TIME-VARYING PERFORMANCE IN THE CROSS-SECTION OF MUTUAL FUND RETURNS

by

RONG HUANG

A thesis submitted to the University of Birmingham
for the degree of DOCTOR OF PHILOSOPHY

Department of Economics
Birmingham Business School
College of Social Sciences
University of Birmingham
September 2016

Abstract

This thesis examines the central question of whether actively managed mutual funds generate returns beyond those offered by passively managed funds. Using a non-parametric change point test and a cross-sectional bootstrap technique, this study conducts the first comprehensive examination of mutual fund performance that explicitly controls for possible time-variation in both alpha and betas of the extended versions of the CAPM, without imposing any specific functional form on the nature of the time-variation in these parameters. We further use an FDR technique that is able to quantify the number of skilled and unskilled managers.

Our empirical analysis using these techniques reaches the following conclusions: first, after controlling for time-variation, there is more evidence of manager skill for both net and gross returns than previously documented in the literature (Chapter 4). Second, the estimated proportion of skilled funds in our sample is 8.4%, and 34.1% of the funds are identified as unskilled (Chapter 5). Third, different rules in selecting funds could lead to contradictory inferences on fund performance (Chapter 6).

This thesis is dedicated to

my parents

for their endless support and love

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor Dr. Frank Strobel for his continuous support of my Ph.D study and research. His guidance helped me throughout my time of research and the writing of this thesis. I could not have finished this work without his help.

My thanks also go to my supervisor Mr Nick Horsewood, for his valuable guidance and support during my PhD study, and for his many helpful comments on this thesis. I would also like to thank my supervisor Dr. William Pouliot for the many stimulating discussions we had, and for his precious help in the writing of this thesis.

I am grateful to my examiners Prof. Jose Olmo and Prof. David Dickinson for their insightful comments on this thesis. Also I thank Prof. David Maddison, and Prof. Robert Elliot for interviewing me and granting me the opportunity to enter this PhD programme.

I thank my colleagues: Yi Liu, Dr. Liyun Zhang, Bing Liu, Dr. Yaoyao, Yuqian Zhao, Dr. Charlie Rahal, Yanan Zhang, Dr. Mary Daywood, Dr. Ruanmin Cao, Danying Li and Aima Tahir, for their help and friendship during the last four years.

Last but not the least, I would like to thank my family, especially my parents' endless and selfless support throughout my life.

Contents

List of Tables

List of Figures

List of Abbreviations

3F-CAPM ... Fama-French three-factor capital asset pricing model

4F-CAPM ... Carhart's four-factor capital asset pricing model

AUM .... Assets under management

BH .... Benjamini and Hochberg (1995)

EW .... Equal-weight

F&F .... Fama and French (2010)

KTWW ...... Kosowski, Timmermann, Wermers and White (2006)

BSW .... Barras, Scaillet and Wermers (2010)

CAPM .... Capital asset pricing model

CDF .... Cumulative density function

CRSP .... The Center for Research in Security Prices

CUSUM ..... Cumulative sum control chart

FDR .... False discovery rate

FWER .... Family-wise error rate

HML .... High minus low

MOM .... Momentum

MSE .... Mean squared error

NAV .... Net asset values

OLS .... Ordinary least squares

pFDR .... Positive false discovery rate

SMB .... Small minus big

U.S. .... United States

VW .... Value-weight

Chapter 1

Introduction

Whether actively managed mutual funds offer returns beyond those offered by passively managed funds has, over the years, generated much debate, especially given the growing popularity of passively managed funds, mainly index funds, in recent years. By the end of 2015, 32% of mutual fund household investors held at least one equity index fund. Recent empirical studies lead by Wermers (2000), Kosowski, Timmermann, Wermers and White (2006) (hereafter KTWW) and Barras, Scaillet and Wermers (2010) (hereafter BSW) confirm the presence of skilled mutual fund managers who generate returns that exceed passive benchmarks. Nonetheless, studies by French (2008) and Fama and French (2010) (hereafter F&F) provide evidence to the contrary. According to F&F, there is no widespread manager skill and actively managed funds cannot generate returns net of costs that outperform the passive benchmark portfolios.

The research undertaken here is motivated by these recent contributions to this debate. Our study aims to revisit the important question of whether there is a group of skilled managers and to shed new light on this issue. This question is not without importance for society in general since many pension plans and investment retirement accounts hold mutual funds that are actively managed, with those funds advertising themselves as superior to passively managed funds. Regarding the contribution offered here, a discussion of the key approaches to performance evaluation will be provided. This discussion will raise several problems with the methodology and data used in these studies. We aim to provide answers to these problems and to offer a systematic way of addressing them. The final part of this introductory chapter will discuss how this thesis is structured in addressing these problems.

1.1 Research Problems

Jensen (1968) is widely recognized as the first study to examine mutual fund performance. In particular, he assumes that fund managers maintain on average a given level of risk (cf. page 395). Under this assumption, the risk factor loadings, i.e. the betas of extended versions of the capital asset pricing model (CAPM), will be time-invariant. He notes that if a manager is a superior forecaster, she can systematically select securities which will realize a return that is higher than what is predicted by the risk factors. As such, the CAPM for Jensen includes an intercept which is popularly referred to as Jensen's alpha. This alpha, the intercept of the CAPM, will be positive for a manager who earns returns above those predicted by the benchmark portfolios.

It is obvious then that providing accurate assessment of manager skill depends critically on consistent estimation of these risk factors as well as alpha. For this to be achieved, suitably specified benchmarks, i.e. the best possible econometric models are critical. There are two other popular models apart from the CAPM: the Fama French three-factor model (3F-CAPM) and Carhart's extension of this model to a four-factor version (4F-CAPM). As mentioned previously, these models are presented as having time-invariant risk factors, which, however, may not necessarily be appropriate when evaluating fund performance for actively managed funds. When the composition of a fund's portfolio changes, so too will the overall risk of the portfolio, i.e. the portfolio risk factor loadings will change.

This leads to our first set of research questions which aim to address the possible time-variation issue in the risk factor loadings. The extended CAPM models (the 3F-CAPM and the 4F-CAPM) most often employed in earlier studies assume that the risk factors are time invariant; however, it is well known that dynamic trading strategies of fund managers can lead to variation in these risk factors. Because of this, a large number of conditional versions of the extended CAPM have been developed. The widely used conditional model proposed by Ferson and Schadt (1996) allows the market portfolio beta to be time-varying by letting it depend linearly on some lagged public information variables. Ferson and Schadt (1996) confirm that these information variables are statistically significant, and the use of their conditional model improves the performance of funds in their sample. However, later studies (Blake and Timmermann, 1998; Kacperczyk et al., 2005; Kosowski et al., 2006; Barras et al., 2010; Cuthbertson et al., 2008, 2012) find that results produced from the Ferson-Schadt conditional model do not differ much from those generated by the traditional unconditional models. This suggests that the reduced form approach of Ferson and Schadt (1996), i.e. allowing the market beta to be a linear function of the information variables, may not be an appropriate way to model time-variation in the market beta. Hence, to properly capture time-variation in the betas remains a problem that needs to be solved. Indeed, F&F confirm that time-variation in the regression slopes poses "thorny" problems to their study and they leave this important issue to future research.

Motivated by this, the method developed here takes an agnostic view on the nature of the relation between risk factors and the information variables. To implement this, we employ a non-parametric statistical test developed by Pouliot (2016) to detect time-variation in the risk factors. Determining when the risk factors change will be central to the methodology implemented here and will be developed in detail in later chapters. Most importantly, we assume that Jensen's alpha can also change over time and the test implemented is able to detect time-variation in fund alphas as well. For instance, a mutual fund may have many managers of differing abilities throughout its entire history, hence the manager skill (Jensen's alpha) of a fund can be time-varying. According to Evans (2010) (cf. page 1599), the average manager tenure of mutual funds is around five years, hence manager replacement happens frequently. Furthermore, studies such as Khorana (2001) and Clare et al. (2014) find significant improvements of post-replacement performance relative to past performance for those underperforming funds. Even if a fund does not have manager replacement or has many managers of equivalent levels of abilities, Berk and Green (2004) have shown that managers exhibit decreasing returns in deploying their superior skill and excess returns should not persist. This is because investors provide capital elastically and capital will move to funds with positive excess returns and leave those funds with negative excess returns. Therefore, the aforementioned evidence and arguments suggest that manager skill can vary over time, hence the fund alpha can change.

Previous studies do not explicitly recognize and model the effect of manager change, nor the fact that managers exhibit decreasing returns in deploying their superior skill. This study conducts the first comprehensive examination of mutual fund performance that explicitly controls for possible time-variation in both alpha and betas of the extended versions of the CAPM without imposing any specific functional form on the nature of the time-variation in these parameters.

Jensen (1968) also assumes that the idiosyncratic shocks that affect returns are normally distributed. KTWW question the appropriateness of this assumption in the context of measuring fund performance. Indeed, they reject the assumption of normality for about half of individual fund alphas (cf. page 2555) that they estimate. More importantly, KTWW question the appropriateness of testing the existence of skilled managers using the standard parametric t-tests of individual funds in previous studies, given possible cross-fund correlations in returns and extreme returns that could be gained by chance. To capture any cross-fund correlations in returns and to control for the role of luck, they implement a cross-sectional bootstrap, which is the measure implemented in this study as well, to infer the existence of skilled and unskilled managers.

KTWW's bootstrap approach essentially involves resampling of the regression residuals of the extended versions of the CAPM to produce empirical cross-sectional distributions of fund alpha estimates. They show that there are skilled managers who produce excess returns net of all expenses and costs. F&F argue that the bootstrap approach of KTWW does not capture the cross-fund dependency during simulations since they resample from the regression residuals separately for each fund, hence they propose a bootstrap method that resamples for all funds simultaneously. In this way, F&F overturn most of the results of KTWW and show that few funds exhibit superior skill in generating excess returns after costs. Furthermore, to capture possible autocorrelation effects in the time series, BSW use a block bootstrap method and the false discovery rate (FDR) approach of Storey (2002), and find some evidence of skilled managers over the short run. They resample from blocks of residuals so that the original time series structure within a block can be preserved. Given the different bootstrap methods used in these studies, it is of interest to examine how the different bootstrap procedures affect the empirical cross-sectional distributions of fund alpha estimates. This is crucial as our inferences of skilled and unskilled fund managers depend on the values of alpha estimates in the upper and lower quantiles of the cross-fund alpha distributions.

Our last research problem arises in the data used in previous studies (KTWW, BSW and F&F), and used in this research as well, which consists of monthly returns from the Center for Research in Security Prices (CRSP). The CRSP mutual fund database contains monthly data on net returns for each share class of every open-end mutual fund since January 1962, with no minimum survival requirement for funds to be included. While this database includes alive and delisted funds, it does not have complete investment objectives of every fund. This is not without consequence for our study, as well as those of others, as to measure manager skill requires knowledge of those funds that invest primarily in equities and also advertise themselves as actively managed. Most of the studies, including those mentioned previously (KTWW, F&F and BSW) do not provide clear guidance on how they identify funds that are actively managed and invest primarily in equities.

We find that previous studies using CRSP tend to have different rules in selecting funds. For instance, F&F include funds in their sample as long as the fund has at least eight months of return history, Pástor and Stambaugh (2002) require funds in the sample to have at least 12 months of return history, Wermers (2000) include a fund only if it has survived for at least 24 months, and funds in the sample of Sirri and Tufano (1998), KTWW and BSW have a minimum 60 months of return history. This suggests that these previous studies have different levels of survivor bias in their samples, which could lead to different results. Indeed, F&F argue that the rules used to select funds could affect the cross-sectional distribution of fund alpha ($t(\alpha)$) estimates for their sample. Furthermore, previous studies also have different treatments regarding newly established funds and incubated return history of funds, and this will be elaborated further in later chapters. Given these controversies in selecting funds, it is of interest to examine how these differences in selection rules can affect the inferences on the existence of skilled and unskilled managers.

1.2 Structure of the Thesis

To offer a systematic way of addressing the various problems listed so far, the thesis is structured as follows:

Chapter 2 sets out to provide some background data for this research. We explore some recent trends in the U.S. mutual fund industry, especially the growing demand of index funds in recent years. Then, the chapter introduces the concept of active and passive management. Furthermore, we explain the theoretical justification of the returns-based measure Jensen's alpha, which is the measure of performance used in this study. Since fund alphas are estimated using both net and gross returns so as to disentangle a fund's expense policy from examination of the manager skill, we also discuss the interpretation of manager skill in terms of net and gross returns.

Chapter 3 gives a more detailed survey of the literature that this study is based on. Specifically, we review the relevant literature on the examination of abnormal performance of mutual funds, where we show existing controversies in previous studies regarding the existence of skilled fund managers. Next, we review some extended models used in the literature to address time-variation in the regression parameters of the factor models.

We also show why our treatment of time-varying parameters, i.e. using a non-parametric change point test to detect time-variation in fund alphas and betas, is superior to other existing approaches. The last section of this chapter provides a review of the use of change point analysis in the economic literature and shows why the non-parametric change point test of Pouliot (2016) is useful in addressing time-variation issues in our context.

In Chapter 4, to answer the question whether there is a group of skilled or unskilled fund mangers, a cross-sectional bootstrap technique developed by KTWW and F&F is used to control for the role of luck in the cross-sectional distribution of fund alphas. Differently from the use of the 3F-CAPM in their study, which assumes the alpha and factor loadings are time-invariant, we use the change point test, which is able to detect multiple discrete changes in the alpha and factor loadings. We also explore the possible underlying reasons that lead to changes in betas as well as in alpha, which provides justification for the use of the change point test in our study. To capture changes in the fund alpha and betas, we build a multi-regression model that essentially involves detecting changes in these parameters first, and then estimating fund alphas on subintervals of return series that are separated according to the dates of changes. Furthermore, we compare the results of our model to those of the traditional models. We have shown that after allowing for time-variation in the regression parameters, the cross-sectional distribution of alpha ($t(\alpha)$) estimates shifts rightwards, which suggests more evidence of manager skill than previously documented in the literature. To confirm that the pessimistic results of F&F are mainly due to the assumption of time-invariant regression parameters, we construct the sample ending in 2006 as in F&F and redo their analysis using the multi-regression model and the traditional 4F-CAPM, respectively.

In Chapter 5, knowing that there is a group of skilled managers, we further examine how many actively managed fund managers within the entire sample have truly abnormal performance that produces positive, negative or zero alphas using the FDR approach of BSW in Chapter 5. Differently from the study of BSW, we use the change point test of Pouliot (2016) to detect time-variation in the fund alpha and betas, and estimate fund alphas using the multi-regression model as described in Chapter 4. As a result, more skilled and unskilled funds are detected relative to the results of BSW. For example, we find that 8.5% of the actively managed fund managers are able to outperform the passive benchmarks, relative to a proportion of 0.7% in BSW. We also examine where skilled funds are located in the cross-sectional distribution of alpha ($t(\alpha)$) estimates. Finally, to examine whether the investment objectives of funds have an impact on fund performance, this chapter examines the performance of funds in different investment styles by dividing the entire sample into three commonly defined investment styles, namely aggressive growth, growth and growth & income.

Chapter 6 aims to examine three important issues that are relevant to mutual fund studies in general. These issues remain controversial in the literature and are closely related to Chapters 4 and 5. First, we show different criteria applied in previous studies to select funds in their samples. Then, we construct several datasets based on different selection rules used in these studies, and use them to estimate fund alphas respectively. This shows how different ways of constructing datasets may lead to different inferences on the existence of skilled and unskilled managers. Furthermore, this chapter aims to examine how the reduced form approach of Ferson and Schadt (1996) performs using our data. We show that the Ferson-Schadt conditional model produces similar alpha estimates compared to those using the unconditional models for our sample. Finally, we compare several bootstrap methods, namely bootstrap by residuals, bootstrap by pairs and block bootstrap, that are widely used in mutual fund studies.

Chapter 7 serves as a conclusion of this thesis. Our main contributions to the literature are summarized as follows: First, this study applies a non-parametric change point test to detect possible time-variation in fund alphas and factor loadings, which provides a novel perspective on how to address time-variation in these parameters, and enables us to overturn most of the pessimistic results in the literature, indicating strong evidence of manager skill. Second, our study is the first to address data issues in existing studies systematically. This is not without importance as different selection rules used to include funds can affect the inferences on the existence of skilled managers. Third, we compare several popular bootstrap methods used in previous mutual fund studies, providing guidance to future research in terms of which method best fits the mutual fund data. Fourth, we examine performance for funds in different investment styles, which provides a view on the relationship between funds' investment objectives and their performance. Fifth, we examine fund performance using both net and gross returns, to disentangle a fund's expense policy from examination of manager skill, providing new evidence on the role of expense policy in the underperformance of mutual funds. Finally, our study examines fund performance over an updated period from 1984 to 2015, which is most important to today's investors given the much changed investment climate in the wake of the global financial crisis.

Chapter 2

Background of the Mutual Fund Study

To appreciate the importance of mutual funds to U.S. investors, it is imperative to discuss their role in and their contribution to the U.S. financial industry. This chapter provides some insightful background data on their role, and how U.S. investors have come to rely on them. In particular, Section 2.1 reports some interesting trends in the U.S. mutual fund industry. This provides justification for our examination of actively managed equity mutual funds. Section 2.2 discusses active and passive management of mutual funds, and Section 2.3 explains the measure of fund performance that will figure prominently in coming chapters.

2.1 Recent Trends in the U.S. Mutual Fund Industry

Mutual funds can trace their origins to the Netherlands, where King William I established the first closed-end investment company in 1822. The next wave of near-mutual funds included the creation of an investment trust in Switzerland in 1849 and investment trusts in Scotland in the 1880s. Investment trusts were attractive as it brought the idea of pooling assets and spreading risks. Their success led the establishment of the first

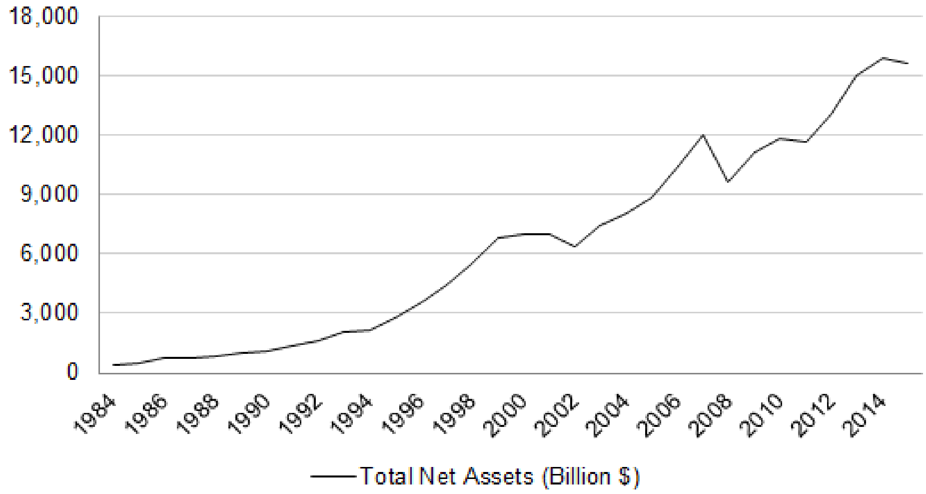

Figure 2.1: Total Net Assets of U.S. Mutual Funds From 1984 to 2015

Source: 2016 Investment Company Fact Book

investment trust in England and then to the U.S. in the 1890s. In 1924, the first modern mutual fund was established in Boston. Over the last few decades, mutual funds have become a popular investment vehicle, especially for U.S. households. They offer households a cost effective way to invest in a large portfolio of equities and bonds. More recently, their appeal has increased because employers have moved away from defined benefit pension plans to defined contribution plans. This shift in defined contribution pension plans has arisen largely because of the unfunded liabilities that frequently arise with defined benefit plans. Defined contribution plans transfer risks associated with investing to the pensioner away from the company to employees, who receive upon retirement the value of their contributions plus any investment gains or losses. This movement has increased households' demand for investments that can generate good returns without excessive risk. In response to this demand, a number of investment companies have entered the mutual fund market.

According to the latest Investment Company Fact Book, U.S. households held 89%

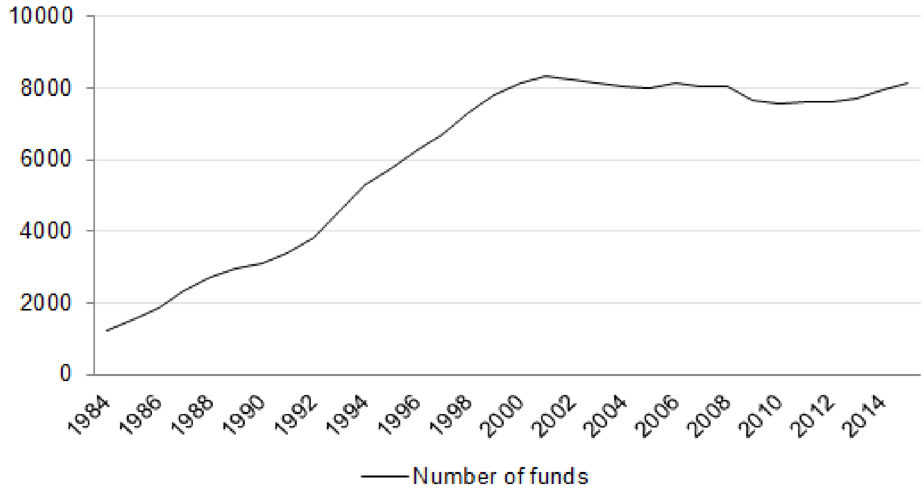

Figure 2.2: Number of U.S. Mutual Funds from 1984 to 2015

Source: 2016 Investment Company Fact Book

of mutual fund assets at the year-end 2015. 54% of defined contribution plan assets were invested in mutual funds, and mutual funds accounted for 48% of investment retirement accounts assets. U.S. investors have found mutual funds attractive, and this is reflected in size of the U.S. mutual fund industry. At the year-end of 2015, the U.S. mutual fund industry remains the largest in the world, with around $16 trillion in assets under management. This wealth accounts for almost half of the $33.4 trillion worldwide assets in mutual funds. U.S. mutual fund industry in terms of assets under management expanded significantly through the latter half of the 1980s, and this continued until 2014. The number of mutual funds also increased over this period with the peak occurring in 2014 and then leveling off, as shown in Figure 2.2.

Among these funds, as of year-end 2015, equity mutual funds made up around half of the U.S. mutual funds assets under management as shown in Figure 2.3. Money market funds account for 17%, bond funds for 22% and hybrid funds for 9%. Among equity mutual funds, actively managed funds account for 78% in terms of assets under

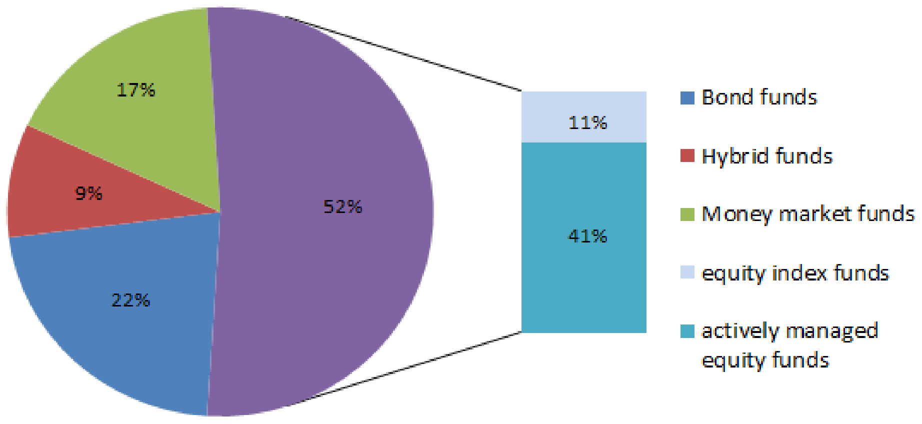

Figure 2.3: Share of Assets under Management for Different Types of Funds, Year-end 2015

Source: 2016 Investment Company Fact Book

management, and passively managed funds, mainly index funds, account for 22%.

Given the significant share of assets invested in equity funds relative to all mutual funds' assets, it is important that we examine their performance and explore their recent trends. The majority of these equity mutual funds are actively managed, and regularly advertise themselves as generating returns greater than those passively managed funds. Actively managed funds hire managers to select and to invest in a portfolio of investment products that aims to achieve a particular investment objective. In contrast, passively managed funds, mainly index funds, aim to track the performance of a certain benchmark or index. Although the majority of the equity mutual funds employ an active investment strategy, we have seen a growing demand of passively managed funds, mainly index funds in recent years, with share of assets invested in index equity funds increased to 22% relative to all equity mutual funds' assets at the year-end 2015, as shown in Figure 2.4. Given the growing popularity of index funds, more investors are advised to switch from

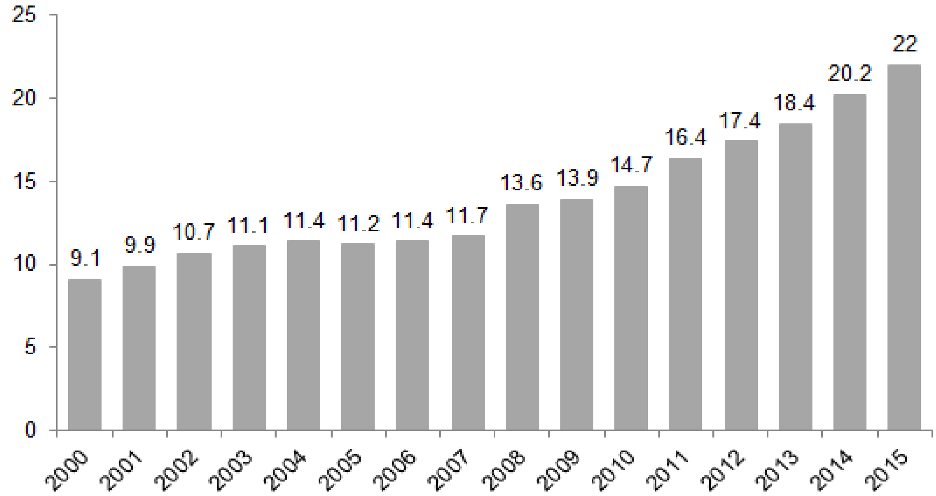

Figure 2.4: Share of Assets Invested in Index Funds from 2000 to 2015

Source: 2016 Investment Company Fact Book

actively managed funds to index funds.

The popularity of index funds is also reflected in the net new cash flow to index equity funds relative to all equity mutual funds. As shown in Figure 2.5, the equity mutual funds in total experienced several significant outflows from 2008 and realized a net total outflow of approximately $412 billion from 2008 to 2015.

In contrast, index equity funds received a net total inflow of around $473 billion. With growing demand of index funds in recent years, the total net assets held by index funds totaled $2.2 trillion in 2015 (Figure 2.6). Lower-than-average expense ratios of index funds play an important role in the growing popularity of index funds. The passive strategy followed by index funds means that trading costs are lower as are other costs, hence they tend to have lower-than-average expense ratios. Furthermore, index funds are averagely larger than active funds so that they benefit more from economies of scale, which also reduces their expense ratios. For instance, in 2015, the average index equity funds held $ 5.1 billion in assets under management, whereas the actively managed equity

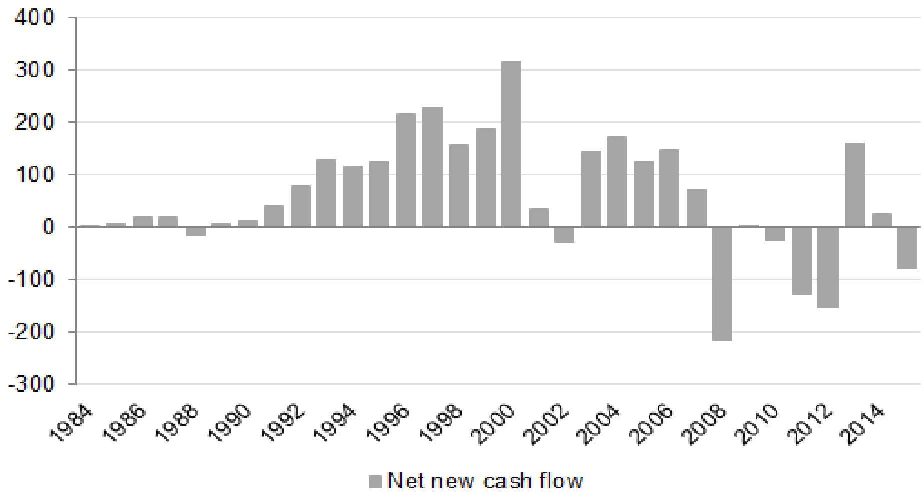

Figure 2.5: Net New Cash Flow of the U.S. Equity Mutual Funds from 1984 to 2015

Source: 2016 Investment Company Fact Book

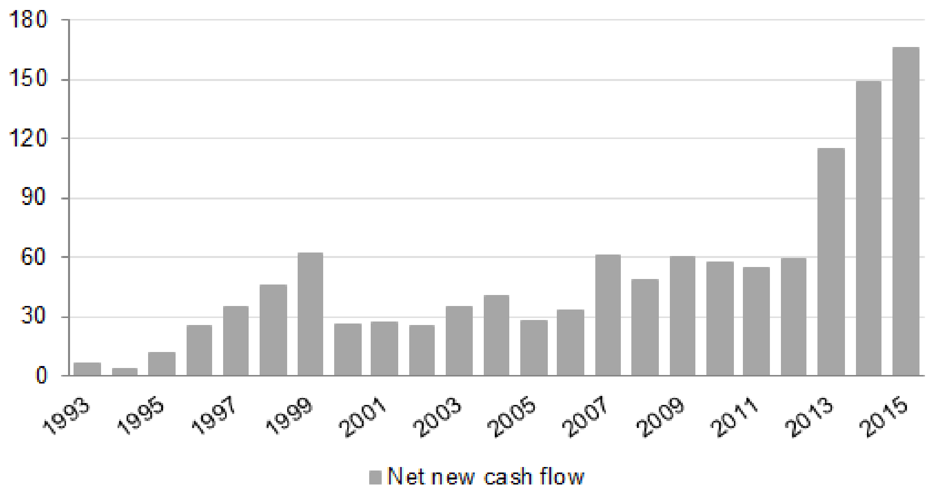

Figure 2.6: Net New Cash Flow of the U.S. Equity Index Mutual funds from 1993 to 2015

Source: 2016 Investment Company Fact Book funds on average had $1.4 billion in assets.

2.2 Active Management vs. Passive Management

Since our study aims to address the question of whether a group of skilled managers with stock selecting abilities exists, it is important to distinguish between passive and active management. For actively managed funds, managers create a portfolio consisting primarily of equities with the aim to outperform passively managed funds. An actively managed fund tries to invest in equities that offer superior returns, hence they incur higher trading costs as well as other administration costs. In contrast, a passively managed fund minimizes trading costs and mimics the performance of an index, which offers lower returns than those actively managed counterpart but they generally have lower management expenses.

Advocates of active management argue that the market can be inefficient and by selecting specific investment portfolios, fund managers are able to obtain higher-than-market returns. However, proponents of passive investing argue that, before all costs, passive investor must earn, in the aggregate, the market return as that is what their objective is. Active investors, however, are constrained by the algebra of active management (Sharpe, 1991), and so in the aggregate can earn only the market return. Therefore, after costs, passive investors earn the return of the market portfolio minus their expenses, and so do active investors. Given the average expenses of active investors are higher than those of passive investors [1], passive investors on aggregate earn higher net returns than those of active investors.

[1] According to 2016 Investment Company Fact Book, by the year-end 2015, the asset-weighted average expense ratio of actively managed equity funds is 84 basis point, whereas the expense ratio of index equity funds is 11 basis point.

In mutual fund studies, a comparison of active with passive management boils down to a debate on whether there is group of fund managers who have sufficient skill to outperform the passive benchmarks. Several studies confirm the underperformance of actively managed fund managers, and argue that there is no widespread manager skill (Jensen (1968), Henriksson (1984), Elton et al. (1993), Malkiel (1995), Carhart (1997), French (2008), and Fama and French (2010)). These authors conclude from their studies that investors would be better served if they switched from actively managed funds to passively managed funds with lower expense ratios. However, other studies conducted by Grinblatt and Titman (1989), Grinblatt and Titman (1992), Hendricks et al. (1993), Daniel et al. (1997), Kosowski et al. (2006), Wermers (2000), and Barras et al. (2010) argue that there are managers that, through actively selecting stocks, can outperform passive benchmarks. Because of this lack of agreement in the existing literature, it is therefore important to re-evaluate some of the more important contributions to research on this topic. By doing so, this thesis hopes to shed new light on this issue and hopefully can come up with more salient conclusions.

2.3 Performance Evaluation of Mutual Funds

To examine mutual fund performance requires an appropriate measure. Generally, there are two widely used approaches to measure manager skill. One is referred to as a portfolio holdings-based measure, and the second is returns-based. Even though the portfolio holdings-based approach allows performance to be measured at any time, it requires up-to-date information on fund holdings. Disclosure of this information may not be readily available or may come with long lags. Hence this method is not frequently used. Because of these concerns, the more frequently used method is the returns-based measure. The most widely used performance measure can be calculated from historical returns, and it is referred to as Jensen's alpha, named after the individual who first developed this measure.

2.3.1 Jensen's alpha and the Baseline Models

Before Jensen's alpha, fund performance as usually were measured by ranking returns on portfolios, Jensen referred to this as a relative measure of fund performance. He argues that it is better to have an absolute measure that is able to tell whether funds outperform relative to some benchmarks. For example, given two funds A and B, we not only would like to know if fund A outperforms fund B, but also whether fund A and B are better relative to a certain benchmark, hence use of an absolute measure allows more flexibility in the examination of mutual fund performance. The key idea of Jensen's alpha is to determine the abnormal return of a fund over the theoretical expected return, where the theoretical return is based on the asset pricing theory (the CAPM, cf. Equation 2.1) developed by Sharpe (1964). From an investment perspective, the slope (beta) on the explanatory returns in the CAPM describes a diversified portfolio of passive benchmarks that replicates the exposures of a fund to market returns, and the regression intercept (alpha) then implies the average return provided by the fund beyond the return on a comparable passive portfolio. Therefore, we interpret a positive alpha as good performance, and a negative one implies bad performance.

Dybvig and Ross (1985) provide a formal justification for this measure of fund performance. This justification is given in Theorem 5 of this paper. Specifically, given a risk-free security, this theorem indicates that if the intercept (alpha) of the CAPM for a certain fund is positive, there is a portfolio with positive weight on this fund plus the portfolio of the explanatory benchmarks that has a higher Sharpe ratio than the portfolio of the passive benchmarks. Similarly, if the intercept is negative, there is a portfolio with negative weight on this fund that has a higher Sharpe ratio than the portfolio of the explanatory benchmarks. Hence, the abnormal performance in Sharpe's sense implies the abnormal performance in Jensen's alpha.

Furthermore, the inferences on the performance of mutual funds can be sensitive to a choice of the passive benchmarks (Roll, 1978; Grinblatt and Titman, 1994; Daniel et al., 1997). Therefore, we also estimate fund alphas using multi-factor asset pricing models such as the Fama-French three-factor model (hereafter, 3F-CAPM, cf. Equation 2.2) and Carhart's four-factor model (hereafter, 4F-CAPM, cf. Equation 2.3). The 3F-CAPM has a size factor (SMB) and a value factor(HML) added to the CAPM based on the observation that stocks with small caps and low price-to-book ratio tend to perform better than the rest. Carhart (1997) extends the 3F-CAPM to a 4F-CAPM by adding a momentum factor (MOM) to account for the tendency of stock prices to increase further if it is going up and to continue decreasing if it is declining. We would like to estimate Jensen's alpha based on these factor asset pricing models to see if the alpha estimate is sensitive to the choice of factor models. More importantly, we aim to address the time-variation issue in the regression parameters of these factor models, which is a major contribution to the literature, and will be discussed in detail in Chapter 4.

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \varepsilon_{i,t}, \tag{2.1}$$

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \varepsilon_{i,t}, \tag{2.2}$$

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{i,t}, \tag{2.3}$$

2.3.2 Net Returns vs Gross Returns

In addition to the choice of different factor models to estimate Jensen's alpha, we have two types of returns in our study: net returns and gross returns. Net returns are returns net of all costs, hence results based on which allow us to examine whether fund managers have sufficient skill to generate returns that can recover all the costs. Furthermore, net returns are returns received by investors, the examination of performance on net returns is consistent with the fact that investors seek for actively managed funds to earn a large surplus net of all costs. For a mutual fund, costs incurred are consisted of trading costs, management fees, administrative costs, sales and distribution fees, investor servicing fees, fund accounting and pricing fees, and other operating costs. The CRSP database reports a fund's expenses as expense ratios. Most of the costs mentioned above are included in the expense ratios, however, there are still costs not counted in the reported expense ratios, mainly trading costs.

Given that a fund's expense policy may not be controlled by fund managers, but more likely by the board of directors of the fund's management company, we also examine manager skill based on gross returns. Gross returns are returns with expense ratios added back to net returns, and gross returns to some extent come closer to test if managers have any skill. For example, one manager may have enough skill to more than just cover the trading costs, but the surplus after trading costs can be offset by the management company that overcharges or generates other costs inefficiently (costs in the expense ratios), hence tests on gross returns enables us to disentangle a fund's expense policy from the examination of manager skill, and evaluation of fund performance on gross returns tells us whether fund managers have enough skill to generate returns to cover the costs missing from the expense ratios, mainly trading costs.

One may want the returns measured before all costs including trading costs so that funds are put on the same pure return basis as the passive benchmark portfolios, however, it is often too complicated to compute trading costs. Estimating trading costs requires information on all the selling and buying activities of fund managers, which is usually not disclosed to the public. Studies that attempt to use trading volumes to proxy trading costs are often criticized for their inaccuracy. Furthermore, the definition of skill on gross returns in our context seems fairly reasonable since one would expect a skilled manager to produce returns that can recover trading costs. In other words, the use of gross returns excluding trading costs does not affect the interpretation of our results.

Chapter 3

Literature Survey

In this chapter, we mainly review three streams of previous studies on mutual fund performance that are closely related to our research undertaken here. We first review studies that aim to examine whether funds are able to provide abnormal performance in Section 3.1. Then, we show how previous studies develop extended models to estimate fund alphas in Section 3.2. Finally, a review of studies of change point analysis is given in Section 3.3.

3.1 Review of Studies on Abnormal Performance

Two recurring themes in the evaluation of mutual fund performance have been explored in past studies. The first is whether funds are able to deliver abnormal performance and whether it is positive, negative or neutral. The second is whether this abnormal performance can be predicted and for how long it will persist. Our study focuses on the first issue. Previous studies tend to give contradictory results on whether actively managed mutual funds deliver abnormal performance or not. Many researchers believe that few funds succeed in repeatedly obtaining positive alpha, unless by chance, i.e.

mutual funds do not exhibit significant stock picking skill.

Jensen (1968) is among the earliest studies that examine mutual fund performance. He estimates the predictability of performance on a portfolio of 115 mutual funds during the period 1945 to 1964 using the CAPM, and finds that the mutual fund managers are not able to outperform a randomly chosen benchmark portfolio with equivalent risk. Besides, the results indicate that future good performance does not necessarily follow past good performance. Consistent with the findings of Jensen (1968), later studies conducted by Elton et al. (1993), Malkiel (1995) and Carhart (1997) confirm the underperformance of active fund managers.

Specifically, Elton et al. (1993) investigate the performance of mutual funds for the period from 1965 to 1984 using an extended version of the CAPM. They argue that several later studies, such as Ippolito (1989), have reached different conclusions from those of Jensen (1968) is because the variation of non-S&P stocks in different sample periods. Hence by adding a portfolio of non-S&P equity index into the passive benchmarks to control for the effects of non-S&P stocks, they show that fund managers underperform passive portfolios. However, these studies are criticized for only including a small number of mutual funds. Malkiel (1995) uses a dataset from 1971 to 1991 containing historical returns from all mutual funds that exist in each year of the period, and is among the earliest contributions that systematically addresses the issue of survivor bias and incubation. They estimated alpha estimates in the context of a CAPM framework and found that investors would be better off buying low expense index funds.

Carhart (1997) uses an earlier version of a CRSP survivor-bias free mutual fund dataset from 1962 to 1993, containing a total of 1892 equity mutual funds, and applies both the CAPM and his four-factor model to estimate fund alphas. By forming portfolios based on funds' lagged one-year return, he examined the short-term and long-term persistence in the performance of the cross-section of funds. His results generally indicate no evidence of skilled managers and suggest that persistence of performance detected in the literature is mainly due to momentum strategies rather than superior fund management. While he finds that the top-percentile funds do produce surplus that can cover investment costs, most funds underperform by about the cost of expenses.

Given the underperformance of actively managed mutual funds, these studies find the continuing investment into active mutual funds puzzling. It is argued that investors may behave irrationally and there are misperceptions about investing choices of investors influenced by the market. For instance, the overconfidence of investors in their ability to produce excess returns and the exaggerated advertising of active mutual funds by financial firms may also influence the choice of investors eventually. Malkiel (1995) discussed the possibility of a fund management company starting a number of "incubator" funds with different managers and waiting to see which ones succeed. If some of these "incubator" funds outperform the passive funds, the fund management company will advertise their success and drop the other ones that underperform the market.

Some investors are also found to be insensitive to a fund's prospect. These performance-insensitive investors still remain after the underperformance of funds and could be unaware that they have been overcharged (Christoffersen and Musto, 2002). Investors may also be constrained to invest in certain funds. For example, since many U.S. employers have moved away from defined benefit to defined contribution pension plans, as discussed in the background chapter, Elton et al. (2006) examined the investment choices offered by 401(k) pension plans [1] and found that almost half of the investors are not offered an

[1] A 401(k) plan is a defined contribution pension account established by employers to allow employees to direct their own investments given a group of investment choices to choose from. The earnings in this

adequate set of options. Elton et al. (2007) further examined the performance of these investment choices on offer and showed that they did not outperform index funds.

However, other studies have come to a different conclusion. To be specific, Grinblatt and Titman (1989) used a eight-portfolio benchmark based on past returns, dividend yield, and firm size constructed by Grinblatt and Titman (1988). They argue that forming benchmarks on portfolios grouped by securities characteristics as proxies for factors is more appropriate relative to the CAPM, given that firm characteristics are correlated with factor loadings. They have shown that Jensen's measure on this new benchmark for a sample of 274 funds from 1975 to 1984 supports the evidence of skilled funds, especially prevalent among growth funds, aggressive-growth funds or funds that have the smallest Net Asset Values (NAV) before costs. However, they also confirm that investors seldom gain from the abnormal performance of managers due to high expenses, i.e. funds do not show superior abilities in terms of net returns. Grinblatt and Titman (1992) further test the persistence of performance using this eight-portfolio benchmark. They argue that the persistence should be attributed to the abnormal performance of fund managers, not to persistence in the benchmarks that are related to size effect, dividend yield, etc. Hendricks et al. (1993) investigate the performance of mutual funds during the period of 1975-1988 by dividing their sample into eight performance-ranked portfolios and track their performance on different evaluation periods using Jensen's alpha. They claim that investors can gain from investing in previous winner funds, particularly those growth-oriented funds. However, these studies are usually criticized for having survivor-ship bias since delisted funds are not included in the dataset of Grinblatt and Titman (1992), and for including a small number of funds over a short period, as well as for inefficient

account are tax-deferred.

benchmarks (Elton et al., 1996; Daniel et al., 1997; Carhart, 1997).

Results of studies on performance of mutual funds in other countries are conflicting as well. Blake and Timmermann (1998) examine UK mutual fund performance based on a sample of 2375 funds during the period 1972-1995 using Jensen's measure. Their results suggest that UK funds on average underperform the benchmarks by about 1.8 percent. Ferreira et al. (2012) extend fund performance analysis to a worldwide scope based on a sample of actively managed funds in 27 countries. They use Carhart's four-factor model to estimate Jensen's alpha and the differences in the determinants of fund performance between U.S. and the other countries are controlled using fund characteristics such as fund size, management company structure, expenses, etc. They find that mutual funds around the world overall underperform the market. However, studies lead by Otten and Bams (2002) and Cuthbertson et al. (2008) confirm the existence of outperforming funds in European countries. Otten and Bams (2002) use a sample of 506 funds from 5 different countries, apply Carhart's four-factor model to estimate Jensen's alpha and conclude that most European mutual funds are able to produce positive abnormal performance. Similarly, Cuthbertson et al. (2008) examine UK mutual fund performance and use the cross-section bootstrap technique developed by Kosowski, Timmermann, Wermers and White (2006) (hereafter KTWW) to distinguish between skill and luck; they find that a small percentage of UK top performing funds have stock picking ability.

In this context, William Sharpe published an article in 1991, arguing that no empirical tests need to be taken to see if active funds on average outperform passive portfolios. Based on what he calls "the arithmetic of active management", one can have two statements on performance of mutual funds: (i) given that the passive investors mimic the market portfolio, and all U.S. stocks are held by either active investors or passive investors, the returns of active investors must also be equal to that of the market portfolio before any costs; (ii) Since the returns of both active and passive investors are equal to that of the market portfolio, and active investors generally pay higher fees and expenses for active investing, we can conclude that active investors underperform passive investors after costs. Fama and French (2010) (hereafter F&F) call this arithmetic of active investing "equilibrium accounting" and argue that active investing in the aggregate is a zero-sum game before all expenses and costs, and a negative-sum game after costs.

Sharpe (1991) also pointed out several reasons that could lead to the spurious evidence of active funds beating the index funds in the literature. He argues that some index funds may not be truly passive. For example, instead of investing in the market portfolio, some index funds may allocate their portfolios in resampled market choices. Another reason is that active funds may not be equivalent to the "non-passive" part of the market. For instance, some active individual investors are usually not included in the assessment of active investing in the literature. Additionally, many previous studies use an equal-weighted portfolio to represent all active funds and to evaluate their overall performance, which can be biased towards average outperformance if those small funds perform well during that period.

While accepting that mutual funds on average underperform passive benchmarks net of fees, whether there is a group of skilled funds that can outperform the market still remains an interesting question. Recent studies have shown contradictory results on this matter. By estimating the costs of active investing, French (2008) finds that the aggregate of all active funds produces lower return than the passive market portfolio, and irrational behavior of investors may explain why active investing is still popular. In F&F, they examine the performance of both the aggregate portfolio and individual funds, but find no evidence of manager skill on average. If there are fund managers with enough skill to cover costs, they are balanced by the many managers with insufficient skill. Opposing findings can be found in Daniel et al. (1997), Wermers (2000), KTWW, and Barras, Scaillet and Wermers (2010) (hereafter BSW), which agree the existence of managers' skill to generate a positive alpha, especially among growth-oriented fund managers.

In Daniel et al. (1997), a different benchmark portfolio is used by directly matching the characteristics of the component equities of the portfolio being examined. They argue that this benchmark can better capture the characteristics of stocks held by the portfolios. Specifically, they use passive benchmark portfolios that have similar characteristics like market capitalization, book-to-market, and prior-year returns compared to those stocks held by the portfolios that are evaluated. Their results suggest that mutual funds, on average, do not outperform passive funds, but there are a few funds, particularly aggressive-growth funds showing some stock picking abilities. Wermers (2000) decomposed fund returns and costs into several components such as stock-picking talent, style, transactions costs, and expenses. Their results of Jensen's alpha indicate that mutual funds held stock portfolios that outperform the benchmark index by 1.3 percent per year. KTWW used a cross-section bootstrap procedure to distinguish skill from luck to infer the existence of skilled managers and found that a few fund managers are able to produce a positive alpha sufficient to cover their costs and that this stock picking skill persists.

In BSW, an FDR technique is used to examine how many funds are able to produce excess returns that are positive, negative or zero. Although they confirm that the proportion of skilled managers decreased rapidly over the last 20 years, most actively managed funds are still able to produce either positive or zero net-of-expense alphas.

They attribute the underperformance of active portfolios to a minority of continuously underperforming funds. However, whether survivorship-bias exists in the latter stream of studies remains in doubt. For example, in BSW, funds are kept only if they have a return history of at least 60 months, so there is some survivorship-bias in their study since funds with short return histories of less than 60 months are excluded from the sample.

A well-known paper by Berk and Green (2004) provides another insightful view on the existence of skilled managers. They argue that the provision of capital to funds provided by investors is competitive, which means that investors are fully aware of funds' past performance and move capital to funds with positive excess returns and leave those with negative excess returns. Fund managers are assumed to exhibit decreasing returns to scale and in deploying their superior ability, so that they will continue to expand the size of funds and their compensation until the excess returns become zero. Hence, in the model of Berk and Green (2004), managers on aggregate do not outperform the passive benchmarks and their outperformance does not persist. However, this is not because fund managers do not have skill or they are unrewarded for their superior ability, but because investors provide capital to funds competitively. Motived by these controversies in the literature, our study aims to revisit the issue of whether there is a group of skilled or unskilled fund managers in later chapters.

3.2 Review of Studies on Extended Models in Addressing Time-variation

Although the above traditional factor models, including the CAPM, the 3F-CAPM and the 4F-CAPM, are straightforward to implement and interpret, they generally neglect the possibility that fund alpha and beta estimates can vary over time. In fact, conditional performance evaluation approaches have been widely examined for the past few decades to address the time-variation issue. The appeal of conditional models to estimate fund alphas can be mainly attributed to the fact that the traditional models are not able to capture the dynamic trading strategies of fund managers.

Studies that attempt to incorporate time-variation into the assessment of mutual fund performance can be dated back to the 1970s. Among the earliest studies, Kon and Jen (1978) address the time-variation in a fund's systematic risk by using a "switching regression model" to allow for time-varying betas. They prove that the switching regression estimates of mutual fund performance are unbiased risk-adjusted measures. This type of conditional model usually incorporates business cycles or different regimes into the evaluation of fund performance. Instead of setting two or three regimes arbitrarily, Fabozzi and Francis (1979) examine mutual fund manager skill over different market conditions, in particular, betas are compared for bull and bear market, which is motivated by the assumption that fund managers may adjust risk exposures (beta) in anticipation of a coming bull market.

A more recent study by Kosowski (2011) develops a regime-switching model that accounts for the alpha difference between recession and expansion periods, which is similar to the setting of bull and bear market. He argues that traditional unconditional models fail to adjust fund performance in the recession periods when the marginal utility of wealth invested is relatively high for investors. Therefore, while mutual funds may underperform in the expansion periods, they are able to add value by providing a diversified equity position that does not underperform benchmarks in recession periods. Kacperczyk et al. (2013) conducted a study to test time-varying fund manager skill by conditioning on the states of business cycles suggested by the National Bureau of Economic Research (NBER) recession indicator. They find that managers with time-varying skill generally outperform the passive benchmarks by 50 to 90 basis points per year, processing information on either firm-specific or economy-wide shocks.

A widely used conditional model developed by Ferson and Schadt introduces lagged information variables to the traditional CAPM and allows the market portfolio beta to be time-varying to account for business cycles. Information variables are added in a way that the market beta depends linearly on them. They assume that fund managers may use public information to form their portfolio strategies and indeed find evidence that risk exposures tend to change in response to public information, and that these information variables are statistically and economically significant. Therefore they argue that the previous results indicating average poor performance of mutual funds (Jensen, 1968; Henriksson, 1984; Elton et al., 1993; Carhart, 1997) are mainly attributable to the setting of time-invariant factor loadings. Using their conditional approach, the distribution of fund alpha estimates shifts to the right and is centered near zero, which differs from the prior results that alphas are negative more often than positive.

Kacperczyk et al. (2005) use the Ferson–Schadt conditional measure to study the relation between industry concentration and performance of U.S. active mutual fund performance. They find that managers that hold portfolios concentrated in a few industries tend to perform better, and using the conditional measure of performance yields stronger and statistically more significant results than the unconditional model. However, other studies argue that this conditional measure produces remarkably similar results to those using the traditional unconditional measures. KTWW examine fund performance based on 15 models, including the Ferson–Schadt conditional model. They find that conditional four-factor alphas produce similar rankings of funds to those based on unconditional four-factor alphas. Similarly, Blake and Timmermann (1998), Kacperczyk et al. (2005), Cuthbertson et al. (2008), BSW and Cuthbertson et al. (2012) also implement the Ferson-Schadt conditional model to control for time-varying exposure to market returns. They argue that introducing time-varying betas provides similar results to those obtained from the traditional unconditional models, which shows that alphas are negative for each group of funds specified in their study.

This suggests that the reduced form approach of Ferson and Schadt (1996) may not be an appropriate way to model time-variation in the alpha and risk factors. Hence the method we use here takes an agnostic view on the nature of the relation between risk factors and the information variables. To do this, we employ a non-parametric statistical method to detect time-variation in these parameters. For this purpose, we also provide a brief review on the studies of change point analysis, and explore the reasons why we use the particular test we choose over other ones.

3.3 Review of Studies on Change Point Analysis

As shown in the previous section, many studies have attempted to develop more advanced models to account for time-variation in the regression parameters, and they usually require a prior setting of business cycles or economic states. Here, we explore an alternative approach to account for time-variation in the regression parameters that is based on change point analysis.

According to Horváth and Rice (2014), change point analysis is to detect and estimate changes in a quantity of interest based on a sample of data. Change point analysis originated in the 1940s and previously focused on quality control issues (Page, 1954). Later, it was developed and extended to a variety of fields such as economics, finance, engineering, biology, etc. and has become an important tool to understand the non-stationarity in time series. Over the last 70 years, change point analysis literature has witnessed huge developments in terms of discovering new methods and applications. Among the many applications, it is of importance and interest to detect multiple parameter changes that may occur in linear regression models, which is the case that our study is concerned with. It is widely acknowledged that the parameters in the linear regression models may change over time due to abrupt events such as policy changes, oil price and technology shocks. Similarly, various sources could lead to changes in parameters in the factor models implemented in this study to estimate mutual fund alphas and will be discussed in detail in later chapters.

The earliest test to detect changes in a time series context, specifically in the linear regressions, is the Chow test (Chow, 1960). It is introduced to detect single change in mean for a linear model that is known *ex ante*. The test is motivated by the question of whether an economic relationship is stable over two periods of time or for two groups of economic units when it is represented by a linear regression model. The idea of the Chow test is to see if two sets of observations belong to the same regression model. If not, then a change is confirmed and two linear regressions with two subsets of coefficients are used so as to better capture the economic relationship.

It should be noted that the Chow test requires knowledge of the location of the change while modern tests don't usually have this requirement. Furthermore, the Chow test focuses on the case of a single change; however, if there are multiple changes or if the dates of changes are unknown in a analysis, this test is no longer feasible. Hence, a large amount of work has been done on testing a parameter with an unknown change point, such as Andrews (1993), Andrews and Ploberger (1994), and Bai (1994). Studies such as Andrews et al. (1996), Garcia and Perron (1996), Liu et al. (1997), Bai and Perron (1998), Bai and Perron (2003), and Aue et al. (2009) aim to address the case of multiple changes.

In order to detect changes in both regression slopes and the intercept, which fits the scenario of our study, we implement the test developed recently by Pouliot (2016), a non-parametric test based on traditional cumulative sum control chart (CUSUM) tests. The CUSUM technique is one of the widely used approaches for detecting change points; it is originated from Page (1954), and a large amount of CUSUM tests are widely available (Brown et al., 1975; Andrews, 1993; Andrews and Ploberger, 1994). However, unlike existing CUSUM tests (Andrews, 1993; Andrews and Ploberger, 1994; Horváth and Rice, 2014), the test of Pouliot (2016) is able to distinguish a change in intercept from one in the slope parameter of the regression model. This is important in our context since the intercept is the fund performance measure that is referred to as Jensen's alpha, and the slopes are betas representing risk exposures to factors considered in the benchmarks. A change in the alpha would lead to very different economic interpretations from that in the slope parameters.

Furthermore, Pouliot (2016)'s test has some additional appealing features. Unlike some existing tests (Andrews, 1993; Andrews and Ploberger, 1994; Bai and Perron, 1998), it avoids trimming. The test also develops a weight function so that changes occurring at the limits of the sample can be detected. More importantly, different from other tests such as Gombay (2010) and Horváth and Rice (2014), this test does not require prior knowledge of which parameters might have changed.

It should be noted that BSW introduce time-varying betas by partitioning their data into six 5-year subintervals, and evaluate fund performance over the short run, which is similar to our way of treating time-variation in the regression parameters. As a result, they find that a small proportion of funds show skill over the short run relative to the fact that long-term skill does not exist. However, any attempt to partition the time-series into successive intervals would impose the prior assumption that parameters are constant for such subintervals (Kon and Jen, 1978). To overcome this problem, we estimate fund alphas over subsamples separated by the estimated dates of changes detected in the regression parameters, which is more rigorous and intuitive [2]. The closest approach to ours is proposed by Bollen and Whaley (2009), in which they develop a optimal change point regression based on Andrews et al. (1996) optimal change point test to allow for time-varying beta, finding 40% of the funds experienced significant changes in factor loadings. However, they only allow for one change for a fund in the betas over the entire sample period, which may not apply to some funds that have multiple changes or, more importantly for our purposes, funds that have changes in the alpha.

[2]Note that our way of treating time-variation in the regression parameters does not change the nature of Jensen's alpha measure, i.e. the various versions of the CAPM are still implemented and an ordinary least squares (OLS) estimation is undertaken to get alpha estimates, but on subsamples of data that are divided according to the estimated dates of changes in the parameters.

Chapter 4

Luck vs. Skill in Time-varying Mutual Fund Performance

4.1 Introduction

As shown in the literature survey chapter, whether there are any active mutual funds that are able to outperform the market is an interesting research topic that has been repeatedly examined over the past 50 years, going back as far as Jensen (1968). More recent studies tend to come to contradictory conclusions on this question. Berk and Green (2004), Wermers (2000), and Barras, Scaillet and Wermers (2010) (hereafter BSW) believe that mutual funds hold equity portfolios that generally outperform the passive benchmarks, indicating that there is a proportion of skilled fund managers who are able to achieve abnormal returns. French (2008) and Fama and French (2010) (hereafter F&F), argue that there is no evidence of certain fund managers being able to outperform the market, with skilled fund managers "hidden" by massive underperforming funds. Motivated by these existing controversies in past studies, we revisit the important question of whether manager skill exists here using a rich panel of 5006 U.S. actively managed equity mutual funds for the period from January 1984 to March 2015.

We aim to contribute to the literature by addressing the time-variation in the regression parameters in the various versions of the CAPM that are commonly used to estimate fund alphas. In previous studies, the traditional models with time-invariant regression parameters are often used. However, given the fact that fund managers may adopt dynamic trading strategies, for instance, the factor loadings (betas) can vary over time. In particular, F&F confirm that time-variation in betas poses a "thorny" issue that needs to be addressed in future research. To capture time-varying market beta, Ferson and Schadt (1996) develop a conditional evaluation approach that allows the market beta to depend linearly on some information variables. However, later studies such as Carhart (1997), Kacperczyk et al. (2005), Kosowski, Timmermann, Wermers and White (2006) (hereafter KTWW), BSW and Cuthbertson et al. (2012) apply the Ferson-Schadt conditional model to estimate fund alphas and get similar alpha estimates to those generated by the unconditional models.

More importantly, we argue that fund alphas can also change. For example, a mutual fund may have many managers of differing abilities throughout its entire history, hence the manager skill (measured by Jensen's alpha) of a fund can be time-varying, which is often neglected in previous studies. Furthermore, Berk and Green (2004) argue that fund managers exhibit decreasing returns to scale and in deploying their superior ability due to the fact that the provision of capital to funds provided by investors is competitive, hence manager skill can be time-varying. Given that various sources could lead to the time-variation in the alpha and betas, this study implements a non-parametric test to detect changes in them without imposing any specific functional form on the nature of the time-variation in these parameters. In particular, we apply a change point test developed by Pouliot (2016) to detect possible changes in the alpha and betas; this test has many attractive features over other existing tests. It allows us to distinguish between a change in the slope and a change in the intercept without trimming, which is important in mutual fund studies given that the intercept (alpha) measures manager skill.

After allowing for time-variation in this way, we find that there is around 17% of the funds in our entire sample having at least one change in either the alpha or the betas. When these changes are incorporated into the regression model and a cross-sectional bootstrap technique of F&F is implemented, more skilled fund managers are shown to have sufficient skill to more than just cover the costs, and the distribution of the cross-sectional estimated $t(\alpha)$ estimates shifts rightwards for tests on both net and gross returns. Hence, we can conclude that results of prior studies based on the traditional models can be biased, and the failure to capture time-variation in the regression parameters biases the inferences of F&F towards more pessimistic conclusions that fund managers underperform the passive benchmarks net of all costs. Our use of change point analysis to address time-variation in the regression parameters sheds new light on this issue, which is of utmost importance in measuring time-varying fund performance.

The rest of this chapter is organized as follows: Data used in this chapter are briefly described in Section 4.2. Section 4.3 introduces the test of Pouliot (2016) to detect change points and the multi-regression model framework. Section 4.4 explains the simulation setup. Section 4.5 then presents the results. In Section 4.6, we contrast our results and those of F&F restricting the sample period to 1984-2006 as in F&F. The final section concludes this chapter.

4.2 Mutual Fund Data

Data used here as well as in Chapters 5 and 6 are taken from the CRSP mutual fund database for the period from January 1984 to March 2015. Appendix A provides a detailed account of how the final dataset is fashioned, so we shall be brief here. To summarize, first, to focus on active management, we include only actively managed mutual funds that invest primarily in U.S. common stocks. We exclude all the other funds including international funds, global funds, municipal funds, bond funds, money market funds and all sector funds since these kinds of funds generally invest a small portion into domestic equities. Furthermore, all kinds of passively managed funds, mainly index funds are excluded from our sample.

Then, for each fund, we merge different share-classes of the same fund into a single fund by value weights. In order to lessen the effects of "incubation" bias, we follow F&F's methodology to limit our attention to those funds that hold assets under management (AUM) of at least 5 million 2006 dollars. Accordingly, a fund is included in the sample, for example, after adjusting for inflation, if it reaches 2.5 million dollars in AUM in 1984, or if it has at least 6 million dollars in AUM in 2015. According to F&F, this minimizes the "incubation" bias because funds in the pre-release period tend to have low AUM. To further examine if our results are biased towards small funds, we also perform regressions on funds that pass the $250 million and $1 billion AUM bound. Our panel is limited to funds that exist in CRSP at least 5 years before the end of our sample period (March 2015) so as to remove new funds with too short return histories. We also exclude funds that have less than 8 months of return history. Overall, we have 5006 funds that span a total of 375 months in the $5 million AUM group, 2376 funds in the $250 million AUM group, and 1093 funds in the $1 billion AUM group.

4.3 The Model with Time-varying Parameters

4.3.1 Basic Assumptions

The appeal of a model to capture time variation in the regression parameters in our context of mutual fund study can be explained with the following hypothetical scenarios. Rationales for time-varying betas are well documented in the literature, such as the dynamic trading strategy a manager may carry out (Ferson and Schadt, 1996). Therefore, calculating the average alpha based on fixed beta estimates through the entire fund history does not account for the fact that factor loadings may vary over time, and consequently will lead to unreliable results.

As to the intercept (Jensen's alpha), it is reasonable to assume that the performance of funds may also vary over time. One possible hypothesis for time-variation in fund alphas is the replacement of fund managers. For instance, according to BSW, some fund managers with superior track performance left the sample for hedge funds during their period of analysis. Throughout the entire history of a mutual fund, it may have many managers of differing abilities and hence the performance of the fund can be time-varying. In this case, estimating one alpha to reflect fund performance would be incorrect and the estimated alpha would be inconsistent since the alpha would have changed with each manager.

There is a rich literature on whether the characteristics of managers have an effect on the performance of mutual funds, such as gender, education, and experience of managers. There is evidence that education and experience are positively correlated with fund performance, and that male managers tend to perform better than female ones. Khorana (2001) use a sample of U.S. equity and bond fund managers that were replaced over the

1979-1991 period and found significant improvements of post-replacement performance relative to the past performance for those underperforming funds. Based on a sample of UK mutual funds, Clare et al. (2014) found similar evidence.

Their results suggest that manager replacement does affect the performance of mutual funds, and that there is a significant improvement of fund performance after manager change for those underperforming funds. On the contrary, new managers of outperforming funds are usually unsuccessful at obtaining equivalent performance as the replaced well-performing managers. This evidence implies that, firstly, for some funds, their performance varies due to replacements of managers; secondly, if one alpha is estimated to reflect manager skill of this fund, the achievements of the replaced outperforming managers are "buried" by the underperformance of new managers. The traditional factor models will therefore measures not the skill of a manager but average performance that is achieved throughout a fund's life. However, whether a fund manager can deliver abnormal performance is of more importance to investors. Hence, the assumption of time-varying manager skill, in our case, alpha estimates that are allowed to have discrete changes, disentangles the examination of fund performance when a fund performs relatively well from that when it performs badly.

Another important notion is the competitiveness in fund flow supply from investors. In the model of Berk and Green (2004), it is argued that investors provide capital to funds competitively, i.e. capital is provided elastically and moves to funds with positive excess returns and leaves those funds with negative excess returns. Managers increase assets under management until excess returns are zero. Therefore, in this model, managers exhibit decreasing returns in deploying their superior skill and excess returns do not persist. This does not therefore suggest a lack of skilled managers or lack of rewards for their differential ability, but it merely implied that manager skill is compensated accordingly on past excess returns. Therefore, a test for change points allows us to evaluate fund performance over the short run before investors compete away their performance with capital inflows, i.e. the multi-regression model evaluates fund performance before and after fund inflow changes separately.

4.3.2 Model Framework

To capture time-variation in the alpha and betas, we construct a multi-regression model that is based on a non-parametric change point test developed by Pouliot (2016). To discuss the non-parametric statistical procedures, we need a baseline model to perform the change point test. The framework of the baseline models are shown in Chapter 2, Section 2.3.1 and we briefly reproduce them here:

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{i,t}, \qquad (4.1)$$

where $R_{i,t}$ is the return for month t of fund i, R_t^f is the risk-free rate, which is the 1-month U.S. Treasury bill rate, R_t^m is the market return constructed with a VW portfolio of NASDAQ, NYSE and Amex stocks, SMB_t and HML_t are the size and value-growth returns constructed following Fama and French (1993), MOM_t is the momentum return of Carhart (1997), α_i is the return left unexplained by this benchmark model, i.e. the extra return generated by funds and $\varepsilon_{i,t}$ is the regression residual. The above model without momentum factor is the 3F-CAPM. All the benchmark factors are obtained from Kenneth French's website.[1]

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

In the framework of F&F, it is assumed that the parameters in Equations 4.1 are fixed over the entire sample period. Therefore, Jensen's alpha (α) is estimated as the past average excess returns, minus the average excess returns of the benchmark portfolio times a fixed vector of betas. To detect possible changes in the parameters of Equation 4.1, we use Pouliot (2016)'s test.

Appendix 4.A shows a detailed account of how the test statistics are constructed and how the test is implemented. To summarize, first, he has constructed a process based on a *U-statistic* (unbiased statistic) that can be interpreted as comparing the residual sum of squares under the null hypothesis of no change. This statistic then yields an estimator of a change point. Since this statistic does not tell which parameter changes, to further distinguish between rejections in the slope and in the intercept, another two auxiliary *U-statistic* type processes are constructed, in which the first statistic is sensitive to a one-time change in the slope while remaining robust to a possible change in the intercept and the second statistic is sensitive to a one-time change in the intercept that is robust to a possible change in the slope. In order to improve the power of this test to make it sensitive to change points that occur at the limits of the sample period, weights are added to the statistics through a weight function so that the test has power over the whole (0,1) range.

This test can also be implemented to detect multiple changes. To achieve that, a dynamic programming approach is adopted here. We follow Vostrikova (1981)'s hierarchic binary segmentation algorithm, which begins with a test statistics, like CUSUM in this context, to detect one change point over the entire sample. If the null of no changes in the parameters cannot be rejected, we conclude that no change points are found in the data. However, if the test rejects the null hypothesis of no change in parameters, then the date of the change point is estimated. The sample data is then divided into two segments: one consisting of data up to the estimated date of change, and the second consisting of data after the estimated date of change. Then the test is applied to these two segments to determine if an additional change occurs in each segment. If no additional change point is detected, then the algorithm stops. However, if the test rejects the null, we estimate the date of the change point within the segment, and the data is separated further, and tests are applied again. The test is implemented sequentially in this way until no further changes are found. We find this approach efficient and fast with little computation cost and this is also confirmed by Bai and Perron (2003).

This particular test has three main advantages in the application in our context. First, this test is able to detect more than one parameter change within a linear regression model using the binary segmentation. This is useful given that some funds may have multiple changes in their factor loadings and alpha. Second, this test allows us to distinguish between a change in the slope and a change in the intercept while remaining robust to each other. This is particularly important as the intercept and the slope has very different interpretations in our study. Information on this tells us which changes are due to time-varying skill and which ones occurred because of shifts in the factor loadings. Third, the test has power against parameter changes that are located at the limits of the sample period by using weight functions, which implies that this test is able to capture changes that occur early or late in the alpha and betas.

A multi-regression model is further constructed after dates of change points for the regression parameters are estimated. We show the framework of this model according to the following different cases:

If the test for change points does not reject the null hypothesis of no change in α or

βs, then the traditional 4F-CAPM is used:

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{i,t}. \quad (4.2)$$

If however, the null hypothesis of no change is rejected, and there is one change in the fund alpha, based on the estimated date of the change, the sample is divided into two segments: one segment up to and including the estimated date of the change, and a second segment after this date. Hence, there would be two equations corresponding to Equation 4.2 and the following multi-regression model is estimated:

$$R_{i,t} - R_t^f = \begin{cases} \alpha_i^{(1)} + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{i,t} \\ \hfill t_0 \leq t \leq \widehat{t^*} \\ \alpha_i^{(2)} + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{i,t} \\ \hfill \widehat{t^*} + 1 \leq t \leq t_T. \end{cases} \quad (4.3)$$

Instead, if there is only one change in the slope β^2, the following model is estimated:

$$R_{i,t} - R_t^f = \begin{cases} \alpha_i + \beta_{1i}^{(1)}(R_t^m - R_t^f) + \beta_{2i}^{(1)}SMB_t + \beta_{3i}^{(1)}HML_t + \beta_{4i}^{(1)}MOM_t + \varepsilon_{i,t} \\ \hfill t_0 \leq t \leq \widehat{t^*} \\ \alpha_i + \beta_{1i}^{(2)}(R_t^m - R_t^f) + \beta_{2i}^{(2)}SMB_t + \beta_{3i}^{(2)}HML_t + \beta_{4i}^{(2)}MOM_t + \varepsilon_{i,t} \\ \hfill \widehat{t^*} + 1 \leq t \leq t_T. \end{cases} \quad (4.4)$$

For more complicated cases, if there is one change in the intercept and one change in the

[2]It should be noted that although the test is able to tell a change in the slope from one in the intercept, it cannot tell which one of the regression slopes changes. Hence, when there is a change in the slope, we assume all the regression slopes change. This setting still gives us consistent estimates of betas, only that the estimates are not efficient.

slope, the following model is estimated:

$$R_{i,t} - R_t^f = \begin{cases} \alpha_i^{(1)} + \beta_{1i}^{(1)}(R_t^m - R_t^f) + \beta_{2i}^{(1)} SMB_t + \beta_{3i}^{(1)} HML_t + \beta_{4i}^{(1)} MOM_t + \varepsilon_{i,t} \\ \hfill t_0 \leq t \leq \widehat{t}_1^* \\ \alpha_i^{(2)} + \beta_{1i}^{(1)}(R_t^m - R_t^f) + \beta_{2i}^{(1)} SMB_t + \beta_{3i}^{(1)} HML_t + \beta_{4i}^{(1)} MOM_t + \varepsilon_{i,t} \\ \hfill \widehat{t}_1^* + 1 \leq t \leq \widehat{t}_2^* \\ \alpha_i^{(2)} + \beta_{1i}^{(2)}(R_t^m - R_t^f) + \beta_{2i}^{(2)} SMB_t + \beta_{3i}^{(2)} HML_t + \beta_{4i}^{(2)} MOM_t + \varepsilon_{i,t} \\ \hfill \widehat{t}_2^* + 1 \leq t \leq t_T \end{cases} \quad (4.5)$$

A similar process is followed when we have more than two parameter changes. For brevity, we only list the above four cases. As we can see, after a change point is detected, fund return history is divided into subsamples according to the estimated date of the change and one regression is estimated on each of these subsamples. It should be noted that there might be changes very close to the prior change or located near the beginning or the end of the sample, which leaves us with only a few observations in the subsamples, which is not sufficient to obtain accurate parameter estimates. Therefore, we assume there is no change and proceed to estimate based on only one equation.

As mentioned previously, Ferson and Schadt (1996) model time-varying parameters as a reduced form so that market beta depends linearly on the information variables. However, the correct relationship between beta and these information variables is difficult to be tested. Financial theory offers little guidance here and there is difficulty in estimating nonlinear models. Therefore, instead of taking a particular view on the precise nature of this relationship, we leave it unspecified and allow for discrete changes in the regression parameters and build a multi-regression model that includes piecewise linear regressions. Our approach is arguably more appealing given that it allows for more general types of time-varying regression parameters.

4.4 Simulations

As argued in the literature survey chapter, actively managed funds on average underperform passively managed funds after costs according to "the arithmetic of active management" (Sharpe, 1991). However, there might be funds with skilled managers that are "hidden" by the mass of unskilled managers. Therefore, it is necessary to look at the performance of funds individually to detect the existence of skilled and unskilled funds. However, as argued by F&F and KTWW, the difficulty of examining individual fund performance is to distinguish manager skill from luck. Given the magnitude of the funds in the sample, there might be extreme returns obtained by chance. In order to separate skill from luck to see if there are some skilled managers, we use simulations, following the bootstrap procedure developed by KTWW and F&F.

To put it simply, we first obtain the actual α estimate for each fund using funds' historical monthly returns. Equation 4.1 is estimated for each fund in the sample individually, for the 3F-CAPM and the 4F-CAPM, for net returns and gross returns, and for \$5 million, \$250 million and \$1 billion AUM groups. Therefore, we have 12 groups of funds in the simulations. For each group, a cross-section of actual α estimates is obtained by estimating alpha of every fund in the sample. Then, we assume the skill (true α) is zero in the simulations. To achieve that, we subtract the actual α estimate from each fund's monthly returns to get the adjusted returns to reflect that. After that, we resample jointly from the adjusted monthly returns and all regressors to produce a sequence of resampled returns with the same length as that of the actual returns. To preserve possible cross-sectional correlations among funds, the resampling is taken for all funds simultaneously to produce a sequence of resampled returns for each fund. Equation 4.1 is estimated again for each fund on the adjusted returns to obtain the simulated

α estimates.

For each simulation run, a cross-section of simulated α estimates is obtained. Instead of using α estimates, t-statistics of α estimates are used to control for differences in the number of months funds are in a simulation run, and in residual variance. Finally, 10,000 simulation runs are conducted to get 10,000 cross-sections of $t(\alpha)$ estimates for each one of the 12 different groups, for a world where true α is zero. Hence, these $t(\alpha)$ estimates are not obtained from regressions on actual returns, but on adjusted returns in which the actual α estimates are subtracted. In this way, a "resampled world" or "simulated world" is constructed where the true skill is set to zero and only luck exist.

Having set true α equal to zero has different interpretations of skill for net returns and gross returns. In terms of net returns, zero true α for a fund means that the manager has just enough skill to compensate for all the costs. As to gross returns, zero true α implies a world where every manager has just sufficient skill to cover costs that are missed in the expense ratios, mainly trading costs.

For each group, the cross-section of actual $t(\alpha)$ estimates can then be ordered into a cumulative density function (CDF). After having generated 10,000 cross-sections of simulated $t(\alpha)$ estimates, each cross-section of $t(\alpha)$ estimates can also be ordered into a CDF. To infer the existence of skilled and unskilled funds, we produce and compare qualitatively the percentiles of empirical CDF of actual $t(\alpha)$ estimates and the average percentiles of CDFs of simulated $t(\alpha)$ estimates across 10,000 simulation runs.

Finally, we calculate how many of the 10,000 simulation runs generate smaller values of $t(\alpha)$ estimates at the selected percentiles than those obtained from actual fund returns. Percentages of the runs against a total 10,000 simulation runs are further computed. The percentages, in other words, likelihoods, help to infer more formally whether the tails of the cross-section of $t(\alpha)$ estimates from simulations are extreme compared with what we observe for actual fund returns.

It should be noted that no further change point tests are conducted in the simulations. Since bootstrap requires resampling of individual fund returns, and to allow for time-variation in the regression parameters in the simulations, we estimate the dates of changes in the regression parameters and partition the sample into segments according to the estimated dates of the changes, and then treat each segment as a separate "fund". Therefore, we have more "funds" in the simulations, only some of the "funds" are not real ones but subintervals of funds' return histories. Accordingly, the sequence of fund returns we resample from is constructed using the full return series of funds with no changes and segments of those with changes in the regression parameters. In this way, we are able to deal with the existence of several alpha and betas of the same fund (due to having obtained change points in alpha and betas) when generating the cross-section of bootstrapped $t(\alpha)$ estimates.

The advantage of this simulation approach is that it tells us whether there is a group of skilled managers without trying to identify which funds are skilled. The main idea of this bootstrap design is to construct and compare two sets of world: (i) one "real world" of $t(\alpha)$ estimates on actual returns in which manager skill and luck coexist, which prohibits us from telling whether managers produce extreme returns by skill or chance; (ii) the other one referred to as a "simulated world" of $t(\alpha)$ estimates on resampled returns, where we set the skill to zero (zero α) *ex ante* and only luck is left.

Comparing these two sets of worlds gives us inference on whether there is a group of skilled or unskilled funds. If the percentiles of actual $t(\alpha)$ estimates are larger than the average values of percentiles from the simulation runs, we have reason to believe that the positive alphas are not due to luck alone. Accordingly, if the percentiles of actual $t(\alpha)$ estimates are smaller than the average values of those from the simulation runs, then there are unskilled funds that underperform the passive benchmarks. If manager skill is zero, we expect the $t(\alpha)$ estimates on actual returns to be similar to the average of those from the simulations.

4.5 Results

4.5.1 Results on Tests of Change Points

The aforementioned change point test is performed on each mutual fund in the sample, and there are no more than three changes during the entire sample period in either slope or intercept, and no more than seven changes in parameters altogether. This is consistent with the argument of Berndt (1991), in which he states that econometrics studies based on monthly data such as the returns from the New York Stock Exchange are often found to have relatively stable βs from the 4F-CAPM over a five-year time span. The summary statistics of results on the tests for change points are shown in Table 4.1 and Table 4.2.

Table 4.1: Summary of Results on Tests for Change Points in the Slopes and the Intercept: 1984-2015 (4F-CAPM, $5 Million AUM)

	First test	Forward tests			Backward tests		
		1st	2nd	3rd	1st	2nd	3rd
Net returns							
Number of funds							
Change in intercept & slopes	288	43	2	0	12	3	0
Change in slopes	206	50	15	1	36	2	0
Change in intercept	322	132	31	7	47	11	2
No change	4190	4781	4958	4998	4911	4990	5004
Proportions							
Change in intercept & slopes	5.7	0.9	0.0	0.0	0.2	0.0	0.0
Change in slopes	4.1	1.0	0.3	0.0	0.7	0.0	0.0
Change in intercept	6.4	2.6	0.6	0.0	0.9	0.2	0.0
No change	83.8	95.5	99.1	100.0	98.2	99.8	100.0
Gross returns							
Number of funds							
Change in intercept & slopes	302	49	5	0	12	2	0
Change in slopes	202	45	13	1	35	3	0
Change in intercept	335	137	32	8	45	10	1
No change	4167	4775	4956	4997	4914	4991	5005
Proportions							
Change in intercept & slopes	6.0	1.0	0.0	0.0	0.2	0.0	0.0
Change in slopes	4.0	0.9	0.3	0.0	0.7	0.0	0.0
Change in intercept	6.7	2.7	0.6	0.0	0.9	0.2	0.0
No change	83.3	95.4	99.1	100.0	98.2	99.8	100.0

Table 4.2: Statistics on Different Cases of Combinations of Change Points: 1984-2015 (4F-CAPM, $5 Million AUM Group)

Different cases of combinations	Net returns		Gross returns	
	funds	Pct(%)	funds	Pct(%)
Total number of funds	5006			
No change	3993	79.8	3979	79.5
1 change in both βs and α	270	5.4	283	5.7
1 change in βs and no change in α	212	4.2	196	4.0
1 change in α and no change in βs	387	7.7	399	8.0
2 changes in βs and no change in α	16	0.3	16	0.3
2 changes in βs and 1 change in α	21	0.4	21	0.4
2 changes in βs and 2 changes in α	18	0.4	22	0.4
1 change in βs and 2 changes in α	43	0.9	44	0.9
0 change in βs and 2 changes in α	34	0.7	31	0.6
3 changes in βs and 0 change in α	1	0.0	1	0.0
3 changes in βs and 1 change in α	0	0.0	1	0.0
3 changes in βs and 2 changes in α	1	0.0	0	0.0
3 changes in βs and 3 change in α	1	0.0	2	0.0
2 changes in βs and 3 changes in α	5	0.1	3	0.0
1 change in βs and 3 changes in α	4	0.1	7	0.1
0 change in βs and 3 changes in α	0	0.0	0	0.0

As shown in Table 4.1, around 17% of the total funds in the sample have at least one parameter change. When we run the subsequent tests both forwards and backwards from the estimated date of the first detected change point, we find further changes in the parameters. For example, for net returns, around 4.5% of the total funds are found to have further parameter change when we run the test on the subsample prior to the estimated date of the first change, and around 2% of the total funds have further parameter change when the test in run on the subsample after the date of the first change. Finally, no more than three changes are found in either slopes or intercept as we continue to run the tests until no more changes are found. Table 4.2 summarizes the number of funds in different scenarios with different combinations of parameter changes. For instance, for net returns, there are 270 funds with only one change in the intercept and one change in the slope, and they account for 5.4% of all the funds in the sample. After all these changes have been detected, we construct the multi-regression model introduced in Section 4.3.2 to estimate fund alphas.

4.5.2 Results of the Simulations

4.5.2.1 Net Returns

As presented in Table 4.3, the left tail percentiles of actual $t(\alpha)$ estimates from net returns are much smaller than the average values from simulations at the same percentiles. For example, for the 3F-CAPM, the 5^{th} percentiles of the actual $t(\alpha)$ estimates are -3.01, -2.86, and -2.79 for the $5 million, $250 million, and $1 billion AUM group, respectively, which are far smaller than the corresponding average values from the simulations, which are -1.67, -1.68, and -1.69. If we look at the right tail percentiles, below the 90^{th} percentile, the actual $t(\alpha)$ estimates on net returns are still smaller than the average values from the simulations. However, from the 90^{th} percentile onwards, the actual $t(\alpha)$ estimates start to be greater than the average values from the simulations for all AUM groups.

Hence, there is evidence of manager skill above the 90^{th} percentile. For instance, the $t(\alpha)$ estimates for the 3F-CAPM and for the $5 million AUM group, are 2.23, 2.44, 2.66, 2.96, 3.42 at the 95^{th}, 96^{th}, 97^{th}, 98^{th}, 99^{th} percentile, which are much larger than the average values from the simulation runs (1.67, 1.78, 1.93, 2.12, 2.45 correspondingly). When the momentum factor is included in the model, the results are similar to those using the 3F-CAPM. The left tail percentiles produce much smaller $t(\alpha)$ estimates than those from simulations. The evidence of sufficient manager skill to cover costs above the 90^{th} percentile does not change for the $5 million AUM group, and is slightly weaker for the \$250 million and \$1 billion AUM group.

Table 4.3: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2015 (Model with Time-varying Parameters)**

The table shows values of $t(\alpha)$ at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for actual (Act) net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We bootstrap by resampling the fund and factor returns jointly as a pair. The sample covers the period from January 1984 to March 2015. Results are shown for the 3F-CAPM and the 4F-CAPM for the \$5 million, \$250 million, and \$1 billion AUM groups. There are 5006 funds in the \$5 million group, 2376 funds in the \$250 million group, and 1093 funds in the \$1 billion group.

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
3F-CAPM, Net Returns									
1	-4.04	-2.44	0.00	-4.04	-2.47	0.07	-3.81	-2.50	0.62
2	-3.62	-2.12	0.00	-3.56	-2.14	0.02	-3.38	-2.16	0.11
3	-3.35	-1.92	0.00	-3.31	-1.94	0.00	-3.16	-1.96	0.06
4	-3.13	-1.78	0.00	-3.03	-1.79	0.00	-3.01	-1.81	0.03
5	-3.01	-1.67	0.00	-2.86	-1.68	0.00	-2.79	-1.69	0.03
10	-2.47	-1.29	0.00	-2.34	-1.30	0.01	-2.24	-1.30	0.02
20	-1.82	-0.84	0.00	-1.68	-0.85	0.01	-1.63	-0.85	0.03
30	-1.34	-0.52	0.00	-1.18	-0.53	0.06	-1.17	-0.53	0.05
40	-0.93	-0.25	0.01	-0.82	-0.26	0.12	-0.78	-0.26	0.14
50	-0.56	0.00	0.01	-0.50	-0.01	0.25	-0.45	-0.01	0.24
60	-0.21	0.25	0.09	-0.17	0.24	0.66	-0.17	0.24	0.41
70	0.18	0.52	1.52	0.18	0.51	2.32	0.23	0.51	3.73
80	0.69	0.84	20.74	0.69	0.83	23.46	0.72	0.83	27.82
90	1.52	1.29	88.16	1.40	1.28	74.50	1.37	1.28	69.98
95	2.23	1.67	98.92	2.09	1.66	95.96	1.91	1.66	86.11
96	2.44	1.78	99.45	2.26	1.77	97.01	2.18	1.78	94.09
97	2.66	1.93	99.61	2.50	1.91	98.20	2.37	1.93	94.66
98	2.96	2.12	99.76	2.74	2.10	98.35	2.64	2.13	95.72
99	3.42	2.45	99.74	3.14	2.43	98.33	2.89	2.47	90.55

continued

Table 4.3 – continued

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
4F-CAPM, Net Returns									
1	-4.15	-2.51	0.00	-4.15	-2.57	0.30	-3.48	-2.64	4.32
2	-3.62	-2.16	0.00	-3.49	-2.18	0.06	-3.26	-2.23	0.81
3	-3.33	-1.95	0.00	-3.20	-1.97	0.04	-3.11	-2.01	0.18
4	-3.12	-1.80	0.00	-2.97	-1.81	0.03	-2.87	-1.85	0.11
5	-2.97	-1.68	0.00	-2.83	-1.69	0.01	-2.65	-1.73	0.15
10	-2.44	-1.29	0.00	-2.27	-1.30	0.01	-2.17	-1.32	0.05
20	-1.80	-0.84	0.00	-1.64	-0.84	0.01	-1.59	-0.85	0.04
30	-1.32	-0.52	0.00	-1.24	-0.52	0.00	-1.21	-0.53	0.02
40	-0.95	-0.25	0.00	-0.88	-0.25	0.00	-0.82	-0.26	0.12
50	-0.59	0.00	0.00	-0.58	0.00	0.00	-0.54	0.00	0.13
60	-0.26	0.25	0.00	-0.23	0.25	0.16	-0.20	0.25	0.23
70	0.12	0.52	0.31	0.15	0.52	1.09	0.16	0.52	1.32
80	0.60	0.85	6.92	0.61	0.84	9.10	0.70	0.85	22.34
90	1.38	1.30	67.38	1.34	1.30	61.09	1.32	1.31	54.67
95	2.08	1.70	95.32	1.93	1.69	85.47	1.88	1.71	78.07
96	2.29	1.82	97.36	2.14	1.81	91.51	1.99	1.84	75.72
97	2.51	1.97	98.36	2.37	1.96	94.35	2.22	1.99	82.37
98	2.79	2.18	98.75	2.54	2.17	91.90	2.35	2.21	72.22
99	3.31	2.54	99.01	2.88	2.56	87.32	2.80	2.61	75.02
3F-CAPM, Gross Returns									
1	-3.34	-2.43	0.30	-3.22	-2.47	1.91	-3.05	-2.51	6.71
2	-2.98	-2.11	0.13	-2.88	-2.14	0.90	-2.64	-2.16	5.42
3	-2.72	-1.92	0.12	-2.64	-1.94	0.82	-2.50	-1.95	3.13
4	-2.50	-1.78	0.18	-2.41	-1.79	1.23	-2.28	-1.80	3.85
5	-2.36	-1.66	0.18	-2.25	-1.68	1.52	-2.13	-1.69	4.29
10	-1.83	-1.29	0.41	-1.70	-1.29	3.58	-1.60	-1.30	8.28
20	-1.15	-0.84	4.28	-1.03	-0.84	16.03	-1.00	-0.85	20.09
30	-0.65	-0.52	20.44	-0.58	-0.53	37.33	-0.58	-0.53	36.29
40	-0.28	-0.25	41.99	-0.19	-0.26	63.82	-0.16	-0.26	70.84
50	0.09	0.00	71.37	0.14	-0.01	79.94	0.13	-0.01	80.03
60	0.44	0.25	88.02	0.46	0.25	89.83	0.43	0.24	86.58
70	0.84	0.52	96.98	0.82	0.51	94.82	0.77	0.51	92.70
80	1.35	0.84	99.56	1.29	0.83	98.51	1.29	0.83	98.75
90	2.16	1.29	99.99	2.05	1.28	99.93	1.93	1.28	99.60
95	2.80	1.67	100.00	2.72	1.66	99.98	2.54	1.66	99.91
96	2.96	1.78	100.00	2.87	1.77	99.97	2.73	1.78	99.93
97	3.14	1.93	100.00	3.00	1.91	99.95	2.87	1.92	99.85
98	3.43	2.12	100.00	3.22	2.11	99.93	3.14	2.12	99.77
99	3.95	2.45	99.99	3.74	2.43	99.91	3.41	2.46	98.62

continued

Table 4.3 – continued

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
4F-CAPM, Gross Returns									
1	-3.43	-2.52	0.34	-3.27	-2.58	3.31	-3.09	-2.65	12.62
2	-2.96	-2.16	0.26	-2.86	-2.19	1.60	-2.63	-2.23	10.20
3	-2.69	-1.95	0.27	-2.48	-1.97	3.28	-2.39	-2.01	8.79
4	-2.46	-1.80	0.52	-2.28	-1.82	3.93	-2.20	-1.85	9.01
5	-2.31	-1.69	0.57	-2.11	-1.69	4.74	-2.03	-1.72	10.64
10	-1.77	-1.30	1.39	-1.63	-1.30	6.27	-1.56	-1.32	12.74
20	-1.14	-0.84	4.77	-1.01	-0.84	17.53	-1.00	-0.85	20.89
30	-0.66	-0.52	19.98	-0.61	-0.52	29.69	-0.61	-0.53	31.25
40	-0.27	-0.25	45.36	-0.26	-0.25	46.54	-0.21	-0.25	60.26
50	0.07	0.00	66.75	0.08	0.00	68.67	0.11	0.00	74.71
60	0.41	0.25	83.38	0.44	0.25	86.18	0.41	0.25	81.85
70	0.77	0.52	92.87	0.80	0.52	93.35	0.75	0.52	89.13
80	1.29	0.84	98.89	1.25	0.84	97.65	1.23	0.85	97.16
90	2.04	1.30	99.93	2.01	1.30	99.78	1.92	1.31	99.34
95	2.70	1.70	99.98	2.60	1.70	99.94	2.51	1.72	99.73
96	2.87	1.82	99.99	2.76	1.82	99.95	2.59	1.84	99.41
97	3.07	1.97	99.99	2.93	1.97	99.93	2.65	2.00	98.18
98	3.35	2.18	99.99	3.14	2.18	99.85	2.97	2.22	98.46
99	3.89	2.54	99.94	3.62	2.56	99.60	3.26	2.62	94.40

4.5.2.2 Gross Returns

For net returns, we define a fund as skilled fund if it is able to produce positive excess returns net of all costs. The results show that there are a few fund managers that are able to more than cover the costs at the right tail, especially above the 90^{th} percentile. The percentiles in the extreme left tail of actual $t(\alpha)$ estimates are much smaller than the corresponding percentiles from the simulated $t(\alpha)$ estimates, which suggests that a large group of unskilled managers exists. However, it could be the fact that excess returns generated by managers are balanced out by the expenses, which managers may have little control of. To disentangle manager skill from funds' expense policies, we add back the costs reported in expense ratios so as to test more closely if fund managers have any skill.

Generally, as shown in Table 4.3, the percentiles produced using gross returns are larger than those using net returns. This is consistent with the argument of Carhart (1997) that expense ratios are significantly and negatively related to fund performance. It should be noted that the average values from simulations for gross and net returns are similar. This is not surprising since we set true α to zero in the simulation runs. The actual $t(\alpha)$ estimates on gross returns are still smaller than the average values from simulations for the percentiles below the 30^{th}. For example, for the 3F-CAPM, the 30^{th} percentile of the actual $t(\alpha)$ estimates are -0.65, -0.58, and -0.58 for the $5 million, $250 million, and $1 billion AUM group, which are smaller than the corresponding average values from simulations, -0.52, -0.53, and -0.53. Above the 30th percentile, there is evidence of widespread manager skill sufficient to cover costs missed in the expense ratio, since the average of $t(\alpha)$ estimates from simulations are much smaller than those on actual returns for all AUM groups. Furthermore, as shown in Table 4.3, results using the 4F-CAPM are similar to those using the 3F-CAPM.

4.5.2.3 Interpretation of Likelihoods

By comparing the percentiles of actual $t(\alpha)$ estimates on fund returns to those of the simulated averages, we can conclude that when returns are measured before all costs (net returns), there is strong evidence of unskilled managers whose actions result in negative alphas, and weak evidence of skilled managers located above the 90^{th} percentile. When returns are measured before expenses, the results are more promising. More fund managers are found to be able to cover the costs that are not included in the expense ratios, showing evidence of skill above the 30th percentile. Table 4.3 also shows the likelihoods, to be specific, the percentage of actual $t(\alpha)$ estimates on fund returns larger than the 10,000 simulation runs at the selected percentiles. These likelihoods offer a more intuitive way for us to judge whether percentiles of actual $t(\alpha)$ estimates on fund returns are extreme compared with the averages from simulations.

Following F&F's argument, if large fractions of the percentiles of $t(\alpha)$ estimates in the left tail in the 10,000 simulation runs are larger than those produced using actual fund returns, or low fractions of the percentiles of $t(\alpha)$ estimates in the right tail for the actual returns are larger than those in the simulation runs, we say that there are some unskilled managers that lack sufficient skill to cover costs. For the right tail percentiles, we infer that some managers are skilled, i.e. are able to generate expected returns that beat the benchmarks if large fractions of the percentiles of $t(\alpha)$ estimates on actual returns are larger than those from the simulation runs.

From the likelihoods on net returns, we confirm that few percentiles of $t(\alpha)$ estimates on actual returns are able to exceed those from simulation runs for the left tail; hence, a large group of fund managers present negative skill that cannot cover the costs. The likelihoods are extreme at the very right tail percentiles for all AUM groups, suggesting strong evidence of positive skill. For example, for the \$5 million AUM group, the 3F-CAPM $t(\alpha)$ estimates for actual returns are larger than those from simulations in more than 80% of the net return simulation runs above the 90^{th} percentile. For the 99^{th} percentile, 99.74%, 98.33% and 90.55% of the $t(\alpha)$ estimates from the 10,000 simulation runs are smaller than those from actual returns for the \$5 million, \$250 million, and \$1 billion AUM group. We also observe a slight drop of likelihoods when we switch to the 4F-CAPM.

When returns are measured before expenses, recall that the examination of manager skill on gross returns asks whether managers have sufficient skill to cover the part of costs missed in the expense ratios, mainly transaction costs. The likelihoods for tests on gross returns are more promising compared with those on net returns. For all AUM groups, the fractions of the 3F-CAPM $t(\alpha)$ estimates on actual returns larger than those from the simulation runs are more than 90% for percentiles that are above the 70^{th}. For example, at the 70^{th} percentile, the 3F-CAPM actual $t(\alpha)$ estimates for gross returns, for the \$5 million, \$250 million, and \$1 billion AUM group, are larger than those from simulation runs at 96.98%, 94.82%, 92.7% respectively. Furthermore, adding a momentum factor to the model only leads to trivial effects on the results.

Overall, the likelihoods suggest that most funds show sufficient skill to produce positive alphas when returns are measured before expenses; however, the positive excess returns are balanced out by the expenses, leaving only a small group of fund managers who have enough skill to cover all costs. Our results also indicate that there is a significant subgroup of unskilled managers, which makes it puzzling that investors continue to invest in those underperforming funds. As suggested in the literature, some investors are unsophisticated or constrained to unskilled funds, or disadvantaged to move to other funds due to taxes (Gruber, 1996; Christoffersen and Musto, 2002; Elton et al., 2006, 2007).

4.5.3 Time-varying vs. Time-invariant Parameters

By implementing the multi-regression model to address the time-variation in the parameters, we find evidence of skill in the net returns and strong evidence of a group of skilled managers when returns are measured before expenses. To show that capturing time-variation in the alpha and betas is important as it greatly affects the extreme right tail of the CDF of actual $t(\alpha)$ estimates, we compare the above results using the multi-regression model with those obtained using the traditional 4F-CAPM.

Table 4.4: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2015 (Time-varying vs. Time-invariant Parameters)**

This table This table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We bootstrap by resampling fund and factor returns jointly as a pair. The sample covers the period from January 1984 to March 2015. Results are shown for the 4F-CAPM with time-invariant parameter and the multi-regression model with time-varying parameters, for the $5 million AUM group. There are 5006 funds in the $5 million AUM group in our sample.

	Time-invariant parameters			Time-varying Parameters			Time-invariant parameters			Time-varying Parameters		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
Net Returns							Gross Returns					
1	-3.79	-2.46	0.01	-4.15	-2.51	0.00	-2.99	-2.47	3.40	-3.43	-2.52	0.34
2	-3.26	-2.12	0.01	-3.62	-2.16	0.00	-2.53	-2.13	5.80	-2.96	-2.16	0.26
3	-2.99	-1.93	0.01	-3.33	-1.95	0.00	-2.32	-1.93	5.84	-2.69	-1.95	0.27
4	-2.86	-1.78	0.01	-3.12	-1.80	0.00	-2.11	-1.79	8.59	-2.46	-1.80	0.52
5	-2.71	-1.67	0.01	-2.97	-1.68	0.00	-1.99	-1.68	8.46	-2.31	-1.69	0.57
10	-2.23	-1.29	0.01	-2.44	-1.29	0.00	-1.54	-1.29	11.80	-1.77	-1.30	1.39
20	-1.68	-0.84	0.02	-1.80	-0.84	0.00	-0.93	-0.85	31.68	-1.14	-0.84	4.77
30	-1.24	-0.52	0.02	-1.32	-0.52	0.00	-0.53	-0.53	48.10	-0.66	-0.52	19.98
40	-0.91	-0.25	0.03	-0.95	-0.25	0.00	-0.22	-0.26	57.56	-0.27	-0.25	45.36
50	-0.61	-0.00	0.03	-0.59	0.00	0.00	0.09	-0.01	69.44	0.07	0.00	66.75
60	-0.32	0.25	0.04	-0.26	0.25	0.00	0.38	0.25	77.57	0.41	0.25	83.38
70	0.01	0.52	0.09	0.12	0.52	0.31	0.71	0.52	84.39	0.77	0.52	92.87
80	0.40	0.84	0.40	0.60	0.85	6.92	1.10	0.84	90.30	1.29	0.84	98.89
90	1.00	1.29	6.57	1.38	1.30	67.38	1.73	1.29	97.14	2.04	1.30	99.93
95	1.49	1.68	19.20	2.08	1.70	95.32	2.26	1.68	98.61	2.70	1.70	99.98
96	1.62	1.80	22.14	2.29	1.82	97.36	2.40	1.79	98.76	2.87	1.82	99.99
97	1.85	1.94	37.98	2.51	1.97	98.36	2.63	1.94	99.10	3.07	1.97	99.99
98	2.12	2.14	50.18	2.79	2.18	98.75	2.89	2.14	99.36	3.35	2.18	99.99
99	2.56	2.49	63.18	3.31	2.54	99.01	3.40	2.49	99.67	3.89	2.54	99.94

4.5.3.1 Percentiles of $t(\alpha)$ Estimates

The percentiles of $t(\alpha)$ estimates on the actual returns and the average values of $t(\alpha)$ estimates from the simulations, for model with time-varying parameters (the multi-regression model) and that with time-invariant parameters (the 4F-CAPM) are shown in Table 4.4. The top left block shows the percentiles of $t(\alpha)$ estimates based on Equation 4.2 with time-invariant regression parameters, and on net returns. The percentiles of $t(\alpha)$ estimates using the multi-regression model with time-varying regression parameters described in Section 4.3.2 are shown in the top right block. Table 4.4 also presents the percentage of actual $t(\alpha)$ estimates on fund returns that are larger than those from the 10,000 simulation runs at selected percentiles. The $t(\alpha)$ estimated with time-invariant and time-varying parameters on gross returns are shown in the bottom right and left block. Since the results using the 4F-CAPM and the 3F-CAPM are similar as shown in Section 4.5.2, we only present results using the baseline model the 4F-CAPM here, and only for the $5 million AUM group. Percentiles of $t(\alpha)$ estimates for other AUM groups and using the 3F-CAPM, and the percentiles of $t(\alpha)$ estimates for different AUM groups based on the 4F-CAPM are shown in Table 4.12 in Appendix 4.B.

Net returns As shown in Table 4.4, when we allow for time-varying regression parameters on actual net returns, the dispersion of the CDF percentiles of actual $t(\alpha)$ estimates increases. The first and 99^{th} percentile of actual $t(\alpha)$ estimates for regressions with time-invariant parameters are -3.79 and 2.56, versus the corresponding actual $t(\alpha)$ estimates for regressions with time-varying parameters, -4.15 and 3.31.

It should be noted that for the left tail, the percentiles of actual $t(\alpha)$ estimates for regressions with time-varying parameters are not significantly different compared to those from regressions with time-invariant parameters. For example, the 10^{th}, 20^{th}, 30^{th}, 40^{th} and 50^{th} percentiles of actual $t(\alpha)$ estimates for regressions with time-invariant parameters are -2.23, -1.68, -1.24, -0.91, and -0.61, and they are not very different from the corresponding percentiles of $t(\alpha)$ estimates for regressions with time-varying parameters, which are -2.44, -1.80, -1.32, -0.95, and -0.59. This is not surprising given that bad performance persists no matter which model we use.

However, for the right tail, the percentiles of actual $t(\alpha)$ estimates for regressions with time-varying parameters do differ from those based on regressions with time-invariant parameters. For example, the 95^{th}, 96^{th}, 97^{th}, 98^{th}, and 99^{th} percentiles of actual $t(\alpha)$ estimates for regressions with time-invariant parameters are 1.49, 1.62, 1.85, 2.12, and 2.56, and they are far smaller than the corresponding percentiles of $t(\alpha)$ estimates for regressions with time-varying parameters, which are 2.08, 2.29, 2.51, 2.79, and 3.31.

If we look at the likelihoods, the fractions of actual $t(\alpha)$ estimates on net returns that are larger than those from the simulation runs for the right tail percentiles, increase dramatically when we switch from the traditional model to the multi-regression model with time-varying parameters. For example, the likelihoods of 95^{th}, 96^{th}, 97^{th}, 98^{th} and 99^{th} percentiles of actual $t(\alpha)$ estimates on net returns that are larger than those in the simulation runs based on regressions with time-varying parameters are 95.32%, 97.36%, 98.36%, 98.75%, and 99.01%, which are much larger than those using the traditional model with time-invariant parameters: 19.20%, 22.14%, 37.98%, 50.18%, and 63.18%. Therefore, we confirm that using the multi-regression model with time-varying parameters allows us to better detect the existence of superior and inferior performance of mutual fund managers, suggesting that is a group of skilled managers who can generate higher expected returns relative to the passive benchmarks, net of all costs.

Gross Returns As shown in Table 4.4, using gross returns, when we allow the regression slopes and intercept to be time-varying, the results are enhanced generally compared to the model with time-invariant parameters. First, the dispersion of the percentiles of CDF of actual $t(\alpha)$ estimates becomes larger when we account for time-varying parameters. The 1^{st} and 99^{th} percentile of actual $t(\alpha)$ estimates for regressions with time-varying parameters are -2.99 and 3.40, versus the corresponding $t(\alpha)$ estimates for regressions with time-variant parameters, -3.43 and 3.89.

Second, the right tail percentiles of actual $t(\alpha)$ estimates are more extreme than the average values from the simulations, which suggests the existence of skilled managers who are able to deliver higher expected returns relative to the passive benchmarks. Third, the likelihoods for the right tail percentiles are very close to 100%, which strongly supports the argument that there are skilled managers distributed in the right tail percentiles. Therefore, using the multi-regression model with time-varying parameters on gross returns, we are more confident in concluding that there are skilled managers located in the right tail percentiles who have enough skill to generate positive true α before expenses.

4.5.3.2 CDF Plots of $t(\alpha)$ Estimates

By implementing the multi-regression model with time-varying parameters, we find strong evidence of skill on net and gross returns. To provide a more intuitive understanding of the differences of results between using the 4F-CAPM and the multi-regression model, we plot empirical CDFs of the $t(\alpha)$ estimates on both actual returns and averaged values from simulations. Figure 4.1 and Figure 4.2 compare CDF plots of the cross-section $t(\alpha)$ estimates of the multi-regression model with those from the 4F-CAPM:

Specifically, Figure 4.1 shows the CDF plots for both the cross-sectional actual $t(\alpha)$

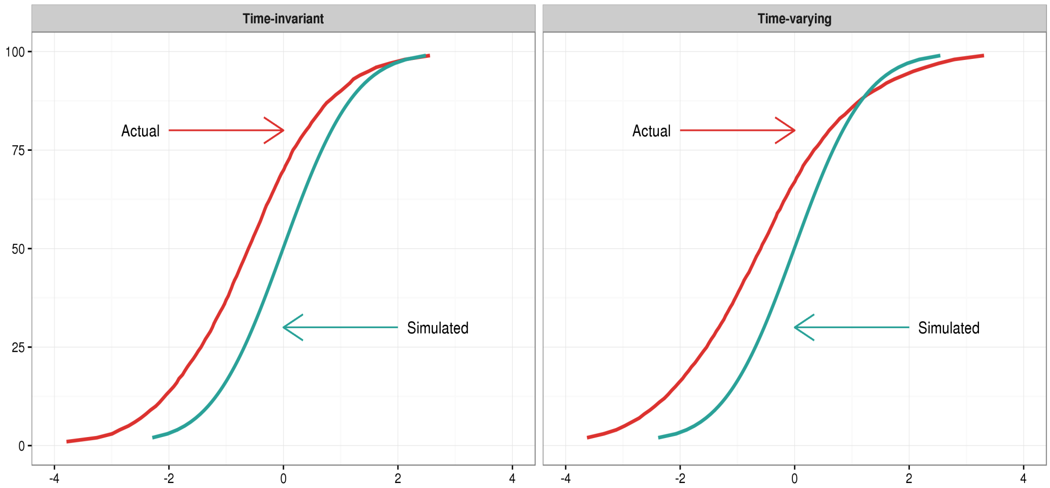

Figure 4.1: Simulated, Actual CDF Plots of $t(\alpha)$ Estimates for Net Returns, for the 4F-CAPM and the Multi-regression Model

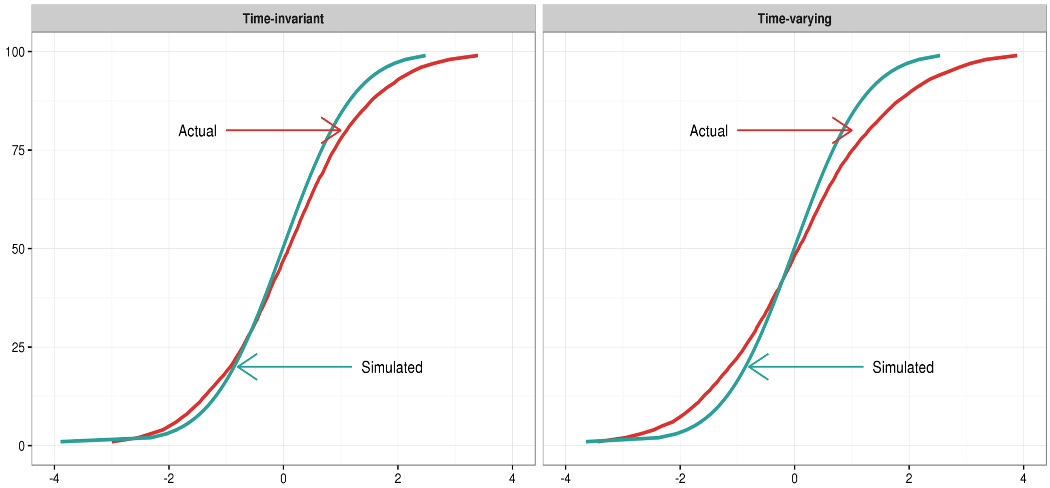

Figure 4.2: Simulated, Actual CDF Plots of $t(\alpha)$ Estimates for Gross Returns, for the 4F-CAPM and the Multi-regression Model

estimates on net returns and the averages from the simulations, for the 4F-CAPM and the multi-regression model. Figure 4.2 shows the comparison of CDFs on gross returns.

As shown in Figure 4.1, for net returns and for the 4F-CAPM, the distribution of actual cross-section $t(\alpha)$ estimates is almost completely to the left of the distribution of average $t(\alpha)$ estimates from simulations. Recall that the actual $t(\alpha)$ estimates represent the "real world" that has both skill and luck so that the distribution of $t(\alpha)$ estimates is in fact a mixed distribution of luck and skill. The simulated $t(\alpha)$ estimates capture the "simulated world" where only luck exist and skill is set to zero, hence the distribution of $t(\alpha)$ estimates is in fact a distribution of luck. The fact that the CDF of actual $t(\alpha)$ estimates is to the left of CDF of simulated $t(\alpha)$ estimates implies the evidence of no manager skill.

When the multi-regression model with time-varying parameters is used, the CDF of actual $t(\alpha)$ estimates shifts rightwards, with part of its right tail moving to the right of simulated $t(\alpha)$ estimates, from the 90th percentile. This implies that there are skilled managers in the extreme right tail percentiles. The CDF of simulated $t(\alpha)$ estimates only shifts a bit to the right, and this is not surprising given that in the simulations skill was set to zero so that no manager was skilled. We observe that the ends of the tails using the multi-regression model are fatter than those from the 4F-CAPM, indicating that more extreme alphas, negative as well as positive, are generated.

As to gross returns based on the 4F-CAPM with time-invariant parameters, as shown in Figure 4.2, the left tails of the CDF of actual $t(\alpha)$ estimates are to the left of the average from the simulations, which suggests there are unskilled managers whose actions result in negative true alphas relative to the benchmarks. The right tails of CDFs indicates that there are a group of skilled managers who enhance expected returns to beat the benchmarks. When time-variation of regression parameters is taken into account, the CDF of actual $t(\alpha)$ estimates shifts rightwards and the ends of the tails are fatter due to more extreme alphas. Therefore, after time-variation of regression parameters is allowed for in the multi-regression model, more skilled managers are found to have sufficient skill to generate positive true alphas relative to the passive benchmarks.

4.5.4 Robustness Checks

By resampling the entire monthly returns for all funds together, the simulations are able to capture the fact that returns might be correlated across funds. Jointly sampling fund and explanatory returns implicitly treats the regressors as random, thus the functional form of regression model fit to the data does not have to be correct. However, by resampling months in a simulation run, any effect of autocorrelation is lost. Therefore, we turn to block bootstrap to control for autocorrelation effect. Further details on different bootstrap methods are discussed in Chapter 6, Section 6.4.

To perform a block bootstrap, we divide the data series into several blocks, hence the original time series structure within a block can be preserved. The difficulty here is to determine the optimal length of blocks given that the optimal block size depends on the context in applications. Here we choose a block length equal to $T^{\frac{1}{5}}$ proposed by Hall et al. (1995). [3].

We sample with replacement from blocks of fund and factor returns and combine them to form the bootstrapped time series. Then we fit Equation 4.2 to this bootstrapped time series to get a cross-section of $t(\alpha)$ estimates that can be then ordered into a CDF. The entire process was repeated 10,000 times, so that we have the average values of $t(\alpha)$

[3]Details of how to implement the block bootstrap for our data are shown in Chapter 6, Section 6.4.2 and further details are documented in Efron and Tibshirani (1994) (cf. page 99-102).

estimates across the 10,000 simulation runs at the selected percentiles. It should be noted that we still set true α to zero in the simulation by subtracting the alpha estimates from funds' monthly returns. The results of block bootstrap are shown in Table 4.13 in Appendix 4.B. Generally, the results are not significantly different from those obtained from our baseline bootstrap described in Section 4.4. This is consistent with the results of KTWW and BSW, who find that estimates of statistics from block bootstrap are almost identical to those using the baseline bootstrap. Therefore, it seems safe to conclude that autocorrelation effect is only a minor issue in this study.

4.6 Fama and French (2010)

The closest precursor study to ours is F&F. They used a bootstrap technique that jointly resamples from monthly returns and factor returns for all funds simultaneously for the period of January, 1984 to September, 2006. Their results lead to a pessimistic assessment regarding the existence of manager skill. Specifically, by comparing the percentiles of actual $t(\alpha)$ estimates with those of simulation runs, they argue that few funds show sufficient skill to cover all costs and that there is evidence of skill, positive as well as negative, when returns are measured on gross returns. Specifically, using the 4F-CAPM, their results show that the $t(\alpha)$ estimates for simulations are larger than those on net returns of funds in the \$5 million AUM group for all percentiles. Their actual $t(\alpha)$ estimates for gross returns of funds are above those from the simulations runs only for the extreme right tail percentiles. However, they perform all their tests under the assumption that regression parameters are time-invariant and they indeed point out this limitation of their research as quoted (cf. F&F, p.1925): "Capturing time variation in the regression slopes poses thorny problems, and we leave this potentially important issue for future research."

Our study aims to address this time variation issue, the relevance of which is confirmed by our change point test results. We argue that prior results based on the traditional models are biased and the failure to account for time-variation in the regression parameters biases the inferences of F&F towards more pessimistic conclusions. To test this hypothesis, we estimate fund alphas using (i) the multi-regression model framework described in this study, and (ii) the 4F-CAPM used in F&F over their sample period January, 1984 to September, 2006. There are 3174 funds during this period in the $5 million AUM bound group. We perform the change point tests on this sample to obtain the estimated dates of the change points in the regression parameters. Results of the change point test are shown in Table 4.5 and Table 4.6.

4.6.1 Results of Change Point Tests

To summarize, around 17% of the funds in the sample have at least one parameter change. Based on the estimated dates of changes in the alpha and betas, we follow the methodology described in Section 4.3.2 to build the multi-regression model to estimate fund alphas and that described in Section 4.4 to perform simulations.

Table 4.5: Summary of Results on Tests for Change Points in the Slopes and the Intercept: 1984 to 2006 (4F-CAPM)

	First test	Forward tests			Backward tests		
		1st	2nd	3rd	1st	2nd	3rd
Number of funds							
Change in intercept & slopes	233	16	1	0	4	0	0
Change in slopes	130	25	0	0	22	1	0
Change in intercept	164	49	3	0	26	12	3
No change	2647	3084	3170	3174	3122	3161	3171
Proportions							
Change in intercept & slopes	7.3	0.5	0.0	0.0	0.1	0.0	0.0
Change in slopes	4.1	0.8	0.0	0.0	0.7	0.0	0.0
Change in intercept	5.2	1.5	0.1	0.0	0.8	0.4	0.1
No change	83.4	97.2	99.9	100.0	98.4	99.6	99.9

Table 4.6: Statistics on Different Cases of Combinations of Change Points: 1984 to 2006 (4F-CAPM)

Different cases of combinations	Number of funds	Percentage (%)
No change	2555	80.50
1 change in both βs and α	229	7.21
1 change in βs and no change in α	149	4.69
1 change in α and no change in βs	190	5.99
2 changes in βs and no change in α	4	0.01
2 changes in βs and 1 change in α	4	0.01
2 changes in βs and 2 changes in α	5	0.01
1 change in βs and 2 changes in α	23	0.07
0 change in βs and 2 changes in α	13	0.04
3 changes in βs and 0 change in α	0	0
3 changes in βs and 1 change in α	0	0
3 changes in βs and 2 changes in α	0	0
3 changes in βs and 3 change in α	1	0
2 changes in βs and 3 changes in α	1	0
1 change in βs and 3 changes in α	0	0
0 change in βs and 3 changes in α	0	0

4.6.2 Results of the Simulations

4.6.2.1 Net Returns

The upper block of Table 4.7 shows the results for net returns, for our multi-regression model and the 4F-CAPM. [4] The lower block shows results on gross returns. For regression results on net returns, when the alpha and betas are assumed to be constant, the left tail percentiles of actual $t(\alpha)$ estimates on net returns are much smaller than the corresponding average values from the simulations. Actual $t(\alpha)$ estimates on net returns becomes positive only after the 70^{th} percentile, and the actual $t(\alpha)$ estimates on net returns are always smaller than the average values from the simulations for all percentiles. This suggests that there is no evidence of widespread skill net of all costs, especially for the left tail percentiles. Furthermore, the likelihoods for net returns confirm the fact that most managers do not have sufficient skill to cover costs. Below the 90^{th} percentile, the actual $t(\alpha)$ estimates on net returns are larger than those from the simulation runs in less than 1.00% of the 10,000 simulation runs. All these results so far are consistent with those of F&F.

[4]Results on the other AUM groups, and based on multi-regression model and the 3F-CAPM are shown in Appendix 4.B.

Table 4.7: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2006 (Time-invariant vs. Time-varying Parameters)**

This table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We bootstrap by resampling fund and factor returns jointly as a pair. The sample covers the period from January 1984 to September 2006 and there are 3174 funds in the sample. Results are shown for the 4F-CAPM and the multi-regression model with change points.

	Time-invariant parameters			Time-varying parameters			Time-invariant parameters			Time-varying parameters		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
	Net Returns						Gross Returns					
1	-4.00	-2.44	0.00	-4.34	-2.44	0.00	-3.23	-2.44	0.96	-3.52	-2.44	0.17
2	-3.49	-2.11	0.01	-3.80	-2.11	0.00	-2.73	-2.10	1.71	-3.14	-2.10	0.13
3	-3.18	-1.91	0.01	-3.44	-1.91	0.00	-2.46	-1.91	2.43	-2.75	-1.91	0.35
4	-2.95	-1.77	0.01	-3.20	-1.77	0.00	-2.26	-1.76	3.17	-2.55	-1.76	0.46
5	-2.77	-1.66	0.03	-2.99	-1.66	0.00	-2.10	-1.65	4.29	-2.37	-1.65	0.60
10	-2.27	-1.28	0.01	-2.42	-1.28	0.01	-1.62	-1.27	6.50	-1.79	-1.27	1.77
20	-1.68	-0.83	0.01	-1.77	-0.83	0.01	-1.05	-0.82	13.76	-1.19	-0.82	4.49
30	-1.24	-0.51	0.02	-1.33	-0.51	0.01	-0.62	-0.50	26.30	-0.74	-0.50	12.58
40	-0.88	-0.24	0.06	-0.93	-0.24	0.02	-0.28	-0.23	40.21	-0.38	-0.23	22.51
50	-0.59	0.01	0.08	-0.60	0.01	0.06	0.00	0.02	46.03	-0.04	0.02	38.80
60	-0.31	0.26	0.08	-0.26	0.26	0.17	0.29	0.27	56.39	0.31	0.27	59.04
70	0.00	0.53	0.18	0.11	0.53	0.84	0.63	0.54	68.31	0.68	0.54	76.70
80	0.41	0.86	0.79	0.62	0.86	13.15	1.01	0.86	77.05	1.18	0.86	91.95
90	1.02	1.31	9.89	1.47	1.31	75.53	1.57	1.32	85.92	1.93	1.32	98.79
95	1.50	1.70	22.71	2.21	1.70	96.11	2.05	1.70	90.77	2.63	1.70	99.78
96	1.61	1.82	23.28	2.37	1.82	96.79	2.20	1.82	91.95	2.80	1.82	99.81
97	1.83	1.96	33.38	2.59	1.96	97.65	2.38	1.97	92.48	2.94	1.97	99.70
98	2.10	2.16	43.92	2.82	2.16	97.58	2.62	2.17	93.38	3.30	2.17	99.85
99	2.45	2.51	44.75	3.40	2.51	98.83	2.99	2.52	92.45	3.82	2.52	99.89

As shown in Table 4.7, when we allow for time-varying parameters, the dispersion of the CDF of actual $t(\alpha)$ estimates increases. The first and 99^{th} percentile of actual $t(\alpha)$ estimates for regressions with time-invariant parameters are -4.00 and 2.45, versus the corresponding actual $t(\alpha)$ estimates for regressions with time-varying parameters, -4.34 and 3.40. For the right tail percentiles, the actual $t(\alpha)$ estimates for regressions with time-varying parameters do differ from those based on regressions with time-invariant parameters. For example, the 95^{th}, 96^{th}, 97^{th}, 98^{th}, and 99^{th} percentiles of actual $t(\alpha)$ estimates for regressions with time-invariant parameters, 1.50, 1.61, 1.83, 2.10, and 2.45, are far smaller than the corresponding percentiles of $t(\alpha)$ estimates for regressions with time-varying parameters, 2.21, 2.37, 2.59, 2.82, and 3.40.

If we look at the likelihoods, the fractions of $t(\alpha)$ estimates on actual net returns are larger than those from the simulation runs for the right tail percentiles, which increase dramatically when we switch from regressions with time-invariant parameters to time-varying parameters. For example, the likelihoods of 95^{th}, 96^{th}, 97^{th}, 98^{th} and 99^{th} percentiles of actual $t(\alpha)$ estimates on net returns based on regressions with time-varying parameters are larger than those in the simulation runs, 96.11%, 96.79%, 97.65%, 97.58%, and 98.83%, which are much greater than those with time-invariant parameters, 22.71%, 23.28%, 33.38%, 43.92%, and 44.75%. This confirms our argument that allowing for time-variation in the regression parameters produces more positive alpha estimates net of all costs compared to those using the traditional model with time-invariant parameters.

4.6.2.2 Gross Returns

As for the gross returns, as shown in Table 4.7, when regression slopes and intercept are assumed to be constant, the dispersion of the CDF of actual $t(\alpha)$ estimates on gross returns decreases compared to that on net returns. The left tail percentiles of actual $t(\alpha)$ estimates on gross returns are still to the left of the corresponding average values from the simulations, which suggests that there are unskilled managers who generate negative true α relative to the benchmarks. In contrast, the right tail percentiles suggest that there are skilled managers whose actions lead to positive true α relative to the benchmarks given that the right tail percentiles of actual $t(\alpha)$ estimates on gross returns are greater than those average values from simulations. Therefore, in terms of gross returns, there are a few skilled managers with sufficient skill to more than just cover costs missed in expense ratios, which is also consistent with F&F.

When we allow the regression parameters to be time-varying, the results become even more promising generally. First, the dispersion of the CDF of actual $t(\alpha)$ estimates becomes larger when we account for time-varying parameters. The fist and 99^{th} percentile of actual $t(\alpha)$ estimates for regressions with time-varying parameters are -3.52 and 3.82, versus the corresponding $t(\alpha)$ estimates for regressions with time-invariant parameters, -3.23 and 2.99.

If we turn to the likelihoods, for the right tail percentiles, likelihoods are very close to 100%, which strongly support the argument that there are outperforming managers distributed in the right tail percentiles. For example, the 97^{th}, 98^{th} and 99^{th} percentile of actual $t(\alpha)$ estimates, 2.94, 3.30, and 3.82, all exceed in more than 99% of the corresponding $t(\alpha)$ estimates from the 10,000 simulation runs. Therefore, using the model with time-varying parameters on gross returns, we are more confident in concluding that there are skilled managers distributed in the right tail percentiles who indeed have enough skill to generate positive true α. Generally, after accounting for time-variation in the alpha and betas, we are able to challenge several of F&F's results. Importantly, we argue that there is strong evidence of manager skill for both net and gross returns, whereas F&F only confirm the weak evidence of skill for gross returns.

4.7 Conclusions

Whether there is a group of skilled and unskilled fund managers is an interesting research topic that has been repeatedly examined over the past 50 years. Given the existing controversies in the literature, we revisited this issue using monthly return data of a rich sample of 5006 mutual funds obtained from CRSP, and estimated the alpha of individual funds to infer the existence of skilled and unskilled fund managers. In order to distinguish skill from luck, we follow F&F's approach by constructing two sets of world using a bootstrap technique. Furthermore, to capture time-variation in the regression parameters in the baseline models, which was considered as a "thorny" problem in F&F, we conducted a non-parametric test to detect and to estimate the date of changes in both the alpha and betas for each fund in the sample.

Our results show that around 17% of the total funds in the sample have at least one parameter change. We then build a multi-regression model, which estimates fund alphas using separate regressions on the subsamples of return series divided according to the estimated dates of the changes. Generally, after allowing for time-varying parameters in the regressions, the cross-sectional distribution of $t(\alpha)$ estimates shifts rightwards for both net returns and gross returns, which suggests more evidence of manager skill than previously documented in the literature.

Specifically, for net returns, there are skilled fund managers located in the right tail above the 90^{th} percentile. For gross returns, the percentiles of actual $t(\alpha)$ estimates are larger than those from the simulations in more than 70% of simulation runs for the whole right tail, especially for those above the 90th, which shows that the percentiles of actual $t(\alpha)$ estimates are larger than those average simulated values in more than 99% of 10,000 simulation runs. Using the multi-regression model with time-varying parameters, we are able to challenge several of F&F's results, showing in particular that the pessimistic results of F&F are mainly due to the fact that they assume the parameters in the regression model are time-invariant. Hence, after correcting for this, we confirm that there is a group of U.S. actively managed mutual fund managers who outperform those passively managed funds.

There are two main reasons why we think the multi-regression model outperforms the existing models: first, funds may change managers throughout its entire history, after allowing for discrete changes in fund alphas, the part of history when a fund performs relatively well is separated from the period when it performs badly, so that outperformance of managers will not be balanced out by the underperformance of managers prior to or after the replacement of managers; second, investors' capital inflows to mutual funds can be competitive so that time-varying regression parameters can capture the outperformance of managers over the short run before investors compete away their performance with capital inflows. By treating each subsample as a separate "fund" record, our multi-regression model works well in the simulations, and hence has wide applications in finance issues where both luck and time-variation need to be controlled for.

Appendix 4.A Details on the Construction and Derivation of the Test for Change Points

We use the test for change points proposed by Pouliot (2016). To construct the test, first, he assumes a random sample $\{(Y_t, \mathbf{X}_t)\}_{t=1}^{T}$ that satisfies the equation below with an unknown break in parameters at time t^*:

$$Y_t = \begin{cases} \alpha^{(1)} + \mathbf{X}_t\boldsymbol{\beta}^{(\mathbf{1})} + \sigma\varepsilon_t, & 1 \leq t \leq t^*, \\ \alpha^{(2)} + \mathbf{X}_t\boldsymbol{\beta}^{(2)} + \sigma\varepsilon_t, & t^* < t \leq T, \end{cases} \tag{4.6}$$

where the ε_ts should be independent and identically distributed (iid) random variables which satisfy the following moment conditions:

$$\varepsilon_t = 0, \varepsilon_t^2 = 1 \text{ and } |\varepsilon_t|^4 < \infty,\ t = 1, \ldots, T. \tag{4.7}$$

It should be noted that equation (4.6) can be expressed in a way that the terms $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ are $K \times 1$ parameter vectors, $\mathbf{X}_t$ is a $1 \times K$ vector of explanatory variables. He assumes that all components of $\mathbf{X}_t$ and the dependent variable Y_t are stationary. However, non-stationary Y_t and $\mathbf{X}_t$ are also allowed as long as the non-stationary variables in (4.6) are co-integrated. This assumption is mainly satisfied in order to assure the weak convergence of results.

In Specific, to detect parameter instability, the null hypothesis is:

$$H_0 : t^* \geq T$$

versus the alternative hypothesis of having at-most-one change in regression slopes or intercept:

$$H_1 : 1 \leq t^* < T.$$

Under the alternative hypothesis of a break in either intercept or slope parameters, he assumes that at least $\alpha^{(1)} \neq \alpha^{(2)}$ or $\boldsymbol{\beta}^{(1)} \neq \boldsymbol{\beta}^{(2)}$.

The discussion of rest of the test is divided into two scenarios: i) when parameters in (4.6) are known and ii) when parameters are unknown. Apparently, the latter should be the more realistic setting, however, given that the asymptotic distributions of their test statistics are easier to establish under assumption i), we mainly show the derivations of scenario i) here for brevity. It can be proved that the statistics derived under assumption i) are asymptotically equivalent to the those derived under assumption ii).

4.A.1 Parameters Known

For the main test, it builds a process originated from a U-statistic (Csörgő and Horváth (1988)). He defines process $M_T(\tau)$ to be a function of τ where $\tau \in (0,1)$, such that

$$M_T(\tau) \;:=\; T^{-1/2}\left\{\sum_{t=1}^{(T+1)\tau}(Y_t - \alpha^{(1)} - \mathbf{X}_t\boldsymbol{\beta}^{(1)})^2 - \tau\sum_{t=1}^{T}(Y_t - \alpha^{(1)} - \mathbf{X}_t\boldsymbol{\beta}^{(1)})^2\right\}. \quad (4.8)$$

The above process can be interpreted as comparing the variance before a change in parameters to the variance after a change in parameters. The interest here is how large this process is across the interval of τ, and to suggest that the variance before is significantly different from the variance after. A suitable functional to achieve that is the supreme, which results in the test statistic:

$$\sup_{0<\tau<1} |M_T(\tau)|. \tag{4.9}$$

A useful by-product of this statistic is that it indirectly produces an estimate of the break point τ^* as well. Gombay (1996) prove the convergence to the supreme of a Brownian bridge of this statistic. However, this statistic is not able to distinguish between rejections in the intercept and the slopes. Therefore, he constructs two auxiliary U-statistic type processes to tackle this issue:

$$M_T^{(A)}(\tau) \;:=\; T^{-1/2}\left\{\sum_{t=1}^{(T+1)\tau}(Y_t-\alpha-\mathbf{X}_t\boldsymbol{\beta}^{(1)})^2-\tau\sum_{t=1}^{T}(Y_t-\alpha-\mathbf{X}_t\boldsymbol{\beta}^{(1)})^2\right\} \tag{4.10}$$

$$M_T^{(B)}(\tau) \;:=\; T^{-1/2}\left\{\sum_{t=1}^{(T+1)\tau}(Y_t-\mathbf{X}_t\boldsymbol{\beta})-\tau\sum_{t=1}^{T}(Y_t-\mathbf{X}_t\boldsymbol{\beta})\right\}, \tag{4.11}$$

where in $M_T^{(A)}(\tau)$ set

$$\alpha = \begin{cases} \alpha^{(1)}, & t \le t^* \\ \alpha^{(2)}, & t > t^* \end{cases}$$

and in $M_T^{(B)}(\tau)$ set

$$\boldsymbol{\beta} = \begin{cases} \boldsymbol{\beta}^{(1)}, & t \le t^* \\ \boldsymbol{\beta}^{(2)}, & t > t^*. \end{cases}$$

where the first statistic is sensitive to a one-time change in $\boldsymbol{\beta}^{(1)}$ while remains robust to any possible changes in the intercept. The second is sensitive to a one-time change in the intercept while remains robust to possible changes in slopes. To enhance the power of this test so that a break at any time can be detected, the following test statistics are built:

$$\sup_{0<\tau<1} \frac{|M_T^{(i)}(\tau)|}{q(\tau)}, \tag{4.12}$$

for $i = A,\ B,$, where $q(\tau)$ is a weight function designed to improve the test for detecting any changes in parameters that occur within the sample. While other similar tests for change points only have power in a compact range within $(0,1)$, this test can be used to detect changes over the whole $(0,1)$ range, in particular, those changes close to the beginning or end of the sample period. In order to achieve that, $q(\tau)$ should satisfy the following two assumptions:

A.1: The function $q(\cdot)$ defined on (0,1) is such that $\inf_{\delta\le\tau\le 1-\delta} q(\tau) > 0$ for all $\tau \in (0,1)$ and $\delta \in (0, 1/2)$.

A.2: $I(q,c) = \int_0^1 \frac{1}{\tau(1-\tau)} \exp^{-\frac{cq^2(\tau)}{(\tau(1-\tau))}} d\tau < \infty$ for some constant $c > 0$.

And the weight function is given by:

$$q(\tau) = q(\tau;\nu) := \{(\tau(1-\tau))^{\nu};\ 0 \le \nu < 1/2\}. \tag{4.13}$$

This kind of functions satisfies A.1 and A.2 for all $c > 0$, and can be sensitive to a change in parameters of linear regression models that occurs near the beginning or end of the sample period. The derivation of the asymptotic distribution of these processes in (4.12) are shown below: under H_0, if the process (4.6) satisfies the conditions detailed in (4.6) and $q(\tau)$ satisfies A.1 and A.2, then as $T \to \infty$, the following two statements can be proved:

(i) If in A.2 the integral holds for all $c > 0$ rather than for some $c > 0$, a sequence of Brownian bridges $\{\mathcal{B}_T(\tau)\}$ can be constructed such that the following result holds:

$$\sup_{0<\tau<1} \frac{\left|\frac{1}{\Delta^{(i)}} M_T^{(i)}(\tau) - \mathcal{B}_T(\tau)\right|}{q(\tau)} = o_P(1);$$

(ii) and if in A.2 the integral holds for some $c > 0$ rather than for all $c > 0$, then a sequence of Brownian bridges $\{\mathcal{B}_T(\tau)\}$ can be constructed such that the following result holds:

$$\sup_{0<\tau<1} \frac{|\frac{1}{\Delta^{(i)}} M_T^{(i)}(\tau)|}{q(\tau)} \xrightarrow{D} \sup_{0<\tau<1} \frac{|B(\tau)|}{q(\tau)},$$

where $\Delta^{(A)} = \sigma^2\sqrt{Var(\varepsilon_1^2)}$, $\Delta^{(B)} = \sigma$ and $\mathcal{B}_T(t) := \frac{W(T\tau)}{\sqrt{T}} - \tau\frac{W(T)}{\sqrt{T}}$. The proof of this can be found in OP, page 7.

Interest here centers on a bivariate process derived from the two processes given in (4.10) and (4.11). Under H_0, if the process (4.6) satisfies conditions detailed in (4.7) and $q(\tau)$ satisfies A.2 for all $c > 0$, then as $T \to \infty$, it can be shown that:

$$\begin{bmatrix} \sup\limits_{0<\tau<1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} \\ \sup\limits_{0<\tau<1} \frac{1}{\Delta^{(B)}} \frac{|M_T^{(B)}(\tau)|}{q(\tau)} \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \sup\limits_{0<\tau<1} \frac{|\mathcal{B}^{(A)}(\tau)|}{q(\tau)} \\ \sup\limits_{0<\tau<1} \frac{|\rho\mathcal{B}^{(A)}(\tau) + (1-\rho^2)^{1/2}\mathcal{B}^{(B)}(\tau)|}{q(\tau)} \end{bmatrix}, \tag{4.14}$$

with $\xrightarrow{D}$ representing convergence in distribution and $\mathcal{B}^{(A)}(\tau)$ and $\mathcal{B}^{(B)}(\tau)$ denoting two independent copies of Brownian bridges and $\rho = \frac{\mathrm{E}[\varepsilon_1^3]}{\sqrt{Var(\varepsilon_1^2)}}$. Detailed proof can be found in Pouliot (2016).

These distributions are based on unknown parameters, i.e. the variance, skewness and kurtosis of the error terms. Given the symmetry of the distribution of the residuals of the regression model, when $\rho = 0$,[5] the bivariate process in (4.14) converges in distribution to two independent copies of a weighted Brownian bridge. In this case the test for parameter changes in either intercept or slopes is equivalent to comparing the test statistics in (4.10) and (4.11) with the critical values at a γ significance level b_γ, which can be derived from the corresponding tabulated asymptotic distribution. For example, if the following condition is satisfied:

$$\frac{1}{\Delta^{(A)}} \sup_{0<\tau<1} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} > b_\gamma > \sup_{0<\tau<1} \frac{1}{\Delta^{(B)}} \frac{|M_T^{(B)}(\tau)|}{q(\tau)},$$

then only the slope parameter has changed. If, however, the condition:

$$\frac{1}{\Delta^{(A)}} \sup_{0<\tau<1} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} > b_\gamma \text{ and } \frac{1}{\Delta^{(B)}} \frac{|M_T^{(B)}(\tau)|}{q(\tau)} > b_\gamma$$

is satisfied, then both intercept and slope have changes. Finally, if the critical value, at a γ significance level, is greater than both statistics, then null hypothesis of no change points in the model can be rejected.

This test also considers the case when the residuals are asymmetrically distributed. In this case, given the dependence of the two processes, it is difficult to construct the relevant asymptotic critical values. Therefore, the test reformulates the statistics such that, under the same assumptions illustrated before, as $T \to \infty$,

$$\begin{bmatrix} \sup\limits_{0<\tau<1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} \\ \sup\limits_{0<\tau<1} \frac{|\mathcal{M}(\tau)|}{q(\tau)} \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \sup\limits_{0<\tau<1} \frac{|\mathcal{B}^{(A)}(\tau)|}{q(\tau)} \\ \sup\limits_{0<\tau<1} \frac{|\mathcal{B}^{(B)}(\tau)|}{q(\tau)} \end{bmatrix} \tag{4.15}$$

where $\mathcal{M}(\tau) = -\rho((1-\rho^2)\sigma^4 Var(\varepsilon_1^2))^{-\frac{1}{2}} M_T^{(A)}(\tau) + ((1-\rho^2)\sigma^2)^{-\frac{1}{2}} M_T^{(B)}(\tau)$. This bivariate process and its corresponding asymptotic theory allow this test to build two different test

[5] $\rho = 0$ if and only if the distribution of the residuals in our regression (**??**) is symmetrically distributed around zero.

statistics under the null hypothesis of no changes in parameters of the model given by equation (**??**). The statistic in the first row of this bivariate process detailed in (4.15) remains fixed given a change in α, *i.e.* it remains robust to a change in the intercept. The second test statistic, however, will be sensitive to both changes in slopes and intercept, so that a joint hypothesis can be developed:

$$H_0 : \alpha^{(1)} = \alpha^{(2)} \text{ and } \boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}. \tag{4.16}$$

The alternative hypothesis in our joint test implies a change in at least one parameter, either α or $\boldsymbol{\beta}$. A value of the test statistic greater than b_γ implies the rejection of H_0. This test can be considered as a simultaneous test because it runs simultaneously on the two test statistics in equation (4.15). The statistic in the first row of this vector tests the null hypothesis $H_{0,\text{slope}} : \boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(2)}$, while the test statistic in the second row tests the null hypothesis H_0. Therefore, this simultaneous test is proved to be more informative than those traditional tests for changes in parameters since it gives information on which parameter has changed: whether α or $\boldsymbol{\beta}$. As a test for changes in the parameters, it has one drawback. If there is a change in $\boldsymbol{\beta}$, it is no longer clear on a possible change in α. However, this does not weaken the power of this test in our context since we consider a change in α also happened in this case in our test. Thus, if in fact there is no change in α, and we split the sample into two subsamples and run regressions separately, it does not hurt since these two regressions on two subsamples will just give us identical α estimates.

Let us explore further possible outcomes of this test. It is possible, though unlikely, that

$$\sup_{0<\tau<1} \frac{|\mathcal{M}(\tau)|}{q(\tau)} > b_\gamma > \sup_{0<\tau<1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)}. \tag{4.17}$$

In this case, we can conclude that there is one change in slope only and H_0 is rejected. If, however, the more likely case happens:

$$\begin{aligned} b_\gamma &< \sup_{0<\tau<1} \frac{|\mathcal{M}(\tau)|}{q(\tau)} \\ b_\gamma &> \sup_{0<\tau<1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)} \end{aligned}$$

then we reject H_0 and conclude that there has been a change in α only. If, on the other hand, we have:

$$\begin{aligned} b_\gamma &< \sup_{0<\tau<1} \frac{|\mathcal{M}(\tau)|}{q(\tau)} \\ b_\gamma &< \sup_{0<\tau<1} \frac{1}{\Delta^{(A)}} \frac{|M_T^{(A)}(\tau)|}{q(\tau)}. \end{aligned}$$

then we can conclude that only a change in $\boldsymbol{\beta}$ has occurred. Here, the test is not clear about a change in α. It is necessary to run an auxiliary test based on the statistic $M_T^{(B)}(\tau)/q(\tau)$ to determine whether α has changed as well.

4.A.2 Parameters Unknown

The processes defined in (4.10) and (4.11) are based on unknown parameters. Ordinary Least Squares (OLS) will produce consistent estimators of α_i and $\boldsymbol{\beta}_i$ for $i = 1,\ 2$ under H_0 and H_1. Let these sequences of estimators be denoted $\{\widehat{\alpha}_T^{(i)}\}_{T=1}^{\infty}$ and $\{\widehat{\boldsymbol{\beta}}_T^{(i)}\}_{T=1}^{\infty}$ for $i = 1,\ 2$. When these sample estimates are replaced by population parameters, the following slightly altered sequence of partial sum processes should occur:

$$\widehat{M}_T^{(A)}(\tau) \;:=\; T^{-1/2}\left\{\sum_{t=1}^{(T+1)\tau}(Y_t-\widehat{\alpha}_T-\mathbf{X}_t\widehat{\boldsymbol{\beta}}_T)^2-\tau\sum_{t=1}^{T}(Y_t-\widehat{\alpha}_T-\mathbf{X}_t)^2\widehat{\boldsymbol{\beta}}_T\right\} \quad (4.18)$$

$$\widehat{M}_T^{(B)}(\tau) \;:=\; T^{-1/2}\left\{\sum_{t=1}^{(T+1)\tau}(Y_t-\mathbf{X}_t\widehat{\boldsymbol{\beta}}_T)-\tau\sum_{t=1}^{T}(Y_t-\mathbf{X}_t\widehat{\boldsymbol{\beta}}_T)\right\}. \quad (4.19)$$

In $\widehat{M}_T^{(A)}(\tau)$ set

$$\widehat{\alpha}_T = \begin{cases} \widehat{\alpha}_T^{(1)}, & t \le \widehat{t}^* \\ \widehat{\alpha}_T^{(2)}, & t > \widehat{t}^*, \end{cases}$$

and $\widehat{\boldsymbol{\beta}}_T = \widehat{\boldsymbol{\beta}}_T^{(1)}$ and in process $\widehat{M}_T^{(B)}(\tau)$ set

$$\widehat{\boldsymbol{\beta}}_T = \begin{cases} \widehat{\boldsymbol{\beta}}_T^{(1)}, & t \le \widehat{t}^* \\ \widehat{\boldsymbol{\beta}}_T^{(2)}, & t > \widehat{t}^*, \end{cases}$$

where $\widehat{t}^*$ is some consistent estimator of t^*. One consistent estimator of t^* that has been widely used in the literature is defined as follows:

$$\widehat{t}^* \;:=\; \frac{1}{T}\min\left\{k : \frac{|\widehat{M}_T(\frac{k}{T})|}{q(\frac{k}{T})} = \max_{1\le i<T}\frac{|\widehat{M}_T(\frac{i}{T})|}{q(\frac{i}{T})}\right\} \quad (4.20)$$

$$\widehat{M}_T(t) \;=\; T^{-1/2}\left\{\sum_{t=1}^{(T+1)\tau}(Y_t-\widehat{\alpha}_{LS}-\widehat{\boldsymbol{\beta}}_{LS}'\mathbf{X}_t)^2-\tau\sum_{t=1}^{T}(Y_t-\widehat{\alpha}_{LS}-\widehat{\boldsymbol{\beta}}_{LS}'\mathbf{X}_t)^2\right\},$$

where the subscript LS denotes the least squares estimator of α and $\boldsymbol{\beta}$ using all T observations. The asymptotic properties of this estimator have been proved by Antoch et al. (1995). They also prove the asymptotic properties of bootstrap approximation to this distribution.

Pouliot (2016) also proves the asymptotic equivalence between test statistics based on $M_T^{(i)}(\tau)$ and ones based on $\widehat{M}_T^{(i)}(\tau)$, for $i = A, B$. He proves that all properties built regarding the processes $M_T^{(i)}(\tau)$ still hold for $\widehat{M}_T^{(i)}(\tau)$, for $i = A, B$, by replacing estimators with the population parameters of the linear regression model. They also hold when parameters ρ, σ and $Var(\varepsilon_1^2)$ are replaced by any sequence of consistent estimators. He proves that if $\{\widehat{\alpha}_T\}_{T=1}^{\infty}$ and $\{\widehat{\boldsymbol{\beta}}_T\}_{T=1}^{\infty}$ are assumed to be sequences of consistent estimators of the parameters in (4.6), then under the same conditions as before, the following can be proved:

$$\sup_{0<\tau<1} \frac{|M_T^{(i)}(\tau) \; - \; \widehat{M}_T^{(i)}(\tau)|}{q(\tau)} \; = \; o_P(1),$$

for $i = A, B$, as $T \to \infty$. Details of proof can be found in Olmo et al. (2014), Lemma 2.1 page 11.

Furthermore, Pouliot (2016) discusses the asymptotic power of the test by Corollaries 3.1 and 3.2, which prove that tests based on these statistics are asymptotically consistent.These statistics will detect changes in coefficients of the 4F-CAPM when at least one of the parameters changes with a probability approaching one.

Appendix 4.B Additional Results on Change Point Tests and Simulations

This appendix shows the summary of results on change points, specifically Table 4.8 shows the summary of change points tests based on the 3F-CAPM for different AUM groups from 1984 to 2015; Table 4.11 shows the summary of change points tests based on the 4F-CAPM for different AUM groups from 1984 to 2015; Table 4.10 shows the different combinations of change points scenarios based on the 3F-CAPM for different AUM groups from 1984 to 2015; Table 4.11 shows the different combinations of change points scenarios based on 4F-CAPM for different AUM groups from 1984 to 2015.

This appendix also shows the simulation results, specifically, Table 4.12 shows the percentiles of the actual and simulated $t(\alpha)$ estimates for different AUM groups, using the 3F-CAPM and the 4F-CAPM with time-invariant parameters and a dataset from 1984 to 2015; Table 4.13 shows the percentiles of the actual and simulated $t(\alpha)$ estimates using the multi-regression model based on the 3F-CAPM and the 4F-CAPM, using a block bootstrap method and a dataset from 1984 to 2015; Table 4.14 shows the aggregate performance of mutual funds based on a dataset from 1984 to 2006; Table 4.15 shows the percentiles of the actual and simulated $t(\alpha)$ estimates for different AUM groups, using the 3F-CAPM and the 4F-CAPM with time-invariant parameters, using a dataset from 1984 to 2006; Table 4.16 shows the percentiles of the actual and simulated $t(\alpha)$ estimates for different AUM groups, using the 3F-CAPM and the 4F-CAPM with time-invariant parameters, using a block bootstrap and a dataset from 1984 to 2006.

Table 4.8: Summary of Results on Tests for Change Points in the Slopes and the Intercept: 1984-2015 (3F-CAPM)

$5 million AUM	First test	Forward tests			Backward tests		
		1st	2nd	3rd	1st	2nd	3rd
Net returns							
Number of funds							
Change in intercept & slopes	364	53	5	1	6	0	0
Change in slopes	222	56	9	0	29	3	1
Change in intercept	336	117	17	3	34	6	3
No change	4084	4780	4975	5002	4937	4997	5002
Proportions							
Change in intercept & slopes	7.3	1.1	0.1	0.0	0.1	0.0	0.0
Change in slopes	4.4	1.1	0.2	0.0	0.6	0.1	0.0
Change in intercept	6.7	2.3	0.3	0.0	0.7	0.1	0.0
No change	81.6	95.5	99.4	100.0	98.6	99.8	100.0
Gross returns							
Number of funds							
Change in intercept & slopes	370	52	3	0	8	0	0
Change in slopes	232	57	11	0	33	5	1
Change in intercept	339	114	19	2	38	6	3
No change	4065	4783	4973	5004	4927	4995	5002
Proportions							
Change in intercept & slopes	7.4	1.0	0.1	0.0	0.2	0.0	0.0
Change in slopes	4.6	1.1	0.2	0.0	0.7	0.1	0.0
Change in intercept	6.8	2.3	0.4	0.0	0.8	0.1	0.0
No change	81.2	95.6	99.3	100.0	98.3	99.8	100.0

continued

Table 4.8 – continued

$250 million AUM	First test	Forward tests 1st	Forward tests 2nd	Forward tests 3rd	Backward tests 1st	Backward tests 2nd	Backward tests 3rd
Net returns							
Number of funds							
Change in intercept & slopes	98	21	2	0	1	0	0
Change in slopes	74	15	2	0	12	1	0
Change in intercept	125	43	6	0	7	1	0
No change	2079	2297	2366	2376	2356	2374	2376
Proportions							
Change in intercept & slopes	4.1	0.9	0.1	0.0	0.0	0.0	0.0
Change in slopes	3.1	0.6	0.1	0.0	0.5	0.0	0.0
Change in intercept	5.3	1.8	0.3	0.0	0.3	0.0	0.0
No change	87.5	96.7	99.5	100.0	99.2	100.0	100.0
Gross returns							
Number of funds							
Change in intercept & slopes	102	22	1	0	3	0	0
Change in slopes	77	16	2	0	12	1	0
Change in intercept	131	44	7	0	6	1	0
No change	2066	2294	2366	2376	2355	2374	2376
Proportions							
Change in intercept & slopes	4.3	1.0	0.0	0.0	0.1	0.0	0.0
Change in slopes	3.2	0.7	0.1	0.0	0.5	0.0	0.0
Change in intercept	5.5	1.9	0.3	0.0	0.3	0.0	0.0
No change	87.0	96.4	99.6	100.0	99.1	100.0	100.0

continued

Table 4.8 – continued

$1 billion AUM	First test	Forward tests			Backward tests		
		1st	2nd	3rd	1st	2nd	3rd
			Net returns				
Number of funds							
Change in intercept & slopes	42	6	1	0	1	0	0
Change in slopes	31	8	1	0	3	1	0
Change in intercept	46	17	3	0	0	0	0
No change	974	1062	1088	1093	1089	1092	1093
Proportions							
Change in intercept & slopes	3.8	0.5	0.1	0.0	0.1	0.0	0.0
Change in slopes	2.8	0.7	0.1	0.0	0.3	0.1	0.0
Change in intercept	4.2	1.6	0.3	0.0	0.0	0.0	0.0
No change	89.2	97.2	99.5	100.0	99.6	99.9	100.0
			Gross returns				
Number of funds							
Change in intercept & slopes	41	6	0	0	1	0	0
Change in slopes	35	10	2	0	5	1	0
Change in intercept	46	20	4	0	0	0	0
No change	971	1057	1087	1093	1087	1092	1093
Proportions							
Change in intercept & slopes	3.8	0.5	0.0	0.0	0.1	0.0	0.0
Change in slopes	3.2	1.0	0.2	0.0	0.5	0.1	0.0
Change in intercept	4.2	1.8	0.4	0.0	0.0	0.0	0.0
No change	88.8	96.7	99.4	100.0	99.4	99.9	100.0

Table 4.9: Summary of Results on Tests for Change Points in the Slopes and the Intercept (4F-CAPM)-$250 Million Group & $1 Billion Group

$250 million AUM	First test	Forward tests			Backward tests		
		1st	2nd	3rd	1st	2nd	3rd
				Net returns			
Number of funds							
Change in intercept & slopes	84	17	0	0	3	1	0
Change in slopes	56	12	6	1	8	0	0
Change in intercept	124	46	9	3	5	4	1
No change	2112	2301	2361	2372	2360	2371	2375
Proportions							
Change in intercept & slopes	3.5	0.7	0.0	0.0	0.1	0.0	0.0
Change in slopes	2.4	0.5	0.3	0.0	0.3	0.0	0.0
Change in intercept	5.2	1.9	0.4	0.1	0.2	0.1	0.0
No change	88.9	96.9	99.3	99.9	99.4	99.9	100.0
				Gross returns			
Number of funds							
Change in intercept & slopes	88	17	1	0	4	1	0
Change in slopes	57	15	5	1	9	0	0
Change in intercept	134	42	8	3	6	3	1
No change	2097	2302	2362	2372	2357	2372	2375
Proportions							
Change in intercept & slopes	3.7	0.7	0.0	0.0	0.2	0.0	0.0
Change in slopes	2.4	0.6	0.2	0.0	0.4	0.0	0.0
Change in intercept	5.6	1.8	0.3	0.0	0.3	0.1	0.0
No change	88.3	96.9	99.5	100.0	99.1	99.9	100.0

continued

Table 4.9 – continued

$1 billion AUM	First test	Forward tests			Backward tests		
		1st	2nd	3rd	1st	2nd	3rd
			Net returns				
Number of funds							
Change in intercept & slopes	35	5	0	0	0	0	0
Change in slopes	27	4	2	0	7	0	0
Change in intercept	49	13	3	2	4	0	0
No change	982	1071	1088	1091	1082	1093	1093
Proportions							
Change in intercept & slopes	3.2	0.5	0.0	0.0	0.0	0.0	0.0
Change in slopes	2.5	0.4	0.1	0.0	0.6	0.0	0.0
Change in intercept	4.5	1.2	0.3	0.0	0.4	0.0	0.0
No change	89.8	97.9	99.6	100.0	99.0	100.0	100.0
			Gross returns				
Number of funds							
Change in intercept & slopes	37	6	0	0	1	0	0
Change in slopes	25	6	2	0	7	0	0
Change in intercept	51	13	3	2	4	0	0
No change	980	1068	1088	1091	1081	1093	1093
Proportions							
Change in intercept & slopes	3.3	0.5	0.0	0.0	0.1	0.0	0.0
Change in slopes	2.3	0.5	0.2	0.0	0.6	0.0	0.0
Change in intercept	4.7	1.2	0.3	0.0	0.4	0.0	0.0
No change	89.7	97.8	99.5	100.0	98.9	100.0	100.0

Table 4.10: Statistics on Different Combinations of Change Points: 1984-2015 (3F-CAPM)

Different cases of combinations	$5 million AUM group				$250 million AUM group				$1 billion AUM group			
	Net returns		Gross returns		Net returns		Gross returns		Net returns		Gross returns	
	funds	Pct(%)	funds	Pct(%)	funds	Pct(%)	funds	Pct(%)	funds	Pct(%)	funds	Pct(%)
Total number of funds	5006				2376				1093			
No change	3929	78.5	3909	78.1	2036	85.7	2020	85.0	954	87.3	947	86.6
1 change in both βs and α	330	6.6	325	6.5	91	3.8	88	3.7	39	3.6	39	3.6
1 change in βs and no change in α	217	4.3	229	4.6	66	2.8	70	2.9	30	2.7	35	3.2
1 change in α and no change in βs	386	7.7	388	7.8	133	5.6	144	6.1	55	5.0	55	5.0
2 changes in βs and no change in α	15	0.3	22	4.4	5	0.2	8	0.3	3	0.3	4	0.4
2 changes in βs and 1 change in α	33	0.7	31	0.6	7	0.3	7	0.3	2	0.2	2	0.2
2 changes in βs and 2 changes in α	21	0.4	24	0.5	11	0.5	14	0.6	5	0.5	5	0.5
1 change in βs and 2 changes in α	41	0.8	45	0.9	10	0.4	9	0.4	5	0.5	5	0.5
0 change in βs and 2 changes in α	22	0.4	22	0.4	12	0.5	11	0.5	0	0.0	1	0.1
3 changes in βs and 0 change in α	1	0.0	1	0.0	1	0.0	1	0.0	0	0.0	0	0.0
3 changes in βs and 1 change in α	2	0.0	1	0.0	2	0.1	1	0.0	0	0.0	0	0.0
3 changes in βs and 2 changes in α	1	0.0	0	0.0	1	0.0	0	0.0	0	0.0	0	0.0
3 changes in βs and 3 change in α	1	0.0	1	0.0	0	0.0	1	0.0	0	0.0	0	0.0
2 changes in βs and 3 changes in α	2	0.0	3	0.1	0	0.0	0	0.0	0	0.0	0	0.0
1 change in βs and 3 changes in α	4	0.1	4	0.1	1	0.0	2	0.1	0	0.0	0	0.0
0 change in βs and 3 changes in α	1	0.0	1	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Table 4.11: Statistics on Different Cases of Combinations of Change Points: 1984 to 2015 (4F-CAPM)-$250 Million Group & $1 Billion Group

Different cases of combinations	$5 million AUM group				$250 million AUM group				$1 billion AUM group			
	Net returns		Gross returns		Net returns		Gross returns		Net returns		Gross returns	
	funds	Pct(%)	funds	Pct(%)	funds	Pct(%)	funds	Pct(%)	funds	Pct(%)	funds	Pct(%)
Total number of funds	5006				2376				1093			
No change	3993	79.8	3979	79.5	2046	86.1	2035	85.6	961	79.8	958	87.6
1 change in both βs and α	270	5.4	283	5.7	85	3.6	89	3.7	33	5.4	33	3.0
1 change in βs and no change in α	212	4.2	196	4.0	58	2.4	58	2.4	33	4.2	31	2.8
1 change in α and no change in βs	387	7.7	399	8.0	150	6.3	156	6.6	52	7.7	53	4.8
2 changes in βs and no change in α	16	0.3	16	0.3	3	0.1	2	0.1	1	0.3	2	0.2
2 changes in βs and 1 change in α	21	0.4	21	0.4	5	0.2	7	0.3	1	0.4	2	0.2
2 changes in βs and 2 changes in α	18	0.4	22	0.4	2	0.1	7	0.3	2	0.4	3	0.3
1 change in βs and 2 changes in α	43	0.9	44	0.9	12	0.5	8	0.3	5	0.9	5	0.5
0 change in βs and 2 changes in α	34	0.7	31	0.6	10	0.4	10	0.4	4	0.7	5	0.5
3 changes in βs and 0 change in α	1	0.0	1	0.0	1	0.0	1	0.0	0	0.0	0	0.0
3 changes in βs and 1 change in α	0	0.0	1	0.0	1	0.0	1	0.0	0	0.0	0	0.0
3 changes in βs and 2 changes in α	1	0.0	0	0.0	1	0.0	1	0.0	0	0.0	0	0.0
3 changes in βs and 3 change in α	1	0.0	2	0.0	0	0.0	0	0.0	0	0.0	0	0.0
2 changes in βs and 3 changes in α	5	0.1	3	0.0	2	0.1	1	0.0	0	0.1	0	0.0
1 change in βs and 3 changes in α	4	0.1	7	0.1	0	0.0	0	0.0	1	0.1	1	0.1
0 change in βs and 3 changes in α	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Table 4.12: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2015 (Model with Time-invariant Parameters)**
The table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We bootstrap by resampling fund and factor returns jointly as a pair. The sample covers the period from January 1984 to March 2015. Results are shown for the 3F-CAPM and the 4F-CAPM with time-invariant parameters, for the $5 million, $250 million, and $1 billion AUM groups. There are 5006 funds in the $5 million group, 2376 funds in the $250 million group, and 1093 funds in the $1 billion group.

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
3F-CAPM, Net Returns									
1	-3.71	-2.40	0.02	-3.89	-2.45	0.17	-3.53	-2.51	1.64
2	-3.29	-2.09	0.01	-3.20	-2.12	0.18	-3.03	-2.14	0.97
3	-3.01	-1.90	0.00	-2.90	-1.92	0.18	-2.79	-1.94	0.65
4	-2.85	-1.77	0.00	-2.78	-1.78	0.11	-2.56	-1.79	0.69
5	-2.70	-1.66	0.00	-2.59	-1.67	0.12	-2.44	-1.68	0.49
10	-2.23	-1.28	0.00	-2.10	-1.29	0.10	-2.03	-1.29	0.24
20	-1.63	-0.84	0.00	-1.51	-0.84	0.16	-1.50	-0.84	0.19
30	-1.23	-0.53	0.03	-1.08	-0.52	0.37	-1.08	-0.53	0.35
40	-0.86	-0.26	0.07	-0.78	-0.25	0.28	-0.70	-0.26	0.94
50	-0.57	-0.01	0.09	-0.49	-0.00	0.30	-0.44	-0.01	0.69
60	-0.27	0.24	0.15	-0.20	0.25	0.46	-0.18	0.24	0.60
70	0.07	0.51	0.34	0.10	0.52	0.93	0.13	0.51	1.09
80	0.46	0.83	1.48	0.49	0.83	3.16	0.57	0.83	8.64
90	1.00	1.27	7.51	1.06	1.28	14.86	1.08	1.27	17.96
95	1.47	1.65	20.29	1.48	1.65	23.61	1.42	1.66	15.15
96	1.65	1.76	33.35	1.61	1.77	27.87	1.54	1.77	18.03
97	1.84	1.90	42.24	1.78	1.91	32.99	1.68	1.91	19.09
98	2.11	2.09	57.49	2.08	2.09	51.91	1.94	2.11	29.23
99	2.48	2.41	64.43	2.46	2.41	61.14	2.32	2.46	36.55

continued

Table 4.12 – *continued*

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
4F-CAPM, Net Returns									
1	-3.79	-2.46	0.01	-3.92	-2.54	0.40	-3.35	-2.65	6.11
2	-3.26	-2.12	0.01	-3.15	-2.16	0.31	-3.10	-2.21	1.47
3	-2.99	-1.93	0.01	-2.89	-1.95	0.23	-2.78	-1.99	1.10
4	-2.86	-1.78	0.01	-2.75	-1.80	0.14	-2.56	-1.83	1.00
5	-2.71	-1.67	0.01	-2.62	-1.69	0.14	-2.40	-1.71	1.03
10	-2.23	-1.29	0.01	-2.11	-1.30	0.10	-1.98	-1.31	0.45
20	-1.68	-0.84	0.02	-1.51	-0.84	0.26	-1.44	-0.85	0.51
30	-1.24	-0.52	0.02	-1.14	-0.52	0.18	-1.08	-0.53	0.41
40	-0.91	-0.25	0.03	-0.83	-0.25	0.19	-0.76	-0.25	0.45
50	-0.61	-0.00	0.03	-0.56	-0.00	0.15	-0.54	-0.00	0.17
60	-0.32	0.25	0.04	-0.26	0.25	0.23	-0.23	0.25	0.27
70	0.01	0.52	0.09	0.06	0.52	0.40	0.08	0.52	0.46
80	0.40	0.84	0.40	0.47	0.84	1.96	0.52	0.84	3.85
90	1.00	1.29	6.57	1.00	1.29	7.98	1.05	1.30	11.43
95	1.49	1.68	19.20	1.49	1.68	21.00	1.47	1.70	17.02
96	1.62	1.80	22.14	1.59	1.80	19.51	1.58	1.82	16.85
97	1.85	1.94	37.98	1.74	1.95	21.47	1.73	1.97	18.15
98	2.12	2.14	50.18	1.93	2.15	19.96	1.91	2.19	16.66
99	2.56	2.49	63.18	2.37	2.51	34.67	2.23	2.60	15.33
3F-CAPM, Gross Returns									
1	-2.90	-2.40	3.78	-2.88	-2.45	8.82	-2.64	-2.50	28.66
2	-2.52	-2.09	4.77	-2.42	-2.12	13.16	-2.41	-2.14	16.92
3	-2.28	-1.90	5.81	-2.20	-1.92	13.43	-2.14	-1.94	21.42
4	-2.12	-1.77	6.49	-2.05	-1.78	13.58	-1.96	-1.79	23.94
5	-2.00	-1.66	6.85	-1.94	-1.67	12.71	-1.84	-1.68	23.98
10	-1.53	-1.28	11.61	-1.39	-1.29	30.20	-1.39	-1.29	29.88
20	-0.92	-0.84	31.60	-0.83	-0.84	49.31	-0.82	-0.84	51.75
30	-0.51	-0.52	50.99	-0.43	-0.53	67.22	-0.45	-0.53	63.89
40	-0.19	-0.26	64.45	-0.11	-0.26	78.05	-0.08	-0.26	85.26
50	0.13	-0.01	76.66	0.17	-0.01	83.62	0.16	-0.01	83.87
60	0.43	0.24	85.28	0.45	0.24	86.99	0.43	0.24	85.85
70	0.75	0.51	90.15	0.76	0.51	89.86	0.73	0.51	88.82
80	1.14	0.83	93.48	1.13	0.83	92.17	1.12	0.83	92.49
90	1.72	1.28	96.94	1.68	1.27	95.97	1.67	1.28	95.53
95	2.23	1.65	98.50	2.20	1.65	98.03	2.09	1.66	94.99
96	2.38	1.76	98.72	2.31	1.76	97.77	2.18	1.77	93.76
97	2.55	1.90	98.79	2.56	1.90	98.80	2.31	1.91	92.65
98	2.81	2.09	99.13	2.80	2.09	98.85	2.65	2.12	95.68
99	3.26	2.41	99.44	3.19	2.41	98.83	3.05	2.47	94.59

continued

Table 4.12 – continued

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
4F-CAPM, Gross Returns									
1	-2.99	-2.47	3.40	-2.79	-2.54	19.31	-2.62	-2.66	45.39
2	-2.53	-2.13	5.80	-2.34	-2.16	23.40	-2.27	-2.22	39.22
3	-2.32	-1.93	5.84	-2.13	-1.95	22.79	-2.03	-2.00	40.46
4	-2.11	-1.79	8.59	-2.02	-1.80	18.47	-1.90	-1.84	36.61
5	-1.99	-1.68	8.46	-1.89	-1.68	18.11	-1.72	-1.72	45.83
10	-1.54	-1.29	11.80	-1.42	-1.29	26.81	-1.37	-1.31	37.35
20	-0.93	-0.85	31.68	-0.84	-0.84	47.58	-0.84	-0.85	51.20
30	-0.53	-0.53	48.10	-0.49	-0.52	54.74	-0.46	-0.53	64.64
40	-0.22	-0.26	57.56	-0.19	-0.25	62.35	-0.15	-0.26	71.88
50	0.09	-0.01	69.44	0.13	0.00	75.06	0.13	-0.01	78.78
60	0.38	0.25	77.57	0.43	0.25	82.35	0.40	0.25	81.59
70	0.71	0.52	84.39	0.73	0.52	85.32	0.70	0.52	84.38
80	1.10	0.84	90.30	1.10	0.84	88.82	1.12	0.84	91.53
90	1.73	1.29	97.14	1.68	1.30	94.38	1.68	1.30	94.39
95	2.26	1.68	98.61	2.19	1.68	97.08	2.14	1.70	94.99
96	2.40	1.79	98.76	2.36	1.80	97.70	2.28	1.82	95.05
97	2.63	1.94	99.10	2.55	1.95	98.02	2.46	1.97	95.27
98	2.89	2.14	99.36	2.82	2.15	98.28	2.59	2.19	90.69
99	3.40	2.49	99.67	3.19	2.52	97.43	2.91	2.61	82.86

Table 4.13: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2015 (Block Bootstrap with Time-varying Parameters)**

This table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We bootstrap by resampling fund and factor returns as a pair. The sample covers the period from January 1984 to March 2015. Results are shown for the multi-regression model with time-varying parameters based on the 3F-CAPM and the 4F-CAPM, for the $5 million, $250 million, and $1 billion AUM groups. There are 5006 funds in the $5 million group, 2376 funds in the $250 million group, and 1093 funds in the $1 billion group.

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%<	Act	Sim	%<	Act	Sim	%<
3F-CAPM with time-varying parameters, net returns									
1	-4.04	-2.56	0.15	-4.04	-2.60	0.45	-3.81	-2.65	2.20
2	-3.62	-2.20	0.03	-3.56	-2.23	0.22	-3.38	-2.26	1.02
3	-3.35	-1.99	0.01	-3.31	-2.01	0.11	-3.16	-2.04	0.54
4	-3.13	-1.84	0.01	-3.03	-1.86	0.15	-3.01	-1.88	0.30
5	-3.01	-1.72	0.00	-2.86	-1.73	0.14	-2.79	-1.75	0.32
10	-2.47	-1.32	0.00	-2.34	-1.33	0.07	-2.24	-1.33	0.17
20	-1.82	-0.85	0.00	-1.68	-0.86	0.06	-1.63	-0.86	0.12
30	-1.34	-0.53	0.00	-1.18	-0.54	0.16	-1.17	-0.54	0.14
40	-0.93	-0.26	0.00	-0.82	-0.26	0.22	-0.78	-0.26	0.32
50	-0.56	-0.01	0.00	-0.50	-0.01	0.37	-0.45	-0.01	0.72
60	-0.21	0.24	0.14	-0.17	0.24	0.94	-0.17	0.24	0.86
70	0.18	0.51	1.88	0.18	0.51	3.69	0.23	0.51	5.97
80	0.69	0.84	24.09	0.69	0.83	27.35	0.72	0.83	31.44
90	1.52	1.30	84.14	1.40	1.29	70.46	1.37	1.29	65.48
95	2.23	1.70	96.66	2.09	1.69	92.09	1.91	1.69	79.56
96	2.44	1.82	97.84	2.26	1.81	93.51	2.18	1.82	88.84
97	2.66	1.97	98.27	2.50	1.95	95.37	2.37	1.97	89.76
98	2.96	2.18	98.72	2.74	2.16	95.40	2.64	2.18	90.92
99	3.42	2.54	98.66	3.14	2.51	94.89	2.89	2.55	82.83

continued

Table 4.13 – continued

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%<	Act	Sim	%<	Act	Sim	%<
4F-CAPM with time-varying parameters, net returns									
1	-4.15	-2.61	0.10	-4.15	-2.64	0.57	-3.48	-2.71	7.07
2	-3.62	-2.22	0.02	-3.49	-2.24	0.36	-3.26	-2.29	1.87
3	-3.33	-2.00	0.00	-3.20	-2.01	0.28	-3.11	-2.05	0.77
4	-3.12	-1.85	0.01	-2.97	-1.85	0.23	-2.87	-1.89	0.72
5	-2.97	-1.72	0.00	-2.83	-1.73	0.14	-2.65	-1.76	0.84
10	-2.44	-1.32	0.00	-2.27	-1.32	0.11	-2.17	-1.34	0.43
20	-1.80	-0.85	0.00	-1.64	-0.85	0.08	-1.59	-0.86	0.34
30	-1.32	-0.53	0.00	-1.24	-0.52	0.06	-1.21	-0.53	0.19
40	-0.95	-0.26	0.00	-0.88	-0.25	0.05	-0.82	-0.26	0.38
50	-0.59	-0.01	0.01	-0.58	0.00	0.08	-0.54	-0.01	0.32
60	-0.26	0.24	0.04	-0.23	0.24	0.36	-0.20	0.24	0.60
70	0.12	0.51	0.58	0.15	0.51	2.25	0.16	0.51	2.65
80	0.60	0.83	11.38	0.61	0.84	14.76	0.70	0.84	27.94
90	1.38	1.30	66.85	1.34	1.30	59.51	1.32	1.31	55.85
95	2.08	1.70	91.92	1.93	1.71	79.63	1.88	1.73	73.89
96	2.29	1.83	95.02	2.14	1.83	85.74	1.99	1.85	71.33
97	2.51	1.99	96.26	2.37	1.99	89.01	2.22	2.02	76.88
98	2.79	2.20	96.58	2.54	2.21	85.78	2.35	2.24	66.22
99	3.31	2.59	97.41	2.88	2.60	80.95	2.80	2.66	68.09
3F-CAPM with time-varying parameters, gross returns									
1	-3.34	-2.56	2.64	-3.22	-2.59	7.04	-3.05	-2.65	17.48
2	-2.98	-2.20	1.37	-2.88	-2.22	4.02	-2.64	-2.26	14.55
3	-2.72	-1.98	1.35	-2.64	-2.01	3.49	-2.50	-2.04	9.41
4	-2.50	-1.83	1.47	-2.41	-1.85	4.31	-2.28	-1.88	10.39
5	-2.36	-1.71	1.39	-2.25	-1.73	4.80	-2.13	-1.75	10.91
10	-1.83	-1.31	1.69	-1.70	-1.33	7.67	-1.60	-1.34	14.25
20	-1.15	-0.85	6.57	-1.03	-0.86	21.22	-1.00	-0.87	25.95
30	-0.65	-0.53	22.85	-0.58	-0.54	39.62	-0.58	-0.54	39.54
40	-0.28	-0.25	42.38	-0.19	-0.26	63.87	-0.16	-0.27	70.90
50	0.09	0.00	71.11	0.14	-0.01	79.32	0.13	-0.02	79.94
60	0.44	0.25	86.74	0.46	0.24	88.55	0.43	0.23	85.34
70	0.84	0.52	95.53	0.82	0.51	93.64	0.77	0.50	90.71
80	1.35	0.84	98.82	1.29	0.83	97.25	1.29	0.83	97.48
90	2.16	1.30	99.79	2.05	1.29	99.42	1.93	1.29	98.59
95	2.80	1.70	99.93	2.72	1.69	99.77	2.54	1.69	99.31
96	2.96	1.82	99.93	2.87	1.81	99.73	2.73	1.81	99.33
97	3.14	1.97	99.91	3.00	1.96	99.63	2.87	1.97	99.17
98	3.43	2.18	99.91	3.22	2.16	99.47	3.14	2.18	99.08
99	3.95	2.53	99.88	3.74	2.51	99.49	3.41	2.54	97.12

continued

Table 4.13 – *continued*

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%<	Act	Sim	%<	Act	Sim	%<
4F-CAPM with time-varying parameters, gross returns									
1	-3.43	-2.60	2.08	-3.27	-2.65	7.12	-3.09	-2.72	18.89
2	-2.96	-2.22	1.64	-2.86	-2.24	4.81	-2.63	-2.29	17.13
3	-2.69	-2.00	1.59	-2.48	-2.02	7.53	-2.39	-2.06	15.70
4	-2.46	-1.84	1.92	-2.28	-1.86	8.40	-2.20	-1.89	16.09
5	-2.31	-1.72	1.88	-2.11	-1.73	9.46	-2.03	-1.76	17.58
10	-1.77	-1.31	3.15	-1.63	-1.32	10.29	-1.56	-1.34	19.04
20	-1.14	-0.85	7.80	-1.01	-0.85	21.31	-1.00	-0.86	25.35
30	-0.66	-0.52	22.27	-0.61	-0.53	32.41	-0.61	-0.53	33.26
40	-0.27	-0.25	45.44	-0.26	-0.26	49.11	-0.21	-0.26	59.28
50	0.07	-0.01	66.49	0.08	-0.01	69.36	0.11	-0.01	73.90
60	0.41	0.24	82.09	0.44	0.24	85.43	0.41	0.24	81.09
70	0.77	0.51	91.62	0.80	0.51	91.90	0.75	0.52	87.32
80	1.29	0.83	98.05	1.25	0.83	96.21	1.23	0.84	95.11
90	2.04	1.30	99.71	2.01	1.30	99.18	1.92	1.32	98.01
95	2.70	1.71	99.89	2.60	1.70	99.48	2.51	1.74	98.79
96	2.87	1.83	99.89	2.76	1.83	99.47	2.59	1.86	98.06
97	3.07	1.99	99.87	2.93	1.98	99.30	2.65	2.03	96.26
98	3.35	2.21	99.85	3.14	2.20	99.10	2.97	2.26	96.70
99	3.89	2.59	99.81	3.62	2.59	98.84	3.26	2.68	91.42

Table 4.14:

Intercepts and Slopes in Variants of the CAPM for Equal-Weight (EW) and Value-Weight (VW) Portfolios of Actively Managed Mutual Funds (1984-2006)

This table provides the annualized intercepts ($12^*\hat{\alpha}$) and t-statistics for the intercepts ($t(Coef)$) for the CAPM, the 3F-CAPM, and the 4F-CAPM estimated on an equal-weight (EW) and value-weight (VW) portfolio, and on net and gross returns based on a panel of actively managed mutual funds that invest mainly in equities. Regressions slopes ($\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$) for the corresponding factors ($RM - Rf$, SMB, HML, and MOM) are also shown in the table. It should be noted that for the market slope, $t(Coef)$ tests the null hypothesis that β is equal to 1. Net returns are returns reported in the CRSP mutual fund database, and the gross returns are net returns plus $1/12^{th}$ of a fund's expense ratio at that year end. When a fund's expense ratio for a year is missing, we replace it with the expense ratio of the fund that has the same investment style and has similar assets under management (AUM). As for those funds with zero expense ratios, we manually check some of their annual reports and confirm that they in fact have positive expense ratios. Hence, we treat zero expense ratios as missing values. Our sample covers the period from January 1984 to September 2006. There are 1360 funds on average every month and their average AUM is $705 million.

	$12^*\hat{\alpha}$ (%)						
	Net	Gross	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2
VW Returns							
CAPM $Coef$	-1.02	-0.03	0.96				0.97
$t(Coef)$	-2.14	-0.07	-4.58				
3F-CAPM $Coef$	-0.97	0.01	0.95	0.10	0.00		0.98
$t(Coef)$	-2.25	0.03	-5.83	8.38	0.32		
4F-CAPM $Coef$	-1.08	-0.10	0.95	0.09	0.01	0.01	0.98
$t(Coef)$	-2.45	-0.22	-5.60	8.25	0.41	1.17	
EW Returns							
CAPM $Coef$	-1.00	0.39	0.96				0.95
$t(Coef)$	-1.47	0.57	-3.11				
3F-CAPM $Coef$	-1.03	0.37	0.94	0.20	0.03		0.98
$t(Coef)$	-2.04	0.73	-5.67	15.40	1.58		
4F-CAPM $Coef$	-1.03	0.37	0.94	0.20	0.03	0.00	0.98
$t(Coef)$	-2.00	0.71	-5.60	15.31	1.57	0.02	

Table 4.15: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2006 (Model with Time-invariant parameters)**
This table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We bootstrap by resampling fund and factor returns as a pair. The sample covers the period from January 1984 to September 2006. Results are shown for the 3F-CAPM and the 4F-CAPM, for the $5 million, $250 million, and $1 billion AUM groups. There are 3174 funds in the $5 million group, 1466 funds in the $250 million group, and 686 funds in the $1 billion group.

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
3F-CAPM, Net Returns									
1	-4.00	-2.39	0.01	-3.84	-2.43	0.35	-3.53	-2.51	2.27
2	-3.53	-2.08	0.01	-3.32	-2.10	0.20	-3.01	-2.12	1.27
3	-3.23	-1.90	0.01	-3.12	-1.90	0.12	-2.88	-1.92	0.51
4	-3.03	-1.76	0.02	-2.91	-1.76	0.12	-2.78	-1.78	0.23
5	-2.89	-1.65	0.02	-2.70	-1.65	0.15	-2.62	-1.66	0.27
10	-2.37	-1.28	0.02	-2.25	-1.28	0.11	-2.15	-1.28	0.23
20	-1.76	-0.84	0.01	-1.58	-0.84	0.12	-1.52	-0.83	0.24
30	-1.28	-0.52	0.02	-1.10	-0.52	0.32	-1.05	-0.52	0.55
40	-0.90	-0.25	0.02	-0.79	-0.25	0.29	-0.68	-0.25	1.13
50	-0.63	0.00	0.02	-0.50	0.00	0.30	-0.42	0.00	1.02
60	-0.27	0.25	0.08	-0.19	0.25	0.79	-0.12	0.25	1.83
70	0.09	0.52	0.43	0.12	0.52	1.49	0.20	0.52	4.64
80	0.48	0.84	2.04	0.46	0.83	2.51	0.52	0.84	5.43
90	1.03	1.29	11.34	0.98	1.28	8.91	1.04	1.28	15.20
95	1.53	1.67	31.84	1.39	1.66	14.29	1.42	1.67	18.50
96	1.65	1.78	33.97	1.52	1.77	17.05	1.66	1.78	35.61
97	1.87	1.92	46.96	1.71	1.91	23.80	1.78	1.93	34.56
98	2.15	2.10	61.17	1.92	2.11	27.91	2.05	2.13	44.82
99	2.48	2.42	61.96	2.23	2.44	29.25	2.32	2.50	37.53

continued

Table 4.15 – *continued*

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
4F-CAPM, Net Returns									
1	-4.00	-2.44	0.00	-3.98	-2.50	0.24	-3.63	-2.64	3.70
2	-3.49	-2.11	0.01	-3.24	-2.13	0.14	-3.13	-2.17	0.90
3	-3.18	-1.91	0.01	-2.96	-1.92	0.09	-2.70	-1.95	1.39
4	-2.95	-1.77	0.01	-2.73	-1.77	0.16	-2.59	-1.80	0.72
5	-2.77	-1.66	0.03	-2.59	-1.66	0.13	-2.43	-1.68	0.92
10	-2.27	-1.28	0.01	-2.11	-1.27	0.16	-1.95	-1.28	0.88
20	-1.68	-0.83	0.01	-1.47	-0.82	0.48	-1.43	-0.83	0.83
30	-1.24	-0.51	0.02	-1.08	-0.51	0.56	-1.01	-0.51	1.50
40	-0.88	-0.24	0.06	-0.77	-0.24	0.74	-0.71	-0.24	1.61
50	-0.59	0.01	0.08	-0.50	0.01	0.63	-0.39	0.01	2.59
60	-0.31	0.26	0.08	-0.22	0.26	0.85	-0.09	0.26	4.24
70	0.00	0.53	0.18	0.10	0.52	1.43	0.17	0.53	4.17
80	0.41	0.86	0.79	0.48	0.84	3.94	0.61	0.85	14.69
90	1.02	1.31	9.89	0.97	1.29	8.43	1.13	1.31	26.23
95	1.50	1.70	22.71	1.47	1.68	22.74	1.59	1.70	37.11
96	1.61	1.82	23.28	1.54	1.80	18.26	1.76	1.82	44.84
97	1.83	1.96	33.38	1.64	1.95	14.14	1.94	1.97	49.07
98	2.10	2.16	43.92	1.80	2.15	11.55	2.19	2.19	53.47
99	2.45	2.51	44.75	2.06	2.51	8.20	2.45	2.70	40.57
3F-CAPM, Gross Returns									
1	-3.16	-2.39	1.28	-2.92	-2.43	8.28	-2.75	-2.52	24.82
2	-2.75	-2.08	1.69	-2.54	-2.10	8.00	-2.46	-2.13	15.87
3	-2.48	-1.90	2.32	-2.33	-1.91	7.73	-2.21	-1.92	17.21
4	-2.31	-1.76	2.69	-2.17	-1.77	8.07	-2.15	-1.78	10.89
5	-2.18	-1.65	2.83	-2.04	-1.65	8.19	-1.96	-1.67	14.12
10	-1.72	-1.28	3.91	-1.60	-1.28	9.64	-1.57	-1.28	12.59
20	-1.11	-0.84	9.32	-0.99	-0.84	22.92	-1.00	-0.84	22.53
30	-0.67	-0.52	21.19	-0.58	-0.52	37.06	-0.57	-0.52	38.82
40	-0.35	-0.25	28.79	-0.28	-0.25	43.38	-0.19	-0.25	61.85
50	-0.03	0.00	43.86	0.03	0.00	58.27	0.05	-0.01	61.70
60	0.31	0.26	62.95	0.30	0.25	62.52	0.38	0.24	77.52
70	0.65	0.53	75.74	0.61	0.52	68.86	0.66	0.51	78.33
80	1.03	0.85	82.20	0.96	0.84	73.62	0.95	0.83	73.38
90	1.58	1.29	89.35	1.50	1.28	82.76	1.47	1.28	78.95
95	2.04	1.67	92.88	1.92	1.66	83.66	1.94	1.66	83.89
96	2.18	1.78	93.43	2.06	1.78	85.33	2.06	1.78	83.57
97	2.40	1.92	95.04	2.26	1.92	88.13	2.29	1.92	88.41
98	2.62	2.11	95.34	2.50	2.11	89.19	2.39	2.12	80.78
99	2.99	2.42	95.85	2.80	2.44	85.76	2.66	2.49	71.13

continued

Table 4.15 – *continued*

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
4F-CAPM, Gross Returns									
1	-3.23	-2.44	0.96	-2.78	-2.50	18.20	-2.78	-2.62	29.64
2	-2.73	-2.10	1.71	-2.50	-2.13	9.88	-2.30	-2.17	31.94
3	-2.46	-1.91	2.43	-2.22	-1.92	13.21	-2.02	-1.95	38.41
4	-2.26	-1.76	3.17	-2.05	-1.77	13.81	-1.88	-1.80	35.83
5	-2.10	-1.65	4.29	-1.92	-1.65	14.27	-1.73	-1.68	39.21
10	-1.62	-1.27	6.50	-1.50	-1.27	16.04	-1.42	-1.28	27.18
20	-1.05	-0.82	13.76	-0.90	-0.82	34.57	-0.90	-0.83	37.30
30	-0.62	-0.50	26.30	-0.55	-0.51	41.09	-0.52	-0.51	47.89
40	-0.28	-0.23	40.21	-0.24	-0.24	49.26	-0.17	-0.24	64.55
50	0.00	0.02	46.03	0.01	0.01	51.98	0.10	0.01	68.96
60	0.29	0.27	56.39	0.31	0.26	62.91	0.42	0.25	79.09
70	0.63	0.54	68.31	0.64	0.52	72.57	0.67	0.52	76.17
80	1.01	0.86	77.05	0.95	0.84	70.11	1.09	0.85	85.93
90	1.57	1.32	85.92	1.46	1.29	76.47	1.62	1.30	89.04
95	2.05	1.70	90.77	1.93	1.68	83.21	2.13	1.70	92.24
96	2.20	1.82	91.95	2.08	1.80	85.12	2.27	1.82	92.24
97	2.38	1.97	92.48	2.25	1.95	85.85	2.34	1.97	87.79
98	2.62	2.17	93.38	2.43	2.15	82.53	2.56	2.19	85.97
99	2.99	2.52	92.45	2.78	2.51	79.68	2.83	2.70	74.22

Table 4.16: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2006 (Block Bootstrap)**

This table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We do block bootstrap using a block length of $T^{1/5}$ in which T is the number of months in the sample. The sample covers the period from January 1984 to September 2006. Results are shown for the 3F-CAPM and the 4F-CAPM, for the \$5 million, \$250 million, and \$1 billion AUM groups. There are 3174 funds in the \$5 million group, 1466 funds in the \$250 million group, and 686 funds in the \$1 billion group.

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
3F-CAPM, Net Returns									
1	-4.00	-2.46	0.14	-3.84	-2.47	1.04	-3.53	-2.55	4.51
2	-3.53	-2.13	0.09	-3.32	-2.13	0.74	-3.01	-2.16	3.82
3	-3.23	-1.93	0.11	-3.12	-1.93	0.47	-2.88	-1.95	1.98
4	-3.03	-1.78	0.11	-2.91	-1.78	0.46	-2.78	-1.80	1.25
5	-2.89	-1.67	0.09	-2.70	-1.66	0.51	-2.62	-1.68	1.18
10	-2.37	-1.28	0.06	-2.25	-1.28	0.23	-2.15	-1.29	0.87
20	-1.76	-0.83	0.02	-1.58	-0.83	0.27	-1.52	-0.83	0.80
30	-1.28	-0.51	0.02	-1.10	-0.51	0.43	-1.05	-0.52	1.20
40	-0.90	-0.24	0.03	-0.79	-0.24	0.30	-0.68	-0.25	1.80
50	-0.63	0.01	0.02	-0.50	0.01	0.35	-0.42	0.00	1.57
60	-0.27	0.26	0.04	-0.19	0.25	0.70	-0.12	0.24	2.74
70	0.09	0.53	0.35	0.12	0.52	1.94	0.20	0.51	6.98
80	0.48	0.85	2.64	0.46	0.84	3.54	0.52	0.83	8.68
90	1.03	1.31	13.16	0.98	1.29	11.28	1.04	1.28	19.61
95	1.53	1.70	31.62	1.39	1.68	16.57	1.42	1.68	22.21
96	1.65	1.81	32.87	1.52	1.80	19.17	1.66	1.79	37.55
97	1.87	1.96	43.87	1.71	1.94	25.17	1.78	1.94	35.65
98	2.15	2.16	54.35	1.92	2.14	27.90	2.05	2.15	44.15
99	2.48	2.50	53.32	2.23	2.48	29.18	2.32	2.54	36.47

continued

Table 4.16 – *continued*

Pct	5 Million Act	Sim	%>	250 Million Act	Sim	%>	1 Billion Act	Sim	%>
4F-CAPM, Net Returns									
1	-4.00	-2.48	0.16	-3.98	-2.50	0.51	-3.63	-2.59	4.17
2	-3.49	-2.13	0.13	-3.24	-2.13	0.57	-3.13	-2.17	2.03
3	-3.18	-1.92	0.15	-2.96	-1.92	0.48	-2.70	-1.95	3.30
4	-2.95	-1.77	0.14	-2.73	-1.77	0.56	-2.59	-1.80	1.90
5	-2.77	-1.65	0.13	-2.59	-1.65	0.50	-2.43	-1.67	1.96
10	-2.27	-1.26	0.07	-2.11	-1.26	0.34	-1.95	-1.27	1.87
20	-1.68	-0.81	0.03	-1.47	-0.81	0.69	-1.43	-0.82	1.46
30	-1.24	-0.49	0.02	-1.08	-0.50	1.01	-1.01	-0.51	2.47
40	-0.88	-0.23	0.04	-0.77	-0.24	1.10	-0.71	-0.24	2.51
50	-0.59	0.02	0.05	-0.50	0.01	1.01	-0.39	0.00	4.51
60	-0.31	0.27	0.10	-0.22	0.25	1.34	-0.09	0.24	6.82
70	0.00	0.54	0.20	0.10	0.51	2.68	0.17	0.51	7.19
80	0.41	0.85	1.63	0.48	0.83	6.81	0.61	0.83	20.37
90	1.02	1.31	14.29	0.97	1.28	13.17	1.13	1.29	32.40
95	1.50	1.70	26.58	1.47	1.67	27.75	1.59	1.69	41.27
96	1.61	1.82	26.84	1.54	1.79	23.04	1.76	1.81	48.19
97	1.83	1.98	36.26	1.64	1.95	18.35	1.94	1.97	51.61
98	2.10	2.18	44.18	1.80	2.16	14.92	2.19	2.19	55.02
99	2.45	2.55	43.74	2.06	2.52	11.01	2.45	2.61	42.30
3F-CAPM, Gross Returns									
1	-3.16	-2.46	4.14	-2.92	-2.47	13.06	-2.75	-2.55	28.29
2	-2.75	-2.13	4.03	-2.54	-2.13	12.76	-2.46	-2.16	20.95
3	-2.48	-1.93	4.75	-2.33	-1.93	12.01	-2.21	-1.95	22.72
4	-2.31	-1.78	4.99	-2.17	-1.78	11.84	-2.15	-1.80	15.45
5	-2.18	-1.67	4.84	-2.04	-1.66	11.85	-1.96	-1.68	19.13
10	-1.72	-1.28	5.35	-1.60	-1.28	12.57	-1.57	-1.29	16.89
20	-1.11	-0.83	10.31	-0.99	-0.83	24.14	-1.00	-0.83	25.14
30	-0.67	-0.51	21.98	-0.58	-0.51	35.85	-0.57	-0.52	40.18
40	-0.35	-0.24	29.60	-0.28	-0.24	42.32	-0.19	-0.25	60.77
50	-0.03	0.01	44.42	0.03	0.01	56.60	0.05	-0.01	61.32
60	0.31	0.26	62.26	0.30	0.25	61.38	0.38	0.24	75.79
70	0.65	0.53	73.79	0.61	0.52	68.28	0.66	0.51	76.63
80	1.03	0.85	79.11	0.96	0.84	72.25	0.95	0.82	72.29
90	1.58	1.31	85.08	1.50	1.29	79.74	1.47	1.28	76.70
95	2.04	1.70	87.87	1.92	1.68	80.16	1.94	1.67	80.88
96	2.18	1.82	88.45	2.06	1.79	81.86	2.06	1.79	80.25
97	2.40	1.96	91.16	2.26	1.94	84.48	2.29	1.94	84.97
98	2.62	2.16	91.15	2.50	2.14	85.74	2.39	2.15	76.56
99	2.99	2.50	90.63	2.80	2.48	81.26	2.66	2.53	66.88

continued

Table 4.16 – *continued*

	5 Million			250 Million			1 Billion		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
4F-CAPM, Gross Returns									
1	-3.23	-2.54	3.87	-2.78	-2.50	21.90	-2.78	-2.59	29.81
2	-2.73	-2.17	4.84	-2.50	-2.13	14.89	-2.30	-2.17	33.98
3	-2.46	-1.96	5.85	-2.22	-1.92	18.14	-2.02	-1.96	40.46
4	-2.26	-1.81	6.74	-2.05	-1.77	18.63	-1.88	-1.80	37.98
5	-2.10	-1.69	8.17	-1.92	-1.65	18.85	-1.73	-1.68	40.58
10	-1.62	-1.29	10.62	-1.50	-1.26	19.84	-1.42	-1.28	30.51
20	-1.05	-0.83	18.34	-0.90	-0.81	34.95	-0.90	-0.83	38.35
30	-0.62	-0.50	30.70	-0.55	-0.50	40.60	-0.52	-0.51	47.49
40	-0.28	-0.23	42.56	-0.24	-0.24	48.66	-0.17	-0.25	63.45
50	0.00	0.02	47.75	0.01	0.01	52.33	0.10	-0.01	68.73
60	0.29	0.27	55.84	0.31	0.25	63.38	0.42	0.24	78.77
70	0.63	0.55	66.15	0.64	0.51	72.75	0.67	0.50	76.67
80	1.01	0.87	73.41	0.95	0.83	70.64	1.09	0.82	84.81
90	1.57	1.34	80.88	1.46	1.28	76.04	1.62	1.28	86.90
95	2.05	1.74	85.05	1.93	1.67	81.38	2.13	1.68	90.00
96	2.20	1.86	86.34	2.08	1.79	82.91	2.27	1.81	90.17
97	2.38	2.01	86.80	2.25	1.94	83.13	2.34	1.97	85.40
98	2.62	2.22	87.61	2.43	2.15	80.34	2.56	2.19	83.65
99	2.99	2.60	85.87	2.78	2.52	77.36	2.83	2.61	72.95

Chapter 5

False Discoveries in Time-varying Mutual Fund Performance

5.1 Introduction

In Chapter 4, we have shown that, using a multi-regression model with time-varying regression parameters to estimate fund alphas, the distributions of fund $t(\alpha)$ estimates shifts rightwards, suggesting more evidence of manager skill for both net and gross returns than previously documented in the literature. While Chapter 4 examined fund performance qualitatively, this chapter aims to quantify the number of skilled and unskilled managers. In particular, we aim to find how many actively managed fund managers within the entire sample have truly abnormal performance that produces positive, negative or zero alphas, and where these funds are located in the cross-sectional distribution of $t(\alpha)$ estimates. This can be important for investment purposes since knowing the proportion of skilled funds in the population and their location within this cross-section distribution provides important information for investors to find the funds that offer excess returns.

To control for the role of luck, this chapter uses a False Discovery Rate (FDR) technique so that the false discoveries (luck) can be controlled in a multiple hypothesis test; in this study the multiple hypothesis to be tested is that each fund has a zero alpha. Given the multitude of funds in our sample, we are in a situation of testing a number of hypotheses, each at a fixed significance level. While individual test error rate can be easily controlled at the significance level, it doesn't control the experiment-wise error rate. Using the FDR rate to estimate the proportion of lucky (unlucky) funds, Barras, Scaillet and Wermers (2010) (hereafter BSW) are able to estimate the number of U.S. mutual funds whose managers generate excess returns due to their superior stock selecting skill.

However, the regression parameters (fund alphas and betas) in the extended versions of the CAPM used by BSW are assumed to be time-invariant. As was pointed out in Chapter 4, Section 4.3.1, this does not hold in these models for a number of reasons that are discussed there. Hence, to extend BSW's study to allow for time-variation in the alpha and betas, we use a non-parametric change point test introduced in Chapter 4, Section 4.3.2 to detect possible time-variation in these regression parameters. Using this novel test is better than using the conditional method of Ferson and Schadt (1996) as BSW point out that using the Ferson-Schadt conditional model to control for time-varying beta does not alter their results.

Using the information provided by the change point tests, in particular which of the parameters changed and the estimated date of that change, we are able to estimate a multi-regression model that reflects the time-variation. After allowing for time-variation in the regression parameters, the estimated proportion of skilled and unskilled managers increases dramatically for both net and gross returns, relative to those using the traditional model with time-invariant parameters. More specifically, the estimated proportion of skilled funds - funds that have superior selecting skill to generate positive excess returns net of all costs, is 8.4%. When returns are measured before expenses, i.e. fund manager skill is disentangled from the fund's expense policy, more funds are found to be skilled. This overturns several results of BSW who find that only 0.6% of the funds in their sample are skilled.

This chapter also evaluates performance of funds within different investment styles. These styles are classified as aggressive growth, growth, and growth & income. Since these different styles of funds employ different strategies and take different levels of risk, it is important to know the proportion of funds that are skilled which is compared across the different styles. According to our results, aggressive growth funds have the highest proportion of skilled fund managers net of all costs (12.8%), and growth & income funds have the highest proportion of skilled fund managers before expenses (22.1%).

The rest of the chapter is arranged as follows: Section 5.2 reviews past studies that have used the FDR technique. Section 5.3 describes the categories classified based on fund performance and investment styles. Section 5.4 describes the mutual fund data used in this chapter. Section 5.5 briefly explains the FDR approach. Section 5.6 reviews the model with time-varying alpha and betas introduced in more detail in Chapter 4 Section 4.3.2. In Section 5.7, we present the results and robustness checks. The final section concludes.

5.2 Review of Studies on the FDR Technique

In statistical testing environments, often a number of hypotheses are tested simultaneously, each with a given type I error rate. In such cases, the problem of controlling the experiment-wise error rate is more complicated. The traditional approach is to control the family-wise error rate (FWER) when testing many hypotheses; FWER is defined as the probability of incorrectly rejecting at least one of the null hypotheses. However, FWER is often criticized for being too stringent and conservative since it controls the probability of falsely rejecting the null hypothesis in at least one of the tests. Because of this, the probability of falsely rejecting an individual test can be too small so that none of the individual hypotheses are rejected.

In pioneering work, Benjamini and Hochberg (1995) (BH) proposed the false discovery rate (FDR) as an alternative measure to be controlled; it can be interpreted as the expected proportion of falsely rejected null hypotheses among all rejections. The controlling procedure of Benjamini and Hochberg (1995) involves testing all the null hypotheses and sorting their corresponding p-values in an increasing order, and these ordered p-values are generally written as $P_1, P_2, ..., P_m$ ($P_1 \leq P_2 \leq ... \leq P_m$), with m denoting the number of hypotheses. Then, the BH procedure goes from a small p-value to a large one until the largest p-value (P_k) is found, such that $P_k \leq \frac{k}{m} q^*$, hence FDR is controlled at q^*. They show via a simulation that controlling the FDR can have substantial power gains when there are a large number of hypotheses to test. Because of these gains, controlling the FDR can be superior to controlling the FWER.

After that, the FDR approach has been developed further. Storey (2003) introduces the positive false discovery rate (pFDR), with the term "positive" relating to the fact that it conditions on the presence of at least one positive finding. pFDR has a simple form making it tractable in applications. To overcome some weaknesses of the BH procedure, Storey (2002) proposes the use of a fixed rejection region when estimating the error rate. He argues that the BH procedure requires the estimation of $\hat{k}$ in the P_k so that the procedure can perform badly in practice if the estimate of $\hat{k}$ becomes variable, i.e. the reliability of $\hat{k}$ is in doubt. Furthermore, the BH procedure does not use any information on the number of true null hypotheses in the data, which can be estimated from the observed p-values. Storey (2002) proposes a measure that fixes the rejection region (k) and then estimates the significance level. Storey (2002) has shown that doing so produces well-behaved estimates for both FDR and pFDR. Storey et al. (2004) further prove that the BH procedure of Benjamini and Hochberg (1995) and the fixed rejection region approach of Storey (2002) are essentially equivalent under both finite sample and asymptotic settings.

Generally, the measures listed above require independent test statistics, which is not often the case in practice. Hence, several studies explored the asymptotics of FDR under dependent test statistics. These studies have proved that FDR can be used under some specific dependence structures, such as positive regression dependency in which test statistics are positively dependent (Benjamini and Yekutieli, 2001), and weak dependence where test statistics can be dependent in types such as dependence in finite blocks or mixed distributions (Storey et al., 2004), many of which are not covered in positive regression dependency. Clarke and Hall (2009) demonstrated that FDR controlled used dependency is asymptotically equivalent to that under independence when the null distributions of test statistics are light-tailed, for example, when they approximate normal or t-distributions. These explorations of asymptotics of test statistics under certain dependence structures allow a wider application of the FDR technique in practice.

In the economics literature, FDR has applications in several contexts such as finance as it leads to tests with more power than comparable tests that control family-wise error rates, such as predictive ability of linear structural exchange rate models (McCracken and Sapp, 2005), data mining in stock returns (Bajgrowicz and Scaillet, 2012), connected trading in investment banks (Griffin et al., 2012), and hedge funds (Criton and Scaillet, 2011).

BSW[1] is the first study to apply the FDR technique to examine mutual fund performance. They separate 2076 US mutual funds as three groups based on manager's skill which is classified into skilled, unskilled or zero skilled. They use the FDR method to deal with multiple tests performed on the alphas of the factor models estimated. As argued in the previous paragraph, control of the FDR can lead to a testing procedure that has greater power than those tests that control only the family-wise error rates. After adjusting for false discoveries - false positive and negative - BSW find that 75% of U.S. mutual fund managers have zero skill - on average they produce zero excess return. Only 0.6% of the U.S. mutual fund managers are skilled - able to generate excess returns. The rest are unskilled - generate negative excess returns. They also split their entire sample into several five-year subintervals so as to examine fund performance over the short periods. When they do this, they find that the number of skilled fund managers increases to 2.4%.

Using U.S. mutual fund data, Kim and In (2012) employ the FDR approach to examine volatility timing ability of mutual fund managers. Cuthbertson et al. (2012) and Bredin et al. (2014) follow a similar procedure as BSW, but look for skilled managers of UK investment trusts - this includes open-ended and closed-ended equity trusts. 3.7% of open-end funds and 16% of closed-end UK funds are able to produce an alpha that more than covers their costs. Studies on mutual fund performance in other countries like Germany (Otamendi et al., 2008), China (Tang et al., 2011) and Korea (Suh and Hong, 2011) follow similar procedures as BSW to account for false discoveries.

Based on previous studies, this chapter aims to estimate the proportions of skilled, unskilled or zero-alpha U.S. actively managed fund managers using the FDR approach of

[1]BSW paper first become available on-line as a working paper in 2005.

Storey (2002) and BSW. Differently from BSW, we relax the assumption of time-invariant regression parameters in the factor models applied in previous studies and estimate fund alphas using the model described in Chapter 4, Section 4.3.2 to allow for time-variation in betas as well as in alpha. Furthermore, we use an updated dataset compared to BSW, which is more relevant for today's investors.

5.3 Fund Categories

Following BSW, we split the funds into different categories according to fund performance and funds' investment styles. We explained the difference in interpretation of the results when returns are measured net and gross of managers skill in Chapter 2, Section 2.3.2. For convenience, this interpretation is repeated here. When returns are measured net of expense and transaction costs, the fund alpha estimates tell us whether managers have sufficient skill to cover all costs, whereas the ones based on gross returns disentangle a manager's skill from a fund's expense policy so that we know if managers are able to cover those costs missing from the expense ratios, mainly trading costs. Three groups of fund are then defined based on manager performance as follows:

1. Unskilled funds: funds where managers lack sufficient stock-picking skill to cover all their costs, resulting in a negative alpha ($\alpha < 0$),
2. Zero-alpha funds: funds that have managers with just sufficient stock-picking skill to cover all their costs ($\alpha = 0$),
3. Skilled funds: funds that have managers with stock-picking skill that more than cover all costs ($\alpha > 0$).

When returns are measured gross of expenses, i.e. expenses are added back to returns, performance categories are defined as:

1. Unskilled funds: funds where managers lack sufficient stock-picking skill to cover costs excluded from the expense ratios, resulting in a negative alpha ($\alpha < 0$),
2. Zero-alpha funds: funds with managers that have just sufficient stock-picking skill to cover the costs excluded from the expense ratios ($\alpha = 0$),
3. Skilled funds: funds that have managers with stock-picking skill that results in a positive alpha, more than simply covering the costs excluded from the expense ratios ($\alpha > 0$).

Furthermore, funds can also be divided into different groups according to their self-declared investment styles. It is of interest to perform tests on these different groups of funds so that investors know how many funds generate abnormal performance in different investment styles, and more importantly which category has the highest proportion of skilled or unskilled funds. Previous studies tend to come to different conclusions on the existence of skilled managers in different investment styles. Many studies have shown that growth-oriented funds, especially aggressive growth funds, are able to outperform the passive benchmarks, and aggressive growth funds have the highest proportion of skilled funds among all styles of funds.

Specifically, Grinblatt and Titman (1989) and Elton et al. (1996) show that aggressive growth and growth funds generate, on average, positive returns. Hendricks et al. (1993), Daniel et al. (1997), Wermers (2000), Kosowski et al. (2006) and Huij and Verbeek (2007) find that stock picking skill is more prevalent among growth-oriented funds. However, Becker et al. (1999) argue that growth funds and income funds have no major difference in average returns. Therefore, we revisit this issue by applying the FDR approach used by BSW to estimate the proportions of skilled, unskilled and zero-alpha managers, for three styles of funds: aggressive growth, growth and growth & income. Based on their self-declared investment styles, or key words in their names[2], we group funds into three categories:

1. Aggressive growth funds: funds that aim to achieve the highest capital gains and generally have the greatest risk. They usually hold investments with high growth potential as well as great price volatility.
2. Growth funds: funds that aim to achieve capital appreciation and often have above average risks. By investing in growth stocks instead of those paying dividends, they gain the benefit of large capital increases.
3. Growth & income funds: funds that value both capital growth and dividend income. They invest in a range of products that are able to provide investors income through dividends as well as capital gains.

5.4 Mutual Fund Data

We use monthly return data obtained from CRSP. To focus on the performance of active managers, we include only actively managed mutual funds that invest primarily in U.S. common stocks. More information on how to identify funds that are actively managed and mainly invest in U.S. equities can be found in Appendix A. Differently from the dataset used in Chapter 4, in order to compare our results to those of BSW, we keep a fund in this sample if it has at least 60 months of return history. [3] Hence, we have fewer funds in this chapter and there are 3673 funds in the sample.

[2] Appendix A gives a full explanation on how to identify funds' investment styles.

[3] In Chapter 4, we included a fund in the sample if it has at least 8 months of return history. This sample selection issue will be further elaborated in Chapter 6, Section 6.2.

To examine performance in different investment styles, we split our funds into three categories according to their self-declared investment styles. Our sample of 3673 funds contains 306 aggressive growth funds, 2458 growth funds and 1419 growth & income funds.[4]

5.5 False Discovery Rate

5.5.1 Hypothesis

This study aims to examine how many actively managed U.S. mutual funds outperform or underperform the market, or have zero skill, when compared to passive benchmarks. Previously, to assess the overall performance of the mutual fund industry, studies suggested a procedure that counts the number of observed significant funds, without controlling for luck (Ferson and Schadt, 1996); therefore the FDR is in fact assumed to be zero in those studies. BSW first try to account for false discoveries using the FDR approach in a multiple hypothesis test to examine mutual fund performance. We will briefly explain the FDR methodology here. To assess the performance of a single fund (fund i), the null hypothesis is that the fund has no abnormal performance (a zero-alpha fund), and the alternative is that the fund provides either positive (a skilled fund) or negative abnormal performance (an unskilled fund):

$$H_0 : \alpha_i = 0, H_A : \alpha_i \neq 0; \tag{5.1}$$

[4]The number of funds from all three categories in total is more than the number of funds in the entire sample. This is because funds may change their investment styles during their life. When the entire sample is split into different investment categories. There are some funds that may have parts of their return history in two or even three different investment categories. This has little impact on the interpretation of our results.

In the above case, the type I error (false positive) is controlled at a predetermined significance level. To assess performance in the sample, this test should be applied across all funds simultaneously. Therefore, a joint hypothesis test is established, which is given in Equation 5.2:

$$H_{0,1} : \alpha_i = 0, H_{A,1} : \alpha_i \neq 0, \tag{5.2}$$

$$... : ... \tag{5.3}$$

$$H_{0,N} : \alpha_N = 0, H_{A,N} : \alpha_N \neq 0; \tag{5.4}$$

5.5.2 Computation of Statistics

To account for false discoveries, the FDR approach of Storey (2002) is used here. Appendix 5.A, Section 5.A.1 shows a detailed account of how luck is controlled for in a multiple fund setting using the FDR approach, so we shall be brief here. To begin with, at a given significance level γ, the probability of a true zero-alpha fund exhibiting luck equals $\gamma/2$. Accordingly, the probability of a true zero-alpha fund being unlucky is also $\gamma/2$ [5]. Suppose that the proportion of true zero-alpha funds in the population of N funds is π_0, then the expected proportion of true zero-alpha funds being lucky or unlucky is:

$$E(F_\gamma^+) = E(F_\gamma^-) = \pi_0(\gamma/2); \tag{5.5}$$

Knowing the proportions of funds which are lucky or unlucky, i.e. generate abnormal performance by chance but are truly zero-alpha funds; we are able to estimate the proportion of truly skilled (unskilled) funds by subtracting this proportion of false positives (negatives) from the population of funds. Given a significance level γ, we denote a fund as

[5] Refer to BSW, Figure 1 for detailed explanation.

a "significant fund" if the t-statistic is smaller than the negative threshold (t_γ^-) or greater than the positive threshold (t_γ^+). After estimating the expected proportion of significant funds ($E(S_\gamma^+)$) in the sample, we can determine the expected proportion of skilled funds ($E(T_\gamma^+)$) and unskilled funds ($E(T_\gamma^-)$) by adjusting $E(S_\gamma^+)$ and $E(S_\gamma^-)$ for the existence of lucky and unlucky funds in the following way:

$$E(T_\gamma^+) = E(S_\gamma^+) - E(F_\gamma^+) = E(S_\gamma^+) - \pi_0 \times \gamma/2, \tag{5.6}$$

$$E(T_\gamma^-) = E(S_\gamma^-) - E(F_\gamma^-) = E(S_\gamma^-) - \pi_0 \times \gamma/2; \tag{5.7}$$

The variable of choice here is the significance level γ. Using different levels of γ allows us to see how the proportion of truly skilled or unskilled funds increases with γ. For instance, if the expected proportion of skilled funds ($E(T_\gamma^-)$) increases slightly as we change the value of γ from 0.1 to 0.2, then most of the skilled funds should lie in the extreme right tail since the majority of the additional significant funds are merely lucky. However, if skilled funds are dispersed randomly throughout the right tail, then we would expect to see a more significant increase in $E(T_\gamma^-)$ as γ increases.

5.5.3 Estimation of π_0

To compute the above statistics, the key is to estimate the proportion of truly zero-alpha funds (π_0) in the population of N funds. To calculate π_0, we use the fact that the truly skilled or unskilled funds should have small p-values (alpha) since their estimated t-statistics tend to be more extreme, i.e. far from zero (the blue area in Figure 5.1). Alternatively, p-values of those truly zero-alpha funds should be uniformly distributed over [0,1] (cf. BSW page 187, footnote 8). This fact allows us to estimate π_0 without

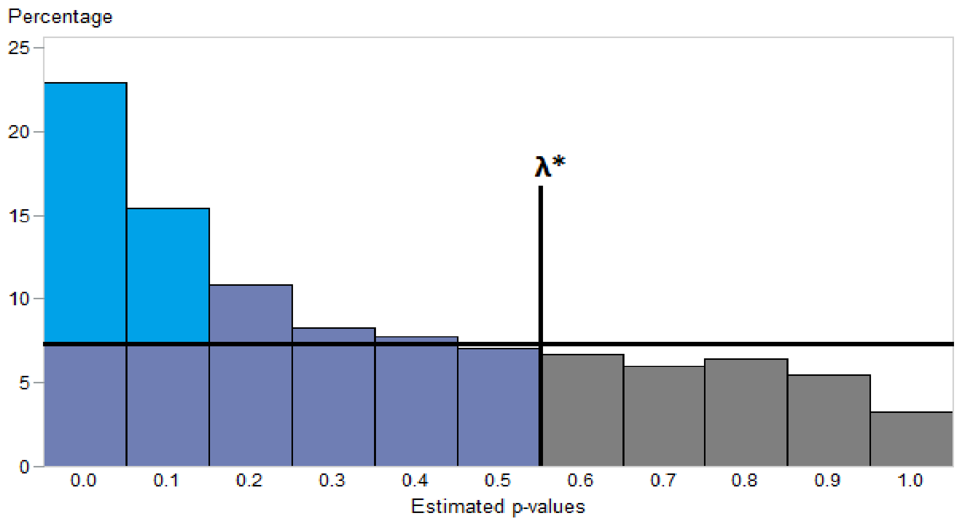

Figure 5.1: Histogram of Fund p-values

knowing how p-values of skilled and unskilled funds are distributed. Therefore, π_0 is estimated as follows:

$$\hat{\pi}_0(\lambda^*) = \frac{\hat{W}(\lambda^*)}{N} \times \frac{1}{(1-\lambda^*)}; \tag{5.8}$$

Where λ^* is a threshold (optimal λ to be determined later), after which the histogram of p-values as shown in Figure 5.1 becomes flat reflecting the fact that the p-values greater than the threshold (the grey area) should all come from the group of truly zero-alpha funds. $\frac{\hat{W}(\lambda^*)}{N}$ is the part of the histogram after the threshold λ^*. To extrapolate the area $(\frac{\hat{W}(\lambda^*)}{N})$ over the entire interval from zero to one, we divide it by $(1-\lambda^*)$. To estimate π_0, we have to compute the p-value for each fund first. In fact, p-values of funds are the only inputs to estimate π_0. Given that the t-statistic is more likely to be asymmetric in our case, as shown in the CDF distribution of t-statistic in Chapter 4, Table 4.3, a bootstrap procedure is used to compute the fund p-values. Specifically, we compute fund p-values following the approach proposed by Davidson and MacKinnon (2004):

$$\widehat{p_i} = 2 \times min\left(\frac{1}{B}\sum_{b=1}^{B} I\left\{\widehat{t_i^{*b}} > \widehat{t_i}\right\}, \frac{1}{B}\sum_{b=1}^{B} I\left\{\widehat{t_i^{*b}} < \widehat{t_i}\right\}\right); \tag{5.9}$$

Where B is the number of bootstrap runs (B=1000), and $\left\{\widehat{t_i^{*b}} > \widehat{t_i}\right\}$ takes the value one if the estimated t-statistic $(\widehat{t_i})$ is lower than the bootstrapped t-statistic $(\widehat{t_i^{*b}})$.

After having the bootstrapped p-values, we implement the bootstrap approach proposed by Storey (2002) to choose the optimal λ and then to estimate the proportion of truly zero-alpha funds in the population using Equation 5.8. To select the optimal λ, first, $\widehat{\pi_0(\lambda)}$ is computed for each of the λ values (λ=0.30, 0.35, 0.40, ..., 0.70). Then, we draw with replacement from the $N \times 1$ vector of all fund p-values to form 1000 bootstrapped samples. For each bootstrapped sample, the bootstrapped $\widehat{\pi_0^b(\lambda)}$ (b=1,2,...,1000) is computed. Finally, the optimal λ (λ^*) is chosen to minimize the estimated Mean Squared Error (MSE), which is as follows:

$$\widehat{MSE}(\lambda) = \frac{1}{1000}\sum_{b=1}^{1000}\left[\widehat{\pi}_0^b(\lambda) - \underset{\lambda}{min}\widehat{\pi}_0(\lambda)\right]^2; \tag{5.10}$$

Now that we have the estimated value of π_0, all the statistics listed below can be computed:

$$\hat{F}_\gamma^+ = \hat{F}_\gamma^- = \hat{\pi}_0(\gamma/2); \tag{5.11}$$

Then the estimated proportion of skilled and unskilled funds with respect to the significance level γ is:

$$\hat{T}_\gamma^+ = \hat{S}_\gamma^+ - \hat{F}_\gamma^+ = \hat{S}_\gamma^+ - \hat{\pi}_0 \times \gamma/2, \tag{5.12}$$

$$\hat{T}_\gamma^- = \hat{S}_\gamma^- - \hat{F}_\gamma^- = \hat{S}_\gamma^- - \hat{\pi}_0 \times \gamma/2; \tag{5.13}$$

Finally, we have the estimated proportion of skilled and unskilled funds in the population of N funds as:

$$\hat{\pi}_A^+ = \hat{T}_{\gamma^*}^+, \tag{5.14}$$

$$\hat{\pi}_A^- = \hat{T}_{\gamma^*}^-; \tag{5.15}$$

Where γ^* is the optimal significance level that has the minimum MSE of $\hat{\pi}_A^+$ and $\hat{\pi}_A^-$. To choose the optimal level of γ, a bootstrap approach similar to the process of choosing optimal λ is used. First, for a range of γ values ($\gamma = 0.05, 0.10, ..., 0.50$), we compute the proportion of skilled funds ($\hat{\pi}_A^+(\gamma)$). Then, we resample with replacement from the $N \times 1$ vector of p-values to produce 1000 bootstrap replications of $\hat{\pi}_A^+(\gamma)$. Finally, the estimated MSE of $\hat{\pi}_A^+(\gamma)$ is computed as :

$$\widehat{MSE}^+(\gamma) = \frac{1}{1000} \sum_{b=1}^{1000} \left[\widehat{\pi}_A^{b+}(\gamma) - \underset{\gamma}{max} \widehat{\pi}_A^+(\gamma) \right]^2; \tag{5.16}$$

We choose optimal γ^+ to minimize $\widehat{MSE}^+(\gamma)$. Accordingly, we also compute $\widehat{MSE}^-(\gamma)$ for each significance level γ. If $min_\gamma \widehat{MSE}^+ \gamma < min_\gamma \widehat{MSE}^-(\gamma)$, then we choose γ to minimize $\widehat{MSE}^+(\gamma)$. To preserve the equality ($1 = \pi_0 + \pi_A^+ + \pi_A^-$), $\widehat{\pi}_A^-(\gamma^*) = 1 - \widehat{\pi}_0 - \widehat{\pi}_A^+(\gamma^*)$. Otherwise, we choose the optimal γ that minimizes $\widehat{MSE}^-(\gamma)$ and set $\widehat{\pi}_A^+(\gamma^*) = 1 - \widehat{\pi}_0 - \widehat{\pi}_A^-(\gamma^*)$. The computation of standard errors of the above mentioned estimates is shown in Appendix 5.A, Section 5.A.2.

5.6 The Model Framework with Time-varying Parameters

In Section 5.5, we have shown that the only inputs to estimate π_0, the proportion of truly zero-alpha funds, are the bootstrapped individual fund p-values. To consistently estimate fund alphas, we apply the change point test of Pouliot (2016) to detect time-variation in the parameters of the factor models considered here. Then we construct a multi-regression model that essentially involves estimating different 4F-CAPMs over the subsamples formed from the original time-series and divided according to the dates of the changes in alpha and betas. Chapter 4, Section 4.3 provided a detailed account of the multi-regression model, so we will be brief here.

Since we have shown in Chapter 4, Section 4.5.2 that different versions of the CAPM produce similar fund alpha estimates, we use the 4F-CAPM here as the baseline model to perform the change point test. For convenience, the functional form of the 4F-CAPM is repeated here:

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i} SMB_t + \beta_{3i} HML_t + \beta_{4i} MOM_t + \varepsilon_{i,t}, \qquad (5.17)$$

where $R_{i,t}$ is the return for month t of fund i, R_t^f is the risk-free rate, which is the 1-month U.S. treasury bill rate, R_t^m is the market return constructed with a VW portfolio of NASDAQ, NYSE and Amex stocks, SMB_t and HML_t are the size and value-growth returns constructed following Fama and French (1993), MOM_t is the momentum return of Carhart (1997), α_i (the estimate of α_i) is the return left unexplained by this benchmark model, i.e. the extra return generated by funds and $\varepsilon_{i,t}$ is the regression residual. All the benchmark factors are obtained from Kenneth French's website.[6]

[6] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Details of how this test is implemented and why this test is useful in our context are described in Chapter 4, Section 4.3.2. After estimating the 4F-CAPM on each of the funds in the sample, the change-point method of Pouliot (2016) is used to test for changes in the alpha and betas. If a change in one or more of these parameters is detected then the date of these changes are estimated. Using information on the dates of change, different versions of the 4F-CAPM are estimated on the fund that reflect one or more changes in these parameters as shown in Chapter 4, Section 4.3.2.

It should be noted that BSW examine both the long-term and short-term performance of mutual funds. Specifically, they estimate fund p-values over the entire sample period to evaluate the long-term performance, and then the sample is further split into several five-year sub-intervals to examine performance over the short run. Each of these five-year sub-intervals is then treated as a separate fund record in their study. As a result, they find higher proportions of skilled funds over the short run than over the long run. The rationale of this approach to examine short-term performance is similar to our multi-regression model. However, according to Kon and Jen (1978), any attempt to partition the time-series into successive intervals would impose the prior assumption that parameters are constant for such subintervals. We divide sample series according to the estimated dates of the change points inherently existing in the data, rather than set them arbitrarily at a five-year span, which is more reasonable.

5.7 Results: Time-varying vs. Time-invariant Parameters

5.7.1 Results of Change Points Tests

We begin the empirical analysis by running the change point test on each fund in the sample. Generally, we find no more than three changes in either alpha or betas over the entire sample period, and no more than five changes in all parameters. The summary statistics of results of the tests for change points are shown in Table 5.1 and Table 5.2.

Table 5.1: Summary of Results on Tests for Change Points in the Slopes and the Intercept: 1984-2015 (4F-CAPM)

	First test	Forward tests			Backward tests		
		1st	2nd	3rd	1st	2nd	3rd
Net returns							
Number of funds							
Change in intercept & slopes	307	45	4	0	12	3	0
Change in slopes	204	57	14	1	36	2	0
Change in intercept	302	127	32	7	56	11	2
No change	2860	3444	3623	3665	3569	3657	3671
Proportions							
Change in intercept & slopes	8.0	1.0	0.0	0.0	0.0	0.0	0.0
Change in slopes	6.0	2.0	0.0	0.0	1.0	0.0	0.0
Change in intercept	8.0	3.0	1.0	0.0	2.0	0.0	0.0
No change	78.0	94.0	99.0	100.0	97.0	100.0	100.0
Gross returns							
Number of funds							
Change in intercept & slopes	317	52	5	0	15	2	0
Change in slopes	202	52	12	1	37	3	0
Change in intercept	320	133	32	8	52	11	1
No change	2834	3436	3624	3664	3569	3657	3672
Proportions							
Change in intercept & slopes	9.0	1.0	0.0	0.0	0.0	0.0	0.0
Change in slopes	5.0	1.0	0.0	0.0	1.0	0.0	0.0
Change in intercept	9.0	4.0	1.0	0.0	1.0	0.0	0.0
No change	77.0	94.0	99.0	100.0	97.0	100.0	100.0

From Table 5.1, around 22% of the total funds in the sample have at least one parameter change for the net returns. Tests for gross returns have a similar number of change points. This is not surprising since gross returns are returns with expense ratios added back to net returns. As long as there are not many changes in the expense policy of mutual funds, we should see a similar number of change points in these two return series.

As we run the test again, both forward and backward from the change point, we find further changes in the parameters. For example, around 6% of the total funds are found to have further parameter change when we run the test on the subsample prior to the first change, and around 3% of the total funds have further parameter change when the test is run on the subsample after the first change. Finally, no more than three changes are found in either alpha or betas as we continue to run the tests until no more changes are found. Next, we show the number of funds in different scenarios with different combinations of parameter changes in Table 5.2. For example, for net returns, there are 276 funds with only one change in alpha and one in the betas, and they account for 7.5% of all the funds in the sample.

Table 5.2: Statistics on Different Combinations of Change Points: 1984 to 2015 (4F-CAPM)

Different cases of combinations	Net returns		Gross returns	
	funds	Per(%)	funds	Per(%)
Total number of funds	3673		3673	
No change	2663	72.5	2639	71.8
1 change in both βs and α	276	7.5	289	7.9
1 change in βs and no change in α	216	5.9	210	5.7
1 change in α and no change in βs	358	9.7	373	10.2
2 changes in βs and no change in α	15	0.4	14	0.4
2 changes in βs and 1 change in α	24	0.7	22	0.6
2 changes in βs and 2 changes in α	22	0.6	25	0.7
1 change in βs and 2 changes in α	46	1.3	50	1.4
0 change in βs and 2 changes in α	39	1.1	36	1.0
3 changes in βs and 0 change in α	1	0.0	1	0.0
3 changes in βs and 1 change in α	1	0.0	1	0.0
3 changes in βs and 2 changes in α	1	0.0	0	0.0
3 changes in βs and 3 change in α	1	0.0	2	0.1
2 changes in βs and 3 changes in α	4	0.1	2	0.1
1 change in βs and 3 changes in α	5	0.1	6	0.2
0 change in βs and 3 changes in α	1	0.0	1	0.0

After we have the results of the change points test, we split the data according to the estimated dates of the change points and run the separate models based on different scenarios listed in Chapter 4, Section 4.3.2 to estimate fund alphas (t-statistics). After that, p-values of individual funds are computed using Equation 5.9. Once the p-values are calculated, we estimate the proportion of truly zero-alpha funds (π_0) and further compute all the rest statistics ($\hat{F}_\gamma^-$, $\hat{F}_\gamma^+$, $\hat{T}_\gamma^-$ and $\hat{T}_\gamma^+$). Results based on net returns are shown in Table 5.3 and those on gross returns are presented in Table 5.4.

5.7.2 Net Returns

As shown in Table 5.3, in the upper block of Panel A, we show that for the traditional 4F-CAPM, over the entire life of funds, 67.2% of the funds are estimated to be zero-alpha funds. These funds deliver skill just sufficient to cover all the costs. Furthermore, we find 32.0% of funds are unskilled funds, which means this proportion of funds are unable to produce expected returns to cover their costs, echoing the results in BSW that 24% of funds are truly unskilled funds. BSW find it puzzling that a large proportion of negative-alpha funds exists, and that their underperformance does persist for a long time period, as investors should have enough time to identify them as underperforming funds. Therefore, BSW argue that these investors either lack sufficient information to know they have been overcharged or they have been prohibited from switching to other funds. Finally, 0.8% of funds in the sample are skilled funds.

Table 5.3:
Proportions of Skilled, Unskilled, and Zero-alpha Funds: 1984-2015 (Net Returns, Time-varying vs. Time-invariant Parameters)
Performance of funds is estimated using the 4F-CAPM over the entire sample period from January 1984 to March 2015. Panel A shows the estimated proportions of zero-alpha ($\hat{\pi}_0$), unskilled ($\hat{\pi}_A^-$) and skilled ($\hat{\pi}_A^+$) funds using the FDR approach based on the entire sample of 3673 funds and on net returns. Panel A also presents the detailed results including significant funds ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$) across the distribution of funds in the left tail and right tail using four different significance levels ($\gamma = 0.05, 0.1, 0.15, 0.2$). Furthermore, significant funds are decomposed into unlucky ($\hat{F}_\gamma^-$) and unskilled ($\hat{T}_\gamma^-$) funds in the left tail, and lucky ($\hat{F}_\gamma^+$) and skilled ($\hat{T}_\gamma^+$) funds in the right tail. At the bottom of Panel A, we show the FDR in the left and right tail at different significance levels. Panel B shows results using the multi-regression model with time-varying regression parameters.

Panel A: Model with time-invariant parameters

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportions	67.2 (0.02)	32.0 (0.02)	0.8 (0.00)
Number of funds	2469	1176	28

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	14.0	19.5	24.6	28.5	7.3	5.9	4.3	2.5	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.7	3.4	5.0	6.7	6.7	5.0	3.4	1.7	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	12.3	16.1	19.6	21.8	0.6	0.9	1.0	0.8	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
FDR$^-$	12.0	17.2	20.5	23.6	91.8	85.3	77.6	68.6	FDR$^+$

Panel B: Model with time-varying parameters

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled($\hat{\pi}_A^+$)
Proportions	57.5 (0.02)	34.1 (0.02)	8.4 (0.01)
Number of "funds"	2726	1618	399

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	19.0	24.5	29.1	32.6	14.2	12.5	10.5	7.9	Signif.$\hat{S}_\gamma^+$(%)
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.4	2.9	4.3	5.7	5.7	4.3	2.9	1.4	Lucky $\hat{F}_\gamma^+$(%)
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	17.6	21.6	24.7	26.8	8.4	8.2	7.7	6.5	Skilled $\hat{T}_\gamma^+$(%)
	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	
FDR$^-$	7.5	11.7	14.8	17.7	40.6	34.5	27.3	18.1	FDR$^+$

In the lower block of Panel A, as we switch to different significance levels ($\gamma = 0.05, 0.10, 0.15, 0.20$), we show the proportion of significant funds in the left and right tails ($\hat{S}_{\gamma}^{-}$ and $\hat{S}_{\gamma}^{+}$), funds being unlucky and lucky ($\hat{F}_{\gamma}^{-}$ and $\hat{F}_{\gamma}^{+}$), and funds that are truly unskilled or skilled ($\hat{T}_{\gamma}^{-}$ and $\hat{T}_{\gamma}^{+}$). Consistent with BSW, we have found a significant proportion of significant alpha funds in the right tail, with the highest level of 7.3% when γ is 0.20. However, most of these significant alpha funds are not truly skilled, but merely lucky to have extreme returns. This leads to the observation that only a small proportion of funds are found to be truly skilled and this proportion of skilled funds does not change much for different significance levels, ranging from 0.6% to 1.0%. Therefore, most of the significant alpha funds in the right tail are truly zero-alpha funds who appear to be lucky and none of these funds present skill to outperform the market given that $\hat{T}_{\gamma}^{+}$ is not significantly different from zero at any significance level. As for the left tails, the proportion of unskilled funds increases as we switch to higher significance levels.

Generally, all the above results based on the traditional 4F-CAPM are consistent with BSW. However, we have shown in Chapter 4 that the traditional model is not valid and may lead to biased conclusions. In fact, results change dramatically as we switch to the model that allows for time-variation in betas and alpha. After incorporating the change points of the regression parameters into the model, the proportion of the truly zero-alpha funds decreases to 57.5%, with fewer funds identified as zero-alpha funds whose managers present skill just enough to cover all the costs. Instead, we find more unskilled and skilled funds, 34.1% and 8.4% respectively.

Our results confirm that our multi-regression model with time-varying parameters is able to separate parts of a fund's history when managers outperform relative to parts of its history when it performs badly, or, it allows the evaluation of fund managers performance over the short run before investors compete away their performance with capital inflows. Based on the results from the multi-regression model, we argue that there is indeed a significant proportion of skilled fund managers in the population. Therefore, prior results based on the traditional models with time-invariant parameters are biased towards more pessimistic conclusions, i.e. no managers are able to deliver expected returns that outperform the market. Specifically, when we decompose the significant alpha funds into lucky and skilled funds, as shown in the lower block of Panel B, the proportion of skilled funds increases slowly as we move to higher significance levels. This suggests that most of the skilled funds should locate in the extreme right tail since the majority of the additional significant alpha funds are lucky funds as γ is increased from 0.10 to 0.20.

5.7.3 Gross Returns

For net returns, a fund is identified as a skilled fund if the manager is able to generate a positive alpha net of all costs. Since parts of mutual fund costs are reflected in the expense ratios (advisory fees, administrative costs, other operating expenses, etc.), the rest of fund costs missing from the expense ratios are mainly trading costs. Given that mutual fund managers may have little control of a fund's expense policy, it seems more interesting to examine the results before expenses are deducted so that fund managers' stock picking skill can be distinguished from funds' expense policy. To achieve that, we estimate the proportions of skilled, unskilled and zero-alpha funds using the FDR technique using gross returns, which are created by adding back the monthly expenses to the net returns. Generally, results for gross returns are more promising.

As shown in Table 5.4, the estimated proportion of zero-alpha funds is 80.3%, higher than that of net returns, which is reasonable given that fund managers are only required to cover costs missed in the expense ratios to achieve zero alphas for gross returns. Further, the estimated proportion of skilled funds increases to the level of 12.6% compared to 0.8% based on net returns. This suggests that the pessimistic results based on net returns are mainly due to high fees charged by funds. Some fund managers may be able to deliver positive alphas to more than compensate for costs missed in the expense ratios, mainly trading costs, but the surplus is offset by the high expenses of fund companies. Using gross returns, we also see a sharp decrease of the proportion of unskilled funds (7.1%), which again echoes our argument that expenses are the main reason why we observe a large group of unskilled fund managers (32.0%) in terms of net returns.

When we incorporate change points into the 4F-CAPM to allow for time-variation in the alpha and betas, the estimated proportions of skilled and unskilled funds increase compared to those using the 4F-CAPM with constant alpha and betas. Specifically, the skilled funds account for 19.8% of the entire sample and the proportion of unskilled funds increases to 11.9% in this case. Additionally, the decomposed results at different significance levels show that the estimated proportions of both skilled and unskilled funds increase slowly as the value of γ increases. Therefore, both skilled funds and unskilled funds should be located in the extreme tails, which is consistent with the findings on net returns.

Table 5.4:
Proportions of Skilled, Unskilled, and Zero-alpha Funds: 1984-2015 (Gross Returns, Time-varying vs. Time-invariant Parameters)

Performance of funds is measured using the 4F-CAPM over the entire sample period from January 1984 to March 2015. Panel A shows the estimated proportions of zero-alpha ($\hat{\pi}_0$), unskilled ($\hat{\pi}_A^-$) and skilled ($\hat{\pi}_A^+$) funds using the FDR approach based on the entire sample of 3673 funds and on gross returns. Panel A also presents the detailed results including significant funds ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$) across the distribution of funds in the left tail and right tail using four different significance levels ($\gamma = 0.05, 0.1, 0.15, 0.2$). Furthermore, significant funds are decomposed into unlucky ($\hat{F}_\gamma^-$) and unskilled ($\hat{T}_\gamma^-$) funds in the left tail, and lucky ($\hat{F}_\gamma^+$) and skilled ($\hat{T}_\gamma^+$) funds in the right tail. At the bottom of Panel A, we show the FDR in the left and right tail at different significance levels. Panel B shows results using the multi-regression model with time-varying regression parameters.

Panel A: Model with time-invariant parameters

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)	
Proportion	80.3 (0.01)	7.1 (0.01)	12.6 (0.01)	
Number of funds	2951	261	461	

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	4.1	6.6	9.3	11.1	19.6	16.1	12.9	8.3	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	2.0	4.0	6.0	8.0	8.0	6.0	4.0	2.0	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	2.1	2.5	3.2	3.1	11.6	10.1	8.9	6.3	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	
FDR^-	48.9	61.2	65.1	72.1	40.9	37.3	31.1	24.1	FDR^+

Panel B: Model with time-varying parameters

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled($\hat{\pi}_A^+$)	
Proportion	68.3 (0.01)	11.9 (0.01)	19.8 (0.01)	
Number of "funds"	3258	568	947	

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	9.0	12.3	15.1	17.6	24.6	21.8	18.7	14.5	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	
Unlucky $\hat{F}_\gamma^-$(%)	1.7	3.4	5.1	6.8	6.8	5.1	3.4	1.7	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	7.3	8.8	10.0	10.7	17.7	16.7	15.3	12.8	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	
FDR^-	19.0	27.8	33.9	38.9	27.8	23.5	18.2	11.8	FDR^+

5.7.4 Performance for Different Investment Styles of Funds

Mutual fund managers usually have different investment styles that can best suit the investors' objective. We categorize funds in the sample into three investment categories: aggressive growth, growth and growth & income funds. Aggressive growth funds usually take the highest risk to achieve the highest capital gains and tend to be rewarded with more extreme returns. More moderate than aggressive growth funds, growth funds generally have above average risk and return volatility that targets fast capital appreciation. In the last category, growth & income funds pursue a dual strategy of income generation and capital gain. We divide funds into the three above mentioned investment styles to examine whether different investment styles of funds differ in their performance.

First of all, we run the three versions of the CAPM on equal and value weighted portfolios of the three categories of funds; results are shown in Table 5.5. We can see that for both EW and VW portfolios, aggressive growth funds have the lowest value of estimated fund alpha on net returns, -1.69 and -1.86 respectively, and growth & income funds have the highest estimated alpha. However, estimated fund alphas on net returns for all categories, for both EW and VW portfolios are negative, indicating that none of these categories are able to beat the passive benchmarks after all costs. For pre-expense returns, growth and growth & income funds on average are able to deliver positive alphas, for both EW and VW portfolios, ranging from 0.3 to 0.86; however, the estimated coefficients are insignificantly different from zero.

As to the factor loadings, aggressive growth funds have the highest exposures to the size factor compared to the other styles. This is because aggressive growth funds target high capital appreciation and tend to invest in small cap companies that have the highest potential growth. We see the highest exposures to the value (book-to-market) factor for growth & income funds, which is consistent with the fact that growth & income funds target both capital gain and income generation, and invest mainly in value stocks.

Next, we turn to individual funds to estimate the proportion of skilled and unskilled funds in each investment style. When the traditional 4F-CAPM is estimated, as shown in Table 5.6, in terms of net returns received by investors, aggressive growth funds have the highest proportion of skilled funds (6.3%), which is consistent with previous literature. Growth funds show similar results to the entire population of funds: 64.5% are truly zero-alpha funds, 34.6% are unskilled funds and only 0.9% of funds are skilled. It is surprising that we find a higher proportion of skilled funds in growth & income funds than in the growth funds category, which is different from the findings in BSW where no growth & income funds are skilled. We argue that the results are reasonable given that fees are the main reason to attribute to fund underperfomance, and growth & income funds usually charge the lowest fees [7].

When we switch to pre-expense returns, the number of zero-alpha and skilled funds for all three categories increases and we see fewer unskilled funds. Growth & income funds have the highest proportion of skilled funds (17.8%) and aggressive growth funds have the lowest level (11.0%). This is reasonable due to the fact that aggressive growth funds aim to achieve the highest capital gains and usually have the highest turnover (Grinblatt and Titman, 1989; Wermers, 1999). Therefore, for pre-expense returns, when fund managers are required to only cover trading costs to achieve zero-alpha, aggressive growth funds could perform worse than the other two styles of funds.

After controlling for time-variation in the regression parameters, i.e. the change point test is undertaken and the multi-regression model is estimated, the proportions of skilled

[7] This is reflected in Figure A.7.

and unskilled funds increase for all three categories of funds and we see fewer zero-alpha funds. Specifically, in terms of net returns, aggressive growth funds have the highest proportion of skilled funds (12.8%). The proportions of skilled funds for growth funds and growth & income funds are 8.8% and 10.0%, respectively. When returns are measured before expenses, growth & income funds, rather than aggressive growth funds, have the highest proportion of skilled funds (22.1%), which can be explained by the fact that aggressive growth funds tend to have higher transaction costs (higher turnover) than the other two styles of funds.

Generally, our results confirm that aggressive growth funds have the highest proportion of skilled managers net of all costs, which is consistent with the literature (Grinblatt and Titman, 1989; Hendricks et al., 1993; Daniel et al., 1997). However, we find growth & income funds outperform the other two categories of funds when returns are measured before expenses, i.e. when fund performance is disentangled from its expense policy. This is contrary to the results of BSW, who argue that no growth & income funds are skilled. We believe that our results are reasonable given that growth funds usually have more turnover, thus more transaction costs than income-generating funds. Therefore, when returns are measured after costs missed in expense ratios, mainly transaction costs, i.e. fund managers are required to cover trading costs only to achieve zero-alpha, growth & income funds are very likely to perform better than those growth-oriented funds.

Table 5.5:
Intercepts and Slopes in the 4F-CAPM for Equal-Weight (EW) and Value-Weight (VW) Portfolios of Actively Managed Mutual Funds (1984-2015, Different Investment Styles)

This table provides the annualized intercepts (12*$\hat{\alpha}$) and t-statistics for the intercepts ($t(Coef)$) for the 4F-CAPM estimated on an equal-weight (EW) and value-weight (VW) portfolio, and on net and gross returns based on three panels of actively managed mutual funds that invest mainly in equities. Regressions slopes ($\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$) for the corresponding factors ($RM - Rf$, SMB, HML, and MOM) are also shown in the table. It should be noted that for the market slope, $t(Coef)$ tests the null hypothesis that β is equal to 1. Net returns are returns reported in the CRSP mutual fund database, and the gross returns are net returns plus $1/12^{th}$ of a fund's expense ratio at that year end. When a fund's expense ratio for a year is missing, we replace it with the expense ratio of the fund that has the same investment style and has similar assets under management (AUM). As for those funds with zero expense ratios, we manually check some of their annual reports and confirm that they in fact have positive expense ratios. Regression results are shown for funds in different categories, namely aggressive growth, growth and growth & income funds.

	12*$\hat{\alpha}$ (%)						
	Net	Gross	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2
EW Returns							
Aggressive growth funds							
4F-CAPM $Coef$	-1.69	-0.32	1.03	0.45	-0.18	0.06	0.96
$t(Coef)$	-2.58	-0.48	2.7	24.89	-9.09	4.78	
Growth funds							
4F-CAPM $Coef$	-0.56	0.79	0.98	0.26	-0.08	0.02	0.98
$t(Coef)$	-1.29	1.82	-2.48	21.13	-6.27	2.97	
Growth & income funds							
4F-CAPM $Coef$	-0.33	0.86	0.91	0.07	0.25	-0.04	0.98
$t(Coef)$	-0.77	1.99	-9.81	6.08	20.21	-4.85	
VW Returns							
Aggressive growth funds							
4F-CAPM $Coef$	-1.86	-0.80	1.06	0.38	-0.26	0.05	0.95
$t(Coef)$	-2.36	-1.01	3.83	17.53	-11.09	3.21	
Growth funds							
4F-CAPM $Coef$	-0.72	0.30	0.98	0.14	-0.11	0.02	0.98
$t(Coef)$	-1.57	0.64	-1.99	11.16	-8.19	2.54	
Growth & income funds							
4F-CAPM $Coef$	-0.50	0.37	0.90	-0.05	0.25	-0.04	0.98
$t(Coef)$	-1.31	0.97	-12.3	-4.54	22.62	-6.00	

Table 5.6:
Proportions of Skilled, Unskilled, and Zero-alpha Funds: 1984-2015 (Different Investment Styles, Time-varying vs. Time-invariant Parameters)
Performance of funds is measured using the 4F-CAPM with time-invariant parameters and the multi-regression model with time-varying parameters, over the entire sample period from January 1984 to March 2015. The table shows the estimated proportions of zero-alpha ($\hat{\pi}_0$), unskilled ($\hat{\pi}_A^-$) and skilled ($\hat{\pi}_A^+$) funds on net and gross returns using the FDR approach for each category of funds, namely aggressive growth, growth, growth & income funds.

Investment styles	Net returns						Gross returns					
	Time-invariant parameters			Time-varying parameters			Time-invariant parameters			Time-varying parameters		
	$\hat{\pi}_0$	$\hat{\pi}_A^-$	$\hat{\pi}_A^+$	$\hat{\pi}_0$	$\hat{\pi}_A^-$	$\hat{\pi}_A^+$	$\hat{\pi}_0$	$\hat{\pi}_A^-$	$\hat{\pi}_A^+$	$\hat{\pi}_0$	$\hat{\pi}_A^-$	$\hat{\pi}_A^+$
Aggressive Growth												
Proportion	68.1	25.6	6.3	59.6	27.6	12.8	73.7	15.3	11.0	59.6	23.1	17.3
Number of funds	209	78	19	217	100	47	226	47	33	217	84	63
Growth												
Proportion	64.5	34.6	0.9	58.1	33.1	8.8	85.1	3.0	11.9	72.9	11.0	16.1
Number of funds	1586	851	21	1809	1030	274	2091	74	293	2287	315	505
Growth & Income												
Proportion	66.7	29.9	3.4	59.9	30.1	10.0	80.2	2.0	17.8	70.0	7.9	22.1
Number of funds	947	424	48	1060	533	177	1138	28	253	1247	140	394

5.7.5 Robustness Checks

In order to test if our results are robust to different versions of the CAPM that are used to estimate fund alphas, we also estimate the proportion of skilled, unskilled and zero-alpha funds based on the 3F-CAPM. The results are shown in Table 5.7. As we can see, generally, the results based on the 3F-CAPM are similar to those of the 4F-CAPM. For instance, in terms of net returns received by investors, the estimated proportions of skilled, unskilled and zero-alpha funds for the 3F-CAPM are 9.1%, 33.0% and 57.9%, similar to those based on the 4F-CAPM: 8.4%, 34.1% and 57.5%.

When returns are measured before expenses, we have 23.2% out of the entire sample as skilled funds, 12.4% as unskilled funds, and 64.4% as zero-alpha funds based on the estimates from the 3F-CAPM, which are similar to the proportions based on the 4F-CAPM. If we turn to decomposed results for different significance levels, the results of the 3F-CAPM are also similar to those of the 4F-CAPM. Specifically, the proportions of significant funds ($\hat{S}_\gamma(\%)$), lucky funds ($\hat{F}^+_\gamma(\%)$), unlucky funds ($\hat{F}^-_\gamma$), skilled funds ($\hat{T}^+_\gamma$) and unskilled funds ($\hat{T}^-_\gamma$) are similar to the ones based on the 4F-CAPM. We also observe a slight increase in the proportions of skilled funds and unskilled funds when we switch from lower significance levels to higher ones, which implies that skilled and unskilled funds are more likely to locate in the extreme tails. These results using the 3F-CAPM are similar to those using the 4F-CAPM, hence, we conclude that our results are robust to different versions of the CAPM.

Table 5.7:
Proportions of Skilled, Unskilled, and Zero-alpha Funds: 1984-2015 (3F-CAPM, Time-varying Parameters)
Performance of funds is measured using the multi-regression model based on the 3F-CAPM. Panel A shows the estimated proportions of zero-alpha ($\hat{\pi}_0$), unskilled ($\hat{\pi}_A^-$) and skilled ($\hat{\pi}_A^+$) funds using the FDR approach on net returns. The lower block of Panel A displays the detailed results including significant funds ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$) across the distribution of funds in the left tail and right tail using four different significance levels ($\gamma = 0.05, 0.1, 0.15, 0.2$). Furthermore, significant funds are decomposed into unlucky ($\hat{F}_\gamma^-$) and unskilled ($\hat{T}_\gamma^-$) funds in the left tail, and lucky ($\hat{F}_\gamma^+$) and skilled ($\hat{T}_\gamma^+$) funds in the right tail. The bottom of the Panel A shows the FDR for each significance level. Panel B shows results on gross returns.

Panel A: Net Returns

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	57.9 (0.02)	33.0 (0.02)	9.1 (0.01)
Number of "funds"	2814	1605	440

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	19.6	25.0	29.1	32.2	14.9	13.4	11.5	8.9	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.4	2.9	4.3	5.8	5.8	4.3	2.9	1.4	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	18.2	22.1	24.8	26.4	9.1	9.1	8.6	7.5	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	
FDR$^-$	7.4	11.6	14.9	18.0	38.9	32.4	25.2	16.2	FDR$^+$

Panel B: Gross Returns

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	64.4 (0.02)	12.4	23.2
Number of "funds"	3140	603	1131

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	9.6	12.7	15.6	18.1	25.0	22.1	19.3	15.4	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	
Unlucky $\hat{F}_\gamma^-$(%)	1.6	3.2	4.8	6.4	6.4	4.8	3.2	1.6	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	8.0	9.5	10.8	11.7	18.5	17.2	16.1	13.8	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	
FDR$^-$	16.7	25.4	30.9	35.6	25.8	21.9	16.7	10.5	FDR$^+$

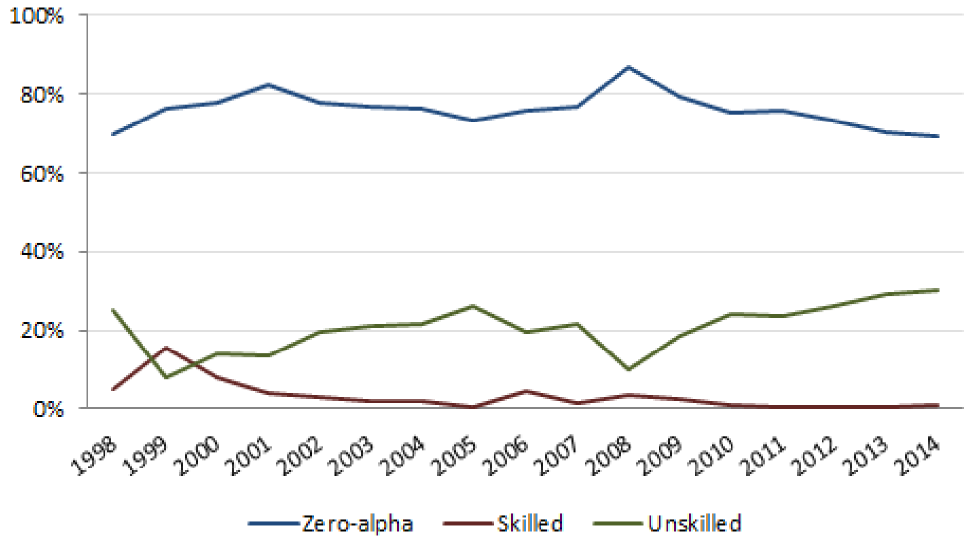

Figure 5.2: Proportions of Zero-alpha, Skilled and Unskilled Funds over Time

5.7.6 Evolution of Mutual Fund Performance over Time

Since we confirmed that some funds have time-varying manager skill based on the results of change point tests, it is of interest to examine how the proportions of skilled, unskilled and zero-alpha funds evolve over time. To do that, we build subsamples that include funds until the end of each year from 1998 to 2014. Then, we estimate the proportions of zero-alpha, skilled and unskilled funds for each subsample using the FDR technique and plot the values in Figure 5.2.

As shown in the figure, the proportion of zero-alpha funds remains relatively constant, fluctuating from 70% to 80%. The proportion of skilled funds continues to drop over these years, from a peak of 15.5% in 1999 to 0.7% in 2014. Accordingly, the proportions of unskilled funds increase from around 10% in the early years to around 30% in recent years. This is generally consistent with the findings of BSW. It is interesting that we find a dramatic drop of the proportion of unskilled funds, and an increase of proportion of skilled funds, during the financial crisis period, which echoes the argument of Kosowski

(2011) that funds tend to generate higher alpha in recession periods than in expansion periods. Generally, these results have shown that the proportions of skilled and unskilled managers vary dramatically over the years. This also provides justification of using the change point test to detect time-variation in fund alphas in our study as manager skill is shown to vary over time.

5.8 Conclusions

Using monthly CRSP mutual fund data, this chapter applied the false discovery rate technique to estimate the proportion of skilled, unskilled and zero-alpha funds over a sample of 3673 actively managed mutual funds that invest primarily in U.S. equities. Differently from the study of BSW, where the traditional 4F-CAPM with time-invariant regression parameters is applied, we implemented a non-parametric change point test to detect possible time-variation in fund alphas and betas. Results from the test show that around 20% of funds in the sample have at least one change in either the alpha or the betas. Based on the results of the change point test, the data series is then split into several subintervals according to the estimated dates of the changes detected, which is referred to as a non-linear multi-regression model.

After factor loadings and fund alphas are allowed to have changes, more skilled and unskilled funds are found: 8.5% of the funds are shown to be skilled and 34.4% of the funds are unskilled. The proportion of zero-alpha funds is 57.1%. These results go counter to many of BSW's conclusions; the way we account for time-variation is the main difference of our study from theirs. We show that the multi-regression model is able to capture time-variation in the regression parameters and hence to estimate fund alphas more accurately. Furthermore, a robustness check using the 3F-CAPM show that the estimated proportions of skilled, unskilled and zero-alpha funds are similar to those using the 4F-CAPM.

When we divide our sample into different investment styles and examine their performance respectively, aggressive growth funds have the highest proportion of skilled funds (12.8%) after all costs. Growth funds and growth & income funds have 8.8% and 10.0% skilled funds, respectively. When returns are measured before expenses, growth & income funds, rather than aggressive growth funds, have the highest proportion of skilled funds (22.1%). This is reasonable given that aggressive growth funds aim to achieve the highest capital gains and usually have the highest turnover. Therefore, for pre-expense returns, when fund managers are required to cover the costs missed in expenses, mainly trading costs, to achieve zero-alpha, aggressive growth funds could perform worse than the other two styles.

To summarize, this study conducted a change point test to detect time-variation in the regression parameters, which remains a problem in the literature. Our results go counter to many of the conclusions of previous studies including BSW, that are based on the traditional factor models. We confirm that there is a significant proportion of skilled managers out there that can produce excess returns to beat the passive benchmarks; this, to some extent, may explain why investors continue to invest in actively managed mutual funds. However, the proportion of skilled managers decreases dramatically when returns are measured after trading costs as well as expenses, indicating that a fund's expense policy can play an important role in the underperformance of mutual funds.

Appendix 5.A Details on the False Discovery Rate Approach

This appendix explains how luck is controlled for using the False Discovery Rate (FDR) approach of Barras, Scaillet and Wermers (2010) (hereafter, BSW), which is based on the approach of Storey (2002). This appendix also shows the equations to compute the standard errors of relevant estimates. The main objective of the FDR approach is to build a framework to estimate the proportions of mutual funds that truly outperform the passive benchmarks. Rather than controlling for the rate of false discoveries, which is what this method is originally designed for, BSW focus on estimating the proportions of false discoveries. In other words, the approach implemented here does not offer "control" of FDR. This may seem counter-intuitive in the traditional multiple test setting, however, it can make sense in this context. In fact, we are more interested in obtaining the proportion of funds that falsely exhibit significant estimated alphas, than in controlling for it. Having this information, we can then estimate the proportions of mutual funds that truly outperform the benchmarks.

5.A.1 The Impact of Luck in a Multiple Fund Setting

As shown in Chapter 5, Section 5.3, mutual funds can be categorized into three performance groups: unskilled, zero-alpha and skilled funds. Of course, we are not able to observe the true alphas of funds in the population. Therefore, to infer the existence of each of the aforementioned performance groups, we base on individual funds' performance estimates. To begin with, we use the t-statistic ($\hat{t}_i = \hat{\alpha}_i / \hat{\sigma}_{\hat{\alpha}_i}$) of funds as the performance measure. Second, a significance level γ (e.g. 5%) is chosen to observe whether $\hat{t}_i$ lies outside the thresholds implied by γ ($t_\gamma^- = -1.65$ and $t_\gamma^+ = 1.65$) and we label it as "significant" if it is such an outlier. As we have 3673 fund in the sample, this procedure applied across all funds simultaneously, becomes a multiple hypothesis test, which is given in Equation 5.18:

$$H_{0,1} : \alpha_i = 0, H_{A,1} : \alpha_i \neq 0, \tag{5.18}$$

$$... : ... \tag{5.19}$$

$$H_{0,N} : \alpha_N = 0, H_{A,N} : \alpha_N \neq 0; \tag{5.20}$$

However, a simple count of significant $t(\alpha)$ funds does not adjust for the role of luck in such a multiple hypothesis test setting as many funds can have significant estimated $t(\alpha)$ estimates due to luck. To control for luck in this multiple test context, i.e. false discoveries, we base on the fact that given a significance level γ, the probability of a zero-alpha fund showing luck equals $\gamma/2$, and the probability of a zero-alpha fund being unlucky is also $\gamma/2$. These lucky and unlucky funds have significant estimated $t(\alpha)$ estimates, but all are zero true alpha funds and hence are "false discoveries". Therefore, the main virtue of the FDR approach is simple: the only parameter needed to determine the proportions of false discoveries, is the proportion of zero-alpha funds in the population (denoted π_0). Previous studies tend to impose a prior assumption on π_0, BSW estimate it using a straightforward approach that is based on the p-values of individual fund alpha estimates alone.

Once the proportion of true zero-alpha funds (π_0) is estimated, we can compute the proportion of funds being unlucky (having significant negative alpha estimate, but are true zero-alpha funds) or lucky (having significant positive alpha estimate, but are true zero-alpha funds). Hence, we can further compute the proportions of truly skilled (unskilled) funds by subtracting the proportion of lucky (unlucky) funds from significant positive (negative) alpha estimates funds.

5.A.2 Standard Errors of Statistics

To get the standard errors of the estimates in Chapter 5, Section 5.5.3, the following equations are used to compute $\widehat{\sigma}_{\widehat{\pi}_0}$, $\widehat{\sigma}_{\widehat{F}^+_\gamma}$, $\widehat{\sigma}_{S^+_\gamma}$, $\widehat{\sigma}_{\widehat{T}^+_\gamma}$ and $\widehat{\sigma}_{\widehat{\pi}^+_A}$. Formal justification for these formulas can be found in the Internet Appendix of BSW:

$$\widehat{\sigma}_{\widehat{\pi}_0} = \left(\frac{\widehat{W}(\lambda^*)(N - \widehat{W}(\lambda^*))}{N^3(1-\lambda^*)^2} \right)^{\frac{1}{2}}, \tag{5.21}$$

where $\widehat{W}(\lambda^*)$ denotes the number of funds that have p-values greater than λ^*. Then we have:

$$\widehat{\sigma}_{\widehat{F}^+_\gamma} = (\gamma/2)\widehat{\sigma}_{\widehat{\pi}_0}, \tag{5.22}$$

$$\widehat{\sigma}_{S^+_\gamma} = \left(\frac{S^+_\gamma(1 - S^+_\gamma)}{N} \right)^{\frac{1}{2}}, \tag{5.23}$$

$$\widehat{\sigma}_{\widehat{T}^+_\gamma} = \left(\widehat{\sigma}_{\widehat{F}^+_\gamma} + (\gamma/2)^2 \widehat{\sigma}_{\widehat{\pi}_0} + 2\frac{(\gamma/2)}{1-\lambda^*} S^+_\gamma \frac{\widehat{W}(\lambda^*)}{N^2} \right)^{\frac{1}{2}}, \tag{5.24}$$

and if $\gamma^* = \gamma^+$, the standard errors of $\widehat{\pi}^+_A$ is computed as $\widehat{\sigma}_{\widehat{T}^+_{\gamma^*}}$. To reserve equality, $\widehat{\sigma}_{\widehat{\pi}^-_A}$ is given by:

$$\widehat{\sigma}_{\widehat{\pi}^-_A} = \left(\widehat{\sigma}^2_{\widehat{\pi}^+_A} + \widehat{\sigma}^2_{\widehat{\pi}_0(\lambda^*)} - 2(\frac{1}{1-\lambda^*}) S^+_{\gamma^*} \frac{\widehat{W}(\lambda^*)}{N^2} - 2(\lambda^*/2)\widehat{\sigma}^2_{\widehat{\pi}_0} \right)^{\frac{1}{2}}; \tag{5.25}$$

In the case where $\gamma^* = \gamma^-$, the superscripts are reversed to obtain all the standard deviations.

Chapter 6

Data Selection, Conditional Models and Bootstrap Methods in Mutual Fund Studies: A Critical Evaluation

6.1 Introduction

This chapter discusses three important issues related to mutual fund studies in general. These issues remain controversial in the literature, hence we revisit these issues in this chapter, and aim to provide a critical evaluation that sheds new light on these issues.

First, this chapter looks at sample selection issues associated with mutual fund data used to evaluate performance of U.S. mutual funds. These issues can be grouped into three types: the first of this taxonomy concerns the minimum number of months a fund must have before it is included in the sample; the second is the incubated returns of funds, i.e. the attractive return history established when they are given the seed money by their fund families; and the last type addresses the newly established funds in the sample. Different rules are used in previous studies to include funds in their samples, and as a consequence some of them have come to contradictory conclusions. For example, Fama and French (2010) (hereafter F&F) include a fund into their sample if it has at least 8 months of return history, whereas Barras, Scaillet and Wermers (2010) (hereafter BSW) require a fund to have at least 60 months of returns history to be included. Hence, it is of interest to examine how different ways of constructing datasets lead to different conclusions.

Using the CRSP survivor-bias free data, we construct several different datasets that allow us to examine the effects of different treatments regarding incubated returns, new funds, and funds with short return history, on the cross-sectional distribution of $t(\alpha)$ estimates. The results from the regressions show that including incubated returns produces slightly smaller individual fund alpha estimates compared to the other funds. More importantly, the use of a longer minimum return history a fund has, produces a larger alpha estimate on an aggregate equal-weighted portfolio, and shift the cross-sectional distribution of alpha ($t(\alpha)$) estimates to the right. This implies that the requirement of a longer minimum return history would induce more survivor-bias given that funds with short return history are excluded from study. The effect of this is to bias findings towards existence of managers with stock selecting skills.

The second issue concerns the methods used in conditional performance evaluation. It has been traditional to measure fund performance by the average excess return, net of a fixed benchmark return over some historical period. However, it is well-known that fund managers may apply dynamic trading strategies including adjusting portfolio weights to respond to the changing market, hence betas can be time-varying. Based on this, Ferson and Schadt (1996) developed a conditional CAPM where the market beta is a linear function of some public information variables. Although they confirm that these information variables are statistically significant, and the use of their conditional model improves the performance of funds in their sample, recent studies by Blake and Timmermann (1998), Kacperczyk et al. (2005), Kosowski, Timmermann, Wermers and White (2006) (hereafter, KTWW), Cuthbertson et al. (2008), BSW and Cuthbertson et al. (2012) argue that the Ferson-Schadt conditional model produces estimates of fund alphas that not significantly different from unconditional versions of these models. This motivates us, as described in Chapter 4, to use a novel approach to capture time-variation in the alpha and betas.

Since Ferson and Schadt (1996) and later studies that used the Ferson-Schadt conditional model come to differ, we aim to revisit this model and estimate fund alphas with it on our data. Results produced here suggest that Ferson-Schadt conditional model yields similar $t(\alpha)$ estimates relative to those from the unconditional models. Ferson and Schadt (1996) takes a reduced-form approach to model time-variation in the market beta of the CAPM, i.e. the market beta is a linear function of the macroeconomic variables. While this simplifies the model and facilitates estimation, this specification may not capture other forms of time-variation that cause the market beta to vary over time. Because of this, it is argued here that a more appropriate method to model this time-variation is to test for changes in the regression parameters of the various versions of the CAPM, using a nonparametric test, as shown in Chapters 4 and 5. With this information, it is a simple process to estimate those factor models that allows the regression parameters to take discrete change over a fund's time-series of returns. This allows us to take an agnostic view on the nature of the relation between risk factors and the information variables.

A further issue of concern are the bootstrap methods that have been widely used in mutual fund studies. We have seen different bootstrap methods applied in previous studies. The most frequently used are bootstrap by residuals, bootstrap by pairs, and block bootstrap. KTWW and BSW use bootstrap by residuals that resamples from the residuals of fitted regression models to produce bootstrapped cross-sectional alpha ($t(\alpha)$) estimates. Given that resampling from residuals is undertaken for each fund separately, F&F argue that bootstrap by residuals would lose any effect of cross-fund correlation in the simulation. Instead, F&F use bootstrap by pairs that resamples from a fund's returns and explanatory returns jointly as a pair, and resamples for all funds simultaneously. Furthermore, to capture possible autocorrelation effects in the time series, BSW use a block bootstrap method that resamples from blocks of residuals.

Different choices of resampling procedure has an important impact on the cross-sectional distribution of $t(\alpha)$ estimates and could affect the inferences made regarding the existence of skilled managers. To examine this impact, we redo the analysis of F&F and BSW on our data using the aforementioned three bootstrap methods, respectively. Results from the simulations show that, overall, bootstrap by residuals tends to produce less extreme simulated $t(\alpha)$ estimates, and autocorrelation effect is only a minor issue in our study.

The rest of the chapter is organized as follows: Section 6.2 explores the sample selection issues. Section 6.3 examines the Ferson-Schadt conditional model and Section 6.4 evaluates the bootstrap methods. The final section concludes.

6.2 Data Selection Issues

Although the CRSP mutual fund database has been widely used in the literature (Carhart, 1997; Daniel et al., 1997; Kosowski et al., 2006; Wermers, 2000; Ferson and Schadt, 1996;

Berk and Green, 2004; Elton et al., 2007; French, 2008; Evans, 2010; Fama and French, 2010; Barras et al., 2010), previous studies have applied different rules when selecting funds to include in their samples. For example, F&F include funds in their sample as long as the fund has at least eight months of return history, Pástor and Stambaugh (2002) require funds in the sample to have at least 12 months of return history, Wermers (2000) include a fund only if it has survived for at least 24 months, and funds in the sample of Sirri and Tufano (1998), KTWW and BSW have a minimum 60 months of return history.

In this section, we examine data selection issues based on two recent studies: F&F and BSW, where they differ in their data lies in which funds from CRSP that they include in their research. Specifically, as described previously, F&F and BSW include funds that have different minimum return history, and fewer funds are included in BSW's data (60-month rule) compared to F&F's (8-month rule). A consequence of this is that the results of BSW will trade off the effects of more survivor bias for more accurate estimates of parameters. They also have different treatments regarding incubated returns of funds. F&F argue that "incubation" bias arises because funds are often given seed money by a management company to develop an appealing return history, and these returns are included in the sample. F&F offset the effects of this bias on their results by setting an asset under management (AUM) threshold that a fund must exceed in order to be selected, whereas BSW do not correct for this bias. A last source of their differences in the data used in their research is due to different treatments of new funds that appear less than five years before the end date of the sample. F&F exclude these new funds, whereas BSW did not exclude them. Therefore, to examine how different methods of constructing datasets lead to different estimates of fund alphas in F&F and BSW, two datasets have been constructed here: one based on F&F's selection rules and the other one based on BSW.

6.2.1 Regression Framework

Firstly, we reproduce the baseline models used to estimate fund alphas. Three baseline models are used here to estimate fund alphas, and these are the CAPM, the 3F-CAPM and the 4F-CAPM. Their respective forms are given below:

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \varepsilon_{i,t}, \tag{6.1}$$

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \varepsilon_{i,t}, \tag{6.2}$$

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{i,t}, \tag{6.3}$$

where $R_{i,t}$ is the return for month t of fund i, R_t^f is the risk-free rate, which is the one-month U.S. Treasury bill rate, R_t^m is the market return constructed with a value-weighted portfolio of NASDAQ, NYSE and Amex stocks, SMB_t and HML_t are the size and value-growth returns constructed along the lines of Fama and French (1993), MOM_t is the momentum return of Carhart (1997), α_i (the estimate of α_i) is the return left unexplained by this benchmark portfolio, i.e. the extra return generated by funds and $\varepsilon_{i,t}$ is the regression residual. The four factors that appear in the augmented versions of the CAPM are obtained from Kenneth French's website[1].

6.2.2 Aggregate Performance on Datasets Using Different Selection Rules

First, interest here concerns aggregate fund performance of these two datasets. In particular, fund monthly returns are averaged in two ways to characterize aggregate fund

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

performance. An equal-weight portfolio of monthly fund returns is formed by averaging monthly returns for funds in the sample. This will allow assessment of fund performance on average. Also of interest is the performance in terms of return of aggregate wealth invested in U.S. mutual funds. This is assessed by a weighted average of monthly returns of funds where the weights are the values of funds' beginning-of-month assets under management. We run regressions using the CAPM, the 3F-CAPM and the 4F-CAPM and results from these regressions are presented in Table 6.1.

Table 6.1:
A Comparison of Intercepts and Slopes in Variants of Regression for Equal-Weight (EW) and Value-Weight (VW) Portfolios of Actively Managed Mutual Funds (F&F vs. BSW Datasets)

This table provides the annualized intercepts (12*$\hat{\alpha}$) and t-statistics for the intercepts ($t(Coef)$) for the CAPM, the 3F-CAPM, and the 4F-CAPM estimated on an equal-weight (EW) and value-weight (VW), net and gross returns based on two panels of actively managed mutual funds. One panel of funds is established following the methods in F&F and the other one following BSW. Regressions slopes ($\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$) for the corresponding factors ($RM - Rf$, SMB, HML, and MOM) are also shown in the table. It should be noted that for the market slope, $t(Coef)$ tests the null hypothesis that β is equal to 1. Net returns are returns reported in the CRSP mutual fund database, and the gross returns are net returns plus $1/12^{th}$ of a fund's expense ratio at that year end. When a fund's expense ratio for a year is missing, we replace it with the expense ratio of the fund that has the same investment style and has similar assets under management (AUM). As for those funds with zero expense ratios, we manually check some of their annual reports and confirm that they in fact have positive expense ratios. Hence, we treat zero expense ratios as missing values. Our sample covers the period from January 1984 to March 2015.

	12*$\hat{\alpha}$ (%)						
	Net	Gross	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2 (%)
Section A: Comparison of Results on EW Returns							
F&F Dataset							
CAPM $Coef$	-0.88	0.41	0.99				0.96
$t(Coef)$	-1.60	0.74	-0.87				
3F-CAPM $Coef$	-0.87	0.42	0.96	0.21	0.02		0.98
$t(Coef)$	-2.17	1.03	-4.89	18.10	1.53		
4F-CAPM $Coef$	-0.88	0.41	0.96	0.21	0.02	-0.00	0.98
$t(Coef)$	-2.14	1.00	-4.73	18.06	1.51	0.06	
BSW Dataset							
CAPM $Coef$	-0.80	0.51	1				0.96
$t(Coef)$	-1.50	0.95	-0.41				
3F-CAPM $Coef$	-0.78	0.53	0.97	0.20	0.02		0.98
$t(Coef)$	-2.05	1.38	-4.52	18.75	1.34		
4F-CAPM $Coef$	-0.80	0.51	0.97	0.20	0.02	0.00	0.98
$t(Coef)$	-2.05	1.31	-4.32	18.71	1.34	0.24	

continued

Table 6.1 – *continued*

	$12^*\hat{\alpha}$ (%)						
	Net	Gross	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2 (%)
Section B: Comparison of Results on VW Returns							
F&F Dataset							
CAPM *Coef*	-1.01	-0.03	0.98				0.98
$t(Coef)$	-2.59	-0.08	-3.24				
3F-CAPM *Coef*	-0.96	0.01	0.96	0.09	-0.00		0.98
$t(Coef)$	-2.72	0.04	-5.53	9.04	-0.12		
4F-CAPM *Coef*	-0.98	-0.01	0.96	0.09	-0.00	0.00	0.98
$t(Coef)$	-2.73	-0.02	-5.29	9.02	-0.06	0.33	
BSW Dataset							
CAPM *Coef*	-1.02	-0.05	0.98				0.98
$t(Coef)$	-2.76	-0.14	-2.93				
3F-CAPM *Coef*	-0.97	0.00	0.97	0.09	-0.00		0.99
$t(Coef)$	-2.88	0.00	-5.32	9.10	-0.35		
4F-CAPM *Coef*	-0.99	-0.02	0.97	0.09	-0.00	0.00	0.99
$t(Coef)$	-2.90	-0.07	-5.07	9.07	-0.30	0.37	

As shown in Table 6.1, in terms of net returns received by investors, the performance of funds is generally not promising. For gross returns, fund alphas are positive for EW returns and negative for VW returns, both being insignificant from zero.

Regarding the differences of results for the F&F and BSW datasets, for EW portfolios where funds are weighted equally, we see that the BSW dataset produces slightly more promising results than that of F&F generally. The annualized alphas for the F&F dataset are from -0.88 to -0.87, and the BSW dataset has annualized alphas that range from -0.78 to -0.80. When returns are measured before expenses, annualized alphas based on F&F are between 0.40 and 0.41, which are smaller than alphas using the BSW dataset, being 0.51-0.53. As to VW portfolios, the annualized alphas are similar for both datasets. This is not surprising given that fund returns that are not included in the BSW dataset are incubated returns and funds with short return histories; these funds tend to have small assets under management, so their contribution to an value-weighted aggregate portfolio is much smaller compared to an equal-weighted portfolio.

6.2.3 A Further Examination of Sample Selection Issues

To explore further the sample selection issues, in particular, the exact source that leads to the different results on aggregate portfolios, we construct a further three panels of funds, along with the F&F dataset, where: (i) Panel A is the F&F dataset, or the benchmark dataset, which excludes newly established funds within 5 years before the end of our sample period March 2015, incubated returns of funds when they do not pass a $5 million AUM bound, and funds that have less than 8 months of returns history; the construction of the other three panels follows F&F's methodology, but (ii) Panel B includes new funds; (iii) Panel C includes incubated returns of funds; (iv) Panel D excludes funds that have less than 60 months of return history. Estimates for the three versions of the CAPM using EW and VW portfolio of funds in the respective sample are reported in Table 6.2.

Table 6.2:

Intercepts and Slopes in Variants of the CAPM for Equal-Weight (EW) and Value-Weight (VW) Portfolios of Actively Managed Mutual Funds (Different Sample Selection Rules)

This table provides the annualized intercepts ($12^*\hat{\alpha}$) and t-statistics for the intercepts ($t(Coef)$) for the CAPM, the 3F-CAPM, and the 4F-CAPM estimated on an equal-weight (EW) and value-weight (VW), net and gross returns based on four different panels of actively managed mutual funds that are built according to different sample selection rules. Regressions slopes ($\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$) for the corresponding factors ($RM - Rf$, SMB, HML, and MOM) are also shown in the table. It should be noted that for the market slope, $t(Coef)$ tests the null hypothesis that β is equal to 1. Net returns are returns reported in the CRSP mutual fund database, and the gross returns are net returns plus $1/12^{th}$ of a fund's expense ratio at that year end. When a fund's expense ratio for a year is missing, we replace it with the expense ratio of the fund that has the same investment style and has similar assets under management (AUM). As for those funds with zero expense ratios, we manually check some of their annual reports and confirm that they in fact have positive expense ratios. Hence, we treat zero expense ratios as missing values. All the panels cover the period from January 1984 to March 2015.

	$12^*\hat{\alpha}$ (%)						
	Net	Gross	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2 (%)
Panel A: F&F dataset (Benchmark dataset)							
EW Returns							
CAPM $Coef$	-0.88	0.41	0.99				0.96
$t(Coef)$	-1.60	0.74	-0.87				
3F-CAPM $Coef$	-0.87	0.42	0.96	0.21	0.02		0.98
$t(Coef)$	-2.17	1.03	-4.89	18.10	1.53		
4F-CAPM $Coef$	-0.88	0.41	0.96	0.21	0.02	-0.00	0.98
$t(Coef)$	-2.14	1.00	-4.73	18.06	1.51	0.06	
VW Returns							
CAPM $Coef$	-1.01	-0.03	0.98				0.98
$t(Coef)$	-2.59	-0.08	-3.24				
3F-CAPM $Coef$	-0.96	0.01	0.96	0.09	-0.00		0.98
$t(Coef)$	-2.72	0.04	-5.53	9.04	-0.12		
4F-CAPM $Coef$	-0.98	-0.01	0.96	0.09	-0.00	0.00	0.98
$t(Coef)$	-2.73	-0.02	-5.29	9.02	-0.06	0.33	

continued

Table 6.2 – *continued*

	12*$\hat{\alpha}$ (%)						
	Net	Gross	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2 (%)
Panel B: Including new funds							
EW Returns							
CAPM $Coef$	-0.89	0.40	0.99				0.98
$t(Coef)$	-1.62	0.72	-0.93				
3F-CAPM $Coef$	-0.88	0.41	0.96	0.20	0.02		0.98
$t(Coef)$	-2.20	1.01	-4.98	18.09	1.51		
4F-CAPM $Coef$	-0.89	0.40	0.96	0.20	0.02	0.00	0.98
$t(Coef)$	-2.17	0.98	-4.81	18.05	1.50	0.06	
VW Returns							
CAPM $Coef$	-1.01	-0.04	0.98				0.98
$t(Coef)$	-2.61	-0.09	-3.28				
3F-CAPM $Coef$	-0.97	0.01	0.96	0.09	0.00		0.98
$t(Coef)$	-2.74	0.03	-5.58	9.02	-0.14		
4F-CAPM $Coef$	-0.99	-0.01	0.96	0.09	0.00	0.00	0.98
$t(Coef)$	-2.75	-0.03	-5.33	8.99	-0.08	0.32	
Panel C: Including incubated returns							
EW Returns							
CAPM $Coef$	-1.02	0.32	0.99				0.98
$t(Coef)$	-1.85	0.58	-1.12				
3F-CAPM $Coef$	-1.02	0.32	0.96	0.21	0.02		0.98
$t(Coef)$	-2.54	0.79	-5.21	18.20	1.67		
4F-CAPM $Coef$	-0.99	0.35	0.96	0.21	0.02	0.00	0.98
$t(Coef)$	-2.43	0.85	-5.14	18.18	1.56	-0.39	
VW Returns							
CAPM $Coef$	-1.01	-0.03	0.98				0.98
$t(Coef)$	-2.59	-0.08	-3.23				
3F-CAPM $Coef$	-0.96	0.01	0.96	0.09	0.00		0.98
$t(Coef)$	-2.72	0.04	-5.54	9.05	-0.12		
4F-CAPM $Coef$	-0.98	-0.01	0.96	0.09	0.00	0.00	0.98
$t(Coef)$	-2.74	-0.02	-5.29	9.02	-0.06	0.33	

continued

Table 6.2 – *continued*

	$12^*\hat{\alpha}$ (%)						
	Net	Gross	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2 (%)
	Panel D: Using the 60-month rule						
EW Returns							
CAPM *Coef*	-0.70	0.56	1.00				0.98
$t(Coef)$	-1.30	1.04	-0.10				
3F-CAPM *Coef*	-0.70	0.57	0.97	0.21	0.02		0.98
$t(Coef)$	-1.80	1.47	-4.06	18.93	1.60		
4F-CAPM *Coef*	-0.74	0.53	0.97	0.21	0.02	0.00	0.98
$t(Coef)$	-1.88	1.34	-3.80	18.88	1.68	0.58	
VW Returns							
CAPM *Coef*	-1.02	-0.05	0.98				0.98
$t(Coef)$	-2.74	-0.14	-2.86				
3F-CAPM *Coef*	-0.97	0.00	0.97	0.09	0.00		0.98
$t(Coef)$	-2.89	-0.01	-5.25	9.24	-0.27		
4F-CAPM *Coef*	-0.99	-0.02	0.97	0.09	0.00	0.00	0.98
$t(Coef)$	-2.90	-0.07	-5.01	9.21	-0.20	0.36	

As shown in Table 6.2, a comparison of results based on the other three panels to those based on the benchmark sample (F&F dataset) suggests the following conclusions: compared to the benchmark sample, (i) the sample including new funds (Panel B) generates similar alpha estimates; (ii) including incubated returns (Panel C), i.e. without excluding pre-release returns, gives more negative net alpha and less positive gross alpha estimates for the EW portfolio; (iii) using the 60-month rule (Panel D) gives less negative net alpha and more positive gross alpha estimates for the EW portfolio.

It should be noted that, in terms of the incubated fund returns, our results are contrary to the findings of Evans (2010) and F&F where they assert that incubated funds should have enhanced performance, hence, including incubated returns would lead to greater values of alpha estimates. However, Panel C that includes incubated fund returns produces more negative net alpha and less positive gross alpha estimates relative to the benchmark sample. This suggests that the included incubated funds underperform the other funds. Therefore, we show that incubated funds in aggregate do not deliver enhanced performance, and excluding incubated funds, i.e. the part of return history when funds do not pass a certain AUM bound, does not correct for "incubation" bias as suggested by F&F. We argue that this is reasonable given that funds could underperform in "incubated" times as they just enter the market and usually have lower levels of assets under management, hence are more vulnerable to risks or distresses.

Table 6.3: **Percentiles of Actual $t(\alpha)$ Estimates: 1984-2015 (Different Sample Selection Rules)**
This table shows the values of $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns, for four panels using different sample selection rules. Panel A is a replication of F&F dataset and the other panels follow the methodology of F&F but (i) Panel B does not exclude new funds; (ii) Panel C includes incubated returns; (iii) Panel D use a 60-month rule instead of 8-month rule for the requirement of minimum return history. Panel A that uses the F&F dataset serves as a benchmark dataset. Fund $t(\alpha)$ estimates are estimated using the 4F-CAPM. There are 5006 funds in Panel A, 5435 funds in Panel B, 5392 funds in Panel C and 3553 funds in Panel D. All panels cover the period from January 1984 to March 2015.

	Net returns				Gross returns			
Pct	Panel A	Panel B	Panel C	Panel D	Panel A	Panel B	Panel C	Panel D
1	-3.79	-3.76	-3.80	-3.52	-2.99	-3.00	-3.03	-2.66
2	-3.26	-3.26	-3.34	-3.13	-2.53	-2.56	-2.54	-2.33
3	-2.99	-2.99	-3.01	-2.95	-2.32	-2.32	-2.31	-2.05
4	-2.86	-2.86	-2.88	-2.81	-2.11	-2.11	-2.11	-1.93
5	-2.71	-2.70	-2.73	-2.68	-1.99	-2.00	-1.99	-1.84
10	-2.23	-2.23	-2.23	-2.20	-1.54	-1.55	-1.54	-1.37
20	-1.68	-1.67	-1.66	-1.65	-0.93	-0.94	-0.96	-0.78
30	-1.24	-1.24	-1.24	-1.23	-0.53	-0.55	-0.55	-0.39
40	-0.91	-0.91	-0.92	-0.90	-0.22	-0.23	-0.23	-0.07
50	-0.61	-0.61	-0.61	-0.59	0.09	0.08	0.08	0.24
60	-0.32	-0.32	-0.32	-0.30	0.38	0.37	0.36	0.54
70	0.01	0.01	0.01	0.04	0.71	0.70	0.71	0.87
80	0.40	0.40	0.41	0.47	1.10	1.09	1.12	1.29
90	1.00	1.00	1.00	1.09	1.73	1.71	1.73	1.92
95	1.49	1.47	1.50	1.56	2.26	2.23	2.24	2.39
96	1.62	1.60	1.62	1.68	2.40	2.38	2.38	2.56
97	1.85	1.82	1.83	1.92	2.63	2.59	2.58	2.75
98	2.12	2.11	2.08	2.15	2.89	2.86	2.89	3.08
99	2.56	2.52	2.53	2.58	3.40	3.33	3.43	3.51

We also estimate the alpha for individual funds to see if the distribution of cross-sectional $t(\alpha)$ estimates vary among the above four panels. As shown in Table 6.3, compared to the F&F dataset (Panel A), Panel B that includes new funds, and Panel C that includes incubated returns, produce similar percentiles of cross-sectional $t(\alpha)$ estimates. Panel D that uses a 60-month rule generates less negative $t(\alpha)$ estimates in the left tail and more positive $t(\alpha)$ estimates in the right tail compared to the F&F dataset that uses a 8-month rule. This confirms our argument that requiring a longer minimum return history for a fund to be included in the sample would induce more survivor bias given that funds with short return history are excluded from the sample. Hence, using a 60-month rule as suggested by BSW biases their conclusions towards more promising inferences on the manager skill of mutual funds.

6.2.4 A Re-examination of F&F's and BSW's Analysis Using Different Sample Selection Rules

6.2.4.1 Different Minimum Return History Requirements

Knowing that different selection rules could lead to different results, we further redo the analysis of F&F based on a dataset using the 60-month rule, and the analysis of BSW based on a dataset using the 8-month rule. This allows us to further understand possible consequences of applying different rules to select funds in their analysis. Since various version of the CAPM tend to give similar results as shown in Section 6.2.2, for brevity, we use the 4F-CAPM to estimate fund alphas here.

To replicate F&F's analysis based on a dataset using the 60-month rule, we construct a new dataset according to F&F's methodology except we exclude funds that have less than 60 months return history, and perform the simulations described in Chapter 4, Section 4.4 on this new dataset. Results are shown in Table 6.4. Generally, using the 60-month rule produces more positive actual $t(\alpha)$ estimates in the right tail and less negative actual $t(\alpha)$ estimates in the left tail, for net and gross returns, compared to those using the 8-month rule. Similarly, the likelihoods are greater for the dataset using the 60-month rule. This suggests that different rules of selecting funds in the sample can alter the inferences on the existence of skilled and unskilled funds. A rule that requires longer minimum return history that funds should survive before they can be included in the sample, rules out those funds with short return histories; it hence produces larger $t(\alpha)$ estimates, indicating larger probabilities of funds being skilled in the right tail and smaller probabilities of funds being unskilled in the left tail of the cross-sectional distribution of $t(\alpha)$ estimates.

Table 6.4: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2015 (8-month vs. 60-month Rule)**
This table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We bootstrap by resampling fund and factor returns jointly as a pair. Panel A shows the comparison of percentiles of net return $t(\alpha)$ estimates on a dataset using an inclusion rule of 8 months to those on a dataset using an inclusion rule of 60 months, and the comparison for gross returns are shown in Panel B. There are 5006 funds in the sample using an inclusion rule of 8 months; when funds are required to have at least 60 months of return history to be included in the sample, we have 3553 funds in the dataset. Both datasets cover the period from January 1984 to March 2015.

	Panel A: net returns						Panel B: gross returns					
	8-month rule			60-month rule			8-month rule			60-month rule		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
1	-3.79	-2.46	0.01	-3.52	-2.32	0.10	-2.99	-2.47	3.40	-2.66	-2.33	13.00
2	-3.26	-2.12	0.01	-3.13	-2.05	0.10	-2.53	-2.13	5.80	-2.33	-2.06	16.05
3	-2.99	-1.93	0.01	-2.96	-1.88	0.10	-2.32	-1.93	5.84	-2.05	-1.89	23.85
4	-2.86	-1.78	0.01	-2.81	-1.75	0.05	-2.11	-1.79	8.59	-1.93	-1.76	23.00
5	-2.71	-1.67	0.01	-2.68	-1.65	0.10	-1.99	-1.68	8.46	-1.84	-1.65	21.30
10	-2.23	-1.29	0.01	-2.20	-1.28	0.20	-1.54	-1.29	11.80	-1.37	-1.29	33.00
20	-1.68	-0.84	0.02	-1.65	-0.84	0.30	-0.93	-0.85	31.68	-0.78	-0.85	59.30
30	-1.24	-0.52	0.02	-1.23	-0.53	0.45	-0.53	-0.53	48.10	-0.39	-0.53	72.85
40	-0.91	-0.25	0.03	-0.90	-0.26	0.75	-0.22	-0.26	57.56	-0.07	-0.26	79.95
50	-0.61	-0.00	0.03	-0.59	0.00	0.75	0.09	-0.01	69.44	0.24	-0.01	85.95
60	-0.32	0.25	0.04	-0.30	0.25	0.85	0.38	0.25	77.57	0.54	0.25	89.00
70	0.01	0.52	0.09	0.04	0.52	1.55	0.71	0.52	84.39	0.87	0.52	91.40
80	0.40	0.84	0.40	0.48	0.84	6.05	1.10	0.84	90.30	1.29	0.83	95.95
90	1.00	1.29	6.57	1.09	1.28	24.55	1.73	1.29	97.14	1.92	1.27	98.60
95	1.49	1.68	19.20	1.56	1.65	41.20	2.26	1.68	98.61	2.39	1.64	99.05
96	1.62	1.80	22.14	1.68	1.75	44.10	2.40	1.79	98.76	2.56	1.74	99.20
97	1.85	1.94	37.98	1.92	1.88	58.60	2.63	1.94	99.10	2.75	1.87	99.55
98	2.12	2.14	50.18	2.15	2.06	67.05	2.89	2.14	99.36	3.08	2.04	99.80
99	2.56	2.49	63.18	2.58	2.33	81.25	3.40	2.49	99.67	3.51	2.32	99.90

We further replicate the analysis of BSW to quantify the proportion of skilled and unskilled funds based on another new dataset, which is built according to BSW's methodology except that funds are included if they have more than 8 months of return history.

We estimate fund alphas using the 4F-CAPM, and perform the FDR technique as described in Chapter 5, Section 5.5 to estimate the proportion of skilled, unskilled and zero-alpha funds. The results are shown in Appendix 6.A, Table 6.9.

For net returns, the proportion of skilled funds increases slightly from 0.7% to 1.7% when we switch from an inclusion rule of 60 months to 8 months. We also look at the decomposed results as shown in the lower block of Panels A and B, i.e. the proportion of unskilled funds at different levels of γ. Compared to Panel A (14.0%, 19.5%, 24.6%, 28.5% at significance level $\gamma = 0.05, 0.1, 0.15, 0.2$, respectively), Panel B that uses the 8-month rule indeed produces smaller proportions of unskilled funds (17.9%, 23.4%, 28.3%, 32.1% at significance level $\gamma = 0.05, 0.1, 0.15, 0.2$, respectively). For gross returns, the proportion of skilled funds decreases dramatically from 12.7% to 6.6% and the proportion of unskilled funds increases from 3.1% to 8.4%. Generally, these results suggest that when funds are required to have a longer minimum return history to be included in the sample, more unskilled funds are discovered. This is consistent with our argument that funds which survived less than 60 months tend to underperform the passive benchmarks, hence most of them will be identified as unskilled funds if they are included in the sample.

6.2.4.2 Different Treatments on Incubated Fund Returns

To examine the impacts of different treatments on incubated returns, we redo the analysis of F&F and BSW based on a dataset with incubated returns included in the sample, and compare to the dataset that has the incubated returns opted out. Recall that the incubated returns are return history of funds when their assets under management (AUM) do not pass a certain bound, and the definition of the AUM bound is given in Appendix A, Section A.7.

Results of the comparison based on F&F's analysis are shown in Table 6.5. Generally, the dataset with the incubated returns produces similar actual $t(\alpha)$ estimates, and slightly more negative simulated $t(\alpha)$ estimates in the left tail and more positive simulated $t(\alpha)$ estimates in the right tail at selected percentiles. As a results, the likelihoods are slightly greater for the left tail percentiles and smaller for the right tails for the dataset with the incubated returns, although the difference is trivial.

Table 6.5: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2015 (With and without Incubated Returns)**
This table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We bootstrap by resampling fund and factor returns jointly as a pair. Panel A shows the comparison of percentiles of net return $t(\alpha)$ estimates on a dataset without incubated returns to those on a dataset with incubated returns, and the comparison for gross returns are shown in Panel B. There are 5006 funds in the sample without incubated returns and 5390 funds in the dataset with incubated returns from January 1984 to March 2015.

	Panel A: net returns						Panel B: gross returns					
	Without incubated returns			With incubated returns			Without incubated returns			With incubated returns		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
Net Returns							Gross Returns					
1	-3.79	-2.46	0.01	-3.82	-2.54	0.09	-2.99	-2.47	3.40	-3.03	-2.53	4.64
2	-3.26	-2.12	0.01	-3.34	-2.18	0.07	-2.53	-2.13	5.80	-2.55	-2.17	7.14
3	-2.99	-1.93	0.01	-3.01	-1.97	0.07	-2.32	-1.93	5.84	-2.31	-1.97	7.79
4	-2.86	-1.78	0.01	-2.88	-1.83	0.07	-2.11	-1.79	8.59	-2.11	-1.82	10.42
5	-2.71	-1.67	0.01	-2.73	-1.71	0.07	-1.99	-1.68	8.46	-1.99	-1.70	10.19
10	-2.23	-1.29	0.01	-2.23	-1.32	0.06	-1.54	-1.29	11.80	-1.54	-1.31	13.75
20	-1.68	-0.84	0.02	-1.66	-0.86	0.03	-0.93	-0.85	31.68	-0.96	-0.86	28.32
30	-1.24	-0.52	0.02	-1.24	-0.53	0.04	-0.53	-0.53	48.10	-0.55	-0.53	42.70
40	-0.91	-0.25	0.03	-0.92	-0.26	0.05	-0.22	-0.26	57.56	-0.23	-0.26	53.61
50	-0.61	-0.00	0.03	-0.61	0.00	0.06	0.09	-0.01	69.44	0.08	0.00	67.23
60	-0.32	0.25	0.04	-0.32	0.26	0.06	0.38	0.25	77.57	0.36	0.26	72.66
70	0.01	0.52	0.09	0.01	0.53	0.08	0.71	0.52	84.39	0.71	0.53	82.66
80	0.40	0.84	0.40	0.41	0.86	0.34	1.10	0.84	90.30	1.12	0.86	89.87
90	1.00	1.29	6.57	1.00	1.32	4.64	1.73	1.29	97.14	1.73	1.32	96.36
95	1.49	1.68	19.20	1.50	1.72	16.34	2.26	1.68	98.61	2.24	1.72	97.89
96	1.62	1.80	22.14	1.62	1.84	18.12	2.40	1.79	98.76	2.38	1.84	98.01
97	1.85	1.94	37.98	1.83	1.99	25.84	2.63	1.94	99.10	2.59	1.99	98.43
98	2.12	2.14	50.18	2.08	2.20	34.41	2.89	2.14	99.36	2.89	2.20	98.92
99	2.56	2.49	63.18	2.55	2.58	51.10	3.40	2.49	99.67	3.44	2.57	99.08

We further replicate the analysis of BSW to quantify the proportion of skilled and unskilled funds based on this two datasets, and the results are shown in Appendix 6.A,

Table 6.10. Generally, the results using the dataset with the incubated returns included are similar to those using the dataset with the incubated returns opted out, except that for the gross returns, when the incubated returns are included, we find a slightly smaller proportion of skilled funds.

Knowing that different rules in constructing data could lead to different inferences on fund performance, and to avoid any biases caused by data issues, we mainly follow F&F's methodology to construct our data, and the details are shown in Appendix A.

6.3 The Conditional Approach of Examining Fund Performance

6.3.1 The Ferson-Schadt Model Framework

The second issue of interest is the conditional approach of estimating Jensen's alpha. We aim to test whether the market beta of mutual funds is correlated with the public information variables, and whether using the Ferson-Schadt conditional model produces different alpha estimates from those using the unconditional ones. Given that different versions of the CAPM generate similar alpha estimates, as shown in Section 6.2.2, for brevity, we construct the Ferson-Schadt conditional model based on the 4F-CAPM only. Specifically, the conditional model assumes that the coefficient (beta) before the market portfolio depends linearly on a set of lagged public information variables Z_{t-1}. Recall that the functional form of the 4F-CAPM is:

$$R_{i,t} - R_t^f = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{i,t}; \qquad (6.4)$$

In the model of Ferson and Schadt (1996), the market beta depends linearly on the information variables:

$$\beta_{1i} = \hat{b}_0 + B' z_{t-1} + \varepsilon^b_{i,t}; \tag{6.5}$$

where $\hat{b}_0$ is the unconditional beta, z_{t-1} is a vector of deviations of the lagged information variables Z_{t-1} from its unconditional mean, and B' is a vector of coefficients. The remaining variables are as described earlier in Equation 6.3. According to Ferson and Schadt (1996), the information variables are: (i) the 1-month Treasury bill yield; (ii) the dividend yield of the CRSP value-weighted stock index, which is calculated as the sum of dividend payments on the CRSP value-weighted index for the previous 12 months, divided by the price level of the index at the end of the previous month; (iii) the term spread, which is calculated as the difference in yield between the 3-month Treasury bill and the 10-year Treasuries; (iv) the default spread, which is proxied by the difference between yields on Moody's Baa-rated and Aaa-rated corporate bonds. Data on these information variables are obtained from CRSP. A simple calculation yields the following functional form of the conditional beta model:

$$\begin{aligned} R_{i,t} - R^f_t =& \alpha_i + \hat{b}_0(R^m_t - R^f_t) + \beta_{2,i} SMB_t + \beta_{3,i} HML_t + \beta_{4,i} MOM_t \\ &+ B'(z_{t-1} \cdot R^m_t) + \varepsilon_{i,t}. \end{aligned} \tag{6.6}$$

We first estimate the aggregate alpha estimate of the equal-weighted portfolio of funds using the Ferson-Schadt conditional model and the 4F-CAPM, and the results are shown in Table 6.6. Generally, the alpha estimates of the conditional model are slightly more positive relative to those based on the unconditional 4F-CAPM, but the difference is trivial. For example, The aggregate equal-weighted alpha estimates for net and gross returns are -0.82 and 0.51 for the conditional model, and -1.01 and 0.33 for the 4F-

CAPM, respectively.

Table 6.6:
Intercepts and Slopes in 4F-CAPM for Equal-Weight (EW) and Value-Weight (VW) Portfolios of Actively Managed Mutual Funds (Conditional vs. Unconditional Model)
The table provides the annualized intercepts ($12^*\hat{\alpha}$) and t-statistics for the intercepts ($t(Coef)$) estimated on an equal-weight (EW) and an value-weight (VW), net and gross returns based on the 4F-CAPM and the Ferson-Schadt conditional model, respectively. Regressions slopes ($\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$) for the corresponding factors ($RM - Rf$, SMB, HML, and MOM) are also shown in the table. It should be noted that for the market slope, $t(Coef)$ tests the null hypothesis that β is equal to 1. Net returns are returns reported in the CRSP mutual fund database, and the gross returns are net returns plus $1/12^{th}$ of a fund's expense ratio at that year end. When a fund's expense ratio for a year is missing, we replace it with the expense ratio of the fund that has the same investment style and has similar assets under management (AUM). As for those funds with zero expense ratios, we manually check some of their annual reports and confirm that they in fact have positive expense ratios. Hence, we treat zero expense ratios as missing values. Our sample covers the period from January 1984 to March 2015.

	$12^*\hat{\alpha}$ (%)						
	Net	Gross	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	R^2 (%)
EW Returns							
4F-CAPM							
$Coef$	-1.01	0.33	0.96	0.21	0.02	0.00	0.98
$t(Coef)$	-2.46	0.81	-5.22	18.16	1.54	-0.38	
Conditional model							
$Coef$	-0.82	0.51	0.96	0.20	0.00	0.00	0.98
$t(Coef)$	-2.05	1.27	-4.93	17.36	0.13	0.21	
VW Returns							
4F-CAPM							
$Coef$	-0.99	-0.01	0.96	0.09	0.00	0.00	0.98
$t(Coef)$	-2.75	-0.03	-5.34	8.99	-0.08	0.32	
Conditional model							
$Coef$	-0.80	0.18	0.96	0.08	-0.02	0.01	0.98
$t(Coef)$	-2.26	0.51	-5.54	8.08	-1.86	1.30	

6.3.2 A Re-examination of F&F's and BSW's Analysis Using the Conditional Approach

To examine whether the Ferson-Schadt conditional model provides different cross-sectional distributions of $t(\alpha)$ estimates, we redo the analysis of F&F and BSW using the Ferson-Schadt conditional model, and compare the results to those of the 4F-CAPM. Specifically, we follow F&F and perform the simulations described in Chapter 4, Section 4.4 to produce a cross-sectional distribution of $t(\alpha)$ estimates on fund returns and 10,000 empirical cross-sectional distributions of $t(\alpha)$ estimates on bootstrapped returns. The alpha estimates are estimated using the 4F-CAPM (Equation 6.3) and the conditional models (Equation 6.6), respectively. There are 4910 funds [2] in the sample covering the period from January 1984 to March 2015.

F tests on the joint significance of the coefficients in the Ferson-Schadt conditional model suggest that around 35% of the funds have at least one estimate of the parameter associated with the information variable that is statistically significant. To reduce the effect of model misspecification, we only include information variables with statistically significant parameter estimates.

Selected percentiles of the cross-sectional distribution of $t(\alpha)$ estimates on actual returns and of the averaged cross-section of $t(\alpha)$ estimates on bootstrapped returns, for the conditional model and the 4F-CAPM, are shown in Table 6.7. Panel A shows the comparison between the conditional model and the 4F-CAPM on net returns and Panel B on gross returns. This table also shows the percentages (likelihoods) of the 10,000 simulation runs that produce smaller values of $t(\alpha)$ at the selected percentiles than

[2] Since a maximum of four information variables are included in the conditional model, in order to obtain accurate alpha estimates, we include a fund in the sample if it has at least 12 months of return history

those observed for actual fund returns. Generally, the Ferson-Schadt conditional model produces similar $t(\alpha)$ estimates, and likelihoods on the selected percentiles, relative to those from the 4F-CAPM. Only for the extreme right tail percentiles, we see that the Ferson-Schadt conditional model tends to produce slightly more positive $t(\alpha)$ estimates and greater likelihoods.

Table 6.7: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2015 (Conditional vs. Unconditional Model)**

The table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. We bootstrap by resampling fund and factor returns jointly as a pair. Panel A shows the comparison of selected percentiles of $t(\alpha)$ estimates distribution between the unconditional 4F-CAPM model and the Ferson-Schadt conditional model on net returns, and Panel B shows the comparison on gross returns. The sample has 4910 funds who have at least 12 months of return history, covering the period from January 1984 to March 2015.

	Panel A: net returns						Panel B: gross returns					
	Unconditional			Conditional			Unconditional			Conditional		
Pct	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>	Act	Sim	%>
1	-3.75	-2.42	0.06	-3.75	-2.44	0.02	-2.96	-2.43	3.63	-3.04	-2.44	2.14
2	-3.24	-2.11	0.11	-3.35	-2.12	0.03	-2.51	-2.12	6.53	-2.62	-2.13	3.24
3	-2.99	-1.93	0.11	-3.06	-1.93	0.03	-2.32	-1.93	6.45	-2.33	-1.93	5.77
4	-2.86	-1.79	0.09	-2.93	-1.79	0.03	-2.10	-1.79	9.90	-2.14	-1.79	7.50
5	-2.71	-1.68	0.09	-2.77	-1.68	0.03	-1.99	-1.68	9.62	-2.02	-1.68	7.52
10	-2.23	-1.30	0.08	-2.26	-1.30	0.03	-1.54	-1.30	13.30	-1.56	-1.30	11.25
20	-1.68	-0.85	0.06	-1.66	-0.85	0.03	-0.93	-0.85	33.02	-0.92	-0.85	33.59
30	-1.25	-0.53	0.07	-1.23	-0.53	0.05	-0.53	-0.53	48.05	-0.51	-0.53	51.65
40	-0.92	-0.25	0.07	-0.89	-0.25	0.07	-0.22	-0.26	56.20	-0.19	-0.26	64.09
50	-0.62	0.00	0.05	-0.58	0.00	0.07	0.09	0.00	68.42	0.14	0.00	78.73
60	-0.33	0.26	0.02	-0.29	0.25	0.07	0.39	0.25	76.38	0.44	0.25	84.82
70	-0.01	0.53	0.02	0.06	0.52	0.32	0.72	0.53	83.41	0.80	0.52	91.55
80	0.39	0.85	0.36	0.48	0.85	1.81	1.11	0.85	89.17	1.22	0.85	96.13
90	1.00	1.31	6.54	1.13	1.30	21.05	1.74	1.30	96.27	1.88	1.30	99.08
95	1.49	1.69	19.62	1.66	1.68	49.57	2.26	1.69	98.38	2.41	1.69	99.62
96	1.62	1.80	23.60	1.88	1.80	67.72	2.41	1.80	98.56	2.65	1.80	99.81
97	1.85	1.94	39.14	2.05	1.94	71.35	2.63	1.94	99.04	2.85	1.94	99.83
98	2.12	2.13	52.23	2.33	2.13	81.34	2.92	2.13	99.44	3.14	2.14	99.89
99	2.56	2.45	69.18	2.71	2.45	84.95	3.40	2.44	99.75	3.65	2.46	99.94

We also apply the Ferson-Schadt conditional model to BSW's analysis to quantify the proportion of skilled, unskilled, and zero-alpha funds. Specifically, we estimate individual fund alphas using the Ferson-Schadt conditional model and the 4F-CAPM, and then perform the FDR technique as described in Chapter 5 Section 5.5. The results are shown in Appendix 6.A, Table 6.11. Our results suggest that the proportions of skilled, unskilled, and zero-alpha funds using the conditional model are overall similar to those using the 4F-CAPM. For example, after costs, the proportions of skilled, unskilled and zero-alpha funds are 3.2%, 29.9% and 66.9%, respectively, for the conditional model, relative to proportions of 0.8%, 32.0% and 67.2% for the unconditional 4F-CAPM. Using the FDR approach of BSW, the conditional model is able to locate a few more percentages of skilled funds.

Therefore, we confirm that, in general, the Ferson-Schadt conditional model produces similar results, either on the aggregate alpha estimates or the cross-section of $t(\alpha)$estimates, relative to those of the unconditional approach. Hence, using the conditional model does not alter the inferences of F&F and BSW on manager skill. This implies that allowing the market beta to depend linearly on the information variables may not capture its time-variation appropriately. Hence, to better control for time-varying betas, as shown in Chapters 4 and 5, we left the nature of the relation between risk factors and the information variables undefined, and implemented a non-parametric change point test to detect time-variation in the betas. More importantly, we also applied the change point test to capture possible time-variation in fund alphas, which is not considered in the Ferson-Schadt conditional model.

6.4 Bootstrap Methods in Mutual Fund Studies

6.4.1 Review of Studies Using Bootstrap Methods

According to the results on the aggregate performance of mutual funds in Section 6.2, we can see that active funds on average do not outperform the market. However, underperformance in the aggregate does not suggest there is no evidence of individual skilled fund managers. In order to track skilled and unskilled fund managers, previous mutual fund studies turn to estimation of individual fund alphas, which is more interesting than inferences on the performance of aggregate fund portfolios.

However, as argued by KTWW and F&F, the difficulty of examining individual fund performance is to control for cross-fund correlation as well as to contain luck. Previous studies that rely on p-values of individual fund alphas to identify skilled funds do not specifically control for any cross-sectional correlations in returns given that mutual funds may hold similar stocks. Therefore, KTWW use a cross-sectional bootstrap to model the empirical cross-sectional distribution of fund alphas to infer the existence of skilled and unskilled funds in the tails. More importantly, the use of bootstrap simulations does not require prior knowledge of the distribution of fund alphas, hence controlling for possible non-normalities in the cross-sectional distribution of fund alphas.

Furthermore, previous studies on mutual fund performance do not explicitly account for the role of luck in the outcomes. Given the multitude of funds and the long period of fund history included in modern mutual fund studies, there might be funds with extreme returns by chance. Therefore, in order to detect the existence of skilled and unskilled fund managers, it is crucial to distinguish manager skill from luck. The use of a cross-sectional bootstrap technique helps construct two sets of world, as shown in Chapter 4, Section 4.4, so that luck can be contained when comparing the actual cross-sectional distribution of $t(\alpha)$ estimates to the simulated ones. KTWW argue that by applying the bootstrap method, some results of Carhart (1997) are rejected, in which the standard parametric t-tests are used. Differently from the pessimistic conclusions of Carhart (1997), they find that skilled managers with sufficient skill to generate excess returns do exist, especially for growth-oriented funds. Cuthbertson et al. (2008) apply this bootstrap approach to UK equity mutual funds and confirm the existence of stock picking skill for the top performing UK funds that account for 5% to 10% of the population.

The bootstrap approach of KTWW requires resampling with replacement of the residuals obtained from regressing returns on the 4F-CAPM, and then constructing a time-series of excess returns by randomly attaching resampled residuals to factor returns. F&F improve the bootstrap technique of KTWW to account for possible cross-fund correlations. They argue that by resampling residuals for each fund separately, any potential correlation across funds in the simulation is lost. Therefore, to capture this effect, F&F resample with replacement of the fund returns and factor returns jointly as a pair, and for all funds simultaneously. They find that few funds present manager skill net of all costs.

Given the controversies in the literature, this section examines the above mentioned two bootstrap methods: bootstrap by residuals and bootstrap by pairs. The bootstrap method was first proposed by Efron (1979) and the basic idea of the bootstrap is to get inference of a population from the sample through resampling the sample data. Generally speaking, a bootstrap is useful when a theoretical distribution of a statistic is either unknown or too complex. Therefore, in our case, the bootstrap measure allows us to examine the fund performance without imposing prior assumptions on the parametric distribution of fund returns. To bootstrap regression models, one can either treat the regressors as fixed (bootstrap by residuals) and then resample from the residuals of the fitted regression model, or treat the regressors as random (bootstrap by pairs) and then resample from the observations (cf. Efron and Tibshirani (1994), Section 9, page 113). We compare these two bootstrap methods of bootstrapping regression models in this chapter to see if the results of simulations are sensitive to the choice of bootstrap techniques.

6.4.2 Simulations

To understand the consequences of applying different bootstrap methods, we redo the analysis of F&F by resampling the residuals, and compare the results to those by resampling the pairs of regressors.

6.4.2.1 Bootstrap by Residuals

First, following the methodology of KTWW, we show how bootstrap by residuals is implemented to obtain the simulated cross-sectional distribution of $t(\alpha)$ estimates. Recall that the 4F-CAPM is as follows:

$$r_{i,t} = \alpha_i + \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{i,t}; \qquad (6.7)$$

definition of the variables are as described in Equation 6.3. For each fund i, we draw with replacement of the residuals after estimating the above 4F-CAPM equation, creating a resampled time-series of residuals, $\varepsilon^b = \varepsilon_1^b, \varepsilon_2^b, \varepsilon_3^b, ..., \varepsilon_n^b$, where b is the index number for the bootstrapped sample (for example, $b = 1$ for bootstrapped sample 1) and n is the number of funds in the entire sample. Since we resample with replacement, there might be residuals that appear more than once in the bootstrapped time series. Next, a time-series of adjusted monthly excess returns for fund i ($r_{i,t}^b$) is created by adding the bootstrapped residuals ε^b to the factor returns, imposing the null hypothesis of true alpha and the intercept α_i is not added to adjusted excess returns to reflect this:

$$r_{i,t}^b = \beta_{1i}(R_t^m - R_t^f) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{i,t}^b. \qquad (6.8)$$

Next, we estimate the 4F-CAPM using the adjusted returns to generate the simulated $t(\alpha)$ estimates. This process is conducted for every fund in the sample ($n = 5006$), and is repeated 10,000 times ($b = 10,000$) for each fund. As a result, we have 10,000 sequences of cross-sectional empirical distribution of simulated $t(\alpha)$ estimates. Finally, we average across the 10,000 simulation runs to obtain the average cross-section of simulated $t(\alpha)$estimates, and we show some selected percentiles of these $t(\alpha)$ estimates in Table 6.8.

As argued in Chapter 4, Section 4.4, Equation 6.8 suggests that this series of adjusted returns has a true alpha equaling zero by construction. Therefore, there is no skill and only luck exists in the simulated samples. If the estimated $t(\alpha)$ estimates on actual monthly excess returns are far more extreme than the simulated $t(\alpha)$ estimates, we can conclude that the extreme returns we observe are not due to luck alone.

6.4.2.2 Bootstrap by Pairs

As to bootstrap by pairs, it is referred to as the bootstrap technique that resamples the fund returns and factor returns jointly as a pair to generate simulated cross-sections of $t(\alpha)$ estimates. First, we obtain alpha estimates by regressing on the actual monthly excess returns ($r_{i,t}$) using Equation 6.7. Next, we impose the null hypothesis of zero alpha by subtracting the alpha estimates from the monthly excess returns to obtain a series of adjusted returns to reflect it:

$$r_{i,t}^{ad} = r_{i,t} - \alpha_i. \tag{6.9}$$

After creating the adjusted excess returns $r_{i,t}^{ad}$, we resample from replacement of the adjusted excess returns and factor returns jointly for all funds simultaneously. By regressing the adjusted returns with the factor returns using the 4F-CAPM, we get a cross-sectional distribution of simulated $t(\alpha)$ estimates. By repeating this process 10,000 times, 10,000 cross-sections of simulated $t(\alpha)$ estimates can be obtained. Finally, we average across the 10,000 simulation runs to get the averaged simulated cross-section of $t(\alpha)$ estimates and their selected percentiles as shown in Table 6.8.

6.4.2.3 Block Bootstrap

Since we resample with replacement from the regression residuals by bootstrapping the residuals, and from the adjusted returns and factor returns jointly by bootstrapping the pairs, any correlation within the time-series is lost. To capture possible autocorrelation effects, we implement the block bootstrap that resamples with replacement of a block of observations so that correlation of time series within that block can be preserved. Here, we choose the block length equal to $T^{\frac{1}{5}}$ proposed by Hall et al. (1995). Given that we have 375 months over the period from January 1984 to March 2015, the length is approximately equal to 3. For example, suppose $s_t(t = 1, 2, 3, 4, ..., T)$ is the time series that we aim to resample from, we select observations s_1, s_2, s_3 to form the first block and s_2, s_3, s_4 to form the second block. The blocks keep moving forward until the last observation of this time-series is selected. Given $T = 375$ in our sample, we have 373 blocks in total with blocks having observations that overlap.

Next, we resample with replacement from these 373 blocks to form the bootstrapped time series s_t^b that has the same length as the original time series. Hence the bootstrapped series s_t^b will have 125 ($375/3 = 125$) blocks and 375 observations. After obtaining the bootstrapped adjusted returns ($\alpha = 0$) and factor returns, we estimate the 4F-CAPM to generate the simulated cross-section of $t(\alpha)$ estimates. Similarly, this process is repeated for 10,000 times and we average across these 10,000 simulation runs to get the average simulated cross-section of $t(\alpha)$ estimates and their selected percentiles are shown in Table 6.8.

Table 6.8: **Percentiles of Actual and Simulated $t(\alpha)$ Estimates: 1984 to 2015 (Different Bootstrap Methods)**
This table shows values of actual (Act) $t(\alpha)$ estimates at selected percentiles (Pct) of the distribution of $t(\alpha)$ estimates for net and gross returns. The table also shows the fractions of the actual $t(\alpha)$ estimates that are greater than the average values of $t(\alpha)$ estimates from the 10000 simulation runs at selected percentiles (%>). Sim is the average value of $t(\alpha)$ at the selected percentiles from the simulations. Panel A shows the comparison of selected percentiles of $t(\alpha)$ estimates distribution between bootstrap by pairs and bootstrap by residuals on net returns, and Panel B shows the comparison on gross returns. The sample has 5006 funds, covering the period from January 1984 to March 2015. $t(\alpha)$ estimates are estimated using the 4F-CAPM. It should be noted that the actual $t(\alpha)$ estimates would be the same regardless of which bootstrap method we use.

	Panel A: net returns							Panel B: gross returns						
		By pairs		By residuals		By blocks			By pairs		By residuals		By blocks	
Pct	Act	Sim	%>	Sim	%>	Sim	%>	Act	Sim	%>	Sim	%>	Sim	%>
1	-3.79	-2.46	0.01	-2.22	0.00	-2.54	0.24	-2.99	-2.47	3.40	-2.22	0.00	-2.62	13.34
2	-3.26	-2.12	0.01	-1.96	0.00	-2.19	0.29	-2.53	-2.13	5.80	-1.96	0.00	-2.24	17.13
3	-2.99	-1.93	0.01	-1.79	0.00	-1.98	0.30	-2.32	-1.93	5.84	-1.79	0.00	-2.03	15.86
4	-2.86	-1.78	0.01	-1.67	0.00	-1.83	0.18	-2.11	-1.79	8.59	-1.67	0.00	-1.88	19.90
5	-2.71	-1.67	0.01	-1.57	0.00	-1.71	0.17	-1.99	-1.68	8.46	-1.57	0.00	-1.75	19.01
10	-2.23	-1.29	0.01	-1.22	0.00	-1.31	0.07	-1.54	-1.29	11.80	-1.22	0.00	-1.35	21.24
20	-1.68	-0.84	0.02	-0.80	0.00	-0.85	0.04	-0.93	-0.85	31.68	-0.80	0.00	-0.88	38.73
30	-1.24	-0.52	0.02	-0.50	0.00	-0.53	0.07	-0.53	-0.53	48.10	-0.50	3.85	-0.54	50.60
40	-0.91	-0.25	0.03	-0.24	0.00	-0.26	0.06	-0.22	-0.26	57.56	-0.24	88.25	-0.27	57.48
50	-0.61	-0.00	0.03	0.00	0.00	-0.01	0.03	0.09	-0.01	69.44	0.00	100.00	-0.01	67.77
60	-0.32	0.25	0.04	0.24	0.00	0.24	0.05	0.38	0.25	77.57	0.24	100.00	0.25	74.43
70	0.01	0.52	0.09	0.50	0.00	0.52	0.12	0.71	0.52	84.39	0.50	100.00	0.53	80.07
80	0.40	0.84	0.40	0.80	0.00	0.84	0.84	1.10	0.84	90.30	0.80	100.00	0.86	85.14
90	1.00	1.29	6.57	1.22	0.00	1.30	10.29	1.73	1.29	97.14	1.22	100.00	1.33	92.98
95	1.49	1.68	19.20	1.57	0.23	1.70	22.66	2.26	1.68	98.61	1.57	100.00	1.74	95.37
96	1.62	1.80	22.14	1.67	5.54	1.82	25.06	2.40	1.79	98.76	1.67	100.00	1.86	95.54
97	1.85	1.94	37.98	1.79	95.76	1.97	37.86	2.63	1.94	99.10	1.79	100.00	2.01	96.64
98	2.12	2.14	50.18	1.96	100.00	2.18	47.19	2.89	2.14	99.36	1.96	100.00	2.23	97.12
99	2.56	2.49	63.18	2.23	100.00	2.53	57.57	3.40	2.49	99.67	2.23	100.00	2.60	97.80

6.4.3 A Re-examination of F&F's and BSW's Analysis Using Different Bootstrap Methods

As shown in Table 6.8, overall, bootstrap by residuals produces less negative $t(\alpha)$ estimates in the left percentiles and less positive $t(\alpha)$ estimates in the right percentiles than those from bootstrapping by pairs. We also show the likelihoods, which are the percentages of the 10000 simulation runs that produce smaller values of $t(\alpha)$ at the selected percentiles than those observed for actual fund returns, in Table 6.8. Since bootstrap by residuals tends to produce less negative $t(\alpha)$ estimates in the left percentiles and less positive $t(\alpha)$ estimates in the right percentiles, we see greater percentages (likelihoods) in the right tail and smaller percentages (likelihoods) in the left tail by resampling residuals, for both net and gross returns. This implies that using bootstrap by residuals tends to give slightly less extreme simulated $t(\alpha)$ estimates, hence leading to stronger inference on the existence of skilled and unskilled fund managers than that of bootstrap by pairs. It is consistent with the argument of F&F that resampling the fund returns and explanatory returns jointly for all funds simultaneously can capture any cross-fund correlations in the returns, whereas resampling the residuals from the fitted regression model is conducted for each fund separately, hence losing any effect of cross-fund correlations during simulations. Furthermore, percentiles of simulated $t(\alpha)$ estimates from block bootstrap are similar to those using bootstrap by pairs. Therefore, autocorrelation effect is only a minor issue for our data.

To further understand the differences in applying these bootstrap techniques, we quantify the skilled and unskilled fund managers using the FDR approach of BSW. Specifically, we apply the aforementioned two bootstrap techniques to compute p-values of individual funds: one that resamples from the residuals of the fitted regression model to estimate bootstrapped t-statistics of fund alphas, whereas the other one resamples from the fund returns and explanatory returns jointly. Details on the computation of fund p-values, proportions of zero-alpha funds (π_0), skilled funds ($\hat{T}_\gamma^+$) and unskilled funds ($\hat{T}_\gamma^-$) are shown in Chapter 5, Section 5.5. The results are shown in Appendix 6.A, Table 6.12. When the bootstrap technique that resamples the residuals is used, we see slightly greater proportions of skilled and unskilled fund managers at different γ levels, for both net and gross returns, relative to those based on bootstrap by pairs, although the difference seems to be trivial.

Furthermore, the appeal of using bootstrap by pairs in our scenario can also be explained by the underlying assumptions required by bootstrapping residuals. According to Efron and Tibshirani (1994) (cf. page 113, Chapter 9), by resampling the residuals and creating adjusted returns using Equation 6.7, one implicitly assumes that the functional form of the regression is correct. Furthermore, by randomly attaching resampled residuals to the factor returns, the errors terms should be identically distributed and not be correlated with the factor returns. This is a strong assumption that may fail in many cases. Bootstrap by pairs, however, does not require these assumptions given that it treats the regressors (factor returns) as random rather than fixed so that the functional form of the regression needs not be correct.

Despite the advantages mentioned above, it should be noted that resampling for all funds simultaneously has a cost. There is a possibility that a fund is not in the tests for the entire sample given that we resample with replacement, and for the same reason, it is also possible that a certain month's return of a fund occurs repeatedly in a simulation run. However, this effect can more or less be alleviated through considerably large numbers of simulation runs. Therefore, generally, bootstrap by pairs is more appropriate to use when exploring the finite sample behavior in the cross-section of fund returns.

6.5 Conclusions

In this chapter, we addressed three issues that are important for mutual fund studies in general and are also closely related to the research undertaken here. These issues remain controversial in the literature and motived by this, we aim to shed new light on these issues.

First, since different studies construct the portfolio of actively managed equity mutual funds according to different criteria, we examined how different ways of constructing datasets may lead to different inferences on fund performance. To do this, we constructed two datasets based on two recent studies of F&F and BSW, respectively, and further built another three panels to examine the effects of including incubated returns, new funds, and excluding funds with short return history, on the aggregate and cross-sectional distribution of alpha ($t(\alpha)$) estimates.

Our results suggest that, generally, including new funds from the sample would generate similar alpha estimates to one excluding the new funds, and not correcting for "incubation" bias gives more negative net alpha and less positive gross alpha for the EW portfolio. More importantly, using a longer minimum return history for a fund to be included in the sample gives less negative net alpha and more positive gross alpha estimates for the EW portfolio. Hence, the way BSW constructed their data in general would lead to more positive inference on aggregate performance of mutual funds. As for the cross-section $t(\alpha)$ estimates, the requirement of a longer minimum return history for a fund to be included in the sample produces more positive actual $t(\alpha)$ estimates in the right tail and less negative actual $t(\alpha)$ estimates in the left tail, for net and gross returns, hence indicating larger probabilities of funds being skilled in the right tail and smaller probabilities of funds being unskilled in the left tail of the cross-sectional distribution of $t(\alpha)$ estimates. Therefore, to avoid any biases caused by sample selection issues, we mainly follow F&F's methodology to construct data.

The second issue we examined is the widely used Ferson-Schadt conditional model based on public information variables. In line with the literature, we have shown that the aggregate alpha estimates on equal-weighted portfolio returns based on the Ferson-Schadt conditional model are similar to those using the traditional unconditional models such as the 4F-CAPM. We also performed the simulations of F&F using the Ferson-Schadt conditional model, and show that the conditional model produces similar cross-section $t(\alpha)$ estimates and likelihoods on the selected percentiles, relative to those from the 4F-CAPM. A further analysis of BSW using the FDR technique also suggests that the estimated proportions of skilled, unskilled, and zero-alpha funds in the sample using the conditional model are similar to those using the 4F-CAPM. Therefore, we argue that the reduced form of the Ferson-Schadt conditional approach may not properly capture time-variation in the market beta. This motivates us to develop a non-parametric statistical approach to detect time-variation in these regression parameters, which has been shown in detail in Chapter 4.

Finally, we compared three bootstrap methods to bootstrap regression models, which are widely used in previous mutual fund studies to explore finite sample properties in the cross-sectional distribution of fund alphas. Specifically, we performed the simulations of F&F and compared selected percentiles of these averaged simulated $t(\alpha)$ estimates using different bootstrap methods. Generally, the bootstrap approach that resamples the regression residuals tends to give less negative $t(\alpha)$ estimates in the extreme left tail and less positive estimates in the extreme right tail, relative to those using the bootstrap approach that resamples the fund returns and explanatory returns jointly for all funds simultaneously. This is consistent with the argument of F&F that resampling residuals cannot capture any cross-fund correlation in the simulation. We further examined the underlying assumptions of bootstrap by residuals. To avoid any biases caused by violations of these assumptions and to capture possible cross-fund correlations in the simulation, bootstrap by pairs is preferred in our context.

Appendix 6.A Additional Results on Estimated Proportions of Skilled, Unskilled, and Zero-alpha Funds

This appendix shows the results using the FDR techinque, specifically, Table 6.9 shows the proportions of skilled, unskilled and zero-alpha funds based on different rules of including funds in the sample; Table 6.10 shows the proportions of skilled, unskilled and zero-alpha funds based on the dataset with incubated returns and the one without incubated returns; Table 6.11 shows the proportions of skilled, unskilled and zero-alpha funds using the traditional unconditional 4F-CAPM and the Ferson-Schadt conditional model; Table 6.12 shows the proportions of skilled, unskilled and zero-alpha funds using different bootstrap methods.

Table 6.9:

Proportions of Skilled, Unskilled, and Zero-alpha Funds: 1984-2015 (Different Minimum Return Histories)

Performance of funds is estimated using the 4F-CAPM. Panel A and Panel B show the estimated proportions of zero-alpha ($\hat{\pi}_0$), unskilled ($\hat{\pi}_A^-$) and skilled ($\hat{\pi}_A^+$) funds using the FDR approach, based on a dataset of 3673 funds using the inclusion rule of 60 months, compared to those based on a dataset of 5750 funds using the inclusion rule of 8 months, for net returns. Panel C and Panel D show the comparison of results on gross returns. Each panel also shows significant funds ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$) across the distribution of funds in the left tail and right tail using four different significance levels ($\gamma = 0.05, 0.1, 0.15, 0.2$). Furthermore, significant funds are decomposed into unlucky ($\hat{F}_\gamma^-$) and unskilled ($\hat{T}_\gamma^-$) funds in the left tail, and lucky ($\hat{F}_\gamma^+$) and skilled ($\hat{T}_\gamma^+$) funds in the right tail. At the bottom of each panel, we show the FDR in the left and right tail at different significance levels. Both dataset cover the period from January 1984 to March 2015.

Panel A: net returns, 60-month rule

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	69.6 (0.02)	29.7 (0.02)	0.7 (0.00)
Number of funds	2557	1090	26

	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	14.0	19.5	24.6	28.5	7.3	5.9	4.3	2.5	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.8	3.5	5.3	7.0	7.0	5.3	3.5	1.8	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	12.2	16.0	19.4	21.5	0.3	0.6	0.8	0.7	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
FDR$^-$	12.5	17.8	21.4	24.4	95.1	88.3	80.4	71.0	FDR$^+$

Panel B: net returns, 8-month rule

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	70.5 (0.02)	27.8 (0.02)	1.7 (0.00)
Number of funds	4050	1600	100

	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	17.9	23.4	28.3	32.1	7.9	6.6	5.2	3.5	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00		
Unlucky $\hat{F}_\gamma^-$(%)	1.8	3.5	5.3	7.0	7.0	5.3	3.5	1.8	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
Unskilled $\hat{T}_\gamma^-$(%)	16.1	19.9	23.0	25.1	0.8	1.4	1.7	1.7	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00		
FDR$^-$	9.9	15.0	18.7	21.9	89.4	79.5	68.0	50.4	FDR$^+$

continued

Table 6.9 – *continued*

Panel C: gross returns, 60-month rule

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)	
Proportion	84.1 (0.02)	3.1 (0.01)	12.7 (0.02)	
Number of funds	3092	113	468	

	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	4.1	6.6	9.3	11.1	19.6	16.1	12.9	8.3	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	2.1	4.1	6.2	8.2	8.2	6.2	4.1	2.1	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	2.1	2.4	3.1	2.9	11.4	10.0	8.8	6.3	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	
FDR^-	50.1	62.8	66.8	74.0	42.0	38.3	31.9	24.7	FDR^+

Panel D: gross returns, 8-month rule

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)	
Proportion	85.4 (0.01)	8.4 (0.01)	6.2 (0.01)	
Number of funds	4909	481	360	

	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	8.4	11.4	14.4	16.6	18.0	15.0	12.3	8.6	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.00	0.00	0.01	0.00	0.00		
Unlucky $\hat{F}_\gamma^-$(%)	2.1	4.3	6.4	8.5	8.5	6.4	4.3	2.1	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
Unskilled $\hat{T}_\gamma^-$(%)	6.3	7.2	8.0	8.0	9.5	8.6	8.0	6.5	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.00		
FDR^-	25.5	37.4	44.4	51.5	47.4	42.8	34.7	24.8	FDR^+

Table 6.10:
Proportions of Skilled, Unskilled, and Zero-alpha Funds: 1984-2015 (Incubated Returns)
Performance of funds is measured using the 4F-CAPM. Panel A and Panel B show the estimated proportions of zero-alpha ($\hat{\pi}_0$), unskilled ($\hat{\pi}_A^-$) and skilled ($\hat{\pi}_A^+$) funds using the FDR approach, based on a dataset of 3673 funds with incubated returns, compared to those based on a dataset of 3475 funds without incubated returns, for net returns. Panel C and Panel D show the comparison of results on gross returns. Each panel also shows significant funds ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$) across the distribution of funds in the left tail and right tail using four different significance levels ($\gamma = 0.05, 0.1, 0.15, 0.2$). Further, significant funds are decomposed into unlucky ($\hat{F}_\gamma^-$) and unskilled ($\hat{T}_\gamma^-$) funds in the left tail, and lucky ($\hat{F}_\gamma^+$) and skilled ($\hat{T}_\gamma^+$) funds in the right tail. At the bottom of each panel, we show the FDR in the left and right tail at different significance levels. Both datasets cover the period from .January 1984 to March 2015.

Panel A: net returns, with incubated returns

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	69.6 (0.02)	29.7 (0.02)	0.7 (0.00)
Number of funds	2557	1090	26

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	14.0	19.5	24.6	28.5	7.3	5.9	4.3	2.5	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.8	3.5	5.3	7.0	7.0	5.3	3.5	1.8	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	12.2	16.0	19.4	21.5	0.3	0.6	0.8	0.7	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
FDR^-	12.5	17.8	21.4	24.4	95.1	88.3	80.4	71.0	FDR^+

Panel B: net returns, without incubated returns

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	68.7 (0.02)	30.7 (0.02)	0.6 (0.00)
Number of funds	2386	1066	23

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	14.2	19.7	24.5	28.4	7.1	5.7	4.3	2.4	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.7	3.4	5.1	6.9	6.9	5.1	3.4	1.7	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	12.5	16.3	19.4	21.6	0.3	0.5	0.9	0.7	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
FDR^-	12.1	17.4	21.0	24.1	96.2	90.4	79.5	71.9	FDR^+

continued

Table 6.10 – *continued*

Panel C: gross returns, with incubated returns									
	Zero alpha ($\hat{\pi}_0$)		Unskilled ($\hat{\pi}_A^-$)		Skilled ($\hat{\pi}_A^+$)				
Proportion	84.1 (0.02)		3.1 (0.01)		12.7 (0.02)				
Number of funds	3092		113		468				
	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	4.1	6.6	9.3	11.1	19.6	16.1	12.9	8.3	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	2.1	4.1	6.2	8.2	8.2	6.2	4.1	2.1	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	2.1	2.4	3.1	2.9	11.4	10.0	8.8	6.3	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	
FDR$^-$	50.1	62.8	66.8	74.0	42.0	38.3	31.9	24.7	FDR$^+$

Panel D: gross returns, without incubated returns									
	Zero alpha ($\hat{\pi}_0$)		Unskilled ($\hat{\pi}_A^-$)		Skilled ($\hat{\pi}_A^+$)				
Proportion	82.9 (0.02)		3.5 (0.01)		13.6 (0.01)				
Number of funds	2880		123		472				
	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	4.2	6.9	9.6	11.5	19.2	15.8	12.8	8.2	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	2.0	4.0	6.1	8.1	8.1	6.1	4.0	2.0	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	2.2	2.8	3.6	3.4	11.1	9.7	8.7	6.2	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	
FDR$^-$	48.4	58.7	63.1	70.6	42.0	38.4	31.6	24.5	FDR$^+$

Table 6.11:
Proportions of Skilled, Unskilled, and Zero-alpha Funds: 1984-2015 (Conditional vs. Unconditional Model)

Performance of funds is estimated using the 4F-CAPM with time-invariant parameters and the Ferson-Schadt conditional model. Panel A and Panel B show the estimated proportions of zero-alpha ($\hat{\pi}_0$), unskilled ($\hat{\pi}_A^-$) and skilled ($\hat{\pi}_A^+$) funds using the FDR approach, based on a dataset of 3673 funds using the 4F-CAPM, compared to those based on the Ferson-Schadt conditional model, for net returns. Panel C and Panel D show the comparison of results on gross returns. Each panel also shows significant funds ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$) across the distribution of funds in the left tail and right tail using four different significance levels ($\gamma = 0.05, 0.1, 0.15, 0.2$). Furthermore, significant funds are decomposed into unlucky ($\hat{F}_\gamma^-$) and unskilled ($\hat{T}_\gamma^-$) funds in the left tail, and lucky ($\hat{F}_\gamma^+$) and skilled ($\hat{T}_\gamma^+$) funds in the right tail. At the bottom of each panel, we show the FDR in the left and right tail at different significance levels. The sample covers the period from January 1984 to March 2015.

Panel A: net returns, unconditional model

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportions	67.2 (0.02)	32.0 (0.02)	0.8 (0.00)
Number of funds	2469	1176	28

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	14.0	19.5	24.6	28.5	7.3	5.9	4.3	2.5	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.7	3.4	5.0	6.7	6.7	5.0	3.4	1.7	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	12.3	16.1	19.6	21.8	0.6	0.9	1.0	0.8	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
FDR^-	12.0	17.2	20.5	23.6	91.8	85.3	77.6	68.6	FDR^+

Panel B: net returns, conditional model

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	66.9 (0.02)	29.9 (0.02)	3.2 (0.00)
Number of funds	2458	1097	118

Impact of Luck in the Left and Right Tails

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	13.9	19.3	24.1	27.9	9.5	7.8	5.8	3.8	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.7	3.3	5.0	6.7	6.7	5.0	3.3	1.7	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	12.2	15.9	19.0	21.2	2.8	2.8	2.5	2.2	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
FDR^-	12.0	17.4	20.8	24.0	70.2	64.0	57.4	43.6	FDR^+

continued

Table 6.11 – *continued*

Panel C: gross returns, unconditional model									
	Zero alpha ($\hat{\pi}_0$)		Unskilled ($\hat{\pi}_A^-$)		Skilled ($\hat{\pi}_A^+$)				
Proportion	80.3 (0.01)		7.1 (0.01)		12.6 (0.01)				
Number of funds	2951		261		461				
	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	4.1	6.6	9.3	11.1	19.6	16.1	12.9	8.3	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	2.0	4.0	6.0	8.0	8.0	6.0	4.0	2.0	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	2.1	2.5	3.2	3.1	11.6	10.1	8.9	6.3	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	
FDR^-	48.9	61.2	65.1	72.1	40.9	37.3	31.1	24.1	FDR^+
Panel D: gross returns, conditional model									
	Zero alpha ($\hat{\pi}_0$)		Unskilled ($\hat{\pi}_A^-$)		Skilled ($\hat{\pi}_A^+$)				
Proportion	69.8 (0.02)		5.1 (0.01)		25.1 (0.01)				
Number of funds	2565		187		921				
	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	4.3	6.6	8.8	11.0	22.7	19.5	15.9	11.1	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	
Unlucky $\hat{F}_\gamma^-$(%)	1.7	3.5	5.2	7.0	7.0	5.2	3.5	1.7	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	2.6	3.1	3.6	4.0	15.7	14.3	12.4	9.4	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	
FDR^-	40.3	53.0	59.2	63.5	30.8	26.9	21.9	15.7	FDR^+

Table 6.12:
Proportions of Skilled, Unskilled, and Zero-alpha Funds: 1984-2015 (Different Bootstrap Methods)

Fund alphas are estimated using the 4F-CAPM over the entire sample period from January 1984 to March 2015. Panel A and Panel B show the estimated proportion of zero-alpha ($\hat{\pi}_0$), unskilled ($\hat{\pi}_A^-$) and skilled ($\hat{\pi}_A^+$) funds using the FDR approach, using the bootstrap technique that resamples from the residuals compared to those using the bootstrap technique that resamples from the fund returns and explanatory returns jointly for all funds simultaneously. Panel C and Panel D show the comparison of results on gross returns. Each panel also shows significant funds ($\hat{S}_\gamma^-$, $\hat{S}_\gamma^+$) across the distribution of funds in the left tail and right tail under four different significance levels ($\gamma = 0.05, 0.1, 0.15, 0.2$). Further, significant funds are decomposed into unlucky ($\hat{F}_\gamma^-$) and unskilled ($\hat{T}_\gamma^-$) funds in the left tail, and lucky ($\hat{F}_\gamma^+$) and skilled ($\hat{T}_\gamma^+$) funds in the right tail. At the bottom of each panel, we show FDR rate in the left and right tail at different significance levels.

Panel A: net returns, bootstrap by residuals

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	67.2 (0.02)	32.0 (0.02)	0.8 (0.00)
Number of funds	2469	1176	28

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	14.0	19.5	24.6	28.5	7.3	5.9	4.3	2.5	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.7	3.4	5.0	6.7	6.7	5.0	3.4	1.7	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	12.3	16.1	19.6	21.8	0.6	0.9	1.0	0.8	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
FDR$^-$	12.0	17.2	20.5	23.6	91.8	85.3	77.6	68.6	FDR$^+$

Panel B: net returns, bootstrap by pairs

	Zero alpha ($\hat{\pi}_0$)	Unskilled ($\hat{\pi}_A^-$)	Skilled ($\hat{\pi}_A^+$)
Proportion	69.6 (0.02)	29.7 (0.02)	0.7 (0.00)
Number of funds	2557	1090	26

	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	14.0	19.5	24.6	28.5	7.3	5.9	4.3	2.5	(%)Signif.$\hat{S}_\gamma^+$
	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	1.8	3.5	5.3	7.0	7.0	5.3	3.5	1.8	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	12.2	16.0	19.4	21.5	0.3	0.6	0.8	0.7	(%)Skilled $\hat{T}_\gamma^+$
	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	
FDR$^-$	12.5	17.8	21.4	24.4	95.1	88.3	80.4	71.0	FDR$^+$

continued

Table 6.12 – continued

Panel C: gross returns, bootstrap by residuals									
	Zero alpha ($\hat{\pi}_0$)		Unskilled ($\hat{\pi}_A^-$)		Skilled ($\hat{\pi}_A^+$)				
Proportion	80.3 (0.01)		7.2 (0.01)		12.5 (0.01)				
Number of funds	2951		261		461				
	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	4.1	6.6	9.3	11.1	19.6	16.1	12.9	8.3	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	2.0	4.0	6.0	8.0	8.0	6.0	4.0	2.0	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	2.1	2.5	3.2	3.1	11.6	10.1	8.9	6.3	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	
FDR$^-$	48.9	61.2	65.1	72.1	40.9	37.3	31.1	24.1	FDR$^+$
Panel D: gross returns, bootstrap by pairs									
	Zero alpha ($\hat{\pi}_0$)		Unskilled ($\hat{\pi}_A^-$)		Skilled ($\hat{\pi}_A^+$)				
Proportion	84.1 (0.02)		3.1 (0.01)		12.7 (0.02)				
Number of funds	3092		113		468				
	Impact of Luck in the Left and Right Tails								
	Left Tail				Right Tail				
Signif.Level(γ)	0.05	0.10	0.15	0.20	0.20	0.15	0.10	0.05	Signif.Level(γ)
Signif.$\hat{S}_\gamma^-$(%)	4.1	6.6	9.3	11.1	19.6	16.1	12.9	8.3	(%)Signif.$\hat{S}_\gamma^+$
	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	
Unlucky $\hat{F}_\gamma^-$(%)	2.1	4.1	6.2	8.2	8.2	6.2	4.1	2.1	(%)Lucky $\hat{F}_\gamma^+$
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Unskilled $\hat{T}_\gamma^-$(%)	2.1	2.4	3.1	2.9	11.4	10.0	8.8	6.3	(%)Skilled $\hat{T}_\gamma^+$
	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.00	
FDR$^-$	50.1	62.8	66.8	74.0	42.0	38.3	31.9	24.7	FDR$^+$

Chapter 7

Conclusions

This study set out to explore the central question of whether there is a group of skilled and unskilled fund managers, which has generated much debate over the years and is of considerable importance for both academic research and for practical decision-making. Motivated by the contradictory conclusions on the existence of skilled managers in previous studies, we revisited this issue and attempted to make improvements to several existing problems prevalent in the literature. In this concluding chapter, we summarize our main findings and their implications, and show what existing views we challenged. The final part of the chapter discusses several limitations of our study, and provides some ideas for future research.

7.1 Main Findings and Implications

First, we extended the study of Fama and French (2010) in Chapter 4, by addressing the time-variation in the regression parameters of factor models, which is considered a "thorny" problem in Fama and French (2010). We estimated individual fund alphas to form a cross-sectional distribution of $t(\alpha)$ estimates, and constructed two sets of worlds through bootstrap simulations: one referred to as the "real world" where skill and luck coexist and $t(\alpha)$ estimates are obtained from regressing on fund returns; the other one referred to as the "simulated world" in which true skill is assumed to be zero by setting the true α to zero to reflect. By comparing selected percentiles of the cross-sectional distribution of $t(\alpha)$ estimates in these two worlds, we show that the simulated runs generate far smaller $t(\alpha)$ estimates relative to those on actual returns at selected percentiles, hence there are managers who are able to generate excess returns net of all costs, which cannot be merely explained by luck.

Differently from the studies of Kosowski et al. (2006) and Fama and French (2010), we developed a novel approach to address the time-variation issue. By implementing a non-parametric change point test, we are able to capture any form of time-variation in the regression parameters. Results from the test contain information on which parameter changes and on the estimated date of the change. Using this information, we further constructed a multi-regression model that enables us to estimate fund alphas using separate regressions on subintervals of the fund return series that are divided according to the dates of the changes. Results from these regressions and simulations show that for net returns, there are skilled fund managers located in the right tail above the 90^{th} percentile. For gross returns, the actual $t(\alpha)$ estimates are greater than those from simulations in more than 70% of simulation runs for all right tail percentiles, especially for those above the 90th, which show that actual $t(\alpha)$ estimates beat the average values in more than 99% of 10,000 simulation runs.

This implies strong evidence of manager skill for both net and gross returns. By further comparing the results using our multi-regression model with time-varying parameters to those using the traditional versions of the CAPM, we show that our model significantly improves the performance of funds in our sample. After conducting the simulations of Fama and French (2010) based on their sample period from January 1984 to September 2006, and comparing the results using our multi-regression model to theirs, we show that the pessimistic results of Fama and French (2010) are mainly due to their assumption of time-invariant alpha and betas.

Second, to further quantify the proportions of skilled and unskilled managers, an FDR technique is implemented in Chapter 5 and allows us to control for false positives, i.e. those truly zero-alpha funds that are falsely identified as funds with positive alphas. Results show that 8.4% of the funds in the sample are skilled and 34.1% of the funds are unskilled funds. Differently from the study of Barras et al. (2010), we implemented the aforementioned change point test and the multi-regression model to control for time-variation in the regression parameters. Our results overturn most conclusions of Barras et al. (2010) where the estimated proportion of skilled funds is 0.7% in their sample.

The analysis in Chapters 4 and 5 challenged the standard assumption of time-invariant alpha and betas in the traditional models, including the 3F-CAPM and the 4F-CAPM. Indeed, our results from the change point test suggest that around 20% of funds in our sample have at least one change in either alpha or betas. Hence, inferences on the existence of skilled managers using the traditional models in previous studies can be biased. Furthermore, we have shown that the Ferson-Schadt conditional model produces similar alpha estimates compared to those using the traditional models, hence their way of capturing time-variation in the market beta may be inappropriate and it is better to leave the nature of the relationship between the market beta and the information variables undefined. Previous studies on extended models to capture time-varying parameters usually specify the source of time-variation, such as changing business cycles and economic conditions, whereas our multi-regression model with time-varying alpha and betas is able to detect multiple changes in the parameters without specifying the functional form of the nature of the time-variation, which is more powerful and has wider applications in addressing time-variation issues in mutual fund studies.

Third, we have seen a significant difference between results on net and gross returns in Chapters 4 and 5. Generally, results on gross returns are much more promising than those based on net returns. For example, we have shown in Chapter 4 that, when returns are measured after trading costs only, the estimated proportion of skilled managers is 19.8%; this proportion decreases dramatically to 8.4% when returns are measured after all expenses and costs. This is reasonable given that manager skill defined on gross returns mainly requires managers to cover the trading costs, whereas for net returns, fund mangers are identified as skilled when they have sufficient skill to cover all expenses and costs.

Recalling that gross returns are returns with expense ratios added back to net returns, this suggests that funds' expense policy plays an important role in the underperformance of funds. Some fund managers may have enough skill to generate excess returns, however, fund companies overcharge so that these funds generally underperform after all expenses and costs. Although we have shown in the summary statistics of data that the assets-weighted expense ratios are declining over the years, this is mainly due to the decline of expense ratios of aggressive growth funds. Expense ratios of growth funds and growth & income funds remain rather constant over these years, and these two kinds of funds account for more than 90% of the entire sample. Therefore, there could be room for improvement in fund performance if expenses can be reduced. Our results on net and gross returns provide a view about the relationship between mutual fund expenses and fund performance, which practitioners and policy makers may find useful.

Fourth, when fund performance is examined for different investment styles in Chapter 5, our results show that aggressive growth funds have the highest proportion of skilled managers net of all expenses and costs. We also find that growth & income funds outperform the other two styles of funds when returns are measured before expenses, i.e. when examination of fund performance is disentangled from funds' expense policy. This is contrary to the results of Barras et al. (2010), who argue that no growth & income funds are skilled. Our results are reasonable given that growth funds usually have more turnover, thus more transaction costs than income-generating funds. Therefore, when returns are measured after transaction costs only, i.e. fund managers are required to cover trading costs to achieve a zero-alpha, growth & income funds may perform better than growth-oriented funds.

Fifth, in Chapter 6, by estimating fund alphas using different datasets based on different sample selection rules, we show that including incubated returns of funds produces slightly smaller individual fund alpha estimates compared to the case where incubated returns are excluded. This is contradictory to the findings of Fama and French (2010) and Evans (2010), who argue that funds have enhanced performance in their incubated periods as incubated funds are provided with seed money from their management companies to establish appealing return history. We argue that our results are reasonable given that incubated funds are usually in the early stage of their return history, and are vulnerable to risks and distresses, hence they could underperform relative to the other funds. Furthermore, we have shown that using a longer minimum return history funds must have before they are included, produces a larger alpha estimate for an aggregate equal-weighted portfolio, and shifts the cross-sectional distribution of $t(\alpha)$ estimates to the right. This implies that the requirement of a longer minimum return history would induce more survivor-bias given that funds with short return histories are excluded from the sample. Therefore, different criteria in selecting funds could lead to contradictory inferences on fund performance. Indeed, we have shown that the results of Barras et al. (2010) would be less promising if they chose to include funds with minimum 8 months of return history, instead of 60 months. For example, the estimated proportion of skilled funds after trading costs is 12.7% using the 60-month rule, compared to 6.6% using the 8-month rule.

Finally, regarding different bootstrap techniques used in previous studies, our results in Chapter 6 show that the bootstrap technique that resamples from regression residuals tends to produce less extreme simulated $t(\alpha)$ estimates, although the difference is trivial for our data. Bootstrap by residuals requires that the functional form of the regression model is correct. In particular, the resampling of residuals implicitly requires that the regression residuals should be uncorrelated with the explanatory returns and should not be correlated with each other. This is a strong assumption that can easily fail if heteroskedasticity is prevalent in the data, hence leading to poor finite sample behaviour of the bootstrapped distribution of $t(\alpha)$ estimates. Therefore, to avoid this strong assumption and to capture any cross-fund correlation in the simulations, the bootstrap technique that resamples from fund returns and explanatory returns jointly as a pair for all funds simultaneously is suggested in applications. Furthermore, results using the bootstrap technique that resamples from blocks of residuals suggest that autocorrelation is a minor issue in our study.

7.2 Limitations and Future Research

Despite all the aforementioned contributions to the literature, our study has several limitations. First, as explained in the background chapter, there are two major approaches to measure manager skill: returns-based approaches and portfolio holdings-based approaches. Our study is mainly based on historical returns. Although returns are of much higher frequency than portfolio holdings data and are easier to obtain, portfolio holdings-approaches allow a security-level examination and a more precise construction of benchmarks. As shown in our study, fund managers may have time-varying skill and changing risk loadings, hence, information on portfolio holdings would provide insights on the dynamic trading strategies of fund managers. Furthermore, a portfolio holdings based approach allows an examination of fund performance before all costs, including trading costs. Therefore, it is worth developing more powerful approaches that combine historical returns and portfolio holdings together in the evaluation of mutual fund performance when more detailed portfolio holdings data are available.

Another issue that could be addressed in future research is to explore the role of expenses in fund performance. As we have shown in the results on net and gross returns, alpha estimates obtained using gross returns improve dramatically relative to those using net returns. This implies that funds' expenses play an important role in the underperformance of funds. Hence, it would be of importance to further analyze the components of expenses so as to figure out which source of expenses mainly impacts the performance of mutual funds.

Finally, although we have shown that there is a group of fund managers who are able to outperform the market, it is surprising that we also see a large number of fund managers who underperform the passive benchmarks. As shown in Chapter 5, the proportion of unskilled funds tends to increase over time. Given that investors should have enough information to identify these unskilled funds over time and then switch to other funds, it makes the existence of these unskilled funds puzzling. Previous studies argue that investors may be unsophisticated, or that some investors are restricted to move to other funds. Hence, further analysis of the behavior of investors and the constraints faced by them when investing in mutual funds would help solve this puzzle.

Appendix A

Construction of CRSP U.S. Survivor-bias Free Actively Managed Equity Mutual Fund Dataset

This appendix documents the details on how to construct a panel of equity U.S. mutual funds that are actively managed. Specifically, Section A.1 will list the variables used to construct the dataset. Section A.2 explains the way to identify U.S. equity mutual funds. Section A.3 provides the detailed steps to exclude those passively managed funds. Section A.4 assigns portfolio identifiers to single funds with different shareclasses. Section A.5 merges funds characteristics variables with return history. Section A.6 creates valued weighted net returns for a single fund, then funds are categorized into three investment styles and gross returns are generated. Section A.7 further screens funds to reduce incubation bias effects. Finally, we show some statistics of our data in Section A.8.

A.1 Summary of Variables from CRSP

We list all the variables that are used to construct the panel of active U.S. equity mutual funds in Table A.1. A brief description is also provided in the table. It should be noted that, some variables do not span our entire sample period (January 1984 to March 2015) so that we also mark their availability.

Table A.1: Summary of Variables from CRSP

Variable Name	Description of the Variables	Variable type	Availability
crsp_funno	Unique identifier for a fund shareclass	Num	
caldt	Date	Date	
fund_name	Fund name	Char	
crsp_portno	Unique identifier for a fund	Num	
retail_fund	Identify if a fund is a retail fund	Char	12/1999-
index_fund_flag	Identify if a fund is an index fund	Char	06/2008-
inst_fund	Identify if a fund is an institutional fund	Char	12/1999-
et_flag	Identify if a fund is an ETF or ETN	Char	
per_com	The proportion of fund's investments into common stocks	Char	
wbrger_obj_cd	Wiesenberger objective code	Char	1984-1993
policy	Main type of securities held by fund	Char	1984-1989
lipper_class	Lipper classification code	Char	12/1999-
lipper_obj_cd	Lipper objective code	Char	12/1998-
crsp_obj_cd	CRSP objective code	Char	
si_obj_cd	Strategic insight objective code	Char	1993-1998
exp-ratio	Annual expense ratios of a fund	Num	
mtna	Monthly Total Net Assets	Num	
mnav	Monthly Net Asset Value	Num	
mret	Total return per share as of month end	Num	

A.2 Selecting the U.S. Equity Mutual Funds

Since we focus on a portfolio of U.S. equity mutual funds to measure fund performance, all other funds should be excluded from our sample, including mortgage-backed funds, balanced funds, international funds, global funds, municipal funds, bond funds, money market funds and funds that invest in precious metals, as these kinds of funds generally invest only a small portion into U.S. equities. In F&F, four variables are used to identify U.S. equity mutual funds:

1. *wbrger_obj_cd*: Weisenberger fund types, available from 1962 to 1993, but is argued to be imprecise in French (2008);

2. *policy*: Policy codes, which are not available after 1990;

3. *si_obj_cd*: Strategic insights objective codes, previously known as Standard & Poor's objective codes, only available after 1993. They are argued to be the most reliable in French (2008);

4. Area codes: They are no longer available in the CRSP database.

Except for area codes, *wbrger_obj_cd*, *policy* and *si_obj_cd* still remain in the CRSP database. Furthermore, there are two new variables on fund style and objectives that are available in CRSP now. In our study, we use the following five variables, including the above three available variables, to select the U.S. equity mutual funds out of all the funds in the CRSP database, namely *wbrger_obj_cd*, *si_obj_cd*, *lipper_obj_cd*, *crsp_obj_cd* and *policy*. As described in Table A.1, not a single variable of them spans the entire sample period, hence our selection of equity funds is based on the following steps:

1. Keep only observations that have at least one of the above 5 variables that is not missing;

2. Funds that belong to at least one of the categories in Table A.2 are included in our sample database;

3. According to French (2008), the most reliable source of fund objectives are strategic insight codes. Therefore, we exclude any funds that are declared as non-U.S. or non-equity by strategic insight codes.

Table A.2: Fund Style or Objective Names Used to Identify U.S. Equity Mutual Funds

Weisenberger objective codes (*wbrger_obj_cd*)	G, GCI, IEQ, LTG, MCG, SCG
Strategic insight objective codes (*si_obj_cd*)	AGG, GMC, GRI, GRO, ING, SCG
Lipper objective codes (*lipper_obj_cd*)	CA, EI, G, GI, MC, MR, SG
CRSP style codes (*crsp_obj_cd*)	all that begin with ED or EDC or EDY, exclude EDYH, EDYS
Policy codes (*policy*)	CS, Flex, I-S, TF, TFE

After the selection, 21459 shareclasses remain in the sample. It should be noted that, there is one set of variables that tells the percentage of a fund's assets invested in a certain instrument in the CRSP database. For example, *per_com* tells the percentage of a fund's assets invested in common stocks. Previous studies use this variable to select U.S. equity mutual funds, but with different standards. For instance, Evans (2010) requires that only funds with an average of 90% or more assets held in common stocks can be included in the sample based on the variable *per_com* provided in CRSP. However, Kacperczyk et al. (2008) exclude funds that invest less than 80% or more than 105% of their assets in common stocks. Furthermore, the variable *per_com* is not complete throughout the sample period, with lots of missing observations. We therefore use a more moderate standard to exclude funds as stated in Kosowski et al. (2006). Specifically, we exclude funds that have less than 50% of their assets invested in domestic equities during the majority of their return history.

Apart form the use of *per_com*, we mainly select equity funds based on their self-declared investment objectives and fund names. It should be noted that there are missing values for funds' self-declared investment objectives. Some studies replace missing values with the prior non-missing value and then select funds based on complete self-declared investment styles. For example, Barras et al. (2010) carry forward the prior non-missing objective if an investment objective is missing. However, funds can change their investment style, and according to our observations, there are 1835 fund shareclasses that changed their investment styles during their lifetime, accounting for around 9% of all equity shareclasses. Therefore, we only include a fund for a certain year if we are sure it is a U.S. equity fund for that year, i.e. the fund should have a self-declared investment style indicating it invests primarily in U.S. common stocks at that fiscal year end.

A.3 Excluding Index Funds

To focus on active management, we need to exclude those funds that are passively managed. After section A.2, we have all the U.S. equity mutual funds. Among these equity funds, some are passively managed, mainly index funds and ETFs (Exchange-traded funds). CRSP does provide a variable named *index_fund_flag* to identify index funds, with character "D" representing pure index fund, character "E" representing enhanced index funds and character "B" representing index-based funds. Another variable named *et_flag* is to identify ETFs (Exchange Traded Funds) or ETNs (Exchange Traded Notes). These funds are normally based upon the performance of a market index and are passively managed, so we need to exclude them from our sample. However, these two variables are only available from the year 2003, which does not fully cover our sample period. Furthermore, they are incomplete with missing values. Therefore, we select actively managed funds based on variables *index_fund_flag* and *et_flag*, as well as occurrences of key words in fund names. To summarize, there are overall five kinds of funds included in our dataset so far:

1. Actively managed funds: Funds that are actively managed to achieve a higher return to beat the market;

2. Ordinary index funds: Funds that are passively managed to mimic the unleveraged performance of a benchmark index, such as the S&P 500. For brevity, we classify ETFs and ETNs as ordinary index funds as well;

3. Leveraged index funds: Funds that aim to mimic the leveraged performance of a benchmark index, for example, to achieve 300% of daily return of the S&P500;

4. Target-Date funds: Funds that automatically choose an asset portfolio of stocks, bonds and cash equivalents based on a selected time period that is appropriate for investors. For example, funds that target the date of retirement;

5. Enhanced index funds: Funds that select securities based on a broad-based index, such as the S&P 500. Their managers apply the core theory of index investing,

but usually go further to outperform the tracking index. Whether enhanced index funds are actively managed or not is still a debate.

To better focus on active management, we exclude ordinary index funds, leveraged index funds, target-date funds and enhanced index funds from our dataset. First, ordinary index funds are set to mimic the performance of a market index to achieve low or zero management expense ratio, so they are typically passively managed funds. Second, instead of mimicking index returns like ordinary index funds, leveraged index funds use derivatives to deliver a relatively higher investment exposure. Since they operate differently from other normal equity mutual funds, we do not include them in our actively managed funds sample. As for target-date funds, some of these can be actively managed to some extent, however the identifiers of actively managed target-date funds are not available in the CRSP database. Therefore we exclude this kind of index funds from our sample. For similar reasons, we exclude enhanced index funds as they are possibly semi-actively managed, which means managing enhanced index funds involves limited stock picking skills. Therefore we prefer to exclude them from our actively managed funds sample.

To exclude the above passively managed funds from our sample, we need to identify them first. Table A.3 shows the key words we use to identify index funds:

Table A.3: Key Words to Identify Index Funds

index	indx	idx	s& p	schwab 1000
nasdaq-100	dow	jones	etf	ishare
profound	russell	proshare	powershare	viper
spider	spdr	wilshare	etn	
exchange traded		exchange-traded		

Generally, any funds that have these key words in their names are very likely to be index funds and hence are passively managed. However, there are exceptions. We put together another list of key words in Table A.4, and any funds that have these key words in their names are more likely to be enhanced index funds even if they have some of the key words in Table A.3:

Table A.4: Key Words to Reject an Index Fund as a Passively Managed Fund

hedge	manage	enhance	plus

Next, we identify target-date funds based on the key words in Table A.5:

Table A.5: Key Words to Identify Target-date Funds

1970	1975	1980	1985	1990	1995
2000	2005	2010	2015	2020	2025
2030	2035	2040	2045	2050	2055
2060	target	lifestyle			

The key words to identify leveraged index funds from ordinary index funds are shown in Table A.6:

Table A.6: Key Words to Identify Leveraged Index Funds

inverse	short	ultra	1.25x	1.5x	2x
2.5x	3x	4x	5x	6x	7x
8x	9x	0x			

Any funds that have some of these key words in Table A.6 are likely to be leveraged index funds. However, if a fund contains any of the following key words in Table A.7, it is not a leveraged index fund even if it has some of the key words in Table A.6:

Table A.7: Key Words to Reject a Fund as a Leveraged Index Fund

short term	short tm	short bond	short bnd	short bd	lg short
long/short	long-short	long short			

Therefore, we identify a fund as a leveraged index fund if the fund contains at least one word from Table A.6 but does not contain any word from Table A.7.

To summarize, we identify funds in our dataset following the rules below:

1. We label a fund as a target-date fund if the fund

    (a) contains a word in Table A.5;

(b) does not contain the word “russell 2000”, since “2000” is in Table A.5, but funds with the key word“russell 2000” may be funds that mimic the performance of the russell 2000 index.

2. We label a fund as a leveraged index fund if the fund

   (a) contains a word in Table A.6;

   (b) does not contain any word from Table A.5 or Table A.7.

3. We label a fund as an enhanced index fund if the fund

   (a) contains a word in Table A.3;

   (b) contains a word in Table A.4.

4. We label a fund as an ordinary index funds if the fund

   (a) contains a word in Table A.3;

   (b) does not contain any word from Table A.4, Table A.5, Table A.6 or Table A.7.

5. If the fund is not labeled as the above four types, then we identify it as an actively managed fund.

It should be noted that we assign labels to each observation of the funds, which means there might be cases when the fund changed its index status during its lifetime, i.e. multiple labels are assigned to the same fund. This is arguably sensible given that some funds may change their investment strategies. After identifying index funds, we have some suspect funds that

1. are not labeled as ordinary index funds by our key words but are suggested to be index funds by the CRSP variable *index_fund_flag* with character “D”;

2. are not labeled as enhanced index funds by our key words but are suggested to be enhanced index funds by the variable *index_fund_flag* with character "E";

3. are not labeled as either index or enhanced index funds by our key words but are suggested to be index-based funds by the variable *index_fund_flag* with character "B";

4. are not labeled as index funds by our key words but are suggested to be ETFs by the variable *et_flag* with character "F";

5. are not labeled as index funds by our key words but are suggested to be ETNs by the variable *et_flag* with character "N";

6. are labeled as ordinary index funds by our key words but suggested to be non-index funds by the variable *index_fund_flag* with null character.

The above six types of funds are excluded from our sample temporarily. Then, we store all these funds in a separate file and do robustness checks using the popular stock indexes that index funds usually target. For those suspect funds and all the active managed funds identified according to our classification explained above, we perform manual checks. We also calculate the correlation between their historical monthly returns and returns of those target stock indexes, such as S&P 500. If a correlation of more than 99% is found, then we consider the fund as an index fund and exclude it from our sample. It should be noted that we leave the robustness check to be done after we merge the fund characteristic variables with returns in Section A.4.

After robustness checks, several funds are known to be misspecified as index funds by our standards and are added back to our sample. These funds usually have the key word “russell” or “dow” or “jones”. However, they are funds from “Russell Investment Company” or “Endowments Trust Company” or “Badson Shadow Stock Fund” or “Jones Villalta Opportunity Fund” and are in fact actively managed. Some funds that failed to be identified as index funds by our standards but are in fact passively managed are excluded from our sample. These funds usually have a correlation with the target stock index of more than 99%.

A.4 Merging in Historical Returns

To merge the selected U.S. actively managed equity mutual fund shareclasses with the monthly returns of these shareclasses into a single file, we focus on return history of each shareclass, i.e. only merge in observations of the year/month when return data is available. For variables such as investment objectives that are only available on a yearly or quarterly basis, we fill in the monthly observations based on their year-end information. As explained in French (2008), mutual fund shareclasses are selected only if they are U.S. and equity shareclasses for that specific year for certain. For those shareclasses that have quarterly information on investment styles, we only merge in observations of the monthly returns if we are sure it is a U.S. equity fund shareclass by the end of that quarter. For example, if a shareclass has an investment style for the first two quarters of a certain year indicating that it is a U.S. active equity shareclass, then only returns of the first six months are merged into our sample since we are not sure about its investment style for the remaining two quarters of that year.

For monthly returns, they need not be contiguous, so missing returns [1] are excluded

[1] It should be noted that, most missing returns are denoted as “.” in CRSP, however, there are some missing returns denoted as “R” that need to be excluded as well.

from our sample. We also exclude the following-month return when we observe a missing return as CRSP fills this with the cumulated return using the last non-missing Net Asset Value (NAV). [2] [3]

A.5 Shareclasses and Portfolios

As the mutual fund industry evolves, funds start to provide different classes of shares so as to attract investors with different preferences. Conferring ownership under the same fund management company, the main differences among these shareclasses are fees and expenses charged. Generally speaking, shareclasses that attract long-term investors may offer a high load fee and a low expense ratio, while a low load fee and a high expense ratio are often offered to attract short-term investors. Usually, shareclasses are categorized into "Class A", "Class B", "Class C", etc. for funds that have more than one shareclass. Class A shares typically charge a front-end fees, Class B shares tend to impose a contingent deferred sales charge, which means investors need to pay if they sell their shares within a certain period, and Class C shares normally have no front-end sales charge, but only small charges if investors sell their shares within short periods.

Since performance should be measured at the fund level to infer the existence of skilled or unskilled fund managers, we need to merge different shareclasses into a single fund. A new portfolio code called *calc_portcode* is assigned to different shareclasses of a fund in our sample. This variable should be similar to *crsp_portno* provided by CRSP, but more

[2] Net Asset Values (NAV) per share is computed based on the closing market prices of that shareclass, not to be confused with Total Net Asset Values (TNA), which is the total amount of all investor dollars invested in that shareclass.

[3] According to *Survivor-bias-free US Mutual Fund Guide* (2014), on page 14 section of "Notes about returns ", monthly return values are calculated as a change in NAV including reinvested dividends from one period to the next. Therefore, if there is a missing NAV, both the return for that month and the following-month return should be excluded.

complete and uniquely identifies a portfolio/fund.

For funds that have more than one shareclass, their names in CRSP usually have two segments. The first segment contains the actual name of the fund, while the second segment contains the name of the specific subclass, and these two segment are often separated by the symbol " ; " or " / ". For example, the fund named as "Spectra Fund; Class N Shares" has the first segment "Spectra Funds" which is the fund name, and after the symbol ";", it has the second segment "Class N Shares" as the shareclass name. Generally, subclasses of the same fund should have identical names in their first segments. So one can build an algorithm to group together different shareclasses based on their first segments of their names. Specifically, the following steps are conducted to assign the portfolio codes to subclasses of the same fund:

1. Exclude observations with missing fund names;

2. Modify the shareclass names for each year and update them to the newest name so that for the same shareclass it has only one name throughout its entire return history. It is common that the shareclass may change its name slightly, however, this makes it difficult for us to assign the portfolio code if the names are different;

3. Compress the names to remove the spaces between two letters, and capitalize the names so that they are not case sensitive;

4. Extract the first segment of the observation names out of the whole name so that the second segment of names containing the subclass information is removed and only the part containing the actual name of the fund is retained. This step can be done based on the fact that the two segments of names are often separated by the

character " ; " or " / " in CRSP; [4]

5. For each fund, find out if the fund has been renamed. If so, we change the old name to the most recent updated name. Repeat this process until there are no name discrepancies remaining within a single fund;

6. We use the "Levenshtein edit distance" to measure the distance between two strings (names) and compare it with a predefined cutoff (cutoff = 2) to identify similar fund names to avoid name typos in CRSP;

7. Finally, we check and make sure that the portfolio code we assign is not different within the same shareclass.

After the portfolio codes are assigned to all the subclasses, we proceed to check their credibility according to the variable *crsp_portno*. *crsp_portno* exists in the CRSP database as a portfolio identifier for a single fund that has multiple shareclasses. Unfortunately, this variable is missing for most funds, and that is why we need to construct our own portfolio identifier to merge different shareclasses of a single fund into one portfolio. However, we can still use those funds that have the *crsp_portno* available to check if our portfolio identifier *calc_portcode* correctly assigns a unique code to a single fund portfolio, i.e. for those funds that have non-missing *crsp_portno*, variables *crsp_portno* and *calc_portcode* should simultaneously identify the same fund portfolio. In other words, there should not be cases where the same *crsp_portno* is assigned to more than one *calc_portcode*. Finally,

[4]It should be noted that the position of the character are searched from right to left in case the shareclasses names contain more than one of these two characters, e.g. there is one shareclass called "ABN ARMO Funds:ABN ARMO/Montag & Caldwell Growth Fund; Class R Shares". This name has both symbols "/" and ";" in it. If we search from left to right like most search statements in software would do, then the position of "/" is returned as the result and then the segments of the name would be incorrect. Instead, we should search from the right to the left. For example, we use the software SAS 9.4 in our study so that we choose the "findc" statement with modifier "b" denoted as searching from right to left to do the job.

we manually check funds with more than 6 shareclasses to make sure that the correct portfolio codes are assigned. It should be noted that, after assigning portfolio codes, our dataset should have the following characteristics:

1. It has a new variable called *calc_portcode* added into the dataset;

2. *crsp_fundno* uniquely identifies a single shareclass, and *calc_portcode* uniquely identifies a single fund portfolio.

A.6 Fund Style, Value Weighted Net and Gross Returns

Each monthly return for a single fund is calculated by weighting the net return of its component share classes by their beginning-of-month total Net Asset Values (TNA). After shareclasses are assigned portfolio codes, they are weighted by their total TNA of that shareclass at the beginning of that month and are merged into a single fund. Before doing that, we fill the missing values of the TNA with the average of adjacent non-missing values before and after the missing ones. [5]

To further generate gross returns, we add monthly expense ratios back to the net returns. Since only yearly expense ratios are available, to generate monthly gross returns, we add $1/12^{th}$ yearly expense ratios to net returns. However, expense ratios can be missing for some funds in certain fiscal years. When a fund's expense ratio for a year is missing, it is assumed that the missing expense ratio is the same as other funds that are actively managed and have similar Assets Under Management (AUM). Therefore, the crucial step is to locate "similar funds", which we define as funds in the same investment style category with similar AUM. We assign three categories of investment styles to funds

[5] It should be noted that, in CRSP, there are some TNA observations values shown as "-99" in the database and are in fact missing values. We treat them as missing observations.

in our sample according to Barras et al. (2010): Aggressive growth funds, growth funds, and growth & income funds. Aggressive growth funds usually take the highest risk to achieve the highest capital gains. Growth funds tend to target fast capital appreciation and are less moderate than aggressive growth funds. Growth & income funds pursue a dual strategy of income generation and capital gain and are the least risky funds among the three categories. We assign investment style categories to funds based on the following variables:

1. CRSP style code (*crsp_obj_cd*);
2. Lipper class code (*lipper_class*);
3. Lipper objective code (*lipper_obj_cd*);
4. Strategic insights objective codes (*si_obj_cd*);
5. Wiesenberger objective code (*wbrger_obj_cd*);
6. Fund names (*fund_name*).

To be specific, we use style or objective codes to classify funds according to the following standard in Table A.8:

Table A.8: Fund Investment Styles Classified According to Objectives

	Growth & income	Growth	Aggressive growth
CRSP style codes (*crsp_obj_cd*)	EDYB, EDYI	EDYG	
Weisenberger objective codes (*wbrger_obj_cd*)	GCI, IEQ	LTG, SCG	MCG
Strategic insight objective codes (*si_obj_cd*)	GRI, ING	GRO	AGG
Lipper objective codes (*lipper_obj_cd*)	GI, EI	G, CA	
Lipper class codes (*lipper_class*)	EIEI, LCVE, MCVE, MLVE, SCVE	G, LCGE, MCGE, MLGE, SCGE	

Fund names can contain key words that indicate funds' investment styles. We use the key words in Table A.9 to identify growth & income funds:

Table A.9: Key Words to Identify Growth & Income Funds

growth and income	growth & income	income & growth
growth with income	income	value
income and growth	income with growth	val

Key words used to identify growth funds are shown in Table A.10:

Table A.10: Key Words to Identify Growth Funds

growth	capital appreciation	appreciation	capital apprec	grth

Generally, any funds that have these key words in their names are very likely to be growth funds. However, there are exceptions so we put together another list of key words in Table A.11 to reject a fund as a growth fund:

Table A.11: Key Words to Reject a Fund as a Growth Fund

aggressive growth	aggressive	aggress growth	growth and income
growth & income	income and growth	income & growth	growth with income
income with growth	value	income	

Finally, we use the following key words in Table A.12 to identify aggressive growth funds:

Table A.12: Key Words to Identify Aggressive Growth Funds

aggressive growth	aggressive	aggress growth	aggr growth

To summarize, we use fund names to classify funds according to the following standard:

1. A fund is denoted as a growth and income fund if its name contains one of the key words in Table A.9;

2. A fund is denoted as a growth fund if its name contains one of the key words in Table A.10, but does not have the key words in Table A.11;

3. A fund is denoted as an aggressive growth fund if its name contains one of the key words in Table A.12.

After the above steps, the new variable *fund_style* should indicate the fund investment style of funds in our sample. We denote fund types as missing when funds cannot be identified using either style codes or fund names. Next, we fill in those missing expense ratios with non-missing expense ratios from similar funds defined as funds in the same investment styles and similar AUM. Finally, we create monthly gross returns by adding in $1/12^{th}$ expense ratios to monthly net returns.

A.7 Excluding New Funds and Correcting Incubation Bias

To avoid having many new funds in our sample, we only keep funds that have at least five years of return history before the end of our sample period so as to remove new ones with short return history. Since CRSP report return history at shareclass level, we only exclude a new fund when the history of all of its shareclasses are less than five years.

We also correct for "incubation" bias which refers to the fact that funds are often provided seed money to establish an appealing return history by fund management companies in their pre-release period. Incubation bias arises when these returns are included in our analysis. We exclude the part of return history of funds that are below a certain AUM bound. This lessens the effects of "incubation" bias given that funds usually have low AUM in their pre-release period. Evans (2010) discussed this issue and suggests that by using returns only after funds receive a NASDAQ ticker symbol, “incubation” bias can be minimized. However, F&F argue that this ticker symbol information is only available after 1998 and their results accounting for NASDAQ ticker dates turn out to be no different from their original results.

Therefore, to solve the “incubation” bias problem, we follow F&F’s methodology to limit tests to those funds that hold AUM of at least 5 million 2006 dollars. We set the AUM bound to $6 million in 2015 after adjusting for inflation since F&F set the AUM bound to $5 million in the year 2006. Similarly, the AUM bound in each year is adjusted according to inflation rates, for example, a fund is included in 1984 if it has more than $2.5 million of assets under management. Furthermore, to obtain accurate estimates of fund α, we only keep funds that have at least 8 months of return history in our sample. In the end, our sample has 5006 funds that span a period of 32 years with a maximum range of 375 months from January 1984 to March 2015.

A.8 Summary Statistics of Data

To have a general picture of our data, we show some characteristics of the data including number of funds, average total net assets of funds and average expense ratios. Since our sample period covers the global financial crisis and the recovery of it, it is interesting to see how these characteristics evolve over time. First, we split our sample data into several sub-samples that contain the funds existing in each year from 1984 to 2015. Then, we construct subsamples that have funds with return history up to the end of each year from 1998 to 2015. The number of funds in our sample existing in each year and number of funds within the sample until a certain year are shown in Figure A.1 and Figure A.2.

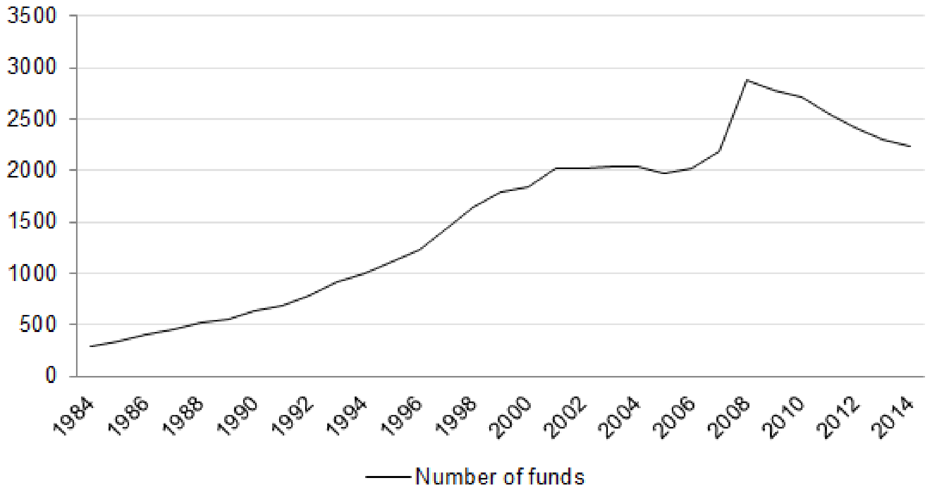

Figure A.1: Number of Mutual Funds per Year in Our Sample

Information on the number of funds existing in each year allows us to see how many fund were in the market that year after some funds left the market and the entry of new funds. As shown in Figure A.1, the number of funds existing in each year gradually increases and stops growing during the 2007-2008 financial crisis period, dropping to a level below 2000. After the crisis, the market experienced a boom with the number of mutual funds increasing sharply, peaking at approximately 3000. However, after 2010, the number of funds in our sample continued to decrease every year, possibly due to the ever-increasing market competition, especially the prevalence of index funds in recent years. When the curve representing the number of funds till a certain year is plotted, we see a similar pattern in the number of funds, dead and alive altogether, increases steadily from 1998 to the point of the global financial crisis and then goes up after the crisis at a higher rate. From 2010 to the end of our sample period, the total number of funds in the sample remains relatively stable.

Furthermore, we also show the average Total Net Assets (TNA), assets-weighted average expense ratios of the funds existing in each year in our sample. As shown in

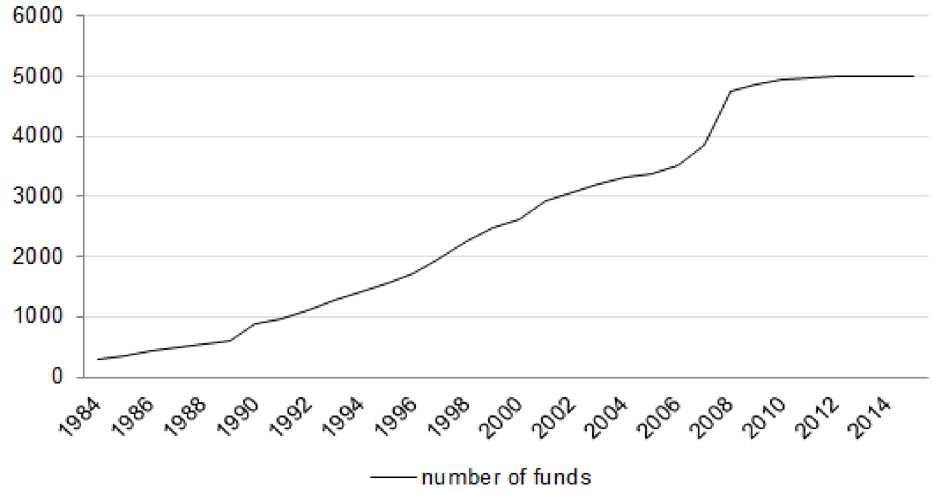

Figure A.2: Number of Mutual Funds up to a Certain Year in Our Sample

Figure A.3, the average assets held by funds experienced two major reductions during the Dot-com bubble in the early 2000s and the 2007-2008 financial crisis period, respectively. At other times, the data show that mutual funds keep expanding in terms of total assets. As to the expense ratios, Figure A.4 shows that overall, assets-weighted expense ratios are declining over the years, although they fluctuate frequently during the entire sample period.

Since funds have different investment styles: aggressive growth, growth and growth & income, it is natural to wonder if different investment style categories show different average TNA and expense ratios. Funds are assigned to different investment style categories based on their self-declared investment objectives and their fund names, details of which are explicitly explained in Section A.6. As shown in Figure A.5, the number of growth funds and growth & income funds keeps growing until the start of the global financial crisis and then drops substantially. Growth funds have the highest number of funds among the three groups, peaking at around 1600 in 2009. Before 1990, there are no growth & income funds in the sample and their number increases from around 200 in

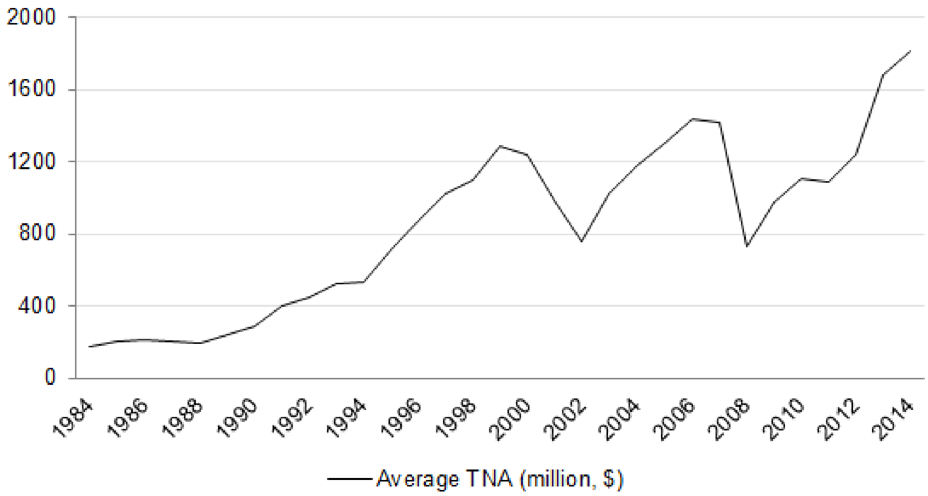

Figure A.3: Average TNA of Mutual Funds per Year in Our Sample

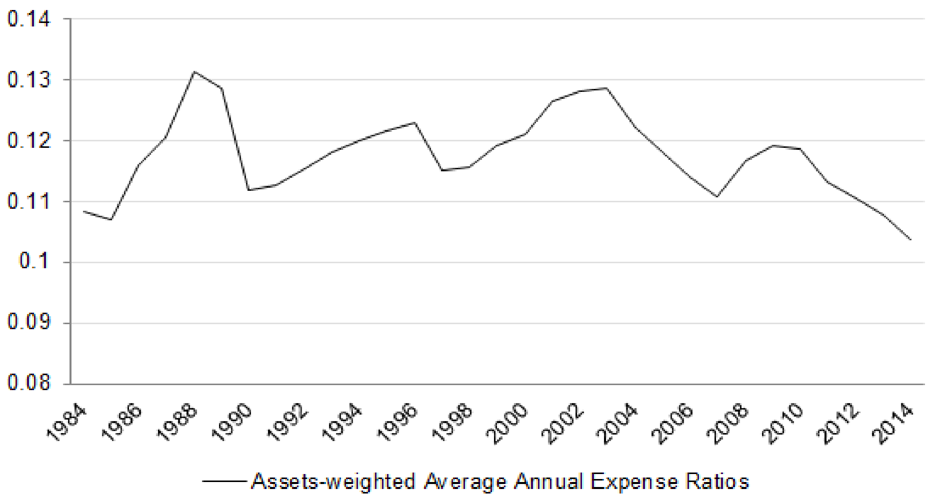

Figure A.4: Average Annual Expense Ratios of Mutual Funds per Year

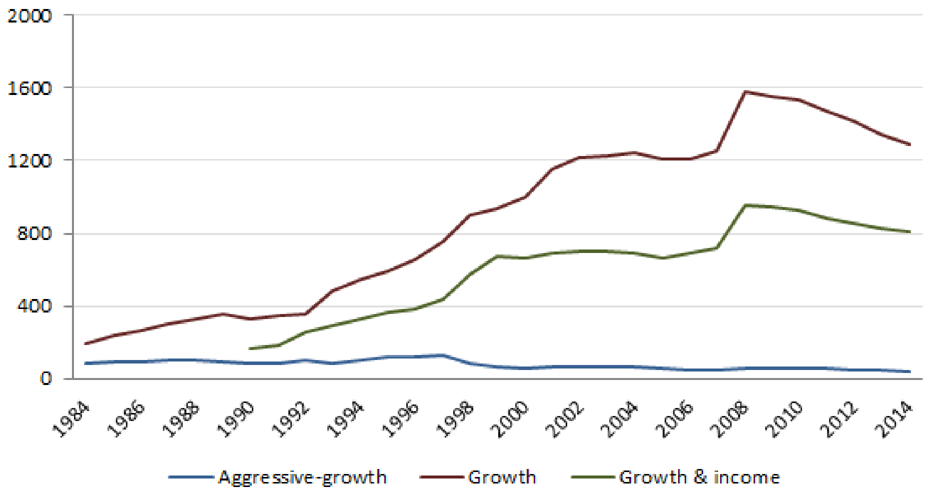

Figure A.5: Number of Mutual Funds in Different Investment Styles per Year in Our Sample

1990 to around 1000 in 2008 and then drops to 800 till 2015. The number of aggressive growth funds remains relatively constant over these years, fluctuating around 100 and is the lowest among the three groups of funds.

Regarding assets held by the funds, growth & income funds generally have the largest amount of average assets under management, and aggressive growth funds have the lowest.

As to expense ratios, before the early 2000s, aggressive growth funds have the highest annual assets-weighted expense ratios peaking at a value of around 0.16 and growth funds have the second highest expense ratios. After around 2004, expense ratios of aggressive growth funds start to decline and become the lowest among all investment style groups from around 2007. Hence, we can see that aggressive growth funds play an important role in reducing the expense ratios of active mutual funds in general in recent years.

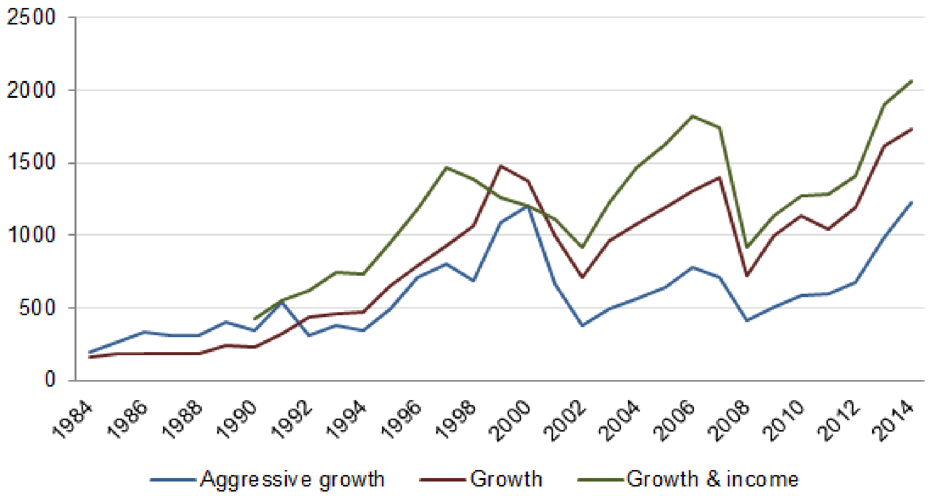

Figure A.6: Average TNA of Mutual Funds in Different Investment Styles per Year in Our Sample

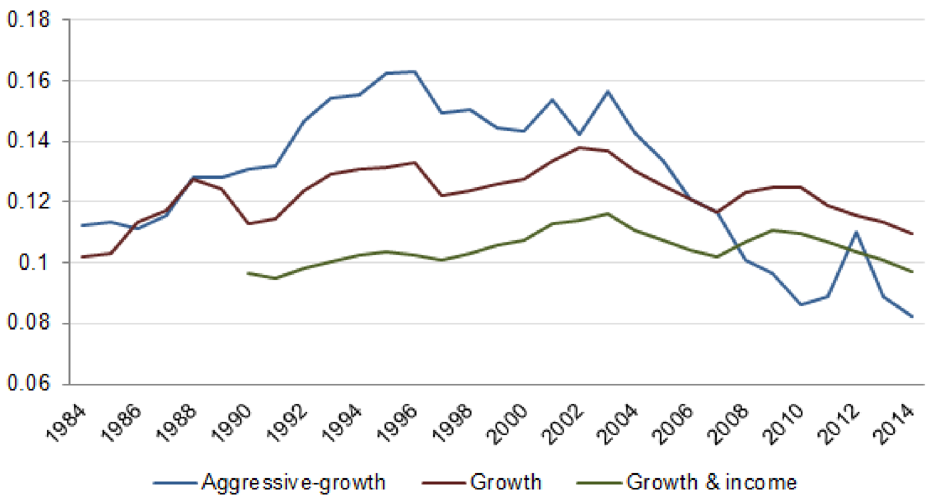

Figure A.7: Average Expense Ratios of Mutual Funds in Different Investment Styles per Year in Our Sample

List of References

Andrews, Donald WK (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica: Journal of the Econometric Society* **61**(4), 821–856.

Andrews, Donald WK, Inpyo Lee and Werner Ploberger (1996), 'Optimal change point tests for normal linear regression', *Journal of Econometrics* **70**(1), 9–38.

Andrews, Donald WK and Werner Ploberger (1994), 'Optimal tests when a nuisance parameter is present only under the alternative', *Econometrica: Journal of the Econometric Society* **62**, 1383–1414.

Antoch, Jaromir, Marie Hušková and Noël Veraverbeke (1995), 'Change-point problem and bootstrap', *Journaltitle of Nonparametric Statistics* **5**(2), 123–144.

Aue, Alexander, Siegfried Hörmann, Lajos Horváth, Matthew Reimherr et al. (2009), 'Break detection in the covariance structure of multivariate time series models', *The Annals of Statistics* **37**(6B), 4046–4087.

Bai, Jushan (1994), 'Least squares estimation of a shift in linear processes', *Journal of Time Series Analysis* **15**(5), 453–472.

Bai, Jushan and Pierre Perron (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* **66**(1), 47–78.

Bai, Jushan and Pierre Perron (2003), 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics* **18**(1), 1–22.

Bajgrowicz, Pierre and Olivier Scaillet (2012), 'Technical trading revisited: False discoveries, persistence tests, and transaction costs', *Journal of Financial Economics* **106**(3), 473–491.

Barras, Laurent, Olivier Scaillet and Russ Wermers (2010), 'False discoveries in mutual fund performance: Measuring luck in estimated alphas', *The Journal of Finance* **65**(1), 179–216.

Becker, Connie, Wayne Ferson, David H Myers and Michael J Schill (1999), 'Conditional market timing with benchmark investors', *Journal of Financial Economics* **52**(1), 119–148.

Benjamini, Yoav and Daniel Yekutieli (2001), 'The control of the false discovery rate in multiple testing under dependency', *Annals of Statistics* **29**(4), 1165–1188.

Benjamini, Yoav and Yosef Hochberg (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 289–300.

Berk, Jonathan B and Richard C Green (2004), 'Mutual fund flows and performance in rational markets', *Journal of Political Economy* **112**(6), 1269–1295.

Berndt, Ernst R (1991), *The practice of econometrics: classic and contemporary*, Addison-Wesley Reading, MA.

Blake, David and Allan Timmermann (1998), 'Mutual fund performance: evidence from the UK', *European Finance Review* **2**(1), 57–77.

Bollen, Nicolas PB and Robert E Whaley (2009), 'Hedge fund risk dynamics: Implications for performance appraisal', *The Journal of Finance* **64**(2), 985–1035.

Bredin, Don, Keith Cuthbertson, Dirk Nitzsche and Dylan C Thomas (2014), 'Performance and performance persistence of UK closed-end equity funds', *International Review of Financial Analysis* **34**, 189–199.

Brown, Robert L, James Durbin and James M Evans (1975), 'Techniques for testing the constancy of regression relationships over time', *Journal of the Royal Statistical Society. Series B (Methodological)* **37**(2), 149–192.

Carhart, Mark M (1997), 'On persistence in mutual fund performance', *The Journal of Finance* **52**(1), 57–82.

Chow, Gregory C (1960), 'Tests of equality between sets of coefficients in two linear regressions', *Econometrica: Journal of the Econometric Society* **28**(3), 591–605.

Christoffersen, Susan EK and David K Musto (2002), 'Demand curves and the pricing of money management', *Review of Financial Studies* **15**(5), 1499–1524.

Clare, Andrew, Nick Motson, Svetlana Sapuric and Natasa Todorovic (2014), 'What impact does a change of fund manager have on mutual fund performance?', *International Review of Financial Analysis* **35**, 167–177.

Clarke, Sandy and Peter Hall (2009), 'Robustness of multiple testing procedures against dependence', *The Annals of Statistics* **37**(1), 332–358.

Criton, Gilles and Olivier Scaillet (2011), Time-varying analysis in risk and hedge fund performance: How forecast ability increases estimated alpha, *in* 'SSRN working paper, March'.

Csörgő, Miklós and Lajos Horváth (1988), 'Invariance principles for changepoint problems', *Journal of Multivariate Analysis* **27**(1), 151–168.

Cuthbertson, Keith, Dirk Nitzsche and Niall O'Sullivan (2008), 'UK mutual fund performance: Skill or luck?', *Journal of Empirical Finance* **15**(4), 613–634.

Cuthbertson, Keith, Dirk Nitzsche and Niall O'Sullivan (2012), 'False discoveries in UK mutual fund performance', *European Financial Management* **18**(3), 444–463.

Daniel, Kent, Mark Grinblatt, Sheridan Titman and Russ Wermers (1997), 'Measuring mutual fund performance with characteristic-based benchmarks', *The Journal of Finance* **52**(3), 1035–1058.

Davidson, Russell and James G MacKinnon (2004), *Econometric theory and methods*, Vol. 5, Oxford University Press New York.

Dybvig, Philip H and Stephen A Ross (1985), 'The analytics of performance measurement using a security market line', *The Journal of Finance* **40**(2), 401–416.

Efron, Bradley (1979), 'Bootstrap methods: another look at the jackknife', *The Annals of Statistics* **7**(1), 1–26.

Efron, Bradley and Robert J Tibshirani (1994), *An introduction to the bootstrap*, CRC press.

Elton, Edwin J, Martin J Gruber and Christopher R Blake (2006), 'The adequacy of investment choices offered by 401 (k) plans', *Journal of Public Economics* **90**(6), 1299–1314.

Elton, Edwin J, Martin J Gruber and Christopher R Blake (2007), 'Participant reaction and the performance of funds offered by 401 (k) plans', *Journal of Financial Intermediation* **16**(2), 249–271.

Elton, Edwin J, Martin Jay Gruber and Christopher R Blake (1996), 'Survivor bias and mutual fund performance', *Review of Financial Studies* **9**(4), 1097–1120.

Elton, Edwin J, Martin Jay Gruber, Sanjiv Das and Matthew Hlavka (1993), 'Efficiency with costly information: A reinterpretation of evidence from managed portfolios', *Review of Financial Studies* **6**(1), 1–22.

Evans, Richard B (2010), 'Mutual fund incubation', *The Journal of Finance* **65**(4), 1581–1611.

Fabozzi, Frank J and Jack C Francis (1979), 'Mutual fund systematic risk for bull and bear markets: an empirical examination', *The Journal of Finance* **34**(5), 1243–1250.

Fama, Eugene F and Kenneth R French (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

Fama, Eugene F and Kenneth R French (2010), 'Luck versus skill in the cross-section of mutual fund returns', *The Journal of Finance* **65**(5), 1915–1947.

Ferreira, Miguel A, Aneel Keswani, António F Miguel and Sofia B Ramos (2012), 'The determinants of mutual fund performance: A cross-country study', *Review of Finance* **0**(0), 1–43.

Ferson, Wayne E and Rudi W Schadt (1996), 'Measuring fund strategy and performance in changing economic conditions', *The Journal of Finance* **51**(2), 425–461.

French, Kenneth R (2008), 'Presidential address: The cost of active investing', *The Journal of Finance* **63**(4), 1537–1573.

Garcia, René and Pierre Perron (1996), 'An analysis of the real interest rate under regime shifts', *The Review of Economics and Statistics* **78**(1), 111–125.

Gombay, Edit (2010), 'Change detection in linear regression with time series errors', *Canadian Journal of Statistics* **38**(1), 65–79.

Gombay, Lajos Horváth, Marie Husková (1996), 'Estimators and tests for change in variances', *Statistics & Decisions* **14**(2), 145–159.

Griffin, John M, Tao Shu and Selim Topaloglu (2012), 'Examining the dark side of financial markets: Do institutions trade on information from investment bank connections?', *The Review of Financial Studies* **00**(0), 1–34.

Grinblatt, Mark and Sheridan Titman (1988), *The evaluation of mutual fund performance: An analysis of monthly returns*, John E. Anderson Graduate School of Management, University of California.

Grinblatt, Mark and Sheridan Titman (1989), 'Mutual fund performance: An analysis of quarterly portfolio holdings', *Journal of Business* **62**(3), 393–416.

Grinblatt, Mark and Sheridan Titman (1992), 'The persistence of mutual fund performance', *The Journal of Finance* **47**(5), 1977–1984.

Grinblatt, Mark and Sheridan Titman (1994), 'A study of monthly mutual fund returns and performance evaluation techniques', *Journal of Financial and Quantitative Analysis* **29**(03), 419–444.

Gruber, Martin J (1996), 'Another puzzle: The growth in actively managed mutual funds', *The Journal of Finance* **51**(3), 783–810.

Hall, Peter, Joel L Horowitz and Bing-Yi Jing (1995), 'On blocking rules for the bootstrap with dependent data', *Biometrika* **82**(3), 561–574.

Hendricks, Darryll, Jayendu Patel and Richard Zeckhauser (1993), 'Hot hands in mutual funds: Short-run persistence of relative performance, 1974-1988', *The Journal of Finance* **48**(1), 93–130.

Henriksson, Roy D (1984), 'Market timing and mutual fund performance: An empirical investigation', *Journal of Business* **57**(1), 73–96.

Horváth, Lajos and Gregory Rice (2014), 'Extensions of some classical methods in change point analysis', *Test* **23**(2), 219–255.

Huij, Joop and Marno Verbeek (2007), 'Cross-sectional learning and short-run persistence in mutual fund performance', *Journal of Banking & Finance* **31**(3), 973–997.

Ippolito, Richard A (1989), 'Efficiency with costly information: A study of mutual fund performance, 1965-1984', *The Quarterly Journal of Economics* **104**(1), 1–23.

Jensen, Michael C (1968), 'The performance of mutual funds in the period 1945–1964', *The Journal of Finance* **23**(2), 389–416.

Kacperczyk, Marcin, Clemens Sialm and Lu Zheng (2005), 'On the industry concentration of actively managed equity mutual funds', *The Journal of Finance* **60**(4), 1983–2011.

Kacperczyk, Marcin, Clemens Sialm and Lu Zheng (2008), 'Unobserved actions of mutual funds', *Review of Financial Studies* **21**(6), 2379–2416.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh and Laura Veldkamp (2013), 'Time-varying fund manager skill', *The Journal of Finance* **69**(4), 1455–1484.

Khorana, Ajay (2001), 'Performance changes following top management turnover: Evidence from open-end mutual funds', *Journal of Financial and Quantitative Analysis* **36**(03), 371–393.

Kim, Sangbae and Francis In (2012), 'False discoveries in volatility timing of mutual funds', *Journal of Banking & Finance* **36**(7), 2083–2094.

Kon, Stanley J and Frank C Jen (1978), 'Estimation of time-varying systematic risk and performance for mutual fund portfolios: an application of switching regression', *The Journal of Finance* **33**(2), 457–475.

Kosowski, Robert (2011), 'Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions', *The Quarterly Journal of Finance* **1**(03), 607–664.

Kosowski, Robert, Allan Timmermann, Russ Wermers and Hal White (2006), 'Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis', *The Journal of Finance* **61**(6), 2551–2595.

Liu, Jian, Shiying Wu and James V Zidek (1997), 'On segmented multivariate regression', *Statistica Sinica* **7**(2), 497–525.

Malkiel, Burton G (1995), 'Returns from investing in equity mutual funds 1971 to 1991', *The Journal of Finance* **50**(2), 549–572.

McCracken, Michael W and Stephen G Sapp (2005), 'Evaluating the predictability of exchange rates using long-horizon regressions: Mind your p's and q's!', *Journal of Money, Credit and Banking* **37**(3), 473–494.

Olmo, Jose, William Pouliot et al. (2014), 'Tests to disentangle breaks in intercept from slope in linear regression models with application to management performance in the mutual fund industry', *Department of Economic University of Birmingham Discussion Paper Series* .

Otamendi, Javier, Luis Miguel Doncel, Pilar Grau and Jorge Sainz (2008), 'An evaluation on the true statistical relevance of Jensen's alpha through simulation: An application for Germany', *Economics Bulletin* **7**(10), 1–9.

Otten, Roger and Dennis Bams (2002), 'European mutual fund performance', *European Financial Management* **8**(1), 75–101.

Page, ES (1954), 'Continuous inspection schemes', *Biometrika* **41**(1/2), 100–115.

Pástor, L'uboš and Robert F Stambaugh (2002), 'Mutual fund performance and seemingly unrelated assets', *Journal of Financial Economics* **63**(3), 315–349.

Pouliot, William (2016), 'Robust tests for change in intercept and slope in linear regression models with application to manager performance in the mutual fund industry', *Economic Modelling* **58**, 523–534.

Roll, Richard (1978), 'Ambiguity when performance is measured by the security market line', *The Journal of Finance* **33**(4), 1051–1069.

Sharpe, William F (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance* **19**(3), 425–442.

Sharpe, William F (1991), 'The arithmetic of active management', *Financial Analysts Journal* **47**(1), 7–9.

Sirri, Erik R and Peter Tufano (1998), 'Costly search and mutual fund flows', *The Journal of Finance* **53**(5), 1589–1622.

Storey, John D (2002), 'A direct approach to false discovery rates', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(3), 479–498.

Storey, John D (2003), 'The positive false discovery rate: a Bayesian interpretation and the q-value', *The Annals of Statistics* **31**(6), 2013–2035.

Storey, John D, Jonathan E Taylor and David Siegmund (2004), 'Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(1), 187–205.

Suh, Sangwon and Kyttack Hong (2011), 'Control of luck in measuring investment fund performance*', *Asia-Pacific Journal of Financial Studies* **40**(3), 467–493.

Survivor-bias-free US Mutual Fund Guide (2014).
URL: *http://www.crsp.com/files/MF_Sift_Guide.pdf*

Tang, Ya, YU-Jane Liu and Yi Yang (2011), 'Measuring luck in mutual fund performance: Evidence from false discovery ratio method', *Available at SSRN 1942862* .

Vostrikova, L (1981), 'Detecting disorder in multidimensional random processes', *Soviet Mathematics Doklady* **24**, 55–59.

Wermers, Russ (1999), 'Mutual fund herding and the impact on stock prices', *The Journal of Finance* **54**(2), 581–622.

Wermers, Russ (2000), 'Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses', *The Journal of Finance* **55**(4), 1655–1703.