

MUTUAL FUND PERFORMANCE MEASUREMENT AND FUND RATING

GAO ZHANGPENG



SCHOOL OF HUMANITIES AND SOCIAL SCIENCES

A thesis submitted to the Nanyang Technological
University in fulfillment of the requirement for the degree
of Doctor of Philosophy

2006

HG
4530

Acknowledgement

The completion of this dissertation would not have been possible without the invaluable advice and assistance of many people; and it is both a duty and a pleasure to express and record my appreciation.

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. SHAHIDUR Rahman for his dedicated supervision, encouragement, and support throughout the study at Nanyang Technological University. His extensive knowledge, serious research attitude, and kindness were extremely valuable to me. He not only imparted knowledge but also taught me how to be kind to other people.

My appreciation also goes to the members of my Ph.D. dissertation panel: Dr Cao Yong and Dr. Chandrasekhar Krishnamurti, for their valuable advice and comments.

Beyond the committee members, I am also indebted to Dr. Randolph Tan for his patient guidance through my second year, and all the professors who have taught me various subjects and/or touched my life here at NTU. Among them are: Dr. Chen Kang, Dr. Huang Weihong, Dr. Li Zhi-Feng, Michael, Dr. Liu Yunhua, Professor LIM Chong Yah, Dr. Joseph Dennis ALBA, Dr TAN Gee Kwang, Randolph, and other staff members for their assistance throughout the PhD

program. Many thanks must also be expressed to Ms. Carol Loke, Ms. Wendy and Ms. Tan Yam Hua for their help in the administrative affairs.

I am also indebted to Nanyang Technological University for providing me with financial support and first-class resources. The discussions and friendship with my fellow students have also contributed substantially to the completion of this thesis. In particular, I would like to thank Liu Xinyi, Li Ting, Zhang Xin, Huang Huamei, Yu Lei, Yi Yuandong and Wu Zhan.

Last but not least, I want to express my deepest love and gratitude to my father, Mr. Gao Wenzheng, my mother, Ms. Hu Mingzhu, and my wife, Ms. Ju Danfeng for their continuous support and encouragement. It is their understanding and love that have given me strength and determination to complete this research work.

Table of Contents

Acknowledgement	2
Table of Contents.....	4
Abstract.....	7
Chapter I Introduction.....	9
1.1 Preamble	9
1.2 Objectives	10
1.3 Contributions.....	12
1.4 Structure of the Thesis	15
Chapter II Literature Review of Fund Performance Measures.....	18
2.1 Introduction.....	18
2.2 Unconditional Measures	18
2.3 Market Timing	23
2.4 Conditional Measures of Alpha	26
2.5 Concluding Remarks.....	30
Chapter III Performance Measure Based on Return-based Style Analysis	33
3.1 Introduction.....	33
3.2 Return-based Style Analysis.....	35
3.3 RBSA Measure by Quadratic Programming.....	38
3.3.1 Rationale of RBSA Measure.....	39
3.3.2 Formulation of RBSA Measure in Quadratic Programming	41
3.3.3 Optimality Conditions of Quadratic Programming.....	43

3.3.4 Interior Point Method.....	45
3.4 Concluding Remarks.....	46
Chapter IV A Comparative Simulation Experiment of Performance Measures.....	47
4.1 Introduction.....	47
4.2 Setup of Simulation Experiment.....	47
4.3 Simulation Results and Analysis	50
4.3.1 Simulation Results and Analysis of Alpha and R^2	50
4.3.2 Simulation Results and Analysis of Style Coefficients (betas)	56
4.4 Conclusion	61
Chapter V A New Direction of Fund Rating Based on Finite Normal Mixture Model.....	63
5.1 Introduction.....	63
5.2 Motivation of Finite Normal Mixture Model.....	66
5.3 Parametric Formulation of Finite Normal Mixture Distribution Model.....	69
5.4 EM Algorithm for Finite Normal Mixture Model	73
5.5 Determination of the Number of Components by Parametric Bootstrap Procedures.....	80
5.6 Fund Rating Procedures.....	86
Chapter VI Empirical Study of Fund Performance and Rating	89
6.1 Introduction.....	89
6.2 Data Description	90
6.3 Fund Styles and Performance	91
6.4 Nonparametric Estimation of Density Function and Normality Check.....	102
6.4.1 Nonparametric Estimation of Density Function	102
6.4.2 Formal Normality Test.....	104

6.5 Specification of Finite Normal Mixture Model	122
6.5.1 Large Growth Funds	122
6.5.2 Large Blend Funds	125
6.5.3 Large Value Funds	126
6.5.4 Medium Funds	127
6.5.5 Small Funds	127
6.5.6 Conclusion	128
6.6 Estimation of Finite Normal Mixture Model	129
6.6.1 Large Growth Funds	133
6.6.2 Large Blend Funds	136
6.6.3 Large Value Funds	138
6.6.4 Medium Funds	140
6.6.5 Small Funds	142
6.6.6 Conclusion	144
6.7 Fund Rating	145
6.8 Fund Performance Persistence	146
Chapter VII Conclusion	150
Appendix A: Modified Newton Method	157
Appendix B: Summary of Fund Rating	161
Reference:	163

Abstract

How to measure and rate the performance of funds is the main question that perplexes the fund industry. We show that the measure proposed by Sharpe, which is derived from return-based style analysis, is superior to other measures by a comparative simulation study. We formally develop the econometric methodology on how to implement it, and show that the measure has several advantages because of its quadratic programming estimation techniques instead of the regression method of traditional measures.

The natural step after fund performance measurement is to rate the funds. We rate the funds from a new perspective, which is based on the distribution of alphas instead of the preset percentiles as Morningstar. The multimodal shape and formal normality tests prompt us to model the distribution by finite normal mixture model. We assume that the difference of expected fund performance arises from the segmented market information and the differentiated ability of managers to acquire and analyze the information. We introduce the parametric bootstrap procedure to determine the number of performance groups in the model, i.e. to specify the model. Then EM algorithm is exploited to estimate the model. Based on the estimated posterior probabilities of the fund, we can assign the rating of the fund. We outline our fund rating procedures after developing the methodology.

Finally, we empirically study the mutual funds in US. We construct a unique database by merging the equity mutual fund data from Bloomberg and Morningstar, which consists of all the funds that have existed for at least nine years at November 2004. The results show that the average performance of these funds in the last three years is around zero after adding back management fees. But small-cap funds could deliver positive performance even after deducting management fees. At the same time we observe that the higher performance of small-cap funds is associated with the higher investment risk in that group of funds. The study of the number of performance groups over nine years shows a tendency of a merge of information sets for all five categories of funds, which indicates that the market became more efficient from 1995 to 2004. We show the mechanism of how the fund market becomes more efficient over time from the perspective of fund rating. Finally, we study the fund performance persistence based on our fund rating results. We don't observe the persistence over the whole sample period, but we do observe the persistence from the first period (1995-1998) to the second period (1998-2001) for large-cap funds.

Mutual Fund Performance Measurement and Fund Rating

Chapter I Introduction

1.1 Preamble

Numerous mutual funds are available in the market. The funds differ in asset allocations, investment strategies, risk levels, and returns. When investors choose from the pool of funds, they usually consider the following two questions. Firstly, what is the performance of the fund? Investors would like to know whether and how much the fund managers could obtain abnormal returns above passive portfolios through stock selection and market timing. This question poses a challenging problem of how to measure fund performance. Various performance measurement models have been proposed since the original paper by Jensen (1968), for example multi-factor measures by Fama and French (1993) and Carhart (1997) and conditional measures by Ferson and Scadt (1996) among others. But they are subject to some limitations. The most obvious one is benchmark inefficiency, which may lead to totally different inferential results from the same dataset using different measures or benchmarks. This leads to a hot debate about whether fund managers can deliver abnormal returns over the decades, for example Ippolito

(1989) and Elton, et al. (1993). As the measures become more refined and sophisticated, the question is still open for further investigation.

Secondly, suppose we have correctly measured fund performance, then which funds are superior funds and which are inferior? This is actually a fund rating issue. We need to classify the funds into several performance groups. If the market is efficient where information is well absorbed and transmitted, then there is no need to do fund rating. The differences among funds are just results of "Luck". For the inefficient market, the performance may differentiate because of the different information that the fund managers have or the different ability to analyze the available information. A simple way of fund rating is to rank funds according to the risk-adjusted performance (e.g. Jensen alpha, 1968), and then arbitrarily divide them into several groups based on preset percentiles and the fixed number of performance groups, for example Morningstar's method. It has several limitations in practice. In addition, there is little theoretical basis to rate funds in this way.

1.2 Objectives

The purpose of this thesis is to contribute to the fund performance measurement and fund rating literature theoretically and empirically by studying the above two questions. We show that the measure proposed by Sharpe, which is derived from return-based style analysis, is superior to other measures by a comparative simulation study. We term it as RBSA measure. The measure differs from traditional measures, such as Jensen's (1968) measure, at the underlying estimation

techniques, which implies the rationales of the two methods even though the models look similar. Traditional measures use the least squares method to estimate the alpha and betas under the framework of a linear regression model. It is to minimize the sum of squares of residuals in the regression model. On the other hand, the RBSA measure exploits the mathematical algorithm of the interior point method to estimate betas under the framework of a convex quadratic programming problem. It is to minimize the variance of residuals, not the absolute value of residuals in regression models. In other words the RBSA measure attempts to identify a set of style exposures that best mimics the return behavior of the fund.

Secondly, we intend to develop a new theoretical framework for fund rating, which is based on the cross-sectional distribution of the alphas of all the funds. The framework is based on the crucial assumption that the expected fund performance may be different and the difference of the expected fund performance arises from the segmented market information and/or the differentiated ability of managers to acquire and analyze the information. We model the distribution as a finite normal mixture model. Under this framework, the number of performance groups is not fixed, but determined by exploring the distribution of alphas of all the funds. For example, if the alphas closely cluster around only one value, we may conclude that there is only one performance group. If alphas cluster around three values, then we may conclude that there are three performance groups. We introduce the parametric bootstrap procedures to formally test the number of performance groups in each period. After we specify and estimate the model, we have the posterior

probabilities that the fund belongs to each performance group, which provides a straightforward way to determine the number of funds in each performance group and funds rating. It also provides an intuitively appealing interpretation of the estimates. We can assess the expected performance and investment risk of each performance group.

With the methodologies developed on fund performance measurement and rating, finally we intend to empirically study the US equity mutual funds in terms of fund styles, performance, rating, and performance persistence from 1995 to 2004. In addition we explore the market efficiency issue and its endogenous mechanism from the perspective of fund rating.

1.3 Contributions

The thesis contributes to the literature in the following aspects. First of all, we develop a new theoretical framework for fund rating study, which is based on the actual distribution of alphas instead of preset percentiles. The new method models the distribution of alphas under the framework of a finite normal mixture model. Until now, there is little academic literature that deals with the fund rating issue. The mainstream method is Morningstar's method, a commercial fund rating method, where it arbitrarily presets the number of performance groups, the cut-off percentiles, and the number of funds in each performance group. However Morningstar's method is subject to some limitations, which will be shown in detail in Chapter V. We assume that there is a group-structure in the distribution of

alphas, which is caused by an inefficient market where information is not well transmitted and absorbed in the market. If the market information is segmented or fund managers have a different ability to acquire and analyze information, the expected performance of managers may be different according to their information that they have. We first determine the number of performance groups by parametric bootstrap procedures, and then use the Expectation-Maximization (EM) algorithm to estimate the finite mixture normal model under group-structure assumption of alphas. Therefore, the number of performance groups is dynamically determined by the distribution of alphas of all the funds in the market. The performance group of the fund is dynamically determined by the fund's posterior probabilities, which are updated simultaneously with the performance information of all the funds. In addition, we can also estimate the expected performance and investment risk of each performance group for each fund asset category.

Secondly, we show that the measure proposed by Sharpe, which is derived from return-based style analysis, is superior to other measures by a comparative simulation study. We formulate the performance measurement problem as a convex quadratic programming problem. The measure has several advantages over the traditional measures, such as Jensen's (1968) measure and the Fama-French (1993) three-factor measure. First, it does not rely on the normal assumption of residuals in the model. Under this framework of a convex quadratic programming problem, we do not need any assumption about variables' distributions in the model, which is different from traditional measures where we assume that the disturbance ε is

normally distributed in the regression model. Second, we include all the investable indexes in the model. This way we circumvent the benchmark inefficiency problems. But we require that the indexes should be exhaustive and exclusive, which are not difficult to accommodate with a lot of indexes developed by commercial companies. Third, we can estimate fund styles with this measure. Fund styles provide more detailed information about fund risk than a simple beta or standard deviation of returns. We further conduct a comparative simulation experiment to test the measure and compare it with other measures. We find the measure is more accurate, efficient and robust than traditional measures. In the simulation, we show that Jensen's measure is not appropriate to measure small-cap funds and market-timing ability. Fama-French three-factor measures are not able to identify funds styles of well-diversified funds.

Thirdly, we empirically study the mutual funds in US. We construct a unique database by merging the mutual fund data from Bloomberg and Morningstar, which consist of all the funds that have existed for at least nine years till 2004. The results show that the average performance of funds in the last three years is around zero after adding back management fees. But small-cap funds could deliver positive performance even after deducting management fees. At the same time we observe that the higher performance of small-cap funds is compensated by the higher investment risk. The study of the evolvement of the number of performance groups over nine years shows a tendency of a merge of information sets for all five categories of funds, which indicates that the market becomes more efficient from

1995 to 2004. We show the process of how the fund market becomes more efficient over time from the perspective of fund rating. Finally we study the fund performance persistence based on fund rating results. We don't observe the persistence over the whole sample period, but we do observe the persistence from the first period (1995-1998) to the second period (1998-2001) for large-cap funds.

1.4 Structure of the Thesis

The thesis is arranged in the following way. In Chapter II, we critically review current performance measures and discuss their limitations. We review unconditional measures, market-timing measures and conditional measures and discuss some limitations, such as benchmark inefficiency, spurious market timing, and unrealistic normal assumption. Among them, benchmark inefficiency is most devastating, which may lead to totally different inferential results from the same dataset using different measures or benchmarks, for example, the research by Ippolito (1989) and Elton et al. (1993).

Admitting the limitations of current fund performance measures, in Chapter III we examine the measure proposed by Sharpe, which is derived from return-based style analysis. It is formulated as a convex quadratic programming problem. We illustrate and compare the rationale of this measure with traditional measures. We observe some advantages of this measure. Firstly, it circumvents the benchmark inefficiency. Secondly, it does not rely on a normal assumption. Thirdly, the

estimation results provide information about fund styles. We also show the technical details on how to measure fund performance with this method.

In Chapter IV we set up a comparative simulation experiment to test the accuracy, efficiency, and robustness of the measure, and compared the measure with other mainstream measures reviewed in Chapter II. The results show that the measure is the best in terms of accuracy, efficiency and robustness, while other measures suffer some limitations.

In Chapter V, instead of using preset percentiles to rate funds, we propose a new direction of fund rating, which is based on the cross-sectional distribution of the alphas. We model the distribution as a finite normal mixture model. We show the motivation of the model and justify the model specification. Then we proceed to discuss how to implement the method by EM and parametric bootstrap procedures. Finally we summarize our fund rating method.

In Chapter VI, we empirically study the performance and rating of mutual funds in US. We construct a unique database by merging the data from Bloomberg and Morningstar. Then we study the performance of the funds and the distribution of alphas. The results prompt us to study fund rating from the cross-sectional distribution of alphas. We show the evolvement of the number of information sets, the expected performance and investment risk of the funds from 1995 to 2004.

Finally we present the rating of all the funds over the sample period, and based on the rating we study the performance persistence issue.

In Chapter VII, we summarize the main contributions and findings of our research.

We also show the direction of future research.

Chapter II Literature Review of Fund Performance Measures

2.1 Introduction

In this chapter we review the traditional fund performance measures, most of which are estimated by the regression method and are actually an application of the Capital Asset Pricing Model (CAPM). Depending on their assumptions about the measure of fund performance, the measure of fund risk, and the behavior of fund managers, we classify the measures into three general categories: (i) unconditional measures, where it is assumed that there is no market-timing activity; (ii) market-timing measures, where we control the measurement bias caused by the fund manager's market timing behavior; (iii) conditional measures that control the investment strategies using publicly available macroeconomic information. We critically analyze the measures and point out their limitations, which prompts us to propose the measure in Chapter III and the subsequent analysis.

2.2 Unconditional Measures

Under the framework of unconditional measures, the alpha and betas are static, not depending on the managers' market-timing behavior and their responses to the predetermined public macroeconomic information. In this section we review the

Jensen measure and the extensions of the Jensen measure, that is, multi-factor measures.

The Jensen measure is a single factor measure, which implicitly assumes that the market portfolio could capture all the relevant risk of the fund.

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \varepsilon_{it} \quad (2.1)$$

where r_{it} is the monthly return of fund i at period t , r_{ft} is the risk-free rate at period t , r_{mt} is the monthly return of market portfolio at time t , and ε_{it} is the disturbance. The alpha, α_i , is to measure the performance of fund i during the evaluation period. β_{im} is the covariance of the fund return and market portfolio return, divided by the variance of the market portfolio return. β_{im} is a measure of the fund's systematic risk, i.e. the sensitivity of the fund return to the market portfolio return. According to the CAPM, the fund is expected to be compensated for the systematic risk it bears.

One issue related to the Jensen measure is the difficulty to find a proxy for market portfolio (benchmark inefficiency). This issue has been extensively investigated in the past three decades. Jensen (1968) studied 115 mutual funds from 1955 to 1964 and found that on average the funds earned 1.1% less annually than they should have earned given their systematic risk. An analysis of gross returns with expenses added back indicated that 42% of the funds did better than the overall market on a risk-adjusted basis, whereas the analysis of net returns indicated only 34% of the

funds outperformed the market. Jensen concluded that on average these funds could not beat a buy-and-hold policy, which is known as passive investment strategy. Contrary to Jensen's findings, Ippolito (1989) found that the estimated risk-adjusted return for the mutual fund industry is greater than zero even after accounting for transaction costs and expenses. Ippolito attributes nonnegative alpha to the existence of informed actions by management. In the paper, Ippolito used Standard&Poors (S&P) 500 as the benchmark (the proxy for market portfolio) to study the 143 mutual funds during the period of 1965 to 1984. Elton et al. (1993) used the same data as the Ippolito's data and noticed that returns of S&P stocks, returns of non-S&P stocks, and returns of bonds are significant factors in performance measurement. They argued that Ippolito's conclusions were due to the choice of an inefficient benchmark and that Ippolito's conclusion are reversed after taking account of mutual funds' holdings of non-S&P500 stocks and bonds,.

Empirical studies, particularly those of Lehmann and Modest (1987) and Grinblatt and Titman (1994), stress the sensitivity of the fund performance to the benchmarks chosen. Lehmann and Modest (1987) employed the standard CAPM benchmarks and a variety of Arbitrage Pricing Theory (APT) benchmarks to investigate this question. They found little similarity between the absolute and relative mutual fund rankings obtained from these alternative benchmarks, which suggest that the conventional measures of abnormal mutual fund performance are sensitive to the benchmarks chosen. Grinblatt and Titman (1994) used a sample of 279 mutual funds that exist from 1974 to 1984 and constructed 109 passive

portfolios to investigate the sensitivity of the performance to the benchmarks. They found that the measures generally yield similar inferences when using the same benchmark and inferences can vary, even from the same measure, when using different benchmarks.

Roll (1977, 1978) criticized that it is practically impossible to find a proxy for the market portfolio. This difficulty poses a serious problem when evaluating fund performance. Since the market portfolio used is not a perfect market portfolio, the covariance of the return of the fund and the return of market portfolio can't correctly measure the risk born by the funds. So the alpha derived from the Jensen measure in (2.1) is biased.

Implicitly admitting that the stock market portfolio can not capture all the risk factors, Fama and French (1993) proposed a three-factor model. The alpha is estimated through the following time series regression:

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \beta_{iSMB}r_{SMB,t} + \beta_{iHML}r_{HML,t} + \varepsilon_{it} \quad (2.2)$$

where β_{iSMB} is the sensitivity of the excess return of fund i to the return of the SMB portfolio and β_{iHML} is the sensitivity of the excess return of fund i to HML portfolio. They suggest that securities returns in excess of risk-free rate are explained by the sensitivity of their return to three factors: (i) the excess return on a broad market portfolio, denoted by $(r_{mt} - r_{ft})$; (ii) the difference between the return on a portfolio of small-cap stocks and the return on a portfolio of large-cap stocks, denoted by

$r_{SMB,t}$; (iii) the difference between the return on a portfolio of high book-to-market stocks and the return on a portfolio of low B/M ratio stocks, denoted by $r_{HML,t}$.

Carhart (1997) found another significant risk factor: the momentum factor, which can explain the variation of stock returns. After adding this factor to Fama-French three-factor model, they proposed a four-factor model. We call it the Carhart four-factor model. The additional factor captures the one year momentum anomaly, recognized by Jegadeesh and Titman (1993). The model may also be interpreted as a performance attribution model. The coefficients on the factor-mimicking portfolios, β_{im} , β_{iSMB} , β_{iHML} , and β_{iPRVYR} , indicate the proportion of mean return, attributable to four elementary strategies: high beta stocks versus low beta stocks, large-cap stocks versus small-cap stocks, value stocks versus growth stocks, and one year return momentum stocks versus contrarian stocks. The model is,

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \beta_{iSMB}r_{SMB,t} + \beta_{iHML}r_{HML,t} + \beta_{iPRVYR}r_{PRVYR,t} + \varepsilon_{it} \quad (2.3)$$

where $r_{SMB,t}$, $r_{HML,t}$, and $r_{PRVYR,t}$ are returns on value-weighted zero-investment factor-mimicking portfolios for size, B/M ratio, and one year momentum factors at period t, and ε_{it} is the disturbance.

The multi-factor measures, such as the Fama-French three-factor measure and the Carhart four-factor measure, suffer three problems as they become more refined and complicated. First, they can not overcome the benchmark inefficiency; they can only reduce the inefficiency effect by adding more risk factors. Second, it is

not easy to interpret the coefficients of the risk factors in regression models except for the coefficient of stock market portfolio. The signs of the coefficients may indicate the fund styles, but they provide no information of the asset allocation of the fund to each asset category or sub-asset groups. Third, the measures are biased, if managers adjust asset allocation or loadings of stocks according to the expectation of market movement or the predetermined market information, such as risk-free rate, dividend yield of the stock market and term structure. This bias will be illustrated in detail in the next two sections on market timing and conditional measures.

2.3 Market Timing

When fund managers adopt a market-timing strategy, which is common in fund management, the previous unconditional measures are biased. Market timing means that the fund managers change asset allocations or the risk level of stocks on the basis of her/his expectation of market movement. When managers successfully time the market movement, the measures without controlling market-timing behavior are biased (Ferson and Schadt, 1996). The Treynor-Mazuy model and the Henriksson-Merton model are proposed to deal with this issue.

In the Treynor-Mazuy model (Treynor and Mazuy, 1966), later refined by Bhattacharya & Pfleiderer (1983), they assume that the risk level of the portfolio varies when managers adopt market-timing strategies. The empirical study by Lee & Rahman (1990) indicated that there is some evidence of superior micro- and

macro-forecasting ability on the part of the fund manager. The later study of Coggin, et al. (1993) showed that the average selectivity measure is positive and the average timing measure is negative regardless of the choice benchmark portfolio or estimation model.

In the original model the beta meaning the risk of the fund is a linear function of excess market return: $\beta_i = \beta_{im} + \gamma_i^{TM}(r_{mt} - r_{ft})$. When the expected market return is higher than the risk free rate, the risk of the portfolio is higher in order to obtain a higher expected return. On the other hand, when the expected market return is below the risk free rate, the managers reduce the portfolio's exposure to the market. With this idea, the Jensen model in (2.1) is modified by adding a quadratic term:

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \gamma_i^{TM}(r_{mt} - r_{ft})^2 + \varepsilon_{it} \quad (2.4)$$

where r_{it} , r_{ft} , r_{mt} and β_{im} are the same as what we defined in equation (2.1). γ_i^{TM} measures the managers' market timing ability, where a positive γ_i^{TM} indicates that managers have superior market timing ability. α_i measures the fund performance due to managers' active management after controlling market-timing behavior.

After adding the quadratic market-timing term to multi-factor measures in (2.2) and (2.3), we obtain the following two measures respectively.

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \gamma_i^{TM}(r_{mt} - r_{ft})^2 + \beta_{iSMB}r_{SMB,t} + \beta_{iHML}r_{HML,t} + \varepsilon_{it} \quad (2.5)$$

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \gamma_i^{TM}(r_{mt} - r_{ft})^2 + \beta_{iSMB}r_{SMB,t} + \beta_{iHML}r_{HML,t} + \beta_{iPRVYR}r_{PRVYR,t} + \varepsilon_{it} \quad (2.6)$$

Instead of assuming that the beta is a linear function of excess market return, Henriksson and Merton (1981) proposed an alternative model. They assumed that managers choose two different levels of risk depending on the managers' forecast of the market return. If the excess market return is positive, a higher risk level is chosen. If the excess market return is negative, a lower risk level is chosen. They modified the Jensen measure in (2.1) by adding a term $MAX(0, r_{mt} - r_{ft})$, which gives

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \gamma_i^{HM} MAX(0, r_{mt} - r_{ft}) + \varepsilon_{it} \quad (2.7)$$

where γ_i^{HM} is used to measure the manager's market timing ability. A positive γ_i^{HM} indicates superior market timing ability. Henriksson and Merton (1981) interpret the term $MAX(0, r_{mt} - r_{ft})$ as the payoff to an option on the market portfolio with an exercise price equal to the risk free asset return. α_i is the fund performance in the Henriksson-Merton market-timing model

Here when we add the market-timing term, $MAX(0, r_{mt} - r_{ft})$, into (2.2) and (2.3) respectively, we get:

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \gamma_i^{HM} MAX(0, r_{mt} - r_{ft}) + \beta_{iSMB} r_{SMB,t} + \beta_{iHML} r_{HML,t} + \varepsilon_{it} \quad (2.8)$$

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \gamma_i^{HM} MAX(0, r_{mt} - r_{ft}) + \beta_{iSMB} r_{SMB,t} + \beta_{iHML} r_{HML,t} + \beta_{iPR1YR} r_{PR1YR,t} + \varepsilon_{it} \quad (2.9)$$

In this section, we consider the market-timing in two ways: the Treynor-Mazuy model and the Henriksson-Merton model. However, the results from market-timing models are not satisfactory. Admati et al. (1986) pointed out that it is difficult to separate timing from selective activities. Furthermore when managers invest in options or option-like securities, spurious timing ability and selection ability may be observed, see Jagannathan and Korajczyk (1986). Ferson and Schadt (1996) further pointed out that similar problems arise adopting dynamic strategies if trading takes place more frequently than the time interval of performance measures, for example, one month. The empirical studies also supported the above-mentioned limitations. Cai et al. (1997) constructed 36 passive portfolios and observed that most of passive portfolios' timing coefficients are negative and there is significant market-timing behavior for portfolios including large-cap value stocks. Therefore, the market timing models have problems in application. Correct separation of market timing from selection ability may depend on some impractical constrictions.

2.4 Conditional Measures of Alpha

Ferson and Schadt (1996) further explored the market-timing behavior by including lagged macro information into the model. The conditional model recognizes that the risk and the expected return of the fund may vary over time given some predetermined public macro information. They argue that managers, who obtain higher returns using public information, should not be interpreted as superior performance. Unconditional measures and market-timing measures may confuse the performance from the managers' selection ability and market-timing ability

with the performance from the managers' responses to the changing macro information.

There is a problem associated with the measures that we discussed in section 2 and 3. The alpha and betas are biased when managers respond to information of the last period, such as risk-free interest rate and term structure. Ferson and Schadt (1996) show that,

$$p \lim(\beta_{iu}) = \beta_{ic} + B_i^T Cov((r_{mt} - r_{ft}), z_{t-1}) / var((r_{mt} - r_{ft}))$$

$$p \lim(a_i) = E(r_{mt} - r_{ft})(\beta_{ic} - p \lim(\beta_i)) + Cov((r_{mt} - r_{ft}), B_i^T z_{t-1}) \quad (2.10)$$

where T denotes transpose. β_{iu} is the unconditional measure of fund i . β_{ic} is the true beta in the conditional model, and z_{t-1} is an innovation vector of the lagged information variables Z_{t-1} , i.e. $z_{t-1} = Z_{t-1} - E(Z_{t-1})$. B_i^T are the response coefficients to the innovations of the lagged information variable. From (2.10), we can see β_{iu} is a biased estimator of β_{ic} . The direction of the bias depends on the covariance between the excess market return and the innovations of the lagged macro information. Ferson and Schadt (1996) used four public information variables that were useful to predict the market: the lagged level of short term Treasury bill rate, the January dummy, the lagged dividend yield of the stock index, and the lagged measure of the slope of term structure, which were previously observed by Keim and Stambauch (1986), Fama and French (1988), Ferson and Harvey (1991), and Evans (1994).

In the market-timing models, the models assume that any information that is correlated to the future market return is superior information. The conditional model assumed public macro information is not superior information. The performance resulted from the public information should be separated from the managers' market-timing ability. The beta of the fund is assumed to be a linear function of public information vector z_{t-1} that captures the changing economic conditions:

$$\beta_i(z_{t-1}) = \beta_{im} + \beta_{ic}^T z_{t-1}$$

$$z_{t-1} = Z_{t-1} - E(Z_{t-1}) \quad (2.11)$$

The linear specification on time-varying betas are also used in previous studies, such as Ferson (1985), Shanken (1990), Ferson and Harvey (1993), Cochrane (1996), and Jagannathan and Wang (1996) among others. The linearity assumption is used for fund performance measurement for two reasons: it is motivated by theoretical models of managers' behavior, such as in Admati et al (1986). And, it is easy to interpret as illustrated by Ferson and Schadt (1996). Thus we could modify the Jensen model in (2.1) as,

$$r_{it} - r_{ft} = \alpha_i + \beta_{im}(r_{mt} - r_{ft}) + \beta_{ic}^T z_{t-1}(r_{mt} - r_{ft}) + \varepsilon_{it} \quad (2.12)$$

where β_{im} is the unconditional mean of the conditional beta in (2.11). The elements of β_{ic}^T are the response coefficients with respect to the innovation of the lagged information variables Z_{t-1} .

When incorporating a quadratic market-timing term of (2.4) into the conditional model, the beta of the fund becomes

$$\beta_i(z_{t-1}) = \beta_{im} + \beta_{ic}^T z_{t-1} + \gamma_i^{TM} (r_{mt} - r_{ft}). \quad (2.13)$$

The Jensen measure in (2.1) becomes

$$r_{it} - r_{ft} = \alpha_i + \beta_{im} (r_{mt} - r_{ft}) + \beta_{ic}^T z_{t-1} (r_{mt} - r_{ft}) + \gamma_i^{TM} (r_{mt} - r_{ft})^2 + \varepsilon_{it} \quad (2.14)$$

where β_{ic}^T captures the managers' response to public information Z_{t-1} . The coefficient γ_i^{TM} measures managers' market timing ability. The term $\beta_{ic}^T z_{t-1} (r_{mt} - r_{ft})$ controls the public information effect 1, which could bias fund performance measurement in unconditional market timing measures in section 2.3.

When we apply the conditional measure to the Fama-French three-factor model and the Carhart four-factor model, the models become very complicated, since it consumes too many degrees of freedom that leads to a less efficient estimation of alpha and betas. Although the conditional measure is theoretically justified by Ferson and Schadt (1996), it makes little sense in practice.

All the measures that we reviewed till now are assuming a constant alpha. However, Christopherson, et al. (1998) allow a time-varying alpha in their measure, i.e. they treat alpha as a function of Z_{t-1} ,

$$\alpha_i(z_{t-1}) = \alpha_i + A_i^T z_{t-1} \quad (2.19)$$

where $z_{t-1} = Z_{t-1} - E(Z_{t-1})$, which is the same as we mentioned in (2.10). Allowing a time-varying alpha makes the measures more complicated. It has the same shortcoming as the conditional measure.

2.5 Concluding Remarks

From the above discussion of various measures, we know that traditional measures suffer the following limitations. Firstly, it is difficult to find a proxy for market portfolio (it is called benchmark inefficiency). This difficulty poses a serious problem when evaluating fund performance, because if the market portfolio used is not a perfect market portfolio the covariance of the return of the fund and the return of the market portfolio can not correctly measure the risk born by the fund. Thus the alpha derived from the measure is biased. Later efforts, like the Fama-French three-factor measure and the Carhart four-factor measure, attempted to solve this problem by adding more risk factors into the Jensen measure. Although they could reduce the inefficiency problem to some extent, the inefficiency is still material as noted by Grinblatt and Titman (1994). In addition, the complex multi-factor measures brought two other problems along: it consumes more degrees of freedom, making statistical inference of coefficients unreliable. And it is difficult to interpret the beta coefficients. They provide no quantified information about the fund's asset allocations to each asset category, which is valuable for the in-depth analysis of fund risk level.

Secondly, although market timing and conditional measures are theoretically attractive, it is practically impossible to implement them. When managers invest in options or option-like securities, spurious market-timing ability and selectivity ability may be observed. In addition, when managers trade securities in less than one month, which is common in practice, we could also observe spurious market-

timing ability. The correct separation of market-timing ability from selection ability, denoted by alpha, depends on some impractical constrictions. Regarding conditional measures, the measures are complicated in multi-factor models, making the inference about beta coefficients and alpha unreliable within a three-year evaluation period. And we can not increase the sample size to deal with this problem, because the fund may significantly shift its investment strategy or change fund managers in the longer sample period. It is a common practice to use three-year data to evaluate fund performance, see, for example, Cai et al. (1997), Carhart (1997), Elton et al. (1996), and Kosowski et al. (2001).

Thirdly, all these measures are estimated by the regression method. An underlying assumption is that ε_{it} is normally distributed in order to make hypothesis tests on betas and alpha. But many empirical studies have shown this assumption is not likely true, for example, a recent study by Kosowski et al. (2001), where they used bootstrap analysis to assess the p value of alphas.

In the next chapter, we show that the measure proposed by Sharpe, which is derived from return-based style analysis, is superior to other measures. We term it as the RBSA measure. The measure has several advantages compared to traditional measures estimated by the regression method. Firstly, we do not require that ε_{it} should be normally distributed. ε_{it} can be distributed differently. In addition, the expected value of ε_{it} is not even required to be zero. We interpret the non-zero value of ε_{it} as the management effect, caused by securities selection or asset

rotations. The expected value of ε_{it} is the measure of fund performance, a counterpart of the alpha of traditional measures in this chapter.

Secondly, we circumvent the benchmark inefficiency problem by including all the investable style indexes in the RBSA measure. The only requirements about the style indexes are that they are exhaustive and exclusive of each other. These requirements are easily accommodated by a large amount of indexes publicly available in the market.

Thirdly, the betas estimated in the RBSA measure provide useful information about fund styles. Fund styles are essential for the decomposed-analysis of the fund's risk level by institutional investors.

We illustrate the RBSA measure in the next chapter, and then we test the robustness, accuracy, and efficiency of the measure, and compare the measure with traditional measures by a comparative simulation experiment in Chapter IV.

Chapter III Performance Measure Based on Return-based Style Analysis

3.1 Introduction

As reported in Chapter II, traditional measures suffer from some limitations, which may lead to totally different inferential results from the same dataset using different measures or benchmarks. So we have seen a hot debate about whether fund managers can deliver abnormal returns over the decades for example Ippolito (1989) and Elton, et al. (1993). As the measures become more refined and sophisticated since the original paper by Jensen (1968), the question is still open for further investigation.

In this chapter, we show that the measure proposed by Sharpe (1992), which is derived from the return-based style analysis, is superior to other measures by a comparative simulation study. This measure has several advantages over the measures based on regression that we reviewed in Chapter II. First, it does not rely on the normal assumption of residuals in the model. In the RBSA measure, we formulate the performance measurement issue as a convex quadratic programming problem. Under this framework we do not need any assumption about variables' distributions in the model. This method is different from traditional measures

where we assume that ε is normally distributed in the regression model. Second, we include all the investable indexes in the model. This way we circumvent the benchmark inefficiency problems we encountered in Chapter II. But to achieve this, we require that the indexes be exhaustive and exclusive, which are not difficult to accommodate with many indexes developed by commercial companies. Finally, we could estimate fund styles with this measure. Fund styles provide more detailed information about fund risk than a simple beta value in the Jensen measure in (2.1).

The measure differs from traditional measures at the underlying rationale. Traditional measures use the least square method to estimate the alpha and betas under the framework of a regression model. The objective is to find a set of risk exposures to minimize the sum of squares of ε_i . On the other hand, the RBSA measure exploits the interior point mathematical algorithm to estimate the style exposures under the framework of a convex quadratic programming problem. The objective is to find a set of style index exposures that minimize the variance of ε_i , not the absolute value of ε_i in the regression models. In other words, the RBSA measure attempts to identify a set of style index exposures that best mimics the return behavior of the fund.

We first introduce the return-based style analysis in section 2. In section 3 we extend it to the RBSA measure and formulate the problem as a convex quadratic programming problem.

3.2 Return-based Style Analysis

Sharpe (1992) proposed to identify the fund styles solely from fund the returns instead of analyzing the actual asset holdings of the fund (fundamental analysis). It attracted a lot of attentions since this pioneering work, please see, for example, Buetow, et al. (2000), Christopherson (1995, 1999), Cummisford, et al. (1996), Lieberman (1996), and Mayes, et al. (2000). The model is,

$$r_t = \beta_1 f_{1t} + \beta_2 f_{2t} + \dots + \beta_k f_{kt} + \varepsilon_t \quad (3.1)$$

where r_t is the fund return from period 1 to T. T is the number of observations in the sample period. f_{kt} is the k^{th} index return in period t. f_{1t} to f_{kt} are called style indexes. The style of the fund in our analysis is identified by the coefficients of the style indexes (β_i), which are defined as style exposures. For example, the small-cap growth fund is expected to have a relatively large style coefficient on the small-cap growth style index. The model implicitly assumes that the style of the fund is time-invariant in the period. Therefore, it implies that the estimated style exposures are the average style during the evaluation period if the fund changes its factor loadings of assets.

To make the model estimation efficient, the selection of style indexes is crucial. First, the set of style indexes should be complete, i.e., a complete coverage of all the investable assets. An incomplete coverage may cause a low R^2 , biased estimation of style exposures, and biased fund performance measurement. The bias in this circumstance is similar to the bias of benchmark inefficiency in Chapter II.

To illustrate the consequences of the incomplete style indexes coverage, we provide an example. Suppose that a fund places a portion of assets in foreign stocks and the foreign stocks outperform the domestic US stocks, then ignoring the style index of foreign stocks may cause the performance measurement of the fund to be upwardly biased and style exposures to US stocks spuriously higher. Second, the style indexes must be mutually exclusive. Overlapped indexes may cause the style exposure estimates to become unstable. Since the style analysis is to identify the best combination of style indexes that mimics the return behavior of the fund, overlapped indexes may cause identification difficulty. Under the same rationale, in addition to the requirement of mutually exclusive, we hope that the indexes are different, i.e. they have low correlations. In case they have high correlations, the standard deviations are different. Therefore, ideally the style indexes should be complete, mutually exclusive and different.

The return-based style analysis provides attractive results. It provides an insight of the fund's style. Before the return-based style analysis was introduced, people were using fundamental analysis. It first collects the information of portfolio holdings in each period, then analyzes the securities' styles in the portfolio one by one, and finally groups the securities according to their styles to get the fund's asset allocations. Although fundamental analysis could give us useful information about fund styles, the releases of the fund's portfolio holdings usually have an unwanted time-lag, and for some funds this information is not publicly available. Looking further, even though we have the necessary information, it is costly to conduct

fundamental style analysis to analyze the security holdings in the fund portfolio one by one. Return-based style analysis provides an alternative that is inexpensive and timely. It only needs the fund's return and the returns of indexes that are readily available at monthly or even daily frequency from newspaper or internet resources. Instead of conducting in-depth analysis of the portfolio holdings, it obtains fund style exposures by studying return behavior.

Even though fundamental analysis could be implemented easily and timely, the return-based style analysis still has its value, since it focuses on the return behavior, thus it is not confined by the actual styles of the security holdings in the portfolio. The style estimates more accurately reflect the actual fund style. Suppose that a fund invests in large US companies that have large exposure to foreign economies, so to some extent the fund would behave like foreign stock funds. A rough fundamental analysis may just show that the fund is just a large-cap fund, ignoring the characteristic of the fund with a risk level and expected return similar to an international fund. But return-based style analysis may show the fund has some exposure to foreign stocks. The positive style exposure to foreign stocks is not a limitation of return-based style analysis, on the contrary an advantage that reflects the true behavior of the fund. Therefore in this sense return-based style analysis is superior to fundamental analysis.

3.3 RBSA Measure by Quadratic Programming

Return-based style analysis can be naturally extended to measure fund performance. We term it the RBSA measure. It decomposes the return in (3.1) into two parts. One is, $\beta_1 f_{1t} + \beta_2 f_{2t} + \dots + \beta_k f_{kt}$, attributable to fund styles; the other is attributable to ε_t , due to the active management like securities selection and asset allocation. It is defined as the tracking error at period t .

The expected value of the tracking error, $E(\varepsilon_t)$, is defined as the performance of the fund, Alpha. It is the difference between the realized fund return and the return of passive style indexes. Please notice that the ε_t is out-of-sample tracking error, which is different from the in-sample tracking error. The detailed calculation steps below may help understand the difference:

1. Estimate the fund styles over the period from $t-36$ to $t-1$. (We use the past three-year's monthly fund data to estimate the fund style in the current period, t .)
2. Calculate the return of the passive style indexes with weights (style exposures) estimated in step 1 at period t . This is actually the expected fund return based on the style exposure estimates.
3. Calculate the difference between the fund return at period t and the return of step 2. The difference is the performance of fund at period t .
4. Move one period forward and repeat step 1 to step 3 until the end of evaluation period, period $t+35$.

5. Average and annualize the differences to get the alpha of the RBSA measure.

3.3.1 Rationale of RBSA Measure

Quadratic programming is employed to estimate the style exposures and then the alpha of the RBSA measure, whereas the regression analysis is employed to estimate the alpha of the Jensen measure and other measures which are based on CAPM. The difference of two estimation methods arises from the underlying rationale of two kinds of measures. Traditional measures, such as the Jensen measure, are trying to estimate a line or curve that best fits the fund's return series, so they are minimizing the sum of squares of ε_t . The RBSA measure attempts to estimate a curve that moves mostly like the fund's return, so they are minimizing the variance of ε_t . ε_t is explicitly expressed as the management effect in the RBSA measure. The expected value of ε_t can be positive or negative, where a positive expected value of ε_t indicates superior ability of managers and a negative value of ε_t indicates an inferior ability compared with passive index fund investment strategy. The difference between these two methods is illustrated in figure 1.

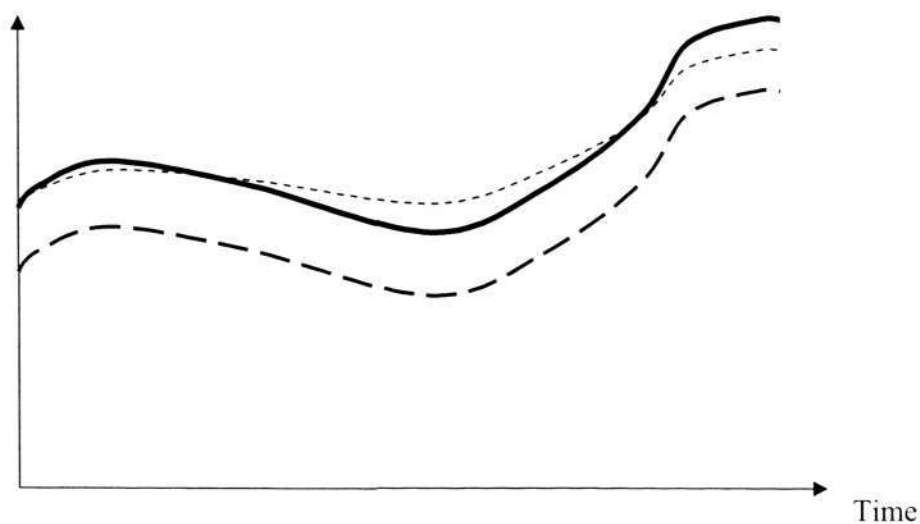


Figure 1: Rationale of the measure

In figure 1 the solid curve represents the actual return series. The short-dashed curve is the estimation result from regression-based measures, which is the best fit of actual return series. The long-dashed curve is from the RBSA measure, a curve that mimics the behavior of actual return series. We observe that the long-dashed curve lies nearly parallel to the actual return series. If they are parallel, the variance of the tracking error is zero, so the alpha is just the difference of the two parallel curves. The positive difference implies that the fund performed better than the passive index fund investment strategy. In the RBSA measure, we mimic the return behavior of the fund, so the estimated curve is not necessarily close to the actual return series. The average of vertical differences between the two curves in the figure is interpreted as the tracking error caused by active management.

3.3.2 Formulation of RBSA Measure in Quadratic

Programming

We may express the rationale of the RBSA measure discussed in section 3.3.1 as an optimization problem, that is,

$$\begin{aligned} \text{Min} \quad & \text{Var}(\varepsilon_t) = \text{Var}(r_t - \beta^T f_t) \\ \text{s.t.} \quad & \beta^T e = 1 \\ & \beta \geq 0 \end{aligned} \tag{3.2}$$

where the notation Min stands for minimization, s.t. stands for subject to, and superscript T of β stands for transpose. ε_t is the tracking error due to the management effect, β is a column vector of style exposures corresponding to style indexes, e is a column vector of ones. r_t is the fund return at period t, and f_t is a column vector of the returns of k style indexes at period t. When minimizing the variance, we impose a portfolio constraint i.e. the sum of the elements of β is one, and non-negativity constraints on β to make estimation results consistent with the actual policy of mutual fund, that is, to invest in securities without short positions. When studying hedge funds where the fund can take long-short position, the non-negativity constraints can be relaxed.

The above minimization problem is actually a convex quadratic programming problem. We rewrite the objective function as,

$$\begin{aligned} \text{Var}(r_t - \beta^T f_t) &= E(r_t - \beta^T f_t)^2 - (E(r_t - \beta^T f_t))^2 \\ &= \frac{1}{N} \sum_{t=1}^N (r_t - \beta^T f_t)^2 - \left[\frac{1}{N} \sum_{t=1}^N (r_t - \beta^T f_t) \right]^2 \end{aligned} \tag{3.3}$$

Let us define R as a column vector with N observations of fund returns, F as return matrix of k style indexes, and e as a column vector of ones to write equation (3.3) as,

$$\begin{aligned} \text{Var}(r_i - \beta^T f_i) &= \frac{1}{N} \|R - F\beta\|^2 - \left(\frac{e^T (R - F\beta)}{N}\right)^2 \\ &= \beta^T \left(\frac{1}{N} F^T F - \frac{1}{N^2} F^T e e^T F\right) \beta - 2\left(\frac{1}{N} R^T F - \frac{1}{N^2} e^T R e^T F\right) \beta + \left[\frac{1}{N} \|R\|^2 - \frac{1}{N^2} (e^T R)^2\right] \\ &= \frac{1}{2} \beta^T \left[2\left(\frac{1}{N} F^T F - \frac{1}{N^2} F^T e e^T F\right)\right] \beta + 2\left(\frac{1}{N^2} e^T R e^T F - \frac{1}{N} R^T F\right) \beta + \left[\frac{1}{N} \|R\|^2 - \frac{1}{N^2} (e^T R)^2\right] \end{aligned} \quad (3.4)$$

Since $\frac{1}{N} \|R\|^2 - \frac{1}{N^2} (e^T R)^2$ is a constant, (3.2) is equivalent to,

$$\text{Min} \quad \frac{1}{2} \beta^T H \beta + g^T \beta \quad (3.5)$$

$$\text{s.t.} \quad \begin{aligned} \beta^T e &= 1 \\ \beta &\geq 0 \end{aligned}$$

where H and g^T are equal to $2\left(\frac{1}{N} F^T F - \frac{1}{N^2} F^T e e^T F\right)$

and $2\left(\frac{1}{N^2} e^T R e^T F - \frac{1}{N} R^T F\right)$ in (3.4) respectively. From (3.5), we see that it is actually a typical quadratic programming problem.

The estimates of the style exposures in our quadratic programming problem are not local optimal solutions, but global optimal solutions. Let us first look at the objective function in (3.5). We take the second order derivative of $\frac{1}{2} \beta^T H \beta$ with respect to β , the unknown coefficients. The result is just H , which is equal to

$2[\frac{1}{N}F^T(I-ee^T)F]$, in which $I-ee^T$ is a positive semi-definite symmetric matrix with eigenvalues 0 and 1, so H is also a positive semi-definite symmetric matrix. Therefore, we know $\frac{1}{2}\beta^T H\beta$ is convex. In addition, $g^T\beta$ is a linear function of β , so the second term of the objective function in (3.5) is also convex. Because the sum of two convex functions is also a convex function, we know the objective function is convex. Further, we know the feasible region in our minimization problem is defined by a set of linear constraint, $\beta^T e=1$ and $\beta \geq 0$. Thus the feasible set is a convex set. When the objective function is a convex function under a convex set, the local minimum is also the global minimum. Thus the problem is just a convex quadratic programming problem.

3.3.3 Optimality Conditions of Quadratic Programming

In order to find out the optimality conditions, let us introduce the Lagrangian multiplier into (3.5), we have

$$\text{Min} \quad \frac{1}{2}\beta^T H\beta + g^T\beta + \lambda(1-\beta^T e) \quad (3.6)$$

$$\text{s.t.} \quad \begin{aligned} \beta^T e &= 1 \\ \beta &\geq 0 \end{aligned}$$

where λ is the Lagrangian multiplier. If there are no non-negativity constraints, the problem is easy to solve. Having the non-negativity constraints, the optimal solution must satisfy the Kuhn-Tucker conditions (Chiang, 1984) below,

$$\begin{aligned}
 H\beta + g - \lambda e &\geq 0 \\
 B(H\beta + g - \lambda e) &= 0 \\
 \beta^T e &= 1 \\
 \beta &\geq 0
 \end{aligned} \tag{3.7}$$

where we define the B as a diagonal matrix with β in its diagonal. We introduce a dummy vector s into (3.7), which is non-negative and has the same dimension as β . Thus equation (3.7) is transformed into,

$$\begin{aligned}
 \lambda e - H\beta - g + s &= 0 \\
 Bs &= 0 \\
 \beta^T e &= 1 \\
 \beta, s &\geq 0
 \end{aligned} \tag{3.8}$$

which is the optimality conditions for our convex quadratic programming problem.

We express the optimality condition in matrix form as,

$$F(\beta, s, \lambda) = \begin{bmatrix} \lambda e - H\beta - g + s \\ Bs \\ 1 - \beta^T e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.9}$$

$$\beta, s \geq 0$$

In a manner similar to the linear programming, optimality conditions in (3.9) can be seen as a collection of conditions:

1. Primal feasibility: $\beta^T e = 1$ and $\beta \geq 0$.
2. Dual feasibility: $\lambda e - H\beta - g + s = 0$, and $s \geq 0$.
3. Complementary slackness: $Bs = 0$.

3.3.4 Interior Point Method

Using the above interpretation of optimality conditions, we may modify the simplex method to solve the convex quadratic programming problem. For most of situations, the performance of the simplex algorithm is satisfactory, but the simplex algorithm has a disadvantage, that is, the amount of time required to solve a problem could grow exponentially in respect to the size of the problem (the number of variables and constraints in the problem). This is the so-called worst-case complexity. To tackle this problem, polynomial-time algorithms were developed, where the running time can be bounded by a polynomial function of the input size. Yudin and Nemirovski (1976) developed the ellipsoid method. The more important development is by Karmarkar (1984), who introduced the interior-point method. In addition to being a polynomial-time algorithm, it is faster than the simplex algorithm at solving some problems. Nesterov and Nemirovski (1994) showed that the interior-point method can be applied to a large class of problems, such as, a convex quadratic programming problem.

We use the interior-point method to estimate β . The interior-point method can be viewed as being based on the Newton method, but is modified to accommodate the inequality constraints. In (3.9), if there is no $Bs=0$, we can solve it using the Gaussian elimination method. If there is no inequality constraints, to find (β, s, λ) that satisfies the optimality condition is a straightforward application of Newton's method. To solve our problem, we firstly identify a point $(\beta^0, s^0, \lambda^0)$ that is in the interior of the feasible region, i.e.:

$$F^0 = \{(\beta, s, \lambda) : \lambda e - H\beta - g + s = 0, Bs = 0, \beta^T e = 1, \beta, s > 0\}. \quad (3.10)$$

This is the same as the optimality conditions except for no binding conditions of β and s . Then we try to generate $(\beta^k, s^k, \lambda^k)$ that is progressively closer to the boundary. This is achieved by an application of modified Newton's method (Nesterov and Nemirovski, 1994) discussed in Appendix A.

3.4 Concluding Remarks

In this chapter we show the rationale and the way to implement it. We find that the performance measurement problem can be formulated as a convex quadratic programming problem. The rationale of the RBSA measure is trying to estimate a set of style index exposures that mimic the behavior of fund return behavior, so the fitted curve is not necessarily close to the actual return series as shown in figure 1.

In the next chapter, we further investigate the robustness of the RBSA measure in different situations, where we simulate four general kinds of funds, that is, large-cap funds, small-cap funds, well-diversified funds with a preference to large-cap stocks, and well-diversified funds with a preference to small-cap stocks. We also compare the RBSA measure with traditional measures that we reviewed in Chapter II.

Chapter IV A Comparative Simulation

Experiment of Performance Measures

4.1 Introduction

In this chapter we conduct a comparative simulation experiment to test the robustness, accuracy and efficiency of the RBSA measure, and compare the measure with traditional measures. We present the setup of the experiment in section 2, then we show the simulation results of alpha, betas and R^2 in section 3, finally we summarize our findings in section 4.

4.2 Setup of Simulation Experiment

The fund returns are simulated from,

$$r_t = a + \beta_1 R_{1t} + \beta_2 R_{2t} + \beta_3 R_{3t} + \beta_4 R_{4t} + \beta_5 R_{5t} + \varepsilon_t \quad (4.1)$$

where a is set at 5% annually. It is possible to change the value of a in the simulation, but the results (not reported) show that the selection of a does not change our conclusions about the accuracy and efficiency of the measures. In (4.1) R_{1t} , R_{2t} , R_{3t} , R_{4t} and R_{5t} stand for three-month Treasury bill rates, Russell Top 200 Growth Index, Russell Top 200 Value Index, Russell 2000 Growth Index, and

Russell 2000 Value Index¹ respectively. These five indexes represent the fund's asset allocation to currency asset, large-cap growth stocks, large-cap value stocks, small-cap growth stocks, and small-cap value stocks. ε_t is a randomly generated residual with a mean of zero and standard deviation calculated from the actual style analysis of more than 1000 US domestic well-diversified equity mutual funds, following normal distribution.

To test the measures' ability to measure fund performance and its styles in different situations, we use four sets of beta coefficients below,

$$\begin{bmatrix} 0.05 & 0.48 & 0.47 & 0 & 0 \\ 0.05 & 0 & 0 & 0.48 & 0.47 \\ 0.05 & 0.35 & 0.35 & 0.13 & 0.12 \\ 0.05 & 0.13 & 0.12 & 0.35 & 0.35 \end{bmatrix} \quad (4.2)$$

The four sets of beta coefficients are to mimic the fund return behavior of four general types of funds: large-cap funds, small-cap funds, well-diversified funds with a preference to large-cap stocks, and well-diversified funds with a preference to small-cap stocks. For example, the first set of beta coefficients, [0.05 0.48 0.47 0 0], means that the simulated funds put 5% of assets in treasury bills, 48% of assets in well-diversified large-cap growth stocks, 47% of assets in well-diversified value stocks, and no assets in small-cap stocks.

¹ The definitions of the indexes are available at <http://www.russell.com/US/Indexes/US-Definitions.asp>.

With the simulated return series of the fund, we are testing the power of the following performance measures that we reviewed and proposed in Chapter II and Chapter III:

1. RBSA measure that is formulated under the framework of a convex quadratic programming problem (RBSA):

$$r_t = \beta_1 f_{1t} + \beta_2 f_{2t} + \dots + \beta_k f_{kt} + \varepsilon_t$$

subject to $\beta' e = 1$ and $\beta \geq 0$

In the simulation setup, the alpha of RBSA measure is simplified as the expected value of the in-sample ε_t .

2. Jensen measure (JS):

$$r_t - r_{ft} = \alpha + \beta_m (r_{mt} - r_{ft}) + \varepsilon_t$$

3. Jensen measure with Treynor-Mazuy market-timing adjustment (JS-TM):

$$r_t - r_{ft} = \alpha + \beta_m (r_{mt} - r_{ft}) + \gamma^{TM} (r_{mt} - r_{ft})^2 + \varepsilon_t$$

4. Jensen measure with Henriksson-Merton market-timing adjustment (JS-HM):

$$r_t - r_{ft} = \alpha + \beta_m (r_{mt} - r_{ft}) + \gamma^{HM} \text{MAX}(0, r_{mt} - r_{ft}) + \varepsilon_t$$

5. Fama-French three-factor measure (FF3):

$$r_t - r_{ft} = \alpha + \beta_m (r_{mt} - r_{ft}) + \beta_{SMB} r_{SMB,t} + \beta_{HML} r_{HML,t} + \varepsilon_t$$

6. Fama-French three-factor measure with Treynor-Mazuy market-timing adjustment (FF3-TM):

$$r_t - r_{ft} = \alpha + \beta_m (r_{mt} - r_{ft}) + \beta_{SMB} r_{SMB,t} + \beta_{HML} r_{HML,t} + \gamma^{TM} (r_{mt} - r_{ft})^2 + \varepsilon_t$$

7. Fama-French three-factor measure with Henriksson-Merton market-timing adjustment (FF3-HM):

$$r_t - r_{ft} = \alpha + \beta_m (r_{mt} - r_{ft}) + \beta_{SMB} r_{SMB,t} + \beta_{HML} r_{HML,t} + \gamma^{HM} \text{MAX}(0, r_{mt} - r_{ft}) + \varepsilon_t$$

where r_t and betas are the same as defined in Chapter II. The risk-free rate r_{ft} is three-month Treasury Bill Rates. The market portfolio r_{mt} is S&P 500, the most frequently used proxy for market portfolio. γ^{TM} and γ^{HM} are market-timing coefficients measured by Treynor-Mazuy method and Henriksson-Merton method respectively. In Fama-French three-factor models, $r_{SMB,t}$ and $r_{HML,t}$ are used to control investment strategies due to size effect and B/M ratio respectively, where $r_{SMB,t}$ is the difference of returns between the monthly return of Russell 1000 index and Russell 2000 index, and $r_{HML,t}$ is the difference of returns between the monthly return of Russell 3000 Value Index and Russell Growth Index.

4.3 Simulation Results and Analysis

4.3.1 Simulation Results and Analysis of Alpha and R^2

Table 1 shows simulation results of alpha and R^2 from seven measures, based on 1000 simulations of randomly generated fund return series under four sets of style coefficients in (4.2). They are presented in table 1 from panel 1 to panel 4. The alpha and R^2 are the average values of the estimation from 1000 simulations. The bias is reported as the difference between the estimated alphas from the measures and the true alpha, which is fixed at 5% in the simulation. To show the efficiency of the performance measurements, we also report the empirical confidence interval at 95% from the simulations. The lower bound is the 5th percentile of the 1000

estimated alphas and the upper bound is the 95th percentile of the 1000 estimated alphas. Because the index return series is possibly not normal due to the cross correlations among stocks in the index portfolios (Kosowski et al., 2001), we construct the confidence intervals from simulation instead of constructing them from t values.

Table 1: Simulation I (Alpha and R²)

Panel I (β 1=0.05, β 2=0.48, β 3=0.47, β 4=0, β 5=0)					
Measures	Alpha	Bias	C. I.	Size	R ²
RBSA	4.76	-0.24	[2.05 7.59]	5.51	0.96
JS	1.13	-3.87	[-1.46 3.75]	5.21	0.94
JS-TM	0.51	-4.49	[-2.83 3.63]	6.46	0.94
JS-HM	-0.32	-5.32	[-4.43 3.68]	8.11	0.94
FF3	2.68	-2.32	[-4.03 9.13]	13.16	0.95
FF3-TM	3.01	-1.99	[-4.20 10.16]	14.36	0.95
FF3-HM	2.62	-2.38	[-5.15 10.41]	15.56	0.95
Panel II (β 1=0.05, β 2=0, β 3=0, β 4=0.48, β 5=0.47)					
Measures	Alpha	Bias	C. I.	Size	R ²
RBSA	5.16	0.16	[2.45 7.82]	5.37	0.97
JS	13.95	8.95	[11.51 16.39]	4.88	0.52
JS-TM	24.03	19.03	[20.89 27.23]	6.34	0.55
JS-HM	27.48	22.48	[23.32 31.87]	8.55	0.54
FF3	1.89	-3.11	[-4.71 9.02]	13.73	0.97
FF3-TM	2.19	-2.81	[-4.87 8.91]	13.78	0.97
FF3-HM	2.21	-2.79	[-5.44 9.54]	14.98	0.97
Panel III (β 1=0.05, β 2=0.35, β 3=0.35, β 4=0.13, β 5=0.12)					
Measures	Alpha	Bias	C. I.	Size	R ²
RBSA	5.1	0.1	[2.31 8.03]	5.72	0.96
JS	4.71	-0.29	[2.04 7.29]	5.25	0.92
JS-TM	6.75	1.75	[3.42 9.98]	6.56	0.92
JS-HM	7.03	2.03	[2.72 11.14]	8.42	0.92
FF3	2.53	-2.47	[-4.17 9.01]	13.18	0.95
FF3-TM	2.75	-2.25	[-4.33 9.75]	14.08	0.95
FF3-HM	2.49	-2.51	[-4.99 10.22]	15.21	0.95
Panel IV (β 1=0.05, β 2=0.13, β 3=0.12, β 4=0.35, β 5=0.35)					
Measures	Alpha	Bias	C. I.	Size	R ²
RBSA	5.02	0.02	[2.31 7.81]	5.5	0.97
JS	10.64	5.64	[8.15 13.13]	4.98	0.67
JS-TM	17.86	12.86	[14.63 21.27]	6.64	0.69
JS-HM	20.09	15.09	[15.67 24.44]	8.77	0.68
FF3	1.87	-3.13	[-4.53 9.06]	13.59	0.96
FF3-TM	2.25	-2.75	[-4.74 9.06]	13.8	0.96
FF3-HM	2.14	-2.86	[-5.21 9.62]	14.81	0.96

Table 1(continued)

Panel V Summary

	RBSA	JS	JS-TM	JS-HM	FF3	FF3-TM	FF3-HM
<u>Alpha</u>							
Panel 1	4.76	1.13	0.51	-0.32	2.68	3.01	2.62
Panel 2	5.16	13.95	24.03	27.48	1.89	2.19	2.21
Panel 3	5.10	4.71	6.75	7.03	2.53	2.75	2.49
Panel 4	5.02	10.64	17.86	20.09	1.87	2.25	2.14
Average	5.01	7.61	12.29	13.57	2.25	2.55	2.36
Bias	0.01	2.61	7.29	8.57	-2.75	-2.45	-2.64
Std.	0.15	4.99	9.20	10.86	0.37	0.34	0.20
<u>R²</u>							
Panel 1	0.96	0.94	0.94	0.94	0.95	0.95	0.95
Panel 2	0.97	0.52	0.55	0.54	0.97	0.97	0.97
Panel 3	0.96	0.92	0.92	0.92	0.95	0.95	0.95
Panel 4	0.97	0.67	0.69	0.68	0.96	0.96	0.96
Average	0.97	0.76	0.78	0.77	0.96	0.96	0.96
<u>Size of C. I.</u>							
Panel 1	5.54	5.21	6.46	8.11	13.16	14.36	15.56
Panel 2	5.37	4.88	6.34	8.55	13.73	13.78	14.98
Panel 3	5.72	5.25	6.56	8.42	13.18	14.08	15.21
Panel 4	5.50	4.98	6.64	8.77	13.59	13.80	14.81
Average	5.53	5.08	6.50	8.46	13.42	14.01	15.14

The table provides simulation results of alpha and R^2 under four sets of beta coefficients presented in (1.2). Fund returns are simulated from an alpha, fixed at 5%, a random error, and five style indexes, that is, three-month Treasury bill rates, Russell Top 200 Growth Index, Russell Top 200 Value Index, Russell 2000 Growth Index, and Russell 2000 Value Index. Beta coefficients correspond to the proportions of assets allocated to Treasury bill and four style indexes. We simulate four types of funds, that is, large-cap funds, small-cap funds, well-diversified funds with a preference of large-cap stocks, and well-diversified funds with a preference of small cap stocks. RBSA stands for RBSA measure by quadratic programming; JS stands for Jensen measure; JS-TM stands for Jensen measure with Treynor-Mazuy market-timing adjustment; JS-HM stands for Jensen measure with Henriksson-Merton market-timing adjustment; FF3 stands for Fama-French three-factor measure; FF3-TM stands for Fama-French three-factor measure with Treynor-Mazuy market-timing adjustment; FF3-HM stands for Fama-French three-factor measure with Henriksson-Merton market-timing adjustment.

C. I. is the empirical confidence interval of alpha estimator based on simulations.

Size is the length of C. I.

Panel I of table 1 shows alpha estimates of the simulated fund with style coefficients $[0.05 \ 0.48 \ 0.47 \ 0 \ 0]$, meaning 5% of fund asset is allocated to currency asset, 48% to well-diversified large-cap growth stocks, 47% to large-cap value stocks, and no asset is allocated to small stocks. We find that RBSA is the most accurate measure with the bias only -0.24% annually. The other measures' accuracy is not comparable to that of the RBSA measure. The biases are larger than 1% as shown in the panel. Using the first set of betas, the three Jensen-based measures, that is, JS, JS-TM, and JS-HM, are less accurate than three FF3-based measures, that is, FF3, FF3-TM, and FF3-HM. The average bias of three JS-based measures is about two times larger than the average bias of three FF3-based measures.

After adjusting market-timing behavior, which actually does not exist in our simulation, with methods suggested by Treynor-Mazuy and Henriksson-Merton, the biases are even larger, except for FF3-TM. Since there is no market-timing in the simulation, we should not observe any change of biases after adding a market-timing term if the market-timing models are solid. We observed spurious market-timing in the simulation. The spurious market timing is also found empirically by Cai, et al. (1997), Glosten and Jagannathan (1994), and Jagannathan and Korajczyk (1986).

The size of confidence interval indicates the efficiency of measures. JS measure has the smallest size, however since the alphas are severely biased, the efficiency

gain has no meaning. The size of RBSA is similar to JS but less biased. The size of confidence interval is largest for FF3-based measures, which are around two times of the size of JS-based measures. This wider confidence interval of FF3 measures is mainly caused by using more variables at the right side of the regression. This kind of correlation may cause inaccurate estimation of alphas in FF3 measures.

We also notice that the R^2 is highest for RBSA measure whose average is 96%. FF3-based measures show a little higher R^2 than JS-based measures. Therefore, using the first set of betas that mimics a large-cap fund we find RBSA measure is less biased and has the largest explanatory power and efficiency.

Panel II shows results using another set of betas. The simulated fund behaves like a small-cap fund according to style coefficients that we set in simulation. The magnitude of the bias of RBSA is similar to what we observed in panel I, but now is upwardly biased. And again RBSA has the smallest bias. But now we observe that bias of JS-based measures is much larger and R^2 is quite low, ranging from 52% to 55%. This is because we are using S&P 500 as the market benchmark, in which most of the stocks are large-cap stocks. This bias clearly illustrates the incapability of JS-based measures in measuring performance when funds invest small-cap securities. FF3-based measures are using the same market benchmark as JS-based measures, but the biases are much smaller, which is due to the explicit incorporation of two risk factors related to size effect and the B/M ratio. We also observe the explanation power of FF3 is comparable to that of the RBSA measure.

Therefore, when a fund is a small-cap fund, JS-based measures are not capable of estimating the true alpha. FF3-based measures are more robust than JS-based measures, because they explicitly consider the size effect in the model. RBSA is still the best measure in this case with high R^2 , small bias and efficient estimation.

In panel III, we randomly generate a fund that widely invests in all the stocks in the market, but leans to large-cap stocks. We notice that the bias of RBSA is 0.1, but JS-based measures also have small biases when evaluating this kind of fund. The average is -0.29. The bias, efficiency and R^2 of FF3-based measures are similar to what we observed in panel I and panel II. In this set of style coefficients, JS-based measures are comparable to RBSA in terms of bias and efficiency but RBSA is more powerful to explain the fund's return behavior with the highest R^2 , 0.96.

In panel IV we generate a fund that widely invests in all the stocks in US market, but leans to small stocks with 70% of assets allocated to small stocks. We find RBSA is very accurate with only a 0.02% bias. The magnitude of bias and R^2 for FF3 measures is stable through the four situations. Regarding JS measures, in panel IV we again observe large bias and low explanation power ranging from 66% to 68%, as we observed in panel II.

From the summary panel of table 1, we find that RBSA unanimously has small biases with an average bias of 0.01% annually, high R^2 accounting for 97% of return variation, and small size of confidence intervals, through the four situations

in table 1. FF3-based measures have high R^2 , stable biases, and stable size of confidence intervals, but the average bias is around 2.5%, which is much larger than the average bias of RBSA. JS-based measures have the largest biases and the biases are volatile depending on the type of the simulated fund. Although the size of the confidence intervals of JS-based measures is relatively small, the biases and variation of estimated alphas make the efficiency not meaningful. Adjusting market timing for JS and FF3 only makes the estimation less efficient, and causes biases larger in JS-based measures. Therefore, from simulation results we may say the RBSA is a better measure in measuring fund performance and explaining the fund return variation compared to other traditional measures.

4.3.2 Simulation Results and Analysis of Style Coefficients (betas)

Table 2 presents simulation results of style coefficients (betas) in four situations. To test the robustness, accuracy, and efficiency of the seven measures in estimating style coefficients, we simulate four types of funds, that is, large-cap funds, small-cap funds, well-diversified funds with a preference of large-cap stocks, and well-diversified funds with a preference of small-cap stocks. The estimates of betas in the table are average betas of 1000 simulations, and the empirical confidence interval is obtained by setting the 5th percentile of the estimates as the lower bound and 95th percentile as the upper bound.

Panel I shows the estimation results when the simulated fund behaves like a large-cap fund. Our estimates of betas using RBSA are very close to the actual betas. The non-negativity constraints of betas may cause a small upward bias when betas are actually zeros and a small downward bias for other positive betas with the same magnitude. When we use traditional measures: JS-based measures and FF3-based measures, we find that the betas of the market benchmark are uniformly above 0.9. Considering the actual asset

Table 2: Simulation II (Style Coefficients)

Panel I ($\beta_1=0.05, \beta_2=0.48, \beta_3=0.47, \beta_4=0, \beta_5=0$)

Measures	$\beta_1=0.05$	$\beta_2=0.48$	$\beta_3=0.47$	$\beta_4=0$	$\beta_5=0$
RBSA	0.05	0.47	0.47	0.01	0.01
	β_m	β_{smb}	β_{hml}	β_{tm}	β_{hm}
JS	0.96				
JS-TM	0.96			0.28 [-0.47 1.07]	
JS-HM	0.92				0.07 [-0.08 0.21]
FF3	0.96	-0.09 [-0.14 -0.04]	-0.01 [-0.11 0.1]		
FF3-TM	0.96	-0.09 [-0.15 -0.04]	-0.01 [-0.11 0.09]	-0.05 [-0.83 0.75]	
FF3-HM	0.96	-0.09 [-0.14 -0.04]	-0.01 [-0.11 0.09]		0.01 [-0.15 0.16368]

Panel II ($\beta_1=0.05, \beta_2=0, \beta_3=0, \beta_4=0.48, \beta_5=0.47$)

Measures	$\beta_1=0.05$	$\beta_2=0$	$\beta_3=0$	$\beta_4=0.48$	$\beta_5=0.47$
RBSA	0.04	0.01	0.01	0.48	0.46
	β_m	β_{smb}	β_{hml}	β_{tm}	β_{hm}
JS	0.92				
JS-TM	0.87			-3.8 [-4.57 -3.02]	
JS-HM	1.19				-0.6 [-0.75 -0.46]
FF3	0.94	0.95 [0.9 1]	0.04 [-0.07 0.14]		
FF3-TM	0.94	0.95 [0.9 1]	0.04 [-0.07 0.14]	-0.1 [-0.9 0.7]	
FF3-HM	0.95	0.95 [0.9 1.01]	0.03 [-0.07 0.13]		-0.01 [-0.17 0.15]

Table 2 (continued)

Panel III ($\beta_1=0.05, \beta_2=0.35, \beta_3=0.35, \beta_4=0.13, \beta_5=0.12$)

Measures	$\beta_1=0.05$	$\beta_2=0.35$	$\beta_3=0.35$	$\beta_4=0.13$	$\beta_5=0.12$
RBSA	0.05	0.34	0.36	0.13	0.11
	β_m	β_{smb}	β_{lmi}	β_{tm}	β_{tim}
JS	0.95				
JS-TM	0.91			-0.79 [-1.57 0.01]	
JS-HM	1				-0.11 [-0.25 0.04]
FF3	0.95	0.18 [0.13 0.24]	0 [-0.1 0.1]		
FF3-TM	0.96	0.18 [0.13 0.24]	0 [-0.11 0.11]	-0.07 [-0.82 0.66]	
FF3-HM	0.95	0.18 [0.13 0.24]	0 [-0.1 0.1]		0 [-0.15 0.15]

Panel IV ($\beta_1=0.05, \beta_2=0.13, \beta_3=0.12, \beta_4=0.35, \beta_5=0.35$)

Measures	$\beta_1=0.05$	$\beta_2=0.13$	$\beta_3=0.12$	$\beta_4=0.35$	$\beta_5=0.35$
RBSA	0.05	0.13	0.12	0.35	0.35
	β_m	β_{smb}	β_{lmi}	β_{tm}	β_{tim}
JS	0.93				
JS-TM	0.9			-2.7 [-3.44 -1.92]	
JS-HM	1.12				-0.42 [-0.58 -0.28]
FF3	0.94	0.68 [0.63 0.73]	0.03 [-0.08 0.13]		
FF3-TM	0.91	0.68 [0.62 0.73]	0.03 [-0.08 0.13]	-0.1 [-0.87 0.68]	
FF3-HM	0.95	0.67 [0.62 0.72]	0.02 [-0.08 0.12]		0 [-0.15 0.14]

The table provides simulation results of alpha and R^2 under four sets of beta coefficients presented in (4.2). Fund returns are simulated from an alpha, fixed at 5%, a random error, and five style indexes, that is, three-month Treasury bill rates, Russell Top 200 Growth Index, Russell Top 200 Value Index, Russell 2000 Growth Index, and Russell 2000 Value Index. Beta coefficients correspond to the proportions of assets allocated to Treasury bill and four style indexes. We simulate four types of funds, that is, large-cap funds, small-cap funds, well-diversified funds with a preference of large-cap stocks, and well-diversified funds with a preference of small-cap stocks. RBSA stands for RBSA measure by quadratic programming; JS stands for Jensen measure; JS-TM stands for Jensen measure with Treynor-Mazuy market-timing adjustment; JS-HM stands for Jensen measure with Henriksson-Merton market-timing adjustment; FF3 stands for Fama-French three-factor measure; FF3-TM stands for Fama-French three-factor measure with Treynor-Mazuy market-timing adjustment; FF3-HM stands for Fama-French three-factor measure with Henriksson-Merton market-timing adjustment.

C.I. is the empirical confidence interval of alpha estimator based on simulations. Size is the length of C.I.

allocation where 95% of assets are invested in large-cap stocks, this beta estimation is acceptable. FF3 measures are capable to capture the style of the fund. We find that β_{smb} is significant in all three cases, indicating a large-cap fund.

When we study the performance measurement of a fund that behaves like a small-cap fund, which is shown in panel II, we have different results. The estimates based on RBSA are similar to the first panel, but we observe spurious market timing, when using JS-based measures. In both JS-TM and JS-HM, we observe significant negative market timing. This may be caused by different return behavior of small-cap stocks from large-cap stocks, because after we control the size effect in FF3-based measures we don't observe market timing behavior of the fund. Again we find that FF3-based measures are capable of capturing the fund style, since β_{smb} is positive and significant, meaning that the fund generally moves in the same direction as the small stocks.

In panel III we investigate the measures' accuracy in measuring a well-diversified equity fund that leans to large-cap stocks. The accuracy in estimation of RBSA is stable as we observed before. But we find that FF3 measures show that the fund is a small-cap fund, which gives a significant positive β_{smb} . The result contradicts the actual asset allocation of the simulated fund, which invests 70% of its assets in large-cap stocks. Therefore, FF3-based measures don't correctly estimate the coefficients in this situation.

Panel 4 gives the estimation results of a well-diversified equity fund that leans to small-cap stocks. The estimates of RBSA are unbiased in this situation. In RBSA, all five estimates of betas are precisely the true values. We again observe the spurious negative market timing in JS-based measures, but no market timing in FF3 measures. The styles from FF3 measures are accurate, which indicates that it is a small-cap fund.

From simulation results of beta estimation, RBSA is quite successful in identifying the true asset allocation no matter whether it is a large-cap fund, a small-cap fund, or a well-diversified fund. FF3-based measures are capable of capturing the true fund style when the fund is exclusively investing in large-cap or small-cap stocks; however, when the fund is a well-diversified fund, FF3-based measures seem difficult to identify the true styles. Another finding is that FF3-based measures may avoid the spurious market timing that we observed in JS-based measures.

4.4 Conclusion

From our simulation results of the performance measurement and style identification, we find that the RBSA measure seems to be the best measure among the seven measures. The RBSA measure is accurate, efficient and robust, and its performance does not depend on the type of the fund in the study. The average bias of alphas is around 0.01% annually, whereas the average biases of other measures range from 2.45% to 8.57% in absolute value. The beta coefficients estimation is also satisfactory, very close to the true betas as shown in table 2. However, the beta

estimation may be upwardly biased when the beta is actually zero. Since we observed that the bias is quite small around 0.01, it does not pose any difficulty in implementation.

The estimates of JS-based measures are unstable, depending on the fund type. When the fund is a large-cap fund, the results are acceptable. However, when funds invest in small-cap stocks, there are some problems. Firstly, it can not identify fund styles, secondly, it shows spurious negative market-timing, and thirdly it captures only a relatively small part of return variations, where R^2 is quite small compared with other measures with R^2 well above 90%.

FF3-based measures have stable estimates, not depending on the fund type. We find that using FF3-based measures we may avoid spurious market-timing that we observed in JS-based measures. However, they are unable to identify the true fund style of a well-diversified equity fund, thus the alpha estimates derived from the measures are also questionable. In addition, the accuracy and efficiency of the measures are not comparable with the RBSA measure.

Therefore, based on the criterion of accuracy, efficiency and robustness of the estimation of alpha and betas, RBSA comes to be superior to other measures. This encourages us to use RBSA to measure fund performance in the following study.

Chapter V A New Direction of Fund Rating Based on Finite Normal Mixture Model

5.1 Introduction

Numerous literatures have been devoted to fund performance study as we have shown in Chapter II, such as Jensen (1968), Cai, Chan and Yamada (1997), Christopherson, Ferson and Glassman (1998), Grinblatt and Titman (1989), Ippolito (1989) among others. However, there is little study on fund rating till now. There are possibly two reasons. One is that the number of funds is small before 1990. Although the fund history is long enough for fund performance study where people usually use three-year time series data such as Carhart (1997), Connor and Korajczyk (1991), and Elton, Gruber and Blake (1996), the number of funds at each point of time is small to do a cross-sectional fund rating study from the perspective of its distribution. Second, there is a lack of statistical methodology to specify and estimate the density model. However, since the introduction of Expectation and Maximization (EM) algorithm into maximum likelihood estimation with missing data by Dempster, Laird and Rubin (1977), finite mixture model becomes popular, e.g. in medical and biological study, in 1980s and 1990s. Furthermore, the parametric bootstrap procedure by McLachlan and Basford (1988) overcomes the model specification difficulty to some extent.

With the proliferation of funds in 1990s and well-developed statistical methodology over the last decade, we propose a new direction of fund rating that is based on the finite normal mixture distribution model. This model is more flexible and provides more interesting results than current fund rating method by Morningstar, which is commercial fund rating method based on the fixed number of performance groups and preset percentiles.

There are several questions regarding Morningstar's method. First, it is not appropriate to fix the number of performance groups before we investigate the distributions of alphas. It is very possible that we only have one performance group if we find later that the difference of alphas may be just a random effect, caused by the "Luck" of managers. Second, the number of funds in each group shouldn't be fixed before rating. We provide a simple example to illustrate its limitations. Suppose we have 100 funds. 50 funds have alphas around 10% and another 50 funds have alphas around 5%. In this situation, it is obviously not appropriate to say we have five fund performance groups and the top 15 funds are rated as superior funds as implied by Morningstar's method. Instead, it is better to say we only have 2 performance groups and the top 50 funds are superior fund group based on the actual distribution of alphas.

Under the rationale of the example, we propose a method, which is based on the cross-sectional distribution of all of the funds' performances, measured by alphas. The method uses a finite normal mixture model to describe the distribution of

funds performances. Doing so, this method partly overcomes the shortcomings of Morningstar's method. First, the number of performance groups is not fixed, but determined by the spectrum of the alphas of all the funds. For example, if the alphas closely cluster around only one value, we may conclude there is only one performance group from estimation; if alphas cluster around three values, then we may conclude there are three performance groups. We exploit the parametric bootstrap procedure to determine the number of groups at that point of time. Second, we can obtain the posterior probability of individual funds after we have specified and estimated the model, so we know the performance group that the fund belongs to by comparing posterior probabilities. Third, after knowing the group of each fund, it is straightforward to know the number of funds in each group. Therefore, the number of performance groups, the number of funds in each performance group, and the performance group that the fund belongs to, are determined by the cross-sectional distribution of alphas, which are not fixed before rating like Morningstar's method.

This chapter is arranged this way. In section 2, we present the motivation and justification of finite normal mixture model. In section 3, we formulate the fund rating issue under the framework of finite normal mixture model. In section 4, we treat the group information of the fund as missing data, so we can estimate the model under EM framework which is more straightforward and intuitive. In section 5, we show how to determine the number of performance groups by parametric bootstrap procedures. Finally we summarize our fund rating procedures.

5.2 Motivation of Finite Normal Mixture Model

Utilization of the finite normal mixture model is motivated due to the multimodal shape of the distribution of alphas and formal normality tests, which will be shown in Chapter VI. The multimodal shape is a strong indication of finite mixture distribution model. In addition, the results from normality tests contradict what we generally believe on the distribution of alphas. The alpha is the management effect, always interpreted as the manager's ability to deliver abnormal return over passive portfolios. It is affected by many factors, like stock selection, the idiosyncratic news shocks to stocks, asset allocation and rotation, and irregular liquidation caused by redemption from investors, so based on the Linderberg-Levy central limit theorem (CLT) it is generally assumed that the fund performance follows a normal distribution. And the statistical inference of alphas in all the traditional measures relies on the assumption that the alpha is normally distributed.

However, if fund managers' decisions are based on different information sets, we have a group structure in the distribution. Therefore, we can not assume the distribution of alphas as a univariate normal distribution, because the expected performance and investment risk will be different for managers who have different information sources. The more information the managers have, the better investment decision they will make. The managers who can make better decisions are expected to deliver higher performance. In this case, the distribution of alphas of all the fund managers, which are from different information sets, will be a finite normal mixture distribution, i.e. an addition of several normal distributions. This

model is a convenient way to model the group structure of the distribution, as shown by McLachlan and Basford (1988), and it is now widely applied in medicine study, for example, Tao, et al. (2004). In the model each component is a normal distribution which is derived from the corresponding information set. The number of components in the finite normal mixture model is actually interpreted as the number of information sets in the fund market. For example, if managers have private information about firms, they are from superior information set which has higher expected performance, so are expected to deliver higher alphas compared to managers that don't have the private information.

There are several sources that may result in different information sets of fund managers. First, there is managers' ability at acquiring private information. The managers who have private information from insiders are in the private information set, while other managers are in the public information set. The managers from the private information set are expected to have higher performance. Second, there is managers' information collection ability. There is huge amount of information today. Collecting all the information is both time-consuming and expensive. The managers that can efficiently collect the relevant information are expected to deliver higher abnormal returns (alphas), since they possess more useful information for investment decision than other managers. Third, there is managers' ability to analyze the information on hand. Only well analyzed and interpreted information can produce higher abnormal returns. Those who correctly analyze the information are actually in a superior information set. Therefore, depending on the information

they have and the ability to collect and analyze the information, we may have several information sets, for example, badly-analyzed public information set, well-analyzed public information set, and private information set. The differentiated ability to acquire private information and analyze public information may lead to more information sets in an inefficient market, where information is not well transmitted and absorbed.

Based on the information set the managers are from, the expected performance and the investment risk in that information set will be different. Assuming that alphas from the same information set follow a normal distribution, alphas of all the managers will thus form a finite normal mixture distribution. This is the possible reason that we observe multimodal shape and non-normal characteristics of the distribution of all the alphas when the alpha is theoretically expected to be normally distributed.

In our model, we still assume that the alphas are normally distributed but arise from different normal distributions that are corresponding to different information sets. Under this assumption we may observe both a multimodal shape and fat tail in the distribution of alphas. As long as we can identify the number of information sets in the distribution, we know the number of performance groups. That is the way that we determine the number of performance groups.

There are some advantages of the funds rating method that we propose. First, the number of performance groups is not arbitrarily fixed. It is estimated from the empirical cross-sectional distribution of alphas. The number of performance groups is interpreted as the number of information sets in the fund market. Second, the performance group of the fund is determined by the posterior probability of the fund in the estimated distribution. So it is not fixed by the preset cut-off percentiles. Third, the number of funds in each performance group is not fixed, which may change from period to period, depending on the distribution of all the alphas in the study.

5.3 Parametric Formulation of Finite Normal Mixture Distribution Model

We assume that the non-normal features are caused by the group-structure of the data. In our fund performance study, we suspect that there are more than one performance group arising from the distinct information sets. In section 2 we have justified that when market information is segmented or exploited at different levels, there are possibly more than one performance group. Thus there exists a group structure in the distribution of alphas. When there is a group structure in the data, finite normal mixture model is a natural way to model the unknown distribution. In this model, it is expected that the more information that the manager has, the higher the alpha.

In this section, we will formulate the finite mixture distribution model based on McLachlan and Basford (1988). Let Y_1, \dots, Y_n denote a random sample of size n . $Y = (Y_1, \dots, Y_n)^T$ is a column vector representing the entire random sample. Y_j is the random variable corresponding to the alpha of fund j . And its probability density function is $f(y_j)$ ($j=1, \dots, n$). A realization of the random sample is denoted by $y = (y_1, \dots, y_n)^T$. y_j is the alpha of fund j that we observed, which is estimated by the RBSA measure.

In finite mixture model, the density function $f(y_j)$ is a summation of finite component densities, $f_i(y_j)$. It is written in the form,

$$f(y_j) = \sum_{i=1}^g \pi_i f_i(y_j) \quad (5.1)$$

where π_i ($i = 1, \dots, g$) is the mixing proportion or can be called component weight.

They are nonnegative and sum to one, i.e.

$$\begin{aligned} 0 \leq \pi_i \leq 1, (i = 1, \dots, g) \\ \sum_{i=1}^g \pi_i = 1. \end{aligned} \quad (5.2)$$

Here g is the number of components, and $f_i(y_j)$ is the component density. $f(y_j)$ is a g -component mixture density, and the corresponding distribution function is denoted by $F(y_j)$.

There is also a component label variable Z_j , which is a vector with g elements. The i th element of Z_j is Z_{ij} ($i = 1, \dots, g; j = 1, \dots, n$), which is an indicator variable being one or zero. In our fund performance study, we assume that there are g performance groups. If the fund j is from performance group i , then Z_{ij} is one, otherwise Z_{ij} is zero. In the model, π_i is interpreted as the proportion of funds that belong to performance group i ($i = 1, \dots, g$). It is straightforward that π_i is also the probability that fund j is generated from performance group i if we don't know the group information. Therefore, Z_j follows a multinomial distribution,

$$pr\{Z_j = z_j\} = \pi_1^{z_{j1}} \pi_2^{z_{j2}} \dots \pi_g^{z_{jg}} \quad (5.3)$$

where z_j is a realization of Z_j . We look at $f(y_j)$ and $f_i(y_j)$ again with the component label variable Z_j . Given the group information that fund j is from group i , so $Z_{ij} = 1$, $f_i(y_j)$ can be viewed as the conditional probability density of Y_j . And $f(y_j)$ can be viewed as the unconditional density without group information.

The finite mixture model in (5.1) can be viewed as a semi-parametric model between the fully parametric model as represented by a single parametric family ($g=1$) and a nonparametric kernel model ($g=n$). But the single parametric model is usually inadequate to describe the actual distribution. In the finite mixture model, $f_i(y_j)$ is from a parametric family and specified by $f_i(y_j; \theta_i)$, where θ_i is a set of unknown parameters in the component density. The finite mixture model thus can be written as,

$$f(y_j; \Psi) = \sum_{i=1}^g \pi_i f_i(y_j; \theta_i) \quad (5.4)$$

where $\Psi = (\pi_1, \dots, \pi_{g-1}, \xi^T)^T$ contains all the unknown parameters in the model, and ξ is a vector containing all the parameters in component densities, from θ_1 to θ_g . Since the summation of $\pi_i (i = 1, \dots, g)$ is one, we only need to estimate $g-1$ mixing proportions. We arbitrarily leave out the g^{th} mixing proportion, π_g . The parameter space of Ψ is denoted by Ω , the parameter space of θ_i is denoted by Θ .

In our study we assume that there are g information sets in the market, arising from the differentiated ability of acquiring and analyzing both public and private information. The different information sets lead to heterogeneous performance groups. We further assume that the alphas of each group follow a normal distribution, denoted as $N(\mu_i, \sigma_i^2)$. The finite mixture model views the alphas of funds as having been generated from one of the g performance groups with mean and variance as,

$$\alpha_j = \mu_i + \varepsilon_{ij}, \text{Var}(\varepsilon_{ij}) = \sigma_i^2, (i = 1, \dots, g; j = 1, \dots, n)$$

where $\alpha_j (j = 1, \dots, n)$ is the performance of fund j . $\mu_i (i = 1, \dots, g)$ is interpreted as the expected performance in this performance group. $\sigma_i^2 (i = 1, \dots, g)$ is interpreted as the investment risk of a fund in performance group i . The higher the σ_i^2 , the higher the risk to invest in this kind of funds. $\varepsilon_{ij} (i = 1, \dots, g; j = 1, \dots, n)$ follows a normal distribution with mean zero and variance σ_i^2 .

With the above assumption, the finite mixture model can be written as,

$$f(y_j; \Psi) = \sum_{i=1}^g \pi_i f_i(y_j; \theta_i)$$

where

$$f_i(y_j; \theta_i) = \phi(y_j; \mu_i, \sigma_i^2) = (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_j - \mu_i)^2 / \sigma_i^2\right\}.$$

(5.5)

The vector Ψ is $(\pi_1, \dots, \pi_{g-1}, \mu_1, \dots, \mu_g, \sigma_1^2, \dots, \sigma_g^2)^T$, $(i = 1, \dots, g; j = 1, \dots, n)$, containing $3g-1$ parameters.

5.4 EM Algorithm for Finite Normal Mixture Model

To obtain parameter estimates in the finite normal mixture model, Redner and Walker (1984) recommended the application of the expectation-maximization (EM) algorithm, synthesized in the celebrated paper by Dempster et al (1977). As pointed out by Woodward and Sain, (2003), the EM algorithm is an effective tool to deal with various missing data problems. In the fund performance study, there exists missing information, i.e. the component label vector Z_i , so we can formulate the maximum likelihood estimation problem under EM framework.

To estimate the maximum likelihood estimator (MLE) of Ψ with observed data, the likelihood function is written as,

$$L(\Psi) = \prod_{j=1}^n \sum_{i=1}^g \pi_i \phi(y_j; \mu_i, \sigma_i^2) \quad (5.6)$$

and its log likelihood is given by,

$$\log L(\Psi) = \sum_{j=1}^n \log \left\{ \sum_{i=1}^g \pi_i \phi_i(y_j; \mu_i, \sigma_i^2) \right\} \quad (5.7)$$

To find MLE, we take first-order derivatives of the log likelihood function,

$$\frac{\partial \log L(\Psi)}{\partial \Psi} = 0 \quad (5.8)$$

Since it is in a summation form of the component density function, it poses a computational difficulty. However, it is straightforward to find MLE, under EM framework.

We introduce the component label vector z_1, \dots, z_n to the observed data. Formulating the finite normal mixture model under the EM framework, the observed data vector $y = (y_1, \dots, y_n)$ is viewed as incomplete data because the component label vectors, $z = (z_1, \dots, z_n)$, are not available. In our study, each y_j of fund j is regarded as being from one of the performance groups, corresponding to one of components in the finite normal mixture model. z_j is a g dimensional indicator vector for fund j . $z_{ij} = 1$ means that the fund j is from performance group i . Zero means that the fund is not from this group. Therefore the complete data vector is

$$y_c = (y, z) \quad (5.9)$$

where the component vector $z = (z_1, \dots, z_n)$ is the realization of the random vector Z_1, \dots, Z_n . As reported in (5.3) the vector Z_j follows a multinomial distribution. Therefore, viewing the component label vector as part of complete data, we can rewrite the likelihood function as,

$$L_c(\Psi) = \prod_{j=1}^n \prod_{i=1}^g \{\pi_i \phi_i(y_j; \mu_i, \sigma_i^2)\}^{z_{ij}}. \quad (5.10)$$

The log likelihood is written as,

$$\log L_c(\Psi) = \sum_{j=1}^n \sum_{i=1}^g z_{ij} \{\log \pi_i + \log \{\phi_i(y_j; \mu_i, \sigma_i^2)\}\} \quad (5.11)$$

where z_{ij} is linear in the log likelihood function. Now it is much easier to calculate it iteratively in (5.11). z_{ij} is treated as missing data when we apply the EM algorithm to the problem. There are two steps: E for expectation step, and M for maximization step. We use E step to deal with the additional missing data z_{ij} . Given the observed data y , we take conditional expectation of the complete data log likelihood $\log L_c(\Psi)$ using $\Psi^{(k)}$, which is MLE of Ψ in the k^{th} iteration. $\Psi^{(0)}$ is the initial value that we specified in the initial step. On the first iteration we need to calculate the expectation of complete data log likelihood given y and $\Psi^{(0)}$. It is expressed as,

$$Q(\Psi; \Psi^{(0)}) = E_{\Psi^{(0)}} \{\log L_c(\Psi) | y\}. \quad (5.12)$$

The subscript under the expectation operator E means that the expectation is also depending on $\Psi^{(0)}$, which changes over time in iterations. After k^{th} iteration, it is written as,

$$Q(\Psi; \Psi^{(k)}) = E_{\Psi^{(k)}} \{\log L_c(\Psi) | y\} \quad (5.13)$$

To calculate the conditional expectation of the complete data log likelihood, we only need to calculate the conditional expectation of Z_{ij} given the observed data y ,

because Z_{ij} is linear in the log like likelihood function (5.13). Here, Z_{ij} is the random variable corresponding to the realized value z_{ij} .

$$E_{\Psi^{(k)}}(Z_{ij} | y) = pr_{\Psi^{(k)}}\{Z_{ij} = 1 | y\} = \tau_i(y_j; \Psi^{(k)}) \quad (5.14)$$

where $\tau_i(y_j; \Psi^{(k)})$ is the posterior probability that fund j belongs to group i given observed data y . Because the expected probability of $Z_{ij} = 1$ is just π_i , according to the Bayesian theorem, it is straightforward to find $\tau_i(y_j; \Psi^{(k)})$, which is given by

$$\begin{aligned} \tau_i(y_j; \Psi^{(k)}) &= \pi_i^{(k)} \phi_i(y_j; \mu_i^{(k)}, \sigma_i^{2(k)}) / f(y_j; \Psi^{(k)}) \\ &= \pi_i^{(k)} \phi_i(y_j; \mu_i^{(k)}, \sigma_i^{2(k)}) / \sum_{c=1}^g \pi_c^{(k)} \phi_c(y_j; \mu_c^{(k)}, \sigma_c^{2(k)}). \end{aligned} \quad (5.15)$$

Since $\phi(y_j; \mu_i, \sigma_i^2) = (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y_j - \mu_i)^2 / \sigma_i^2\}$, after substituting it into the posterior probability in (5.15), we obtain,

$$\tau_i(y_j; \Psi^{(k)}) = \frac{\pi_i^{(k)} \sigma_i^{-1(k)} \exp\{(y_j - \mu_i^{(k)})^2 / \sigma_i^{2(k)}\}}{\sum_{c=1}^g \pi_c^{(k)} \sigma_c^{-1(k)} \exp\{(y_j - \mu_c^{(k)})^2 / \sigma_c^{2(k)}\}}, (k = 0, 1, \dots; i = 1, \dots, g; j = 1, \dots, n). \quad (5.16)$$

Therefore, the conditional expectation of complete data log likelihood is given by

$$Q(\Psi; \Psi^{(k)}) = \sum_{j=1}^n \sum_{i=1}^g \tau_i(y_j; \Psi^{(k)}) \{\log \pi_i + \log \{\phi_i(y_j; \mu_i, \sigma_i^2)\}\}. \quad (5.17)$$

With the observed data y and parameters, which are estimated from k^{th} maximization $\Psi^{(k)}$, we take conditional expectation of complete data log likelihood. This is E step. Then in the M step, we maximize $Q(\Psi; \Psi^{(k)})$ with

respect to Ψ over the parameter space Ω to get the updated $\Psi^{(k+1)}$. The calculation of updated mixing proportions, $\pi_i^{(k+1)}$, of π_i are independent of the calculation of updated parameter, $\xi^{(k+1)}$, of ξ , containing the parameters in component densities. If the missing information, z_{ij} is known, the complete data MLE of π_i is simply,

$$\hat{\pi}_i = \sum_{j=1}^n z_{ij} / n, (i = 1, \dots, g). \quad (5.18)$$

Since z_{ij} is not known, we use, $\tau_i(y_j; \Psi^{(k)})$ to replace z_{ij} in the above estimation, which is the conditional expectation of z_{ij} in complete data log likelihood. The updated mixing proportion is,

$$\pi_i^{(k+1)} = \sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) / n, (i = 1, \dots, g). \quad (5.19)$$

When updating mixing proportions π_i on the $(k+1)^{\text{th}}$ iteration, we sum up all the posterior probabilities that the fund belongs to performance group i . Each y_j contributes to the update.

Regarding the update of ξ on the M step in $(k+1)^{\text{th}}$ iteration, we take the first order derivative of the conditional expectation log likelihood with respect to parameters, and then solve the equations to find out MLE of $\xi^{(k+1)}$ in the $(k+1)^{\text{th}}$ iteration:

$$\frac{\partial Q(\Psi; \Psi^{(k)})}{\partial \xi} = 0, \quad (5.20)$$

gives,

$$\frac{\partial Q(\Psi; \Psi^{(k)})}{\partial \xi} = \sum_{j=1}^n \sum_{i=1}^g \tau_i(y_j; \Psi^{(k)}) \frac{\partial \log \{\phi_i(y_j; \mu_i, \sigma_i^2)\}}{\partial \xi} = 0. \quad (5.21)$$

ξ contains the parameters (μ_i, σ_i^2) , $(i = 1, \dots, g)$.

Since $\phi(y_j; \mu_i, \sigma_i^2) = (2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y_j - \mu_i)^2 / \sigma_i^2\}$, we have,

$$\frac{\partial Q(\Psi; \Psi^{(k)})}{\partial \mu_i} = \sum_{j=1}^n \tau_i(y_j; \Psi^{(k)})(y_j - \mu_i) = 0, (i = 1, \dots, g), \quad (5.22)$$

$$\frac{\partial Q(\Psi; \Psi^{(k)})}{\partial \sigma_i^2} = \sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) \left(\frac{1}{\sigma_i^2} - \frac{(y_j - \mu_i)^2}{\sigma_i^4} \right) = 0, (i = 1, \dots, g).$$

Then, MLE of (μ_i, σ_i^2) , $(i = 1, \dots, g)$ are obtained as,

$$\mu_i = \frac{\sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) y_j}{\sum_{j=1}^n \tau_i(y_j; \Psi^{(k)})}, (i = 1, \dots, g), \quad (5.23)$$

$$\sigma_i^2 = \frac{\sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) (y_j - \mu_i)^2}{\sum_{j=1}^n \tau_i(y_j; \Psi^{(k)})}, (i = 1, \dots, g).$$

Since we know that $\pi_i^{(k+1)} = \sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) / n$, $(i = 1, \dots, g)$ in (5.19), we can

simplify the above two equations as,

$$\mu_i^{(k+1)} = \frac{\sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) y_j}{n\pi_i^{(k+1)}}, (i = 1, \dots, g), \quad (5.24)$$

$$\sigma_i^{2^{(k+1)}} = \frac{\sum_{j=1}^n \tau_i(y_j; \Psi^{(k)})(y_j - \mu_i)^2}{n\pi_i^{(k+1)}}, \quad (i = 1, \dots, g).$$

Note that $k+1$ denotes the updated parameters for the $k+1^{\text{th}}$ iteration. We repeat the E step and M step alternatively until the estimates of parameters in Ω converge. A desirable feature of the EM algorithm is that the solutions are in closed form for the finite normal mixture model.

Procedural steps of EM algorithm are summarized below:

1. Choose initial values of $(\pi_i^{(0)}, \mu_i^{(0)}, \sigma_i^{2^{(0)}}), (i = 1, \dots, g)$, given the g -component finite normal mixture model.
2. Estimate posterior probability that fund j ($j = 1, \dots, n$) belongs to performance group i ($i = 1, \dots, g$) given the observed data y and $\Psi^{(k)}$ that are estimated parameters in k th iteration.

$$\tau_i(y_j; \Psi^{(k)}) = \frac{\pi_i^{(k)} \sigma_i^{-l^{(k)}} \exp\{(y_j - \mu_i^{(k)})^2 / \sigma_i^{2^{(k)}}\}}{\sum_{c=1}^g \pi_c^{(k)} \sigma_c^{-l^{(k)}} \exp\{(y_j - \mu_c^{(k)})^2 / \sigma_c^{2^{(k)}}\}},$$

$$(k = 0, 1, \dots; i = 1, \dots, g; j = 1, \dots, n).$$

3. Update $(\pi_i^{(k)}, \mu_i^{(k)}, \sigma_i^{(k)}), (i = 1, \dots, g)$ by the following equations in sequence,

$$\pi_i^{(k+1)} = \sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) / n, \quad (i = 1, \dots, g),$$

$$\mu_i^{(k+1)} = \frac{\sum_{j=1}^n \tau_i(y_j; \Psi^{(k)}) y_j}{n\pi_i^{(k+1)}}, \quad i = (1, \dots, g),$$

$$\sigma_i^{(k+1)} = \frac{\sum_{j=1}^n \tau_j(y_j; \Psi^{(k)})(y_j - \mu_i)^2}{n\pi_i^{(k+1)}}, i = (1, \dots, g).$$

4. Repeat step 3 and step 4 until the difference between $(\pi_i^{(k+1)}, \mu_i^{(k+1)}, \sigma_i^{(k+1)}), (i=1, \dots, g)$ and $(\pi_i^{(k)}, \mu_i^{(k)}, \sigma_i^{(k)}), (i=1, \dots, g)$ is smaller than the preset tolerance level.

5.5 Determination of the Number of Components by Parametric Bootstrap Procedures

In the procedures outlined in the previous section, we have to specify the number of components in the first step to initiate the iteration. This is a model specification problem. In some situations, the number of components is given as a priori information. However, in other situations, the number of components has to be inferred from observed data along with other estimates of density function. In the fund performance study, the group information is not known and is of particular interest for us to rate funds. The number of components in the finite normal mixture model indicates the number of performance groups that exist among all the funds. In addition, the number of components directly affects the classification of funds. In an extreme example, suppose that our test shows there is only one performance group, and then it is not necessary to group funds. The abnormally high or low alphas are just the results of “Luck”.

There are three approaches to estimate the number of components in the finite normal mixture model. The first method is nonparametric by investigating the

number of modes in an estimated kernel density. We know that multimodal shape is a strong implication of mixture model. Roeder (1994) argued that if there is no priori information about the number of components and component densities, it is appropriate to assess the number of modes. Inferential procedures to assess the number of modes include Titterton et al. (1985) and Silverman (1981, 1986), in which Silverman used a kernel method to estimate the density function and develop a technique to assess the number of modes. Other studies using the number of modes include Hartigan and Mohanty (1992), Wong (1985), and Fisher, et al. (1994). But there is an obvious drawback of this approach. If the means of component densities are not sufficiently separate enough, the number of modes are less than the number of components. Therefore it is difficult to identify the right number of components.

The second stream is based on penalized log likelihood, such as AIC and BIC. As the log likelihood increases with the addition of a component to the finite normal mixture model, the log likelihood is penalized by the subtraction of a term that penalizes the model for the number of parameters in it. Using this method, the results are acceptable as discussed in Solka et al. (1998). But the main purpose of this approach is for density estimation, not the identification of the number of terms. In addition, it produces no confidence of results, so we have no idea of Type I error if we reject the null hypothesis.

In our study, we assess the number of components by hypothesis test, using likelihood ratio as the test statistic. The approach focuses on group finding, and most importantly it provides a p value to assess the confidence about the number of components. In the finite normal mixture model, the likelihood ratio test statistic is,

$$-2 \log(\lambda) = 2\{\log L(\hat{\Psi}_1) - \log L(\hat{\Psi}_0)\}, \quad (5.25)$$

where $\hat{\Psi}_0$ and $\hat{\Psi}_1$ are MLE of Ψ under $H_0 : g = g_0$ and $H_1 : g = g_1$ respectively

Usually we increase the number of components g_0 one by one in sequence to see if the increase in log likelihood starts to fade away after some threshold value g_0 .

After adding a new component into finite normal mixture model, if the increase of log likelihood is not significant, then we can conclude that there is no sufficient evidence to reject the hypothesis that there are g_0 components in the model. From the above analysis, we know that as long as we know the sampling distribution of the likelihood ratio test statistic, $-2 \log(\lambda)$, we can proceed to hypothesis test, and finally identify the number of components.

Unfortunately, in the finite normal mixture model, regularity conditions (Cramer, 1946) do not hold for $-2 \log(\lambda)$ to have its usual asymptotic null distribution of Chi-square, where the degrees of freedom are equal to the difference of the number of parameters under null hypothesis and the number of parameters under alternative hypothesis. In the work by Titterington, et al. (1985) and McLachlan and Basford (1988), it is well discussed that conventional asymptotic results for the null distribution of the likelihood ratio test statistic do not hold because the null

hypothesis lies on the boundary of the alternative hypothesis (in null hypothesis one mixing proportion is specified as zero).

To rescue it, parametric bootstrap procedures proposed by McLachlan (1992) are used to assess the p value of likelihood ratio test statistic, $-2\log(\lambda)$. Simulation is needed in this occasion. Feng and McCulloch (1996) pointed out that this approach leads to valid statistical inference. Wolfe (1971) proposed a modified likelihood ratio test statistic by the rule of thumb, but McLachlan (1987) showed the results may not be applicable in heteroscedastic case where component variances are unequal.

In finite normal mixture model, we test,

$$H_0 : g = g_0 \text{ versus } H_1 : g = g_1.$$

We let $g_1 = g_0 + 1$ in order to find the smallest g that is consistent with the data.

Since the null distribution is unknown, we use parametric bootstrap procedures to assess the p value of likelihood ratio test statistic, $-2\log(\lambda)$. Bootstrap samples are

generated from the finite normal mixture model with Ψ replaced by the MLE, $\hat{\Psi}_0$, which is estimated under null hypothesis by EM algorithm with the observed data.

Then we fit the bootstrap sample under null hypothesis and alternative hypothesis respectively by EM, to obtain the bootstrapped likelihood ratio value, $-2\log(\lambda)^{(b)}$,

where b means the b^{th} likelihood ratio value from the b^{th} bootstrapped sample. We repeat the sampling for a number of times B , so we have a sequence of likelihood ratios $\{-2\log(\lambda)^{(b)}\}$. The sequence of values provides an approximation of the

unknown null hypothesis distribution. Then we refer the original likelihood ratio, computed from the observed data, to the sequence $\{-2 \log(\lambda)^{(b)}\}$. We find the p value of $-2 \log(\lambda)$ as,

$$p = 1 - \frac{j}{B+1} \quad (5.26)$$

where j is the number of replicated likelihood ratio values that are smaller than the original likelihood ratio. If we reject null hypothesis under $g = g_0$, then we can increase the number of components under null hypothesis by one, and move forward to $H_0 : g = g_0 + 1$ versus $H_1 : g = g_0 + 2$ until we don't have sufficient evidence to reject null hypothesis. The threshold g is the number of performance groups in our study.

To facilitate programming, we outlined the parametric bootstrap procedures as follows:

1. Given the observed data y , fit the original data under $H_0 : g = g_0$ and $H_1 : g = g_0 + 1$ respectively by the EM algorithm to get estimates $\hat{\Psi}_0$ and $\hat{\Psi}_1$.
2. Substitute Ψ in finite normal mixture model with the estimated $\hat{\Psi}_0$ and $\hat{\Psi}_1$ to get probability density function under null hypothesis and alternative hypothesis respectively.
3. From the density functions we compute the original likelihood ratio value, $-2 \log(\lambda) = 2\{\log L(\hat{\Psi}_1) - \log L(\hat{\Psi}_0)\}$.

4. Take a bootstrap sample from the finite normal mixture model with parameters, Ψ , replaced by $\hat{\Psi}_0$ that we estimate in step 1.
5. Fit bootstrap sample we obtained in step 4 under $H_0 : g = g_0$ and $H_1 : g = g_0 + 1$ respectively by EM algorithm to get estimate of $\hat{\Psi}_0^{(b)}$ and $\hat{\Psi}_1^{(b)}$. The superscript b represents the estimate from b^{th} bootstrap sample.
6. Substitute Ψ in finite normal mixture model with the estimated $\hat{\Psi}_0^{(b)}$ and $\hat{\Psi}_1^{(b)}$ to get probability density function under null hypothesis and alternative hypothesis respectively.
7. From the density functions we compute the likelihood ratio value, $-2 \log(\lambda)^{(b)} = 2 \{ \log L(\hat{\Psi}_1^{(b)}) - \log L(\hat{\Psi}_0^{(b)}) \}$.
8. Repeat step 4 through step 7 for B times to get a sequence $\{-2 \log(\lambda)^{(b)}\}$ for $b = 1, \dots, B$.
9. Order the sequence of likelihood ratios, and then count the number of values that are smaller than $-2 \log(\lambda)$, which is the original likelihood ratio, in step 3.
10. Find p value of $-2 \log(\lambda)$ as $1 - \frac{j}{B+1}$, where j is the number of counts in step 9.

Usually large B is required to get a precise p value. However, the amount of computation involved is considerable. We choose B as 200 as the number of the bootstrapped samples. The B is sufficient, because our main concern is to see whether we can reject the null hypothesis not to get the precise p values.

5.6 Fund Rating Procedures

We use finite normal mixture model to study the distribution of alphas attempting to find the number of performance groups and assign a rating to each fund. In our research, we provide a new direction of fund rating method that is more flexible and theoretically solid than current fund rating method, like Morningstar's method. The model is implemented in the following steps.

Step One: Normality Check

We will check the normality of the distributions first. If they are normal then no further steps are necessary. It implies that there are no superior or inferior funds in the market. The abnormal negative or positive alphas we observed in the last period are just the consequence of "Luck". In other words, the managers happened to have picked the right stocks and correctly timed the market. If the distributions show non-normal features, such as multimodal shape in kernel density, then it is a good indication of group structure in the data. This may be caused by the different information sets that managers are from. We can also test the normality by formal tests, such as the Jarque-Bera test and Lilliefors test. When we find that the distributions can not be described by a univariate normal distribution, the natural way to model it is a finite normal mixture model. The model provides an intuitively appealing interpretation about the number of components and the expected values and the variances of component densities. They are interpreted as the number of performance groups, the expected performance of the fund, and the expected investment risk of the fund respectively. Note that the expected performance here is not expected alpha of all the funds, instead it is the expected alpha of the funds in

the performance group that the fund is from. In addition, the model provides the posterior probability that the fund belongs to each group. With this information we can group and rate the funds.

Step Two: Specification of the Finite Normal Mixture Model

Before estimating the model parameters, we have to specify the number of components. This is theoretically difficult. A number of approaches are proposed. We use the parametric bootstrap procedures outlined in section 5.5 to identify the number of components. We assess the p value based on the empirical distribution of likelihood ratio. In searching the appropriate number of components we increase g in the null hypothesis gradually one by one until we find the smallest threshold g that is consistent with data.

Step Three: Estimation of the Finite Normal Mixture Model

After having fixed the number of components in the model, then we proceed to estimate the model. We use EM algorithm outline in section 4 to solve out the likelihood function. This is not only for straightforward computation of MLE of Ψ , but also for the intuitive interpretation of group information. We introduced a component label vector Z_j with value of one or zero, indicating whether the fund was generated for the performance group or not. There are two steps in EM algorithm. In the E step we take the conditional expectation of Z_j , given y and current $\Psi^{(k)}$ in finite normal mixture model, to obtain the posterior probability. Then we proceed to the M step to update $(\pi_1, \dots, \pi_g; \mu_1, \dots, \mu_g; \sigma_1^2, \dots, \sigma_g^2)$

sequentially with posterior probabilities of all the funds. We repeat the E step and the M step until estimates converge.

The results have an intuitively appealing interpretation. π_i is interpreted as the proportion of funds in performance group $i, i = 1, \dots, g$. μ_i is interpreted as the expected alpha for performance group $i, i = 1, \dots, g$. σ_i^2 is interpreted as the investment risk of funds in performance group $i, i = 1, \dots, g$. The higher the σ_i^2 , the higher the risk. The high σ_i^2 implies that the performance is volatile in this group. We also have the posterior probability that each fund belong to each group, which provides us with a basis for grouping and rating.

Step Four: Fund Rating

We rank $\mu_i (i = 1, \dots, g)$, which is the expected alpha of performance group i . The funds in the group that has the highest ranking are viewed as superior funds, whereas the funds in the group that has the lowest ranking are viewed as inferior funds. The differences of alphas in the performance group are regarded as random effects. Thus we consider the funds in the same performance group have the same expected performance.

Chapter VI Empirical Study of Fund Performance and Rating

6.1 Introduction

In this chapter we conduct an empirical study of US domestic equity mutual funds, using the RBSA measure and dynamic fund rating procedures outlined in Chapter III and Chapter V respectively. In section 2 we introduce the data used in this chapter. In section 3 we study the effectiveness of the RBSA measure when identifying fund styles and measuring fund performance. With the estimates of alphas from section 3, we further investigate the distributions of alphas in section 4. The non-normal characteristics of the distributions and multimodal shapes motivated us to model the distributions under the finite normal mixture model framework. We specified the model and investigate the group-structure of the distributions by parametric bootstrap procedures in section 5. After the specification of the model, we estimate the model by EM in section 6. The estimates provide an understanding of expected performance and investment risk of each performance group, denoted by μ_i and σ_i^2 respectively. With the estimated parameters in section 6, we can calculate the posterior probabilities of each fund, which is the basis for our fund rating method. The method and results are presented in section 7 and Appendix B respectively. In section 8, we study the fund performance persistence issue based on the rating in the first period.

6.2 Data Description

In our study we have 1145 equity mutual funds. They consist of all the well-diversified US equity funds that are established before November 30, 1995 and have data during the sample period from November 30, 1995 to November 30, 2004. The data is from Bloomberg. The funds are further grouped into nine asset categories according to the capitalization and B/M ratios of the stocks they hold. Funds that hold large-cap stocks are defined as large funds, and funds that hold high B/M ratio stocks are defined as value funds. The nine categories are Large Value (LV), Large Blend (LN), Large Growth (LG), Medium Value funds (MV), Medium Blend funds (MN), Medium Growth funds (MG), Small Value funds (SV), Small Blend funds (SN), and Small Growth funds (SG). The category information of the funds is directly from Morningstar. In our sample, we have 218 LV funds, 272 LN funds, 234 LG funds, 43 MV funds, 50 MN funds, 135 MG funds, 37 SV funds, 58 SN funds, and 98 SG funds. The nine-category classification is for fund style analysis and performance measurement. In this fund-rating study, we combine the nine categories into five categories, because the number of funds is small for the three groups of medium-cap funds and three groups of small-cap funds. We combine MG, MN and MV into M, and combine SG, SN and SV into S. So we have five asset categories, that is, Large Value (LV), Large Blend (LN), Large Growth (LG), Medium-cap (M), and Small-cap (S). We have 228 funds in M and 193 funds in S respectively

To study fund performance, it is a common practice to use three years data, e.g. studies by Cai, et al. (1997), Carhart (1997), Elton, et al. (1996), and Kosowski, et al. (2001) among others. Therefore, to study the funds' performance and performance dynamics over time, we divide the sample period into three sub-periods: from November 30, 1995 to November 30, 1998; from November 30, 1998 to November 30, 2001; from November 30, 2001 to November 30, 2004. Each of them is exactly three years. We use the RBSA measure to evaluate the funds in the three sub-periods respectively by quadratic programming. The alphas of funds in three periods are the basis for fund rating study and the evolution of the number of information sets.

6.3 Fund Styles and Performance

We use the RBSA measure to evaluate fund performance. In Chapter IV, we have showed that the RBSA measure is superior to traditional regression-based measures, like the Jensen measure (Jensen, 1968) and Fama-French three-factor measure (Fama and French, 1993). We firstly formulate the problem as a convex quadratic programming problem, then identify fund styles, β , and finally derive the fund's performance as the difference between fund return and a combination of passive style indexes, represented by $\beta_1 f_{1t} + \beta_2 f_{2t} + \dots + \beta_k f_{kt}$. The model is,

$$\begin{aligned} \text{Min} \quad & \text{Var}(\varepsilon_t) = \text{Var}(r_t - \beta^T f_t) & (6.1) \\ \text{s.t.} \quad & \beta^T e = 1 \\ & \beta \geq 0 \end{aligned}$$

where $r_t = \beta_1 f_{1t} + \beta_2 f_{2t} + \dots + \beta_{10} f_{10t} + \varepsilon_t$ is the fund return. It can be decomposed into two parts: one is from style investment, $\beta_1 f_{1t} + \beta_2 f_{2t} + \dots + \beta_{10} f_{10t}$, which can be easily tracked by investing in passive portfolios; the other is from active management denoted by ε_t . The mean of ε_t during the sample period is interpreted as the fund's performance. β is a column vector of style coefficients on different style indexes. f_t is a matrix of returns of style indexes. In the model we adopt ten exclusive style indexes to cover all the main investable securities for domestic US equity funds: the three-month Treasury Bill Rate, MSCI Government Bond Index, Russell Top 200 Growth Index, Russell Top 200 Value Index, Russell Midcap Growth Index, Russell Midcap Value Index, Russell 2000 Growth Index, Russell 2000 Value Index, MSCI EAFE, and MSCI EMF.

We focus on styles and performance of funds in the last period from 2001 to 2004, since investors are usually most interested in recent performance. They believe that last period fund performance is predictive of the next period fund performance. From 1995 to 2001, the styles and performance of funds are studied in the same way. In the study of fund grouping and rating, we will evaluate the funds over nine years and use the estimation results of alphas from 1995 to 2004.

We want to point out two things. First, the estimated style coefficients depend on style indexes we choose. Second, the interpretation of the results depends on the definition of the style indexes. In our study we define the top 200 stocks, ranked by capitalization, as large-cap stocks. The next 800 stocks are defined as medium-cap

stocks, and the next 2000 stocks are defined as small stocks. We choose this definition according to the Russell's index definition². In fact, a different definition is possible but does not affect our fund performance measurement. In our definition, the style indexes are both publicly available and mutually exclusive. The exclusiveness of indexes is crucial. It makes our estimation more accurate and efficient than overlapped indexes. Since our purpose is to correctly measure fund performance, as long as we can accurately identify the true profile of fund styles, then this definition is feasible. Using a different definition of style indexes only changes the interpretation of style coefficients. It does not affect the identification of the profile of fund styles, so it does not affect fund performance measurement by the RBSA measure.

Figure 2 to figure 10 show the estimated styles of the nine groups of funds. We find that what we obtain from style analysis is consistent with the actual asset compositions of the

² Please see <http://www.russell.com/US/Indexes/US/Definitions.asp>

Figure 2
234 Large Growth Mutual Funds

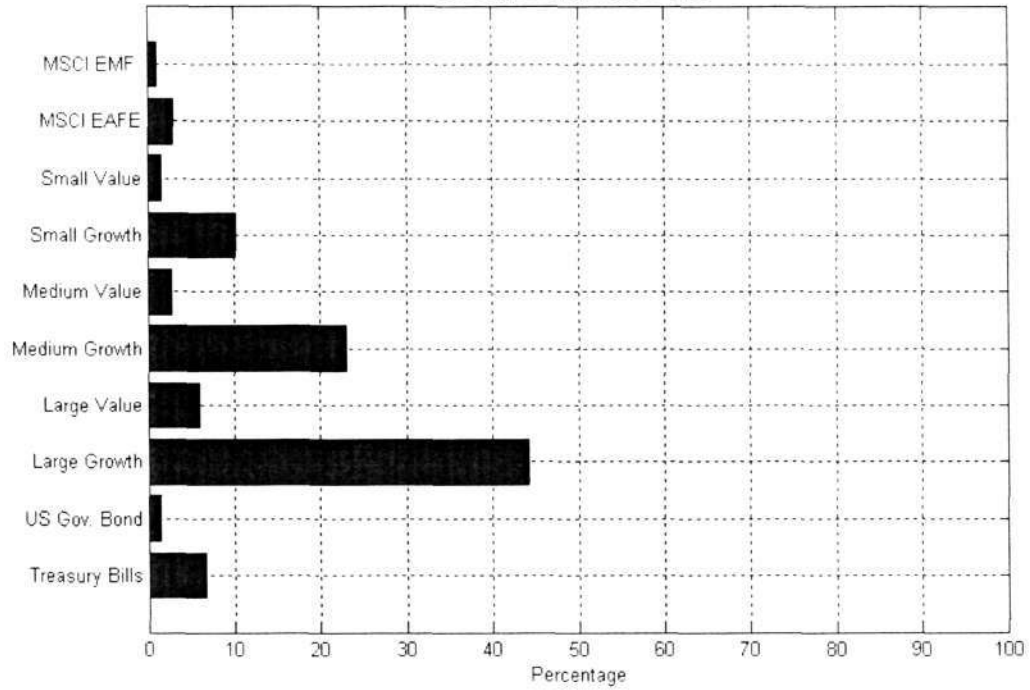


Figure 3
272 Large Blend Mutual Funds

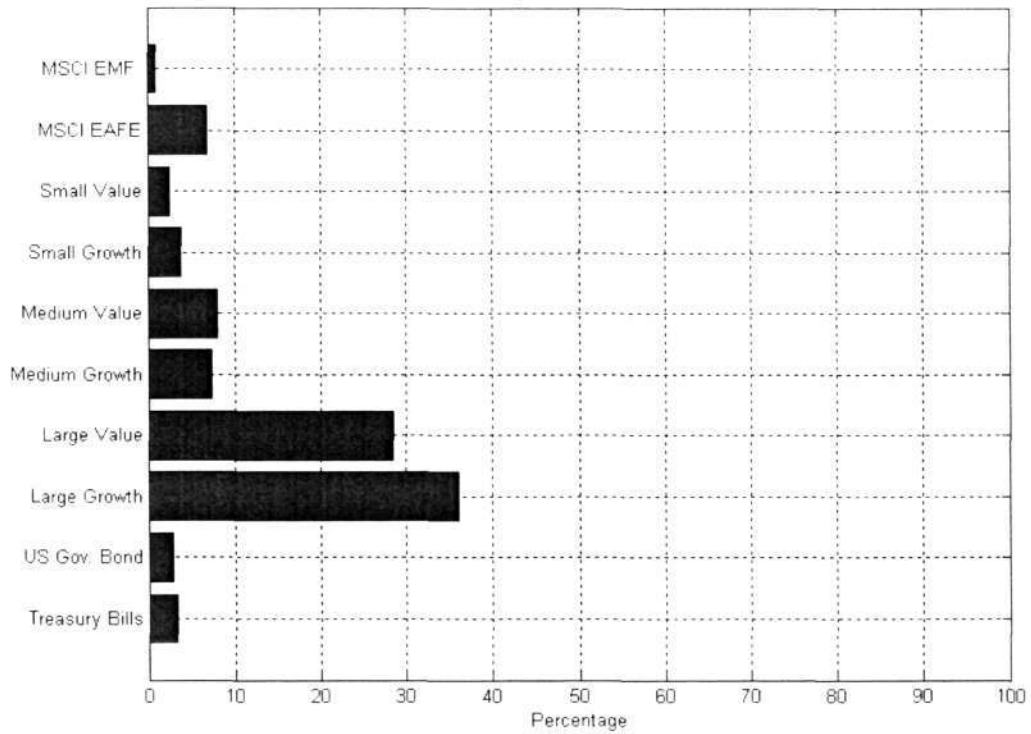


Figure 4
218 Large Value Mutual Funds

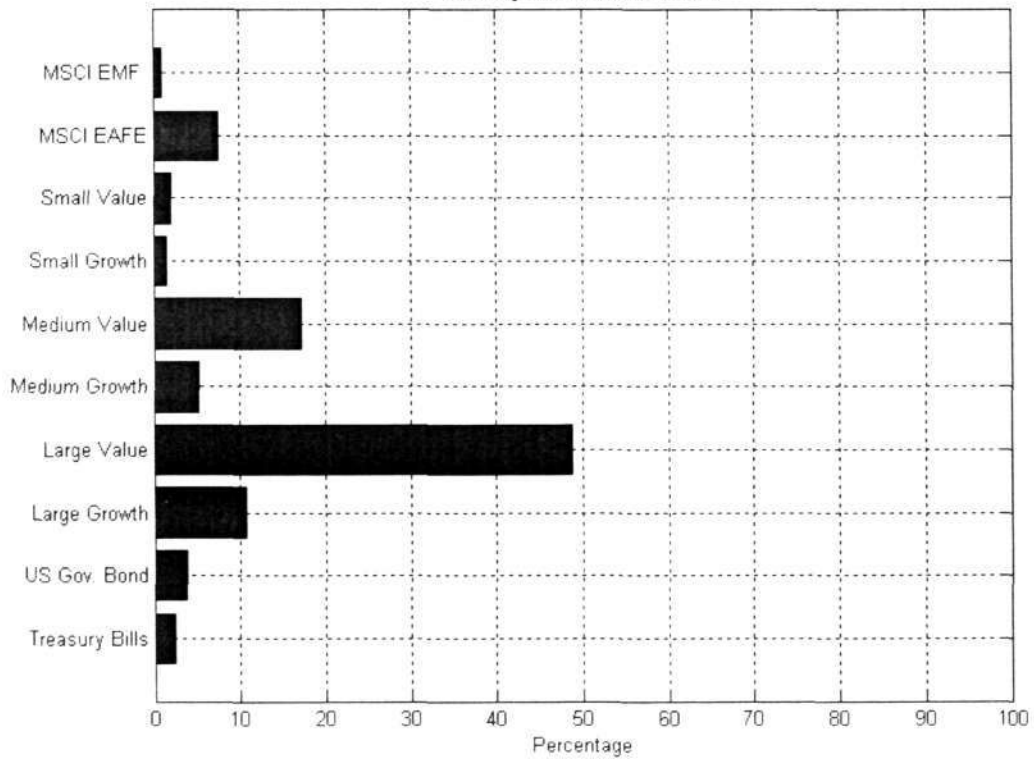


Figure 5
135 Medium Growth Mutual Funds

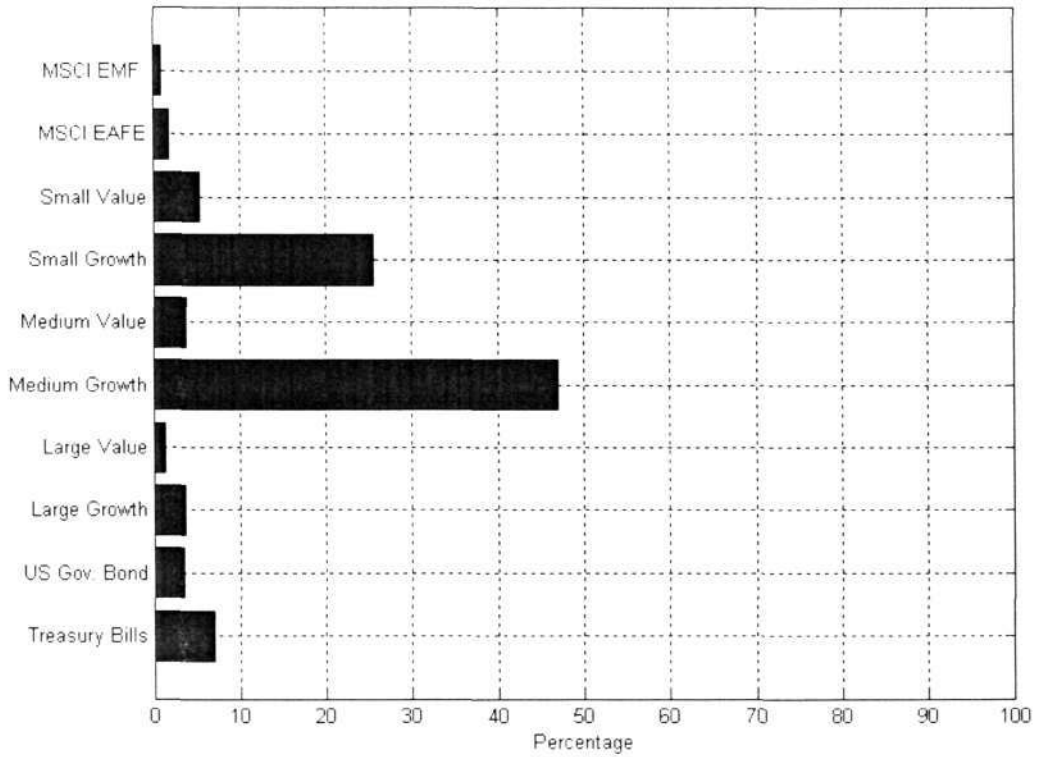


Figure 6
50 Medium Blend Mutual Funds

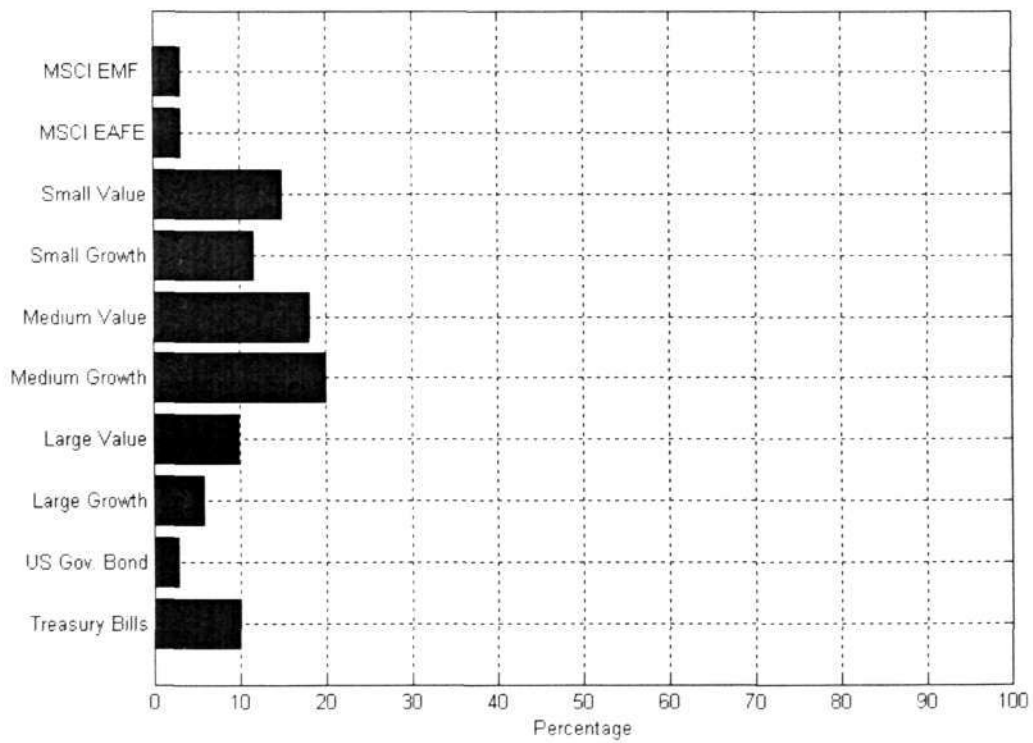


Figure 7
43 Medium Value Mutual Funds

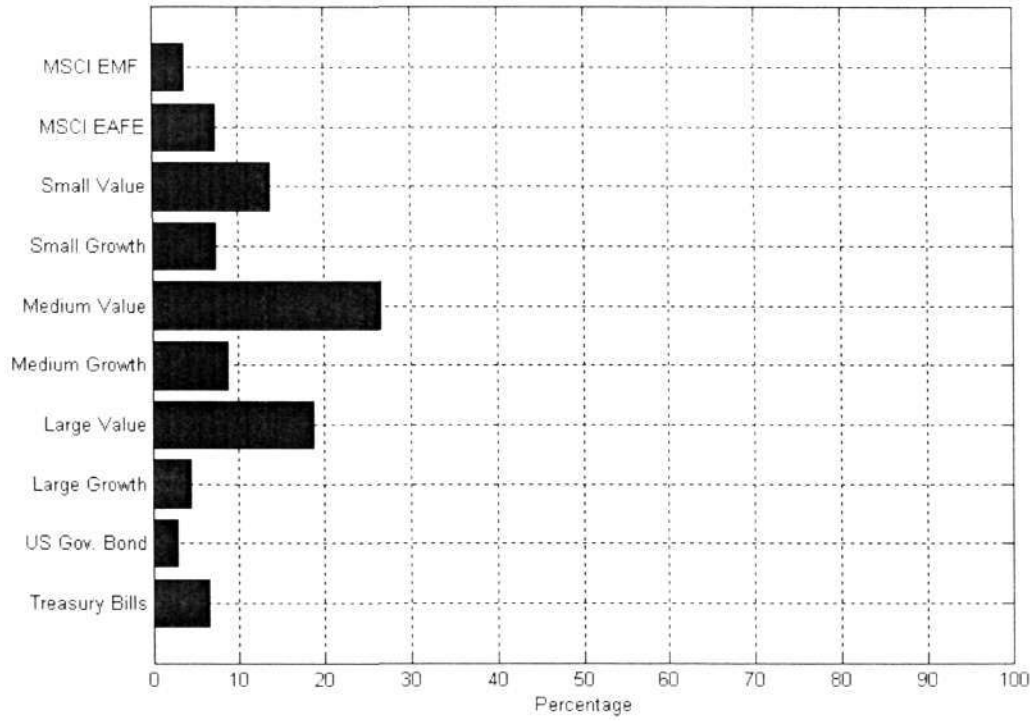


Figure 8
98 Small Growth Mutual Funds

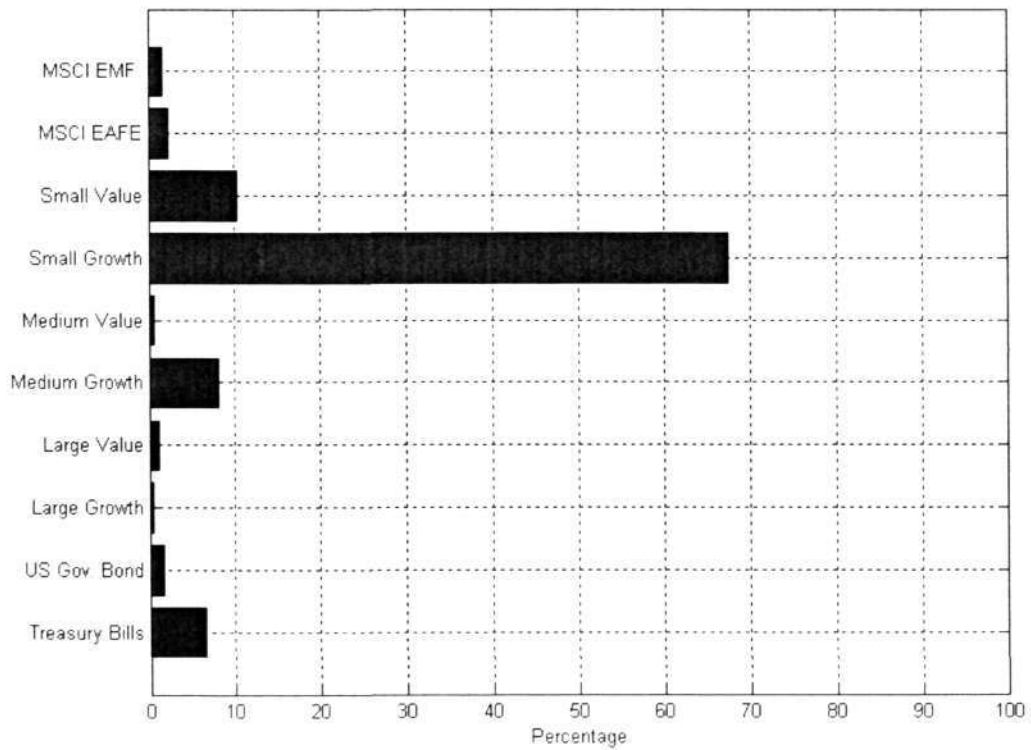


Figure 9
58 Small Blend Mutual Funds

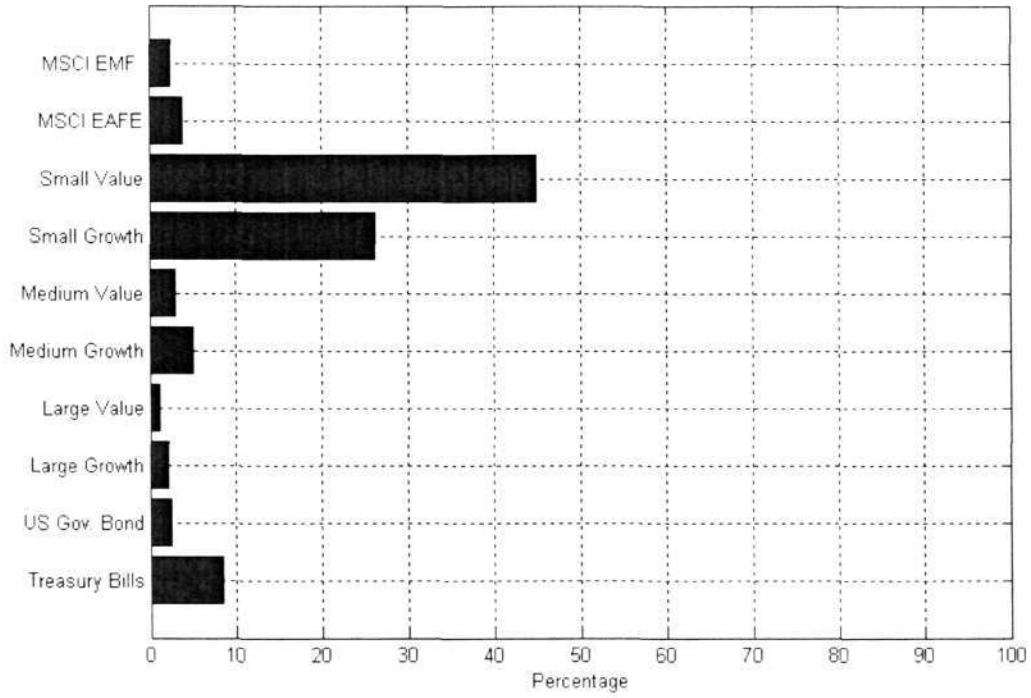
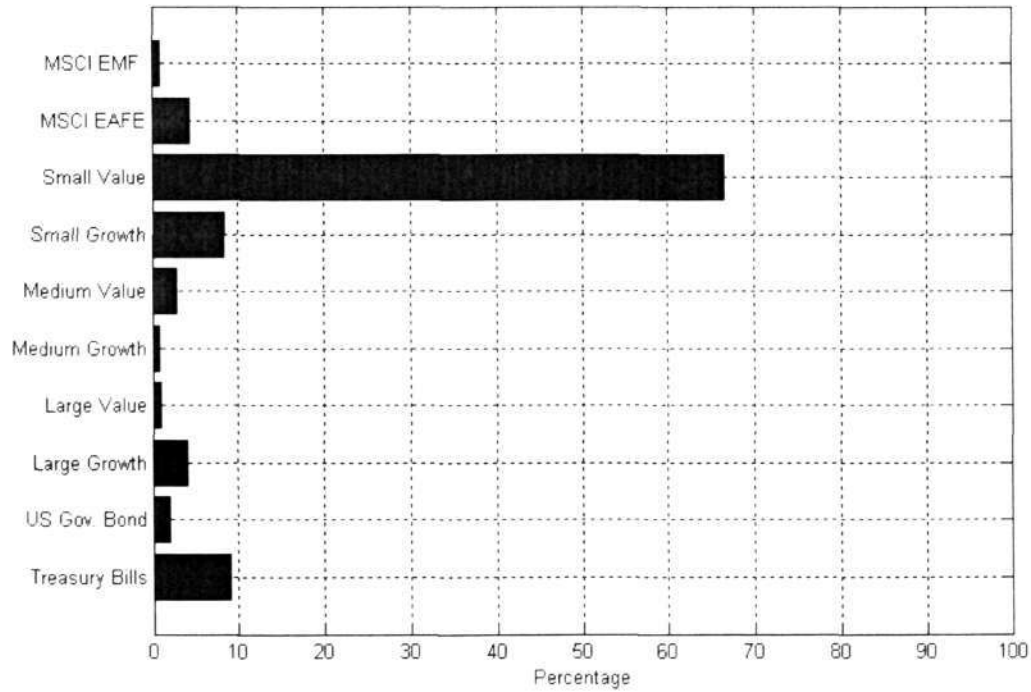


Figure 10
37 Small Value Mutual Funds



funds implied by the asset category information of Morningstar. Figure 2 shows that large-cap growth funds on average put around 45% of assets in large-cap growth stocks, 23% in medium-cap growth stocks and 10% in small-cap growth stocks. So totally the funds put around 75% of assets in growth stocks and around 45% in large-cap growth stocks. We have 234 large-cap growth funds in our study. Figure 3 is the results of large-cap blend funds. From the category information, these funds should invest in large-cap stocks and have no preference on growth stocks and value stocks. Estimation of the funds' styles confirms this kind of investment strategy. The figure shows around 65% of assets are large-cap stocks. The division between value stocks and growth stocks are roughly equal, with around 28% in large value stocks and 36% in large growth stocks. Figure 4 is for large-cap value funds. The value investment strategy is clearly identified in the figure: around 65% of assets is invested in large-cap or medium-cap stocks. We also observe that almost 50% is in large-cap value stocks. Our study of the first three figures clearly supports the style analysis, which can identify the true investment styles of the large-cap funds using quadratic programming techniques.

Figure 5 to figure 7 are estimation results from the medium-cap funds. We observe that the medium-cap growth funds have over 70% of assets in medium-cap growth stocks and small-cap growth stocks. The results show the fund also put a part of their assets in small-cap stocks. This is due to the definition of small-cap stocks and medium-cap stocks, or a high correlation between the medium-cap growth index and small-cap growth index. Regarding the medium-cap blend funds, we find

in figure 6 that the funds widely spread the asset allocation with a some concentration in medium-cap stocks. In figure 7, the medium-cap value funds mainly invest in medium-cap growth funds and also put a large part of their assets in large-cap value stocks and small-cap value stocks.

The style identification of small-cap funds is most significant, which are shown from figure 8 to figure 10. In figure 8, which gives the results of small-cap growth funds, there is approximately 70% in small-cap growth stocks. In figure 8 the small-cap blend funds also have around 70% assets in small-cap stocks, but they have a significant portion in both value stocks and growth stocks. They seem to have preference on small-cap value stocks. Figure 10 we also find around 65% of assets are invested in small-cap value stocks.

Therefore, what we estimate from return-based style analysis using quadratic programming techniques is consistent with the actual styles identified by category information, which directly analyzes the portfolio holdings of the funds. In addition it provides the information of how much asset is allocated to each asset category or the fund's behavior as if it has allocated assets to these asset categories. The empirical results show that the RBSA measure is capable of finding the styles of funds in reality, not just the styles of simulated funds. One point worth mention is that when the fund heavily loads some stocks in one style index, the style estimation may differ from the actual asset holdings. The combination of the style indexes with estimated style coefficients only mimics the return behavior of the

fund. However, as long as we only care about the styles and the return behavior implied by the styles, this difference is not a limitation but an advantage. Because it is able to identify the true styles of the fund, while fundamental style analysis may disguise the true information of styles.

After identifying the styles, we construct a passive portfolio for each fund based on the estimated style coefficients, and then measure the fund's performance as the difference between the fund return and the return of the passive portfolio. The performance of each group is reported in table 3. The performance is the after-expense performance after deducting various fees including management fees. The medium blend funds, small blend funds and small value funds show positive after-expense performance. The small value funds deliver the highest alpha which is 0.55% annually. This is consistent with previous findings, such as Roll (1981) and Chan, et al. (1985), that small value stocks are able to deliver higher abnormal returns. The medium growth funds show the poorest performance which is -2.95%. All the large-cap funds deliver negative performance, which is ranging from -1.06% to -2.74%, but the performance variation for large-cap funds is smaller. If we view the performance variation as the investment risk, we may observe an approximate relationship between average performance and the investment risk. In other words, the higher the investment risk, the higher the average performance for that group of funds. The average performance of small-cap funds is highest, -0.41%, however, the investment risk is also highest which is 4.75%. Although the large-cap funds' performance is worst, the performance is stable. Their

performance variation is the smallest. Medium-cap funds lie between large-cap funds and small-cap funds. With alphas estimated in this section, we can further investigate the fund rating issues in the following sections.

Table 3: Performance Measurement of 1145 US Domestic Equity Funds

Fund groups	No. of Funds	Alpha	Std.	5 th percentile	95 th Percentile
Large Growth	234	-2.71	3.36	-8.47	2.79
Large Blend	272	-1.06	2.36	-5.02	2.53
Large Value	218	-1.53	2.33	-5.60	2.06
Medium Growth	135	-2.95	4.08	-8.59	2.27
Medium Blend	50	0.52	3.82	-5.71	7.19
Medium Value	43	-0.46	3.46	-5.28	5.06
Small Growth	98	-1.89	5.86	-10.04	4.66
Small Blend	58	0.11	4.01	-5.98	5.76
Small Value	37	0.55	4.39	-6.19	7.69
Large-cap Funds		-1.77	2.69		
Medium-cap Funds		-0.97	3.79		
Small-cap Funds		-0.41	1.75		

Alpha is the average alphas of funds in that asset category.

6.4 Nonparametric Estimation of Density Function and Normality Check

The finite normal mixture model is motivated by an empirical check of the shapes of kernel densities and the normality tests of the distributions.

6.4.1 Nonparametric Estimation of Density Function

It is helpful to check the shapes of the distributions by non-parametric method. We estimate the kernel density of the distributions of five groups of alphas, which are shown from figure 11.1 to 15.3. The kernel estimator is given by,

$$\hat{f}_{\text{ker}}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$
(6.2)

where $K(t)$ is kernel and h is the smoothing parameter. In the kernel estimator we use standard normal density as the kernel. Although other kernels are available and more efficient than the normal kernel, such as Epanechnikov, biweight, and triangle (Epanechnikov, 1969), the effects from the choice of kernel are reduced after averaging. The difference arising from the kernel selection is small. Since our purpose here is to have a general idea of the alphas' distributions, we choose standard normal kernel density for convenience.

We evaluate the kernel for each X_i in the domain,

$$K\left(\frac{x - X_i}{h}\right), \quad i = 1, 2, \dots, n$$

(6.3)

where n is the number of alphas in the group. So we have n curves, one for each alpha in the performance group. Then we weight each curve by $1/h$ and average the weighted curves to obtain the kernel density estimate of alphas in the group.

In figure 11.1 to figure 15.3 we also impose the normal distribution on the kernel density curve to see the difference between the actual distribution and the normal distribution. Kernel density is denoted by the red line, whereas the normal fit is denoted by the blue line. From the figures it is obvious that the distributions are not normal. We observe multimodal shapes and relative high frequency around mean.

This shape is a strong indication that the true distribution is possibly a mixture of several component distributions, especially when the sample size is sufficiently large. In our study, the samples used for kernel density estimation are around 200. That is sufficiently large.

6.4.2 Formal Normality Test

To confirm our judgment from kernel density estimation, we also conduct formal normality check using the Jarque-Bera test and Lilliefors test (Conover, 1980). The Jarque-Bera test checks whether the sample skewness and kurtosis are unusually different from their expected values: zero and three, measured by chi-square statistic. Lilliefors test compares the empirical distribution of the sample with a normal distribution having the same mean and variance as the sample. The test is similar to the Kolmogorov-Smirnov test (Conover, 1980), but it adjusts for the fact that the parameters of the normal distribution are estimated from the sample rather than specified in advance. We set the significance level at 5%. We have five groups of funds classified according to asset composition. We further divide the nine-year sample period into three periods, each in three years. From table 4, we find that almost all the distributions show negative skewness and high kurtosis, indicating asymmetry and fat tails about the distributions of alphas. 16 out of 18 Jarque-Bera test statistics show significant values with p values less than 5%, except for large growth funds and large value funds in the period from November 1995 to November 1998. Regarding Lilliefors test, all of the distributions show significant values either at the 5% or 10% level. Our empirical evidence strongly suggests

non-normality of the distributions, which are consistent with the shapes of kernel density that we observed. The shapes of the distributions and the test results encourage us to explore group-structure of distributions by finite normal mixture model.

Figure 11.1
Large Growth Funds (234): Nov. 1995-Nov. 1998

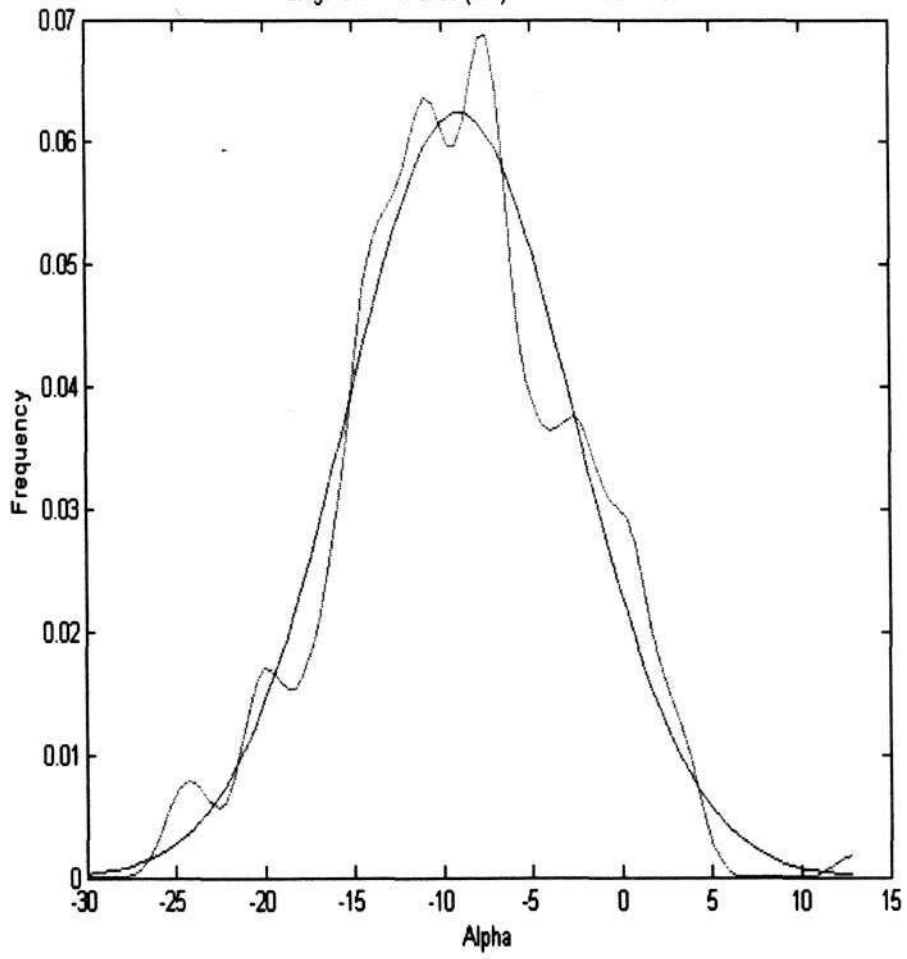


Figure 11.2
Large Growth Funds (234); Nov.1998-Nov.2001

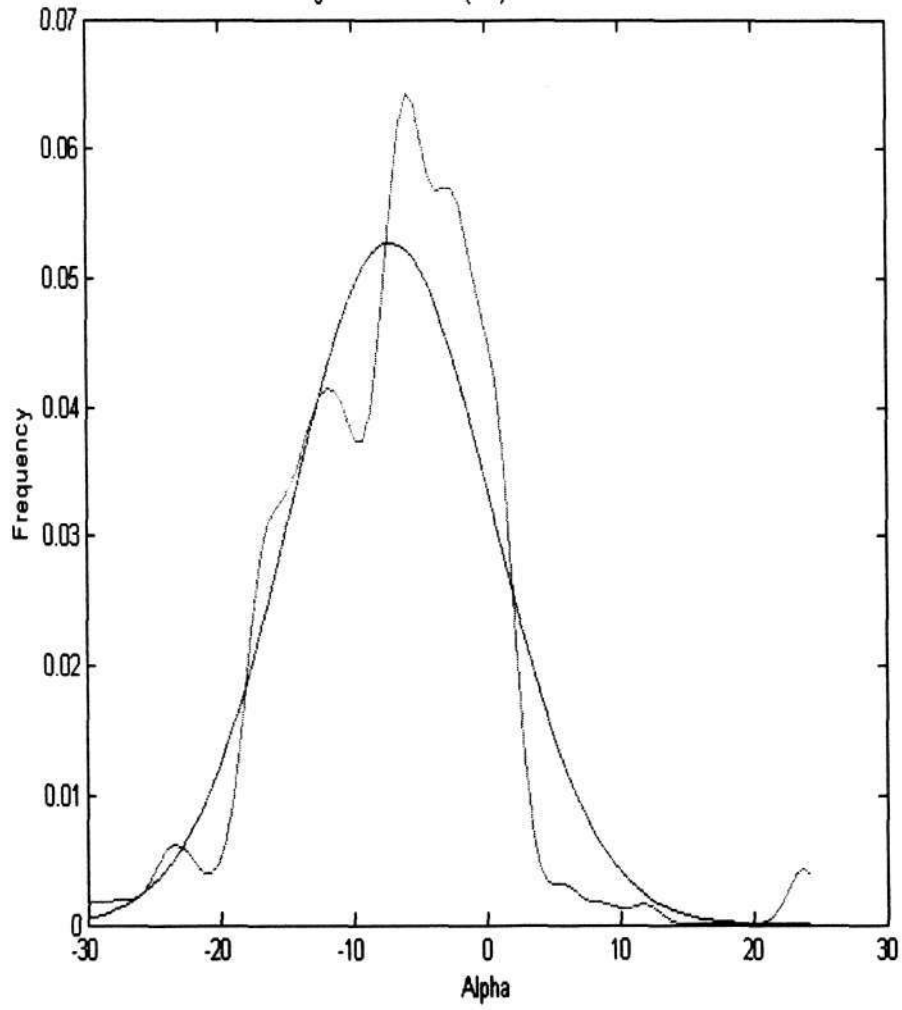


Figure 11.3
Large Growth Funds (234): Nov.2001-Nov.2004

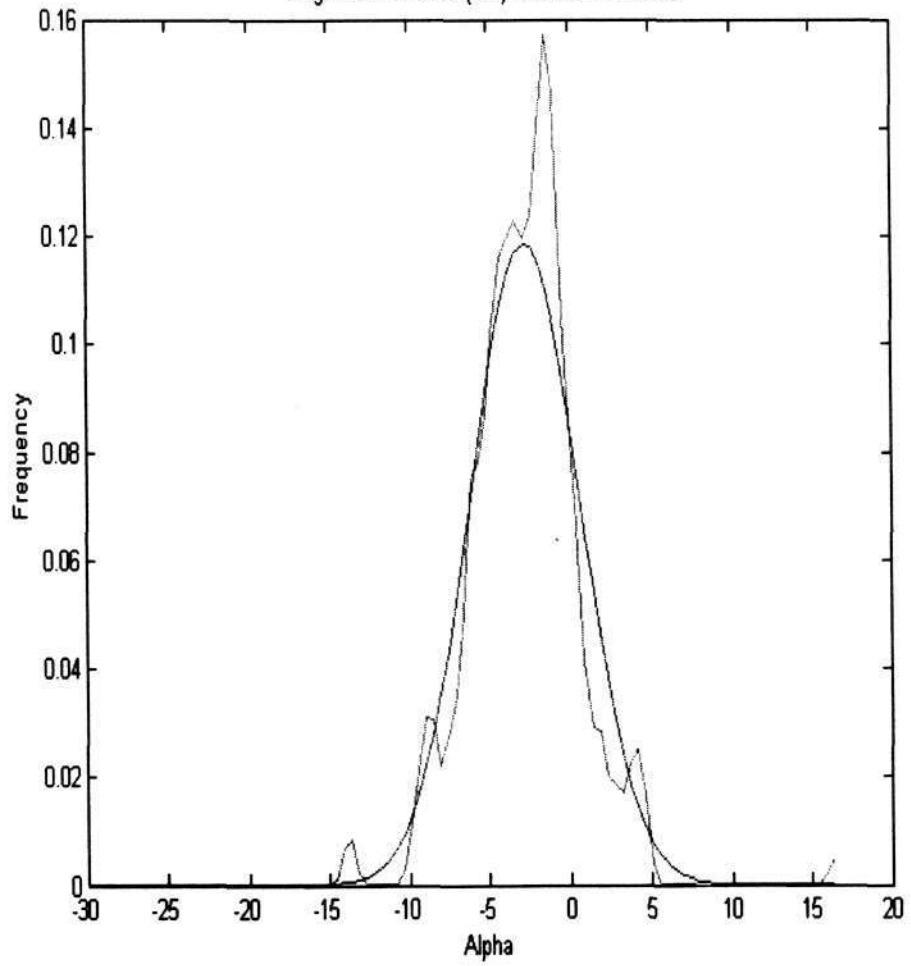


Figure 12.1
Large Blend Funds (272): Nov.1995-Nov.1998

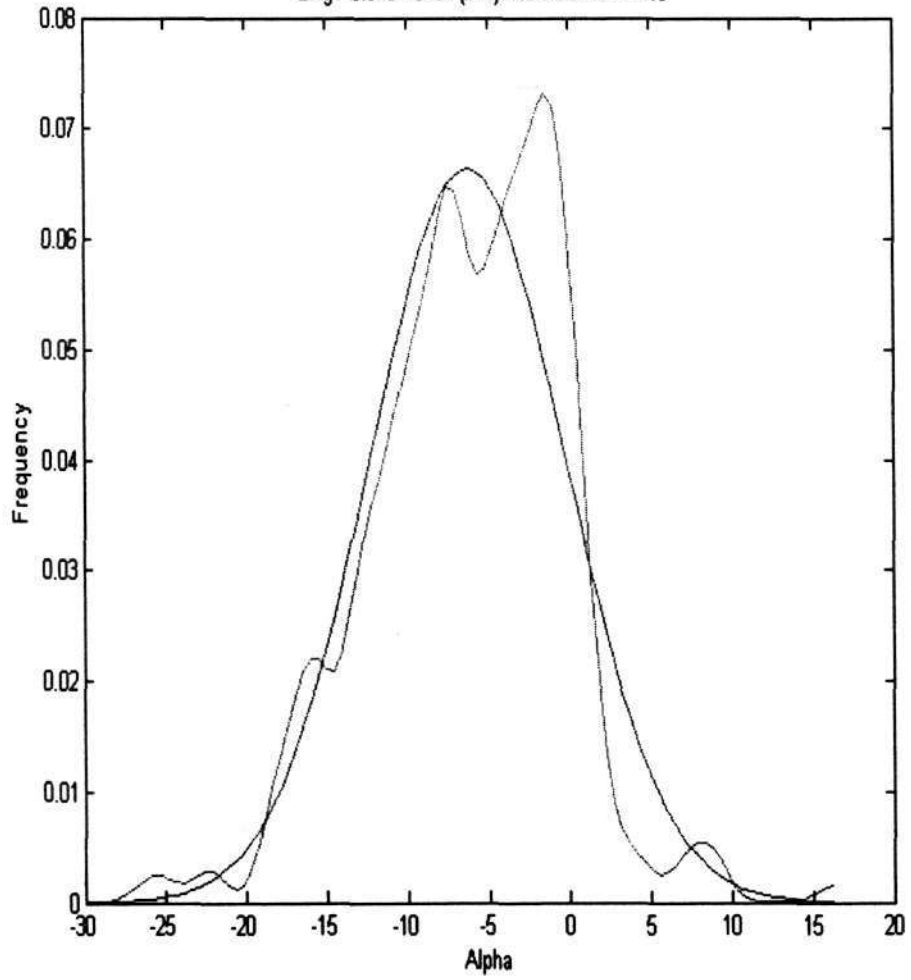


Figure 12.2
Large Blend Funds (272): Nov.1998-Nov.2001

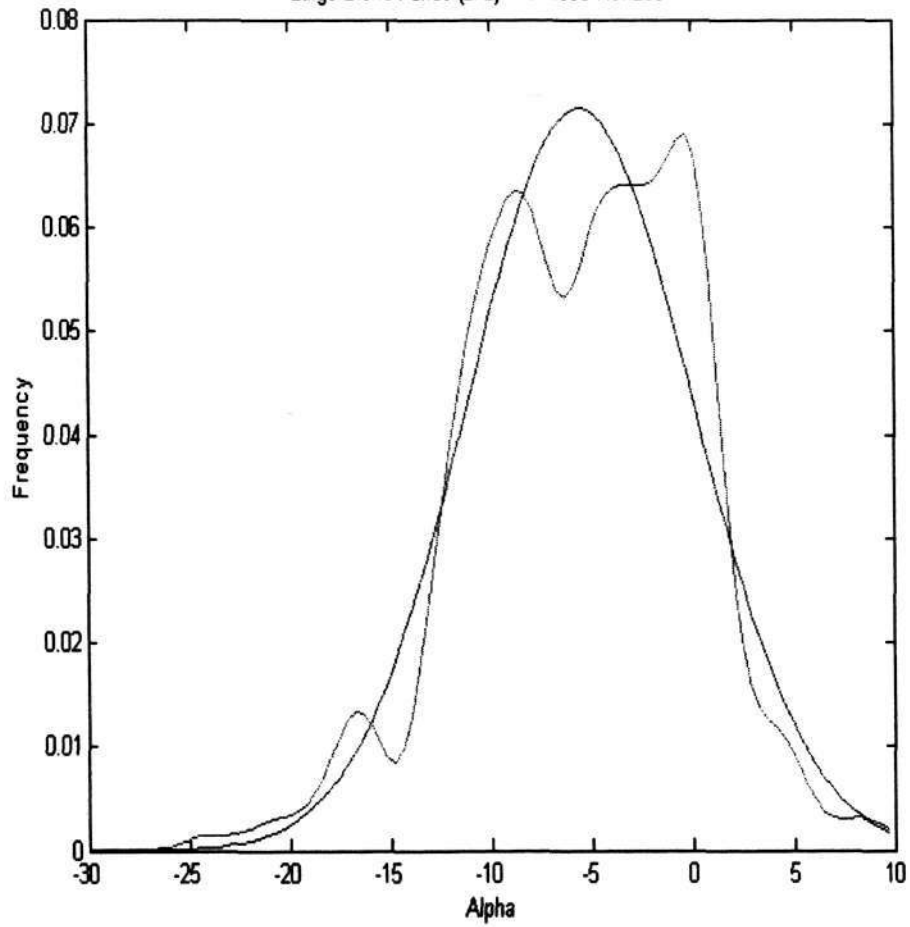


Figure 12.3
Large Blend Funds (272): Nov.2001-Nov.2004

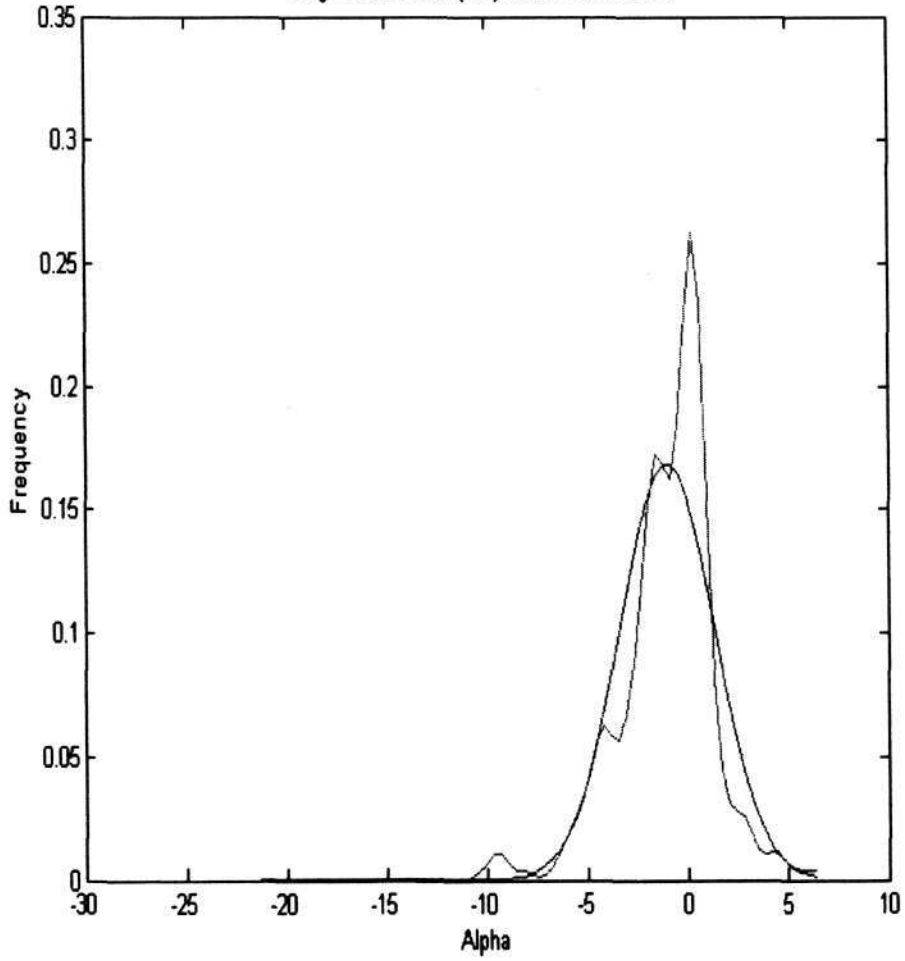


Figure 13.1
Large Value Funds (218): Nov. 1995-Nov. 1998

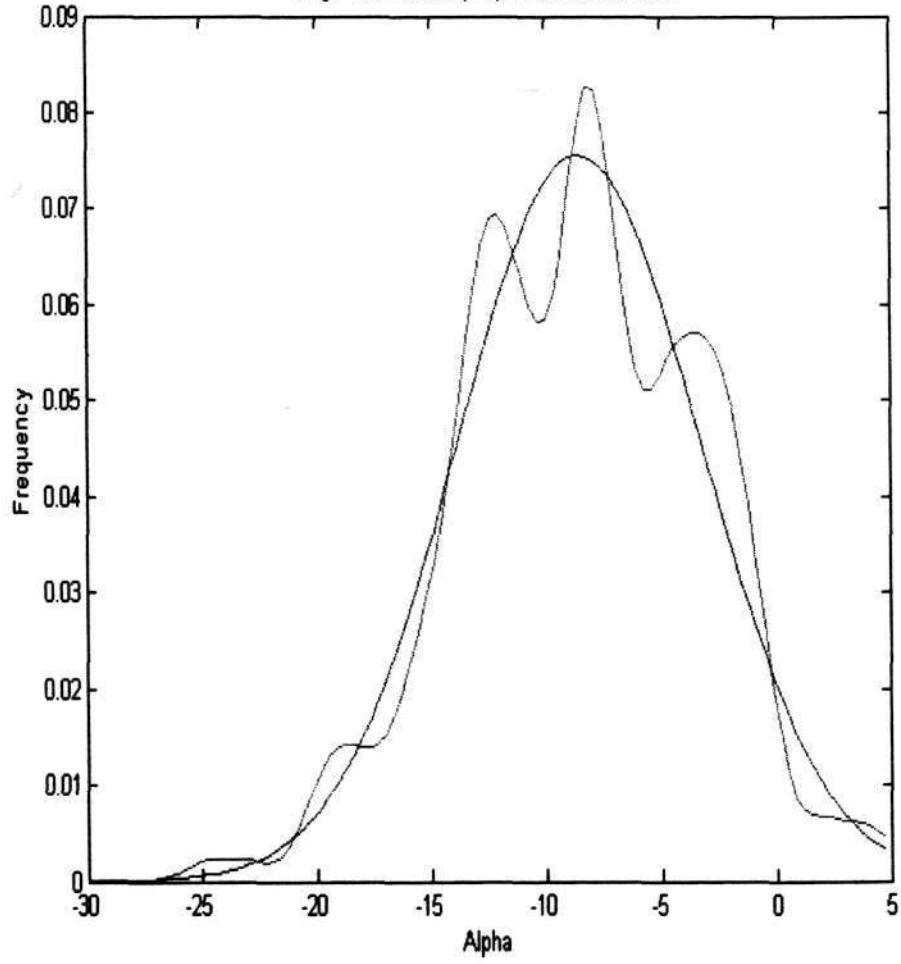


Figure 13.2
Large Value Funds (218): Nov.1998-Nov.2001

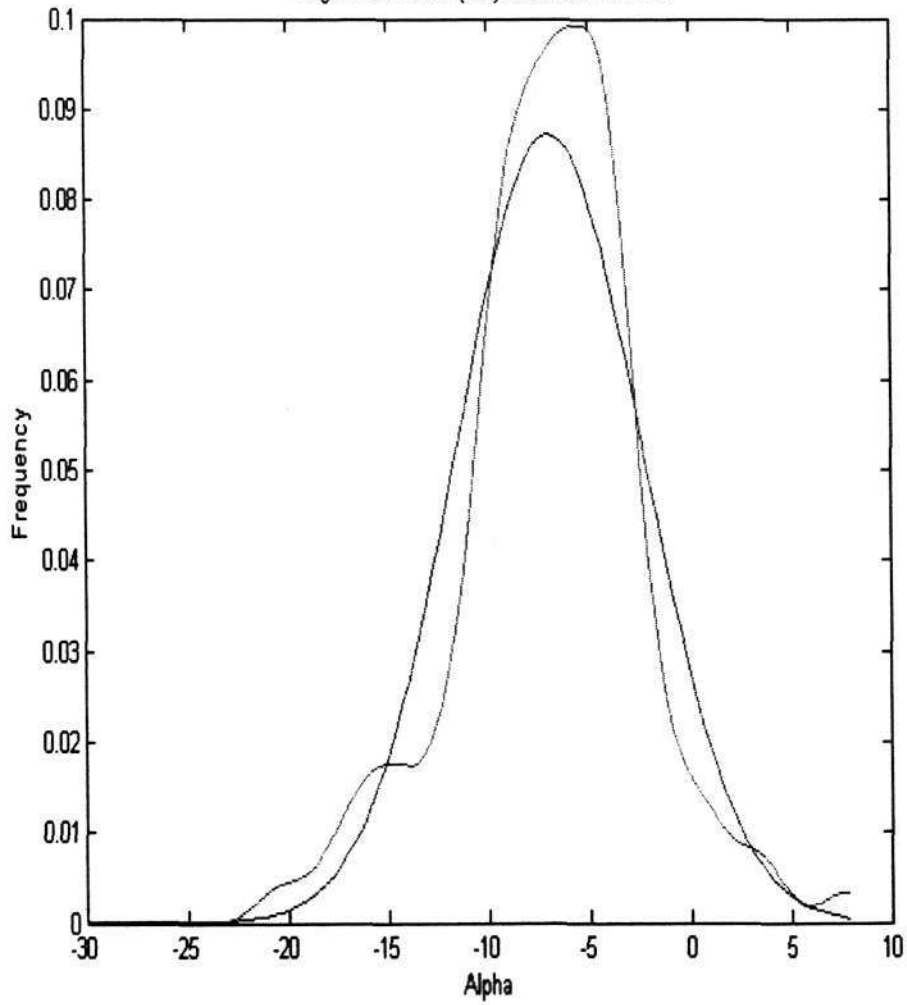


Figure 13.3
Large Value Funds (218): Nov.2001-Nov.2004

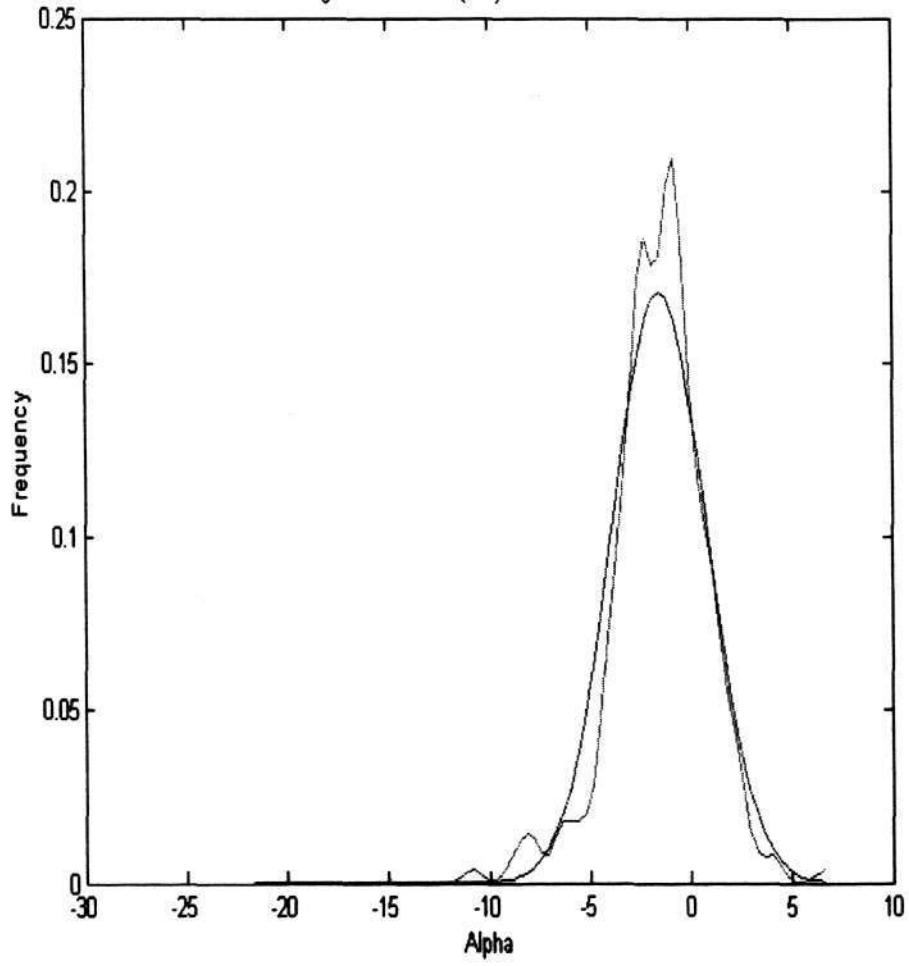


Figure 14.1
Medium Funds (228): Nov. 1995-Nov. 1998

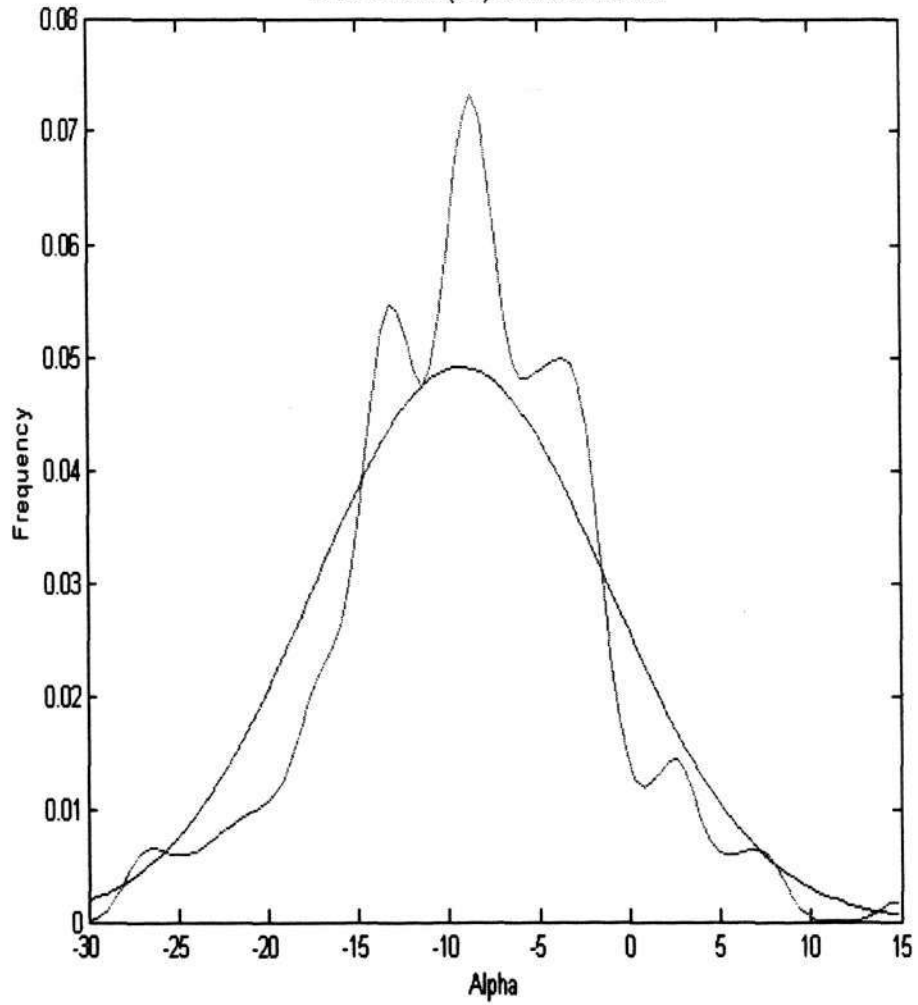


Figure 14.2
Medium Funds (228): Nov.1998-Nov.2001

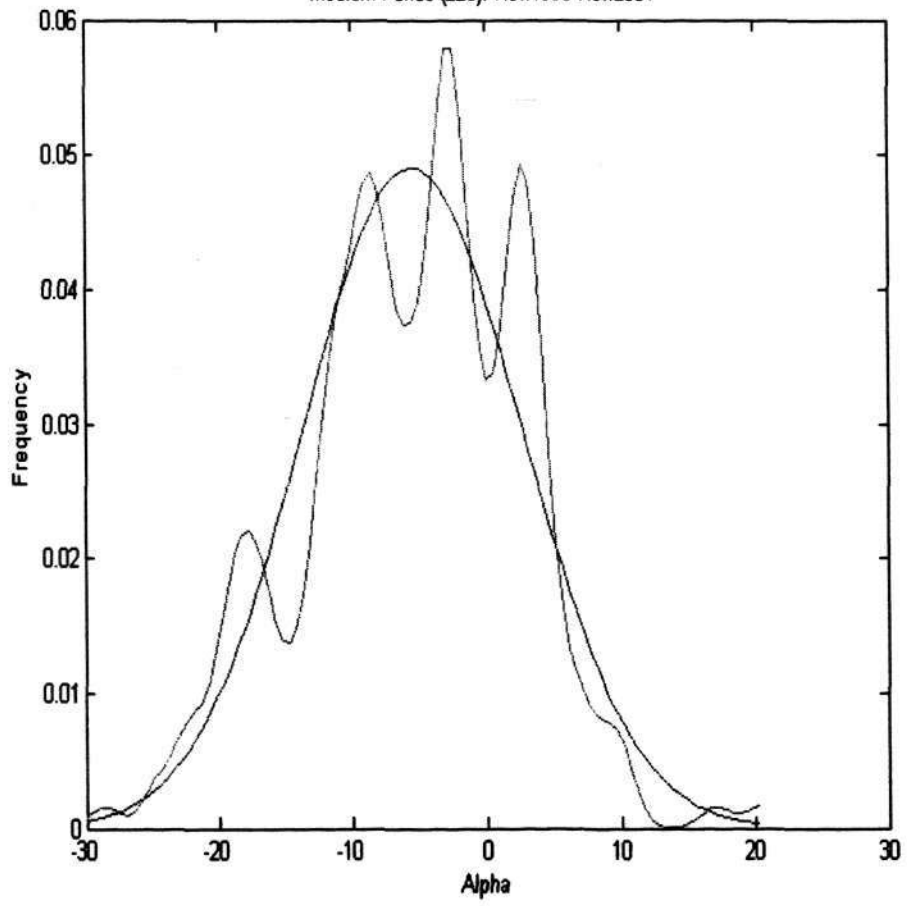


Figure 14.3
Medium Funds (228): Nov.2001-Nov.2004

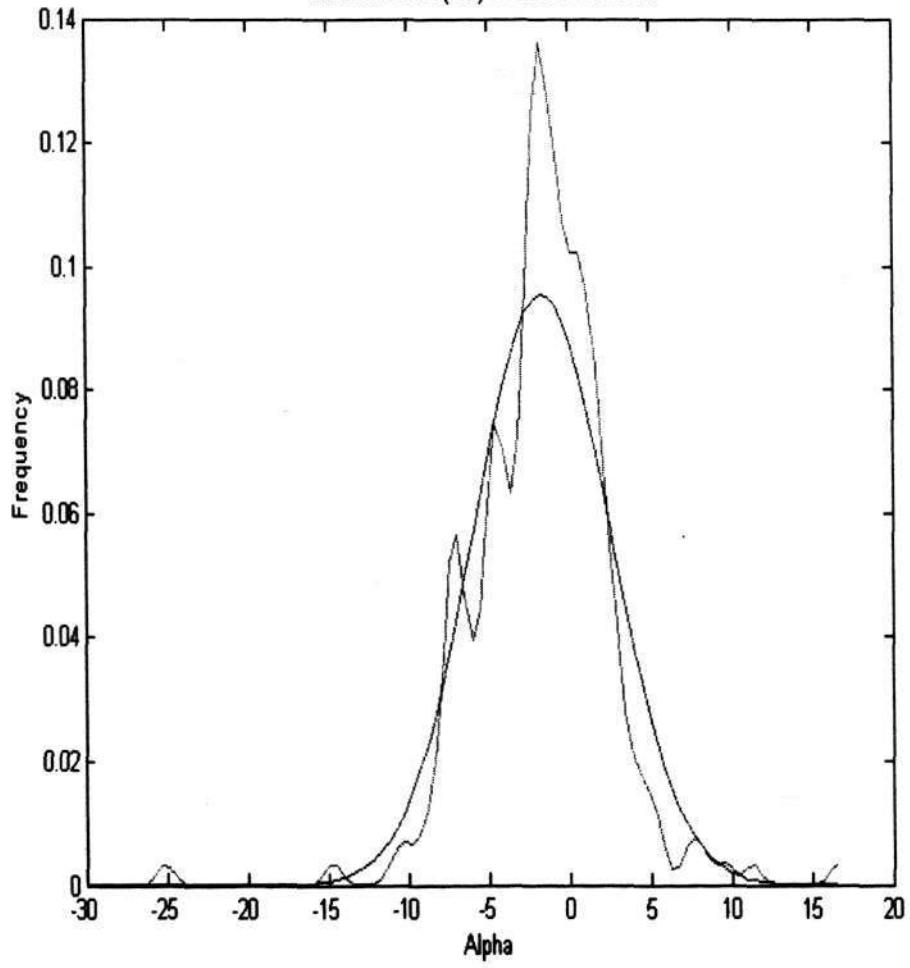


Figure 15.1
Small Funds (193): Nov.1995-Nov.1998

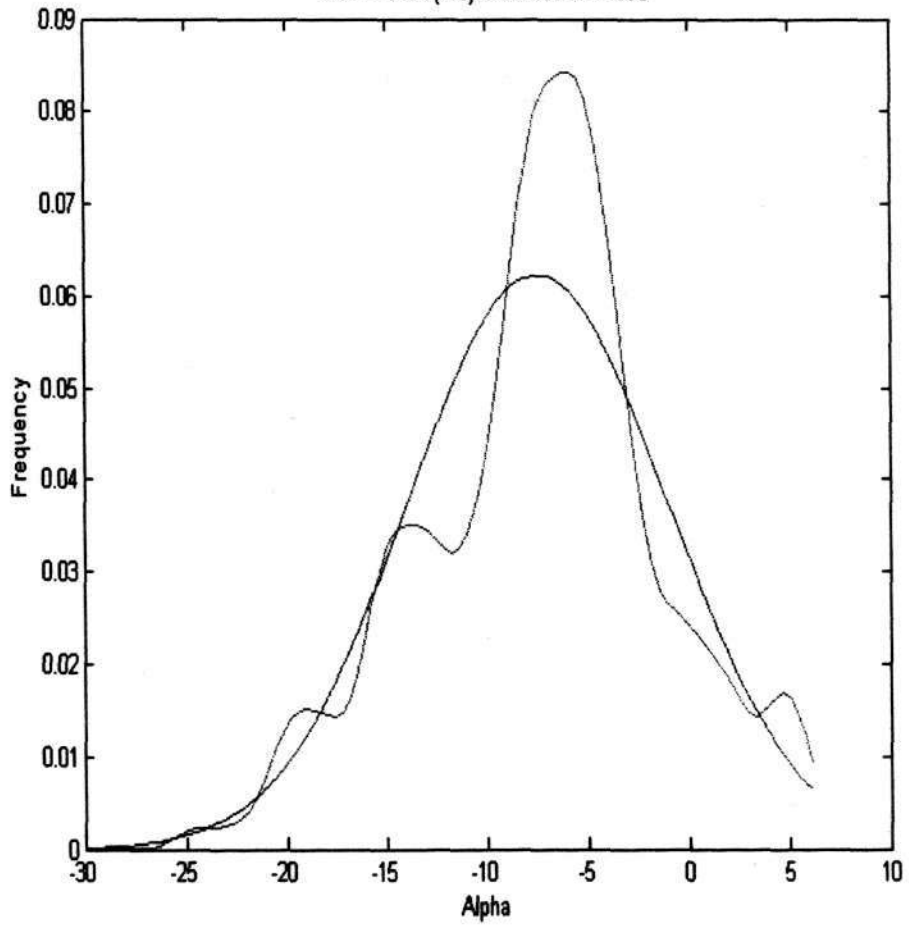


Figure 15.2
Small Funds (193): Nov.1998-Nov.2001

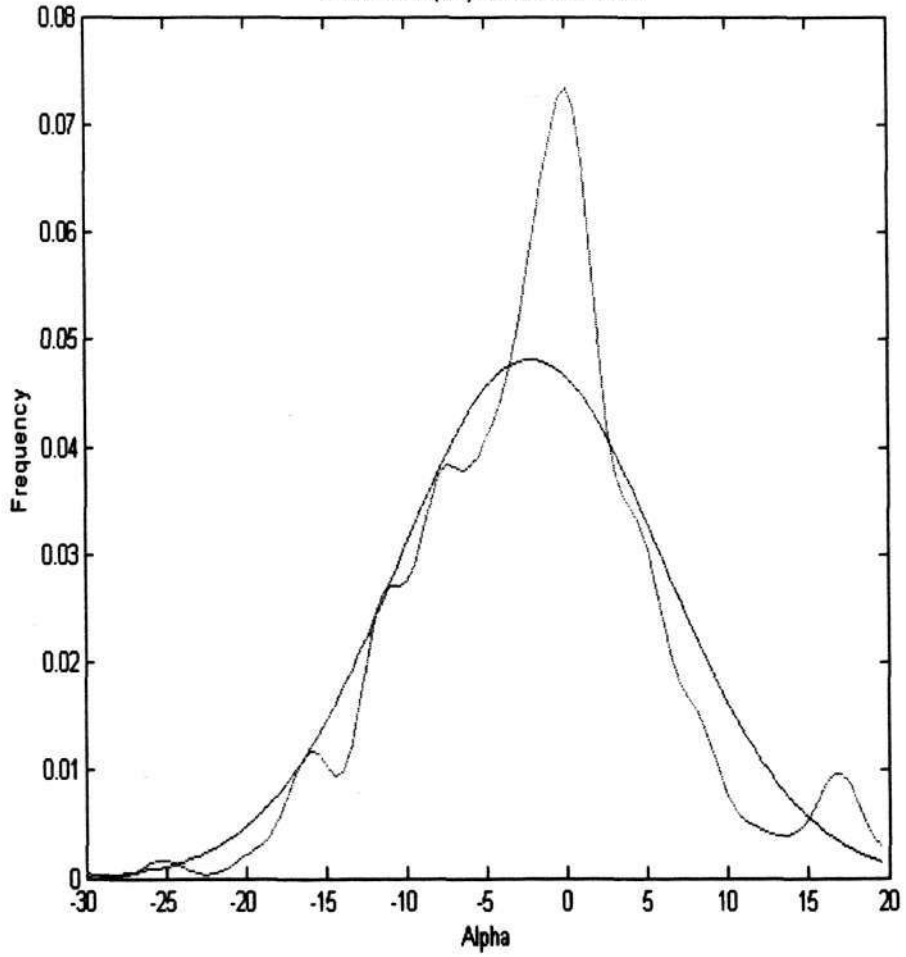


Figure 15.3
Small Funds (193): Nov.2001-Nov.2004

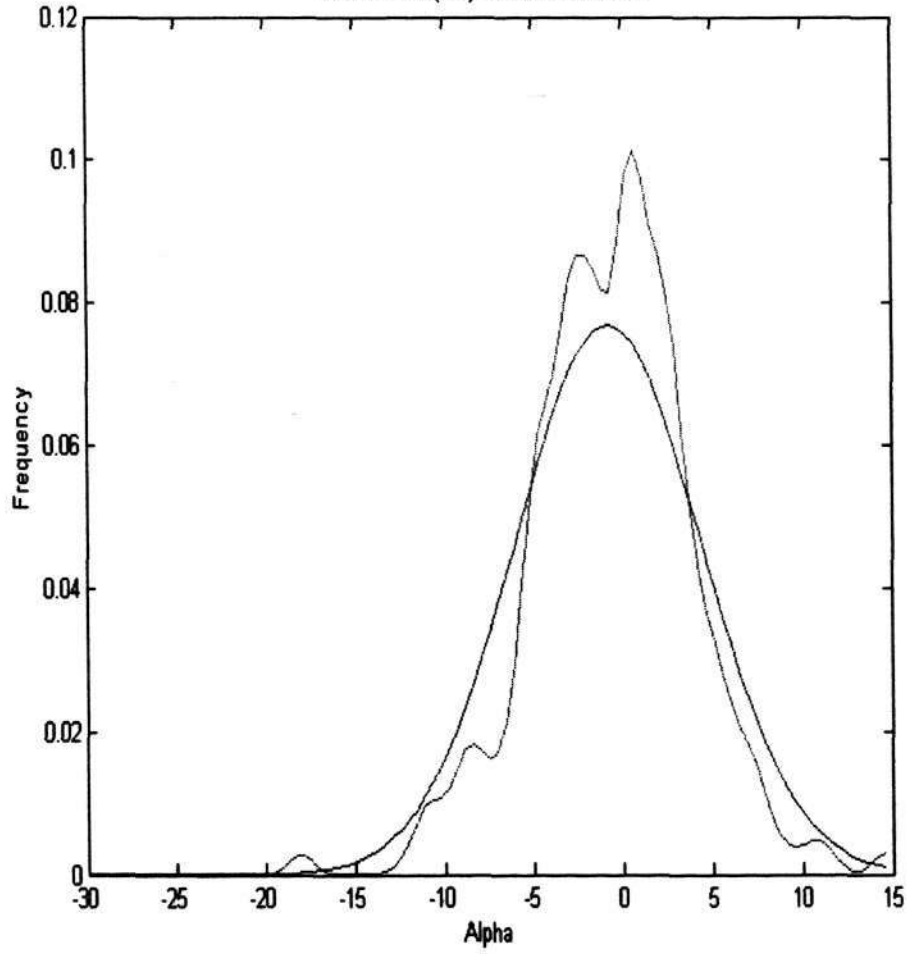


Table 4: Normality Tests for the Distributions of Alphas

Panel I		Large Growth Funds (234)		
Sample period	Skewness	Kurtosis	Jarque-Bera Test Statistic	Lilliefors Test Statistic
Nov. 1995-Nov. 1998	-0.0052	3.0156	0.0047769	0.0114*
Nov. 1998-Nov. 2001	0.4170	5.6357	72.977**	0.0754**
Nov. 2001-Nov. 2004	0.5069	7.2660	182.21**	0.063**
Panel II		Large Blend Funds (272)		
Sample period	Skewness	Kurtosis	Jarque-Bera Test Statistic	Lilliefors Test Statistic
Nov. 1995-Nov. 1998	-0.2128	3.7858	8.5112**	0.0601**
Nov. 1998-Nov. 2001	-0.3187	3.1451	4.7232*	0.0519*
Nov. 2001-Nov. 2004	-0.7043	5.0482	68.079**	0.091**
Panel III		Large Value Funds (218)		
Sample period	Skewness	Kurtosis	Jarque-Bera Test Statistic	Lilliefors Test Statistic
Nov. 1995-Nov. 1998	-0.1660	2.9213	1.089	0.0485*
Nov. 1998-Nov. 2001	-0.1581	4.1165	11.468**	0.0758**
Nov. 2001-Nov. 2004	-0.5287	1.9708	13.689**	0.0712**
Panel IV		Medium Funds (228)		
Sample period	Skewness	Kurtosis	Jarque-Bera Test Statistic	Lilliefors Test Statistic
Nov. 1995-Nov. 1998	-1.6448	13.0710	1043.3**	1606.2**
Nov. 1998-Nov. 2001	-0.1720	3.0039	1.1136	138.11**
Nov. 2001-Nov. 2004	-0.3730	8.2951	264.33**	6201.7**
Panel V		Small Funds (193)		
Sample period	Skewness	Kurtosis	Jarque-Bera Test Statistic	Lilliefors Test Statistic
Nov. 1995-Nov. 1998	-0.5755	4.5955	29.757**	1606.2**
Nov. 1998-Nov. 2001	-0.5124	5.6249	61.293**	138.11**
Nov. 2001-Nov. 2004	-1.8137	14.8750	1209**	6201.7**

* and ** indicate that the test statistics are significant at 10% and 5% levels respectively.

Totally we have 1145 funds under study from November 30, 1995 to November 30, 2004. The sample period is further divided into three sub-periods, from November 30, 1995 to November 30, 1998, from November 30, 1998 to November 30, and from November 30, 2001 to November 30, 2004. The 1145 funds are classified into five groups according to asset composition, that is, Large Growth funds, Large Blend funds, Large Value funds, Medium funds, and Small funds. In each group we have around 200 funds that are sufficient large for density estimation and distribution investigation.

6.5 Specification of Finite Normal Mixture Model

To correctly estimate the finite mixture normal model, we have to specify the number of components in the model first. This is a difficult task, since the likelihood ratio test statistic $-2\log(\lambda)$ does not follow the usual asymptotic Chi-squared distribution under H_0 as shown by Titterington, et al. (1985) and McLachlan and Basford (1988). We use parametric bootstrap procedures to find the empirical distribution under H_0 . With this information we can assess the p value of the likelihood ratio statistics to identify the smallest number of components that are consistent with the data.

Totally we have 1145 funds under study with monthly observations from November 30, 1995 to November 30, 2004. We divide the whole sample periods into three sub-periods. Each of them is exactly three years. We further divide the 1145 funds into five categories according to their asset compositions, i.e. Large Growth funds (234), Large Blend funds (272), Large Value funds (218), Medium funds (228), and Small funds (193). We will study the number of performance groups in each category and their performance dynamics during the whole sample period.

6.5.1 Large Growth Funds

We have 234 funds in this category. In panel I of table 5, during the first sample period, the bootstrapped p value under $H_0 : g = 1$ versus $H_1 : g = 2$ is quite large. It is 0.9805, which implies that there is only one group in this category. This is

consistent with Jarque-Bera test statistics in table 4. However, when using Lilliefors test, we find the null hypothesis of normality testing is rejected. We further check the kernel density of the distribution, which suggests that it is possible to have more than one group in the data, since there is an obvious multimodal shape of the alphas. So we continue the bootstrap procedures and increase the number of components one by one. The p values suggest that we have five performance groups during the first sample period. After identifying the number of components, we further test the normality under $H_0 : g = 1$ versus $H_1 : g = g_0$, where g_0 is 5 in this occasion. We strongly reject the null hypothesis of univariate normal distribution. This confirms our initial judgment that there is possibly more than one group in the data.

For the next two sample periods, the test for normality by bootstrap procedures is consistent with formal normality tests in table 4. We observe that p values are very small, which are 0.0101 and 0.0000 respectively under $H_0 : g = 1$ versus $H_1 : g = 2$. We continue the bootstrap procedure sequentially until we arrive at a relatively large p value. We do not stop the bootstrap procedures until we find a p value larger than 0.15. We have four performance groups and two performance groups respectively in the next two sub-periods. We also observe that from 1995 to 2004 the number of performance groups is becoming smaller and smaller, from 5 groups to 2 groups. It implies that for Large Growth funds, the market becomes more efficient, since the merge of information sets implies that the information have been well transmitted among managers. When market is efficient, the information

Table 5: The Specification of Finite Normal Mixture Models

Panel I Large Growth Funds (234)										
Sample period	$H_{0, g=1}$	$H_{0, g=2}$	$H_{0, g=3}$	$H_{0, g=4}$	$H_{0, g=5}$	$H_{0, g=6}$	$H_{0, g=7}$	$H_{0, g=1}$	$H_{0, g=2}$	
	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	
Nov. 1995–Nov. 1998	0.0517	0.9805	6.0963	0.1362	8.6183	0.0814	8.1209	0.0905	1.1011	0.7322
Nov. 1998–Nov. 2001	22.8962	0.0101	17.4381	0.0632	38.7095	0.0266	4.4990	0.3400		
Nov. 2001–Nov. 2004	22.5703	0.0000	7.5608	0.2000						
	$g_0 = 5$		$g_0 = 4$		$g_0 = 2$		$g_0 = 2$		$g_0 = 2$	
Panel II Large Blend Funds (272)										
Sample period	$H_{0, g=1}$	$H_{0, g=2}$	$H_{0, g=3}$	$H_{0, g=4}$	$H_{0, g=5}$	$H_{0, g=6}$	$H_{0, g=7}$	$H_{0, g=1}$	$H_{0, g=2}$	
	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	
Nov. 1995–Nov. 1998	19.8082	0.0000	21.4733	0.0101	6.3491	0.1137	21.3810	0.0062	3.4281	0.5167
Nov. 1998–Nov. 2001	24.3140	0.0005	13.8799	0.0202	5.6736	0.2205				
Nov. 2001–Nov. 2004	44.2023	0.0053	36.6019	0.0056	43.6893	0.0000	9.4624	0.1728		
	$g_0 = 5$		$g_0 = 3$		$g_0 = 4$		$g_0 = 6$		$g_0 = 4$	
Panel III Large Value Funds (218)										
Sample period	$H_{0, g=1}$	$H_{0, g=2}$	$H_{0, g=3}$	$H_{0, g=4}$	$H_{0, g=5}$	$H_{0, g=6}$	$H_{0, g=7}$	$H_{0, g=1}$	$H_{0, g=2}$	
	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	
Nov. 1995–Nov. 1998	6.6799	0.1616	9.0639	0.0636	13.7195	0.0262	15.3633	0.0000	13.7571	0.0360
Nov. 1998–Nov. 2001	9.1878	0.0571	15.7751	0.0380	7.3046	0.2294				
Nov. 2001–Nov. 2004	18.9036	0.0051	5.2583	0.1910						
	$g_0 = 6$		$g_0 = 3$		$g_0 = 5$		$g_0 = 6$		$g_0 = 3$	
Panel IV Medium Funds (228)										
Sample period	$H_{0, g=1}$	$H_{0, g=2}$	$H_{0, g=3}$	$H_{0, g=4}$	$H_{0, g=5}$	$H_{0, g=6}$	$H_{0, g=7}$	$H_{0, g=1}$	$H_{0, g=2}$	
	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	
Nov. 1995–Nov. 1998	49.0675	0.0051	1.8506	0.3301	8.8283	0.0900	18.8360	0.0106	1.5139	0.7396
Nov. 1998–Nov. 2001	5.4078	0.2113	8.5383	0.1261						
Nov. 2001–Nov. 2004	35.5565	0.0000	1.2821	0.7910						
	$g_0 = 2$		$g_0 = 5$		$g_0 = 2$		$g_0 = 5$		$g_0 = 2$	
Panel V Small Funds (193)										
Sample period	$H_{0, g=1}$	$H_{0, g=2}$	$H_{0, g=3}$	$H_{0, g=4}$	$H_{0, g=5}$	$H_{0, g=6}$	$H_{0, g=7}$	$H_{0, g=1}$	$H_{0, g=2}$	
	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	P Value	$-2\log(\lambda)$	
Nov. 1995–Nov. 1998	17.8413	0.0050	15.6497	0.0319	8.7874	0.1314	11.0463	0.0905	6.5178	0.2482
Nov. 1998–Nov. 2001	17.7571	0.0101	13.5620	0.0785	3.5983	0.4144				
Nov. 2001–Nov. 2004	47.0677	0.0051	2.5999	0.5549						
	$g_0 = 5$		$g_0 = 3$		$g_0 = 2$		$g_0 = 3$		$g_0 = 2$	

Totally we have 1145 funds under study from November 30, 1995 to November 30, 2004. The sample period is further divided into three sub-periods. The 1145 funds are classified into five groups according to asset composition, that is, Large Growth funds, Large Blend funds, Medium funds, and Small funds. In each group we have around 200 funds that are sufficient large for density estimation.

Likelihood ratio test statistics are 2 times of the difference of log likelihoods estimated under null hypothesis and alternative hypothesis, respectively. To assess the p-value of the statistic, parametric bootstrap procedures are applied to formulate the null distribution. We take 200 bootstrap samples from the model fitted under null hypothesis.

is widely disseminated and shared. A consequence of that is a smaller number of information sets we observed in the market.

6.5.2 Large Blend Funds

We find that there are five, three, and four performance groups in three sub-periods, respectively. Under $H_0 : g = 1$ versus $H_1 : g = 2$, all the three likelihood ratios are significant using parametric bootstrap procedures, which are consistent with the results from Jarque-Bera test (Cramer, 1946) and Lilliefors test (Cramer, 1946) in table 4. We double check the univariate normal distribution hypothesis by studying the likelihood ratio under $H_0 : g = 1$ versus $H_1 : g = g_0$, where we find that the statistics are significant in all three occasions with p values as 0.0000, 0.0205, and 0.0000 respectively.

During the first sub-period, we stop the parametric bootstrap procedures when g is five, since the p value is 0.5167. For the second sub-period, we stop the procedure when g is 3, where the p value is obtained as 0.2205. For the third sub-period, we find likelihood ratio is not significant when alternative hypothesis is at $g=5$. So we conclude that in the three periods there are five, three, and four information sets in the market according to sample information. However, we do not observe the significant decrease of the number of information sets from 1995 to 2004 in this asset category.

6.5.3 Large Value Funds

We observe a similar situation here as what we encountered for Large Growth funds. Our result from parametric bootstrap procedure can not reject $H_0 : g = 1$, which is consistent with the result from Jarque-Bera test in table 4, but contradicts the result from Lilliefors test, where we observe a significant statistic as 0.0485. Thus we can not make a conclusion with just this information. The careful check of kernel shows that the distribution is clearly a head-shoulder shape. So we try to increase the number of components using the bootstrap procedure and we find that we actually have six performance groups from 1995 to 1998. When we assess the p value under $H_0 : g = 1$ versus $H_1 : g = 6$, we find that the distribution is not univariate normal. Therefore, we can conclude that the data is a six-component normal fixture model.

In the next two sub-periods, we find that there are three and two performance groups in the market from 1998 to 2001 and from 2001 to 2004 respectively. The p value under $H_0 : g = 2$ versus $H_1 : g = 3$ is 0.0380 in the second sub-period; and the p value under $H_0 : g = 1$ versus $H_1 : g = 2$ is 0.0051 in the third sub-period. So again we find a merge of information sets from 1995 to 2004, which indicates that the market becomes more efficient.

6.5.4 Medium Funds

From 1998 to 2001, we have the same situation as what we observed for Large Growth funds and Large Value funds from 1995 to 1998. We observe non-significance of Jarque-Bera test statistic and likelihood ratio statistic, but a significant Lilliefors test statistic. The shape of the kernel density encourages us to continue the bootstrap procedures, and finally we find that there are actually five performance groups for Medium funds from 1998 to 2001.

Medium funds only have two performance groups from 1995 to 1998, which is different from Large funds where there are around five performance groups during the first sub-period. Although the multimodal shape of the kernel density suggests more components, our formal likelihood ratio tests suggest that the null hypothesis is not rejected significantly. From 2001 to 2004, we again have performance groups that are the same as what we found about Large funds.

6.5.5 Small Funds

Regarding Small funds, we find that there are five, three and two performance groups from 1995 to 1998, 1998 to 2001, and 2001 to 2004 respectively. We clearly see a tendency of smaller number of groups from the first period to the third period. Our result from likelihood ratio is consistent with the formal statistical normality tests and shape of kernel density, which is shown in the last column of table 5

6.5.6 Conclusion

Our empirical findings lead to the following conclusions. Firstly, the number of performance groups is not fixed, it is changing over time. For example, the Large Value funds have six groups during 1995 to 1998, three groups from 1998 to 2001, and 2 groups from 2001 to 2004. Therefore, it is not appropriate to arbitrarily set up five performance groups as Morningstar or just set up two groups, superior performance group and inferior performance group.

Secondly, there is a tendency for merging of information sets. We observe this tendency for all the categories except for Large Blend funds where there is no significant merge of information sets. For the other four categories, there are only two performance groups in the last sub-period, from 2001 to 2004, which means over the years, the fund market is more efficient. There are no obviously separate information sets in the market. On the other hand, the number of performance groups in the first sub-period is the largest for all the five categories, except for Medium funds where the largest number of performance groups occurs from 1998 to 2001.

Thirdly, the p values assessed by parametrically bootstrapping the likelihood ratio statistic, also suggest that the distributions are not univariate normal distributions. It confirms our suspicion that there is group structure in the data. As g_0 is the smallest number of components we identified by parametric bootstrap procedures

under $H_0 : g = 1$ versus $H_1 : g = g_0$, all along we reject the null hypothesis that the distribution is univariate normal.

6.6 Estimation of Finite Normal Mixture Model

After specifying the model, we proceed to estimate finite normal mixture models by EM. In EM, we view the funds as having been generated from a performance group in the asset category. The group information, denoted by component label variable

Table 6: Estimation of Finite Normal Mixture Model

Panel I Large Growth Funds (234)																				
Sample period	Component 1			Component 2			Component 3			Component 4			Component 5			Component 6			Average μ	
	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2		
Nov. 1995-Nov. 1998	0.0251	-21.1860	2.5154	0.0100	-20.1810	0.4105	0.4111	-12.4060	9.5395	0.0726	-7.4637	0.1898	0.1179	-3.7172	20.4730					9.0259
Nov. 1998-Nov. 2001	0.0347	-24.3060	8.4012	0.2830	-13.9370	8.0003	0.6618	-4.0086	15.1250	0.0205	18.5400	47.5960								-7.0613
Nov. 2001-Nov. 2004	0.9655	-2.7522	8.6044	0.0315	-2.4778	85.6620														2.7428
Panel II Large Blend Funds (272)																				
Sample period	Component 1			Component 2			Component 3			Component 4			Component 5			Component 6			Average μ	
	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2		
Nov. 1995-Nov. 1998	0.0145	-23.9790	2.9805	0.0590	-16.6240	1.3638	0.6426	-7.5818	15.8700	0.2452	-1.1960	2.3057	0.0387	6.0602	23.8370					-6.2385
Nov. 1998-Nov. 2001	0.0687	-17.0270	9.1977	0.2596	-9.9311	3.2696	0.6717	-2.6370	14.6770	0.1657	1.8212	1.2155								-5.5191
Nov. 2001-Nov. 2004	0.0146	-9.5706	0.0720	0.7751	-1.3702	4.1363	0.1931	0.3437	0.0587											-1.0355
Panel III Large Value Funds (218)																				
Sample period	Component 1			Component 2			Component 3			Component 4			Component 5			Component 6			Average μ	
	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2		
Nov. 1995-Nov. 1998	0.0092	-23.9180	0.8196	0.0152	-18.9250	0.8102	0.3910	-12.2970	4.6177	0.2076	-7.9121	0.6136	0.3216	-3.1030	3.3992					-8.5389
Nov. 1998-Nov. 2001	0.0921	-15.9000	5.7866	0.8364	-6.3101	8.7666	0.0565	2.6288	9.6406											-6.9229
Nov. 2001-Nov. 2004	0.0381	-7.8953	2.3323	0.9619	-1.2736	3.8660														-1.5256
Panel IV Medium Funds (228)																				
Sample period	Component 1			Component 2			Component 3			Component 4			Component 5			Component 6			Average μ	
	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2		
Nov. 1995-Nov. 1998	0.0250	25.9300	596.8000	0.9750	-8.8561	44.8550														-9.2824
Nov. 1998-Nov. 2001	0.0153	-18.2990	1.3390	0.1111	-9.1117	3.3052	0.6312	-5.8538	75.1740	0.0956	-2.7733	0.3114	0.1131	2.8189	0.8388					-3.308
Nov. 2001-Nov. 2004	0.9313	-1.7569	10.8890	0.0687	-1.2126	105.9200														-1.7216
Panel V Small Funds (193)																				
Sample period	Component 1			Component 2			Component 3			Component 4			Component 5			Component 6			Average μ	
	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2	π_i	μ_i	σ_i^2		
Nov. 1995-Nov. 1998	0.0266	-21.0300	96.9810	0.0579	-19.6880	1.6406	0.1119	-14.1760	2.0973	0.6378	-6.3482	8.1919	0.1179	2.5001	5.2340					-7.5108
Nov. 1998-Nov. 2001	0.0159	-31.3540	50.9000	0.9191	-2.1168	42.7920	0.0317	17.1060	1.3192											-2.2026
Nov. 2001-Nov. 2004	0.0300	-8.9949	269.2200	0.9701	-0.3676	17.1870														-0.82007

Totally we have 1145 funds under study from November 30, 1995 to November 30, 2004. The sample period is further divided into three sub-periods. The 1145 funds are classified into five groups according to asset composition, that is, Large Growth funds, Large Blend funds, Large Value funds, Medium funds, and Small funds. In each group we have around 200 funds that are sufficient large for density estimation.

$Z_j (j = 1, \dots, n)$, is treated as missing information and estimated along with other parameters in the model. The estimates, μ_i and σ_i^2 tell us the expected performance and expected investment risk of the fund if the fund is from performance group $i (i = 1, \dots, g)$. In addition, the estimates of component label variable $Z_j (j = 1, \dots, n)$ are just the posterior probability, which helps us identify the performance group that the fund belongs to, which forms the basis for our fund rating. The estimation results are presented in table 6.

To visualize the underlying structure of finite normal mixture models, we plot dF figures developed by Priebe (1994), along with the curves from the estimated finite mixture models, imposed by kernel density to see the fit of the model. The the dF plot helps us quickly grasp the underlying structure which is not clearly indicated from the model and the fitted curve from the first sight. In dF plot, each circle represents a component. The center of the circle is fixed by mixing proportion π_i (y coordinate) and the expected performance μ_i (x coordinate). The size of the radius of the circle indicates the standard deviation. The structure of finite normal mixture model means the number of components, the component means, and component variances. For example, suppose a three-component finite normal mixture model is,

$$f(y_j; \Psi) = 0.2\phi(y_j; -5, 4) + 0.6\phi(y_j; 0, 1) + 0.2\phi(y_j; 5, 4). \quad (6.4)$$

The dF plot and curve of the model is shown in figures 16.1 and 16.2 respectively.

In the following dF plots, we will use blue curve to denote kernel density estimation, and red curve to denote the fit of the finite normal mixture model.

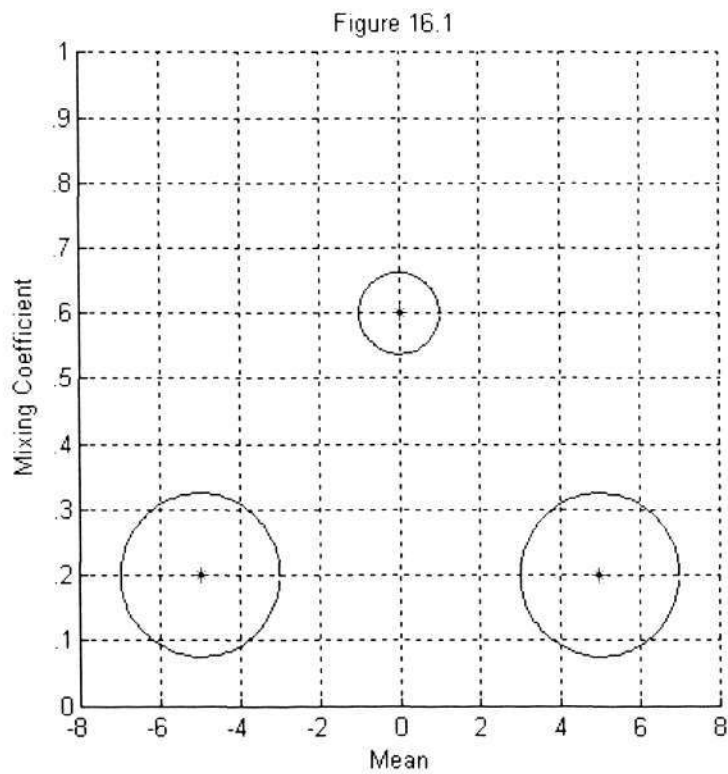
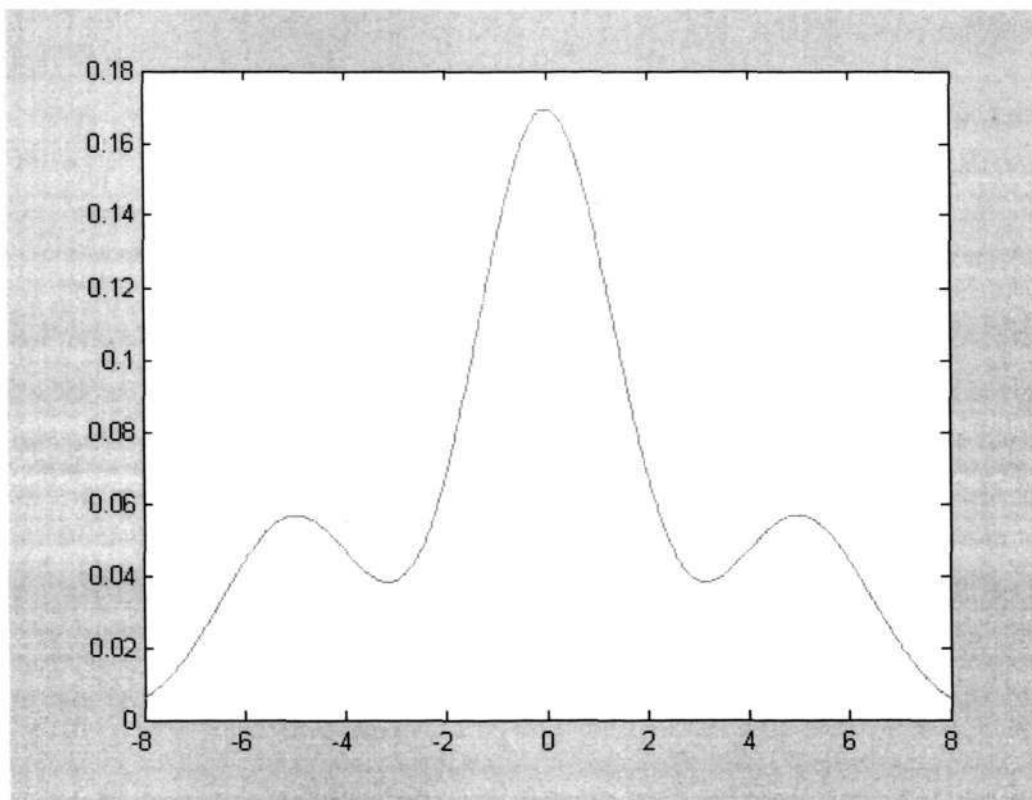


Figure 16.2

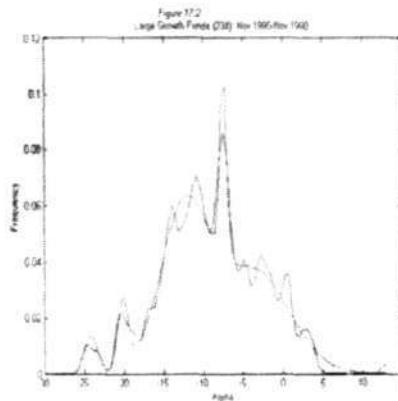
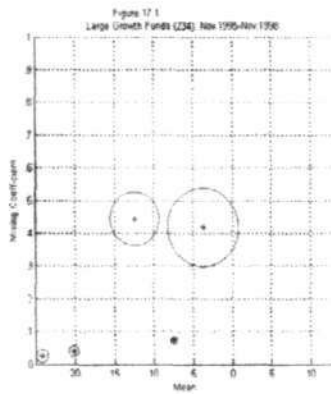


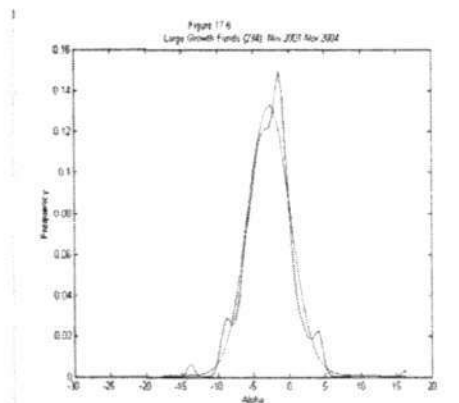
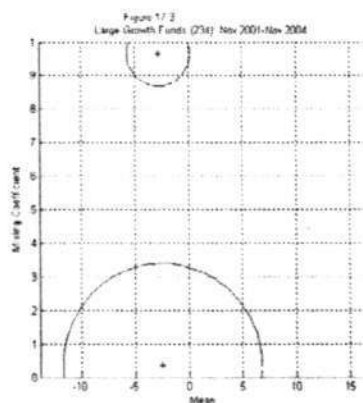
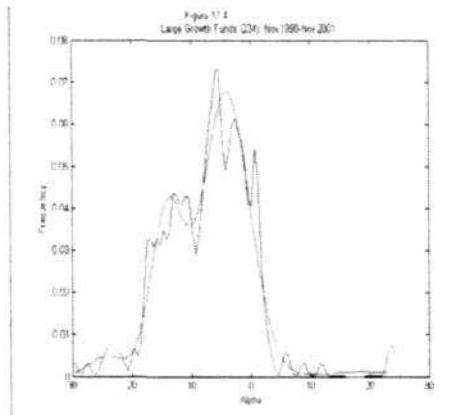
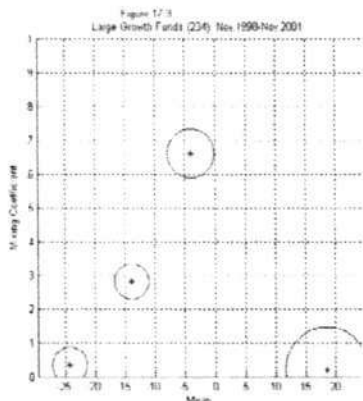
6.6.1 Large Growth Funds

We combine panel I of table 6 and figure 17.1 to figure 17.6 to interpret the results for Large Growth funds. From figures 17.1, 17.3, and 17.5, we can clearly see a merge of information sets, from five groups to two groups. During the first period, we have five performance groups in the market. Among them, there are two inferior ability groups with expected returns of -24% and -20% annually. However, the number of funds of the two groups is small, since the mixing proportions of the two groups are around 2.54% and 4%. The three superior performance groups are closely clustered, with a relatively large variance of 9.55% for the third group and 20.47% for the fifth group. Although there are three superior performance groups in the first period, in practice it should be difficult to separate the three groups.

From figure 17.2, we can see that the model denoted by the red curve is a good fit of the actual distribution, denoted by the blue curve.

Considering the second period from 1998 to 2001, in figure 17.3 the number of information sets decreases to four, and the four performance groups are separated wide enough to practically rate the funds. We have one inferior performance group with an expected return of -24% annually, two average performance groups with expected returns -13.93% for the second group and -4.01% for the third group, and one superior performance group with expected return of 18.54% annually. We also observe that around 90% of funds belong to the two average groups, whereas only a small proportion of funds belong to inferior performance group (3.47%) and superior performance group (2%). This kind of distribution of percentages leads to two fat tails and two separated bumps in





the middle of the shape of distribution of alphas that we observed in the kernel density of figure 17.4.

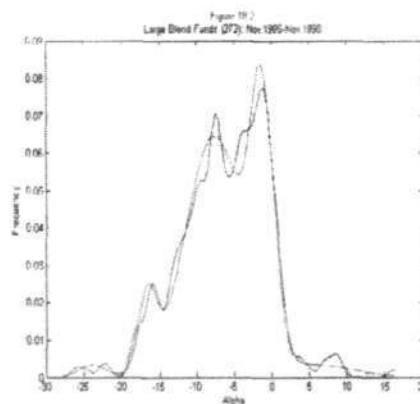
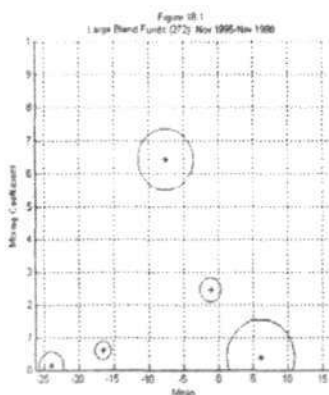
In the third period we observe that all the information sets merge together. Although we have two performance groups, the two groups have almost the same expected returns and only differ in variances, meaning the investment risk of the two groups are different. The group with smaller variance may have more precise private or public information.

From table 6 and figures 18.1 to 18.6, we can say that from 1995 to 2004 the Large Growth fund market become more efficient since we observe a merge of information sets, and the performance of funds are improving from -9.03% to -2.74% annually. Because these alphas are the net performance after expenses the actual performance is around zero from 2001 to 2004 adding back the management fee. This is consistent with efficient market hypothesis as proposed by Fama (1970).

6.6.2 Large Blend Funds

The estimated results are presented in panel II of table 6 and figures from 18.1 to 18.6. We don't observe a significant merge of information sets during the sample period, but we notice that the funds' performance improved from -6.26% to 1.06% annually. In the first period from 1995 to 1998, we have five performance groups, that is, two inferior performance groups with expected returns of -23.98% and -16.62%, two average performance group with expected returns of -7.58% and -1.19%, and one superior performance groups with an expected return of 6.06%.

From the dF plot and estimates in



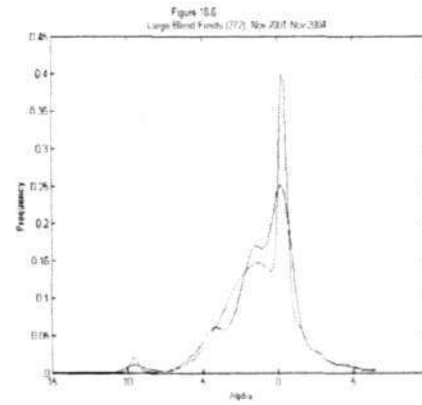
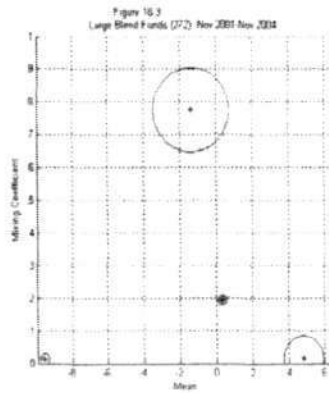
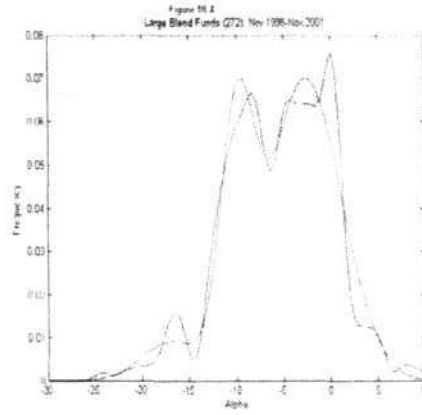
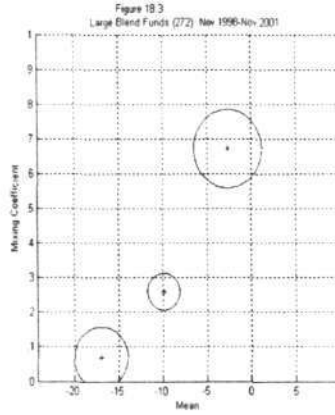


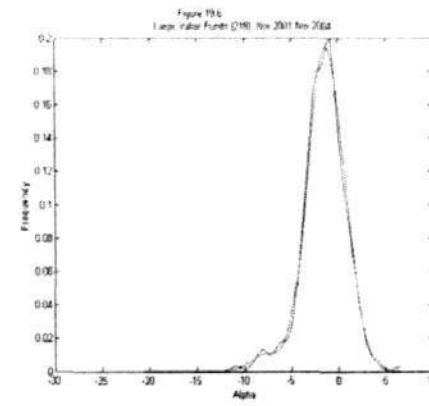
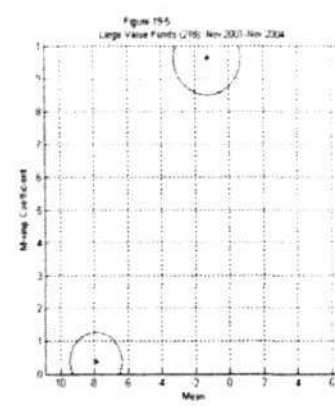
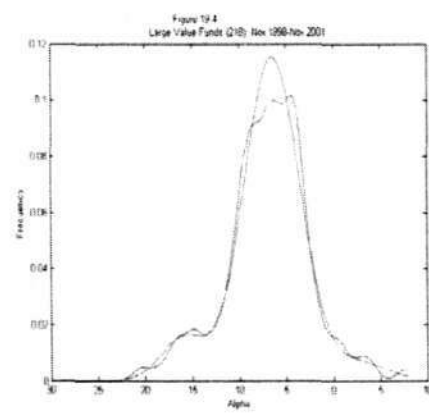
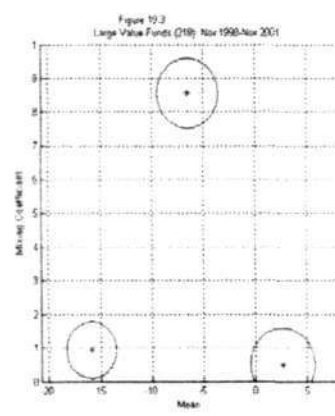
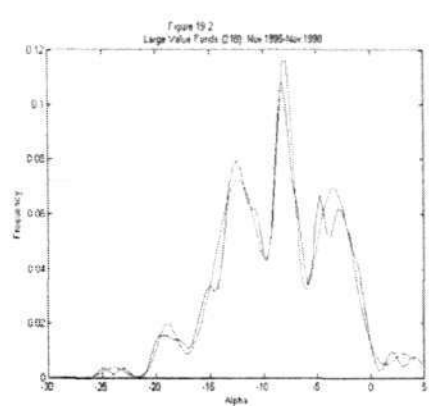
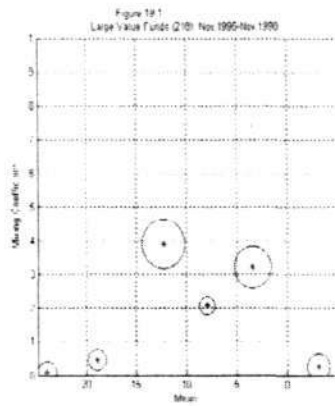
table 6, we know that most funds belong to the average performance groups accounting for around 90% of all the funds in this asset category.

In the second period from 1998 to 2001, we have three performance groups. Among them, we have two inferior performance groups with the expected returns of -17.03% and -9.93%, and one average performance group with an annual return of -2.64%. There is no superior performance group in this period. In the last period

from 2001 to 2004, we find four performance groups. We observe that the groups are widely separated, except for the third group lying between the second group and the fourth group. The number of funds in the inferior and superior performance groups is small, around 3% of funds in these two groups. Most of funds are still in the second and third groups with returns around -1% annually. Although we don't find a merge of information sets, we find the number of funds in inferior and superior groups is smaller. Around 97% of funds belong to the two average performance groups, whereas in the first period we have only around 90% of funds in average performance groups.

6.6.3 Large Value Funds

In this category we observe again a merging of information sets: six performance groups in the first period and two performance groups in the third period. In the first period, we find that the information sets are widely separated. This is also obvious from kernel density estimate in figure 19.2, where we find a head-shoulder shape around the mean of the distribution of alphas, and several bumps in the tails. From dF plot in figure 19.1, we find we have two inferior performance groups with expected returns of - 23.92% and



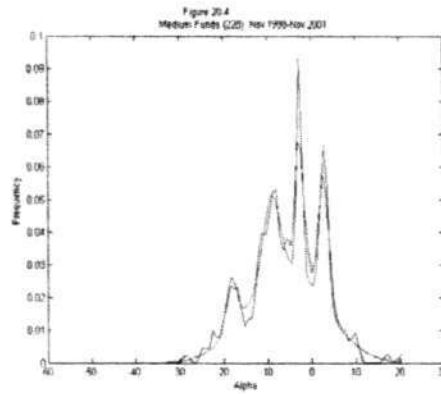
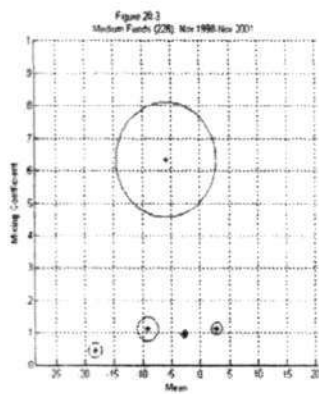
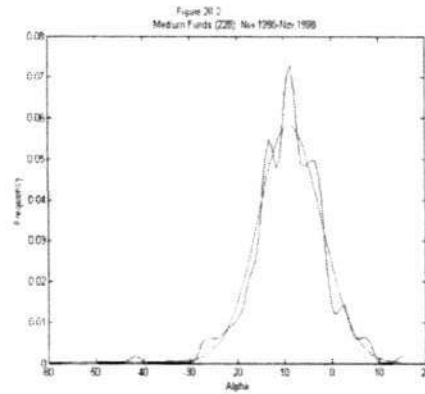
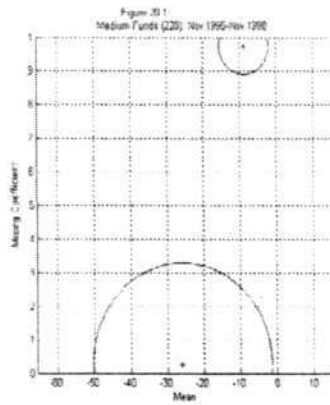
-18.93% accounting for 5.5% of funds, three average performance groups with expected returns of -12.3%, -7.91%, and -3.4% respectively, and one superior performance group with a return of 3.18% accounting for 3.87% of funds. The multimodal shape of kernel density suggests that the market is not efficient. The information in the market is segmented and not well transmitted, which leads to a clear differentiation of the performance that we observed.

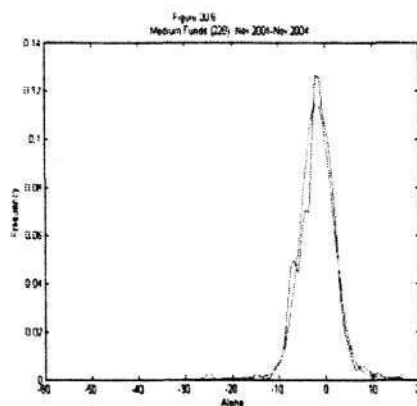
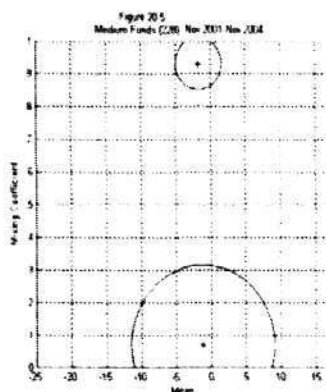
In the second period, the information sets merge into three groups. The first group and the second group of the first period merge into one inferior performance group in the second period with an expected return of -15.9%. The third, the fourth, and the fifth group of the first period merge into one average group in the second period with an expected return of -6.51%. The superior performance group has expected performance of 2.62% annually. In the last period, we find that the superior performance group disappears and more funds are in the average performance group, accounting for 96% of funds in this asset category. In Large Value funds, we also find the funds' performance has improved from -8.54% to -1.53%.

6.6.4 Medium Funds

The situation of Medium funds is very different from Large funds. Firstly, in Medium funds, it is not easy to rate funds practically since the performance groups overlap. Second, there is no clear trend of the merging of information sets. We only

have two performance groups in the first period, five groups in the second period,
and two groups



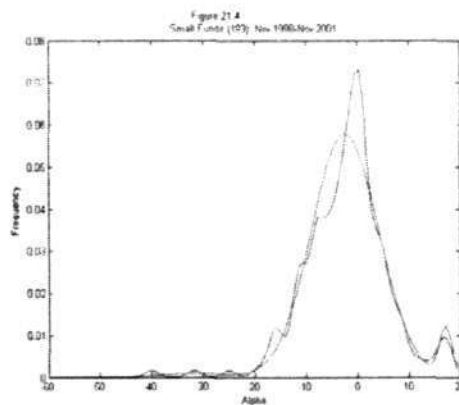
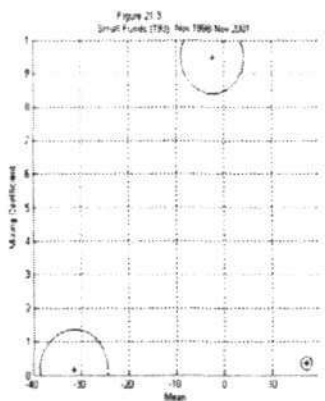
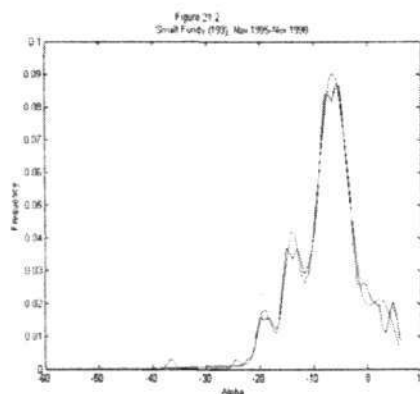
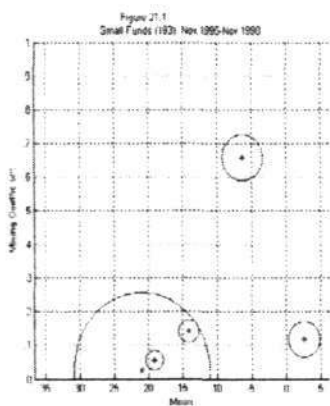


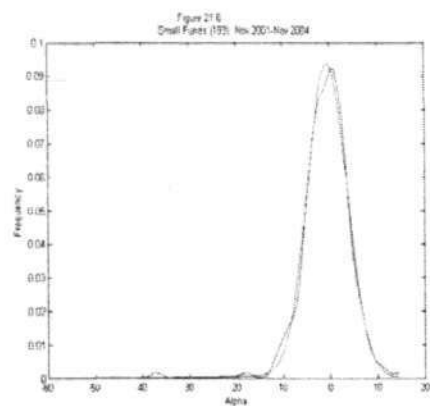
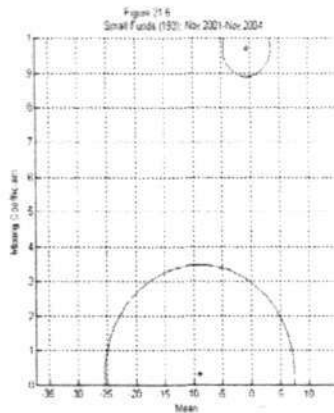
in the last period. In the first period, we observe one inferior performance group with a return of -25.93%, but a small proportion of funds in this group only 2.5%. All the other funds belong to the average performance group with the expected return of -8.86%. In the second period, the performance of average funds decomposed into four small groups. This is confirmed by figure 20.4, where we observe four separated bumps and a long left tail. In the last period, as we observed for Large Growth funds, we find the information sets merge into two closely clustered groups with almost the same expected returns but different variance. We interpret this situation as two groups having the same information but different abilities to analyze the information.

6.6.5 Small Funds

In small funds, we observe a merging of information sets. At the first period, we find three inferior performance groups clustered together with expected returns of -21%, -19%, and -14% respectively. The different variances of the three groups mean that the ability to correctly interpret information is different. We also find an average group with a return of -6.35%, and a superior performance group with a

return of 2.59%. In the second period, the three inferior performance groups merge into a single inferior performance group with a return of -31.55%. However we notice that more funds shift to the average group with a return of -2.42%. The proportion of inferior group decreases from around 23% to 1.59%. In the second period we find a superior performance group with a rather high return of 17.1%. A handful of small funds seem to possess private information from 1998 to 2001. But in the last period, this phenomenon disappears. We again see the merging of





information sets. We only have two performance groups, with an inferior performance group with a return of -8.99%, but in this group only 3% of funds.

6.6.6 Conclusion

From the above analysis of the funds in five asset categories, we can make the following conclusions. First, there is a clear merging of information sets from 1995 to 2004, given the evolvement of dF plots. We view this phenomenon as an improvement of market efficiency. Second, there is an improvement of fund performance over the years from 1995 to 2004. It means the fund's ability to integrate and analyze information is stronger than before. Third, in the last period from 2001 to 2004 the funds' actual performance is around zero after adding back management fees, and we can not observe superior and inferior performance groups except for Large Blend funds. These empirical findings support the definition of market efficiency by Grossman and Stiglitz (1980). It means that funds can not beat down the market, and funds are compensated by the information

they collected and analyzed. Our findings confirmed the findings of Grinblatt and Titman (1989) and Droms and Walker (1996) with other methods.

6.7 Fund Rating

Along with the EM estimates of $\pi_i, \mu_i,$ and σ_i ($i = 1, \dots, g$), we also estimated the posterior probability of each fund j ($j = 1, \dots, n$) from each performance group i ($i = 1, \dots, g$) in each asset category at each period,

$$\tau_i(y_j; \Psi), (i = 1, \dots, g; j = 1, \dots, n) \quad (6.5)$$

where Ψ is the MLE estimator of parameters in finite normal mixture model by EM.

The posterior probability of each fund provides a basis for our fund rating. Firstly, given the estimator Ψ , we estimate the posterior probability of each fund by,

$$\tau_i(y_j; \Psi) = \frac{\pi_i \sigma_i^{-1} \exp\{(y_j - \mu_i)^2 / \sigma_i^2\}}{\sum_{c=1}^g \pi_c \sigma_c^{-1} \exp\{(y_j - \mu_c)^2 / \sigma_c^2\}}, (i = 1, \dots, g; j = 1, \dots, n) \quad (6.6)$$

Then the fund is assigned to the performance group for which it has the largest posterior probability. After that we rank the expected returns of the performance groups from the lowest to the highest. Finally we rate the fund according to the rankings of groups. For example, given $\tau_i(y_j)$, fund 123 has the largest posterior probability from performance group 2, we know fund 123 is from performance group 2. Suppose we identified five performance groups by parametric bootstrap procedures. and performance group 2 is ranked as number five denoting a superior

performance group, based on the expected returns μ_i estimated by EM. Then we rate the fund 123 as number five being a superior ability fund. The detailed information about fund ticker, fund's posterior probabilities from each performance group, and fund's rating, are shown in Appendix B.

6.8 Fund Performance Persistence

Jensen (1968) shows that managers generally are not able to sustain superior performance. However, Hendricks et al. (1993) argue that there is hot-hand effect in short-term. Grinblatt and Titman (1992) observe that past performance is predictive for future performance over period as long as three years. Thus, investigating the persistence issue remains an interesting topic.

Table 7: Fund Performance Persistence Study

Large Growth Funds					
Sample period	1	2	3	4	5
Nov. 1995–Nov. 1998	-24.1790	-20.2450	-12.5060	-7.4872	-2.2431
Nov. 1998–Nov. 2001	-11.8140	-7.9505	-8.1198	-6.7864	-5.2371
Nov. 2001–Nov. 2004	-1.8808	-5.1035	-2.8838	-2.1648	-2.5075

Large Blend Funds					
Sample period	1	2	3	4	5
Nov. 1995–Nov. 1998	-23.9630	-16.6960	-8.0553	-0.93106	7.6086
Nov. 1998–Nov. 2001	-13.4520	-7.7466	-6.2437	-3.79	-0.1577
Nov. 2001–Nov. 2004	-0.6524	-1.0978	-1.3129	-0.76492	0.70242

Large Value Funds						
Sample period	1	2	3	4	5	6
Nov. 1995–Nov. 1998	-23.9180	-18.8250	-12.5000	-7.9467	-3.3677	3.0984
Nov. 1998–Nov. 2001	-10.7820	-5.6445	-7.6147	-6.8428	-6.8	-0.96697
Nov. 2001–Nov. 2004	0.6269	-1.9866	-1.5484	-0.8804	-2.094	-0.0599

Medium Funds		
Sample period	1	2
Nov. 1995–Nov. 1998	-53.5510	-8.8907
Nov. 1998–Nov. 2001	-2.6588	-5.5332
Nov. 2001–Nov. 2004	-2.3762	-1.7158

Small Funds					
Sample period	1	2	3	4	5
Nov. 1995–Nov. 1998	-27.8370	-18.9510	-14.1310	-6.246	2.9259
Nov. 1998–Nov. 2001	8.0766	-10.0290	-3.4611	-1.6703	-0.74892
Nov. 2001–Nov. 2004	-2.3806	1.3405	-1.1186	-1.1634	0.58985

We classify the funds according to the ratings in the first period from 1995 to 1998, then we check the performance of the funds in each group at the next two periods, that is, from 1998 to 2001 and 2001 to 2004. The table shows the average performance of each performance group in each period, expressed as percentages annually. The numbers in the first row of each panel denote the performance group in the first period. 1 denotes the worst performance group, and the largest number in each row denotes the best performance group. For example in the first panel, number 5 denotes the best performance group from 1995 to 1998.

To explore the performance persistence issue of the funds from 1995 to 2004, we classify the funds based on the fund rating in the first period from 1995 to 1998. We calculate the average performance for each performance group in the first period, and then examine the performance of them in the next two periods, that is, from 1998 to 2001 and from 2001 to 2004, to check if there is any persistence pattern. The results are shown in table 7. The numbers in the first row of each panel denotes the number and ranking of the performance groups in the first period. For example, in the first panel for large growth funds, we have 5 performance groups in the first period. 1 denotes the worst performance group and 5 denotes the best performance group. Based on the classification in the first period, we check whether we observe the same performance ranking pattern in the next two periods.

The results show that there may be persistence for large-cap fund from the first period to the second period. For example, for large growth funds, we observe approximately the same pattern of ranking in the second period as the first period. The performance difference between the best performance group and the worst performance group is 22 percent annually in the first period and 6 percent in the second period. The same ranking pattern is observed for large blend funds and large value funds from the first period to the second period. However, in the third period, we don't observe clear persistence among fund performance groups, which implies that the rating in the first period may predict the performance in the second period, but not the performance in the third period. Regarding medium-cap funds and small-cap funds, we don't observe performance persistence pattern over the

sample period. Actually for small cap funds, the worst funds in the first period turn up to be the best funds in the second period, but become the worst funds again in the third period.

To summarize, we don't observe the persistence over the whole sample period, but we do observe the persistence from the first period (1995-1998) to the second period (1998-2001) for large-cap funds. The rating of the funds in the first period generally can not predict the performance in the future, especially the performance in the third period. From the fund rating study in section 5 of this Chapter, we know that the number of performance groups converges in the third period, which implies that the information becomes efficiently transmitted and analyzed in the last period. Fund managers could share the same information set. This may be the reason that the superior funds identified in the first period can not deliver the high performance in the third period. However, we do observe the persistence from the first period to the second period for large-cap funds, especially between the best performance group and the worst performance group, where the performance difference remains significant.

Chapter VII Conclusion

In this research, firstly, we critically review the traditional performance measures. We find that they are subject to some limitations, such as benchmark inefficiency, spurious market timing, and unrealistic normal assumption. Among them, benchmark inefficiency is the most devastating, which may lead to totally different inferential results from the same dataset using different measures or benchmarks, e.g. Ippolito (1989) and Elton et al. (1993). So over the decades we have seen a hot debate on whether fund managers can deliver abnormal returns and whether they possess superior information than ordinary investors. The debate is crucial, since if the managers can't deliver abnormal returns or do not possess superior information, the investors had better invest in index funds (a passive investment strategy) instead of actively managed funds.

Considering the limitations of current fund performance measures reviewed in Chapter II, we show that the measure proposed by Sharpe, which is derived from return-based style analysis, is superior to other measures. We formally explore the measure's mathematical formulation (convex quadratic programming), estimation techniques (interior point method), and rationale. We show that the rationale underlying this measure is more reasonable than current measures, since the target of this measure is attempting to identify a set of style index exposures that mostly mimics the return behavior of the fund, and then calculate the return difference

between the fund returns and the returns from the set of weighted style indexes. The return difference is interpreted as the performance of the fund due to active management like stock selection and asset allocation, whereas the current fund performance measures are just fitting a curve to the actual fund return series by the regression method without considering the actual fund return behavior. We have shown that the RBSA measure possesses several advantages. Firstly, it successfully circumvents benchmark inefficiency. The measure actually includes all the investable indexes in the model. The only requirement is that the indexes should be exclusive, exhaustive and different, which are not difficult to accommodate with a lot of indexes developed by commercial companies, e.g. Russell and Standard&Poor. Secondly, it does not rely on the normal assumption, since the measure is using mathematical algorithm to asymptotically solve out the alpha and betas. Thirdly, it provides style information of the fund that reflects the fund's return behavior, which is valuable for institutional investors.

After developing the methodology and algorithm of this measure, we conducted a comparative simulation experiment to test the accuracy, efficiency, and robustness of the measure and other current measures. The results show that the measure is the best in terms of accuracy, efficiency and robustness. We also find that the results from the Jensen measure are not stable, which are depending on the type of the fund. In addition, the Jensen measures with market-timing adjustments show spurious market –timing behavior and low explanatory power when measuring small-cap funds. We also find that Fama-French three-factor measure can not

correctly capture the true fund styles of well-diversified funds, so the alpha estimated from this measure is also questionable. Our conclusion is that the measure that we proposed could be more appropriate to measure fund performance.

After measuring the performance of funds, a natural extension is to rate funds based on fund performance. Instead of using preset percentiles to rate funds, we provide a new direction of fund rating which is based on the cross-sectional distribution of the alphas. The method is prompted by an empirical finding of the multimodal shapes of the distributions of alphas. We assume that the multimodal shapes that we observed are a consequence of segmented market information or different abilities to acquire and analyze the private and public information by fund managers. The managers who have more insider information or a team of quality analysts are expected to obtain higher abnormal returns. Basically we are assuming that the expected performance of the managers is different, depending on the amount and quality of the information that managers have. Under this assumption, the univariate normal distribution is not appropriate anymore, where the expected performance of all the managers are the same. A flexible way to model this situation is the finite normal mixture model, which is an addition of normal components. Each component has a different expected value. Actually this model has been widely used in medicine and genetics study in the last decade. Before estimating the model, we have to specify the number of components in the distribution model, that is, the number of performance groups in our fund rating study. Unfortunately, it was found that the log likelihood ratio, which is a natural

statistic used to determine the number of components in the model, did not follow the usual asymptotic Chi-square distribution. We exploit parametric bootstrap procedures to assess the p value of log likelihood ratio, and then determine the number of components. After the specification of the model, we estimate the model by EM, assuming fund rating information is part of complete data but missing. Under the EM framework, the estimation of log likelihood is straightforward and more importantly the interpretation of estimation results is intuitively appealing. The number of components, the expected values and the variance of the component density functions are interpreted as the number of performance groups, the expected performance, and the expected investment risk of each component performance group, respectively.

The fund rating method that we proposed is the main contribution of this thesis. We have observed that till now there is little academic literature on this topic. There are possibly two reasons. One is that the number of funds is relatively small before 1990 for distribution-based fund rating study, although the number is large enough for fund performance study, where it only requires that the fund has three-year return data. However, we observed that since 1990 the number of funds proliferated, which makes our fund rating method possible. The other factor that hinders the development of fund rating study is the lack of estimation techniques and software to implement this kind of fund rating method. However, the EM algorithm was introduced into the statistics literature by Dempster, Laird and Rubin in 1976. And during 1980s and 1990s, the technique became widely applied in

maximum likelihood estimation with missing data. The straightforward estimation technique of EM prompted the rapid development of finite normal mixture model in biology and medicine study in late 1990s. But we innovatively noticed that this model was also a natural way to rate funds, given that there are more than one performance groups in the market.

Empirical study of mutual funds in US provides interesting findings. To begin the study, we firstly construct a unique dataset by merging the equity mutual fund data from Bloomberg and Morningstar. Regarding fund performance, we have the following findings. Firstly, the results show that the average performance of these funds in the last three years is around zero after adding back management fees. Secondly, we do observe Medium Blend funds, Small Blend funds, and Small Value funds, can obtain positive abnormal return even after expense. We notice that Small Value funds deliver the highest alpha, which is 0.55% annually. This finding is consistent with previous findings, such as Roll (1981) and Chan, et al. (1985), that small value stocks are able to deliver higher abnormal returns. Thirdly, we observe that fund investment risk is compensated by the average performance of that fund group. The performance of large-cap funds is worst, -1.77% annually, but the variation of performance among the group of funds is smallest. Fourthly, the distribution of alphas is highly non-normal. Mostly noticeable is the multimodal shape of the distribution, which prompts us to model the distribution by the finite normal mixture model.

Regarding fund rating, we have the following findings. Firstly, we observe a clear tendency of the merging of information sets, which implies that the market became more efficient from 1995 to 2004 as information was well transmitted and analyzed. From 1995 to 1998 we identified around 5 separate information sets, while in the last period from 2001 to 2004 we only found 2 separate information sets except for large blend funds. Secondly, we observe an improvement of performance of all the funds from 1995 to 2004. From 1995 to 1998, the performance is the worst, while from 2001 to 2004, the performance is the best.

Besides the empirical findings, one important aspect of the empirical study is that we can show the mechanism of how the fund market becomes more efficient over time from a unique perspective of fund rating with a series of DF plots from 1995 to 2004 for five categories of funds, i.e. large growth funds, large blend funds, large value funds, medium-cap funds, and small-cap funds. This study helps us understand the evolvement of fund market from 1995 to 2004. In addition, from the estimation results of finite normal mixture model, we can tell the proportion of funds, the expected performance, and the expected investment risk of each performance group in each period from 1995 to 2004, which provides an in-depth detailed understanding of each performance group and its risk.

Based on the fund rating in the first period, we also investigate the performance persistence issue in the next two periods. We did not observe the persistence over the whole sample period, but we did find the persistence from the first period to the

second period for large-cap funds. The nonexistence of the persistence in the third period may be due to the more efficient market as we observed from fund rating study that there was a merging of information sets during the last period from 2001 to 2004.

Our research provides a new direction of fund rating. With the availability of high-speed computer, the distribution-based rating method may draw more attention from researchers in fund study in the future as we have seen in biology and medicine study in the last decade. Since our method is a first try to deal with fund rating issue from the perspective of the cross-sectional distribution of alphas, our main purpose is to show the rationale and the framework of this method. There remain some limitations. One is about the selection of initial values in parametric bootstrap procedures. This may be achieved by random starts, hierarchical clustering-based starts, and k-means clustering-based starts. Exploring all of them may take us too far from the essence of the thesis, that is, to provide a new way that is flexible to deal with fund rating difficulty. But exploring may offer us some interesting insights on how to accurately rate funds.

Appendix A: Modified Newton Method

Before introducing modified the Newton Method (Nesterov and Nemirovski, 1994), it is helpful to briefly review the Newton's method first. Newton's method is frequently used to solve nonlinear equations. Suppose we have several equations with several variables.

$$F(x) = 0$$

where x is a vector of n variables and $F(x)$ has n equations,

$$f_1(x_1, x_2, \dots, x_n) = 0$$

$$f_2(x_1, x_2, \dots, x_n) = 0$$

.....

$$f_n(x_1, x_2, \dots, x_n) = 0.$$

Firstly, we approximate F with the first order Taylor's series around current estimate x_k :

$$F(x_k + \delta) \approx F(x_k) + \nabla F(x_k) \delta \quad (\text{A.1})$$

where $\nabla F(x_k)$ denotes the Jacobian of function $F(x)$. We denote the n -dimensional vector x by (x^1, x^2, \dots, x^n) , to represent the Jacobian in matrix form, i.e.

$$\nabla F(x^1, x^2, \dots, x^n) = \begin{pmatrix} \frac{\partial f_1}{\partial x^1} & \dots & \frac{\partial f_1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x^1} & \dots & \frac{\partial f_n}{\partial x^n} \end{pmatrix}. \quad (\text{A.2})$$

Since we are solving $F(x)=0$, from (A.1) the Newton update is determined by

$$F(x_k) + \nabla F(x_k) \delta = 0 \quad (A.3)$$

On the right side of the equation, 0 is an n-dimensional vector. If $\nabla F(x_k)$ is nonsingular, we can solve δ , that is,

$$\delta = -\nabla F(x_k)^{-1} F(x_k) \quad (A.4)$$

Therefore the Newton Update formula is,

$$x_{k+1} = x_k + \delta = x_k - \nabla F(x_k)^{-1} F(x_k) \quad (A.5)$$

For the quadratic programming problem, suppose we have an estimate $(\beta^k, s^k, \lambda^k)$ from F^0 , an interior point from the feasible region (3.10). The Newton step $(\Delta\beta^k, \Delta s^k, \Delta\lambda^k)$ from this point is determined by solving the following system of equations from (A.3),

$$J(\beta^k, s^k, \lambda^k) \begin{pmatrix} \Delta\beta^k \\ \Delta s^k \\ \Delta\lambda^k \end{pmatrix} = -F(\beta^k, s^k, \lambda^k) \quad (A.6)$$

where $(\Delta\beta^k, \Delta s^k, \Delta\lambda^k)$ is also called the search direction from the last estimate $(\beta^k, s^k, \lambda^k)$. $J(\beta^k, s^k, \lambda^k)$ is the Jacobian of $F(\beta, s, \lambda)$. From (3.9) we can find it is,

$$J(\beta^k, s^k, \lambda^k) = \begin{pmatrix} -H & I & e \\ S^k & B^k & 0 \\ e & 0 & 0 \end{pmatrix} \quad (A.7)$$

S^k and B^k are diagonal matrices with the components of the vector s^k and β^k along their diagonals. At the end of solution procedure, we also get s and λ .

Since we are only interested in β , solving them seems a waste of time. However, it

is the most efficient and robust way to implement interior point method, because the information of s and λ helps us make faster and better improvements on iterates of β . Since $(\beta^k, s^k, \lambda^k)$ is from F^0 , from (3.10), we obtain

$$F(\beta^k, s^k, \lambda^k) = \begin{pmatrix} 0 \\ B^k S^k e \\ 0 \end{pmatrix}. \quad (\text{A.8})$$

From (A.8) and (A.7), Newton equation in (A.6) reduces to

$$\begin{pmatrix} -H & I & e \\ S^k & B^k & 0 \\ e & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta\beta^k \\ \Delta s^k \\ \Delta\lambda^k \end{pmatrix} = \begin{pmatrix} 0 \\ -B^k S^k e \\ 0 \end{pmatrix}. \quad (\text{A.9})$$

After we determined the Newton step this way, we can update $(\beta^k, s^k, \lambda^k)$ with Newton update formula, that is,

$$(\beta^{k+1}, s^{k+1}, \lambda^{k+1}) = (\beta^k, s^k, \lambda^k) + (\Delta\beta^k, \Delta s^k, \Delta\lambda^k). \quad (\text{A.10})$$

The above is a straightforward application of Newton's method, but it is possible that the new update $(\beta^{k+1}, s^{k+1}, \lambda^{k+1})$ lies out of feasible region constrained by optimality conditions,

$$F = \{(\beta, s, \lambda) : \lambda e - H\beta - g + s = 0, Bs = 0, \beta^T e = 1, \beta, s \geq 0\}. \quad (\text{A.11})$$

To avoid the violation of the non-negativity constrains $\beta, s \geq 0$, we modify the Newton's method by adding a step-size parameter, $\alpha_k \in (0, 1]$ such that

$$\begin{aligned} \beta^k + \alpha_k \Delta\beta^k &> 0 \\ s^k + \alpha_k \Delta s^k &> 0. \end{aligned} \quad (\text{A.12})$$

The largest possible value of α_k satisfying these restrictions can be found using a procedure similar to the ratio test in the simplex method. Once we determined the step-size parameter, we choose the next iterate as

$$(\beta^{k+1}, s^{k+1}, \lambda^{k+1}) = (\beta^k, s^k, \lambda^k) + \alpha_k (\Delta\beta^k, \Delta s^k, \Delta\lambda^k). \quad (\text{A.13})$$

We summarize the interior point method as follows.

The initial step is to choose $(\beta^0, s^0, \lambda^0) \in F^0$.

For $k=0,1,2,\dots$, repeat the following steps.

1. Solve Newton step $(\Delta\beta^k, \Delta s^k, \Delta\lambda^k)$ from

$$\begin{pmatrix} -H & I & e \\ S^k & B^k & 0 \\ e & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta\beta^k \\ \Delta s^k \\ \Delta\lambda^k \end{pmatrix} = \begin{pmatrix} 0 \\ -B^k S^k e \\ 0 \end{pmatrix}.$$

2. Choose α_k such that

$$\begin{aligned} \beta^k + \alpha_k \Delta\beta^k &> 0 \\ s^k + \alpha_k \Delta s^k &> 0. \end{aligned}$$

3. Update $(\beta^k, s^k, \lambda^k)$ by

$$(\beta^{k+1}, s^{k+1}, \lambda^{k+1}) = (\beta^k, s^k, \lambda^k) + \alpha_k (\Delta\beta^k, \Delta s^k, \Delta\lambda^k).$$

Appendix B: Summary of Fund Rating

The funds are grouped into three general performance groups: inferior performance group, average performance group, and superior performance group, based on the expected group performance and its probabilities.

Groups are numbered from the lowest to the highest, where the lowest group has the lowest expected performance.

Table 8: Summary of Fund Rating

Panel I		Large Growth Funds (234)	
Sample period	Inferior Performance Group	Average Performance Group	Superior Performance Group
Nov.1995-Nov.1998	Group 1 (6) Group 2 (10)	Group 3 (111) Group 4 (26) Group 5 (81)	Group 4(5)
Nov.1998-Nov.2001	Group 1(8)	Group 2(69) Group 3(152)	
Nov.2001-Nov.2004		Group 1(231) Group 2(3)	
Panel II		Large Blend Funds (272)	
Sample period	Inferior Performance Group	Average Performance Group	Superior Performance Group
Nov.1995-Nov.1998	Group 1 (4) Group 2 (20)	Group 3 (157) Group 4 (82)	Group 5 (9)
Nov.1998-Nov.2001	Group 1(16) Group 2(81)	Group 3(175)	Group4(5)
Nov.2001-Nov.2004	Group 1(4)	Group 2(200) Group 3(63)	
Panel III		Large Value Funds (218)	
Sample period	Inferior Performance Group	Average Performance Group	Superior Performance Group
Nov.1995-Nov.1998	Group 1 (2) Group 2 (11)	Group 3 (79) Group 4 (51) Group 5 (69)	Group 6 (6)
Nov.1998-Nov.2001	Group 1(19)	Group 2(190)	Group 3 (9)
Nov.2001-Nov.2004	Group 1(7)	Group 2(211)	
Panel IV		Medium Funds (228)	
Sample period	Inferior Performance Group	Average Performance Group	Superior Performance Group
Nov.1995-Nov.1998	Group 1 (2)	Group 2 (226)	Group 5(26)
Nov.1998-Nov.2001	Group 1(11) Group 2(23)	Group 3(146) Group 4(22)	
Nov.2001-Nov.2004		Group 1(6) Group 2(222)	
Panel V		Small Funds (193)	
Sample period	Inferior Performance Group	Average Performance Group	Superior Performance Group
Nov.1995-Nov.1998	Group 1 (3) Group 2 (12) Group 3 (29)	Group 4 (127)	Group 5(22)
Nov.1998-Nov.2001	Group 1(3)	Group 2(183)	Group 3(7)
Nov.2001-Nov.2004	Group 1(3)	Group 2(190)	

Totally we have 1145 funds under study from November 30,1995 to November 30 2004. The sample period is further divided into three sub-periods, i.e. from 1995 to 1998, from 1998 to 2001, and from 2001 to 2004. The 1145 funds are classified into five asset categories: Large Growth funds, Large Blend funds, Large Value funds, Medium funds, and Small funds.

Reference:

- Admati, A. R., S. Bhattacharya, S. A. Ross, and P. Pfleiderer, 1986, On timing and selectivity, *Journal of Business* 58, 1-26.
- Arnott, A. C., 1996, *Morningstar mutual funds*. December 6.
- Buetow, Gerald W., Jr., Johnson, Robert R., and Runkle, David E. Spring 2000, The Inconsistency of Return-Based Style Analysis. *Journal of Portfolio Management*, 61-77.
- Cai, J, K. C. Chan, and T. Yamada, 1997, The performance of Japanese mutual funds, *Review of Financial Studies* 10, 237-273.
- Carhart, M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57-82.
- Chan, K. C., Nai-fu Chen, and David A. H., 1985, An explanatory investigation of the firm size effect, *Journal of Financial Economics* 14, 451-471.
- Chiang, A. C., 1984, *Fundamental methods of mathematical economics*.
- Conover, W. J., 1980, *Practical Nonparametric Statistics*. Wiley.
- Christopherson, J. A., Spring 1995, Equity Style Classifications. *Journal of Portfolio Management*, 32-43.
- Christopherson, J. A., W. E. Ferson, and D. A. Glassman, 1998, Conditioning manager alphas on economic information: another look at the persistence of performance, *Review of Financial Studies* 11, 111-142.
- Christopherson, J. A. and Sabin, F. C., 1999, How Effective is Effective Mix?, *Russell*

Research Commentary, 1-18.

Cochrane, J., 1996, A cross-sectional test of production-based asset pricing model, *Journal of Political Economy* 104, 572-621.

Cramer, H., 1946. *Mathematical methods of statistics*. (Princeton University Press, Princeton).

Cummisford, R. and Lummer, S., Oct. 1996, Controlling the Limitations of Style Analysis. *Journal of Financial Planning*, 70-76.

Dempster, A.P., N.M. Laird, and D.B. Rubin, 1977, Maximum likelihood estimation from incomplete data via the EM algorithm, *Journal of Royal Statistical Society, Series B* 39, 1-38.

Droms, W. G. , and D. A. Walker, 1996, Mutual fund investment performance, *The Quarterly Journal of Economics* 36, 347-363.

Elton, E. J., M. J. Gruber, and C. R. Blake, 1996, The persistence of risk-adjusted mutual fund performance, *Journal of Business* 69, 133-157.

Elton, E. J., M.. J. Gruber, S. Das, and M. Hlavka, 1993, Efficiency with costly information: a reinterpretation of evidence from managed portfolios, *Review of Financial Studies* 6, 1-22.

Epanechnikov, V. K., 1969, Non-parametric estimation of a multivariate probability density, *Theory of Probability and Applications* 14, 153-158.

Evans, M., 1994, Expected returns, time-varying risk, and risk premia, *Journal of Finance* 40, 655-680.

Fama, E. F., 1970, Efficient Capital Markets: A review of theory and empirical work, *Journal of Finance* 45, 383-417.

- Fama, E. F., and K. R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22.
- Fama, E. F., and K. R. French, 1993, Common risk factors in the returns on bonds and stocks, *Journal of Financial Economics* 33, 3-53.
- Feng, Z.D., and C.E. McCulloch, 1996, Using bootstrap likelihood ratio in finite mixture models, *Journal of the Royal Statistical Society, Series B* 58, 609-617.
- Ferson, W. E., and C. R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy* 99, 385-415.
- Ferson, W. E., 1985, Changes in expected risk premiums and security risk measures, *Proceedings of the European Finance Association*.
- Ferson, W. E., and C. R. Harvey, 1993, The risk and predictability of international equity market, *Review of Financial Studies* 6, 527-566.
- Ferson, W. E., and R. W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425-461.
- Fisher, N.I., E. Mammen, and J.S. Marron, 1994, Testing for multimodality, *Computational Statistics and Data Analysis* 18, 499-512.
- Glosten, L., and R. Jagannathan, 1994, A contingent claims approach to performance evaluation, *Journal of Empirical Finance* 1, 133-166.
- Grinblatt, M., and S. Titman, 1989, Mutual fund performance: an analysis of quarterly portfolio holdings, *Journal of Business* 62, 393-416.
- Grinblatt, M., and S. Titman, 1992, The persistence of mutual fund performance, *Journal of Finance* 47, 1977-1984.

- Grinblatt, M., and S. Titman, 1994, A study of monthly mutual funds returns and performance evaluation techniques., *Journal of Financial and Quantitative Analysis* 29, 419-444.
- Grossman, S. J., and J. E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393-408.
- Hartingan, J., and S. Mohanty, 1992, The run test for multimodality, *Journal of Classification* 9.
- Hendricks, D., J. Patel, and R. Zeckhauser, 1993, Hot hands in mutual funds: short-run persistence of relative performance, 1974-1988, *Journal of Finance* 48, 93-130.
- Henriksson, R. D., and R. C. Merton, 1981, On market timing and investment performance. II. statistical procedures for evaluating forecasting skills, *Journal of Business* 54, 513-534.
- Ippolito, R. A., 1989, Efficiency with costly information: a study of mutual fund performance, 1965-1984., *The Quarterly Journal of Economics* 104, 1-23.
- Jagannathan, R., and R. A. Korajczyk, 1986, Assessing the market timing performance of managed portfolios, *Journal of Business* 59, 217-235.
- Jagannathan, R., and Z. Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3-54.
- Jegadeesh, N., and S. Titman, 1993, Returns to buying winners and selling losers: implication for stock market efficiency, *Journal of Finance* 48, 65-91.
- Jensen, M. C., 1968, The performance of the mutual funds in the period 1945-1964., *Journal of Finance* 23, 389-416.

- Karmarkar, N., 1984, A new polynomial-time algorithm for linear programming, *Combinatorica* 4, 373-395.
- Keim, D.B., and R.F. Stambauch, 1986, Predicting returns in the bond and stock Markets, *Journal of Financial Economics* 17, 357-390.
- Kosowski, R., A. Timmermann, H. White, and R. Wermers, 2001, Can mutual fund "stars" really pick stocks? evidence from a bootstrap analysis, *Working paper*.
- Kosowski, R., A. Timmermann, H. White, and R. Wermers, 2001, Can mutual fund "stars" really pick stocks? new evidence from a bootstrap analysis, *Working paper*.
- Lehmann, B, and D. Modest, 1987, Mutual fund performance : A comparison of benchmarks and benchmarks comparisons, *Journal of Finance* 42, 233-265.
- Lieberman, D. L., Fall 1996, Return-Based Style Analysis: Are Quarterly Returns As Meaningful?, *Journal of Investing*, 51-55.
- Markowitz, H. M., 1987, Mean-variance analysis in portfolio choice and capital markets, Oxford: Basil Blackwell, Inc..
- Mayes, T. R., Jay, N. R., and Thurston, R., August 2000, A Returns-Based Style Analysis Examination of Asset Classes. *Journal of Financial Planning*, 94-104.
- McLachlan, G.J., 1987, On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture, *Applied Statistics* 36, 318-324.
- McLachlan, G.J., 1992, Cluster analysis and related techniques in medical research., *Statistical Method in Medical Research* 1, 27-49.

- McLachlan, G.J., and K.E. Basford, 1988. *Mixture models: inference and applications to clustering* (Marcel Dekker, New York).
- Nesterov, Y., and A. Nemirovski, 1994, Interior-point polynomial algorithms in convex programming, *SIAM*.
- Priebe, C.E., 1994, Adaptive mixture density estimation, *Journal of the American Statistical Association* 89, 796-806.
- Redner, R.A., and H.F. Walker, 1984, Mixture densities, maximum likelihood and Em algorithm., *SIAM Review* 26, 195-239.
- Roeder, K., 1994, A graphical technique for determining the number of components in a mixture of normals., *Journal of the American Statistical Association* 89, 487-495.
- Roll, R., 1977, A critique of the asset pricing theory's tests, *Journal of Financial Economics* 4, 129-176.
- Roll, R., 1978, Ambiguity when performance is measured by security market line, *Journal of Finance* 33, 1051-1069.
- Roll, R., 1981, A possible explanation of the small firm effect, *Journal of Finance* 36, 879-888.
- Shanken, J., 1990, Intertemporal asset pricing: an empirical investigation, *Journal of Econometrics* 45, 99-120.
- Sharpe, W., 1987, An algorithm for portfolio improvement. *Advances in Mathematical Programming and Financial Planning*, 1, 155-170. JAL Press, Inc.

- Sharpe, W., 1988, Determining a fund's effective asset mix, *Investment Management Review*, December, 59-69.
- Sharpe, W., 1992, Asset allocation: management style and performance measurement, *Journal of Portfolio Management* 18, 7-19.
- Sharpe, W., 1998, Morningstar's risk-adjusted ratings. *Financial Analysts Journal* 54, 21-33
- Silverman, B.W., 1981, Using kernel density estimates to investigate multimodality, *Journal of the Royal Statistical Society, Series B* 43, 97-99.
- Silverman, B.W., 1986. *Density estimation for statistics and data analysis* (Chapman and Hall, London).
- Solka, J.L., E.J. Wegman, C.E. Priebe, W.L. Poston, and W. Rogers, 1998, Mixture structure analysis using Akaike criterion and the bootstrap, *Statistics and Computing* 8, 177-188.
- Tao, J., N. Z. Shi, and S. Y. Lee, 2004, Drug risk assessment with determining the number of sub-populations under finite mixture normal models, *Computational Statistics and Data Analysis* 46, 661-676.
- Titterington, D.M., A.F.M. Smith, and U.E. Makov, 1985. *Statistical analysis of finite mixture distributions* (Wiley, New York).
- Treynor, J. L., and K.K. Mazuy, 1966, Can mutual funds outguess the market?, *Harvard Business Review* 44, 131-136.
- Wolfe, J.H., 1971, A Monte Carlo study of sampling distribution of the likelihood ratio for mixtures of multinormal distributions, *Technical Bulletin STB 72*, San Diego: US Naval Personnel and Training Research Laboratory.

- Wong, W.A., 1985, A bootstrap testing procedure for investigating the number of subpopulations., *Journal of Statistical Computation and Simulation* 22, 99-112.
- Woodward, W.A., and S.R. Sain, 2003, Testing for outliers from a mixture distribution when some data are missing, *Computational Statistics and Data Analysis* 44, 193-210.
- Yudin, D.B., and A. Nemirovski, 1976, Information complexity and efficient methods for the solution of convex extremal problems, *Matekon* 13, 3-25.