conditions of a clearly defined benchmark, this performance measure can also be considered to be appropriate and it is used as an alternative to the DPM.

In this thesis, performance evaluation initially will be carried out using the SSR, the DPM and the Decay Rate measures. A comparison of the ranks from a dataset of US fund managers with a clearly identifiable benchmark is reviewed and the size, power and bias of the SSR and DPM are ascertained.

This work is developed further through the DPM and SSR which are estimated using a Bayesian approach where prior beliefs are placed on the ability of fund managers to outperform. This technique will extend the work carried out by Baks et al. (2001) who also use a Bayesian approach by setting prior beliefs of the skill level of a manager in conjunction with observed returns to derive posterior estimates of alpha using a CAPM based regression methodology. In their study, the authors argue that investment in active fund management would be desirable if at least one manager could achieve a positive posterior alpha under the most sceptical of beliefs on the ability of managers to have skill.

As argued earlier, the use of CAPM or multi-factor approaches is thought to be inappropriate under the condition of a clearly defined benchmark. Therefore the work in this thesis will employ a Bayesian approach to evaluate the DPM and the SSR. Through
careful model selection for mathematical convenience, Baks et al. (2001) are able to provide a closed form solution in order to estimate the posterior alphas. This study is wider in scope where through the use of Markov Chain Monte Carlo simulation techniques - more specifically a Gibbs sampling algorithm - it is possible to calculate posterior estimates of the DPM and SSR as well as other variables of interest in the model that is proposed. The resultant posterior estimates will be based on a combination of the data and prior beliefs. In essence, the greater the evidence associated with the data, the less weight will be placed on the prior beliefs. As such, this can be viewed as a technique for separating the skilled from the lucky. These measures will then be used to rank the managers and assess whether investment in active management is desirable. Subsequently, the results will be compared to the ranks of managers under the classical metrics outlined earlier in this Chapter.
Chapter 3

Classical performance evaluation

3.1 Introduction

This Chapter will provide a review of the extant literature on fund performance as well as examine in more detail which performance measures are appropriate under the assumption of a clearly defined benchmark for the fund manager. As described in the introduction, the available literature on modelling of fund manager performance can be broken down into two types: single-factor and multi-factor models. Each of these models will be described in more detail in this Chapter where the performance measures, conditional on the a priori knowledge that a manager is mandated to a specific benchmark, will be reviewed. Under this assumption, it will be argued some performance measures will be preferred over others and a US dataset will be used to find estimates of a variety of the preferred measures. In doing so, it is hoped that it will be possible to assess the suitability of these instruments for evaluating performance of active fund managers. A simulation study using both symmetric and skewed distributions of alpha will also be used to ascertain the power and bias of each of the preferred performance measures.

The arrangement of this Chapter will be as follows. In Section 3.2 an introduction to the analysis of performance using single-factor models will be discussed. In Section 3.3 multi-factor models will be reviewed and in Section 3.4 the most appropriate models
will be applied using the US dataset. Section 3.5 will describe a simulation technique used to assess the power and bias of each of the relevant performance measures and Section 3.6 will describe the simulation results. Section 3.7 will conclude the Chapter.

3.2 Development of single-factor models

A precursor to the use of single-factor models was to assess fund manager performance by using the funds’ raw returns. Hallahan and Faff (2001) support the ranking of fund managers in this way, as raw returns are generally the most visible performance measure reported by the media and, as such, will provide transparency of rank to investors. Moreover, ranking by raw returns alleviates any issues regarding model misspecification, which is problematic in more complex methodologies when assessing fund manager performance. However, this method provides no adjustment for the risk undertaken in making the investment and, as a result, single-factor risk-adjusted measures have been developed.

In the 1960s, three risk-adjusted single-factor models were proposed to rank fund manager performance. These were developed by Treynor (1965), Sharpe (1966) and Jensen (1968). Each of these measures is based on the ability of the fund manager to beat a benchmark once an adjustment has been made for risk. These will be discussed in turn.

Firstly, the Treynor index ranks fund managers by estimating the average excess return over the risk free rate of return per unit of systematic risk, which is defined for fund $i$ as

$$TI_i = \frac{\bar{R}_i - \bar{R}_f}{\beta_i},$$

where $\bar{R}_i$ is the average return for fund $i$ over the performance period at time $t$, $\bar{R}_f$ is the risk free rate of return over the performance period and $\beta_i$ is the systematic risk of fund $i$ with the market portfolio over the performance period, given as
\[
\beta_{it} = \frac{\text{cov}(R_i, R_m)}{\sigma_m^2}.
\]

Under the theory of CAPM, systematic risk is a standardised measure of how the portfolio co-varies with the market. The Treynor Index assumes that fund managers hold well-diversified portfolios and that the only relevant measure of risk is systematic risk, which is measured through beta. As a result, it is thought to be a relevant measure of risk since the non-systematic component of total risk can be diversified away.

Alternatively, the Sharpe Index, which is the second risk-adjusted single-factor model, provides a measure of rank by estimating the average excess performance of the fund above the risk free rate of return per unit of total risk (reward to variability ratio) and is given by

\[
SI_i = \frac{\bar{R}_i - \bar{R}_f}{\sigma_i}.
\]

Here, \(\sigma_i\) is the standard deviation of fund \(i\) over the performance period and \(\bar{R}_i\) and \(\bar{R}_f\) are defined as before. By adjusting for total risk instead of systematic risk, the Sharpe Index effectively penalises fund managers who hold portfolios that contain unsystematic or idiosyncratic risk.

Both the Treynor and Sharpe measures compare the fund performance to the risk free rate of return. Under the assumption of a clearly defined benchmark, the suitability of these measures is questionable since they do not compare risk and performance directly with the benchmark that the fund manager is required to outperform.

The third measure of risk-adjusted performance is Jensen’s alpha. This performance measure estimates the abnormal return (alpha) above the risk-adjusted return that is expected under the theory of the CAPM. This performance measure implicitly assumes that managers hold well-diversified portfolios, therefore the only measure of risk priced
by the market is systematic risk, measured through beta. Jensen's alpha is estimated through the intercept in a time series regression of excess returns of the fund portfolio above the risk free rate of return, on the excess returns of the market portfolio above the risk free rate of return. This is given by

$$R_{it} - R_{ft} = \alpha_i + \beta_i (R_{mt} - R_{ft}) + \epsilon_{it}$$

where $\alpha_i$ is Jensen's alpha for fund $i$, $R_{it}$ is the return for fund $i$ in period $t$, $R_{ft}$ is the return on a risk free asset in period $t$, $R_{mt}$ is the return on the market portfolio in period $t$, $\beta_i$ is the sensitivity of excess returns on fund $i$ to excess returns on market portfolio and $\epsilon_{it}$ is the random error for fund $i$ in period $t$.

Both Jensen's alpha and the Treynor Index rely on the validity of the CAPM in the estimation of their respective performance metrics. Jensen (1972), Admati and Ross (1985) and Dybvig and Ross (1985) found evidence that Jensen's alpha can result in spurious negative abnormal returns for managers possessing market timing. Furthermore, the CAPM theory requires the market portfolio to be mean-variance efficient. Given the market portfolio is unobservable, Roll (1978, 1979) argued that performance measures based on a market portfolio could be biased due to the inefficiency of the proxy. Due to the questionable validity of CAPM based performance measures and the increased focus on a fund manager's performance relative to a specific benchmark, Sharpe (1994) proposed another performance measure known as the Sharpe Selection Ratio (or Information Ratio). This measure estimates the average excess return of the fund over a benchmark, per unit of benchmark risk, where benchmark risk is defined as the standard deviation of excess returns around the benchmark. The benchmark in this performance metric is aligned with the style of the fund and can be viewed as the opportunity cost of choosing active management above a passive return. The Sharpe Selection Ratio for fund $i$ is defined as

$$SSR_i = \frac{R_{it} - R_{ft}}{\sigma_{R_{it} - R_{ft}}}$$
where $R_i$ is the average return for fund $i$ over the performance period, $R_{bt}$ is the average benchmark return over the performance period and $\sigma_{R_i - R_{bt}}$ is the standard deviation of excess returns for fund $i$ over the performance period.

The Sharpe Selection Ratio directly compares risk-adjusted performance relative to a specific benchmark. The benefit of the measure is therefore dependent on the ability to be able to a priori allocate a benchmark to the relevant fund manager. The circumstances where this measure will be useful include the situation when a manager has been mandated to outperform a specific benchmark or when it is clear to the investor that the opportunity cost of investing in the fund will be the performance relative to where the monies would have been invested. However, the Sharpe Selection Ratio implicitly assumes that investors are indifferent to upside and downside risk. If the distribution of excess returns is symmetric this will make no difference to the performance estimate. However, this measure may lead to an incorrect ranking of a fund manager if the distribution of the alphas is skewed. If excess returns are positively skewed the active manager will be unfairly penalised for obtaining large positive excess returns since the standard deviation of the excess returns will be increased, thus reducing the Sharpe Selection Ratio. Conversely, active managers with negatively skewed distributions of excess returns have a higher probability of obtaining a large negative return relative to the benchmark with limited upside potential and consequently should be penalised. This argument revolves around the definition of risk. Bawa (1967) argues that an appropriate measure of risk should only be concerned with underperformance of the fund manager relative to the benchmark. Therefore, Bawa proposed that the risk adjustment measure should be estimated as the squared sum of the negative excess returns only divided by the number of the observations, that is, a second order lower partial moment approach. Under this definition of risk the ranking measure is defined as the Downside Performance Metric and is given by

$$DPM_i = \frac{\bar{R}_i - \bar{R}_{bt}}{\sqrt{LPM_{xs}}}$$
where $DPM_i$ is the Downside Performance Measure for fund $i$, $LPM_{xs}$ is the square root of the sum of the squared deviations below benchmark for fund $i$ and $\bar{R}_l$ and $\bar{R}_b$ are defined as before.

Both the SSR and the DPM are thought to be good measures of performance under the circumstances of a clearly defined benchmark. The DPM is preferred on theoretical grounds as it measures risk as underperformance relative to the predefined benchmark and compensates those managers who have a positively skewed distribution of alpha with a higher performance measure. Under the circumstances of a symmetric distribution of alphas the two measures will be equivalent in their ranking outcomes.

Another single-factor model similar in spirit to the DPM has been proposed in a working paper by Foster and Stutzer (2002). Fund managers are ranked on the basis of the decay in probability of underperforming the benchmark through time. A manager whose underperformance probability decays to zero at a faster rate will be ranked higher than a manager with a lower decay rate. This measure only ranks funds which have a positive probability of exceeding the benchmark as $T \to \infty$ and therefore only includes funds with past performance above benchmark. The measure estimates the decay in

$$\text{Prob} \left[ \frac{1}{T} \sum_{t=1}^{T} (\log R_t - \log R_b) \leq 0 \right]$$

as $T \to \infty$, and the faster the decay the higher the rank of the fund. The Gartner-Ellis Large Deviations Theorem states this probability will decay to zero at a fund specific exponential rate (for all funds with past positive average alphas measured in continuous time). Under the assumption that the continuous alphas are independently and identically distributed, a consistent estimate of the decay rate for each fund can be calculated according to

$$D = \max_{\lambda > 0} \left( \frac{-\lambda}{T} \sum_{t=1}^{T} e^{-\lambda (\log R_t - \log R_b)} \right).$$
The Foster and Stutzer decay rate is another measure that estimates rank and performance relative to a specific benchmark and therefore is considered to be an appropriate measure for this study.

### 3.3 Development of multi-factor models

Other alternatives to single-factor models stem from the Arbitrage Pricing Theory (APT) initially proposed by Ross (1976) and developed into an equilibrium-based model by Conner (1984). These models imply that returns are driven by multiple factors. However, the identity of the factors remains unknown. Performance evaluation is assessed by the intercept in a time series multiple regression of fund returns on the relevant factors as

\[ R_{it} - R_{ft} = \alpha_i + \sum_{k=1}^{J} \beta_{ik} F_{kt} + \epsilon_{it}, \]

where \( \alpha_i \) is the risk adjusted abnormal performance for fund \( i \), \( R_{it} \) is the return for fund \( i \) in period \( t \), \( R_{ft} \) is the return on a risk free asset in period \( t \), \( \beta_i \) is the sensitivity of excess returns on fund \( i \) to factor \( i \) and \( \epsilon_{it} \) is the random error for fund \( i \) in period \( t \).

A number of models have been discussed in the literature, and these will be reviewed briefly in this Section. Fama and French (1992, 1993) suggest a three-factor model for use in performance evaluation. The first factor is defined as excess market returns above the risk free rate of return and stems directly from the CAPM. The differential between returns on a small capitalisation portfolio with a large capitalisation portfolio and a value portfolio with a growth portfolio make up the other two factors. Fama and French argue that the last two factors are empirically significant in driving asset returns and, as such, should be adjusted for in the estimation of performance. Elton, Gruber, Das and Hlavka (1993) also propose a three-factor model where the factors are the excess return of the market index, a small portfolio and a bond index over the risk free rate of return. They assume that fund managers hold assets outside those held in the market index, including fixed income securities. As the market index usually
encompasses the largest securities, returns on non-index securities are assumed to be well approximated by a small portfolio. The third factor - excess return of a bond index above the risk free rate of return - accounts for any fixed income securities held by the fund manager. This last factor is orthogonalised to highlight the marginal effect of non-index assets above that accounted for by the market index.

Elton, Gruber and Blake (1996) and Gruber (1996) expand the model used by Elton et al. (1993) to account for the performance of value versus growth stocks. Elton et al. (1996, p137) include this additional factor “due to the establishment of a number of mutual funds that state either growth or value as an objective and because the growth and value distinction is highly correlated with book-to-market ratios, which have been shown by Fama and French (1993, 1994) to be empirically important in explaining common stock returns”. The authors also adopt a differential approach to the size and value versus growth factors since this method provides indices that are almost completely uncorrelated. Furthermore, this approach provides easy to understand impacts of the indices on risk-adjusted performance since they are all zero investment portfolios.

Carhart (1997) developed a four-factor model extending Fama and French’s (1993) three-factor model to include a one year momentum anomaly as reported by Jegadeesh and Titman (1993). Carhart finds little evidence of multicollinearity between factors and significant reductions in the size and patterns in pricing errors relative to both the CAPM and the three-factor model. Grinblatt and Titman (1989, 1992) adopt an eight-portfolio benchmark for performance measurement. They argue various firm characteristics are correlated with stock factor loadings. As a result, portfolios formed from stocks grouped by security characteristics can be used as proxies for the factors. Their eight-portfolio benchmark, constructed from groupings of passive portfolio returns, consists of four size-based portfolios, three dividend-yield-based portfolios and the lowest past return portfolio.
While the progress of performance evaluation has evolved to multi-factor models, their suitability under the assumption of a clearly defined benchmark is questionable. The issues regarding multi-factor models will be outlined in the following Section.

3.4 Choice of performance measures

In the previous Sections, it has already been discussed that if a fund manager is mandated to a specified benchmark, the suitability and application of some of the aforementioned models may be questionable. Therefore, the choice methods include the SSR, the DPM, and the Foster and Stutzer Decay Rate, all of which allow for a specific benchmark and define performance and risk relative to that benchmark. The multi-factor models do not explicitly have a benchmark return and measure outperformance relative to outperformance of the factors used in a multiple regression model and not relative to the opportunity cost of investment in active management (that is, the passive benchmark where the money would have been invested if an active manager had not been chosen). Furthermore, these multi-factor models do not allow for the "style" bets chosen by fund managers but measure performance purely at a stock specific level. For example, consider a manager who adopts a value style to outperform a broad index benchmark and who may still achieve a negative alpha through the multi-factor models even though overall performance was above benchmark. This is a consequence of the manager choosing "value" stocks that underperform relative to the value factor even though this underperformance may still be higher than the benchmark performance.

In this study, to test the preferred measures, a sample of US large capitalisation and value managers has been selected. Therefore, it is appropriate to use risk measures that are based on a clearly defined benchmark. However, given that the literature on fund performance has evolved to multi-factor models, the Fama and French three-factor multiple regression model is also used for comparative purposes, although this application is limited in any further analysis that is to be carried out in this research. These four methods provide a broad umbrella of measures that are calculated from the
data alone, and are used to ascertain fund outperformance both in academia and industry under the assumption of a clearly defined benchmark. A hypothesis test using the students t-statistic is also used to ascertain whether a manager’s alpha is significantly above benchmark. Managers with a high variance around the mean will have a lower t-statistic and as a result, this value will also measure outperformance directly relative to the benchmark.

In the following Section the data will be described and in Section 3.6 the empirical results are presented.

3.5 Data

The data used in this study was obtained from the Centre for Research in Security Prices (CRSP), a research centre at the University of Chicago Graduate School of Business. Returns are measured in aggregate, adjusting for changes in net asset value, dividends and capitalisation changes.

Fund manager selection was based on US common stock funds where interest is primarily based on fund managers whose chosen style was explicitly stated to be investment in large capitalisation value securities only. A subset of the database was obtained via a search for funds with the words “value” and “large” in their title and these funds were subsequently used. Strong evidence of a value and large capitalisation strategy was required in the filtering process such that the choice of benchmark would be clear. The benchmark index used for this study is the Wilshire Large Cap Value Index. This is an accumulation index that provides total return of large capitalisation value securities in the U.S. This index was obtained from DataStream and is gross of capital appreciation, dividends and capitalisation changes and is based in US dollars.

While other studies generally adopt the Standard and Poor 500 (S&P500) as the return on the market in determining performance, this study specifically requires a benchmark
that can be clearly identified with the fund manager style. The choice of large value fund managers coupled with a large value benchmark achieves this requirement.

The filtering process outlined above left 287 fund managers with monthly return histories spanning from January 1998 to June 2004. Any manager who had a change of name that did not include the words *value* and *large* was assumed to have changed their style and only the previous returns were retained. Summary statistics for the 287 funds are provided in Table 3.1.

A graphical representation of the data is provided in Figure 3.1. The scatter plot showing the average alpha versus the standard deviation of alpha for each fund indicates that those managers who take on more risk will generally have a higher variability in the average alphas. However, there is little evidence of dependence between average alpha and risk since the correlation coefficient is equal to 0.19. Furthermore, from viewing the plot, there is clear evidence of outliers, in particular three funds with an average alpha of around 0.008 per month and a standard deviation of 0.057 per month and one manager with an average alpha of -0.006 and a standard deviation of 0.119. The three outliers with positive average alphas belong to an incubator of funds called “J Hancock Large Cap Value Fund A/B/C” respectively. These funds do not seem to track the Wilshire Large Cap Value Index well and may need an alternative benchmark specification. The other outlier with the negative alpha and large standard deviation, “Value Trend Large Cap Fund”, is likely to be a large cap manager only as the Value seems to refer to the fund manager name.

A histogram and density of the average alphas for the 287 funds are also shown in Figure 3.1. The distribution appears to be unimodal and has a positive skew with only 34% of managers achieving above benchmark average performance. The boxplot and normal quantile-quantile plot provide evidence of platykurtosis with an excess of average alphas in the tails of the distribution.
<table>
<thead>
<tr>
<th>Statistic</th>
<th>Alpha</th>
<th>Average Fund Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations</td>
<td>9622</td>
<td>287</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.0004</td>
<td>-0.0008</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.0152</td>
<td>0.0030</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.2324</td>
<td>-0.0107</td>
</tr>
<tr>
<td>25% Quantile</td>
<td>-0.0091</td>
<td>-0.0026</td>
</tr>
<tr>
<td>Median</td>
<td>-0.0006</td>
<td>-0.0010</td>
</tr>
<tr>
<td>75% Quantile</td>
<td>0.0078</td>
<td>0.0007</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.3587</td>
<td>0.0087</td>
</tr>
<tr>
<td>Interquartile Range</td>
<td>0.0169</td>
<td>0.0034</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.03</td>
<td>0.26</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>31.52</td>
<td>1.75</td>
</tr>
</tbody>
</table>

Table 3.1: Summary statistics of 287 US value and large capitalisation funds over the period January 1998 through to June 2004. Alpha is the monthly fund return minus benchmark.

Figure 3.2 provides a comparison of cumulative performance between the Wilshire Large Capitalisation Value Index (benchmark) and the average fund over the period January 1998 through to June 2004. Initial performance is initially scaled to one, and the evidence indicates the average fund significantly underperforms benchmark.

The empirical results using this dataset are provided in Section 3.6.

3.6 Empirical results

In this Section, the SSR, DPM, decay rate, Fama and French (1992) regression model and a Student’s t-test are used to evaluate and rank the 287 active funds encompassing the large capitalisation value style within the US market. A comparison between the ranks of the different methodologies for the top 50 large capitalisation value managers within the US market (ranked by the Downside Performance Metric in the first column) is provided in Table 3.2. Observation of Table 3.2 suggests that there are both similarities and differences in rank under the various measures. The most obvious difference appears to be the large discrepancy between the ranks assigned by the Fama
Figure 3.1: Distributional properties of the average alphas for 287 large value funds from January 1998 to June 2004. Alpha is measured as fund return minus benchmark (Wilshire Large Cap Value Index).
Figure 3.2: Cumulative absolute performance of the Wilshire Large Capitalisation Value Index against the average performance of the 287 US value and large capitalisation funds over the period January 1998 through to June 2004.
and French regression model and all of the other models. These differences can be attributed to the use of the Wilshire Large Capitalisation Index as the benchmark for the DPM, SSR and Decay rate measures, whilst the Fama and French regression model's performance measure is based upon how the funds performed relative to three factors used in the model, namely, the excess market return, a value/growth and a large/small capitalisation factor. As suggested previously, under the assumption of a clearly defined benchmark, the use of these multi-factor models may be subject to some doubt.

It is of interest to note that only the decay rate measure agreed with the DPM in ranking Manager 133 as the number one manager. If the top 10 managers ranked by the decay rate are compared then it can be seen that 9 out of the 10 appear in the top 10 of the DPM measure, which is clearly more than any of the other measures. However, whilst this observation is true, there are still significant differences in rank with the second and fourth best funds under the DPM being ranked only as the seventh and eighth best when using the decay rate. Note that none of the performance measures ranked the same manager as the best.

The SSR was also similar to the DPM with the top manager ranked third when using the DPM and ranked as second best manager under the DPM. However, 8 of the top 10 appear in the top 10 of the DPM and overall the rankings tend to be very similar to the decay rate.

The Student’s t-test also had a quite similar top 10 relative to the DPM with 7 out of 10 managers appearing in the top 10 of the DPM. However, there is a large discrepancy in the ranking of the top manager (Manager 276) when using the t-statistic where the same manager is only ranked eighteenth by the DPM. This manager also scored well with the SSR and the decay rate measure. To ascertain why this manager did well under some measures and not others it was useful to look at the distribution of returns for this manager.
<table>
<thead>
<tr>
<th>Manager</th>
<th>DPM</th>
<th>SSR</th>
<th>t stat</th>
<th>Decay</th>
<th>F-F Reg (bps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>133</td>
<td>0.632</td>
<td>0.452</td>
<td>3.555</td>
<td>0.052</td>
<td>0.094 (63)</td>
</tr>
<tr>
<td>72</td>
<td>0.586</td>
<td>0.350</td>
<td>3.236</td>
<td>0.031</td>
<td>0.149 (52)</td>
</tr>
<tr>
<td>208</td>
<td>0.539</td>
<td>0.483</td>
<td>3.979</td>
<td>0.050</td>
<td>0.514 (9)</td>
</tr>
<tr>
<td>11</td>
<td>0.486</td>
<td>0.365</td>
<td>3.516</td>
<td>0.030</td>
<td>0.321 (29)</td>
</tr>
<tr>
<td>233</td>
<td>0.479</td>
<td>0.338</td>
<td>2.679</td>
<td>0.028</td>
<td>0.581 (6)</td>
</tr>
<tr>
<td>210</td>
<td>0.456</td>
<td>0.396</td>
<td>3.257</td>
<td>0.034</td>
<td>0.387 (24)</td>
</tr>
<tr>
<td>209</td>
<td>0.450</td>
<td>0.395</td>
<td>3.231</td>
<td>0.033</td>
<td>0.385 (25)</td>
</tr>
<tr>
<td>127</td>
<td>0.443</td>
<td>0.307</td>
<td>2.389</td>
<td>0.022</td>
<td>0.398 (22)</td>
</tr>
<tr>
<td>92</td>
<td>0.413</td>
<td>0.419</td>
<td>2.918</td>
<td>0.036</td>
<td>0.464 (12)</td>
</tr>
<tr>
<td>134</td>
<td>0.381</td>
<td>0.321</td>
<td>2.418</td>
<td>0.023</td>
<td>0.247 (33)</td>
</tr>
<tr>
<td>241</td>
<td>0.359</td>
<td>0.206</td>
<td>2.569</td>
<td>0.012</td>
<td>0.476 (11)</td>
</tr>
<tr>
<td>145</td>
<td>0.358</td>
<td>0.340</td>
<td>3.016</td>
<td>0.024</td>
<td>0.155 (48)</td>
</tr>
<tr>
<td>207</td>
<td>0.336</td>
<td>0.133</td>
<td>0.727</td>
<td>0.004</td>
<td>0.940 (1)</td>
</tr>
<tr>
<td>277</td>
<td>0.322</td>
<td>0.207</td>
<td>1.846</td>
<td>0.011</td>
<td>0.629 (3)</td>
</tr>
<tr>
<td>136</td>
<td>0.325</td>
<td>0.275</td>
<td>2.455</td>
<td>0.016</td>
<td>0.183 (42)</td>
</tr>
<tr>
<td>117</td>
<td>0.319</td>
<td>0.324</td>
<td>2.698</td>
<td>0.022</td>
<td>0.010 (89)</td>
</tr>
<tr>
<td>119</td>
<td>0.318</td>
<td>0.238</td>
<td>2.320</td>
<td>0.013</td>
<td>0.144 (53)</td>
</tr>
<tr>
<td>276</td>
<td>0.314</td>
<td>0.409</td>
<td>5.680</td>
<td>0.040</td>
<td>0.450 (17)</td>
</tr>
<tr>
<td>147</td>
<td>0.302</td>
<td>0.289</td>
<td>2.552</td>
<td>0.018</td>
<td>0.090 (66)</td>
</tr>
<tr>
<td>146</td>
<td>0.301</td>
<td>0.289</td>
<td>2.549</td>
<td>0.018</td>
<td>0.090 (65)</td>
</tr>
<tr>
<td>282</td>
<td>0.279</td>
<td>0.234</td>
<td>2.897</td>
<td>0.014</td>
<td>0.411 (20)</td>
</tr>
<tr>
<td>120</td>
<td>0.272</td>
<td>0.216</td>
<td>1.772</td>
<td>0.010</td>
<td>0.275 (31)</td>
</tr>
<tr>
<td>54</td>
<td>0.272</td>
<td>0.259</td>
<td>1.664</td>
<td>0.014</td>
<td>0.454 (16)</td>
</tr>
<tr>
<td>271</td>
<td>0.272</td>
<td>0.212</td>
<td>2.182</td>
<td>0.010</td>
<td>0.461 (13)</td>
</tr>
<tr>
<td>238</td>
<td>0.258</td>
<td>0.166</td>
<td>1.971</td>
<td>0.007</td>
<td>0.179 (39)</td>
</tr>
<tr>
<td>20</td>
<td>0.251</td>
<td>0.217</td>
<td>2.239</td>
<td>0.011</td>
<td>0.223 (36)</td>
</tr>
<tr>
<td>13</td>
<td>0.243</td>
<td>0.191</td>
<td>1.294</td>
<td>0.006</td>
<td>0.097 (70)</td>
</tr>
<tr>
<td>121</td>
<td>0.227</td>
<td>0.182</td>
<td>1.494</td>
<td>0.007</td>
<td>0.197 (39)</td>
</tr>
<tr>
<td>123</td>
<td>0.216</td>
<td>0.180</td>
<td>1.481</td>
<td>0.007</td>
<td>0.223 (36)</td>
</tr>
<tr>
<td>15</td>
<td>0.206</td>
<td>0.179</td>
<td>1.367</td>
<td>0.007</td>
<td>-0.058 (110)</td>
</tr>
<tr>
<td>124</td>
<td>0.206</td>
<td>0.171</td>
<td>1.160</td>
<td>0.006</td>
<td>0.054 (80)</td>
</tr>
<tr>
<td>125</td>
<td>0.190</td>
<td>0.171</td>
<td>1.348</td>
<td>0.006</td>
<td>0.216 (37)</td>
</tr>
<tr>
<td>128</td>
<td>0.187</td>
<td>0.139</td>
<td>1.053</td>
<td>0.004</td>
<td>0.537 (7)</td>
</tr>
<tr>
<td>41</td>
<td>0.184</td>
<td>0.150</td>
<td>0.629</td>
<td>0.005</td>
<td>0.672 (2)</td>
</tr>
<tr>
<td>1</td>
<td>0.184</td>
<td>0.153</td>
<td>1.040</td>
<td>0.005</td>
<td>-0.003 (93)</td>
</tr>
<tr>
<td>122</td>
<td>0.180</td>
<td>0.140</td>
<td>0.585</td>
<td>0.004</td>
<td>0.611 (4)</td>
</tr>
<tr>
<td>284</td>
<td>0.175</td>
<td>0.140</td>
<td>0.584</td>
<td>0.004</td>
<td>0.609 (5)</td>
</tr>
<tr>
<td>57</td>
<td>0.171</td>
<td>0.127</td>
<td>0.968</td>
<td>0.003</td>
<td>0.516 (8)</td>
</tr>
<tr>
<td>33</td>
<td>0.161</td>
<td>0.142</td>
<td>1.174</td>
<td>0.004</td>
<td>0.236 (35)</td>
</tr>
<tr>
<td>115</td>
<td>0.160</td>
<td>0.134</td>
<td>0.902</td>
<td>0.004</td>
<td>-0.025 (101)</td>
</tr>
<tr>
<td>27</td>
<td>0.155</td>
<td>0.175</td>
<td>1.251</td>
<td>0.006</td>
<td>-0.005 (95)</td>
</tr>
<tr>
<td>135</td>
<td>0.150</td>
<td>0.155</td>
<td>0.824</td>
<td>0.006</td>
<td>-0.892 (287)</td>
</tr>
<tr>
<td>67</td>
<td>0.142</td>
<td>0.124</td>
<td>1.018</td>
<td>0.002</td>
<td>0.212 (38)</td>
</tr>
<tr>
<td>87</td>
<td>0.140</td>
<td>0.112</td>
<td>1.035</td>
<td>0.002</td>
<td>-0.160 (147)</td>
</tr>
<tr>
<td>239</td>
<td>0.139</td>
<td>0.091</td>
<td>1.155</td>
<td>0.001</td>
<td>0.410 (21)</td>
</tr>
<tr>
<td>86</td>
<td>0.137</td>
<td>0.111</td>
<td>1.018</td>
<td>0.001</td>
<td>0.446 (18)</td>
</tr>
<tr>
<td>240</td>
<td>0.125</td>
<td>0.083</td>
<td>1.015</td>
<td>0.001</td>
<td>-0.517 (277)</td>
</tr>
<tr>
<td>85</td>
<td>0.122</td>
<td>0.088</td>
<td>0.666</td>
<td>0.003</td>
<td>0.019 (57)</td>
</tr>
<tr>
<td>259</td>
<td>0.111</td>
<td>0.125</td>
<td>0.967</td>
<td>0.003</td>
<td>0.277 (277)</td>
</tr>
<tr>
<td>281</td>
<td>0.109</td>
<td>0.088</td>
<td>1.094</td>
<td>0.002</td>
<td>0.317 (30)</td>
</tr>
</tbody>
</table>

Table 3.2: Top 50 US value and large capitalisation funds ranked by DPM over the period January 1998 through to June 2004. Relative rankings are also provided using the SSR, t-statistic, decay rate and Fama and French three-factor regression performance measures.
It was found that this particular manager had only six months of return data with one negative return and five positive returns. Therefore, with a sample this small it would be extremely dangerous to make inferences regarding the likelihood that a manager is skilled or unskilled and consequently the risk measure would need to be adjusted to account for this and treated with scepticism.

Table 3.3 shows the Pearson correlation coefficients between the various performance measures for the top 50 funds that were shown in Table 3.2. Correlation measures the strength of the linear relationship between pairs of variables and is bounded between -1 and +1 where -1 implies a perfect negative linear relationship and +1 implies a perfect positive linear relationship. The choice of technique in estimating the correlation between two variables is dependent on whether the data is quantifiable and meaningful or alternatively is measured as a rank. The Pearson's correlation coefficient requires both variables to be measured on an interval or ratio scale and the calculations are based on the actual values as shown below

$$
Cor(x, y) = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}.
$$

Alternatively, the Spearman's rank coefficient requires data that are at least ordinal and the calculation, which is the same as for Pearson correlation is given by

$$
Cor(x, y) = \frac{\sum [rank_x - (n + 1)/2] [rank_y - (n + 1)/2]}{n(n-1)(n+1)/12}.
$$

This is carried out on the ranks of the data and is used to measure how tightly the ranked data cluster around a straight line. It is also bounded between -1 and +1.
Given the data in Table 3.2 provide both a quantifiable performance measure as well as a rank for each manager, either of the two correlation measures could have been adopted. However, as more information is observed in the performance figure relative to the rank, the Pearson's correlation coefficient was adopted.

A perusal of Table 3.3 suggests the SSR and Decay Rate performance measures are the most highly correlated with a correlation coefficient of 0.9765. These measures are also highly correlated with the t-statistic, indicating that these three methods are very similar in identifying rank for the fund managers in this study. The correlation coefficients of the DPM rank with the SSR. The Decay rate and t-statistic are also high but not to the same degree, indicating differences in rank between the DPM and the other three measures. The final performance measure, the Fama and French regression method, has a low correlation with all other measures. This can be attributed to the non-specification of a direct benchmark for comparison of the active fund returns. Both the ranking of the top 50 funds and the correlation coefficients between the ranking methods suggest that multi-factor regression models will provide significantly different results compared to those performance measures that are based directly on a relevant benchmark. While the multi-regression models may be appropriate when a benchmark cannot be identified, in this study, where a benchmark can be clearly identified, the use of such models may be inappropriate.

3.7 Size, power and bias

To ascertain the power and bias of each of the performance measures, a series of simulation studies was conducted using two simulated funds for comparison. A random sample of observations was drawn from a variety of both symmetrical and skewed distributions with known parameters and the performance measures were estimated.
<table>
<thead>
<tr>
<th></th>
<th>DPM</th>
<th>SSR</th>
<th>t stat</th>
<th>Decay</th>
<th>F.F Reg</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPM</td>
<td>1</td>
<td>0.8989</td>
<td>0.7842</td>
<td>0.8994</td>
<td>0.2496</td>
</tr>
<tr>
<td>SSR</td>
<td>1</td>
<td>0.9022</td>
<td>0.9765</td>
<td>0.2655</td>
<td></td>
</tr>
<tr>
<td>t stat</td>
<td>1</td>
<td>0.8955</td>
<td>0.1518</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decay</td>
<td>1</td>
<td></td>
<td>0.1442</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F.F Reg</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3.3: Correlations of performance measures on US value and large capitalisation funds over the period January 1998 through to June 2004.

The first set of simulations assumes that the returns for both funds are drawn from a normal distribution. The returns for Fund A are drawn from a normal distribution with mean alpha equal to 0.1% per month and standard deviation equal to 1.5% per month. The alphas for Fund B are also drawn from a normal distribution with standard deviation equal to 1.5%.

However, the mean is allowed to vary from 0.1% through to 0.5% in increments of 0.05%. The standard deviation used throughout this performance study is representative of the average standard deviation of funds within the CRSP dataset. The number of monthly alphas for each manager, $n$, is set at 15, 50 and 100 and each of the aforementioned performance measures was calculated. This process was repeated 1000 times and the percentage of times that Fund B was ranked higher than Fund A was calculated for each $n$ in order to enable a comparative measure between the ranks of the two simulated funds.

Given that both funds have the same risk profile, one would expect Fund B to have a higher percentage rank than Fund A for all performance measures when the mean of Fund B is higher than Fund A. Furthermore, the power of each performance measure can be ascertained by the proportion of times that Fund B is ranked above Fund A when the alphas of Fund B are drawn from a distribution with a higher mean than Fund A. The power is measured relative to both the changing number of sample values from each manager, $n$, as well as the variation in the mean of Fund B.
The second set of simulations assumes that the alphas for both Fund A and Fund B are drawn from positively skewed distributions. The required skewed distribution is created using a mixture of a normal and a lognormal distribution where the rationale behind this choice is as follows. A lognormal distribution can be created with positive skew properties (given that the appropriate parameters are chosen). However, it is bounded at zero. Even if this distribution were relocated so that negative alphas were made possible, there would still be a lower bound on the possible values that alpha could take. In comparison, a normal distribution allows for any value of alpha on the real number line. Consequently, using a mixture of the lognormal and the normal distribution provides a distribution with a positive skew as well as allowing the minimum possible value of alpha to be non-bounded. The details regarding model and parameter selection are shown in Appendix A. The mean and standard deviation of Fund A are 0.1% and 1.5% respectively, while the standard deviation of Fund B remains fixed at 1.5%, however, the mean is allowed to vary from 0.1% to 0.5% in increments of 0.05%. The sample size of the managers, \( n \), ranged from 2 to 100 and the power of each of the performance measures is again measured as the proportion of times that Fund B is ranked above Fund A over 1000 iterations. A plot of the densities of the normal distribution and mixture distribution is provided in Figure 3.3, both with mean equal to 1.5% and standard deviation equal to 1.5%.

The last set of simulations assumes that the monthly alphas for Fund A are drawn from a normal distribution with a mean of 0.1% and standard deviation of 1.5%; for Fund B they are drawn from a positively skewed distribution where the mean alpha of Fund B is allowed to vary from 0.1% to 0.5% in increments of 0.05% whilst the standard deviation remains fixed at 1.5%. As argued earlier, performance measures such as the SSR and \( t \)-statistic are expected to provide biased results if it is assumed that risk should be treated as performance under benchmark since managers with a positive skew on the alphas are being penalised for achieving high excess returns.
Finally, the bias of the SSR and DPM are assessed relative to their true values over different choices of \( n \). Evidence from previous literature suggests that sample ratios exhibit bias relative to their true values especially when the sample size is small. Consequently, it was deemed prudent to test the level of bias of the SSR and DPM and thus assess their suitability as a superior performance measure.

### 3.8 Simulation results

This Section describes the results that were obtained from each of the simulation studies and considers two aspects: an analysis of the relative bias and of the power of the SSR, DPM, decay rate and the t-statistic as performance measures for each combination of distributions.
3.8.1 Power

Table 3.3 through to Table 3.5 provide the results from the series of simulations studies described in Section 3.7. For comparative purposes $n$ is selected to represent a small, medium and large sample size. As such, values of $n = 15, 50$ and 100 are chosen. The proportion of times that Fund B is ranked above Fund A will be identical for the SSR and t-statistic given that $n$ is constant at each set of iterations. The power of all measures appears to be very similar under these conditions.

Table 3.4 reviews the power of each performance measure when the alphas of both Fund A and Fund B are drawn from a normal distribution. As expected, as the mean alpha of Fund B is increased relative to Fund A and as the sample size becomes large, all performance measures have increasing power in their ability to select the optimal fund. All measures are performing equally well under these circumstances and there appears to be little difference in the power of the four performance measures in choosing the best manager.

Figure 3.4 shows the proportion of times that Fund A outranked Fund B when both funds have the same distribution with equal parameters values. In this case, where both funds have the same mean and variance, it is expected that the percentage of time that Fund A is ranked above Fund B will be 50%, which is shown to be consistent with the trajectory plots in Figure 3.2.

Figure 3.5 shows the trajectory plot of a second set of simulations that is drawn from normal distributions where the mean of Fund B is greater than that of Fund A (with a difference of 0.001). In this situation it is expected that each of the measures will rank Fund B above Fund A as $n \to \infty$. However, the power of each performance measure can be ascertained by the rate of increase in the likelihood of choosing Fund B over Fund A as
Table 3.4: Percentage of times Fund B outranks Fund A where Fund A’s alphas are drawn from a N(0.001, 0.015) and Fund B from a normal distribution with standard deviation equal to 0.015 and mean varying as shown in the table.

\[ n \to \infty \]. Figure 3.3 shows that all four performance measures have similar power in the ability to choose the superior fund, with no measure standing out above the rest.

Table 3.5 reviews the power of each performance measure when the alphas of both Fund A and Fund B are drawn from a positively skewed mixture distribution with a standard deviation of 1.5% per month as the mean increases from 0.1% to 0.5% for \( n = 15, 50 \) and
Figure 3.4: Proportion of times that Fund A is ranked above Fund B where both funds alphas are drawn from a $N(0.001,0.015)$. 
Figure 3.5: Proportion of times that Fund A is ranked above Fund B. Fund A is drawn from a $N(0.001, 0.015)$ and Fund B from a $N(0.002, 0.015)$. 
100 respectively, as described before. Again it is evident there is little difference in the power of the performance measures in choosing the optimal fund.

Table 3.6 reviews the power of each performance measure when the alphas of Fund A are drawn from a normal distribution with a mean and standard deviation of 0.1% and 1.5% per month, respectively, while the alphas for Fund B are drawn from a positively skewed mixture distribution with the same standard deviation. The results are once again shown for \( n = 15, 50 \) and 100 respectively. Under this set of distributional conditions, the DPM provides a more robust measure of outperformance of Fund A relative to Fund B. The power of the DPM is significantly higher than the other performance measures in its ability to choose the optimal fund over finite \( n \).

Figure 3.6 provides a graphical representation of the performance of the DPM relative to the other performance measures when the distribution of Fund B is positively skewed. The alphas for Fund A are drawn from a \( \text{N}(0.001, 0.015) \) and for Fund B, the alphas are drawn from a positively skewed mixture distribution with the same mean and standard deviation as Fund A, as described in Section 3.5. Given the positively skewed distribution has a lower probability of very large losses and a higher probability of very large gains, it is argued that Fund B should be ranked above Fund A over finite \( n \). Figure 3.4 shows that the SSR, t-statistic and decay rate are indifferent to skewness in the distribution of alphas and rank the funds on average equally with the proportion of times that Fund B exceeds Fund A equal to 0.5 as \( n \to \infty \). However, the DPM has a preference to Fund B with the proportion of times Fund A ranking above Fund B falling as \( n \to \infty \).

Given that the DPM provides a higher power in its ability to chose the optimal fund when the distribution of alphas is skewed and similar power when the distribution of alpha is symmetric, overall the DPM is regarded as a better performance measure in its ability to choose the optimal fund. This preference to the DPM is based on the assumption that the appropriate measure of risk is associated with performance under benchmark.
### Table 3.5: Percentage of times Fund B outranks Fund A where Fund A’s alphas are drawn from a positively skewed mixture distribution with mean 0.001 and standard deviation 0.015 and Fund B from a similar mixture distribution with the same standard deviation, however, with varying means as shown in the table.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Mean of Fund B</th>
<th>SSR</th>
<th>t-statistic</th>
<th>DPM</th>
<th>Decay</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 15</td>
<td>0.0010</td>
<td>0.510</td>
<td>0.510</td>
<td>0.508</td>
<td>0.504</td>
</tr>
<tr>
<td></td>
<td>0.0015</td>
<td>0.546</td>
<td>0.546</td>
<td>0.55</td>
<td>0.548</td>
</tr>
<tr>
<td></td>
<td>0.0020</td>
<td>0.596</td>
<td>0.596</td>
<td>0.598</td>
<td>0.606</td>
</tr>
<tr>
<td></td>
<td>0.0025</td>
<td>0.639</td>
<td>0.639</td>
<td>0.639</td>
<td>0.640</td>
</tr>
<tr>
<td></td>
<td>0.0030</td>
<td>0.674</td>
<td>0.674</td>
<td>0.670</td>
<td>0.675</td>
</tr>
<tr>
<td></td>
<td>0.0035</td>
<td>0.698</td>
<td>0.698</td>
<td>0.698</td>
<td>0.696</td>
</tr>
<tr>
<td></td>
<td>0.0040</td>
<td>0.708</td>
<td>0.708</td>
<td>0.693</td>
<td>0.709</td>
</tr>
<tr>
<td></td>
<td>0.0045</td>
<td>0.784</td>
<td>0.784</td>
<td>0.781</td>
<td>0.789</td>
</tr>
<tr>
<td></td>
<td>0.0050</td>
<td>0.802</td>
<td>0.802</td>
<td>0.803</td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>0.0010</td>
<td>0.497</td>
<td>0.497</td>
<td>0.498</td>
<td>0.499</td>
</tr>
<tr>
<td></td>
<td>0.0015</td>
<td>0.554</td>
<td>0.554</td>
<td>0.551</td>
<td>0.551</td>
</tr>
<tr>
<td></td>
<td>0.0020</td>
<td>0.675</td>
<td>0.675</td>
<td>0.68</td>
<td>0.679</td>
</tr>
<tr>
<td></td>
<td>0.0025</td>
<td>0.732</td>
<td>0.732</td>
<td>0.721</td>
<td>0.734</td>
</tr>
<tr>
<td></td>
<td>0.0030</td>
<td>0.775</td>
<td>0.775</td>
<td>0.776</td>
<td>0.775</td>
</tr>
<tr>
<td></td>
<td>0.0035</td>
<td>0.817</td>
<td>0.817</td>
<td>0.811</td>
<td>0.818</td>
</tr>
<tr>
<td></td>
<td>0.0040</td>
<td>0.872</td>
<td>0.872</td>
<td>0.888</td>
<td>0.868</td>
</tr>
<tr>
<td></td>
<td>0.0045</td>
<td>0.907</td>
<td>0.907</td>
<td>0.905</td>
<td>0.910</td>
</tr>
<tr>
<td></td>
<td>0.0050</td>
<td>0.945</td>
<td>0.945</td>
<td>0.936</td>
<td>0.943</td>
</tr>
<tr>
<td>N = 100</td>
<td>0.0010</td>
<td>0.498</td>
<td>0.498</td>
<td>0.501</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td>0.0015</td>
<td>0.598</td>
<td>0.598</td>
<td>0.596</td>
<td>0.601</td>
</tr>
<tr>
<td></td>
<td>0.0020</td>
<td>0.705</td>
<td>0.705</td>
<td>0.695</td>
<td>0.703</td>
</tr>
<tr>
<td></td>
<td>0.0025</td>
<td>0.749</td>
<td>0.749</td>
<td>0.751</td>
<td>0.750</td>
</tr>
<tr>
<td></td>
<td>0.0030</td>
<td>0.851</td>
<td>0.851</td>
<td>0.842</td>
<td>0.849</td>
</tr>
<tr>
<td></td>
<td>0.0035</td>
<td>0.909</td>
<td>0.909</td>
<td>0.904</td>
<td>0.907</td>
</tr>
<tr>
<td></td>
<td>0.0040</td>
<td>0.953</td>
<td>0.953</td>
<td>0.948</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>0.0045</td>
<td>0.967</td>
<td>0.967</td>
<td>0.963</td>
<td>0.966</td>
</tr>
<tr>
<td></td>
<td>0.0050</td>
<td>0.980</td>
<td>0.980</td>
<td>0.975</td>
<td>0.980</td>
</tr>
</tbody>
</table>

3.8.2 Bias

Figures 3.7 and 3.8 assess the bias of the SSR and DPM where the alphas of the funds are drawn from normal distributions and positively skewed distributions respectively. The t-statistic was omitted from this study as the relative bias of the t-statistic will be the same as: the SSR for any given $n$. The bias associated with the respective ratios can be obtained from the two expectations.
<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Mean of Fund B</th>
<th>SSR</th>
<th>t-statistic</th>
<th>DPM</th>
<th>Decay</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 15</td>
<td>0.0010</td>
<td>0.469</td>
<td>0.469</td>
<td>0.490</td>
<td>0.485</td>
</tr>
<tr>
<td></td>
<td>0.0015</td>
<td>0.485</td>
<td>0.485</td>
<td>0.507</td>
<td>0.503</td>
</tr>
<tr>
<td></td>
<td>0.0020</td>
<td>0.565</td>
<td>0.565</td>
<td>0.582</td>
<td>0.567</td>
</tr>
<tr>
<td></td>
<td>0.0025</td>
<td>0.571</td>
<td>0.571</td>
<td>0.601</td>
<td>0.582</td>
</tr>
<tr>
<td></td>
<td>0.0030</td>
<td>0.631</td>
<td>0.631</td>
<td>0.664</td>
<td>0.646</td>
</tr>
<tr>
<td></td>
<td>0.0035</td>
<td>0.664</td>
<td>0.664</td>
<td>0.69</td>
<td>0.679</td>
</tr>
<tr>
<td></td>
<td>0.0040</td>
<td>0.689</td>
<td>0.689</td>
<td>0.714</td>
<td>0.706</td>
</tr>
<tr>
<td></td>
<td>0.0045</td>
<td>0.727</td>
<td>0.727</td>
<td>0.77</td>
<td>0.737</td>
</tr>
<tr>
<td></td>
<td>0.0050</td>
<td>0.770</td>
<td>0.770</td>
<td>0.804</td>
<td>0.786</td>
</tr>
<tr>
<td>N = 50</td>
<td>0.0010</td>
<td>0.500</td>
<td>0.500</td>
<td>0.527</td>
<td>0.507</td>
</tr>
<tr>
<td></td>
<td>0.0015</td>
<td>0.557</td>
<td>0.557</td>
<td>0.597</td>
<td>0.570</td>
</tr>
<tr>
<td></td>
<td>0.0020</td>
<td>0.613</td>
<td>0.613</td>
<td>0.668</td>
<td>0.630</td>
</tr>
<tr>
<td></td>
<td>0.0025</td>
<td>0.674</td>
<td>0.674</td>
<td>0.733</td>
<td>0.689</td>
</tr>
<tr>
<td></td>
<td>0.0030</td>
<td>0.734</td>
<td>0.734</td>
<td>0.781</td>
<td>0.745</td>
</tr>
<tr>
<td></td>
<td>0.0035</td>
<td>0.797</td>
<td>0.797</td>
<td>0.843</td>
<td>0.810</td>
</tr>
<tr>
<td></td>
<td>0.0040</td>
<td>0.868</td>
<td>0.868</td>
<td>0.902</td>
<td>0.878</td>
</tr>
<tr>
<td></td>
<td>0.0045</td>
<td>0.900</td>
<td>0.900</td>
<td>0.923</td>
<td>0.905</td>
</tr>
<tr>
<td></td>
<td>0.0050</td>
<td>0.912</td>
<td>0.912</td>
<td>0.946</td>
<td>0.919</td>
</tr>
<tr>
<td>N = 100</td>
<td>0.0010</td>
<td>0.501</td>
<td>0.501</td>
<td>0.552</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>0.0015</td>
<td>0.564</td>
<td>0.564</td>
<td>0.618</td>
<td>0.569</td>
</tr>
<tr>
<td></td>
<td>0.0020</td>
<td>0.688</td>
<td>0.688</td>
<td>0.752</td>
<td>0.699</td>
</tr>
<tr>
<td></td>
<td>0.0025</td>
<td>0.772</td>
<td>0.772</td>
<td>0.832</td>
<td>0.788</td>
</tr>
<tr>
<td></td>
<td>0.0030</td>
<td>0.838</td>
<td>0.838</td>
<td>0.899</td>
<td>0.846</td>
</tr>
<tr>
<td></td>
<td>0.0035</td>
<td>0.891</td>
<td>0.891</td>
<td>0.925</td>
<td>0.904</td>
</tr>
<tr>
<td></td>
<td>0.0040</td>
<td>0.933</td>
<td>0.933</td>
<td>0.970</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>0.0045</td>
<td>0.956</td>
<td>0.956</td>
<td>0.978</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>0.0050</td>
<td>0.966</td>
<td>0.966</td>
<td>0.99</td>
<td>0.977</td>
</tr>
</tbody>
</table>

Table 3.6: Percentage of times Fund B outranks Fund A where Fund A's alphas are drawn from a $N(0.001, 0.015)$ and Fund B's alphas are drawn from a positively skewed mixture distribution with mean $X$ and standard deviation 0.015, with the mean $X$ varying as shown in the table.
Figure 3.6: Proportion of times that Fund B is ranked above Fund A. Fund A is drawn from a $N(0.001,0.015)$ and Fund B from a positively skewed mixture distribution with same mean and variance.
Figure 3.7: Percentage bias of SSR and DPM. Alpha simulated from \( N(0.001, 0.015) \).

Figure 3.8: Percentage bias of SSR and DPM. Alpha simulated from a positively skewed distribution with mean equal to 0.001 and standard deviation equal to 0.015.
\[ E(\hat{\alpha}) = E(\hat{\alpha} \frac{1}{\hat{\sigma}_a}) = E(\hat{\alpha}) E\left(\frac{1}{\hat{\sigma}_a}\right) \]

\[ \geq \alpha \frac{1}{E(\hat{\sigma}_a)} = SSR \]

and

\[ E(\hat{\alpha} \frac{1}{L\hat{P}M_a}) = E(\hat{\alpha}) E\left(\frac{1}{L\hat{P}M_a}\right) \]

\[ \geq \alpha \frac{1}{E(L\hat{P}M_a)} = DPM \]

These follow under the assumption of independence and normality and by Jensen's inequality on convex functions. These inequalities are strict as long as the sample estimates are not degenerate at the constant values of the population parameters. Miller and Gehr (1978) and Lo (2002) suggest that bias can also result from the misspecification of unbiased population parameters and from inappropriate time aggregation of performance data. However, these effects will not alter the results under the design of the simulation study carried out in this Chapter.

After observing Figure 3.7, it is evident that as \( n \to \infty \) the bias of the SSR measure decays at a faster rate than that of the DPM when the alphas for each of the funds are drawn from a normal distribution. However, this pattern tends to be reversed when a positively skewed mixture distribution is used, as is seen in Figure 3.8.

Given this evidence on the power and bias of the performance metrics, the DPM is regarded as the most appropriate measure of performance. While other performance measures perform equally, or slightly better than the DPM under the assumption of normality, the DPM is argued to be a much better measure of performance when the distribution of alpha is skewed. In the following Section an analysis is conducted on the
distributional properties of alpha to assess the extent of non-symmetry and therefore the importance of using the DPM as the appropriate measure of performance.

3.9 Empirical distributional properties of alpha

In this Section an empirical investigation is conducted on the distributional properties of alpha. If there is any evidence of non-symmetry in the distributions of alpha, the SSR, t-statistic and decay rate can be argued to be inferior performance measures compared to the DPM.

To conduct this analysis a bootstrap is employed to obtain 95% empirical confidence intervals on the skewness of each manager’s alphas. Any measure of skewness will be unreliable without enough data, therefore, all managers with a return history of less than 30 observations are removed from the sample. This filtering process left 158 out of the 287 managers on which the bootstrap is calculated. For each manager a sample equal to the number of observations is drawn with replacement from the observed data and the coefficient of skewness is then calculated. This process is then repeated 1000 times.

The mean and standard error are then estimated from the 1000 values and a 95% confidence interval is calculated by ordering the 1000 calculated coefficients of skewness and selecting the 2.5\textsuperscript{th} and 97.5\textsuperscript{th} percentiles. Out of the 158 managers, there are 11 that have 95% confidence intervals on the coefficient of skewness that do not include zero. Of these managers, 10 exhibit positively skewed distributions with only one exhibiting a negatively skewed distribution. These results are shown in Table 3.7.

The above results provide strong evidence of positive skewness in a significant proportion of the managers used in this study. Approximately 7% of the managers tested show evidence of non-symmetry. Under the assumption that all managers have symmetric returns it is expected that 5% or approximately eight managers would have
Panel A: Managers with Negatively Skewed Distributions of Alpha

<table>
<thead>
<tr>
<th>Number</th>
<th>95% C.I.</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Manager</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(-1.287, -0.015)</td>
<td>-0.730</td>
<td>0.356</td>
<td>269</td>
</tr>
</tbody>
</table>

Panel B: Managers with Positively Skewed Distributions of Alpha

<table>
<thead>
<tr>
<th>Number</th>
<th>95% C.I.</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Manager</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0.002, 1.058)</td>
<td>0.531</td>
<td>0.290</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>(0.016, 1.156)</td>
<td>0.541</td>
<td>0.318</td>
<td>284</td>
</tr>
<tr>
<td>3</td>
<td>(0.027, 1.423)</td>
<td>0.767</td>
<td>0.393</td>
<td>119</td>
</tr>
<tr>
<td>4</td>
<td>(0.030, 1.401)</td>
<td>0.816</td>
<td>0.379</td>
<td>126</td>
</tr>
<tr>
<td>5</td>
<td>(0.036, 1.563)</td>
<td>0.953</td>
<td>0.406</td>
<td>72</td>
</tr>
<tr>
<td>6</td>
<td>(0.037, 1.101)</td>
<td>0.517</td>
<td>0.303</td>
<td>122</td>
</tr>
<tr>
<td>7</td>
<td>(0.040, 1.153)</td>
<td>0.522</td>
<td>0.310</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>(0.140, 1.121)</td>
<td>0.637</td>
<td>0.284</td>
<td>127</td>
</tr>
<tr>
<td>9</td>
<td>(0.157, 1.037)</td>
<td>0.589</td>
<td>0.254</td>
<td>62</td>
</tr>
<tr>
<td>10</td>
<td>(0.166, 1.021)</td>
<td>0.576</td>
<td>0.242</td>
<td>75</td>
</tr>
</tbody>
</table>

Table 3.7: Empirical confidence intervals on managers exhibiting skewness from a bootstrap of the 158 managers with returns history exceeding thirty observations.

95% confidence intervals that do not include the value zero. Of these managers, four would be expected to exhibit negative skewness and four positive skewness. Given only one manager exhibited negative skewness and ten positive skewness, it is argued there is evidence of positive skewness in the distribution of alpha. Consequently, preference is given to the DPM as the preferred performance measure.

3.10 Discussion

This Chapter investigated a number of techniques that can be used to measure fund manager performance. The t-statistic, SSR, decay rate and DPM were then argued to provide superior performance evaluation measures under the assumption of a clearly defined benchmark.

These four measures were then used to rank funds with a large capitalisation value style within the US equity market. The results suggest some variation in rank, depending on
the choice of performance methodology adopted. Simulations were then used to ascertain the power and bias of each performance measure.

The results suggest little difference in power between measures under the assumption of normal or symmetric return distributions. The DPM, however, achieves higher power if the distribution of one of the funds has a positive skew. Furthermore, while the relative bias of the DPM is slightly higher than the SSR under the assumption of normality, this pattern is reversed when the distributions of alpha are skewed.

Given that the empirical bootstrap confidence intervals indicate strong statistical evidence of asymmetry in fund manager returns and the a priori preference for the DPM as a model that has a more appropriate measure of risk, it is argued that the DPM is the optimal performance measure under the assumption of a clearly defined benchmark.
Chapter 4

Bayesian methodology and model

4.1 Introduction

The motivation throughout this thesis is to provide the most appropriate performance metric in order to measure fund manager performance and rank under the assumption of a clearly defined benchmark. Recall that the DPM is argued in Chapter 3 to be a more appropriate measure of performance if active managers are mandated to a specific benchmark. A Bayesian measure of the DPM allows more flexibility in the resulting model of performance. Within a Bayesian framework it is possible to posit prior beliefs on the probability that active fund managers are skilled as well as on their ability to achieve a superior return relative to their mandated benchmark, whilst taking into account the costs that are involved in active management. These prior beliefs can then be used in conjunction with the observed data to form posterior distributions of the DPM for each manager.

The adopted Bayesian approach is analogous to the work of Baks et al. (2001). Prior beliefs on the skill set of fund managers will be used in conjunction with the observed data to obtain posterior estimates of performance measures. While Baks et al. (2001) obtain a closed form solution using a posterior expected value of alpha from a multi-factor regression as their measure of performance, this study will adopt the DPM and the SSR as the underlying measures to be used in calculating the posterior performance.
metrics. As outlined earlier these measures are believed to be more appropriate under the assumption of a clearly defined benchmark. In the absence of a closed form solution, Markov Chain Monte Carlo methods – more specifically a Gibbs sampling algorithm – are used to estimate the posterior distributions of the DPM and SSR.

Baks et al. (2001) advocate the use of Bayesian methods to assess the desirability of investment in active fund management. Unskilled managers are assumed to underperform benchmark by the fees and transaction costs associated with active management as well as any losses from trading with skilled managers. The distribution of the skilled managers is assumed to be half normal above the performance of the unskilled managers. Baks et al. (2001) use this Bayesian model to assess whether investment in active management may be desirable under the most sceptical of prior beliefs on the probability of active management skill.

This Bayesian methodology can be viewed as a technique for separating the lucky managers from the skilled managers when no specific benchmark can be identified. Managers with high returns, longer return histories and less variation in excess returns provide more evidence of skill and will be less influenced by the prior beliefs defined under the Bayesian model. The skill parameters can be adjusted according to the efficiency of the market the manager invests within.

Under the assumption of a clearly identifiable benchmark, the Baks et al. (2001) use of multi-factor regression models is argued to be inappropriate and therefore in this study a Bayesian DPM and SSR are estimated for each manager in the US dataset. The probability a manager has skill as well as the probability a manager has enough skill to exceed the benchmark are allowed to vary and the posterior estimates of the DPM and SSR are estimated for each manager. This analysis provides an alternative technique for ranking the managers, as well as being a tool to test whether investment in active management is desirable, based upon prior beliefs on market efficiency.
Through the use of MCMC methods this study will also allow the prior beliefs on the skills of active managers to vary. Unlike Baks et al. (2001) who place a point estimate on the priors, the Bayesian model in this study allows the prior probabilities of having skill and enough skill to exceed benchmark to have a probability density distribution. This added complexity allows for a richer set of outcomes, depending on the purpose of the analysis and the confidence relating to the prior beliefs. The initial model described above can be viewed as one extreme of the general model where dogmatic prior beliefs are placed on the prior parameters.

At the other extreme, a priori ignorance can be posited relating to the probability a manager has skill and the probability a manager has enough skill to exceed the benchmark return. These parameters can then be estimated, based on the data within the Bayesian model which is used to obtain a posterior DPM and SSR for each manager. This version of the model also has intuitive appeal as it can be viewed as obtaining the priors relating to skill directly from the observed data of all fund managers within the dataset and then using these priors to estimate the posterior performance measures relating to each specific manager.

Baks et al. (2001) did not investigate the different performance measures explicitly. This study will further develop their work and adopt the DPM and the SSR as the underlying measures to be used in calculating the posterior performance metrics. In Chapter 3 it was argued that the use of the DPM alleviates the problem of benchmark mismatch and inappropriate measures of risk as well as providing more power and less bias relative to the SSR. As the SSR does not discriminate whether the variance is induced by a return above or below benchmark, managers with a positively skewed distribution of their excess returns will be penalised for achieving high returns through a higher variance of their excess returns, which in turn will reduce their overall SSR. Nevertheless, the Bayesian SSR is also estimated for comparative purposes.
Furthermore, Baks et al. (2001) assume the likelihood within the Bayesian model is multivariate normal for mathematical simplicity in obtaining a closed form solution for their measure of risk adjusted performance. In this study such restrictions on the likelihood are not necessary, and any distribution can be assumed for alpha which can then be used within the MCMC simulation algorithm. If there is strong empirical evidence of skewness in alpha, the model can be easily adapted by assuming a skewed distribution to represent the likelihood within the Bayesian model.

The outline of this Chapter is as follows. The Bayesian framework will be described in Section 4.2. In Section 4.3, the statistical techniques that are required to estimate the posterior DPM and SSR given our a priori beliefs, known collectively as Markov Chain Monte Carlo methods, will be introduced. The form of the Bayesian DPM and SSR will be briefly described in Section 4.4 and the Bayesian hierarchical model will be broken down into various subsections describing the likelihood, the prior beliefs, and ultimately the posterior distribution, and will be discussed in more detail in Section 4.5. Section 4.6 concludes the Chapter. A more detailed description of the model is given in Appendix B.

4.2 Bayes’ theorem and the posterior distribution

Let us suppose that $y$ denotes the data and that one conditions on the observed data where the parameters, represented by $\theta$, are taken to be random variables. This value $\theta$ could consist of model parameters, or perhaps missing data. Assume $\{f(\theta) : \theta \in S\}$ to be a prior distribution reflecting the uncertainty or the initial beliefs about $\theta$ before the data $y$ is taken into account. The posterior distribution $f(\theta \mid y)$ reflects the updated beliefs about $\theta$ after observing the sample $y$. Applying Bayes’ Theorem, conditional on the data $y$, the posterior distribution of $\theta$ can be expressed as

$$f(\theta \mid y) = \frac{L(\theta \mid y) f(\theta)}{f(y)} \quad (4.1)$$
where \( L(\theta | y) \) is the usual likelihood function and could also be written \( f(y | \theta) \). Note that when \( \theta \) is a continuous random variable, the denominator in Equation 4.1 is the integral \( f(y) = \int f(\theta)f(y | \theta)d\theta \), and when \( \theta \) is discrete, then \( f(y) = \sum_\theta f(\theta)f(y | \theta) \).

The posterior distribution for our model is likely to be complex due to the mixture model inherent in the set up; as such, classical statistical techniques such as maximum likelihood are generally difficult to implement. We propose to use Markov Chain Monte Carlo methods in our analysis and these will be discussed in more detail in the next Section.

### 4.3 Markov Chain Monte Carlo methods

There are several MCMC methods that have been developed. For the application of MCMC techniques, it is only required that the posterior distribution is known up to a constant of proportionality. Therefore, the integral or sum in the denominator in Equation 4.1 does not need to be evaluated. The most common of the MCMC methods is the Gibbs Sampler developed by Geman & Geman (1984). This will be discussed in more detail in the following subsection.

#### 4.3.1 The Gibbs sampling algorithm

The Gibbs sampling algorithm is a natural method to consider whenever it is necessary to sample from densities that exhibit a complicated form. The Gibbs Sampler involves the process of drawing random samples from all full conditional distributions. These have each been extracted from the full posterior distribution in turn, where the posterior distribution need only be derived up to a function of proportionality. At each iteration, parameters are updated consecutively by sampling a new value from their full conditional distribution.
More explicitly, the joint posterior density is defined to be $f(\theta \mid y)$, as before, where $\theta = (\theta_1, \ldots, \theta_k)$ is a $k \times 1$ vector of unknown parameters, estimates of which are desired. The Gibbs Sampler uses the conditional density of each parameter, given the rest of the parameters and the data, $f(\theta_i \mid y, \theta_{-i})$, where $\theta_{-i}$ is the vector $\theta$ with the $i$th element removed. The algorithm begins by assigning initial values to each free parameter, $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)})$ where $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_k^{(0)})$ represents the values of the Gibbs sampling chain at the $r$th iteration. Thus, the following loop is iterated:

1. Sample $\theta_1^{(i)}$ from $f(\theta_1 \mid y, \theta_{-1}^{(i)})$
2. Sample $\theta_2^{(i)}$ from $f(\theta_2 \mid y, \theta_{-2}^{(i)})$

\[ \vdots \]

k. Sample $\theta_k^{(i)}$ from $f(\theta_k \mid y, \theta_{-k}^{(i)})$

The chain for the parameters is updated by sampling a new value from their full conditional distribution at every step in the loop. This newly sampled value is taken to be the current approximation of the parameter, and is then used in the sampling process for the following conditional distribution. As a result, the conditional distributions change from one iteration to the next until overall convergence has been reached for each of the parameters in the loop. Consequently, through the use of the Gibbs sampler, inference can be made simultaneously regarding each parameter of interest, using the output from one entire Gibbs sampling chain.

Asymptotic results that have been developed enable inference to be made using the output; these results will be discussed in more detail in the following subsection.
4.3.2 Posterior inference and asymptotic results

The success of the Gibbs sampler, as well as all of the other MCMC methods, relies heavily on the ability to assess the convergence of a chain to its equilibrium. Once the length of the chain has been established and convergence has become apparent, then the posterior distribution contains all of the relevant information regarding $\theta$. Under suitable regularity conditions, asymptotic results (Geman & Geman, 1984) exist that validate inference that is to be performed on the Gibbs sampling chain. These results, which are to be applied throughout the course of this thesis, are as follows:

1. As $t \to \infty$ the joint density of $\theta^{(i)}$ converges to $f(\theta \mid y)$. Result (1)

$$\frac{1}{t} \sum_{i=1}^{t} g(\theta^{(i)}) \xrightarrow{s} \int g(\theta) f(\theta \mid y) d\theta \text{ as } t \to \infty.$$ Result (2)

Result (1) indicates that for a sufficiently large value of $t$, $\theta^{(i)}$ can be regarded as a value that has been sampled from the posterior distribution of $\theta$. Therefore through the use of result (2), the sample output can be used to derive numerical summaries such as the posterior mean and posterior standard error; this is done through the calculation of the expected value of a particular function of interest, with respect to the posterior distribution.

As well as using point summaries to report information on the posterior distribution, it is also important to report posterior uncertainty. One possible method is to evaluate quantiles of the posterior distribution for the relevant estimands. Interval summaries can also be derived; two interval estimates will be considered in the subsequent analysis: the $100(1-\alpha)\%$ Central Posterior Density Region and the $100(1-\alpha)\%$ Highest Posterior Density Region.

The $100(1-\alpha)\%$ Central Posterior Density region for $\theta$ corresponds to the smallest range of values above and below which lies at most $100(\alpha/2)\%$ of the posterior probability. This is sometimes referred to as a Bayesian credible interval. It is
important that due care be taken in interpreting the form of the interval from a Bayesian perspective; there is an important difference between the Bayesian and the Frequentist interpretations which must be realised when any inference is drawn. A Bayesian will deduce that the parameter in question lies within the interval with a probability of $100(1-\alpha)\%$ after observing the data. In contrast, the frequentist interpretation of a $100(1-\alpha)\%$ confidence interval implies that before the data has been observed, there is a $100(1-\alpha)\%$ chance that the interval will contain the parameter. Apart from the interpretation of each type of interval, the main difference between the methods is that the Frequentist confidence interval does not allow for conditioning on the data that is actually observed.

### 4.4 The Bayesian DPM and SSR

The DPM described in Chapter 1 is defined as

$$DPM = \frac{\mu}{\sigma^2}, \quad (4.2)$$

where

$$\sigma^2 = \begin{cases} \frac{1}{\alpha} \int_{-\infty}^{B} (\alpha_i - B)^2 p(\alpha_i) \, d\alpha_i & \text{in the continuous case;} \\ \frac{1}{n} \sum_{i=1}^{n} I(\alpha_i < B)(\alpha_i - B)^2 & \text{in the discrete case.} \end{cases}$$

In Equation 4.2, $\mu$ represents the average return of the active manager above their designated benchmark $B$ or average alpha. The lower partial moment, $\sigma^2$, is the sum of the squared deviations of the negative alphas only, divided by the total number of observations, $n$.

Throughout this thesis, the benchmark of the funds is assumed to be a passive portfolio of securities that the active manager has been mandated to exceed. Since alpha is
defined as fund return minus benchmark it is appropriate to chose \( B = 0 \) in the above formula. This is deemed appropriate because a positive alpha implies that fund managers are outperforming their mandated passive benchmarks. Therefore, the lower partial moment becomes

\[
\tilde{\sigma}^2 = \begin{cases} 
\int_{-\infty}^{0} \alpha_i^2 p(\alpha_i) \, d\alpha_i & \text{in the continuous case;} \\
\frac{1}{n} \sum_{i=1}^{n} I(\alpha_i < 0)\alpha_i^2 & \text{in the discrete case.}
\end{cases}
\]

Similarly, the SSR can be expressed as

\[
SSR = \frac{\mu}{\sigma}, \tag{4.3}
\]

where

\[
\sigma^2 = \begin{cases} 
\int_{-\infty}^{\infty} (\alpha_i - B)^2 \, d\alpha_i & \text{in the continuous case;} \\
\frac{1}{n} \sum_{i=1}^{n} (\alpha_i - B)^2 & \text{in the discrete case.}
\end{cases}
\]

Once again, in Equation 4.3, \( \mu \) represents the average return of the active manager above their designated benchmark \( B \), and \( \sigma^2 \) is the sum of the squared deviations of the alphas divided by the total number of observations, \( n \). When \( B \) is equal to zero, \( \sigma^2 \) reduces to

\[
\sigma^2 = \begin{cases} 
\int_{-\infty}^{\infty} \alpha_i^2 p(\alpha_i) \, d\alpha_i & \text{in the continuous case;} \\
\frac{1}{n} \sum_{i=1}^{n} \alpha_i^2 & \text{in the discrete case.}
\end{cases}
\]

Both the numerator and the denominator in Equations 4.2 and 4.3 can be calculated using simulated values from the Gibbs sampling algorithm (the entire algorithmic process will be described in more detail in Chapter 5). At each iteration, a value for the
numerator will be retained and used directly in the calculation of the function in Equations 4.2 and 4.3. However, the derivation of the denominator for the DPM requires more care. This formulation is shown in more detail in Appendix B. For the remainder of this Chapter, the model described will refer to the Bayesian DPM in its exposition. However, the calculation of the estimates of the Bayesian SSR is mathematically similar.

4.5 The Bayesian model

The hierarchical structure of the Bayesian model will now be described more explicitly. This model will be used to calculate the joint posterior density of the parameters of interest for two particular cases which will be described in the next Chapter.

4.5.1 The likelihood or data

To obtain sample values from the distribution of the DPM for each individual manager it is first necessary to assume a return distribution for the alphas of each fund, thus describing the likelihood function which is used to obtain the resultant joint posterior distribution. Therefore, assume that alpha, which represents the arithmetic excess return for the fund manager and is equal to active fund returns minus benchmark returns (i.e. \( \alpha_i = r_i - B_i \)), is normally distributed, such that

\[
f(\alpha_i | \mu, \sigma^2) \sim \text{iid } N(\mu, \sigma^2), \ i = 1, \ldots, n, \tag{4.4}
\]

where \( n \) represents the available number of months of recorded return history for the fund manager. Furthermore, \( r_i \) and \( B_i \) denote the total return (including dividends, cashflow adjustments and capitalisation changes) considered for an individual fund manager and for passive benchmark for month \( i \), respectively. It is assumed the alphas have a mean and variance of \( \mu \) and \( \sigma^2 \) respectively.
The assumption of normality in alpha is based upon the bootstrap conducted in Section 3.9. While some managers exhibit positive skew in their excess returns, overall the evidence suggests the distribution of most managers’ alphas are symmetric. If there is strong empirical evidence of skewness in alpha, the model can be easily adapted by assuming the likelihood is also skewed. Thus, through the use of MCMC techniques the Bayesian model in this study provides more flexibility than Baks et al. (2001) who must assume multivariate normality of excess returns to obtain a closed form solution in estimating their risk-adjusted performance measure.

4.5.2 The prior distributions of $\mu$ and $z$

Making some judgement on whether a manager is regarded as skilled or unskilled allows some form of prior belief to be placed on alpha. Under Fama’s semi-strong form of market efficiency, it is expected that on average active management should underperform benchmark by the fees and transactions costs that are associated with active management. Furthermore, the vast majority of extant literature regarding active fund performance suggests that on average, fund performance is below benchmark up to the costs associated with active management. Therefore, it is initially proposed that

$$E(\mu) = -\text{fees-transaction costs} = \bar{\mu}$$

where the expected value of alpha is $\mu$. While the prior belief may be slightly conservative regarding manager outperformance, this is considered to be appropriate since an investor’s perspective in deciding between active and passive management is being taken. More explicitly, a conservative prior belief will imply a lower likelihood of choosing an active fund manager whose outperformance is based purely on chance.

As described above, active fund managers can be broken down into skilled and unskilled managers. Therefore, let $z$ be a random variable representing two states; a skilled manager or an unskilled manager. More formally:
\[ P(\text{manager is skilled}) = P(z = 1) = q \]
\[ P(\text{manager is unskilled}) = P(z = 0) = 1 - q \]

where \( q \) is the probability a manager is skilled. Thus,

\[ (z \mid q, \sigma^2) \sim \text{Bernoulli}(q). \]

Next it is assumed that the average alpha of unskilled managers, \( \tilde{\mu} \), will be below benchmark by their fees and transaction costs, as well as any loss from trading with skilled managers (\( U \)). Therefore, let

\[ \tilde{\mu} = E(\mu \mid z = 0, \sigma^2) = U - \text{fees-transaction costs}. \]

The average alpha of skilled managers is assumed to follow a half normal distribution above the average alpha of the unskilled managers. While any distribution could have been chosen to represent the average alphas of the skilled, it is reasonable to assume there are less highly skilled managers than moderately or low skilled managers. Consequently, any monotonically decreasing distribution function could have been used to depict the average alphas of the skilled managers. Formally, the prior distribution for \( \mu \) can be represented as:

\[
\begin{align*}
    f(\mu \mid z = 0, \sigma^2, q) &= \delta_{\tilde{\mu}} \\
    f(\mu \mid z = 1, \sigma^2, q) &= N(\tilde{\mu}, \sigma_a^2) I(\mu > \tilde{\mu})
\end{align*}
\]

where \( I(\ ) \) is the standard indicator function and \( \delta_{\tilde{\mu}} \) is the Dirac function that indicates the distribution is a point mass (thus, \( \delta_{\tilde{\mu}} = I(\mu = \tilde{\mu}) \)). Alternatively, this may be written as follows:

\[
\begin{align*}
    f(\mu \mid z = 0, \sigma^2, q) &= 1, \quad \mu = \tilde{\mu} \\
    f(\mu \mid z = 1, \sigma^2, q) &= \frac{2}{\sigma_a \sqrt{2\pi}} e^{-\frac{1}{2\sigma_a^2} (\mu - \tilde{\mu})^2}, \mu > \tilde{\mu}
\end{align*}
\]
A graphical representation of the prior distribution of the average alphas is shown in Figure 4.1. As previously described there is a probability of $1 - q$ that a manager is unskilled and will have an average return equal to $\mu$ and there is a probability of $q$ that a manager is skilled and has an average alpha from a half normal distribution centred at $\mu$.

4.5.3 The remaining prior specification

There are two separate but important cases that will be considered in this thesis with respect to the model described here. The first case is where the probability of alpha averaging above benchmark, $p$, and the probability a manager has skill, $q$, are assumed to be constant and known a priori. Therefore, small values of these probabilities are used to ascertain whether it is possible to obtain a positive posterior DPM for at least one manager under the most stringent conditions regarding the informational efficiency of the market. This is the case where the prior probability of an active manager being skilled is assumed to be constant and known or hypothesised through some deduction of the investigator. This can be rationalised by observing that if a manager can achieve a positive posterior alpha and DPM for extremely low prior beliefs on the likelihood of managers having skill, then this suggests that there is evidence that investment in the active manager might be desirable.

A Bayesian approach performed in this way is appropriate because a manager may outperform based on pure chance. For example, if a classical test is conducted at a 95% confidence level to assess whether a manager is skilled, under the assumption that all managers have an a priori alpha of zero, 1 in 20 managers will have an average alpha, and consequently a positive DPM, that is significantly different from zero. This result does not imply that the manager has skill but rather that a random draw from a distribution with a mean of zero has drawn a set of numbers that are significantly statistically different from zero, and done so by chance. Therefore, even under the
assumption that all of the managers have alphas with a mean of zero, some of them will be perceived to be skilful from chance alone. As a result, the use of a Bayesian constant prior on \( p \) and \( q \) allows the likelihood of outperformance to be revised downwards and the posterior estimates of alpha and the DPM to be adjusted due to the stochastic nature of chance.

An alternative specification for this model and the second case to be considered is to allow the probability of a manager being skilled \((q)\) and the probability that a manager can outperform benchmark \((p)\) to be random variables. This approach can address two different research problems. Firstly, similar to the arguments in the previous paragraph where \( p \) and \( q \) are held constant and the posterior distribution of alpha and the DPM are estimated based on these estimates, \( p \) and \( q \) can be set as random such that there is an expectation on the probability of a manager being skilled as well as exceed benchmark. However, assuming that it is not possible to have perfect insight, it is reasonable to allow the estimates of \( p \) and \( q \) to have some level of variance that can be adjusted.
depending on the certainty regarding the confidence in estimating these parameters. Alternatively, if \( p \) and \( q \) are set to be random and \textit{a priori} ignorance is placed on the probability that \( p \) and \( q \) can take on any value, then the Bayesian methodology can be used to estimate the distributions of \( p \) and \( q \), given all data on fund manager performance for each specific manager. As a result, the second model allows \( p \) and \( q \) to be random and by placing no prior belief (or non-informative prior distributions) on these parameters, the data is primarily used to estimate the values of these parameters.

Note that it is necessary to define the model in terms of the parameter \( r \) where

\[
    r = \frac{p}{q} \quad \text{or} \quad p = rq
\]

since the prior density is required to be equal to 1 and, if \( p \) were to be used in the hierarchical model, this would lead to mathematical problems in the chosen approach. However, most inference will be done with respect to \( p \) rather than \( r \) throughout the remainder of this thesis where \( p \) is the probability that a manager exceeds benchmark, and \( r \) is the proportion of skilled managers that exceed benchmark where \( 0 < r < 1 \).

Therefore, in order to complete the prior specification, the remaining parameters in the hierarchical model can be defined as follows

\[
    f\left(\sigma^2, r, q\right) f\left(\sigma^2 | r, q\right) f\left(r | q\right) f\left(q\right)
\]

Furthermore, for reasons of mathematical convenience and conjugacy, a precision parameter, \( \lambda = 1/\sigma^2 \) will be used rather than \( \sigma^2 \). Thus,

\[
    f\left(\lambda | r, q\right) = f\left(\sigma^2 | r, q\right) \frac{d\sigma^2}{d\lambda}
\]

where
\[ \sigma^2 = \frac{1}{\lambda} \quad \text{and} \quad \frac{d\sigma^2}{d\lambda} = -\frac{1}{\lambda^2} \]

and hence

\[ f(\lambda | r, q) - \lambda \left| -\frac{1}{\lambda^2} \right| = \frac{1}{\lambda}, \quad \lambda > 0. \]

### 4.5.4 The posterior distribution for the general case

Firstly, in order to complete the prior specification, let

\[
(z | \lambda, r, q) \sim \text{Bernoulli}(q)
\]

\[ f(\lambda | r, q) \propto \frac{1}{\lambda}, \quad \lambda > 0 \]

\[ f(r | q) \sim \text{Beta}(g, h) \]

\[ f(q) \sim \text{Beta}(a, b) \]

where \(g, h, a\) and \(b\) are specified constants that reflect prior belief regarding \(r\) and \(q\).

The joint posterior density is then defined as follows

\[
f(\lambda, \mu, z, q, r | \alpha) \propto f(\alpha | \lambda, \mu, z, q, r) f(\mu | \lambda, q, z, r) f(z | \lambda, q, r) f(\lambda | q, r) f(r | q) f(q) \\
\propto \prod_{i=1}^{r} \left\{ \frac{\lambda}{2\pi} \exp \left[ -\frac{\lambda}{2} (\alpha_i - \mu)^2 \right] \right\} \\
\times \left\{ (1-q)I(z=0)I(\mu = \bar{\mu}) + (q)I(z=1)I(\mu > \bar{\mu}) \frac{2}{\sigma_\alpha \sqrt{2\pi}} e^{-\frac{1}{2} (\mu - \bar{\mu})^2} \right\} \\
\times \frac{1}{\lambda} f(r, q). \]

where \(\mu \geq \bar{\mu}; z = 0 \text{ or } 1, \lambda > 0 \text{ and } 0 < p < q < 1.\)

When \(p\) and \(q\) are treated as random variables, \(f(r, q) \propto r^{s-1} (1-r)^{b-1} \times q^{s-1} (1-q)^{b-1}\) and this is substituted into the joint posterior density above. However, when \(p\) and \(q\) are assumed to be constant, \(f(r, q) = 1.\)
Note that the probability of exceeding benchmark \((p)\) must be less than the probability that a manager is skilled \((q)\).

### 4.5.5 The priors on \(p\) and \(q\)

In postulating prior distributions for \(q\) and \(r\), a beta distribution was used so as to allow for some flexibility in the specification of \textit{a priori} beliefs about the level of skill of a manager and the proportion of skilled managers who can obtain a return above benchmark, respectively. To see why beta priors on \(r\) and \(q\) are appropriate, consider a random variable \(X\) from a beta distribution with parameters \(\eta\) and \(\xi\). This is described below, where \(X\) will have the following probability density function

\[
f(x) = \frac{x^{\eta-1}(1-x)^{\xi-1}}{B(\eta,\xi)}
\]

where

\[
B(\eta,\xi) = \frac{\Gamma(\eta)\Gamma(\xi)}{\Gamma(\eta+\xi)}.
\]

In order to assume \textit{a priori} ignorance regarding the probabilities \(x\), \(\eta\) and \(\xi\) are both set to equal one. This implies that the prior probability density function of \(X\) is constant over the range 0 to 1. Consequently, if the parameters of the beta distributions for \(q\) and \(r\) are set equal to 1 respectively, then \textit{a priori} ignorance is being placed on the likely distribution of these parameters and therefore allows the data to drive the model and estimate the distribution of these parameters, \(q\) and \(r\).

Alternatively, the parameters \(\eta\) and \(\xi\) can be altered depending on the confidence of the prior beliefs in the fund manager's likelihood on being skilled, or the proportion of skilled managers that will exceed benchmark. In order to see this more clearly, consider that the expected value and variance for the beta distribution are defined as follows:
\[ E(X) = \frac{\eta}{\eta + \xi} \]
\[ \text{Var}(X) = \frac{\eta_\xi}{(\eta + \xi)^2 (\eta + \xi + 1)} \]

Therefore, for any given expected value for \( X \), the parameters \( \eta \) and \( \xi \) can be manipulated to adjust the variance of \( X \). Thus the parameters of the beta distributions for \( r \) and \( q \) can be varied to express any confidence in the possible values of these parameters. For any given expected value of \( r \) and \( q \), higher values of \( \eta \) and \( \xi \) will provide a smaller variance or uncertainty around the expectation.

### 4.5.6 Calculation of \( \tilde{\mu} \) and \( \sigma_a \)

In order to proceed, the remaining parameters, \( \tilde{\mu} \) and \( \sigma_a \) that have been introduced into the prior distribution specification and are unknown need to be calculated. While \( e \), \( p \) and \( q \) remain constant in the model, the values of \( \tilde{\mu} \) and \( \sigma_a \) will also remain fixed. However, if \( p \) and \( q \) are assumed to be random, the values of \( \tilde{\mu} \) and \( \sigma_a \) will also become random and will be updated at every step in the Gibbs sampling algorithm. In order to calculate these values, the following two identities will be introduced:

\[ P(\mu > 0 | \sigma^2, q, r) = p \]
\[ E(\mu | \sigma^2, q, r) = e, \]

where it can be shown that

\[ \sigma_a = \frac{e}{q \sqrt{\frac{2}{\pi} - \Phi^{-1} \left( 1 - \frac{p}{2q} \right)}} \quad (4.5) \]

and

\[ \tilde{\mu} = e - q \sigma_a \sqrt{\frac{2}{\pi}}. \]

The derivation of these results is given in Appendix C.
However a point worthy of comment is that as \( q \) approaches \( p \) (downwards), the denominator in Equation 4.5 becomes positive since the value from the inverse cumulative distribution function of the normal distribution gets closer to zero and \( e \) is always fixed and negative. So, there appears to be a natural limit that has to be imposed on \( p \) as follows:

\[
0 < p < 2q \left[ 1 - \Phi \left( \frac{2}{\sqrt{\pi}} \right) \right].
\]

Figure 4.2 shows the area for which possible combinations of \( p \) and \( q \) are plausible in this model.

Therefore, a better term for \( r \) might be

\[
r = \frac{p}{2q \left[ 1 - \Phi \left( \frac{2}{\sqrt{\pi}} \right) \right]^2}.
\]

But then the interpretation on \( r \) becomes rather untenable and the process to sample \( r \) is changed. This formulation could be of interest for future work.

**4.6 Discussion**

This Chapter has introduced a Bayesian model for estimation of performance. Under this methodology, a range of prior beliefs can be placed on a manager’s skill set and used in conjunction with the observed data to obtain posterior performance metrics.

The flexibility of the Bayesian model is useful in determining both the rank of a fund manager relative to their peers and also the likelihood a manager can outperform their
mandated benchmark. In particular, the lower the variance in the prior distributions of $p$ and $q$, the more weight is placed on the prior relative to the data in estimating the posterior performance metrics. As a result, funds with more observational evidence, *ceteris paribus*, will achieve a higher rank. The performance analyst can therefore adjust the distributions of $p$ and $q$ to suit their beliefs on the importance of a return history in measuring performance.

Furthermore, $p$ and $q$ can be altered to reflect the efficiency of the market. Fund managers operating in highly efficient markets can be assigned low prior beliefs on $p$ and $q$ as it can be assumed that there is a lower likelihood managers have skill. Consequently, more observational evidence is required to obtain positive posterior performance metrics, which would suggest investment in the fund manager was desirable and performance of the fund manager is based upon skill and not luck.
Chapter 5
Bayesian Results

5.1 Introduction

This Chapter provides an alternative approach to measuring performance of active fund managers using the Bayesian model described in Chapter 4. As previously described, prior beliefs will be placed on the likelihood that active fund managers can add value and these assumptions will be used in conjunction with the observed alphas for each manager to obtain posterior performance statistics. Particular emphasis will be placed on the Bayesian DPM as it is has been shown to be the most appropriate performance metric under the assumption of a clearly defined benchmark. However, a Bayesian SSR and other measures are also estimated for comparative purposes.

The first Section of this Chapter (Section 5.2) will allow dogmatic prior assumptions to be placed on the parameters $p$ and $q$ associated with a fund manager’s ability to beat the benchmark. The values selected for $p$ and $q$ are chosen to represent the efficiency of the market that the active manager is investing within. The more efficient the market, the less likely it is that active fund managers will be able to achieve superior returns and consequently the lower the values that should be adopted for $p$ and $q$. These prior beliefs together with the observed alphas will be used to obtain posterior performance metrics which can be used to assess both the rank of the active manager relative to its counterparts as well as its ability to achieve superior returns relative to its designated benchmark. If the posterior performance measures are indicative of outperformance
relative to benchmark for at least one active manager under the most sceptical of prior beliefs on the ability of active managers to add value, then investment in active fund management may be desirable.

The second Section of this Chapter (Section 5.3) will assume no prior knowledge of the skill of active managers or of the efficiency of the market. Instead the observed alphas will be used within the Bayesian framework to estimate these parameters and test the level of skill of active fund management.

5.2 Posterior distribution of performance measures when \( p \) and \( q \) are constant

Initially, it is assumed that the probability that a manager is skilled \( q \), and the probability a manager can average above benchmark returns \( p \), are constant. Four combinations of the parameters are chosen; \( q = 0.05, p = 0.04 \); \( q = 0.05, p = 0.01 \); \( q = 0.01, p = 0.008 \) and finally \( q = 0.01, p = 0.002 \). These parameters are deliberately chosen to be small to represent a relatively efficient market with a low probability of active manager skill. The initial estimate of the expected value of alpha, \( e \), is -14 basis points per month in line with Baks, et al. (2001) who estimate that manager fees and transaction costs are 8 and 6 basis points respectively.

Given these assumptions, values for \( \tilde{\mu} \), \( \sigma_a \) and \( U \) are calculated using the results obtained in Section 4.5.6. These are shown in Table 5.1 for the different values of \( p \) and \( q \), where \( e \) is held fixed at -14 basis points per month. For any given probability that a manager is skilled it is evident that, as the probability of a manager having enough skill to outperform benchmark is increased, the overall expected performance relative to benchmark of the unskilled \( \tilde{\mu} \) decreases. In a world where no manager is skilled we would expect the performance of the active fund managers to be below benchmark by the fees and transaction costs associated with active management (14 basis points). As the number of skilled managers together with an ability to beat benchmark increases,
the unskilled manager’s performance is expected to decrease as they lose through trading with skilled managers. The performance of the unskilled manager $\bar{\mu}$, is a function of both the costs associated with active management as well as the loss from trading with the skilled. Furthermore, as $p$ increases relative to $q$, the half normal distribution of the skilled managers becomes positively dispersed and, as a result, $\sigma_a$ increases.

### 5.2.1 Simulation results

Once the parameters of the model are known, the Gibbs sampling algorithm is applied to obtain samples from the joint conditional posterior distributions for the model outlined in Section 4.5. The Gibbs sampler was run for 5000 iterations for each individual manager and the relevant posterior estimates were calculated using this output. Convergence seemed to have occurred rapidly where this was checked only by sight. As a result of this iterative approach, it is necessary to discard the first 1000 iterations due to burn-in. This is standard practice when using MCMC methods since the chain will not converge automatically and it is necessary to ensure that each variable has converged before the amount of burn-in is estimated and ultimately discarded. Samples from the joint posterior distribution were retained for each of the variables. Those of interest were used to find the posterior distribution of the DPM, SSR, $\bar{\mu}$, $z$ and the LPM. The expected value of each measure was calculated. These results, ranked by

<table>
<thead>
<tr>
<th>$q$</th>
<th>$p$</th>
<th>$\bar{\mu}$</th>
<th>$\sigma_a$</th>
<th>$U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.04</td>
<td>-0.001662</td>
<td>0.006559</td>
<td>-0.003062</td>
</tr>
<tr>
<td>0.01</td>
<td>0.008</td>
<td>-0.001446</td>
<td>0.005706</td>
<td>-0.002846</td>
</tr>
<tr>
<td>0.01</td>
<td>0.002</td>
<td>-0.001409</td>
<td>0.001099</td>
<td>-0.002809</td>
</tr>
</tbody>
</table>

Table 5.1: Values for $\bar{\mu}$, $\sigma_a$ and $U$ for different values of $p$ and $q$. 
the Bayesian DPM, are provided in Tables 5.2 through to Table 5.5. The traditional non-Bayesian DPM is also provided for comparative purposes.

Table 5.2 provides the relevant posterior estimates where it is assumed that \( q = 0.05 \) and \( p = 0.04 \). Under these \textit{a priori} assumptions only 13 of the 287 managers within the dataset had posterior expected values of the DPM and SSR that were positive. The top 20 managers were the same under both performance measures. However, there is some variation in the rank of the managers within the top 20. In particular, while the DPM and SSR ranked the top four managers the same, there is a degree of variation in rank given by the DPM and SSR from the fifth manager downwards. This result is a function of the distribution of the observed alphas with positively skewed distributions achieving a higher rank through the DPM than the SSR. The Bayesian \( z \) represents the percentage of times that \( z=1 \) (probability that a manager is skilled) in the MCMC simulations and is another possible performance measure. However, as this statistic only measures skill and not outperformance relative to benchmark it is not as insightful as the DPM or SSR.

Table 5.3 provides the Bayesian statistics when \( q = 0.05 \) and \( p = 0.01 \). In this case it is assumed that only 20\% of the skilled managers, who represent 5\% of the active manager population, are able to achieve a return above that of benchmark. Under these assumptions, no manager obtained an expected posterior DPM or SSR that is positive and therefore investment in active management would not be advised given these prior beliefs. It should be noted that the rank of the managers in this table should not be used to infer the relative ability of the active fund managers. Managers with high volatility in alpha will have a higher LPM or standard deviation of alpha and will consequently have a DPM or SSR that is closer to zero. They will subsequently be ranked higher than other managers with the same (negative) alpha who have lower dispersion relative to the benchmark. Thus it is only appropriate to rank managers with positive performance metrics when using the DPM or SSR.
Table 5.2: Comparisons of Bayes DPM performance measure relative to other Bayes measures (p=0.05, q=0.04).

<table>
<thead>
<tr>
<th>Manager</th>
<th>Bayes DPM (Rank)</th>
<th>Bayes Mu (Rank)</th>
<th>Bayes SSR (Rank)</th>
<th>Bayes LPM (Rank)</th>
<th>Bayes Z (Rank)</th>
<th>DPM (Rank)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>92</td>
<td>0.5774 (1)</td>
<td>0.0059 (1)</td>
<td>0.2863 (1)</td>
<td>0.0120 (76)</td>
<td>0.8695 (1)</td>
<td>0.4133 (9)</td>
<td>54</td>
</tr>
<tr>
<td>208</td>
<td>0.5193 (2)</td>
<td>0.0033 (2)</td>
<td>0.2272 (3)</td>
<td>0.0096 (157)</td>
<td>0.6654 (4)</td>
<td>0.5395 (3)</td>
<td>30</td>
</tr>
<tr>
<td>145</td>
<td>0.4763 (3)</td>
<td>0.0029 (3)</td>
<td>0.2371 (2)</td>
<td>0.0077 (230)</td>
<td>0.8282 (2)</td>
<td>0.3586 (12)</td>
<td>54</td>
</tr>
<tr>
<td>11</td>
<td>0.4567 (4)</td>
<td>0.0022 (4)</td>
<td>0.2145 (4)</td>
<td>0.0068 (251)</td>
<td>0.7578 (3)</td>
<td>0.4867 (4)</td>
<td>42</td>
</tr>
<tr>
<td>72</td>
<td>0.2609 (5)</td>
<td>0.0012 (8)</td>
<td>0.1089 (8)</td>
<td>0.0082 (218)</td>
<td>0.5416 (8)</td>
<td>0.5860 (2)</td>
<td>36</td>
</tr>
<tr>
<td>147</td>
<td>0.2584 (6)</td>
<td>0.0015 (5)</td>
<td>0.1240 (5)</td>
<td>0.0086 (201)</td>
<td>0.6414 (6)</td>
<td>0.3026 (19)</td>
<td>54</td>
</tr>
<tr>
<td>146</td>
<td>0.2527 (7)</td>
<td>0.0014 (7)</td>
<td>0.1197 (7)</td>
<td>0.0087 (200)</td>
<td>0.6234 (7)</td>
<td>0.3016 (20)</td>
<td>54</td>
</tr>
<tr>
<td>136</td>
<td>0.2515 (8)</td>
<td>0.0014 (6)</td>
<td>0.1220 (6)</td>
<td>0.0084 (210)</td>
<td>0.6551 (5)</td>
<td>0.3254 (15)</td>
<td>59</td>
</tr>
<tr>
<td>117</td>
<td>0.1944 (9)</td>
<td>0.0011 (9)</td>
<td>0.0829 (9)</td>
<td>0.0102 (147)</td>
<td>0.4927 (9)</td>
<td>0.3193 (16)</td>
<td>42</td>
</tr>
<tr>
<td>209</td>
<td>0.1860 (10)</td>
<td>0.0008 (11)</td>
<td>0.0675 (10)</td>
<td>0.0111 (115)</td>
<td>0.3936 (11)</td>
<td>0.4509 (7)</td>
<td>28</td>
</tr>
<tr>
<td>210</td>
<td>0.1747 (11)</td>
<td>0.0008 (12)</td>
<td>0.0630 (12)</td>
<td>0.0111 (118)</td>
<td>0.3936 (11)</td>
<td>0.4562 (6)</td>
<td>28</td>
</tr>
<tr>
<td>127</td>
<td>0.1528 (12)</td>
<td>0.0010 (10)</td>
<td>0.0659 (11)</td>
<td>0.0118 (86)</td>
<td>0.4590 (10)</td>
<td>0.4436 (8)</td>
<td>46</td>
</tr>
<tr>
<td>54</td>
<td>0.0625 (13)</td>
<td>0.0005 (13)</td>
<td>0.0248 (13)</td>
<td>0.0176 (16)</td>
<td>0.3425 (13)</td>
<td>0.2725 (23)</td>
<td>60</td>
</tr>
<tr>
<td>15</td>
<td>-0.0013 (14)</td>
<td>0.0003 (14)</td>
<td>-0.0163 (15)</td>
<td>0.0127 (64)</td>
<td>0.3039 (15)</td>
<td>0.2068 (30)</td>
<td>78</td>
</tr>
<tr>
<td>119</td>
<td>-0.0037 (15)</td>
<td>-0.0004 (15)</td>
<td>-0.0346 (19)</td>
<td>0.0082 (217)</td>
<td>0.3061 (14)</td>
<td>0.3182 (17)</td>
<td>42</td>
</tr>
<tr>
<td>140</td>
<td>-0.0168 (16)</td>
<td>-0.0014 (48)</td>
<td>-0.0121 (14)</td>
<td>0.0856 (1)</td>
<td>0.0462 (73)</td>
<td>-0.0546 (130)</td>
<td>28</td>
</tr>
<tr>
<td>41</td>
<td>-0.0288 (17)</td>
<td>-0.0012 (27)</td>
<td>-0.0216 (16)</td>
<td>0.0421 (2)</td>
<td>0.0719 (40)</td>
<td>0.1849 (34)</td>
<td>36</td>
</tr>
<tr>
<td>122</td>
<td>-0.0292 (18)</td>
<td>-0.0012 (28)</td>
<td>-0.0219 (17)</td>
<td>0.0421 (3)</td>
<td>0.0662 (44)</td>
<td>0.1802 (36)</td>
<td>36</td>
</tr>
<tr>
<td>284</td>
<td>-0.0296 (19)</td>
<td>-0.0012 (29)</td>
<td>-0.0221 (18)</td>
<td>0.0419 (4)</td>
<td>0.0670 (43)</td>
<td>0.1751 (37)</td>
<td>36</td>
</tr>
<tr>
<td>125</td>
<td>-0.0355 (20)</td>
<td>-0.0006 (16)</td>
<td>-0.0388 (20)</td>
<td>0.0122 (75)</td>
<td>0.2432 (16)</td>
<td>0.1909 (32)</td>
<td>72</td>
</tr>
<tr>
<td>Manager</td>
<td>Bayes DPM (Rank)</td>
<td>Bayes Mu (Rank)</td>
<td>Bayes SSR (Rank)</td>
<td>Bayes LPM (Rank)</td>
<td>Bayes Z (Rank)</td>
<td>DPM (Rank)</td>
<td>N</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>---------------</td>
<td>------------</td>
<td>----</td>
</tr>
<tr>
<td>140</td>
<td>-0.0167 (1)</td>
<td>-0.0014 (178)</td>
<td>-0.0119 (1)</td>
<td>0.0855 (1)</td>
<td>0.0431 (186)</td>
<td>-0.0546 (130)</td>
<td>28</td>
</tr>
<tr>
<td>122</td>
<td>-0.0334 (2)</td>
<td>-0.0013 (88)</td>
<td>-0.0241 (2)</td>
<td>0.0421 (4)</td>
<td>0.0590 (83)</td>
<td>0.1802 (36)</td>
<td>36</td>
</tr>
<tr>
<td>284</td>
<td>-0.0334 (3)</td>
<td>-0.0013 (96)</td>
<td>-0.0241 (3)</td>
<td>0.0421 (2)</td>
<td>0.0562 (95)</td>
<td>0.1751 (37)</td>
<td>36</td>
</tr>
<tr>
<td>41</td>
<td>-0.0335 (4)</td>
<td>-0.0013 (98)</td>
<td>-0.0241 (4)</td>
<td>0.0421 (3)</td>
<td>0.0517 (120)</td>
<td>0.1849 (34)</td>
<td>36</td>
</tr>
<tr>
<td>135</td>
<td>-0.0552 (5)</td>
<td>-0.0014 (132)</td>
<td>-0.0405 (5)</td>
<td>0.0285 (5)</td>
<td>0.0506 (127)</td>
<td>0.1507 (42)</td>
<td>5</td>
</tr>
<tr>
<td>207</td>
<td>-0.0568 (6)</td>
<td>-0.0013 (94)</td>
<td>-0.0416 (6)</td>
<td>0.0251 (6)</td>
<td>0.0524 (115)</td>
<td>0.3360 (13)</td>
<td>18</td>
</tr>
<tr>
<td>52</td>
<td>-0.0616 (7)</td>
<td>-0.0014 (170)</td>
<td>-0.0452 (7)</td>
<td>0.0231 (7)</td>
<td>0.0457 (165)</td>
<td>-0.1826 (203)</td>
<td>36</td>
</tr>
<tr>
<td>101</td>
<td>-0.0688 (8)</td>
<td>-0.0013 (113)</td>
<td>-0.0507 (8)</td>
<td>0.0205 (8)</td>
<td>0.0499 (133)</td>
<td>-0.0202 (106)</td>
<td>35</td>
</tr>
<tr>
<td>100</td>
<td>-0.0695 (9)</td>
<td>-0.0014 (149)</td>
<td>-0.0512 (9)</td>
<td>0.0203 (9)</td>
<td>0.0466 (158)</td>
<td>-0.0789 (143)</td>
<td>66</td>
</tr>
<tr>
<td>107</td>
<td>-0.0700 (10)</td>
<td>-0.0014 (184)</td>
<td>-0.0516 (10)</td>
<td>0.0202 (10)</td>
<td>0.0446 (174)</td>
<td>-0.0704 (138)</td>
<td>78</td>
</tr>
<tr>
<td>54</td>
<td>-0.0702 (11)</td>
<td>-0.0013 (16)</td>
<td>-0.0520 (12)</td>
<td>0.0188 (14)</td>
<td>0.0964 (18)</td>
<td>0.2725 (23)</td>
<td>60</td>
</tr>
<tr>
<td>108</td>
<td>-0.0703 (12)</td>
<td>-0.0014 (198)</td>
<td>-0.0518 (11)</td>
<td>0.0201 (11)</td>
<td>0.0415 (198)</td>
<td>-0.1000 (156)</td>
<td>78</td>
</tr>
<tr>
<td>129</td>
<td>-0.0708 (13)</td>
<td>-0.0013 (114)</td>
<td>-0.0523 (13)</td>
<td>0.0199 (12)</td>
<td>0.0510 (124)</td>
<td>0.0218 (86)</td>
<td>43</td>
</tr>
<tr>
<td>102</td>
<td>-0.0718 (14)</td>
<td>-0.0013 (101)</td>
<td>-0.0531 (14)</td>
<td>0.0195 (13)</td>
<td>0.0533 (110)</td>
<td>0.0498 (75)</td>
<td>45</td>
</tr>
<tr>
<td>92</td>
<td>-0.0725 (15)</td>
<td>-0.0012 (8)</td>
<td>-0.0541 (15)</td>
<td>0.0170 (20)</td>
<td>0.1428 (8)</td>
<td>0.4133 (9)</td>
<td>54</td>
</tr>
<tr>
<td>261</td>
<td>-0.0749 (16)</td>
<td>-0.0013 (87)</td>
<td>-0.0555 (16)</td>
<td>0.0186 (16)</td>
<td>0.0559 (97)</td>
<td>0.0858 (60)</td>
<td>54</td>
</tr>
<tr>
<td>50</td>
<td>-0.0783 (17)</td>
<td>-0.0014 (189)</td>
<td>-0.0581 (17)</td>
<td>0.0187 (15)</td>
<td>0.0444 (176)</td>
<td>-0.3586 (272)</td>
<td>12</td>
</tr>
<tr>
<td>69</td>
<td>-0.0795 (18)</td>
<td>-0.0013 (104)</td>
<td>-0.0590 (18)</td>
<td>0.0176 (17)</td>
<td>0.0548 (100)</td>
<td>0.0020 (97)</td>
<td>78</td>
</tr>
<tr>
<td>71</td>
<td>-0.0803 (19)</td>
<td>-0.0013 (112)</td>
<td>-0.0596 (19)</td>
<td>0.0174 (19)</td>
<td>0.0548 (100)</td>
<td>-0.0292 (115)</td>
<td>78</td>
</tr>
<tr>
<td>70</td>
<td>-0.0804 (20)</td>
<td>-0.0014 (137)</td>
<td>-0.0597 (20)</td>
<td>0.0175 (18)</td>
<td>0.0522 (117)</td>
<td>-0.0286 (114)</td>
<td>78</td>
</tr>
</tbody>
</table>

Table 5.3: Comparisons of Bayes DPM performance measure relative to other Bayes measures (p=0.05, q=0.01).
In Table 5.4 the probability that a manager is skilled is reduced to 0.01 and the probability a manager has enough skill to beat the benchmark is 0.008. Under these prior assumptions, only four of the 287 managers achieve positive posterior expected values for the DPM and SSR metrics. The rank of the top four managers is consistent between the DPM and SSR. However, the estimate of $z$ for the manager ranked third is higher than the manager ranked in fourth position. This suggests that there is more confidence that the fourth ranked manager has skill relative to the third ranked manager. Yet, under the assumption of skill, the third ranked manager has more skill than the fourth. This is evidenced by the posterior expected value of average alpha $\mu$ which is higher for the third ranked manager.

Table 5.5 provides the results of the final set of assumptions where $q = 0.001$ and $p = 0.002$. In this case, similar to when $q = 0.05$ and $p = 0.01$, no manager’s posterior expected value of the DPM and SSR is positive and therefore it would not be wise to invest in active management.

Table 5.6 provides the Pearson’s correlation coefficients between the Bayesian DPM, SSR, $\mu$ and $z$ with traditional versions of the DPM and SSR. It is evident that when the ratio of $p$ over $q$ is small the correlation falls between the Bayesian DPM and SSR relative to the traditional measures. This is a result of an increase in the number of managers with negative expected values relating to their posterior DPM and SSR. Due to the functional form of the SSR and DPM, managers with increased variability in alpha may have a higher rank than managers with less variability, conditional on both managers’ posterior average alphas being negative. Consequently, under these circumstances caution must be taken when interpreting the resultant correlation coefficients. Nevertheless, it is of interest to note that the correlation between the Bayesian and classical DPM and SSR measures range between 0.3343 and 0.6412 - significantly less than the relationship between the classical measures. This suggests the Bayesian approach provides significantly ranking results relative to the traditional measures.
<table>
<thead>
<tr>
<th>Manager</th>
<th>Bayes DPM (Rank)</th>
<th>Bayes Mu (Rank)</th>
<th>Bayes SSR (Rank)</th>
<th>Bayes LPM (Rank)</th>
<th>DPM (Rank)</th>
<th>Bayes Z (Rank)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>92</td>
<td>0.2637 (1)</td>
<td>0.0025 (1)</td>
<td>0.1247 (1)</td>
<td>0.0144 (37)</td>
<td>0.4133 (9)</td>
<td>0.4867 (1)</td>
<td>54</td>
</tr>
<tr>
<td>145</td>
<td>0.1290 (2)</td>
<td>0.0005 (2)</td>
<td>0.0452 (2)</td>
<td>0.0094 (171)</td>
<td>0.3586 (12)</td>
<td>0.3685 (2)</td>
<td>54</td>
</tr>
<tr>
<td>208</td>
<td>0.1101 (3)</td>
<td>0.0002 (3)</td>
<td>0.0268 (3)</td>
<td>0.0118 (88)</td>
<td>0.5395 (3)</td>
<td>0.2428 (4)</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td>0.1067 (4)</td>
<td>0.0001 (4)</td>
<td>0.0235 (4)</td>
<td>0.0082 (217)</td>
<td>0.4867 (4)</td>
<td>0.3232 (3)</td>
<td>42</td>
</tr>
<tr>
<td>136</td>
<td>-0.0066 (5)</td>
<td>-0.0004 (5)</td>
<td>-0.0284 (9)</td>
<td>0.0097 (154)</td>
<td>0.3254 (15)</td>
<td>0.2219 (5)</td>
<td>59</td>
</tr>
<tr>
<td>147</td>
<td>-0.0102 (6)</td>
<td>-0.0004 (6)</td>
<td>-0.0310 (10)</td>
<td>0.0100 (152)</td>
<td>0.3026 (19)</td>
<td>0.2039 (6)</td>
<td>54</td>
</tr>
<tr>
<td>146</td>
<td>-0.0137 (7)</td>
<td>-0.0004 (7)</td>
<td>-0.0326 (11)</td>
<td>0.0100 (150)</td>
<td>0.3016 (20)</td>
<td>0.2010 (7)</td>
<td>54</td>
</tr>
<tr>
<td>140</td>
<td>-0.0168 (8)</td>
<td>-0.0014 (76)</td>
<td>-0.0120 (5)</td>
<td>0.0856 (1)</td>
<td>-0.0546 (130)</td>
<td>0.0073 (89)</td>
<td>28</td>
</tr>
<tr>
<td>72</td>
<td>-0.0318 (9)</td>
<td>-0.0006 (8)</td>
<td>-0.0491 (18)</td>
<td>0.0095 (162)</td>
<td>0.5860 (2)</td>
<td>0.1541 (8)</td>
<td>36</td>
</tr>
<tr>
<td>41</td>
<td>-0.0324 (10)</td>
<td>-0.0013 (27)</td>
<td>-0.0235 (6)</td>
<td>0.0423 (2)</td>
<td>0.1849 (34)</td>
<td>0.0153 (36)</td>
<td>36</td>
</tr>
<tr>
<td>122</td>
<td>-0.0330 (11)</td>
<td>-0.0013 (30)</td>
<td>-0.0239 (7)</td>
<td>0.0420 (4)</td>
<td>0.1802 (36)</td>
<td>0.0133 (41)</td>
<td>36</td>
</tr>
<tr>
<td>284</td>
<td>-0.0331 (12)</td>
<td>-0.0013 (36)</td>
<td>-0.0240 (8)</td>
<td>0.0421 (3)</td>
<td>0.1751 (37)</td>
<td>0.0131 (45)</td>
<td>36</td>
</tr>
<tr>
<td>117</td>
<td>-0.0376 (13)</td>
<td>-0.0007 (9)</td>
<td>-0.0451 (16)</td>
<td>0.0115 (104)</td>
<td>0.3193 (16)</td>
<td>0.1290 (9)</td>
<td>42</td>
</tr>
<tr>
<td>127</td>
<td>-0.0438 (14)</td>
<td>-0.0008 (10)</td>
<td>-0.0447 (15)</td>
<td>0.0131 (59)</td>
<td>0.4436 (8)</td>
<td>0.1110 (10)</td>
<td>46</td>
</tr>
<tr>
<td>54</td>
<td>-0.0449 (15)</td>
<td>-0.0009 (13)</td>
<td>-0.0375 (12)</td>
<td>0.0186 (16)</td>
<td>0.2725 (23)</td>
<td>0.0775 (13)</td>
<td>60</td>
</tr>
<tr>
<td>209</td>
<td>-0.0480 (16)</td>
<td>-0.0009 (12)</td>
<td>-0.0516 (20)</td>
<td>0.0124 (68)</td>
<td>0.4509 (7)</td>
<td>0.0899 (12)</td>
<td>28</td>
</tr>
<tr>
<td>210</td>
<td>-0.0497 (17)</td>
<td>-0.0009 (11)</td>
<td>-0.0523 (22)</td>
<td>0.0122 (73)</td>
<td>0.4562 (6)</td>
<td>0.0906 (11)</td>
<td>28</td>
</tr>
<tr>
<td>135</td>
<td>-0.0542 (18)</td>
<td>-0.0013 (35)</td>
<td>-0.0401 (13)</td>
<td>0.0282 (5)</td>
<td>0.1507 (42)</td>
<td>0.0133 (41)</td>
<td>5</td>
</tr>
<tr>
<td>207</td>
<td>-0.0563 (19)</td>
<td>-0.0013 (37)</td>
<td>-0.0415 (14)</td>
<td>0.0252 (6)</td>
<td>0.3360 (13)</td>
<td>0.0131 (45)</td>
<td>18</td>
</tr>
<tr>
<td>52</td>
<td>-0.0626 (20)</td>
<td>-0.0014 (153)</td>
<td>-0.0460 (17)</td>
<td>0.0231 (7)</td>
<td>-0.182 (203)</td>
<td>0.0033 (201)</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 5.4: Comparisons of Bayes DPM performance measure relative to other Bayes measures (p=0.01, q=0.008).

p=0.01, q=0.008
<table>
<thead>
<tr>
<th>Manager</th>
<th>Bayes DPM (Rank)</th>
<th>Bayes Mu (Rank)</th>
<th>Bayes SSR (Rank)</th>
<th>Bayes LPM (Rank)</th>
<th>Bayes Z (Rank)</th>
<th>DPM (Rank)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>-0.0166 (1)</td>
<td>-0.0014 (166)</td>
<td>-0.0119 (1)</td>
<td>0.0855 (1)</td>
<td>0.0088 (167)</td>
<td>-0.0546 (130)</td>
<td>28</td>
</tr>
<tr>
<td>41</td>
<td>-0.0336 (2)</td>
<td>-0.0013 (103)</td>
<td>-0.0242 (2)</td>
<td>0.0421 (2)</td>
<td>0.0113 (94)</td>
<td>0.1849 (34)</td>
<td>36</td>
</tr>
<tr>
<td>284</td>
<td>-0.0336 (3)</td>
<td>-0.0013 (105)</td>
<td>-0.0242 (3)</td>
<td>0.0421 (3)</td>
<td>0.0106 (105)</td>
<td>0.1751 (37)</td>
<td>36</td>
</tr>
<tr>
<td>122</td>
<td>-0.0337 (4)</td>
<td>-0.0014 (148)</td>
<td>-0.0243 (4)</td>
<td>0.0420 (4)</td>
<td>0.0093 (144)</td>
<td>0.1802 (36)</td>
<td>36</td>
</tr>
<tr>
<td>135</td>
<td>-0.0557 (5)</td>
<td>-0.0014 (138)</td>
<td>-0.0408 (5)</td>
<td>0.0282 (5)</td>
<td>0.0097 (127)</td>
<td>0.1507 (42)</td>
<td>5</td>
</tr>
<tr>
<td>207</td>
<td>-0.0569 (6)</td>
<td>-0.0014 (119)</td>
<td>-0.0416 (6)</td>
<td>0.0252 (6)</td>
<td>0.0102 (117)</td>
<td>0.3360 (13)</td>
<td>18</td>
</tr>
<tr>
<td>52</td>
<td>-0.0615 (7)</td>
<td>-0.0014 (201)</td>
<td>-0.0450 (7)</td>
<td>0.0231 (7)</td>
<td>0.0097 (127)</td>
<td>-0.1826 (203)</td>
<td>36</td>
</tr>
<tr>
<td>'101</td>
<td>-0.0690 (8)</td>
<td>-0.0014 (162)</td>
<td>-0.0508 (8)</td>
<td>0.0205 (8)</td>
<td>0.0073 (224)</td>
<td>-0.0202 (106)</td>
<td>35</td>
</tr>
<tr>
<td>100</td>
<td>-0.0695 (9)</td>
<td>-0.0014 (243)</td>
<td>-0.0512 (9)</td>
<td>0.0203 (9)</td>
<td>0.0066 (242)</td>
<td>-0.0789 (143)</td>
<td>66</td>
</tr>
<tr>
<td>107</td>
<td>-0.0696 (10)</td>
<td>-0.0013 (102)</td>
<td>-0.0513 (10)</td>
<td>0.0201 (11)</td>
<td>0.0122 (77)</td>
<td>-0.0704 (138)</td>
<td>78</td>
</tr>
<tr>
<td>108</td>
<td>-0.0697 (11)</td>
<td>-0.0014 (216)</td>
<td>-0.0514 (11)</td>
<td>0.0202 (10)</td>
<td>0.0071 (231)</td>
<td>-0.1000 (156)</td>
<td>78</td>
</tr>
<tr>
<td>129</td>
<td>-0.0709 (12)</td>
<td>-0.0014 (124)</td>
<td>-0.0523 (12)</td>
<td>0.0199 (12)</td>
<td>0.0086 (176)</td>
<td>0.0218 (86)</td>
<td>43</td>
</tr>
<tr>
<td>102</td>
<td>-0.0722 (13)</td>
<td>-0.0013 (90)</td>
<td>-0.0533 (13)</td>
<td>0.0195 (13)</td>
<td>0.0093 (144)</td>
<td>0.0498 (75)</td>
<td>45</td>
</tr>
<tr>
<td>54</td>
<td>-0.0737 (14)</td>
<td>-0.0013 (17)</td>
<td>-0.0545 (14)</td>
<td>0.0188 (14)</td>
<td>0.0188 (20)</td>
<td>0.2725 (23)</td>
<td>60</td>
</tr>
<tr>
<td>261</td>
<td>-0.0756 (15)</td>
<td>-0.0013 (52)</td>
<td>-0.0560 (15)</td>
<td>0.0185 (16)</td>
<td>0.0128 (62)</td>
<td>0.0858 (60)</td>
<td>54</td>
</tr>
<tr>
<td>50</td>
<td>-0.0779 (16)</td>
<td>-0.0014 (192)</td>
<td>-0.0577 (16)</td>
<td>0.0187 (15)</td>
<td>0.0088 (167)</td>
<td>-0.3586 (272)</td>
<td>12</td>
</tr>
<tr>
<td>92</td>
<td>-0.0795 (17)</td>
<td>-0.0013 (7)</td>
<td>-0.0591 (17)</td>
<td>0.0172 (20)</td>
<td>0.0302 (9)</td>
<td>0.4133 (9)</td>
<td>54</td>
</tr>
<tr>
<td>69</td>
<td>-0.0797 (18)</td>
<td>-0.0013 (100)</td>
<td>-0.0591 (18)</td>
<td>0.0176 (17)</td>
<td>0.0119 (81)</td>
<td>0.0020 (97)</td>
<td>78</td>
</tr>
<tr>
<td>71</td>
<td>-0.0804 (19)</td>
<td>-0.0014 (120)</td>
<td>-0.0596 (19)</td>
<td>0.0174 (18)</td>
<td>0.0091 (155)</td>
<td>-0.0292 (115)</td>
<td>78</td>
</tr>
<tr>
<td>70</td>
<td>-0.0805 (20)</td>
<td>-0.0014 (132)</td>
<td>-0.0597 (20)</td>
<td>0.0174 (19)</td>
<td>0.0102 (117)</td>
<td>-0.0286 (114)</td>
<td>78</td>
</tr>
</tbody>
</table>

Table 5.5: Comparisons of Bayes DPM performance measure relative to other Bayes measures (p=0.01, q=0.002).


Panel A: \(q=0.05, p=0.04\)

<table>
<thead>
<tr>
<th></th>
<th>Bayes DPM</th>
<th>Bayes Mu</th>
<th>Bayes SSR</th>
<th>Bayes Z</th>
<th>DPM</th>
<th>SSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes DPM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes Mu</td>
<td>0.8446</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes SSR</td>
<td>0.9448</td>
<td>0.6497</td>
<td>0.6601</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes Z</td>
<td>0.8489</td>
<td>0.9717</td>
<td>0.5827</td>
<td>0.6146</td>
<td>0.9785</td>
<td>1</td>
</tr>
<tr>
<td>DPM</td>
<td>0.6412</td>
<td>0.5602</td>
<td>0.5762</td>
<td>0.5657</td>
<td>0.9785</td>
<td>1</td>
</tr>
<tr>
<td>SSR</td>
<td>0.6162</td>
<td>0.5144</td>
<td>0.5762</td>
<td>0.5657</td>
<td>0.9785</td>
<td>1</td>
</tr>
</tbody>
</table>

Panel B: \(q=0.05, p=0.01\)

<table>
<thead>
<tr>
<th></th>
<th>Bayes DPM</th>
<th>Bayes Mu</th>
<th>Bayes SSR</th>
<th>Bayes Z</th>
<th>DPM</th>
<th>SSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes DPM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes Mu</td>
<td>0.2298</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes SSR</td>
<td>0.9838</td>
<td>0.1922</td>
<td>0.1996</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes Z</td>
<td>0.2404</td>
<td>0.9875</td>
<td>0.3665</td>
<td>0.7273</td>
<td>0.9785</td>
<td>1</td>
</tr>
<tr>
<td>DPM</td>
<td>0.4042</td>
<td>0.6676</td>
<td>0.3903</td>
<td>0.7018</td>
<td>0.9785</td>
<td>1</td>
</tr>
<tr>
<td>SSR</td>
<td>0.4263</td>
<td>0.6315</td>
<td>0.3903</td>
<td>0.7018</td>
<td>0.9785</td>
<td>1</td>
</tr>
</tbody>
</table>

Panel C: \(q=0.01, p=0.008\)

<table>
<thead>
<tr>
<th></th>
<th>Bayes DPM</th>
<th>Bayes Mu</th>
<th>Bayes SSR</th>
<th>Bayes Z</th>
<th>DPM</th>
<th>SSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes DPM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes Mu</td>
<td>0.5632</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes SSR</td>
<td>0.9631</td>
<td>0.3718</td>
<td>0.3826</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes Z</td>
<td>0.5748</td>
<td>0.9715</td>
<td>0.4386</td>
<td>0.498</td>
<td>0.9785</td>
<td>1</td>
</tr>
<tr>
<td>DPM</td>
<td>0.5156</td>
<td>0.4366</td>
<td>0.4514</td>
<td>0.4589</td>
<td>0.9785</td>
<td>1</td>
</tr>
<tr>
<td>SSR</td>
<td>0.5177</td>
<td>0.4046</td>
<td>0.4514</td>
<td>0.4589</td>
<td>0.9785</td>
<td>1</td>
</tr>
</tbody>
</table>

Panel D: \(q=0.01, p=0.002\)

<table>
<thead>
<tr>
<th></th>
<th>Bayes DPM</th>
<th>Bayes Mu</th>
<th>Bayes SSR</th>
<th>Bayes Z</th>
<th>DPM</th>
<th>SSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayes DPM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes Mu</td>
<td>0.1590</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes SSR</td>
<td>0.9842</td>
<td>0.1439</td>
<td>0.1268</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayes Z</td>
<td>0.1419</td>
<td>0.9567</td>
<td>0.3343</td>
<td>0.6666</td>
<td>0.9785</td>
<td>1</td>
</tr>
<tr>
<td>DPM</td>
<td>0.3609</td>
<td>0.5889</td>
<td>0.3593</td>
<td>0.6416</td>
<td>0.9785</td>
<td>1</td>
</tr>
<tr>
<td>SSR</td>
<td>0.3853</td>
<td>0.5677</td>
<td>0.3593</td>
<td>0.6416</td>
<td>0.9785</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.6: Correlations between Bayesian parameters of interest and traditional performance measures (DPM and SSR) under varying prior assumptions relating to \(p\) and \(q\).

The Bayesian model provides two interesting results for when \(p\) and \(q\) are held constant and are assumed known \(a\) priori. Firstly, even under restrictive assumptions on the ability of active managers to have skill (i.e. \(q = 0.01\)) there is evidence that some managers’ posterior distributions of alpha have an expected value that is positive and consequently investment in active management may be desirable. However, these results rely on the assumption that if a manager has skill then there is a high probability
that they have enough skill to beat benchmark, as measured by the ratio of $p$ to $q$. Secondly, there is a significant difference in the rank given to managers based upon the Bayesian model relative to the classical performance measures described in Chapter 3. In particular, Manager 92 was ranked as the best manager under the Bayesian framework but only ninth best under the non-Bayesian DPM. Furthermore, Manager 133, who ranked first by the non-Bayesian DPM and in the top three using the non-Bayesian SSR, $t$-test and decay rate, does not feature at all in the top 20 managers ranked in a Bayesian fashion.

In Figure 5.1 the distribution of these two managers is explored to ascertain why there is such a big discrepancy between rank using classical and Bayesian techniques. This figure shows the observed distributions of alpha for Manager 92 and Manager 133. Both of the managers' observed average alphas are above benchmark with Manager 92 averaging 0.00869 per month and Manager 133 averaging 0.00733 per month. The LPM and standard deviation were 0.01145 and 0.02070 for Manager 92 and 0.00667 and 0.01619 for Manager 133. The traditional DPM and SSR techniques only incorporate these statistics in deriving a performance metric and rank. A comparison of these two managers indicates that while Manager 92 has been able to achieve a higher average alpha, the risk associated with the performance, whether measured using a SSR or LPM approach, exceeded Manager 133. Consequently their overall performance measures were lower.

Figure 5.2 provides the posterior distributions of the DPM and SSR for Managers 92 and 133. It is evident that the posterior distributions of the SSR and DPM for Manager 133 are significantly lower than for Manager 92. This result is a consequence of the number of observed alphas available in relation to each manager. Manager 133 had only six observations while Manager 92 had 54 observations. Under the Bayesian methodology the posterior distribution of the performance measures are a function of an interrelationship between the prior beliefs and the observed data. The greater the
Figure 5.1: Observed and posterior distributions of alpha for two US large capitalisation fund managers over the period January 1998 through to June 2004. Manager 133 had the highest rank based upon classical performance measures and Manager 92 had the highest rank under Bayesian performance measures.

evidence associated with the data, the less weight is placed upon the prior beliefs in obtaining the required posterior distributions of the performance measures. In particular, the higher the number of observations and lower the dispersion associated with the observations, the lower the weight placed upon the prior in estimating the posterior distribution. Thus while Manager 133 achieved a high alpha with relatively low dispersion (because only six observations were available for this manager), the posterior distribution of alpha and consequently the DPM and SSR placed greater weight on the prior belief than the actual data. Thus, given that the prior beliefs imposed suggest that less than 5% of managers had skill (under all sets of assumptions)
Figure 5.2: Posterior distributions of the DPM and SSR for two US large capitalisation fund managers over the period January 1998 through to June 2004. Manager 133 had the highest rank based upon classical performance measures and Manager 92 had the highest rank under Bayesian performance measures.

and 95% were unskilled earning a return below benchmark of the fees and costs associated with active management and the losses associated with trading with active managers, the posterior distributions of the DPM and SSR were much lower. On the other hand, Manager 92 had 54 observations and consequently less weight was placed on the prior beliefs, and the posterior distribution of Manager 92’s performance metrics was higher.
5.3 Posterior distribution of parameters of interest when $p$ and $q$ are random

In this Section no prior assumptions will be posited on the probability that a manager has skill nor on the probability that a manager can outperform benchmark. Instead the Gibbs sampling algorithm will be used to estimate these parameters based on the available data associated with the 287 managers with a large capitalisation value style within the US market.

The Gibbs sampling algorithm for this case will now be described in order to obtain samples from the joint conditional posterior distributions for the model parameters outlined in Section 4.5.4. The main difference between the Gibbs sampling algorithm applied to this model and that considered previously is that at every iteration, a value is sampled for $q$ and then a value for $r$ (where $r = p / q$) from their respective conditional distributions. These values are then used in the next iteration of the algorithm. More specifically, both $q$ and $r$ are assumed to come from a $Beta(g, h)$ distribution with parameters $g$ and $h$ equal to 1. Given these parameter values, $q$ and $r$ have a prior distribution that is uniform in probability over the range 0 to 1 and are therefore a priori unknown. In the previous Section, both $p$ and $q$ were assumed known and constant, and so the additional step associated with placing a distribution of the parameters $q$ and $r$ was not necessary. The remaining conditional distributions will not change. As before, the Gibbs sampler was run for 5000 iterations for all managers simultaneously and the relevant posterior estimates were calculated using this output.

Figure 5.3 provides the posterior distributions of $p$ and $q$ when no prior beliefs are placed on these parameters. Given this specification, the Bayesian model estimates for $q$ and $p$ were 0.99 and 0.27, respectively. Thus, if no prior belief is placed on the ability of active managers to have skill, the evidence suggests that 99% of managers have enough skill to exceed $\bar{\mu}$ which is the benchmark return, less the fees and transaction costs associated with active management as well as any loss from trading with skilled
Figure 5.3: Posterior distributions of $p$ and $q$.

managers. Furthermore, 27% of managers have enough skill to achieve an average alpha above benchmark.

Under the assumption that $q = 0.99$ and $p = 0.27$, the values of $\mu$ and $\sigma_\alpha$ are -0.50% and 0.46% per month, respectively. Figure 5.4 provides a graphical representation of the distribution of average alphas based upon these results. Interestingly, the performance of the unskilled managers is significantly less than the costs of active trading which is estimated as fees minus transaction costs or -0.14% per month. The remaining 0.36% in underperformance can be attributed to the loss in performance from trading with skilled managers.

The remaining 99% of fund managers are estimated to have skill. However, even some of the skillful are seen to underperform benchmark by more than the fees and transaction
Figure 5.4: Distribution of average alpha under Bayesian model where the parameters $q$ and $p$ are estimated from the Gibbs sampling algorithm.

costs associated with active management. This result can be attributed to the loss in performance from trading with even more skilful managers.

Under the assumption of a priori ignorance regarding the parameters $p$ and $q$, the Bayesian DPM and SSR are estimated with the top 20 managers ranked by the Bayesian DPM given in Table 5.7. The Central Posterior Density Region (CPDR) and Highest Posterior Density Region (HPDR) outlined in Section 4.3.2 are also provided. The most significant difference in the rankings between the DPM and SSR can be seen to be associated with Manager 276 who achieved a ranking of 2 under the DPM and 14 under the SSR. On closer inspection, this manager had only one observed negative alpha and was consequently viewed more favourable under the DPM methodology. Interestingly, the fourth ranked manager (Manager 145) had the highest 95% lower confidence levels
under both the CPDR and HPDR. Under an alternative ranking methodology associated with the confidence intervals encompassing the performance metric, this manager could be perceived as optimal.

Table 5.8 provides a comparison of the manager ranks relative to the classical DPM and SSR measures as well as the Bayesian model under $q = 0.05$ and $p = 0.04$. The top ranked manager (Manager 208) under the $p$, $q$ random Bayesian model performs consistently well. However, the second ranked manager (Manager 276) has significant variation in their rank across different methodologies. This manager had only six observations and therefore, when the parameters of the Bayesian model are equal to $q = 0.05$ and $p = 0.04$, the posterior performance measures will be driven towards the priors. However, when the parameters of the Bayesian model were a priori unknown and the distributions estimated within the Gibbs sampling algorithm the expected values were $q = 0.99$ and $p = 0.27$. Under these conditions the priors are not as restrictive in driving the posterior performance measures and are based directly on the evidence within the US dataset. While Manager 276 had only six observations, the evidence suggested very strong performance with very low variation and consequently the manager was ranked highly.

5.4 Discussion

In this Chapter the Bayesian model described in Chapter 4 was used to measure performance of active fund managers. Initially, prior assumptions were placed on the efficiency of the market through the parameter values of the probability a manager is skilled and the probability a manager can exceed benchmark. The rationale behind the model was to assess whether investment in active management was desirable under stringent conditions regarding market efficiency and to provide a more robust
<table>
<thead>
<tr>
<th>Random p,q</th>
<th>Bayesian DPM</th>
<th>Bayesian SSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manager</td>
<td>Expected Value</td>
<td>95% Central Posterior Density Region</td>
</tr>
<tr>
<td>208</td>
<td>0.5265 (1)</td>
<td>(-0.1235, 1.6743)</td>
</tr>
<tr>
<td>276</td>
<td>0.5110 (2)</td>
<td>(-0.5063, 3.3560)</td>
</tr>
<tr>
<td>11</td>
<td>0.5019 (3)</td>
<td>(-0.0554, 1.3427)</td>
</tr>
<tr>
<td>145</td>
<td>0.4521 (4)</td>
<td>(-0.0297, 1.1821)</td>
</tr>
<tr>
<td>92</td>
<td>0.4316 (5)</td>
<td>(-0.0504, 1.2366)</td>
</tr>
<tr>
<td>72</td>
<td>0.4272 (6)</td>
<td>(-0.1180, 1.3190)</td>
</tr>
<tr>
<td>210</td>
<td>0.3710 (7)</td>
<td>(-0.1803, 1.3971)</td>
</tr>
<tr>
<td>209</td>
<td>0.3674 (8)</td>
<td>(-0.0874, 1.0122)</td>
</tr>
<tr>
<td>146</td>
<td>0.3598 (9)</td>
<td>(-0.1828, 1.3947)</td>
</tr>
<tr>
<td>147</td>
<td>0.3567 (10)</td>
<td>(-0.0799, 0.9762)</td>
</tr>
<tr>
<td>117</td>
<td>0.3538 (11)</td>
<td>(-0.1295, 1.1455)</td>
</tr>
<tr>
<td>136</td>
<td>0.3464 (12)</td>
<td>(-0.0858, 0.9690)</td>
</tr>
<tr>
<td>127</td>
<td>0.3079 (13)</td>
<td>(-0.1370, 1.0059)</td>
</tr>
<tr>
<td>119</td>
<td>0.2815 (14)</td>
<td>(-0.1885, 0.9605)</td>
</tr>
<tr>
<td>125</td>
<td>0.1703 (20)</td>
<td>(-0.1543, 0.6024)</td>
</tr>
</tbody>
</table>

Table 5.7: Manager rank using Bayesian DPM and SSR performance measures where $q$ and $p$ are a priori unknown.
<table>
<thead>
<tr>
<th>Manager</th>
<th>Random p,q</th>
<th>p=0.05, q=0.04</th>
<th>Classical</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bayes DPM</td>
<td>Bayes SSR</td>
<td>DPM</td>
<td>SSR</td>
</tr>
<tr>
<td>208</td>
<td>0.5265 (1)</td>
<td>0.2582 (2)</td>
<td>0.5193 (2)</td>
<td>0.2272 (3)</td>
</tr>
<tr>
<td>276</td>
<td>0.5110 (2)</td>
<td>0.1502 (14)</td>
<td>-0.1504 (147)</td>
<td>-0.1904 (246)</td>
</tr>
<tr>
<td>11</td>
<td>0.5019 (3)</td>
<td>0.2591 (1)</td>
<td>0.4567 (4)</td>
<td>0.2145 (4)</td>
</tr>
<tr>
<td>145</td>
<td>0.4521 (4)</td>
<td>0.2409 (3)</td>
<td>0.4763 (3)</td>
<td>0.2371 (2)</td>
</tr>
<tr>
<td>92</td>
<td>0.4316 (5)</td>
<td>0.2301 (4)</td>
<td>0.5774 (1)</td>
<td>0.2863 (1)</td>
</tr>
<tr>
<td>72</td>
<td>0.4272 (6)</td>
<td>0.2216 (5)</td>
<td>0.2609 (5)</td>
<td>0.1089 (8)</td>
</tr>
<tr>
<td>210</td>
<td>0.3710 (7)</td>
<td>0.1888 (10)</td>
<td>0.1747 (11)</td>
<td>0.0630 (12)</td>
</tr>
<tr>
<td>209</td>
<td>0.3674 (8)</td>
<td>0.1869 (11)</td>
<td>0.1860 (10)</td>
<td>0.0675 (10)</td>
</tr>
<tr>
<td>117</td>
<td>0.3598 (9)</td>
<td>0.1973 (6)</td>
<td>0.2527 (7)</td>
<td>0.1197 (7)</td>
</tr>
<tr>
<td>146</td>
<td>0.3567 (10)</td>
<td>0.1960 (7)</td>
<td>0.2584 (6)</td>
<td>0.1240 (5)</td>
</tr>
<tr>
<td>147</td>
<td>0.3538 (11)</td>
<td>0.1900 (9)</td>
<td>0.1944 (9)</td>
<td>0.0829 (9)</td>
</tr>
<tr>
<td>112</td>
<td>0.3464 (12)</td>
<td>0.1923 (8)</td>
<td>0.2515 (8)</td>
<td>0.1220 (6)</td>
</tr>
<tr>
<td>127</td>
<td>0.3079 (13)</td>
<td>0.1688 (12)</td>
<td>0.1528 (12)</td>
<td>0.0659 (11)</td>
</tr>
<tr>
<td>119</td>
<td>0.2815 (14)</td>
<td>0.1534 (13)</td>
<td>-0.0037 (15)</td>
<td>-0.0346 (19)</td>
</tr>
<tr>
<td>54</td>
<td>0.1968 (15)</td>
<td>0.1134 (15)</td>
<td>0.0625 (13)</td>
<td>0.0248 (13)</td>
</tr>
<tr>
<td>120</td>
<td>0.1864 (16)</td>
<td>0.1022 (17)</td>
<td>-0.0439 (22)</td>
<td>-0.0486 (25)</td>
</tr>
<tr>
<td>15</td>
<td>0.1779 (17)</td>
<td>0.1044 (16)</td>
<td>-0.0013 (14)</td>
<td>-0.0163 (15)</td>
</tr>
<tr>
<td>20</td>
<td>0.1775 (18)</td>
<td>0.0751 (21)</td>
<td>-0.1443 (131)</td>
<td>-0.1291 (167)</td>
</tr>
<tr>
<td>271</td>
<td>0.1724 (19)</td>
<td>0.0735 (22)</td>
<td>-0.1406 (118)</td>
<td>-0.1272 (165)</td>
</tr>
<tr>
<td>125</td>
<td>0.1703 (20)</td>
<td>0.0996 (18)</td>
<td>-0.0355 (20)</td>
<td>-0.0388 (20)</td>
</tr>
</tbody>
</table>

Table 5.8: Comparisons of manager rank using DPM and SSR performance measures.
performance measure in selecting the top fund managers based on past performance. The appeal of the Bayesian approach is associated with the weight it places on the data relative to the prior assumptions. Managers with few observations and/or high dispersion in the observed alphas will not rank highly in this framework as the evidence related to performance is low and a higher weight is placed on the prior assumptions.

Consequently, managers with these characteristics will have posterior distributions on their performance metrics that will be lower and closer to the prior assumption of a high probability that the manager is unskilled and, as such, will have an average alpha closer to \( \bar{\alpha} \).

The model described also provides a mechanism for disaggregating the skilled from the lucky. It is argued that the longer the observed performance history and the lower the dispersion in performance, the greater the evidence of skill. The non-Bayesian methodologies outlined in Chapter 3 take no account of these interrelationships and are therefore considered to be inferior performance measures. These results are clearly demonstrated by the high rank of Manager 133 under the non-Bayesian methodologies. Even though this manager's performance was excellent, there were only six observations available and therefore the manager did not rank as highly under the Bayesian framework as there was not enough evidence to suggest significant skill. On the other hand, the top ranked managers based upon the Bayesian model had a longer history and lower dispersion in alpha.

Furthermore, the performance metrics associated with the Bayesian model focused around the Bayesian DPM. Under the assumption of a clearly defined benchmark it was argued in Chapter 3 that the DPM provides a better measure of performance because it assesses return relative to performance under benchmark. Furthermore, the DPM had more power and less bias than other traditional measures when the underlying distribution of alpha was skewed. Other measures including a Bayesian SSR are also
estimated for comparative purposes. However, these measures are thought to be inferior due to the way they measure risk. The focus on a Bayesian DPM provides a more appropriate mechanism for selecting rank of the active fund managers, even though the differences between the Bayesian DPM and SSR are small.

Finally, the parameters $p$ and $q$ in the Bayesian model are assumed to have $Beta(1,1)$ distributions to reflect a priori ignorance. These parameters can then be estimated, based on the data within the Bayesian model, and used to obtain a posterior DPM and SSR for each manager. This version of the model also has intuitive appeal as it can be viewed as obtaining the priors relating to skill directly from the observed data of all fund managers within the dataset and then using these priors to estimate the posterior performance measures relating to each specific manager.
Chapter 6

Summary and discussion

This thesis has sought to determine the most appropriate measure of performance for active fund managers who have clearly identifiable passive investment portfolios as benchmarks. Under these circumstances the use of traditional multi-factor regression models in evaluating performance is thought to be questionable since these measures assess performance relative to a set of factors rather than specifically to the benchmark that the managers are mandated to outperform. To deal with this issue, the performance measures used in this thesis all explicitly identify a benchmark that the active manager is required to outperform. The performance measures that were considered included the Student's t-statistic and the Foster and Stutzer (2002) Decay Rate. However, the main focus of this thesis was a comparison between the SSR and the DPM, which is based upon the lower partial moments approach of Bawa (1967).

A simulation study was used to assess the power and bias of the SSR and DPM based on variations in the size of the dataset and changes in the mean performance of two simulated funds, when it was assumed that the excess returns (alphas) of the funds above benchmark were normally distributed. The simulation was then re-evaluated under the assumption of skewness in the underlying distributions of alpha. The results that were obtained suggest a similar power and bias in the SSR and DPM when the underlying distribution of alpha is normal. However, less bias and more power was present in the DPM when the distribution of alpha was skewed. A bootstrap on the US fund manager dataset was carried out and provided evidence of positive skewness in the
empirical distributions of alpha. Consequently the DPM was advocated as the choice performance measure.

Further to this investigation, a Bayesian approach analogous to the work of Baks et al. (2001) was also employed to evaluate performance. Baks et al. (2001) advocated the use of Bayesian methods to assess the desirability of investment in active fund management. Unskilled managers were assumed to underperform benchmark by the fees and transaction costs associated with active management as well as any losses from trading with skilled managers. Therefore, a Bayesian model, in conjunction with the observed data, was used to assess whether investment in active management is desirable under the most sceptical of prior beliefs on the probability that an active manager has skill. As such, this methodology was viewed as a technique for separating the lucky managers from the skilled managers when no specific benchmark could be identified. Managers with high returns, longer return histories and less variation in excess returns provide more evidence of skill and will be less influenced by the prior beliefs defined under the Bayesian model. The skill parameters were adjusted based on the efficiency of the market that the manager invests within via tweaking of the parameters of the prior distributions.

This study provided a Bayesian model for performance when the active managers had clearly identifiable passive investment portfolios as benchmarks. A Bayesian DPM and SSR were estimated for each manager in the US dataset where the prior probabilities of having skill and enough skill to exceed benchmark were allowed to have probability density distributions. The choice of distributions on the prior parameters allowed for a richer set of outcomes depending on the purpose of the analysis and the confidence placed in the prior beliefs relative to the point estimates on the priors used by Baks et al. (2001). This analysis provides an alternative technique for ranking the managers as well as a tool to test whether investment in active management is desirable based upon prior beliefs on market efficiency. The initial model described above can be viewed as
one extreme of the general model derived in this study where dogmatic prior beliefs are placed on the prior parameters.

At the other extreme a priori ignorance was posited relating to the probability that a manager has skill and the probability that a manager has enough skill to exceed the benchmark return. These parameters were then estimated based on the data within the Bayesian model and used to obtain a posterior DPM and SSR for each manager. This version of the model also had intuitive appeal since it could be viewed as obtaining the priors relating to skill directly from the observed data of all fund managers within the dataset and then using these priors to estimate the posterior performance measures relating to each specific manager. Under this framework, 99% of managers were found to have skill, however, only 27% of managers had enough skill to outperform benchmark net of the fees and transaction costs associated with active managements as well as any losses through trading with other skilled managers. These data-based estimates of the skill set parameters were then used to obtain posterior performance measures for each individual manager as an alternative to traditional performance evaluation techniques provided in the extant literature.

In conclusion, this study assessed the most appropriate measures of performance for active fund managers under the assumption of a clearly identifiable passive investment benchmark portfolio. Initially, the DPM was advocated as the most appropriate measure due to its functional form in quantifying risk as well as its power and reduction in bias when the underlying distribution of the managers’ excess returns are skewed. However, the use of a Bayesian model was argued to provide more flexibility in the measurement of performance based upon the prior beliefs of the performance analyst. The Bayesian methodology can be viewed as a technique for separating the skilled from the lucky and also providing a mechanism to use aggregate information regarding fund manager performance to obtain posterior performance metrics for individual managers.
Appendix A

Formulation of mixture distribution

Recall that the mixture distribution that was used in the simulation study in Chapter 4 was created using a normal distribution and a log normal distribution since it was required that the mixture distribution needed to be unbounded. In order to carry out the simulation study (whereby the mean of the mixture distribution was increased in increments to compare it to both a normal distribution and another mixture distribution to allow comparison of various performance measures), it is necessary to describe how the appropriate parameters are chosen.

A weight of 0.2 was allocated to a normal distribution with some pre-specified mean and standard deviation for use in the mixture distribution and therefore a weight of 0.8 was given to a relocated and rescaled lognormal distribution with some mean value and standard deviation.

Firstly, the expected value and variance for a lognormal distribution, $X$, are given below

$$E(X) = e^{\frac{\mu + \sigma^2}{2}}$$

$$Var(X) = e^{\mu + \sigma^2} \left(e^{\sigma^2} - 1\right)$$

The choice of parameters for $\mu$ and $\sigma^2$ are vital to obtain the required distribution in terms of shape. An empirical investigation suggests values of $\mu = 0$ and $\sigma = 0.5$
provide the required positively skewed distribution. A plot of the density of a lognormal distribution with these parameters is shown in Figure A1.

Thus

\[
E(X) = e^{\mu + \frac{\sigma^2}{2}} = e^{0.52}
\]

\[
Var(X) = e^{2\mu + \sigma^2} \left( e^{\sigma^2} - 1 \right) = e^{0.52} \left( e^{0.52} - 1 \right).
\]

Unfortunately, while this distribution has the correct shape, the mean and standard deviation are incorrect since the distribution is bounded at 0 (where negative alphas must be allowed). To alleviate this problem, the lognormal is relocated and rescaled so that it has, say, a mean of 0 and a standard deviation of \( \sigma_{LN} \). Therefore

\[
E \left( \frac{X - e^{\mu + \frac{\sigma^2}{2}}}{\sqrt{e^{2\mu + \sigma^2} \left( e^{\sigma^2} - 1 \right)}} \times \sigma_{LN} \right) = 0
\]

\[
Var \left( \frac{X - e^{\mu + \frac{\sigma^2}{2}}}{\sqrt{e^{2\mu + \sigma^2} \left( e^{\sigma^2} - 1 \right)}} \times \sigma_{LN} \right) = \sigma_{LN}^2.
\]

The mean alpha is then relocated to some value, \( \mu_{LN} \) per month. Initially this is set to be 0.1% in line with the previous simulations, but this mean will be increased in increments of 0.05% in due course. The calculation for the relocation of the mean is constructed as

\[
E \left( \frac{X - e^{\mu + \frac{\sigma^2}{2}}}{\sqrt{e^{2\mu + \sigma^2} \left( e^{\sigma^2} - 1 \right)}} \times \sigma_{LN} + \mu_{LN} \right) = \mu_{LN}
\]

\[
Var \left( \frac{X - e^{\mu + \frac{\sigma^2}{2}}}{\sqrt{e^{2\mu + \sigma^2} \left( e^{\sigma^2} - 1 \right)}} \times \sigma_{LN} + \mu_{LN} \right) = \sigma_{LN}^2.
\]
Figure A1: Lognormal distribution with parameters $\mu = 0$ and $\sigma = 0.5$.

Mathematically, the mean and variance of a mixture distribution derived from two independent distributions can be estimated by

\[
\mu = w_1 \mu_1 + (1-w_1) \mu_2 \\
\sigma^2 = w_1 \sigma_1^2 + (1-w_1) \sigma_2^2 + w_1 (1-w_1) (\mu_1 - \mu_2)^2.
\]

where $\mu_1$ and $\mu_2$ are the means of the two distributions respectively and $\sigma_1^2$ and $\sigma_2^2$ are the two variances respectively.

Both distributions used to create the mixture model will have identical means (i.e. $\mu_1 = \mu_2$), therefore $w_1 (1-w_1) (\mu_1 - \mu_2)^2 = 0$. 
Therefore, a mixture distribution with a mean $\mu_M$ and standard deviation 1.5\% (in line with the other examples in the simulation study) needs to be obtained. Now, suppose that $X_1 \sim N(\mu_1, \sigma_1^2 = 1.5\%)$, $X_2 \sim LN(\mu_{LN}, \sigma_{LN}^2)$, $w_1 = 0.2$ and $w_2 = 0.8$. Thus,

$$\mu_M = w_1 \mu_1 + (1 - w_1) \mu_2 = 0.2 \mu_1 + 0.8 \mu_2,$$

where $\mu_1$ and $\mu_2$ are known and $\mu_2$ is the mean of the log normal distribution.

Furthermore,

$$\sigma_M^2 = w_1^2 \sigma_1^2 + (1 - w_1)^2 \sigma_2^2 + w_1 (1 - w_1) (\mu_1 - \mu_2)^2 = 0.2^2 \times 0.015^2 + 0.8^2 \times \sigma_2^2 = 0.015^2,$$

and so

$$\sigma_2 = \sqrt{\frac{0.015^2 - 0.2^2 \times 0.015^2}{0.8^2}} = 0.01837.$$

This process is then repeated to obtain the required mixture distribution where the mean is varied in small increments.
Appendix B

Calculation of the posterior DPM

Consider Equation 4.4 shown in Section 4.4. In the Bayesian framework, this will be derived using posterior estimates obtained from the Gibbs sampling algorithm that is applied to the hierarchical model and the data. The numerator is obtained straight from the conditional distribution of $\mu$ given all of the other parameters in the model. However, the functional form of the posterior lower partial moment given the unknown parameters is required to be derived.

In general, the second order lower partial moment is expressed as follows

$$LPM = \int_{-\infty}^{T} (T-r)^2 p(r) dr .$$

Given a benchmark of $\alpha_i = 0$ we have

$$LPM = \int_{-\infty}^{0} \alpha_i^2 p(\alpha_i) d\alpha_i$$

in which

$$\alpha_i \sim N\left(\mu, \sigma^2 = \frac{1}{\lambda}\right).$$
Therefore,

\[
LPM = \int_{-\alpha}^{\infty} \frac{\lambda}{2\pi} e^{\frac{-\lambda}{2}(\alpha - \mu)^2} d\alpha. \tag{B1}
\]

Let

\[ t = \sqrt{\lambda} (\alpha - \mu), \]

then

\[
\frac{dt}{d\alpha} = \sqrt{\lambda} \quad \text{and so} \quad d\alpha = \frac{dt}{\sqrt{\lambda}}.
\]

Next substitute the above functions into Equation (B1), such that

\[
LPM = \int_{-\infty}^{\infty} \left( \frac{t}{\sqrt{\lambda}} + \bar{\mu} \right)^2 \phi(t) dt
\]

\[
= \int_{-\infty}^{\infty} \frac{t^2}{\lambda} \phi(t) dt + 2\bar{\mu} \int_{-\infty}^{\infty} \frac{t}{\sqrt{\lambda}} \phi(t) dt + \bar{\mu}^2 \int_{-\infty}^{\infty} \phi(t) dt
\]

\[
= I_1 + \frac{2\bar{\mu}}{\sqrt{\lambda}} I_2 + \bar{\mu}^2 I_3
\]

for simplicity.

Firstly,

\[
I_1 = \int_{-\infty}^{\infty} \frac{t^2}{\lambda} \phi(t) dt = \frac{1}{\lambda^2} \int_{-\infty}^{\infty} t(t\phi(t)) dt
\]

Then, integrating by parts,

\[
I_1 = \frac{1}{\lambda} \left\{ -t\phi(t) \right\}_{-\infty}^{\infty} - \frac{1}{\lambda} \int_{-\infty}^{\infty} (-\phi(t)) dt
\]

\[
= \frac{\bar{\mu}}{\lambda} \phi(-\bar{\mu}/\lambda) + \frac{1}{\lambda} \Phi(-\bar{\mu}/\lambda)
\]

96
Now,

\[
I_2 = \int_{-\infty}^{-\tilde{\mu}\sqrt{\lambda}} t\phi(t)\,dt
= -\phi(t)|_{-\infty}^{-\tilde{\mu}\sqrt{\lambda}}
= -\phi(-\tilde{\mu}\sqrt{\lambda})
\]
as above.

Finally,

\[
I_3 = \int_{-\infty}^{-\tilde{\mu}\sqrt{\lambda}} \phi(t)\,dt = \Phi(-\tilde{\mu}\sqrt{\lambda}).
\]

Therefore,

\[
LPM = I_1 + \frac{2\tilde{\mu}}{\sqrt{\lambda}} I_2 + \tilde{\mu}^2 I_3
= \frac{\tilde{\mu}\sqrt{\lambda}}{\lambda} \phi(-\tilde{\mu}\sqrt{\lambda}) + \frac{1}{\lambda} \Phi(-\tilde{\mu}\sqrt{\lambda}) - \frac{2\tilde{\mu}}{\sqrt{\lambda}} \phi(-\mu\sqrt{\lambda}) + \tilde{\mu}^2 \Phi(-\tilde{\mu}\sqrt{\lambda})
= \left(\tilde{\mu}^2 + \frac{1}{\lambda}\right) \Phi(-\tilde{\mu}\sqrt{\lambda}) - \frac{\tilde{\mu}}{\sqrt{\lambda}} \phi(-\tilde{\mu}\sqrt{\lambda}).
\]
Appendix C

Calculation of $\tilde{\mu}$ and $\sigma_\alpha$

Recall that in Section 4.5.6 the necessity of evaluating values of $\tilde{\mu}$ and $\sigma_\alpha$ was discussed. Depending on the model that is being considered (whether $p$ and $q$ are treated as constant in the model, or whether $p$ and $q$ are treated as random variables), these will remain the same throughout the simulation process or will be updated at each step in the Gibbs sampling algorithm. However, the actual derivation remains the same and it is only through the posterior distribution that the updating step changes.

The first identity that is considered in Section 4.5.6 is

$$P\left(\tilde{\mu} > 0 \mid \sigma^2, q, r \right) = P\left(z = 0 \mid \sigma^2, q, r \right) P\left(\tilde{\mu} > 0 \mid z = 0, \sigma^2, q, r \right) + P\left(z = 1 \mid \sigma^2, q, r \right) P\left(\tilde{\mu} > 0 \mid z = 1, \sigma^2, q, r \right)$$

where

$$P\left(z = 0 \mid \sigma^2, q, r \right) = 1 - q, \quad P\left(z = 1 \mid \sigma^2, q, r \right) = q \quad \text{and} \quad P\left(\tilde{\mu} > 0 \mid z = 0, \sigma^2, q \right) = 0.$$

The only term that requires some thought is $P\left(\tilde{\mu} > 0 \mid z = 1, \sigma^2, q, r \right)$. This is shown to be equal to

$$P\left(\tilde{\mu} > 0 \mid z = 1, \sigma^2, q, r \right) = 2 P\left( Z > \frac{0 - \tilde{\mu}}{\sigma_\alpha} \right) = 2 \left(1 - \Phi\left(\frac{-\tilde{\mu}}{\sigma_\alpha}\right)\right).$$
Thus,

\[ ((1-q)x0) + \left(q \times 2 \left(1 - \Phi \left(\frac{-\mu}{\sigma_a} \right)\right) \right) = p \]

Therefore,

\[ 2q \left(1 - \Phi \left(\frac{-\mu}{\sigma_a} \right)\right) = p. \]  

(C1)

Next consider the estimation of the average alpha for an individual manager. This can be expressed as follows

\[
E(\mu | \sigma^2, q) = P(z=0 | \sigma^2, q, r) E(\mu | \sigma^2, q, r, z=0) + P(z=1 | \sigma^2, q, r) E(\mu | \sigma^2, q, r, z=1)
\]

\[ = e. \]

where \( P(z=0 | \sigma^2, q, r) = 1 - q \) and \( P(z=1 | \sigma^2, q, r) = q \).

In Section 4.5.2, \( E(\mu | \sigma^2, z=0) = E(\mu | \sigma^2, q, r, z=0) = \tilde{\mu} \) was defined. Therefore, the only unknown part in the expression above is \( E(\mu | \sigma^2, q, r, z=1) \). Now, \( E(\mu | \sigma^2, q, r, z=1) \) is the expected value of a half normal distribution where

\[
f(\mu | z=1, \sigma^2, q, r) \sim N(\tilde{\mu}, \sigma_a^2) \mathbf{1}(\mu > \tilde{\mu}).
\]

Let \( A \) and \( \omega \) be defined such that

\[ A \sim N(\tilde{\mu}, \sigma_a^2) \text{ and } \omega = (A | A > \tilde{\mu}). \]

Then,

\[
f(\mu | A > \tilde{\mu}, \sigma^2, q, r) = \frac{f_A(\omega)}{P(A > \tilde{\mu})} = 2 \frac{\phi \left( \frac{\omega - \tilde{\mu}}{\sigma_a} \right)}{\sigma_a}.
\]
So,

\[
E(\mu | A > \bar{\mu}, \sigma^2, q, r) = \int_\bar{\mu}^\infty \frac{2}{\sigma_a} \phi \left( \frac{\omega - \bar{\mu}}{\sigma_a} \right) d\omega \\
= 2 \int_0^\infty (\bar{\mu} + \sigma_a t) \phi(t) dt \\
= 2\bar{\mu}(1/2) + 2\sigma_a \int_0^\infty \left( t/\sqrt{2\pi} \right) e^{-t^2} dt \\
= \bar{\mu} + 2\sigma_a/\sqrt{2\pi} \\
= \bar{\mu} + \sigma_a \sqrt{2/\pi}.
\]

Consequently,

\[
((1-q) \times \bar{\mu}) + (q \times (\bar{\mu} + \sigma_a \sqrt{2/\pi})) = \bar{\mu} + \sigma_a \sqrt{2/\pi} = e.
\]  

(C2)

From Equation (C2),

\[
\bar{\mu} = e - q\sigma_a \sqrt{2/\pi}
\]  

(C3)

so that Equation (C1) now becomes

\[
2q \left( 1 - \Phi \left( \frac{- (e - q\sigma_a \sqrt{2/\pi})}{\sigma_a} \right) \right) = p.
\]

Therefore,

\[
q \sqrt{\frac{2}{\pi}} - \frac{e}{\sigma_a} = \Phi^{-1} \left( 1 - \frac{p}{2q} \right)
\]

and so,

\[
\sigma_a = \frac{e}{q \sqrt{\frac{2}{\pi}} - \Phi^{-1} \left( 1 - \frac{p}{2q} \right)}.
\]  

(C4)
Note, for the case when $p$ and $q$ are constant, it is possible to estimate $\mu$ and $\sigma_a$ for different values of $p$, $e$ and $q$ using Equations C3 and C4. However, when $p$ and $q$ are treated as random, the only term that is held fixed is $e$, and so from iteration to iteration, a new value is sampled for $p$ and $q$. If the probability of a manager being skilled is 0, that is $q = 0$, then an active manager will have an average alpha of minus fees minus costs. As this probability of skilled management is increased, $\mu$ falls due to the losses from trading with an increasing number of skilled managers. The higher the probability a manager can obtain an alpha higher than benchmark, $p$, the lower $\mu$ as more return must be taken from the unskilled managers.
Appendix D

Details of the Gibbs Sampler

Recall that if it is possible to express the joint posterior distribution of the parameters of interest, then the MCMC method to use is the Gibbs sampling algorithm.

Recall that the joint distribution of \((\alpha, \mu, \lambda, z, q, r)\), assuming for the moment that \(p\) and \(q\) are random, can be written as follows

\[
f(\lambda, \mu, z, q, r, \alpha) \propto f(\lambda, \mu, z, q, r | \alpha)
\]

\[
\propto \prod_{i=1}^{T} \left\{ \frac{\lambda}{2\pi} \exp \left[ -\frac{\lambda}{2}(\alpha_i - \mu)^2 \right] \right\}
\]

\[
\times \left\{ (1-q)I(z=0)I(\mu=\mu) + (q)I(z=1)I(\mu > \mu) \frac{2}{\sigma^2} e^{-\frac{1}{2\sigma^2}(\mu-\bar{\mu})^2} \right\}
\]

\[
\times \frac{1}{\lambda} \times f(r, q)
\]

for \(0 < q < 1, 0 < p < q, z = 0\) or \(1, \lambda > 0, \mu \geq \bar{\mu}\) and \(r = p / q\).

Then, in order to run the Gibbs sampling algorithm, the full conditional distributions of the parameters need to be evaluated. The first three steps for this process are the same for both cases considered. When \(p\) and \(q\) are random, there are extra steps to be carried out in the Gibbs sampler for each of these parameters. Therefore, the first three steps will be discussed in turn, where \(\lambda\) is considered first, and will follow on from this for the second case. Thus,
\[ f(\lambda | \alpha, \mu, z, q, r) \propto \lambda^{(T-z-1)/2} e^{-\frac{1}{2} \sum_{i=1}^{T} (\alpha_i - \mu)^2}, \]

and so,

\[ f(\lambda | \alpha, \mu, z, q, r) \sim \text{Gamma} \left( \frac{T}{2}, \frac{1}{2} \sum_{i=1}^{T} (\alpha_i - \mu)^2 \right). \]

The next step involves drawing a value of \( z \) then of \( \mu \). Therefore,

\[ (z, \mu) \sim f(z, \mu | \alpha, \lambda, q, r), \]

for which a value of \( z \) must first be drawn and, conditional on whether \( z = 0 \) (a manager is unskilled) or \( z = 1 \) (manager is skilled), a value of \( \mu \) can be drawn from the relevant conditional posterior distribution of \( \mu \).

More formally, draw \( z \sim f(z | \alpha, \lambda, q, r) \) and then draw a value of \( \mu \) such that \( \mu \sim f(\mu | \alpha, \lambda, q, r, z) \).

Now

\[ f(z | \alpha, \lambda, q, r) \propto f(z, \alpha, \lambda, q, r) = E_\mu \left( f(z, \alpha, \lambda, q, r, \mu) \right) \]

\[ = \begin{cases} 
(1-q) e^{-\frac{(\lambda-z)}{2} \sum_{i=1}^{T} (\alpha_i - \mu)^2}, & z = 0 \text{ since } \mu = \tilde{\mu} \text{ when } z = 0 \\
q \int_{\tilde{\mu}}^{\infty} e^{-\frac{(\lambda-z)}{2} \sum_{i=1}^{T} (\alpha_i - \mu)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\mu-\tilde{\mu})^2} \ d\mu, & z = 1 
\end{cases} \]

In order to evaluate the integral in the above conditional posterior distribution, let

\[ I_1 = \int_{\tilde{\mu}}^{\infty} \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\mu-\tilde{\mu})^2} \ d\mu = \frac{2}{\sigma \sqrt{2\pi}} \int_{\tilde{\mu}}^{\infty} e^{-\frac{\mu^2}{2\sigma^2}} \ d\mu, \]

where
\[
Q = \lambda \left[ (T-1)s_a^2 + T(\mu - \bar{\alpha})^2 \right] + \frac{1}{\sigma_a^2} (\mu - \bar{\mu})^2
\]

and

\[
\bar{\alpha} = \frac{1}{T} \sum_{i=1}^{T} \alpha_i \quad \text{and} \quad s_a^2 = \frac{1}{T-1} \sum_{i=1}^{T} (\alpha_i - \bar{\alpha})^2 .
\]

Thus,

\[
Q = \lambda (T-1)s_a^2 + T\lambda \left( \mu^2 - 2\mu \bar{\alpha} + \bar{\alpha}^2 \right) + \frac{1}{\sigma_a^2} (\mu^2 - 2\mu \bar{\mu} + \bar{\mu}^2 ) = \lambda (T-1)s_a^2 + \mu^2 \left( T\lambda + \frac{1}{\sigma_a^2} \right) - 2\mu \left( T\lambda \bar{\alpha} + \bar{\mu} \right) + \left( T\lambda \bar{\alpha}^2 + \bar{\mu}^2 \right)
\]

where

\[
a_1 = T\lambda + \frac{1}{\sigma_a^2} \quad b_1 = T\lambda \bar{\alpha} + \frac{\bar{\mu}}{\sigma_a^2}.
\]

Therefore,

\[
Q = \lambda (T-1)s_a^2 + T\lambda \bar{\alpha}^2 + \frac{\bar{\mu}^2}{\sigma_a^2} + a_1 \left( \mu^2 - 2\mu \frac{b_1}{a_1} + \left( \frac{b_1}{a_1} \right)^2 \right) - \frac{b_1^2}{a_1} = a_1 \left( \mu - \frac{b_1}{a_1} \right)^2 + \lambda (T-1)s_a^2 + T\lambda \bar{\alpha}^2 + \frac{\bar{\mu}^2}{\sigma_a^2} - \frac{b_1^2}{a_1}.
\]

Now, let

\[
c_i = \lambda (T-1)s_a^2 + T\lambda \bar{\alpha}^2 + \frac{\bar{\mu}^2}{\sigma_a^2} - \frac{b_1^2}{a_1}.
\]

then,
\[ I_1 = \int_{\bar{\mu}}^{\infty} e^{-\frac{1}{2} \sum_{a_i}(\mu - \bar{\mu})^2} \frac{2}{\sigma_a} e^{\frac{-1}{2\sigma_a^2}(\mu - \bar{\mu})^2} d\mu \]
\[ = \frac{2}{\sigma_a} e^{-\frac{1}{2}} \int_{\bar{\mu}}^{\infty} e^{-\frac{1}{2} a_i(\mu - \bar{\mu})^2} d\mu \]

Now,

\[ \int_{\bar{\mu}}^{\infty} a_i e^{-\frac{1}{2} a_i(\mu - \bar{\mu})^2} d\mu = P(R > \bar{\mu}) \]

where

\[ R \sim N \left( \frac{b_1}{a_1}, \frac{1}{a_1} \right). \]

Thus,

\[ P(R > \bar{\mu}) = P \left( Z > \frac{\bar{\mu} - b_1/a_1}{1/a_1} \right) \]
\[ = 1 - \Phi \left( \frac{\bar{\mu} - b_1}{\sqrt{a_1}} \right). \]

Therefore, in obtaining a sampled value of \( z \), sample from a Bernoulli distribution with probabilities defined as follows:

\[ f(z | a, \lambda, q, r) = \begin{cases} \frac{f_0}{f_0 + f_1}, & z = 0 \\ \frac{f_1}{f_0 + f_1}, & z = 1 \end{cases} \]

where

\[ f_0 = (1-q)e^{-\frac{1}{2} a_i(\bar{\mu} - \alpha)^2} \]

and
The next step is to sample a value for $\mu$ conditional on what value of $z$ is sampled in the previous step. Therefore, if $z = 0$, then let $\mu = \tilde{\mu}$, otherwise, if $z = 1$, then draw $\mu \sim f(\mu | \alpha, \lambda, q, r, z = 1)$ where

$$f(\mu | \alpha, \lambda, q, r, z = 1) \propto e^{-\frac{1}{2} \sum_{m} (\mu-a_{m})^2} e^{-\frac{1}{2} (\mu-\tilde{\mu})^2}, \mu > \tilde{\mu};$$

that is draw a value from

$$\mu \sim N\left(\frac{h}{a_{1}}, \frac{1}{a_{1}}\right) \mathcal{I}(\mu > \tilde{\mu})$$

where $a_{1}$ and $b_{1}$ are defined as before.

It remains to obtain the conditional distributions of $q$ and $r$ and describe how to sample a value of each of these parameters within the Gibbs sampling algorithm.

First, consider $q$. Now,

$$f(q | \alpha, z, \mu, \lambda, r) \propto \begin{cases} q^{a-1} (1-q)^b & \text{if } z = 0 \\ q^a (1-q)^{b-1} & \text{if } z = 1 \end{cases}$$

Therefore,

$$(q | z, \alpha, \mu, \lambda, r) \sim \text{Beta}(a + z, b + 1 - z)$$

Furthermore, in considering any future inference on $p$, this must be done through the parameter $r$ where $r = p/q$ and $0 < p < q$. Therefore,

$$(r | z, \alpha, \mu, \lambda, q) \propto r^{(a-1)} (1-r)^{(b-1)} \times r^\nu(\mu > 0) (1-r)^{\nu(\mu > \tilde{\mu}) - \nu(\mu > 0)}$$
where $N(*)$ refers to the number of times that the condition * is met. Therefore,

$$(r | z, \alpha, \mu, \lambda, q) \sim \text{Beta}(g + N(\mu > 0), h + N(\mu > \bar{\mu}) - N(\mu > 0)).$$

When $p$ and $q$ are constant, the two steps just described for sampling $p$ and $q$ can be omitted from the Gibbs sampler.
Bibliography


