UK MUTUAL FUND PERFORMANCE

Niall Michael O' Sullivan M.A., M.Sc., M.Sc.

Research presented in fulfilment of the requirements of the examination for the degree of

Doctor of Philosophy in Finance

City University, London.

Cass Business School, July 2006.

Supervisors:

Professor Keith Cuthbertson
Dr Dirk Nitzsche

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge the support given to me by various people in the completion of this work. Special thanks go to:

- Professor Keith Cuthbertson and Dr Dirk Nitzsche, my research supervisors, for providing me with an excellent training in research. Both have been extremely generous and helpful in availing me of their considerable expertise. I am very fortunate and appreciative to have been included in their research team.

- Mr Grahame Goodyer, Fenchurch Corporate Services, for providing me with extensive data on UK mutual funds and for his infinite patience in generously assisting me with many queries.

- The many people who have provided useful comments at conferences, seminars etc where this work has been presented including at Cass Business School, Imperial College London, Trinity College Dublin, University College Dublin, Dublin City University and University College Cork.

LIBRARY DECLARATION

Discretionary power is granted to the university librarian to permit this thesis to be copied for study purposes subject to normal conditions of acknowledgement.

ABSTRACT

Using a comprehensive data set on (surviving and non-surviving) UK equity mutual funds (April 1975 – December 2002), this study uses a bootstrap methodology to distinguish between 'skill' and 'luck' in fund performance. This methodology allows for non-normality in the idiosyncratic risks of the funds – a major issue when considering the 'best' and 'worst' funds and these are the funds which investors are most interested in. The study points to the existence of genuine stock picking ability among a relatively small number of top performing UK equity mutual funds (i.e. performance which is not solely due to good luck). At the negative end of the performance scale, the analysis strongly rejects the hypothesis that most poor performing funds are merely unlucky. Most of these funds demonstrate 'bad skill'. The study also examines the economic and statistical significance of persistence. Sorting funds into deciles based on past raw returns or on past 4-factor alphas, strong evidence is found that past loser funds continue to perform badly in terms of their future 4-factor alphas while little evidence is found that past winner funds provide future positive risk adjusted performance. However, on investigating relatively small 'fund-of-fund' portfolios of past winners, evidence of positive persistence is found. Using a cross-section bootstrap approach the study derives the empirical distribution of final wealth at a 10 year horizon and finds that if transactions costs are above 2.5% per fund round trip, a passive strategy seems at least as good as the active strategies examined while with transactions costs of 5% the passive strategy is most probably superior. The study also examines the market timing performance of the funds. Using a nonparametric test procedure the study evaluates both unconditional market timing and timing conditional on publicly available information. A relatively small number of funds (around 1%) are found to successfully time the market while market mistiming is relatively prevalent.

CHAPTER 1

INTRODUCTION

This study examines the performance of open-end mutual funds (Unit Trusts and Open Ended Investment Companies (OEICs)) investing in UK domestic equity during the period April 1975 to December 2002. A data set of 1,620 funds is examined. This represents almost all UK domestic equity mutual funds in existence at the end of the sample period and includes over 200 nonsurviving funds. In contrast to the US mutual fund industry, there have been comparatively few studies of the performance of UK mutual funds. Studies of UK funds have, for the most part, examined issues such as risk adjusted performance, survivor bias and performance persistence.

This study makes a number of contributions to three areas of the existing literature on UK mutual fund performance. First, it distinguishes between skill and luck in fund stock picking ability. Second, it examines persistence in this ability. Third, it evaluates a second aspect of performance, i.e. market timing.

In this first area, this study directly addresses the issue of 'skill versus luck'. It uses alpha, α, and the t-statistic of alpha, t_α, as the measures of risk adjusted performance of mutual funds. However, it does not assume, as many earlier studies do, that a fund's idiosyncratic risk has a known distribution. Instead a bootstrap procedure is applied to determine the *empirical* distribution of idiosyncratic risk. This is not just done fund-by-fund but uses the entire cross-section of funds. The procedure makes it possible to obtain separate performance distributions for funds which are in the tails of the cross-section distribution – precisely the funds that investors are likely to be most interested in (i.e. extreme 'winners' or 'losers'). Indeed the procedure generates separate individual nonparametric sampling distributions under the null hypothesis of no abnormal performance at *every* point in the cross-section distribution of performance, i.e. for the best fund, 2nd best fund, 3rd best fund etc down to the worst fund. This methodology allows for non-normality in the idiosyncratic risks of the funds – a common finding among top and bottom performers. By comparing the bootstrap distribution against the actual distribution of fund performance it is possible to determine whether high α (t_α) funds exceed random sampling variability in the performance measure, i.e. luck. Similarly, it is possible to evaluate whether poor performance is worse than bad luck.

This bootstrap methodology represents an original and significant contribution to the extant literature on UK mutual fund performance. The technique has been applied to US mutual funds in an unpublished working paper by Kosowski, Timmermann, Wermers and White (2004).

We perform a number of bootstrap techniques to account for serial correlation or heteroscedasticity in the idiosyncratic risk of each fund as well as possible cross-section correlation (across funds) in this risk. The bootstrap procedure is robust to possible misspecification but reported results are of course dependent on the chosen performance model. To address this the study examines a wide range of alternative performance measurement models which are divided into three broad classes (i) unconditional models (Jensen 1968, Fama and French 1993, Carhart 1997) (ii) conditional-beta models, in which factor loadings are allowed to change with conditioning public information (Ferson and Schadt 1996) and (iii) conditional alpha-beta models where conditioning information also allows for time varying alphas (Christopherson, Ferson and Glassman 1998).

This work also examines whether skill varies across (i) funds of different investment objectives (income stock funds, general equity funds and small stocks funds) and (ii) funds domiciled onshore versus offshore (although all funds invest only in UK domestic equity). These separate subcategories represent a more homogenous risk group of funds and therefore applying the bootstrap procedure separately controls for risk characteristics in these funds' stockholdings which may not be captured by the performance model. It also enables the study to examine questions such as whether the small stock market is less efficient and whether offshore funds incur an informational asymmetry cost. This is the first UK study to examine this onshore/offshore question.

The study controls for survivor bias by including 216 'nonsurviving' funds in the analysis. Furthermore, the sample period under investigation in this study (28 years) is the longest among similar studies in the literature. Both of these factors mean there is a greater amount of information informing the actual and bootstrap distributions of performance. This reduces any possibility that findings may be survivor biased or may be sample-period specific.

This study also examines the possible role of a momentum effect in stock returns in explaining the cross-section distribution of mutual fund returns. Notwithstanding some work by Fletcher and Forbes (2002), no detailed analysis has been conducted on a momentum effect among UK equity mutual funds. This study makes significant contributions to the extant UK literature in this area. First, it applies a much improved measure of momentum using the comprehensive London Share Price Database: in the construction of the momentum variable the portfolios of past winner and loser stocks calculated at each time period t are based on stocks which were available to fund managers at *that* time in history and *not* just on the historical time series of stocks which continue to exist *at the end* of the sample period. The latter calculation risks introducing a survivorship bias as the momentum factor would be based solely on stocks which have survived to the end of the sample period. Second, the robustness of momentum findings to variations in its definition is also examined.

The second major area of contribution of this study is in its examination of persistence in mutual fund performance. A crucial issue is whether abnormal performance can be identified *ex-ante* and for how long it persists. Persistence in the sense of 'statistical predictability' is usually examined using (rank) correlations or regressions of future on past performance or using a contingency table approach. 'Economic predictability' is usually based on the alpha from a strategy of holding a portfolio of past 'winning' funds (rebalanced periodically) or by observing actual trades of mutual funds (i.e. holdings and buy/sell data) and using a characteristic selectivity measure in an event study framework.

For tests based on 'statistical predictability' the evidence is rather mixed. Early US studies suggest some predictability in performance at both long and short horizons (Grinblatt and Titman 1992, Hendricks, Patel and Zechauser 1993, Brown and Goetzmann 1995, Goetzmann and Ibbotson 1994). On balance, more recent studies suggest that predictability among past winners is rather more tenuous but some predictability is found when a portfolio of past winning funds is rebalanced frequently (e.g. at least once per quarter) and the performance horizon is not longer than about one-year (e.g. Carhart 1997, Wermers 1997, Blake and Morey 2000). For UK data on mutual and pension funds there is little evidence of 'statistical predictability' amongst past winner funds but much stronger evidence that poor performers continue to under-perform (e.g. Blake and

Timmermann 1998, Allen and Tan 1999, Fletcher and Forbes 2002, Blake, Lehmann and Timmermann 1999, Tonks 2004).

US studies examining 'economic predictability' using actual trades of mutual funds find that one-year persistence among past winner funds is due to stocks passively carried over, rather than newly purchased stocks of winner funds performing better than newly purchased stocks of past loser funds (Chen et al 1999, Wermers 2003).

Results from an analysis of persistence are usually interpreted as direct tests of the EMH in a market where entry barriers are relatively low, there are many professional traders who operate in a competitive environment and information is available at relatively low cost – precisely the conditions under which the EMH is expected to be valid. Therefore mutual funds provide a way of testing the behavior of investors against the classic paradigm of finance theory where individuals are assumed to make rational decisions in relatively frictionless and low information cost markets, which leads to the elimination of inferior financial products and the growth in successful ones.

This study contributes to the debate on persistence in several ways. First, the study does not focus on 'statistical predictability' as this does not necessarily result in explicit investment rules that produce future positive risk adjusted returns. The statistical approach measures the average association between the *relative orderings* of funds in a pre-sort and post-sort period using correlation, regression or contingency tables. However, although such tests may provide *prima facie* evidence of persistence it is often not clear how this may be exploited by investors. Measured persistence could be due mainly to repeat losers rather than repeat winners. In addition, (Spearman) rank correlations treat each point in the ranking equally and lack power against the hypothesis that predictability in performance is concentrated in the tails of fund performance. Instead this study concentrates on the *economic* value of persistence by focusing on the approach of recursively rebalancing a portfolio of past winning funds. Here, if persistence is found it may represent an exploitable strategy for investors. Second, as well as analyzing the post-sort performance of quite large and possibly heterogeneous portfolios of funds (e.g. deciles), as done in earlier studies, the study also examines alternative smaller fund-of-fund portfolios which is probably of more practical interest to both professional and retail investors. Third, the study examines not only the risk adjusted average performance of portfolios of past winner/loser funds but also *the distribution* of final wealth from this *ex-ante* strategy, taking account of 'luck' across *all* funds and transactions costs of rebalancing. Here 'luck' is represented by the empirical distribution of all funds' idiosyncratic risks and hence picks up any non-normality and contemporaneous cross-correlations in idiosyncratic risks. As far as can be ascertained, this approach to an analysis of final wealth has not been developed in any previous studies. Since most saving in mutual funds (as a whole) is long-term, investors are interested in the distribution of final wealth (e.g. mean, skewness, kurtosis) from an active strategy, relative to that from alternative strategies such as holding index trackers. These alternatives are examined with respect to the distribution of final wealth, taking account of 'luck' and transactions costs for both strategies. The methodology adopted in this study moves the debate on persistence closer to the practical issues surrounding the implementation of *ex-ante* investment strategies by fund investors. Fourthly, persistence is examined using a wide variety of alternative sorting rules in an attempt to find repeat winners (and losers). Finally, a survivorship bias free and more comprehensive UK data base of mutual funds over a long data period is used. This has not always been the case in earlier persistence studies.

The third major contribution of this study lies in the area of mutual fund market timing. The question of market timing has attracted little attention among UK studies of fund performance. Treatment has been limited to the rather basic regression based tests of Treynor-Mazuy (1966), (TM), and Henrikkson-Merton (1981), (HM) as applied by Fletcher (1995), for example. In this study, a nonparametric test procedure is applied to UK funds for the first time. This procedure has been applied to US mutual funds by Jiang (2003). The nonparametric procedure has several advantages: the technique isolates the quality of a fund manager's timing information from the aggressiveness of his response while the TM and HM methods do not. Here, the quality of the manager's timing signals is of more interest to the investor. In addition, the nonparametric method is more robust to the frequency at which the manager times the market unlike the TM and HM tests which assume the fund's timing frequency is the same as the frequency of the sample data, as discussed in Bollen and Busse (2001) and Goetzmann et al (2000) in the US case. The nonparametric approach is also based on less restrictive underlying assumptions than the TM and HM approaches.

Furthermore, this study then examines whether mutual fund managers are able to add market timing value to investors by the quality of the manager's *private* market timing information (timing signals) in excess of the information quality contained in publicly available information as examined by Becker et al (1999) and Ferson and Khang (2001) in the case of the US. The latter information is already available to the investor and as such may reduce the necessity of availing of mutual fund services.

Finally, the study of fund market timing here evaluates timing skill against a number of alternative benchmark market indices as information on precisely which market index funds attempt to time, if any, is not available.

The findings of this study make important contributions to the academic literature in terms of measuring abnormal performance, persistence and market timing. These findings are clearly useful for investors in search of profitable portfolio strategies in equity mutual funds and have obvious implications for the wider debate on market efficiency. The findings of this study also make an important contribution to current policy issues.

The absolute performance of mutual (and pension) funds and the relative performance of active versus passive (index) funds are central to recent policy debates, particularly in Europe. With increasing longevity and given projected state pensions, a 'savings gap' is predicted for many European countries in 20 years (Turner 2004, OECD 2003). An important question is whether voluntary saving in mutual and pension funds over the next 20 years will be sufficient to fill this gap so that those reaching retirement age have sufficient savings to provide an adequate standard of living.

In recent theoretical and empirical work, the allocation across different asset classes (mainly bonds versus stocks, but in principal across all asset classes) has been examined in an intertemporal framework. The 'rule of thumb' that the percentage investment in risky assets (stocks) should equal '100 minus your age' is not robust either in the face of uncertain income (which gives rise to hedging demands – Bodie, Merton and Samuelson 1992, Campbell and Viceira 1999, Viceira 2001) or when return predictability is present (Brennan et al 1997, Campbell et al 2003) or when there is uncertainty about parameters in the prediction equation for returns (Barberis 2000, Xia 2001). In practice, the lack of a consensus 'model' of asset allocation at both the 'strategic' and 'tactical' level is starkly illustrated by Boots (the UK chemist) switching all its pension fund assets into bonds in 2001 (for strategic not market timing reasons), while most UK pension funds continue to hold around 70% of their assets in stocks. In the US, participants in 401(K) retirement plans (Benartzi and Thaler 2001), when faced with the choice between several funds each of which has alternative proportions of stocks and bonds, tend to use a simple 1/n allocation rule - so the actual allocation to each asset class is not determined by any sophisticated optimization problem and is changed infrequently. Such naïve asset allocation decisions may carry over to investment in UK mutual funds (and even trustees' decisions for UK pension fund asset allocations), so that poor funds survive and exacerbate the savings gap.

The Presidential Commission on Social Security Reform (2001) and the State of the Union Address (2005) envisage the part-privatization of US Social Security. This will increase debate on all aspects of the fund management industry, particularly in the light of the 'market timing' abuses uncovered in the US by New York Attorney General Elliot Spitzer (see also Goetzmann, Ivkovic and Rouwenhorst 2001) - which has reduced confidence in the financial service sector's ability to provide adequate and fair treatment of retail investors. In the UK, the continuing switch from defined benefit to defined contribution pension schemes will strengthen the argument for a closer analysis of active versus passive strategies (as well as the competence and independence of trustee governance arrangements-Myners 2001).

The Financial Services Authority (FSA) in the UK is concerned that (retail) investors may be misled by mutual fund advertising. In its 'comparative tables' it currently does not enter a fund's ranking *vis-a-vis* competitor funds, in terms of (raw) returns. The FSA believes this could encourage more investment in funds which may simply have high returns because they are more risky (Blake and Timmermann 1998 and 2003 and Charles River Associates 2002).

Therefore a key element in these policy debates is the attractiveness of savings products such as mutual funds and in particular the choice between actively managed and passive (or index/tracker) funds. It is important to evaluate the relative performance of actively managed funds to determine the extent to which such funds truly add value to investors/savers as a means of efficiently allocating their scarce resources to saving instruments for the future. To the extent that any 'savings gap' is to be filled by investment in mutual funds, the need to evaluate risk adjusted performance in a tractable and intuitive way, while taking account of the inherent uncertainty in performance measures, will be of increasing importance.

This study proceeds as follows: Chapter 2 presents a review of the existing literature in the areas of mutual fund performance in security selection, persistence and in market timing. It also reviews literature findings in relation to the role of momentum, fund capital flow, fund expenses and other fund characteristics in performance. Chapter 3 presents detailed descriptive statistics of the data used in this study along with information on data sources. It also provides a description of the UK mutual fund industry generally in terms of its size, asset allocation, client type etc. Chapter 4 describes different classes of models of fund performance measurement. It then applies all of these models to the data set of this study. Estimation diagnostics are used to select 'best fit' models from within each class of model for use in subsequent chapters in applying the bootstrap and persistence testing methodologies of the study. The remainder of the study may be sectioned into three parts: Part 1 relates to the bootstrap evaluation of fund performance. Here chapter 5 describes the bootstrap methodology while chapter 6 presents empirical findings from the implementation of the methodology to the data set of UK mutual funds. Part 2 relates to evaluating persistence in fund performance. Here chapter 7 outlines the persistence testing methodology while chapter 8 reports empirical findings from its implementation. Finally, part 3 examines the market timing performance of funds. Here chapter 9 describes the nonparametric market timing tests to be applied while chapter 10 presents the empirical findings. Chapter 11 concludes.

CHAPTER 2

LITERATURE REVIEW

This chapter presents a review of the existing literature in the area of mutual fund performance in relation to stock picking ability, persistence and market timing. It also reviews literature findings in relation to the role of momentum, fund capital flow, fund expenses, fund size and fund turnover in performance. Data issues and methodological issues such as survivorship bias and look-ahead bias are also discussed. While performance is usually thought of as stock picking ability, ie security selection or 'selectivity', this study also examines market timing ability and the extant literature on this subject is also discussed here. One of the core questions to be evaluated in this research is whether stock picking ability exists after correctly controlling for sampling variability in the performance measures at each point in the performance distribution and this recent development in the literature is also presented.

2.1 Survivorship and Look-ahead Bias in Evaluating Mutual Fund Performance

There are two types of bias that have the potential to obscure the findings of performance analyses, ie survivorship bias and look-ahead bias. Survivorship bias is a property of the sample selected. Conditioning performance results on funds which are in existence at the end of the sample period ignores the attrition effect, i.e. to incorporate the performance of funds which disappear at some point prior to the end of the sample period. This sample property commonly arises because many (particularly commercial) mutual fund databases fail to include the returns of 'dead' funds. If poor performance is a cause of nonsurvivorship then conditioning findings on surviving funds may induce an upward bias in average performance findings and in the persistence of performance.

Look-ahead bias arises even when the sample of funds includes nonsurviving funds but where the methodology requires funds to have existed for a minimum length of time in order to be included in the analysis. This restriction improves the statistical properties of estimated performance measures but may bias performance findings upwards as only funds which have been skilled enough to survive for the minimum length of time are included.

Literature findings in relation to survivorship and look-ahead bias are reported where relevant in the discussion to follow.

However, this study employs a database which contains the performance records of both surviving *and* nonsurviving funds. Hence this study provides a full performance analysis of the UK mutual fund industry and controls for survivorship bias. In testing for persistence, look-ahead bias is also mitigated by the selected methodology, (see chapter 7).

2.2 Performance and Persistence Among UK Equity Mutual Funds

An important contribution to the literature on UK equity mutual fund performance is that of Quigley and Sinquefield (1999). The study examines the monthly performance of all UK unit trusts in existence at any time between 1978 and 1997, a total of 752 funds including nonsurviving funds. The study uses returns calculated bid price to bid price, i.e. gross of load fees and other transactions costs such as stamp duty, dealing commissions, and the bid/offer spreads of the underlying securities but returns are net of annual charges to customers. However, the authors later refer to whether out-performance by customers would have been possible gross of annual charges).

Forming an equal weighted portfolio of the unit trusts (surviving and nonsurviving) and estimating a single factor CAPM the authors report an aggregate alpha performance measure of –0.04 basis points per month. This indicates that overall the mutual fund industry under-performed the benchmark market index. This under-performance is worsened when the size and value exposures are taken into account in a Fama and French three-factor model where the alpha measure falls to a statistically significant –0.09 basis points per month. The deterioration in performance on moving to the three-factor model is consistent across all four investment objective classes of funds examined. These investment objectives are growth stock, income stock, general equity and smaller companies funds. Equal weighted portfolios of unit trusts within each of these classes all exhibit negative alphas. The deterioration is most notable among small company funds and income stock funds.

Quigley and Sinquefield (1999) examine both surviving and nonsurviving funds and report the extent of the survivor premium and survivor bias. They report a survivor premium of 2.31%. This is the difference between the annual compound raw returns of equal weighted portfolios of surviving and nonsurviving funds. The reported survivor bias of 0.7% is the difference between the annual compound returns of the surviving funds and the full set of both surviving and nonsurviving funds. This substantial difference highlights the dangers of survivorship bias described above.

Quigley and Sinquefield (1999) examine persistence in fund performance by replicating the methodology of Carhart (1997) and Hendricks, Patel and Zeckhauser (1993). Each year, they form ten equal weighted portfolios of unit trusts based on the decile rankings of the funds' raw returns over the previous year. Each decile portfolios is then held for one year. The procedure is repeated recursively each year. This generates the holding period returns of the ranking period's decile portfolios. The spread in the annual compound raw return between the best and worst portfolio is 3.54%. While this initially seems to point to an easy 'beat the market' strategy, in fact pursuing this strategy involves an annual turnover of 80% in the composition of the top portfolio and with a bid/offer spread of up to 5% in many cases gains would be eliminated. In addition, when these portfolios are adjusted for risk in a three-factor model, the alphas of the ten portfolios do not suggest significant persistence in (abnormal) risk adjusted performance. The alphas from the top two portfolios, while positive, are statistically insignificant at the 5% level. By contrast the negative alphas of the bottom four portfolios are all statistically significant. This finding echoes that of a number of US studies (Carhart 1997 and Malkiel 1995) in that first, pursuing persistence strategies involves high turnover and second, in risk adjusted terms poor performance persists but good performance does not.

Quigley and Sinquefield (1999) further explore persistence in risk adjusted performance using a similar procedure to above where they sort unit trusts each year into decile portfolios by the three-factor alphas estimated over the previous three years and hold portfolios for one year. The spread in the annual compound raw return between the top and bottom portfolios is 2.95% while the spread in the three-factor alphas is 0.27% per month. The ten post-formation alpha measures continue, for the most part, to descend in size. However, only the top two have positive alphas, neither of which is statistically significant at 5%. The remaining eight three-factor alphas are negative and the bottom two are significantly so. Once again this leads to the conclusion that poor (risk adjusted) performance persists but good performance does not.

Interestingly, the authors go on to investigate the patterns of persistence beyond one year. That is, they compare the performance of the above ten portfolios, ranked by three-factor alpha, for different holding periods post formation. By year 3 the pattern of persistence has almost entirely disappeared where the rank correlation between pre and post formation alphas falls to 0.12.

The paper also undertakes a closer examination of the size effect. The authors rank all unit trusts by the loading on the size risk factor in a three-factor model based on the prior three years and divide this ranking into three groups. Within each size group they rank unit trusts into three further subgroups ranked by the three-factor alpha, form equal weighted portfolios and repeat recursively annually to test whether there is a relationship between persistence and size. The authors find that within the category of unit trusts with the highest exposure to size risk, i.e. unit trusts investing in small stocks, the top two portfolios ranked by three-factor alpha continue to have positive and statistically significant alphas in the year after formation. However, the bid/offer spread in these unit trusts is almost three times the size of the alpha measures. Therefore, while the finding provides some modest evidence against market efficiency at the level of the fund manager it is not exploitable by the fund customer. However, this analysis revealed persistence in poor performance irrespective of size exposure. This indicates once again a stronger tendency for losers rather than winners to repeat.

The study also involved tests of the hypothesis that the market for small stocks is less efficient. Unit trusts are again ranked by the loading on the size factor from regressions over the previous three years of data and decile portfolios are formed based on this ranking. These portfolios are held for one year and re-formed. In the case of all ten portfolios ranked by size exposure, the alpha performance estimate is negative. If the market for small company stocks is less efficient it was not exploited. As the authors remark "if the small-company unit trusts were horses, they'd be glue", [p5].

Quigley and Sinquefield (1999) highlights an important point in relation to the specification of the value versus growth risk factor in the three-factor model when applied to UK mutual fund data. The paper ranks trusts by the loading on the value factor from regressions over the previous three years. Ten portfolios are formed based on this ranking. These portfolios are held for one year and this process is repeated. The authors show that the persistence in the relative exposure to the value factor is weak. This suggests that UK unit trusts do not have a consistent exposure to value stocks.

Blake and Timmermann (1998) is a further important contribution to the literature on UK mutual fund performance. The study examines a total of 2,375 mutual funds, not all of which are restricted to investing in UK equity. Nonsurviving funds constitute 973 of the total number of funds. Returns are measured gross of transactions costs and management fees. The research focuses on the sample period February 1972 to June 1995, a slightly earlier sample period than Quigley and Sinquefield (1999).

Blake and Timmermann (1998) report a substantial survivor bias of 0.8% per annum. This is the difference between the monthly return on surviving funds and the weighted mean monthly return on surviving and nonsurviving funds, weighted by the numbers of survivors and nonsurvivors. The average of this monthly time series is taken and then annualized to yield 0.8% per annum. This is broadly consistent with the figure of 0.7% reported by Quigley and Sinquefield (1999), notwithstanding a slight difference in definition and again highlights the danger of survivor bias, particularly in the UK case.

However, on the question of performance persistence the findings of Blake and Timmermann (1998) differ in a number of respects from those of Quigley and Sinquefield (1999). The methodological approaches also differ. First, the risk adjusted return is the α_i coefficient estimated by regressing the fund's excess return (relative to the Tbill rate) on a constant, excess returns on the (broad) stock market index, excess return on a small capitalization stock index over the broad market index and excess returns on a five-year UK government bond:

$$(2.1) \qquad r_{it} - r_{ft} = \alpha_i + \beta_{m,i}(r_{m,t} - r_{f,t}) + \beta_{s,i}(r_{s,t} - r_{m,t}) + \beta_{5,i}(r_{5,t} - r_{f,t}) + \varepsilon_{i,t}$$

Blake and Timmermann restrict their analysis of performance persistence to the UK equity and balanced sectors, ie 855 funds, for which the right hand side variables in (2.1) are good benchmarks. Second, Blake and Timmerman (1998) follow the procedure of Hendricks et al (1993) and sort the funds each month into quartiles based on the abnormal performance from (2.1) over the previous 24 months. They form two equal weighted portfolios of funds from among the top and bottom quartiles and hold these portfolios for one month. This recursive procedure generates the holding period returns of the top and bottom quartile of funds ranked by abnormal performance. The abnormal performance from these two portfolios are in turn derived from (2.1). The authors also analyze performance persistence by forming portfolios of the top and bottom quartiles of funds with the weights suggested by modern portfolio theory, rather than equally weighted. The paper also carries out this procedure separately for funds investing in five sectors: equity growth, equity income, general equity, smaller companies and a balanced sector.

With the exception of the balanced sector, the recursive portfolios derived from the top quartile of funds in all sectors produced positive abnormal returns over the holding period. The recursively formed portfolios of the bottom quartile of funds produced negative abnormal returns over the holding period. (This finding was robust with respect to how the portfolios were weighted). This indicates persistence in performance among both the top and bottom performing funds. This finding of persistence was found to be statistically significant among funds investing in growth stocks and smaller company stocks.

The persistence finding among top performers is in contrast to Quigley and Sinquefield (1999) while the finding of persistence among the worst performing funds is consistent between the two studies. Comparisons between these two studies is complicated by both differing measures of abnormal return and differing frequencies of reforming portfolios. For an investor to exploit this apparent persistence anomaly may require considerable portfolio rebalancing thus incurring significant transactions costs which may eliminate the abnormal return. As Blake and Timmermann (1998) carry out their analysis using returns on a bid to bid basis the returns exclude some transactions costs.

The measure of performance (risk adjusted or otherwise), as a possible source of difference in persistence findings between studies is picked up in Fletcher and Forbes (2002). The authors apply the contingency table testing procedure of Brown and Goetzmann (1995). This involves designating funds as winners or losers according to whether their annual excess returns rank above/below the sample median ranking and testing whether fund rankings are then independent over consecutive periods. The paper finds significant persistence. When returns are measured as excess returns over a market index, however, persistence is driven primarily by repeat underperformance. Switching testing methodologies to the Carhart (1997) procedure and estimating abnormal performance from the CAPM or APT, Fletcher and Forbes (2002) report significant persistence. However, this persistence disappears when performance is evaluated by the Carhart (1997) four-factor model which, in addition to market, controls for size, value and momentum risk attributes. Furthermore, measuring performance by a conditional model, ie time varying betas, suggests a reversal, rather than persistence, in performance.

Blake and Timmermann (1998) highlight a further interesting feature regarding performance: underperformance intensifies as the fund approaches its termination date while the authors find some (weaker) evidence that funds outperform their sector peer groups during their first year of existence. In each month prior to the death of all nonsurviving funds within each sector[1] Blake and Timmermann calculate abnormal return as the difference between the nonsurviving fund's return and the return on an equal weighted portfolio of all funds in existence at that time within the same sector. Abnormal returns are therefore measured relative to the returns of the fund's peer group. Assuming an alignment in time, the authors form equal weighted portfolios of these nonsurviving funds in the months preceding their death and calculate the average return on the portfolios 6, 12 and 24 months prior to the death of the funds. Underperformance is seen to intensify as the termination of the fund approaches and negative returns of –30 basis points per month are common to many sectors. During the final year of a fund's existence it underperforms relative to the universe of funds in existence at the same time by an average of –3.3% per annum. Using a similar method to examine the performance of funds in the months after their launch, the authors report an average abnormal performance (across all sectors) of 0.8% per month in the first year after the launch of new funds.

Fletcher (1997) is a slightly earlier study of the UK unit trust industry which focuses on performance between 1980 and 1989. This paper limits its analysis to funds which have survived for two years or more. The methodology followed by Fletcher in testing for persistence is similar to that of Quigley and Sinquefield (1999) and Blake and Timmermann (1998). However, Fletcher applies a different measure of risk adjusted performance in testing persistence, using Arbitrage Pricing Theory (APT) benchmark portfolios estimated by an asymptotic principal components technique as outlined in

[1] The authors carry out this analysis on 20 separate sectors, some of which include mutual funds investing in foreign assets.

Connor and Korajczyk (1986). Connor and Korajczyk show that that the k factors in an APT framework can be proxied by the first k eigenvectors (corresponding to the largest k eigenvalues) of the cross-product matrix of excess returns $\Omega^n = (1/n)R^{n'}R^n$ where R^n is a $(n*T)$ matrix of n assets with T observations. This is asymptotically valid as $n \rightarrow \infty$ where assets may be all securities in the market. The link to principal components analysis lies in the fact that the first k eigenvalues and their corresponding eigenvectors correspond to the variances and parameters respectively of the first k principal components.

Fletcher (1997) first categorizes a sample of 101 unit trusts by the following characteristics in 1980, (i) market value, (ii) CAPM beta – based on the past five years (iii) skewness - the coefficient on the squared market excess returns in a regression of unit trust excess returns over the previous five years, (iv) the sum of unit trust returns over the previous five years and (v) the variances of unit trust returns over the previous five years. This stems from the intuition of Grinblatt and Titman (1989a) that a stock's factor loadings are correlated with these characteristics of the firm. Within each of these five categories Fletcher sorts all of the unit trusts into ten equal sized portfolios of trusts based on the ranking in ascending order. (the lowest ranking portfolio contains one additional fund). Thus 50 portfolios are created. These portfolios are held for five years and a time series of returns are generated. The process is repeated in 1985 and the resulting portfolios are held for a further 5 years. The first five eigenvectors from the cross product matrix of the returns of these 50 portfolios are taken as the risk factors in a multi-factor model. The constant term in a regression of unit trusts on these factors is taken as the risk adjusted measure of return, ie a multi-factor alpha.

Using this risk adjusted return measure, Fletcher (1997) assesses the performance of funds with investment objectives of growth, income and general equity. Equal weighted portfolios of unit trusts within each investment style were constructed and the alpha measure estimated. None of the alpha measures proved to be significantly different from zero while a chi-square multivariate test also rejected the hypothesis that the performance of the three investment styles were significantly different from one another. Fletcher also tests the hypothesis that performance may be affected by factors such as size, annual charges of the fund and initial charges of the fund. In the case of the latter two factors the hypothesis under examination is that if performance is assessed using gross returns (as is the case in Fletcher's study) then we would expect funds with higher initial and annual charges to earn higher returns to compensate investors. That is, do charges provide a reliable indication to the investor as to the 'quality' of the fund's management. At the beginning of the sample period (1980), within each of the categories of size, annual charges and initial charges five equal weighted portfolios are formed based on a ranking according to these criteria and these portfolios are maintained throughout the entire sample period. None of the risk adjusted return measures of these portfolios proved to be significantly different from zero or significantly different from one another.

To test persistence in performance Fletcher (1997) carries out a battery of tests. All unit trusts are ranked into quintiles according to the risk adjusted return measure described previously and equal sized portfolios of trusts are created within each quintile. Portfolios are formed and evaluated based on five-year and two-year ranking and evaluation periods. Performance is evaluated by the same risk adjusted measure as used for the ranking. Fletcher finds very little evidence to suggest that past performance provides an indication of future performance. None of the portfolios exhibit statistically significant performance persistence regardless of the length of the ranking period. Likewise the chi-square tests of joint equality in performance usually fails to reject the null hypothesis.

The finding by Fletcher (1997) that superior performance does not persist is largely consistent with the findings of Quigley and Sinquefield (1999). However, Fletcher also fails to find evidence of persistence among poor performers. There are two possible explanations for this: first, Fletcher's study is limited to the 1980s while Quigley and Sinquefield's study spans a longer time period including parts of the 1970s and 1990s. Second, Fletcher's measure of risk adjusted return differs considerably from that adopted by Quigley and Sinquefield (1999). Therefore, direct comparisons are complicated. However, Fletcher (1997) does test the robustness of his findings by repeating the performance persistence tests described above using a Jensen's alpha performance measure. Once again, however, none of the Jensen alpha estimates are significantly different from zero irrespective of the length of the ranking period. Overall, Fletcher's study suggests there is little economic significance in the predictive power of past UK unit trust performance.

A difficulty with the performance persistence tests described above in Quigley and Sinquefield (1999), Blake and Timmermann (1998) and Fletcher (1997), i.e. assessing persistence through a recursive portfolio formation scheme, is that it aggregates the data considerably rather than looking at persistence at the individual fund level. This question is picked up in Lunde, Timmermann and Blake (1999) and is also examined by Allen and Tan (1999), the latter evaluating UK investment trusts rather than unit trusts. Lunde et al (1999) first sort the set of UK equity mutual fund returns into quartiles by a peer-group adjusted return measure similar to that used by Blake and Timmermann (1998) above. Funds are sorted based on the previous 36 months of returns. For each quartile, the proportion of funds that fall into a given quartile over the following 36 months is recorded and a contingency table of transitional probabilities is constructed. Under a null hypothesis of no persistence, all of the transitional probabilities should equal 0.25. Lunde at al report that this null is clearly rejected when looking at the full set of both surviving and nonsurviving funds. The probability that the worst performing (bottom quartile) funds will remain in the bottom quartile is reported as 0.332 while the probability of repeated top performance is 0.355. When surviving funds are examined in isolation, ie a more homogenous group, these probabilities fall to 0.284 and 0.317 respectively.

Allen and Tan (1999) examine performance and performance persistence among the (weekly) returns of 131 investment trusts over the period 1989 – 1995. Although Allen and Tan concentrate exclusively on investment trusts, rather than unit trusts and OEICs as in this study, a brief discussion of their findings is nevertheless helpful as the literature on UK fund performance generally is relatively small. Allen and Tan evaluate persistence at the individual fund level. The authors rank funds over a one year period by both raw (cumulative) returns and risk adjusted returns (Jensen's alpha from a CAPM estimation) and select 'winners' and 'losers' according to the median ranking. They then perform a number of tests of persistence in ranking over the following one year period: a contingency table analysis of winners and losers from one period to the next and a number of rank correlation tests. Both sets of tests reject the null hypothesis of no persistence at a 5% significance regardless of whether returns are measured on a raw or risk adjusted basis. The authors provide some weak evidence of negative or reverse persistence if returns are measured over shorter time periods.

Allen and Tan (1999) restrict their analysis to funds which survive for more than one year. This has the disadvantage that it may induce a slight look-ahead bias in the results. However, it has the advantage that it excludes funds with a low number of observations and consequently excludes those funds which may exhibit high sampling variability in the Jensen's alpha estimate. Such high sampling variability makes the conclusions regarding performance and performance persistence less reliable. In an attempt to control for this possible look-ahead bias, however, Allen and Tan postulate that funds with higher variances of returns are less likely to survive but that among surviving funds higher variance funds are likely to be the top performers. To control for this the study ranks funds according to the variance of returns over the entire sample period and sorts funds into high and low variability according to the median. The tests described previously are repeated for the high and low variance sub-groups. The repeat winner phenomenon is found to exist equally among both high and low variance funds leading the authors to conclude that the look-ahead is not a significant weakness.

Leger (1997) examines the performance persistence, *inter alia*, of 72 UK investment trusts in four non-overlapping five-year samples between 1974 and 1993. Without adjusting for survivorship bias in the sample of funds, Leger estimates Jensen's alpha in the four separate sub-samples. Some evidence, albeit relatively weak, is found in support of abnormal performance in the first half of the sample period up to 1984 but this subsequently disappears. Persistence is assessed by calculating the number of funds with positive or negative abnormal performance in two, three and four consecutive sub-periods but is not restricted to looking at performance in the successive period only.[2] Of the 72 funds in the study, none of the funds records a positive abnormal performance in all four consecutive periods, 2 (4) of the funds report positive performance in 3 (2) consecutive periods. No fund records negative abnormal performance in two or more consecutive periods. This represents quite a robust finding against the existence of performance persistence and is at variance with the results reported by Allen and Tan (1999) which also looked at investment trusts but over the shorter time period of 1989 – 1995. This may suggest that the relatively strong findings of persistence in the Allen and Tan study may simply be specific to the shorter sample period. However, direct comparison between the two sets of findings is complicated by different measures of abnormal performance.

The contingency table tests of Lunde at al (1999) and Allen and Tan examine persistence at a more disaggregated level. However, tests based on contingency tables (e.g. log-odds ratio, Wilcox test) involve a 'frequency count' of fractiles of repeat winners WW and losers LL (relative to the number of WL and LW's) in two different periods. But any

[2] Leger (1997) measures abnormal performance in a model which also specifies a market timing variable. Hence the performance measure is not strictly comparable with that of Allen and Tan (1999) as abnormal performance in terms of stock picking and market timing are not necessarily independent: Henriksson (1984).

measured 'persistence' only involves relative frequencies, so we cannot directly assess the economic significance of the results (e.g. in terms of the risk adjusted returns to the persistent winner/loser portfolios) and it is often not clear how this may be exploited by investors. Also, in the contingency table (and rank regression/correlation approaches), measured persistence may be due mainly to repeat losers rather than repeat winners. In addition (Spearman) rank correlations treat each point in the ranking equally and lack power against the hypothesis that predictability in performance is concentrated in the tails of fund performance – an issue taken up in this study. Further, rank correlations or a regression of pre- and post-sort alphas can be used to establish predictability. However, although there may be a high correlation between the alphas of past fractile ranked funds and their subsequent alphas, nevertheless all of the post-sort alphas may be negative, indicating predictability but poor future abnormal performance for all fractile portfolios. The above approaches can be used to establish statistical predictability but investors are presumably more interested in the future absolute performance of both winners and losers (taken separately).

In an unusual approach not commonly seen in this literature, Rhodes (2000) addresses this issue of examining persistence over consecutive periods only. Rhodes examines UK equity unit trusts between 1980 and 1998 with returns measured before costs. Each year Rhodes sorts all the unit trusts into quintiles ranked by raw returns and assigns a score to each quintile. For example, quintile 5 (lowest performance) may be assigned a score of 2 while quintile 1 may be assigned a score of 10, with linear scaling in between. The study relies on a framework where it is assumed that the investor's utility each period is directly related to this score. A utility function provides a utility measure. This utility function is assumed to exhibit diminishing marginal utility. Therefore, higher relative performance increases utility but at a diminishing rate, possibly reflecting the increased risk associated with high relative performance. The concave shape of the utility function implies that the average utility over two periods from a score of say 2 in period 1 and 6 in period 2 is less than the average utility from a score of 4 in both periods even though the average performance is the same in both scenarios. That is, given the same average performance outcome less volatility is preferred to more.[3] For each fund the average utility may be calculated across all the years for which returns are available.

[3] Note there is a simplification in the scoring system in Rhodes (2000). The scoring system implies that the gap in performance between neighbouring quintiles is equal and constant over time. This is not necessarily what we would observe if actual raw returns within the quintiles were examined.

Funds which consistently remain in the higher(est) quintiles will produce the highest average utility. Rhodes demonstrates that if relative performance is random then the cross sectional distribution of average utilities is normal. Hence the null hypothesis of no persistency can be tested by the Jarque- Bera test of normality.

Rhodes (2000) finds that the distribution of average utilities is strongly leptokurtic and rejects normality at 1%. The bunched nature of the utilities suggests that there is less movement by funds in and out of quintiles than would be expected by random relative performance alone.

Rhodes (2000) identifies a change in cross-sectional performance after the stock market turbulence of 1987. In particular, each year Rhodes calculates the relative performance of each fund by dividing its average monthly return by the average return of all funds that year. For a selection of years both pre-1987 and post-1987 the author plots histograms of this relative performance measure. The pre-1987 histograms have wider right tails, ie they exhibit a far greater number of funds with high relative performance compared to the post-1987 histograms. Since more recent persistence information is used by investors to decide which funds to invest in, Rhodes repeats the above normality tests on the average utilities in the post-1987 period only. In this case, the Jarque-Bera test fails to reject normality indicating a lack of evidence of persistence post-1987.

Rhodes (2000) carries out the normality tests with a number of alternative utility function specifications to change the level of risk aversion. The study's conclusions are generally found to be quite robust.

A further issue in the area of performance measurement relates to the choice of benchmark portfolio in a single factor (market) model. Roll (1977 and 1978) argued that the CAPM equilibrium model is untestable since the market portfolio is unobservable. Fletcher (1995) evaluates both selectivity and market timing performance of 101 UK unit trusts over the period 1980 – 1989 employing four separate benchmark portfolios: the Financial Times All Share (FTA) index, Financial Times 100 index, (FT 100), an equally weighted index of securities and the size index of Huberman and Kandel (1985). Fletcher (1995) reports that the selectivity measure is on average positive and almost always significant for all benchmark portfolios. Once again, however, this is a relatively short sample period.

The question of conditional versus unconditional models of performance is addressed by Black, Fraser and Power (1992) in their study of 30 UK unit trusts during the 1980s. The authors cast the single factor CAPM in 'state space' form and examine the time varying characteristics of the market factor loading and performance by applying the Kalman Filter. Only 3 of the 30 funds under consideration are statistically significantly found to have time varying market betas. However, the extent of abnormal performance among the 30 unit trusts is found to be considerably higher (21 out of 30 funds). Also of interest is that among the three funds which are found to have time varying betas, the betas are nonstationary, ie there is no tendency towards mean reversion.

The studies described above encompass the main issues and results that arise in the literature on performance, abnormal performance and performance persistence among UK mutual funds. There is a small number of additional studies which also examine the UK mutual fund industry using similar procedures to those already described and these report broadly similar results. A brief description of these is provided below.

The WM company (1999) examined surviving funds in the UK income and growth sector over the period 1979 – 1998. Using only raw returns the study analysed the persistence of the top quartile of funds over five year periods. Not surprisingly the proportion of funds remaining in the top quartile in subsequent periods quickly declines. When there was no overlap in the sample periods the study indicates that the proportion of funds retaining a top quartile ranking is no more than would be expected by chance alone. The conclusion is that historic relative performance is not a good guide for investors to picking funds.

In addition to unit and investment trusts, the performance of pension funds has also been examined. Blake, Lehman and Timmermann (1998) examined over 306 pension funds over the period 1986 – 1994. This study indicates that persistence inferences are more difficult among pension funds: investment policy is more restricted and performance is determined by the asset mix rather than by stock selection or by market timing. Blake et al report surprisingly little cross sectional variation in (raw) returns to strategic asset allocation, security selection or market timing. Brown, Draper and Mckenzie (1997) and Thomas and Tonks (2001) report broadly similar findings although the latter study does report findings of over and underperformance during sub-periods. In a persistence testing methodology similar to Carhart (1997), Tonks (2004) finds evidence of persistence among UK pension fund managers where a zero net investment portfolio yielded statistically significant abnormal performance.

Comparing the results of UK mutual fund studies regarding performance persistence is complicated by a number of factors: differing sample periods of various lengths, differing measures of risk adjusted return, differing fund ranking schemes (deciles, quartiles, etc), differing ranking and evaluation periods, differing levels of data aggregation, survivorship bias and return measures (gross/net of costs). This notwithstanding, a reasonable characterisation of performance seems to be that repeat performance among top performers is small in effect and relatively short lived. It is doubtful that a significant exploitable persistence anomaly exists at the level of the fund investor net of the charges imposed by the fund. In comparison, the evidence is stronger that poor performance persists. Overall, an analysis of persistence may provide insight into some funds that are best avoided but says less about which funds to select.

Table 2.1 provides an overall summary of findings among studies of the UK mutual fund industry regarding performance persistence.

Table 2.1: Summary of Studies of UK Equity Mutual Fund Performance

Study	Period	Type of Funds	Risk Adjusted Returns	Controls Transactions Costs	Controls for Survivorship Bias	Performance Persistence
Allen and Tan (1999)	1989 – 1995	131 investment trusts	Yes	No	Indirect test is carried out	Yes, among both top and bottom performers.
Blake and Timmermann (1998)	1972 – 1995	2,375 unit trusts divided by 20 sectors.	Yes	No	Yes	Yes
Fletcher (1997)	1980 – 1989	101 unit trusts by sector.	Yes	No	Minimum of 2 years.	No
Fletcher and Forbes (2002)	1982-1996	724 unit trust	Yes	Yes	Yes	No
Leger(1997)	1974 - 1993	72 investment trusts	Yes	No	No	No
Lunde et al (1999)	1972 – 1995	As in Blake and Timmermann (1998)	Yes	No	Yes	Yes
Quigley and Sinquefield (1999)	1978 - 1997	752 unit trusts by sector.	Yes	Partially.	Yes	Yes, among poor performers.
Rhodes(2000)	1980 - 1998	Population of funds.	Yes, using a utility based risk adjustment	No	Minimum of 2 years.	Yes, over full sample. No, post-1987.
WM Company	1979 - 1998	UK income and growth funds	No	No	No	No

2.3 Studies of the US Mutual Fund Industry

Studies of the US mutual fund industry are far greater in number relative to the UK and Europe. The main issues to emerge from the US literature include (i) survivorship bias, (ii) different performance models including different measures of benchmark portfolios representing risk factors, (iii) different persistence testing methodologies including examining persistence from a conditional performance estimate, (iv) whether persistence is attributable to a momentum effect among the stock holdings of funds or whether it is due to persistent selectivity, (v) the changing nature of persistence over time, (vi) market timing and (vii) the relationship between fund performance and fund charges and other fund characteristics. This sections aims to provide a review of these issues from the research on US equity mutual funds.

2.3.1 Survivorship and Look-ahead Bias in Studies of US Mutual Funds.

Some US studies which find evidence of survivorhip bias and look-ahead bias include Brown, Goetzmann, Ibbotson and Ross (1992), Brown and Goetzmann (1995), Malkiel (1995), Elton, Gruber and Blake (1996a) and Shukla and Trzcinka (1992). These studies suggest that the degree of survivor bias can increase the evidence of persistence in performance significantly.

Malkiel (1995) examines a full data set of both surviving and nonsurviving funds between 1982–1990. The attrition rate among funds averages 14% per year. Malkiel records significant differences between the average returns (weighted by fund size) of surviving funds and the full set of funds. The annual average difference is 1.4%. The difference between the annual average returns (equally weighted) of surviving funds and nonsurviving funds is 6.15%.

Similarly, Elton, Gruber and Blake (1996a) examine the bias of presenting performance results based only on surviving funds by tracking all funds that would be classified as nonsurvivor funds between 1977 and 1993 due to name changes, investment policy changes or mergers. Specifically, Elton et al (1996a) apply two definitions of survival: first, that the fund is not merged during the sample period and second that the fund neither merges nor undergoes a change in investment policy during the period. The authors report performance results for both the group of survivor funds and the entire group of funds and reveal substantially higher average fund performance among survivor funds by both definitions of survival. The study also focuses on funds that merge by calculating a weighted average of the performance pre and post merger, weighted by the lengths of time in each state. One can then compare the average performance among this group of funds with the group of exclusively surviving funds. Once again, surviving funds substantially outperform merged funds leading to the conclusion that survivorship bias is present. Briefly, Elton et al (1996a) also estimate the extent of survivorship bias within different fund attributes. The paper finds the bias is particularly large among small stock funds which is consistent with the fact that a larger percentage of small stocks fail to survive relative to large stocks. Similarly, the paper also reports a large bias among growth stocks where the average performance estimate switches from positive to negative when switching from a survivor biased to unbiased estimate.

Brown, Goetzmann, Ibbotson and Ross (1992) examine survivorship bias and persistence using simulated returns. Monthly returns are generated for a four year period (using an assumed return generating process). A single period survival rule is applied, i.e. in *each* of the four years the worst performing funds are dropped from the sample where 'worst performing' varies from the lowest 5% to lowest 20% (under different scenarios examined). Estimating Jensen's alpha from two equal sub-periods and denoting funds as winners and losers according to the sample median ranking the authors test for persistence using a contingency table. Regression tests of the cross section of fund rankings in both sub-periods are also used. The simulation tests demonstrate that if fund returns are independently, but not identically distributed, and if a single period survival rule is applied then conditioning on survivorship induces spurious persistence.

However, in another simulation study, Carpenter and Lynch (1999) continue this line of investigation and conclude that when survival depends on performance over several periods, and with cross-sectional variation in fund volatility, a reversal effect (where funds revert from winner to loser status and *vice-versa*) dominates the spurious persistence effect found by Brown et al (1992). In the multi-period rule the authors drop the worst performing 3.6% of funds each year based on the funds' average four-factor alpha from the previous five years. (This figure is suggested by Carhart (1997) who finds that 3.6% of funds disappear from his sample in an average year). The paper applies a number of persistence tests including contingency tables, rank correlation tests and the recursive portfolio formation test of Carhart (1997), see section 2.3.2.

Carhart et al (2002) also examine the effects of a singe period and a number of multi-period survival rules imposed on a real data set of returns. The authors show that their survival rules (similar to those of the studies above) cause estimates of average annual performance to increase with sample length but at a diminishing rate. The study confirms the existence of performance persistence using raw returns, though this is lessened using a four-factor alpha, but shows that the evidence of persistence is weakened by conditioning on both survivorship and look-ahead bias. (The Carhart (1997) persistence test is used, see section 2.3.2). Carhart et al (2002) rationalize this by noting that a multi-period survival rule removes losers-losers in greater proportion than funds which revert from winner to loser or loser to winner and winner-winner funds. This leaves the sample more heavily comprised of 'reversers'. Grinblatt and Titman (1992a) and Hendricks et al (1993) argued a similar point earlier noting that when survival depends on several periods it will point to reversals because losers must revert to being winners in order to remain in the sample in the first instance.

The discussion of survivorship bias presented here is limited to the above as generally the findings in relation to the effects and extent of survivorship bias are quite consistent across the surveyed studies of the US mutual fund industry. Conditioning on survivorship in mutual fund data sets may lead to spurious findings in support of persistence though this finding is reversed if a longer term survival rule is examined. Critically, however, it should be noted that the above studies implicitly assume that survivorship is directly linked to performance. As was noted previously by Blake and Timmermann (1998), the majority of UK funds labeled 'nonsurvivors' are merged or taken over . In this study, 'nonsurviving' funds as a group are not shown to underperform surviving funds.

2.3.2 Performance and Persistence in the US Mutual Fund Industry

As is the case with studies of the UK mutual fund industry, the question of performance persistence attracts a great deal of attention among studies of the US mutual fund industry. In general there is stronger evidence in support of persistence among US equity funds compared to UK equity funds. (Volkman and Wohar 1994; Elton, Gruber and Blake 1996b; Goetzmann and Ibbotson 1994). However, positive persistence may be relatively short lived (Bollen and Busse 2002) and persistence tends to be more evident among poor performers than among top performers. (Hendricks et al 1993, Christopherson, Ferson and

Glassman 1998; Carhart 1997). However, persistent outperformance may have been more achievable during some time periods more than others, Malkiel (1995), while the source of persistence may lie in stock momentum rather than persistent selectivity, Carhart (1997), Chen, Jegadeesh and Wermers (2000). This section describes these discussions.

The Carhart (1997) study is an important, comprehensive and widely cited contribution to the literature and consequently is discussed in some detail here. Carhart applies the recursive portfolio formation methodology (similar to Hendricks et al, 1993) in his survivor bias free examination of US mutual fund monthly returns during the 1963 – 1993 period. All mutual funds are sorted into deciles based on lagged one-year raw returns, equal weighted portfolios of funds in each decile are formed and held for one year. This procedure is repeated recursively thus generating holding period returns for the decile portfolios. Carhart's fund returns are net of all operating expenses and security level transaction costs. Carhart first applies the CAPM to the above decile portfolios but it is clear that this does not explain the cross-section of returns. The CAPM betas on the top and bottom decile portfolios are almost identical and therefore the resulting Jensen's alphas exhibit as much dispersion as the simple returns.

Carhart (1997) goes on to apply a four-factor model (see chapter 4) to the ranked portfolios. The four-factor model explains much of the spread in performance among the portfolios with size (SMB) and momentum (PR1YR) explaining most of the variation. The top decile portfolios in particular are sensitive to the size factor suggesting that top performers hold more small stocks. The author reports that the momentum factor explains half of the spread between the top and bottom decile portfolios. However, the alpha measures from the four-factor models are negative for all portfolios and are significantly so for all portfolios ranked decile 3 or lower, (where decile 1 represents the top performance portfolio). This leads to the conclusion that on a risk adjusted and net return basis the only evidence that performance persists is concentrated in underperformance and the results do not support the existence of skilled fund managers.

Carhart (1997) also examines persistence ranking funds into deciles by four-factor alphas estimated over the previous three years and holding the resulting portfolios for one year and repeating recursively. The four-factor model is then applied to the decile portfolios of holding period returns. Again, evidence of persistence among top performing funds is not found to be statistically significant while in contrast underperformance is found to persist.

Elton et al (1996b) highlight a caveat in interpreting the persistence results from this recursive portfolio formation procedure. The composition of the top and bottom portfolios will change over time. Indeed Carhart's 1997 study revealed this to be strongly the case. Therefore using unconditional (constant parameter) models to measure risk adjusted performance may be inaccurate.

Carhart (1997) also applies a contingency table approach to test the persistence of one-year mutual fund ranking. This provides some weak evidence that the top and bottom decile funds persist. However, Carhart reports an 80% annual turnover in the composition of the top decile of funds while last year's winning funds frequently become next year's losing funds. However, Carhart demonstrates that whatever performance persistence might exist it is short lived. The returns on the decile portfolios plotted over the subsequent five years quickly converge. (That is where portfolios are not rebalanced annually). Similarly, Bollen and Busse (2002) apply a similar recursive portfolio formation approach using daily data and quarterly ranking and holding periods. Their findings suggest that post-ranking abnormal returns disappear for holding periods longer than one quarter and that superior performance is observable only when funds are evaluated several times per year.

Following evidence from earlier studies that momentum trading strategies can yield abnormal returns (see Jegadeesh and Titman 1993; Grinblatt, Titman and Wermers 1995), Carhart (1997) further investigates the momentum effect in mutual funds. As described, Carhart finds that the momentum factor explains half of the spread between the top and bottom performing portfolios as sorted on one-year lagged returns. This suggests that funds which pursue a momentum strategy may earn above average returns even if the four-factor model alpha is not abnormal. To test this Carhart ranks all funds based on the coefficient on the momentum variable but finds, in fact, that one-year momentum funds do not earn substantially higher returns than contrarian funds[4]. This suggests that the statistically significant coefficients on the momentum variable in the four-factor model does not necessarily imply that funds actively pursue a momentum strategy but rather that funds tend to hold many of the same stocks for a number of consecutive periods. The fact

[4] It is not clear whether this ranking is done recursively each year with resulting portfolios held for one year or whether it is done once over the full sample period.

that the momentum explains some of the cross-section in the ranked portfolios is picking up on a momentum effect in the underlying stocks themselves, which funds may either purposefully or accidentally hold over consecutive periods. This momentum effect in the fund's stock holdings rather than an actively pursued momentum strategy each period as a source of higher fund returns is further supported by the fact that the latter would incur substantial transaction costs which one would expect to detract from returns.

Chen, Jegadeesh and Wermers (2000) pick up this question of persistence and momentum. Their data set includes information on constituent stocks and the buy and sell trades of winning and losing mutual funds between 1975-1995, where winners and losers are defined as the top and bottom quintile of funds as ranked quarterly by past one-year raw returns.

The paper examines the *past* returns of the *current* constituent stock holdings of winning and losing funds and finds that stocks currently held by winning funds have higher past returns, or momentum, than stocks held by losing funds. The raw returns of the winning funds go on to outperform the returns of losing funds for the subsequent two quarters. The risk adjusted returns of winning funds go on to outperform those of losing funds for the subsequent (one) quarter. (Returns are adjusted here by the application of the benchmarks in Daniel et al (1997) discussed further below).

However, Chen et al (2000) also show that the past returns of winning funds' 'buys' are substantially higher than the past returns of losing funds' buys. This indicates that winning funds, relative to losing funds, tend to add past winning stocks to their portfolios. However, Chen at al point out that the buys of winning funds do not outperform the buys of losing funds in subsequent quarters by a statistically significant amount. So while winning funds do appear to pursue a momentum strategy, this does not appear to be the source of their success. These findings differ slightly with the assertion in Carhart (1997) above that winning funds do not actively pursue a strategy of buying past winning stocks but supports Carhart's assertion that a momentum effect in stocks already held contributes to the outperformance of winning funds. On the other hand in the case of losing funds, stocks bought exhibit higher past and future returns than the full set of losing funds' stock holdings. Indirectly, this indicates that losing funds are stuck with past losers which supports Carhart's argument that this hurts their future returns by a momentum effect.

Wermers (2003) also deals extensively with the momentum question but reports strong evidence that fund flow-related buying pushes up stock prices. The stock buys of winning funds, in response to persistent inflows, go on to yield returns which beat their risk characteristics in the subsequent four years. Cross-sectional regressions indicate that these abnormal returns are strongly related to the fund inflows.

The Chen at al (2000) paper also draws some other interesting conclusions from its extensive data base. First, the stocks most widely held by the mutual fund industry generally do not outperform the stocks least widely held. Second, stocks which are newly bought tend to outperform stocks newly sold. This is true of both winning and losing funds and is not unique to particular stock characteristics such as size (market capitalization) or value (book to market equity ratio). Third, stocks actively traded by funds tend to outperform those passively held from prior periods but funds that trade more frequently exhibit at best only marginally better stock selection skills which is clearly linked to associated transactions costs. Finally, the evidence of persistence at the level of raw returns tends to be more concentrated among growth funds rather than income funds.

Jegadeesh and Titman (2001) evaluate the various explanations for the profitability of momentum trading strategies previously identified in the literature. The authors offer evidence to refute the criticism that the momentum anomaly is a product of data mining by demonstrating that profitable momentum strategies persisted in the 1990s after initially being identified in the 1980s (Jegadeesh and Titman 1993). Conrad and Kaul (1998) argue that the apparent momentum arises because of cross-sectional variation in expected returns in adjacent time periods and is simply a compensation for risk. In direct contrast, others such as Daniel, Hirshleifer and Subrahmanyam (1998), Barberis, Shleifer and Vishny (1998) and Hong and Stein (1999) present behavioural models (see Barberis and Thaler 2003 for expanded survey) which argue that the momentum effect arises because of a delayed over-reaction to information that pushes the prices of winners (losers) above (below) their long term values and in subsequent periods the returns of losers should exceed that of winners as prices re-adjust to the over-reaction. Hence such models predict that in the 'postholding' period returns to a momentum strategy should be negative. Jegadeesh and Titman (2001) do indeed find evidence that the performance of a momentum portfolio in the 13 to 60 month postholding period is negative.

The Fama and French (1993) three-factor and the Carhart (1997) four-factor models discussed in chapter 4 examine portfolio performance by controlling for additional risk factors due to size, value versus growth and momentum strategies on the part of the fund. Wermers (2000) provides an alternative approach to decomposing returns attributable to investment style versus stock picking ability. Looking at the 1975–1994 period, Wermers' methodology, which in turn is based on Daniel et al (1997), is to first construct characteristic-based benchmark portfolios based on size, book-to-market equity and momentum characteristics of the underlying stock holdings of the funds. For all stocks listed on the NYSE, AMEX or NASDEQ indices for which data on the characteristics are available, Wermers ranks all stocks into quintiles according to market capitalization. Each quintile is then further subdivided into quintiles sorted by book-to-market equity. Finally, within each of these 25 fractile portfolios, stocks are sorted into quintiles ranked by past 12 month returns. This results in 125 benchmark portfolios with distinct size, book-to-market and momentum attributes. Portfolios are reconstructed recursively each year and value weighted quarterly returns are computed for each of the 125 portfolios. The data set used by Wermers (2000) includes data on the stock holdings of the individual mutual funds under consideration. This enables Wermers to calculate an investment style adjusted selectivity performance measure ("Characteristic Selectivity" (CS)) for each fund as

(2.2)
$$CS_t = \sum_{i=1}^{N} W_{j,t-1}(R_{j,t} - R_t^{bj})$$

where $W_{j,t-1}$ is the portfolio weight on stock *j* at the end of quarter *t-1*, $R_{j,t}$ is the return on stock *j* in quarter *t* and R^{bj}_t is the return on the benchmark portfolio in quarter *t* to which stock *j* is matched.

Wermers (2000) reports that over the universe of funds the cross sectional weighted average, weighted by funds' total net assets, CS performance measure (annualized) is 0.71% and is significantly different from zero at the 95% confidence level. Wermers also finds that over the same period the average mutual fund outperforms a broad market index (the CRSP value-weighted index) by 1.3% per annum in terms of raw returns. Therefore, about 60 basis points of this outperformance is attributable to the style characteristics of the stocks held by the funds while 70 basis points is due to the stock picking skills of the managers in excess of the fund's chosen long term investment style.

Wermers (2000) also examines performance with returns net of the transaction costs incurred by the fund and expenses charged by the fund. The paper applies the Carhart (1997) four-factor model to these net returns and reports an average alpha measure across funds of −1.16% per annum which is significant at the 99% confidence level. To the extent that the two procedures are comparable, the results show that while return anomalies may have existed at the level of the fund manager these were not exploitable by investors. This comparison is complicated, however, by the fact that the CS measure is based exclusively on the funds' stock holdings while the Carhart's alpha measure includes the effect on fund returns of a certain proportion of nonstock holdings within funds, ie primarily cash and bonds.

One of the earliest studies of persistence among US equity mutual funds is that of Grinblatt and Titman (1992a). This study introduces a simple persistence testing methodology and investigates whether its finding of persistence arises from a momentum effect in stock returns or from persistent stock picking ability on the part of the fund. Grinblatt and Titman (1992a) split their 10 year sample period (1974 To 1984) into two 5 year sub-periods, estimate α_i (the intercept in a regression against eight factor portfolios, see Grinblatt and Titman, 1989a) for each fund in each sub-period and performe a cross-section regression of abnormal performance from the second period on those in the first period. A positive and significant t-statistic on the slope coefficient would reject the null hypothesis of no persistence. They also check for possible bias by repeating the above with a control sample of 109 funds which they construct themselves to exhibit a variety of characteristics based on firm size, dividend yield, past returns, beta, interest rate sensitivity and co-skewness with the CRSP market index. These are passive funds rebalanced monthly, i.e. they are constructed 'mechanically' to exhibit the above characteristics but do not involve stock picking. The authors find significant evidence of predictability in alphas for actual funds, while the control sample fails to show persistence.

The sensitivity of performance and persistence findings to the choice of benchmark portfolio for market risk is taken up by Hendricks, Patel and Zeckhauser (1993) in their persistence study. This is important because benchmarks during the 1970s have been shown to be mean variance inefficient with respect to portfolios based on firm size and dividend yield (see De Bondt and Thaler 1989 and Fama 1991). Hendricks et al (1993) provide yet another procedure for testing persistence based on Fama and McBeth (1993) and applied similarly in Jegadeesh (1990). They estimate the equation

$$(2.3) \qquad r_{it} - E_{t-1}(r_{it}) = k_t + \sum_{j=1}^{J} a_{jt} r_{it-j} + u_{it},$$

where r_{it} is the return on fund i in period t, $E_{t-1}(r_{it})$ is the expected return on fund i in period t based on the information set at period t-1. The authors estimate the left hand side of (2.3) as the residual from a market model. Under the null hypothesis of market efficiency and no persistence the a_j should be zero. In order to address the mean variance inefficiency question above Hendricks et al test the robustness of findings by using three separate market indices in the market model estimation. Based on all three, F tests reject the null hypothesis of no positive persistence for approx one year but indicate a reversal to negative persistence thereafter. (Of course there remains a joint-hypothesis problem of regarding whether the model of equilibrium security returns is correct).

Hendricks et al (1993) also test persistence using the recursive portfolio formation procedure as in Carhart (1997) and others. Mutual funds are ranked into octile portfolios based on raw returns over the past 1, 2, 4, and 8 quarters, reconstituted quarterly and evaluated by Jensen's alpha. This is carried out for a number of market portfolio benchmarks in the single factor model. Overall, there is persistence in the relative rankings of the octile ranked mutual funds but the Jensen's alpha for the top ranked funds is statistically insignificant. The evidence point to stronger persistence among poorer performing funds.

Malkiel (1995) examines performance and persistence in a longer sample period from 1971 – 1991 using a contingency table approach ranking returns into winners and losers by the median ranking each quarter, Malkiel identifies significant patterns of persistence during the 1970s but not during the 1980s. This is the case for both raw returns and risk adjusted returns. The author finds that an investment strategy based on persistence would have yielded considerable excess returns relative to the S&P 500 index and relative to the mutual fund industry as a whole during the period 1972 – 1981. However, even ignoring load fees, persistence strategies would not have provided an economically significant abnormal return during the 1980s. The paper highlights the importance of examining sufficiently long sample periods to avoid identifying short term anomalies which do not represent long run exploitable trading strategies.

Elton, Gruber and Blake (1996b) further examine persistence among a survivorship bias free sample of equity mutual funds between 1977 and 1993 by applying a recursive portfolio formation procedure and a four-factor model. Ranking funds into deciles by three-year and one-year ranking periods and evaluating decile performance also over three-year and one-year evaluation periods, the authors find evidence of persistence. They find that portfolios based on optimal weights have a significantly higher return than returns provided by equal weighted portfolios.

In a further assessment of persistence Blake and Morey (2000) study the Morningstar five-star rating service as a predictor of US domestic equity mutual fund performance (where a five-star rating is the best and a one star rating is the worst). As an indicator of the importance of the Morningstar rating service the authors cite that at one time (January – August 1995) 97% of all capital flowing into no load equity funds were invested in funds with a four or five star rating by Morningstar while funds with a less than 3 star rating suffered net outflows over the same period. Blake and Morey report that low ratings by Morningstar generally predict poor future performance but there is little evidence that high to medium ranked funds perform well in the future. Once again this is consistent with previous findings that poor performance persists while good performance does not. These findings are robust to fund age and fund investment styles. Blake and Morey also report that the Morningstar ratings perform, at best, only marginally better than the alternative performance measures as predictors of future performance.

Carhart (1997) and Chen, Jegadeesh and Wermers (2000) as discussed above address the question of whether mutual funds actively pursue momentum strategies or whether they "accidentally" already hold and retain previous period winning stocks in the fund. A closely related question is whether apparent persistence is being driven by 'pure' selectivity skills or whether it is due, at least in part, to 'timing selectivity'. This question in the literature is now discussed.

2.3.3 The Source of Persistence

Grinblat and Titman (1992a) also investigate whether the source of the persistence found in their study lies in a momentum effect in the fund's stock holdings or whether persistent stock picking ability exists. This is done by randomly sorting the 120 monthly observations in their sample period into two 60 observation sub-periods and repeating their persistence testing methodology as described above (see section 2.3.2). This procedure tests whether a fund's performance is relatively consistent at each point in time or whether the persistence finding is capturing momentum in consecutive sub-periods. The findings of persistence from among the randomly ordered time series of return observations are even more significant than from among the chronologically ordered fund returns. This supports 'genuine' persistence.

Hendricks et al (1993) also pursue this question. As described section 2.3.2, every quarter the authors rank mutual fund quarterly returns into octiles and hold for one quarter. They then perform a bootstrap simulation procedure where every quarter the cross-section of fund returns are assigned to octiles randomly but with probability equal to the relative frequency at which the fund ranks in that octile over the entire sample period. (i.e. the frequency with which a particular fund ranks in a particular octile in the actual sample of mutual fund returns equals the frequency with which it ranks in that octile in the simulated data but in any given quarter the fund's assignment to a particular octile varies randomly between simulations). By randomizing the time ordering of fund rankings the bootstrap simulation enables the derivation of the distribution of performance statistics such as the octile alphas from the recursive portfolio formation procedure having controlled for timing. If timing or momentum is an important feature driving persistence then the octile alphas from the actual mutual fund data will lie in the tails of the bootstrap simulated distributions. Hendricks et al record a p value of 0.04 for the 'actual' top octile. This indicates that there is only a 4% chance of observing the level of persistence observed in the top octile if timing or momentum was not a factor in driving persistence. In contrast the significant and persistent underperformance of the bottom octile portfolio appears to be driven by relatively consistent poor stock picking ability through time.

Notwithstanding different methodological approaches, Grinblatt and Titman (1992a) and Hendricks et al (1993) arrive at broadly similar conclusions in citing persistently poor stock picking ability, as opposed to a negative momentum or timing effect, as the main determinant of persistently poor performance.

In a related study of the mutual fund industry between 1976 – 1988 Goetzmann and Ibbotson (1994) also question whether a fund's monthly rank is related to last month's rank or just overall long term rank (perhaps due to risk not adequately captured by the performance model). The authors bootstrap the joint distribution of fund returns where they preserve the cross sectional relationship each month but destroy the time series relationship. For both the actual data and each of the 100 bootstrap simulations the study performs a pooled cross sectional and time series regression of monthly rankings on their one period lagged values. By comparing the regression results of the actual return data against the bootstrap simulated regression results the study concludes that the preceding period's rank has power to predict next month's rank in excess of the effects caused by long term mean rankings.

Before proceeding further it is worth noting an important caveat reported by Kothari and Warner (2001). The numerous performance measures described here including Jensen's alpha, Carhart's alpha and the characteristic selectivity measure applied in Wermers (2000) are used extensively in the literature. However, Kothari and Warner (2001) question the power of these models to detect abnormal performance, particularly performance of large magnitude. The study simulates 348 portfolios of funds which are designed to mimic the characteristics of actual US equity funds. (one portfolio is constructed each month between January 1966 and December 1994). For example, among actual (real) funds, the authors note that the median fund based on the median market capitalization of funds' equity holdings is tilted towards large stocks. Consequently in simulating the 348 funds, stocks are selected into each fund with probability equal to the weight of the stock in the NYSE-AMEX stock index. However, the simulation procedure is repeated for a number of other characteristics including book to market ratio characteristics. Although simulated funds mimic actual equity mutual funds, they are nevertheless random stratified samples of NYSE-AMEX securities. Therefore, a well specified performance model using the NYSE-AMEX index as a benchmark market portfolio should not systematically indicate abnormal performance among the 348 simulated funds. The authors also repeat the simulation procedure where in each set of 348 simulations they supplement returns in order to generate various levels of abnormal performance by design. With these simulated portfolios Kothari and Warner (2001) test whether the performance models typically applied in the literature can in fact detect the abnormal performance in the simulated portfolios. In fact both the regression based performance measure of Jensen (1968) and Carhart (1997) along with the characteristic style measure applied by Wermers (2000) have low power, although the Wermers method is marginally better than the others. Kothari and Warner (2001) recommend an event study type evaluation of funds' actual performance using the techniques applied in Chen, Jegadeesh and Wermers (2000) and Wermers (2000) where portfolio holdings data for each fund permit an evaluation not only of overall return but also of the fund's buy and sell trades.

However, the Kothari and Warner (2001) study does not evaluate the ability of conditional models to describe performance and this discussion of the literature now turns to this class of models.

2.3.4 Performance Persistence under Conditional Models of Performance

The discussion of performance persistence among US equity mutual funds presented so far is based on studies which apply unconditional models of performance measurement. Fletcher (1999) and Christopherson, Ferson and Glassman (1998) re-examine the persistence question employing both unconditional and conditional (or time varying) measures of risk and abnormal performance. Both studies apply the conditional models of Ferson and Schadt (1996) discussed in chapter 4. Although Christopherson et al examine persistence among pension funds it is one of few conditional persistence studies and so is mentioned here. Furthermore, Christopherson et al introduce yet a further persistence testing methodology.

Fletcher (1999) evaluates the performance of UK based unit trusts investing in US equity during the period 1985 - 1996. Fletcher measures performance based on an unconditional market model (CAPM), a conditional market model and conditional multi-factor models where the excess returns of trusts are regressed on the excess returns of the factor portfolios and on the product of each factor portfolio and each public information variable. (The public information variables used by Fletcher are (i) 3 month US Treasury Bill return, (ii) dividend yield on the S&P 500, (iii) US Treasury yield spread and (iv) January dummy variable). Fletcher examines persistence using the recursive portfolio formation methodology similar to Hendricks et al (1993) above using quartile rankings. Overall, Fletcher (1997) finds little evidence in support of time varying abnormal performance. 'Winner' portfolios earn slightly more than 'loser' portfolios in the holding period based on unconditional performance measurements but this is reversed for conditional measures. In both cases, the differences in performance between winner and loser funds is not found to be statistically significant.

The Christopherson, Ferson and Glassman (1998) approach to measuring persistence is based on cross sectional regressions of future fund excess returns on a past performance measure of abnormal returns. ie the authors apply regressions of the form

$$r_i(t,t+\tau) = \gamma_{0,t,\tau} + \gamma_{1,t,\tau}(\alpha_{it}) + u_i(t,t+\tau), \quad i = 1, \ldots n \tag{2.4}$$

where $r_i(t,t+\tau)$ is the (compounded) excess return from period t to $t+\tau$ for manager i. τ denotes the return horizon and is examined for values τ = 1, 3, 6, 12, 24 and 36. The regressor, α_{it}, is a measure of return estimated up to month t. $u_i(t,t+\tau)$ is the regression disturbance term. The authors apply a number of measures of past return including conditional and unconditional alpha measures. The hypothesis that the past value of α_{it} cannot be used to predict future return (ie H_0: no persistence) implies that the expected value of the coefficient $\gamma_{1,t,\tau}$ is zero.

Christopherson et al outline a number of caveats including that the regressions do not account for differences in the risk of future returns and consequently if the alphas are related to risk due to a misspecification in the model then evidence of persistence may reflect persistence in the expected compensation for this risk. This aside, the paper finds stronger evidence of persistence when conditional performance models are used to estimate alpha in (2.4) than when unconditional methods are applied. This finding is robust with respect to investment horizons. Finally, once again predicting future performance is more reliable among negative alpha funds, ie poor performance tends to persist to a greater extent than good performance.

On a related issue Mamaysky, Spiegel and Zhang (2005) question the use of a single model of security returns applied to measure abnormal performance across a wide cross-section of mutual funds, i.e. a 'one model fits all' approach, as is the case in almost all studies of mutual fund performance and persistence. The authors argue that one model cannot capture the risks arising from the wide range of trading strategies pursued by mutual funds. Therefore sorting on estimated alphas may mean sorting on misspecification error, i.e. extreme performers may be those with extreme estimation error. Mamaysky, Spiegel and Zhang apply a variety of different models. They implement a recursive procedure where for a given model a fund is dropped if the model fails a filter rule. Filter rules are that the estimated alphas and betas must lie within specified ranges and the model must correctly predict the sign of a fund's excess return in the previous period. The authors argue that if one wishes to use a model to select winning funds in the future it should demonstrate such success in the past. The authors demonstrate in the case of the Carhart four-factor model that following this trading strategy it was possible to earn abnormal returns of between 3.5% and 7.0% per annum. This study indicates that the failure to find supporting evidence of persistence in the extant literature may be related to estimation error in the model.

Once again, drawing conclusions from the body of US studies regarding performance and persistence is complicated by differing methodologies and sample periods etc. Nevertheless, overall there are some results which show managers outperforming benchmark portfolios in risk adjusted terms with some short term persistence. Consistent with findings from studies of UK mutual funds is the evidence that poor performance has a stronger tendency to persist than good performance. Table 2.2 provides a summary of the main findings from US studies.

Table 2.2: Summary of Studies of US Equity Mutual Fund Performance

Study	Period	Type of Funds	Risk Adjusted Returns	Controls Transactions Costs	Controls for Survivorship Bias	Performance Persistence
Carhart (1997)	1963-1993	All US equity Mutual funds	Yes	Yes	Yes	Yes, among poor funds. No, among top funds
Chen et al (2000)	1975 - 1995	US equity funds	Yes	Uses gross returns	Yes	Yes, due to momentum. No, when risk adjusted.
Christopherson et al (1998)	1979 - 1990	273 US equity pension funds	Yes	Net of trading costs	No	Yes, concentrated among poor performers
Elton et al (1996)	1977 - 1993	188 US equity mutual funds	Yes	Yes	Yes	Yes
Fletcher (1999)	1985 - 1996	85 UK American unit trusts	Yes	net of management fee	Yes	No.
Goetzmann and Ibbotson (1994)	1976 - 1988	728 US mutual funds	Yes	No	No	Yes
Grinblatt and Titman (1992a)	1975 - 1984	279 US mutual funds	Yes	Yes	No	Yes, strongest evidence is among poor performers
Hendricks et al (1993)	1975 - 1988	US equity mutual funds	Only in evaluation period.	Net of management fees	Yes	Yes, strongest evidence is among poor performers
Malkiel (1995)	1971 - 1991	All US equity mutual funds	Yes	Yes	Yes	Yes, during the 1970s. Not 1980s.
Volkman and Wohar (1994)	1980 - 1989	332 US mutual funds	Yes	Yes	Yes	Yes

2.4 Testing Portfolio Performance: Skill Versus Luck

Many of the asset pricing models widely applied in the literature to assess abnormal performance rely on the iid-normality assumption for their (asymptotic) validity. However, empirically, the normality assumption is often violated as asset returns exhibit non-normal characteristics, ie skewness and kurtosis, as well as conditional heteroscedasticity and temporal dependence.

Kosowski et al (2004) introduce a new methodology to test portfolio performance and in particular to distinguish skill from luck in performance. In part, the motivation for this technique is based on the authors' findings of non-normally distributed residuals across a range of both unconditional and conditional performance models. Such findings question the inferences of standard t-tests and F-tests.

For a given model, the study first estimates the model over the entire sample period for each fund. It then applies a bootstrap procedure. For each fund, using a random draw from the fund's regression residuals and using the estimated factor loadings, the authors simulate fund returns under the null hypothesis of zero abnormal performance, i.e. set the fund's alpha to zero in the simulation. This is done for every fund in the sample. The authors then re-estimate each fund's alpha using the *simulated* returns. These alphas measure random sampling variation around zero - by construction. Repeating this bootstrap procedure 1,000 times for each fund and within each simulation ranking the cross-section of fund alphas from highest to lowest generates a distribution of luck at each point in the cross-sectional distribution of performance. For example, the highest 1,000 alphas from each simulation represent a distribution which describes the highest level of performance which may be attributable to luck alone – as each alpha has been generated under the null hypothesis. This luck distribution can then be compared to the actual alpha of the best fund to see how this fund performs relative to luck. This can be repeated for the 2nd highest 1,000 bootstrap alphas and compared to the second highest actual alpha and so on across the cross-section distribution to the worst alpha.

This methodology generates an empirical distribution of performance under the null hypothesis but requires no restrictive assumptions regarding the shape of the distribution, in particular normality. This is critical because the deviation from normality is found to be greatest among the top and bottom funds leading to incorrect inferences concerning the funds of greatest interest to investors.

This study applies the bootstrap procedure of Kosowski et al (2004) to the UK mutual fund industry. As such, a fuller discussion of the methodology is deferred until chapter 5.

Rather than apply the bootstrap procedure to all performance models Kosowski et al (2004) first select representative 'best fit' models within the classes of unconditional and conditional beta models, see chapter 4. A Carhart four-factor model is chosen within each class. The results of the bootstrap technique applied to funds across all investment objectives taken together indicates an empirical p value for the top ranked actual fund of 0.02: the probability of observing this level of performance in the top fund due simply to sampling variability in alpha around a true alpha of zero is only 2%. This is indicates that the top fund has genuine stock picking skill. This finding is largely unchanged using the conditional beta model. The paper also presents bootstrap results at several points and percentiles in the left and right tails of the alpha distribution. However, lower ranked funds around the 80^{th} percentile have performance within the boundary of luck. Towards the extreme low end of the distribution, poor performance is worse than that which may be attributed to bad luck.

The breakdown of bootstrap results by investment style strongly indicates stock-picking ability among growth stock funds but not among growth & income funds, balanced funds and income funds.

A further method to distinguish skill from luck is to compare the performance of actual funds with that of simulated or randomly generated portfolios. Burns (2004) uses randomly generated portfolios to demonstrate, *inter alia*, the sensitivity of performance findings to the construction of benchmarks. Using a database of 191 large and small capitalization US stocks between 1996 and 2004, the author uses the set of stocks to generate both 1,000 random portfolios and three alternative benchmarks. (The random portfolios are created subject to some constrains concerning the weights in which individual and groups of stocks are held and no short sales are permitted). The benchmarks are an equally weighted portfolio of all stocks and two randomly weighted (within certain parameters) portfolios of all stocks. Using information ratios, the study then evaluates the performance of the random portfolios against the three benchmarks. In each quarter, the author compares the out-performance of the random portfolios across the three benchmarks to reveal that performance is highly sensitive to even slight differences in benchmark returns. That such differences exist is not surprising: a benchmark will be hard to beat during periods when heavily weighted constituents perform well and *vice-versa.* However, the extent of the differences is high. It is clear that the more unequal the weights in the benchmark portfolio the greater the dispersion in over- and under-performance through time in the simulated portfolios – which of course are generated under the null of no out-performance. Given that Burns (2004) applies weights in the randomly weighted benchmarks which are not extreme relative to the weights in benchmarks commonly used in practice, the study highlights the sensitivity of findings to the selected benchmark and demonstrate an alternative means via random portfolios by which to assess investment skill. (See Dawson and Young 2003 for further discussion on stochastically generated portfolios).

Randomly generated portfolios have also been put forward as a useful alternative to peer group evaluation, i.e. ranking funds of similar investment styles, as a method of assessing relative fund performance. Surz (1998) outlines numerous disadvantages associated with peer group evaluation: (i) a classification bias arises when some funds are pigeonholed into investment style classifications such as growth or value which are too broadly defined and misrepresent the 'true' objective of the fund, (ii) a composition bias arises in fund databases where there are too few funds in certain classifications to implement a reliable peer group comparison and (iii) a survivorship bias arises due to the attrition of funds in some classifications. Surz (1998) advocates the use of random portfolios, or "cyberclone peer groups" as a means of constructing portfolio opportunity distributions (PODs) to evaluate individual fund performance. Given the precise investment style of a particular 'actual' or 'real' fund, multiple (say1,000) random portfolios are generated by selecting from the relevant defined universe of stocks. The performance measure of these random portfolios are used to construct the POD (under the null hypothesis of no out-performance) and the performance of the actual fund is evaluated relative to this distribution. Composition bias is eliminated as there is no limit to the number of 'peer group' random portfolios one can construct to evaluate the performance of any fund, irrespective of its investment style. In addition, survivorship bias is eliminated as performance comparison is made with a simulated peer group rather than a surviorship biased actual peer group.

The use of random portfolios has also been applied in the discussion on the relative importance of investment choices as determinants of fund returns, see Kritzman and Page (2003, 2002). The former evaluates the relative importance of five investment choices: asset allocation, country allocation, global sector allocation, country sector allocation and security selection. The received doctrine is that asset allocation is the chief determinant of performance. Using a bootstrap simulation procedure to construct random portfolios, Kritzman and Page present findings to the contrary. For example, the data set includes a sample that is weighted 60% stocks, 30% bonds and 10% cash. To examine the importance of global asset allocation the authors generate 10,000 portfolios of 100 assets whose composition vary randomly around a mean asset mix of 60%, 30% and 10% stocks, bonds and cash respectively. As a control all other investment choices are fixed across the simulated portfolios and consequently variation in portfolio return arises purely from variation in asset mix. Similar procedures are followed to examine the importance of the other investment choices. The authors report that dispersion around average performance arising from security selection is substantially greater than for all other investment choices while asset allocation produces the least dispersion. The paper also presents evidence that the apparent emergence of global sector allocation over country allocation as the more important investment choice may be an artifact of the 'tech-stock' bubble as there is little evidence for this outside the 1996-2000 period.

Kritzman and Page (2002) is a very similar study and uses the same bootstrap simulation procedure to assess the relative importance of security selection versus asset allocation investment decisions in Austria, Germany, Japan, the UK and the US. Similar to the authors' 2003 study, this study supports the relative importance of security selection over asset allocation, reporting significantly higher dispersion around average performance arising from the former. Broadly similar simulation procedures are outlined in Bridgeland (2001, 2000) to examine performance attribution and value added arising from alternative portfolio construction strategies such as stock size (capitalization), portfolio size etc.

2.5 Market Timing

In addition to stock picking or security selection skill (or 'selectivity'), a further aspect of fund performance is that of market timing. One form of market timing is the tactical asset allocation approach which keeps the composition of a portfolio of risky assets constant but alters the proportion of the portfolio held in cash (non risky assets) according to the expected future direction of the market. Alternatively, market timing may be practiced by adjusting the beta of a portfolio, i.e. the portfolio's sensitivity to the market, in response to the expected market return. Beta is increased (decreased) in response to an expected bull (bear) market. To test tactical asset allocation requires information on the portfolio's composition over time which is often not available. However, tests of whether the portfolio beta is conditional on a benchmark market return may be conducted with widely available *ex-post* fund and market returns.

Two 'classic' conditional beta models of market timing are Treynor and Mazuy (1966) and Henriksson and Merton (1981). In this study both these models are described in more detail and are estimated for UK mutual funds in chapter 4. However, a brief description is provided here.

Treynor and Mazuy (1966) specify a quadratic regression of the form

(2.8) $$r_{i,t+1} = \alpha_i + \theta_i(r_{m,t+1}) + \gamma_{iu}[r_{m,t+1}]^2 + \varepsilon_{i,t+1}$$

where the coefficient γ_{iu} measures market timing ability. $r_{i,t+1}$ and $r_{m,t+1}$ are the fund and market excess returns respectively. Admati et al (1986) demonstrate that the model is consistent with a manager with constant absolute risk aversion who adjusts the portfolio beta at time *t* according to a private linear signal of the form

(2.9) $$\beta_{it} = \theta_i + \gamma_{iu}[r_{m,t+1}] + \eta_t$$

where η_t is random noise. The hypothesis of no abnormal timing performance implies $\gamma_{iu} = 0$.

Henriksson and Merton (1981) propose a model in which the conditional portfolio beta has two target values in a binary response function depending on the manager's forecast of whether market return will exceed the risk free rate. The authors show that if the manager can successfully time the market then the coefficient γ_{iu} in the following regression will be positive.

$$r_{i,t+1} = \alpha_i + \theta_i(r_{m,t+1}) + \gamma_{iu}[r_{m,t+1}]^+ + \varepsilon_{i,t+1} \tag{2.10}$$

where $[r_{mt+1}]^+$ is defined as $\max(0, r_{mt+1})$. Here $\max(0, r_{mt+1})$ may also be interpreted as the payoff to an option on the market portfolio with a strike price equal to the risk free rate.

Ferson and Schadt (1996) further develop the above models by also specifying the portfolio beta to be a function of public information variables thus controlling for timing skills which may be attributable to public information. The test is then a test of the quality of the fund manager's private timing signal. (The Ferson and Schadt (1996) models are discussed and estimated in chapter 4). See also Becker et al (1999) and Ferson and Khang (2001) for further discussion on the effects of conditioning information on timing performance measures. Of course, a fund manager may also vary the fund's exposure to risk factors other than the market or indeed to other benchmark indices according to their year-to-date performance in response to incentives s/he may face (Chevalier and Ellison, 1997; Brown, Harlow and Starks, 1996).

Other issues in the literature on market timing include the question of spurious timing arising from employing an incorrect benchmark (Jagannathan and Korajczyk, 1986; Breen et al, 1986), distinguishing timing from selectivity in abnormal performance (Admati et al, 1986; Grinblatt and Titman, 1989b) and the power of a test when actual fund timing frequencies differ from data sampling frequencies (Goetzmann et al, 2000; Bollen and Busse, 2001). Many of these issues arise because of the parametric model nature of market timing tests such as, or based upon, Treynor and Mazuy (1966) and Henriksson and Merton (1981). These issues are now discussed.

Breen et al (1986) discuss the issue of heteroscedasticty in the Henriksson and Merton test and demonstrate that the portfolio return of a market timer will exhibit conditional heterscedasticity, (see pp587-588). Using simulation techniques Breen et al (1986) demonstrate that ignoring heteroscedasticity falsely rejects the true null hypothesis too often while the probability of failing to reject the null hypothesis when it is not true also increases significantly. The authors suggest that their conclusions will also apply to the Treynor-Mazuy method.

A further parametric or specification issue that can affect the consistency and power of market timing tests is the separation of market timing skills from security selection skills in performance. (Admati et al, 1986; Grinblatt and Titman, 1989b, Coggin et al, 1993). The bulk of the empirical evidence on timing suggests that statistically significant timing is rare and where it exists it is more likely to be negative, though this is not necessarily irrational for investors with increasing absolute risk aversion. (Kon 1983, Chang and Lewellen 1984, Henriksson 1984, Lockwood and Kadiyala 1985). However, many of these studies also find evidence of a negative cross-sectional correlation between the market timing and selectivity measures of performance. Henriksson (1984) suggests that this may be caused by error-in-variables bias, misspecification of the market portfolio or the use of a single rather than multi-factor asset pricing model. However, Jagannathan and Korajczyk (1986) suggest that this empirical finding arises from the nonlinear pay-off structure of options and option-like securities (such as the common stock of highly levered firms) in portfolios. The authors demonstrate that a portfolio strategy of buying call options on the market portfolio will exhibit positive timing performance and negative security selection performance even where no market or stock picking is being undertaken. Clearly, a long position in call options on the market will provide a high pay-off in a rising market thus exhibiting apparent market timing but in a relatively steady or falling market the reduction in return caused by the option premium appears as negative security selection. A similar rationale applies to holding the common stock of highly levered firms where there is little or no return to equity holders unless there is a high stock return associated with a rising market. As such, funds which hold highly levered stocks, such as perhaps small firm stocks, may show positive timing performance.

Further evidence of specification biases are provided by Kothari and Warner (2001). The authors simulate portfolios by randomly picking stocks, controlling for size and value effects, and periodically adjust the portfolio composition to mimic typical fund turnover. While the portfolios are generated under the null hypothesis of neither market timing nor security selection skills, standard performance measures nevertheless detect abnormal performance. Kothari and Warner conclude the performance measures (including market timing measures) are misspecified.

As noted by Bollen and Busse (2001), spurious poor or even negative market timing may also arise from the cash-flow hypothesis described by Edelen (1999), Warther (1995) and Ferson and Warther (1996). This hypothesis suggests that investors increase subscriptions to mutual funds during periods when the market return in relatively high which results in a high cash position for funds causing a lower portfolio return. For the

Henriksson-Merton (1981) model the market timing coefficient is estimated only when the market (excess) return is positive and so the cash-flow hypothesis is asymmetric: it can bias the coefficient downwards but not upwards. Bollen and Busse (2001) also argue the timing coefficient in the Treynor-Mazuy method is similarly biased downward.

Bollen and Busse (2001) and Goetzmann et al (2000) address the bias that may arise in tests of market timing when the frequency of the researcher's observed return data differs from the frequency of the manager's timing strategy (where the latter may not itself be of a uniform frequency). For example, the standard parametric tests above, typically applied to monthly fund return data, may underestimate the timing skills of daily or weekly market timers. Using a bootstrap simulation technique Bollen and Busse (2001) generate synthetic fund returns which mimic the holdings of actual funds using both daily and monthly data. Treynor-Mazuy and Henriksson-Merton models augmented by multiple risk factors are used. Under the null hypothesis of no market timing in the simulated returns the authors demonstrate that the size of the daily tests are correct in generating the expected number of positive, negative, significant and insignificant timing coefficients. The tests on monthly data, however, are biased. Simulating fund returns under the alternative hypothesis the authors demonstrate a significant increase in the power of the test to detect daily relative to monthly timing. Furthermore, using bootstrap standard errors Bollen and Busse provide evidence of greater market timing ability among actual fund managers, using daily return data, than was previously thought to exist using monthly frequency. Goetzmann et al (2000) similarly demonstrate that the Henriksson-Merton method is weak and biased downwards when applied to the monthly returns of daily timers, and proposes an adjustment without requiring daily data.

A final specification issue which may complicate the assessment of market timing ability arises if/when fund managers attempt to time market volatilities as well as market returns. Busse (1999) shows that funds attempt to reduce market exposure when market volatility is high, although Laplante (2001) questions whether funds are successful in this attempt. If a manager tries to time market volatility s/he may reduce the fund's market exposure even when expected return is high. Thus if market volatility and market return are positively correlated, the market timing measure may under-estimate the manager's market timing ability, ie the quality of his/her timing information. If market volatility and return are uncorrelated then the timing measure remains consistent in the presence of volatility timing. Breen et al 1989, Glosten et al 1993 and Busse 1999 provide evidence that the relation between market return and volatility is weak so this effect should not seriously distort the results of market timing tests.

Jiang (2003) proposes a nonparametric test of market timing in order to address some of the specification issues described above that arise in the parametric approaches. The methodology of Jiang (2003) is described in detail and is implemented empirically for the UK mutual fund industry in chapter 9 and chapter 10 in this study. As such the discussion here is limited to a brief description. For any triplet of market return observations $\{r_{m,t1}, r_{m,t2}, r_{m,t3}\}$ sampled from *any* three time periods (not necessarily in consecutive chronological order) with $r_{m,t1} < r_{m,t2} < r_{m,t3}$ the approach suggests that an informed timer will maintain a higher exposure to the market in the $[r_{m,t2}, r_{m,t3}]$ range than in the $[r_{m,t1}, r_{m,t2}]$ range. For a fund *i* with non-increasing absolute risk aversion and independent timing and selectivity information this implies that

(2.11)
$$\frac{ri,t_3 - ri,t_2}{rm,t_3 - rm,t_2} > \frac{ri,t_2 - ri,t_1}{rm,t_2 - rm,t_1}$$

Assigning a sign function that assumes a value of 1(-1) if the argument is positive (negative) and equals zero if the argument equals zero, the average sign across all triplets taken from *n* observations is given by

(2.12)
$$\hat{\theta}_n = \begin{bmatrix} n \\ 3 \end{bmatrix}^{-1} \sum_{rm,t_1<rm,t_2<rm,t_3} sign\left(\frac{ri,t_3 - ri,t_2}{rm,t_3 - rm,t_2} > \frac{ri,t_2 - ri,t_1}{rm,t_2 - rm,t_1}\right)$$

Jiang (2003) reports the asymptotic distribution of θ_n enabling the statistical significance of a fund's market timing performance measure to be determined.

The Jiang (2003) methodology has several advantages over the specifications of Treynor-Mazuy and Henriksson-Merton and is more robust to some of the problems described above. First, the procedure requires the fund's market exposure simply to be a non-decreasing function of the expected market return which in turn requires non-increasing absolute risk aversion. However, this is less restrictive than the linear and binary response functions required by Treynor-Mazuy and Henriksson-Merton respectively. Second, both the Treynor-Mazuy and Henriksson-Mertson methods incorporate two aspects of timing: (i) the quality of the manager's timing information and (ii) the aggressiveness of the response. The nonparametric method, however, measures how often a manager correctly forecasts a market movement and acts on it but is independent of the aggressiveness of the response. Third, Jiang (2003) reports that simulation results indicate that the nonparametric measure is more robust to the difference between actual manager timing frequency and observed sample data frequency. This is because, unlike the standard regression based approaches, it is less reliant on the 'correct' timing frequency as it examines timing between all pairwise triplets of fund and market returns and not just monthly consecutive pairs. Finally, the asymptotic distribution of the nonparametric timing measure is unaffected by heteroscedasticity in fund returns. However, the nonparametric measure, like the standard parametric tests, is less robust when managers are volatility timing as well as market timing and similarly cannot distinguish between market timing and option related spurious timing.

The empirical findings described here relate to studies of the US mutual fund industry. There has been relatively little research carried out on the market timing skills of UK equity unit and investment trusts. Fletcher (1995) applies both the Chen and Stockum (1986) and Henriksson and Merton (1981) models to test for market timing performance. (The Chen and Stockum (1986) approach is similar to that of Treynor and Mazuy (1966) but embodies a conditional coefficient on the market variable). Evaluating 101 unit trusts between 1980 and 1989, Fletcher reports the cross sectional average timing measures to be negative and strongly significant. This is found to be the case for both models of market timing and alternative market benchmark indices. These results suggest that funds on average reduced their market exposures when market returns were high and *vice-versa*. Leger (1997) also examines the issue of market timing among UK equity investment trusts between 1974 and 1993. Leger produces broadly similar results to Fletcher (1995) where he finds negative and statistically significant timing measures. As indicated previously, negative market timing may be consistent with (i) the cashflow hypothesis, (ii) volatility timing and/or (iii) increasing absolute risk aversion.

2.6 Fund Performance and Other Fund Characteristics

Another issue examined in the literature is the claim by fund managers that expenses, turnover and load charges do not reduce performance since they reflect the quality of the manager's information, which is collated at a cost, and since managers trade only in anticipation of earning returns net of transaction costs. There is also a question regarding whether fund performance is related to the size of the find, as measured by its total net assets.

Transaction costs are trading costs incurred by the fund and typically include spread costs, brokerage fees and in some studies include a market impact effect on the stock price arising from the trade, Wermers (2000). Annual charges imposed by the fund on its investors, expressed as a percentage of the net asset value of the investment and hence often called the expense ratio, is to cover administration fees, audit fees and other annual expenses but does not typically include the load (initial) charge.

Ippolito (1989) finds evidence that mutual funds earn returns sufficient to cover expenses during the period 1965 - 1984. These results are in contrast with many earlier studies (Friend, Blume and Crockett 1970; Jensen 1968; Sharpe 1966). However, Elton, Gruber, Das and Hlavka (1993) suggest that Ippolito's findings may result from an insufficient benchmark indices used to estimate risk adjusted return, ie the S&P 500. They show that correcting for non-S&P 500 stocks in the benchmark market index positive alphas disappear. Malkiel (1995) finds similarly: a small number of mutual fund managers outperform a broad market index at the gross return level but not net of costs. Volkman and Wohar (1995) confirm that high management fee funds do not demonstrate abnormal performance but their study does find positive persistence in some low management fee funds. Droms and Walker (1994) examine international equity mutual funds and find no link between returns and expenses and also find no higher return reward to investors arising from paying a load fee.

To examine the link between performance and size, charges and turnover, Carhart (1997) estimates monthly abnormal performance by the four-factor model where the model loadings are estimated over the prior three years, i.e.

(2.5) $$\alpha_{it} = R_{it} - R_{ft} - b_{1i,t-1}(R_{mt}-R_f) - b_{2i,t-1}(SMB_t) - b_{3i,t-1}(HML_t) - b_{4i,t-1}(PR1YR_t)$$

and each month estimates the cross-section regression

$$\alpha_{it} = \delta_1 + \delta_2(X_{it}) + \varepsilon_{it} \tag{2.6}$$

where X_{it} are fund characteristics: size (in logs), expenses, turnover and load fees for that month. As in Fama and McBeth (1993), each month Carhart (1997) estimates the cross sectional relation in (2.6) and then averages the coefficients across the time series sample period. His results indicate a statistically significant negative relation between performance and all characteristics except size. Of particular interest are the coefficients on the expense ratio and turnover: for every 100 basis point increase in expense ratios and turnover, abnormal return (annualized) drops by 1.54% and 0.95% respectively. This suggests that on average mutual funds do not recoup their investment costs through higher abnormal returns. Carhart also notes that the negative coefficient on the load fees variable contradicts the claim by load funds that the load fee is a charge for skill. These additional fund characteristics appear to further explain abnormal performance, and the persistence therein among poor performers.

Wermers (2000) also examines the link between stock turnover in a fund and performance. The study applies a recursive portfolio formation methodology described previously for Carhart (1997) and others. However, Wermers recursively forms portfolios of funds by ranking funds into deciles (repeated for quintiles) by their levels of turnover during the previous year and holding these portfolios for one year. This enables an analysis of whether top decile turnover funds are persistently top performers.

Examining gross returns, ie gross of transaction costs incurred and expenses imposed by the fund, Wermers reveals that the top turnover decile of funds on average outperforms the bottom decile by 4.3% per annum between 1975 – 1993. This is significant at the 10% level. Wermers also investigates the attribution of this difference and reports that the difference is attributable, in descending order of importance, to funds' investment styles, stock selection ability and market timing ability. In terms of net returns, the difference between top and bottom turnover deciles falls to a statistically insignificant 2.1% while the difference between top and bottom turnover quintiles is 2.7% (significant at the 5% level). In terms of risk adjusted net returns (Carhart's alpha) there is no difference between the performance of high and low turnover funds.

Chalmers, Edelen and Kadlec (1999) is a further comprehensive examination of the costs/performance question. This study evaluates the relationship between trading costs, expense ratio and turnover on the one hand and a number of fund performance measures (raw returns net of expenses and a CAPM and four-factor adjusted return) on the other during the period 1984-1991. Each quarter the authors rank funds by total fund costs into quintiles and calculate the average performance of funds in each quintile. Figures are then averaged over the sample period. This procedure is repeated ranking by (i) trading costs, (ii) expense ratio, and (iii) turnover. In the case of all three costs ranking criteria, there is a strong negative (and statistically significant) relationship between the costs and the performance measures tested. The relationship between turnover and performance is also found to be negative, though not significant. The findings in Chalmers et al (1999) are broadly in line with those of Carhart (1997).

The issue of the relative performance between load and no-load funds is examined in Ippolito and Richard (1989), Elton et al (1993), Grinblatt and Titman (1994), Droms and Walker (1994), Gruber (1996) and Fortin and Michelson (1995). With the exception of Ippolito and Richard (1989), who find that load funds earn rates of return that plausibly offset the load charge, other studies find no significant difference between the performance of load and no-load funds. However, Morey (2003) is perhaps the most comprehensive paper to address this question as Morey is the only paper in the area to examine the load adjusted performances of load and no-load funds. In addition, Morey (2003) examines relative performance within load funds between relatively high load and low load funds. The study evaluates these relative performances using raw returns as well as single and four-factor model alphas. Morey finds that after adjusting for loads in the return data, no-load funds outperform load funds for almost all performance measures examined while within load funds themselves there is little significant difference in performance between high load funds and low load funds.

Elton, Gruber and Blake (1996b) also examine the performance/expenses question. As indicated previously, the fund's fee is almost always stated as a fraction of the fund's total net assets (unitised) and is referred to as the expense ratio. Therefore, managers can increase revenue by increasing either the percentage fee or the total net assets of the fund. The difficulty with the former is that higher expense ratios reduce the post-expense performance possibly leading to some capital outflow and a reduction in total net assets. The authors find a strong negative cross sectional relationship between fund size (total net assets) and expense ratios, although expense ratios are found to rise over the time period in question (1997 - 1991). In particular, Blake et al (1996) examine whether top performing funds (top decile funds) increase their revenues by increasing the fee or by aiming to increase the total net assets of the fund. The study examines fees for up to five years after a fund's top decile ranking and demonstrates that top performing funds increase their fees by no more than the average fund.

Until recently lack of available data has made it difficult to evaluate these questions in the UK case. However, Fletcher and Forbes (2002) examine whether fund attributes such as annual charges, load charges, size and percentage cash flow (percentage change in total net assets) are related to fund performance (net raw returns). Ranking funds into quartiles annually based on past year raw returns, the authors calculate the cross-sectional and then time series average of the characteristics in each quartile. They report little cross-sectional variation in these characteristics and hence they do not explain persistence in performance.

2.7 Mutual Fund Investment Flows and Performance

A further area of investigation in the mutual fund performance literature is that of the relation between fund performance and the capital investment flows in and out of the fund. The direction of causation between fund flows and performance is a matter of debate. Many studies test for a positive relationship in which performance influences subsequent flows, Gruber 1996, Zheng (1999) and Lynch and Musto (2002). However, Edelen (1999) and Berk and Green (2004) examine the reverse causation.

Gruber (1996) defines cash flows as the change in the fund's total net assets minus the appreciation in the fund's assets held at the beginning of the period. The study examines the fund flows in period *t* of deciles of funds ranked by performance in period *t-1*. This is performed recursively for various values of *t*. The study reports a strong and significant correlation between performance and subsequent capital flows. Interestingly, this procedure also provides insight into which performance measures investors use to evaluate funds. While many performance measures prove significant in causing capital inflows/outflows, perhaps surprisingly Gruber (1996) determines that a four factor alpha measure proves particularly significant and robust. This would seem to indicate that investors are quite sophisticated in accounting for market, size, value and momentum attributes in choosing where to invest capital. Equally, however, it may simply reflect that fund rating companies rank funds by this or similar criteria and this is the cause of investors selecting such high alpha funds. Del Guercio and Tkac (2000) also find a significant relation between a fund's Jensen's alpha measure and subsequent fund flow. The authors find this to be the case for both mutual funds and pension funds although in the latter case tracking error is also important in influencing flows. Tracking error is a measure of diversifiable risk and is often measured as the volatility of a portfolio's deviation from a benchmark index.

Having found evidence that good (bad) performance gives rise to subsequent capital inflows (outflows), Gruber (1996) proceeds to evaluate whether investors improve their performance as result of re-directing their capital. Inflows to a fund in quarter *t* are multiplied by the risk adjusted return of the fund in (a number of) subsequent periods. Aggregated over all funds and all time periods and expressed as a percentage of total capital inflows to all funds, the average risk adjusted return on 'new cash' was 29 basis points per annum. A similar procedure applied to fund outflows to measure how much money an investor saves by removing their capital from a fund indicates a saving of 22 basis points per annum.

Lynch and Musto (2002) find in support of a convex relationship between past returns and fund flows, i.e. fund flows are less sensitive to past performance when past performance is low. This is because, intuitively, one might expect less persistence in poor performance rather than high performance because low past returns contain less information about future returns as poor performing funds are more likely to change strategy. Lynch and Musto (2002) find (i) fund strategy changes occur only after poor performance and (ii) poor performers who change strategy enjoy a greater performance improvement relative to poor performers who do not.

In a broadly similar study Zheng (1999) investigates whether "new money flows" predict future returns: does fund flow information represent a smart money signal? Zheng (1999) implements a number of trading strategies based on money flows and tests whether these strategies earn abnormal returns. For example, one investment strategy is to hold a portfolio of funds which exhibit positive fund flow. The portfolio is updated recursively each quarter. Another strategy is to hold funds exhibiting negative fund flow. Another portfolio holds above median cash flow funds while yet another holds below median cash flow funds. In all cases portfolios are reconstituted each quarter. This generates a time series of returns for each portfolio. A single factor and multi-factor performance alpha can then be estimated for these money flow based portfolios. Money flow or cash flow is as defined in Gruber (1996). During the 1970 – 1993 sample period under investigation, Zheng (1999) reports that funds that receive new money significantly outperform those that lose capital. However, the outperformance is relatively short lived. In addition, new money funds are not found to significantly beat the market as a whole.

In the case of the Gruber (1996) and Zheng (1999) studies, the apparent relationship between fund flow and performance may simply be picking up on persistence (if it exists) in fund returns, i.e. positive fund performance attracts capital or 'new money' which in turn earns a high return for investors by benefiting from performance persistence in the fund. Sirri and Tufano (1993) is a further study to look at flows into individual funds. The study broadly finds that money flows into funds with the best past performance but does not flow out of funds with the worst past performance. This is likely to reflect (i) the huge growth in the mutual fund industry generally which means that a large proportion of fund flow is from 'new' capital rather than capital moving between funds and this new capital is attracted to high past return funds and (ii) that there is a cost for existing investors to leave one fund to join another.

Berk and Green (2004) present a model in which persistence in performance is diminished by fund inflow since managers are likely to have decreasing returns to scale in their talents. Edelen (1999) picks up on a similar theme in the fund-flow/performance relationship. Edelen hypothesizes a reverse direction of causation in this relationship to that of Gruber (1996). Specifically, that fund flow has a negative effect on subsequent fund performance. This study distinguishes two functions of a mutual fund: to undertake discretionary trades which will lead to a positive risk adjusted return and to satisfy its investors' equity liquidity demands. Edelen (1999) argues that fund flows, or flow shocks, force the manager to engage in "liquidity motivated trading" which is nondiscretionary. An inward flow shock immediately alters a fund's relative cash/equity holdings and moves the fund from its target portfolio. Sufficiently high fund flow magnitudes would increase the variability of the fund's cash position. First, this complicates the investor's task of making risk/return choices and second it compromises the manager's objective of tracking or beating a benchmark index. Consequently, providing a liquid equity position for the investor triggers marginal trading by the fund. Edelen argues that this liquidity component of the fund's trading plays the role of noise trading and since noise traders face expected losses the fund should experience negative return performance in proportion to the volume of fund flow.

Edelen (1999) partitions a mutual fund's abnormal return between return attributable to liquidity motivated trading and information motivated (discretionary) trading. The paper first estimates liquidity motivated trading as $f_{jt} = c^I.f^I_{jt} + c^O.f^O_{jt}$, where f^I_{jt} and f^O_{jt} denote the volumes of inflows and outflows respectively for fund *j* in time *t* and c^I, c^O are the estimated coefficients in separate bivariate regressions of the volume of fund stock purchases on f^I_{jt} and the volume of fund stock sales on f^O_{jt} respectively, i.e. c^I and c^O are flow-trade response coefficients. The author then constructs a regression of the form

$$AR_{jt} = \lambda.f_{jt} + \delta.d_{jt} + \varepsilon_{jt} \tag{2.7}$$

where AR_{jt} is the abnormal return on fund *j* in period *t*, f_{jt} is the estimated liquidity motivated trading and d_{jt} is the estimated information (discretionary) motivated trading estimated as the combined volume of stock purchases and sales minus f_{jt}, E_{jt} is a random disturbance term. In order to incorporate persistence the author supplements the right hand side of (2.7) with lagged values of AR_{jt}. Finally, in an attempt to avoid inference problems arising from the possible reverse causation between AR_{jt} and f_{jt} cited by Gruber (1996), the author instruments f_{jt} by its lagged value.

Overall, the empirical results in Edelen (1999) provide some evidence in support of the author's hypothesis that fund flow negatively impacts on fund performance.

One overall possible conclusion from combining the findings of Gruber (1996), Zheng (1999) and Sirri and Tufano (1993) on the one hand and Edelen (1999) on the other is that high relative performance attracts capital inflow. However, under pressure to provide a liquid equity investment for investors the fund makes, at least some, poor equity trades. With a higher proportion of the fund now in equity this detracts from subsequent performance.

A further avenue of investigation in examining mutual fund capital flows is to evaluate whether fund flows in aggregate affect stock market returns. This question is taken up by Warther (1995) and is briefly addressed here. Warther (1995) divides fund flows into anticipated and unanticipated flows, using Box-Jenkins procedures to estimate anticipated flows. Monthly unanticipated fund flows are found to strongly correlate with concurrent stock market returns in a regression of the latter on the former. Anticipated fund flows are uncorrelated with stock market returns which is consistent with informationally efficient markets where anticipated flows are 'pre-contemporaneously' discounted in returns. Warther (1995) also tests the feedback hypothesis by reversing the direction of the regression and hypothesizing that fund flows are, at least in part, determined by lagged stock market returns. The author finds no empirical evidence in support of the feedback trader hypothesis, however. Remolona, Kleiman and Gruenstein (1997) also examine both directions of causality in the aggregate fund flow – aggregate stock market returns relationship in order to determine if in a declining stock market the positive feedback theory could lead to a self-sustaining decline in stock prices. However, their analysis suggests that over the 1986 – 1996 period the effect of short term stock market returns on mutual fund flows were weak

Once again, the difficulty in obtaining data in the UK means that the relationship between fund performance and fund flow is comparatively unexplored. However, Fletcher and Forbes (2002) do examine whether the degree of fund flow varies across quartile ranked portfolios of funds and as such whether it is linked to persistence in performance. The authors find that the winning quartile of funds experience the highest degree of cash flows while the worst quartile experience the least cash flow, suggesting little penalty for the relatively poor performance.

Due to paucity of data, Keswani and Stolin (2005) is the only UK study to link new cash inflows and outflows to future performance (as measured by the four-factor alpha over the period 1992-2000, using around 500 funds. With monthly portfolio rebalancing they find that new money earns a higher abnormal return than old money but in each case the abnormal return is negative.

2.8 Mutual Funds versus Mutual Fund Managers

It is also important to note that the findings relating to performance throughout this review of the literature relate to the mutual fund as the entity rather than specifically to the fund manager, i.e. the fund manager is likely to change over the return history of the fund. As such it is also useful to examine the relationship between fund performance and the cross-sectional characteristics of fund managers. This issue is pursued by Chevalier and Ellison (1999) who evaluate whether mutual fund performance is related to fund manager 'skill' as measured by the manager's age, the average SAT score of the manager's undergraduate institution, whether or not the manager held an MBA and the manager's tenure (how long he has been in the position). The performance evaluation is carried out both before and after controlling for other fund risk characteristics. Using a sample of 492 mutual fund managers who had sole responsibility for a fund for at least some part of the 1988 – 1994 period and raw returns, they report that managers with MBAs outperform managers without an MBA by 63 basis points per year. However, controlling for systematic risk (using a single factor alpha as the performance measure), the differential in performance is zero, i.e. managers with an MBA tend to hold more market risk. The authors also find a small risk adjusted performance differential among younger versus older fund managers and suggest this is due to a stronger work ethic among young mangers, who are still establishing their careers and who face a higher probability of dismissal. Finally, the most robust performance differential identified is that managers from universities with higher undergraduate SAT scores obtain higher (risk adjusted) returns. While some of this return is attributable to expense characteristics, the SAT score remains highly significant. They attribute this outperformance to better natural ability, education and professional networks associated with having attended a higher SAT score undergraduate institution.

Unfortunately data on UK equity mutual fund characteristics such as expenses, load charges, turnover and fund size are not available historically in sufficient number of observations to carry out reliable tests of the relationships between performance and such characteristics. Data on Total Expense Ratios by fund are available from *Fitzrovia* but not before 1997 and no fund provides an expense figure more than twice per annum. Data on fund size are available from *Micropal* but again not before 1997. Similarly, data on the characteristics of the individual fund managers operating UK equity mutual funds are also unavailable thus ruling out the type of analysis undertaken by Chevalier and Ellison (1999) above.

2.9 European Studies of Mutual Fund Performance

In order to provide a picture of some of the results to emerge from studies of the European mutual fund industry a brief discussion of a study by Otten and Bams (2000) is provided here. This is one of very few comprehensive studies of the European mutual fund industry and is selected here because of its methodological similarities to many UK studies.

Otten and Bams (2000) focus on the mutual fund industries of France, Germany, Italy and the Netherlands and the UK. In the case of each country only mutual funds which invest exclusively in their domestic equity market are examined. Performance is evaluated on local currency returns over the period 1991–1998 and is limited to funds with two or more years of data. Otten and Bams (2000) report a large volume of results, estimating separately for each country both a single and four-factor model applied to funds of alternative investment objectives. A summary of results is provided here.

Mutual funds in all regions are found to have a high sensitivity to the returns of small company stocks and value stocks. Only funds in the German domestic market generally underperform but this is not significant at 5%. The bulk of the evidence points to outperformance among the UK and Italian funds (at 5% significance) and outperformance among the French and Dutch funds (at 10% significance). Otten and Bams (2000) also carry out the performance tests on fund returns net of annual charges. In this case, most German funds are now found to produce significantly negative alpha measures. While risk adjusted performance in all other regions is positive, it is only significantly so in the UK and furthermore this is found for the most part among funds investing in UK small company stocks.

In terms of performance persistence, Otten and Bams (2000) employ a recursive portfolio formation procedure. The only significant result is observed for UK funds which exhibit a 7.28% annual spread between the top and bottom portfolios. The evidence of persistence is qualitatively similar to that of Blake and Timmerman (1998) who examined the longer sample period 1972 – 1995.

This concludes the review of the literature. The next chapter describes the UK mutual fund industry generally and in particular the data set employed in this study.

CHAPTER 3

DATA DESCRIPTION

This chapter first provides a brief background description of the UK asset management industry as a whole in terms of size, client structure, asset allocation and current industry trends. The chapter then provides a detailed description of the data employed in this study as well as data definitions and sources. Some simple descriptive statistics of the mutual fund returns, benchmark factor portfolios and public information conditioning variables (see chapter 4) are presented. While some basic performance statistics are reported, this chapter is not intended as a rigorous performance analysis.

3.1 The UK Asset Management Industry

This section describes the aggregate UK asset management industry in terms of size, client structure, asset allocation and current trends. This discussion draws heavily from the findings of the Investment Management Association (IMA) *Asset Management Survey 2004*.

3.1.1 Industry Size

By June 2003 the UK fund management industry managed a total of almost £7 trillion in assets. Of that, almost £2 trillion was under management within the UK, i.e. in UK assets, while a further £5 trillion was managed globally. Industry concentration is high but has fallen slightly in recent years. The five largest operators (Legal & General Investment Management, Barclays Global Investors, M&G Investment Management[1], Morley Fund Management and Standard Life Investments), with UK assets ranging from £122bn to £77bn each, manage 28% of the UK market while the ten largest groups account for 46%.

In terms of global assets, investment banks account for the largest proportion of assets under management at 28% followed by insurance companies at 22%, stand alone fund managers at 18% and retail banks at 14% with the remainder comprised of custodians

[1] Fornerly known as Municipal & General Securities

and other diversified financial services. In terms of assets managed in the UK, insurance companies dominate managing 39%, with investment and retail banks at 19% and 18% respectively.

3.1.2 Client Type and Client Mandate

Almost 40% of assets managed in the UK are invested in insurance funds, followed closely by 37% managed in pension funds while the third largest source of clients is the retail fund sector at 15%. The relatively small remainder is made up largely of central banks and private clients. The IMA survey does not filter out unit trusts as a separate category. Many unit trusts are sold to both the retail and institutional sectors, although generally at different unit prices.

The majority of assets managed in the UK (54%) are actively managed against a customized benchmark. Funds holding these assets are mandated to perform relative to a blend of (or single) indexes. 18% of all assets managed in the UK are passively managed, ie are mandated to track an index to within 50 basis points or less. Specialist sectors comprise 16% of assets under management, ie specific asset class or geographic exposure. 8% of assets managed in the UK are held by peer group mandated funds, ie are managed in relation to a group of comparable funds. Finally, absolute return mandated funds make up 4% of assets managed in the UK. These are managed according to a target level of return, for example LIBOR + 2%.

Among retail funds and pension funds there is the broadest spread of mandates but all client types utilize customized benchmarking more than any other type of mandate. This emphasizes the importance in this study of establishing whether such funds have genuine ability to beat a benchmark.

3.1.3 Asset Allocation

Of the entire asset value under management in the UK, equities represents 46%, followed by bonds at 37% and money market investments at 8%. The remainder is in property and venture capital[2]. Of the 46% invested in equities (approx £460bn), 60% is invested in UK

[2] At the time of writing the IMA has 29 separate asset classifications under 3 subheadings as follows. Income Funds: UK Gilts, UK Index Linked Gilts, UK Corporate Bond, UK Other Bond, Global Bonds, UK

domestic equity. Therefore, UK domestic equity is dominant in both total asset allocation and equity allocation which highlights the importance and relevance of examining UK domestic equity mutual funds in this study.

Of total assets managed in the UK, the fixed income allocation of £366bn is dominated by government debt (51%) and investment grade debt (48%). In terms of asset allocation by fund type, bonds dominate over equity among insurance funds but the opposite is true of pension funds. At over 70%, retail funds hold the highest proportion of equity.

3.1.4 Current Industry Trends

In the IMA *Asset Management Survey 2004*, the majority of respondents reported increased interest from private clients, some pension funds and other institutional investors in absolute return mandates and high alpha products. This demand for high alpha products in particular again emphasizes the relevance of establishing whether such fund managers can reliably deliver such a mandate.

The survey also noted an emergence of liability benchmarking among pension funds. This aims to predict a fund's future liabilities and invest in assets with matched characteristics. The impetus comes from companies who are increasingly concerned by the liability represented by their pension schemes. The results is a continued decline in Defined Benefit pension schemes and a shift towards fixed income rather than equity investments. However, a large number of survey respondents continue to expect open Defined Benefit and Defined Contribution Schemes along with open life assurance schemes to continue to generate demand for equities.

3.2 The Mutual Fund Data Set: Definitions and Sources

The mutual fund data set in this study is comprised of 1,620 Unit Trusts and Open Ended Investment Companies (OEICs). These are UK domestic equity funds, i.e. by definition at

---

Equity & Bond Income, UK Equity Income. Growth Funds: Money Market, Protected/Guaranteed Funds, UK All Companies, UK Smaller Companies, Japan, Japanese Smaller Companies, Asia Pacific Including Japan, Asia Pacific Excluding Japan, North America, North American Smaller Companies, Europe Including UK, Europe Excluding UK, European Smaller Companies, Cautious Managed, Balanced Managed, Active Managed, Global Growth, Global Emerging Markets, UK Zeros. Specialist: Specialist, Technology & Telecommunications, Personal Pension.

least 80% of the fund is invested in UK domestic equity. The data set in this study represents almost the entire set of UK domestic equity funds which have existed at any point during the sample period under consideration, April 1975 – December 2002. Unit trusts are pooled investments which enable their investors to enjoy economies of scale in gaining access to well diversified portfolios of securities. However, unit trusts often have different investment objectives as laid down in the trust deed. Unit trusts are 'open ended' mutual funds in the sense that investors can purchase new units in the fund at the going market price per unit, ie the demand for units does not increase the unit price. Unit trusts can only be traded between the investor and the trust manager, there is no secondary market. Unit trusts differ from investment trusts in that the latter may be described as a 'closed end' fund. Although they are still pooled investments, investment trusts are, in effect, companies which are quoted on the stock exchange in their own right whose business it is to trade in securities. Investment trusts have a fixed number of units, just as there are a fixed number of shares in a company. Unlike in the case of unit trusts, demand for investment trusts may push up the price of the trust. Here, it is possible for the price of the investment trust to trade at a premium (discount) where the price is higher (lower) than the value of the underlying assets of the investment trust. A premium, for example, may reflect investor demand for the skills of the investment trust manager. OEICs are constituted as companies so that investors buy shares but the number of shares in issue varies according to demand, again hence the term open ended. This implies that the share price always reflects the underlying asset value and, unlike investment trusts, is not affected by market sentiment towards the OEIC itself. Hence the risk profiles of OEICs are more in line with that of unit trusts than investment trusts. The data set of funds examined in this study includes OEIC's.

The mutual fund returns data have been obtained from Fenchurch Corporate Services using Standard & Poor's Analytical Software and Data. Returns are measured monthly between April 1975 and December 2002. All fund returns are fully contiguous.

The analysis is restricted to funds investing in UK domestic equity so that accurate benchmark factor portfolios may be constructed to estimate fund risk adjusted or abnormal performance. UK equity funds are defined by the Investment Management Association (IMA), formerly the Association of Unit Trusts and Investment Funds (AUTIF), as having at least 80% of the fund invested in UK equity.

Among the database of 1,620 funds, many funds are referred to as 'second units'. Second units arise for the most part when a single fund is sold under different pricing structures to different groups of investors such as retail and institutional. Furthermore, in addition to being offered directly by the fund managers, some funds are also sold under agreed but slightly different pricing structures by life assurance companies etc. Second units do not represent separate independent portfolios and they are not included in the analysis in this study.

Furthermore, 112 of the funds in the database are market tracker funds. As this study is interested in stock selection ability, the performance of tracker funds is of little interest and such funds are also excluded.

Concentrating on independent and non-tracker funds leaves a sample of 842 funds for the analysis.

All equity funds are classified by the investment objective/style of the fund. For example, funds which specialise in investing in small company stocks, funds which invest in income stocks etc. These investment classes (styles) are declared by the funds themselves but are certified initially and subsequently monitored monthly by the IMA in the UK. The IMA classifications for the UK domestic equity funds in this study are (i) Equity Income, (ii) Equity and (iii) Smaller Company. The IMA defines Equity Income funds as aiming to have a dividend yield in excess of 110% of the yield of the FT A All Share Index. Equity funds are defined as those having 80% of the fund invested in UK equity but are not restricted to a particular type of equity. Smaller Company funds have at least 80% of the fund invested in UK equities which form the bottom 10% of all equities by market capitalisation. Pre-1999 many Equity funds were subcategorized as either "Growth" funds or "Growth and Income" funds but these classes were merged in April 1999. Funds failing to remain within their stated investment objective for a period of three months are required by the IMA to either realign the fund or change the stated investment objective.

Within these classes of Equity Income funds, Equity funds and Smaller Company funds, more specialised IMA sub-classifications exist. These include funds investing in: Special Situations and Recovery, Funds of Funds, Mid-Cap stocks, Blue Chip stocks, Ethical stocks (eg non-tobacco related etc), Technology stocks, Finance stocks and

Ecology stocks (environmentally-friendly). These also include some capital protected investments. However, in many cases there are too few funds within each of these more specialised subcategories for them to be treated separately in a comparison between fund styles. Therefore this study uses the three broader classifications. There are 162 equity income funds, 553 equity funds and 127 small stock funds.

As of August 2005, the total value of unit trusts and OEICs under management in the UK was £277bn and is growing quickly. Sales in 2004 amounted to £54.5bn up from £18bn in 1995. Sales are approximately evenly divided between institutional and retail investors. Funds under management in the UK domestic equity sector total £117bn.

In order to control for survivorship bias, the data set includes both surviving funds (626) and nonsurviving funds (216). A nonsurviving fund is one which has existed for some time during the sample period but has not 'survived' until the end of the sample period. Nonsurviving funds may cease to exist because they were merged with (or taken over by) other funds or they may have been forced to close due to bad performance. Because of the latter scenario, it is critical to include nonsurviving funds in any performance analysis of the mutual fund industry as failure to do so may bias performance findings upwards. The inclusion of nonsurviving funds, which represent 25% of the total sample of funds examined, is a significant strength of this study.

Standard and Poors assign a unique identification number to all funds: the S&P ID. The S&P ID is helpful in tracing a fund's history through name changes and mergers with other funds etc. Name changes can be problematic in this type of research as it can lead to the inadvertent inclusion of the same fund twice. In this study, the S&P ID code of every fund was examined and name changes and mergers were identified in order to avoid this error. Examination of the S&P ID codes reveals that many funds whose history cease before the end of the sample period, ie 'nonsurvivors', were in fact taken over by other funds. The reason for the takeover may well be due to a fund's strong performance and therefore it would be wrong to assume that all nonsurvivors die due to poor performance. This is examined further in later chapters.

In addition, funds are also categorized by the location from where the fund is domiciled: onshore UK operated funds (662) and offshore funds (180). Offshore funds are comprised mainly of those operated from Dublin, Luxembourg, the Channel Islands, the

Isle of Man and some other European locations. In this study, all offshore fund returns are denominated in sterling for comparability with onshore funds.

All fund returns are calculated bid price to bid price with (gross) income reinvested. As the bid/offer spread captures the initial charge or load fee, as well as stamp duty, dealing charges, other commissions and bid/offer spreads of the underlying securities, the returns are gross of, or before, the load fee. However, returns are net of the annual management fees imposed. Fund returns are measured gross of taxes on dividends and capital gains. This is appropriate as the returns on the benchmark factor portfolios (for market, size, value and momentum) against which funds are measured are also gross of such taxes.

3.3 A Breakdown of the Data Set

In this section, a description of the data set of mutual funds used in this study is provided. The numbers of funds in various categories and some preliminary descriptive statistics of the fund returns are also presented. All fund returns discussed here are excess returns (ie are measured in excess of the one-month UK TBill rate) and are raw returns (not risk adjusted). Unless otherwise stated, the statistics presented in this section are not restricted to funds with a minimum number of observations

3.3.1 Number of Funds by Various Categories

This section presents breakdowns and cross-tabulations of the numbers of funds by investment objective, by survivorship and by location. Table 3.1 presents a contingency table of funds by investment objective and survivorship. Equity funds represent the largest single investment class with 553 funds in total and this is also the largest sub-class of funds among both surviving and nonsurvivng funds. A chi-square test of independence of the contingency table rejects the null hypothesis of independence between investment style and survivorship at 5% significance. This lends some support to the hypothesis that the survival status of funds is dependent on investment style. For example, nonsurviving funds comprise the highest proportion (33%) of equity income funds but a smaller 22% of equity funds. While the chi-square test does not reveal the cause of the dependence between survivorship and investment style, one obvious hypothesis is that performance differs between investment styles and in turn this alters the probability of a fund dying and becoming a nonsurvivor. To anticipate the results of later chapters, there is strong evidence to suggest that 'luck-adjusted' performance differs significantly between investment styles but poor performance is not the cause of many funds becoming nonsurvivors.

Table 3.1 Contingency Table of Investment Objective and Survivorship

	Equity Income	Equity	Small Stock	Total
Survivor	109	427	90	626
Nonsurvivor	53	126	37	216
Total	162	553	127	842

In Table 3.2 another contingency table is presented which reports the breakdown of the numbers of funds by investment objective and location: onshore/offshore domicile. 180 (21%) of the funds are operated offshore. A chi-square test of independence between fund location and fund investment style is strongly rejected at 1% significance. This may be the result, for example, of a perceived informational asymmetry by offshore fund managers in certain investment objectives. From Table 3.2, offshore funds represent a relatively large 26% of the investment in equity funds compared to only 10% and 13% in equity income and small stock funds respectively. This may reflect a preference by offshore fund managers to invest in broadly defined fund objectives rather than more specialized (and therefore possibly more informationally asymmetric) funds. The hypothesis that offshore funds may underperform relative to onshore funds, because of informational asymmetry or other factors, is examined in later chapters.

Table 3.2 Contingency Table of Investment Objective and Location

	Equity Income	Equity	Small Stock	Total
Onshore	146	406	110	662
Offshore	16	147	17	180
Total	162	553	127	842

3.3.2 Numbers of Funds and Performance Through Time

This section presents a breakdown of the numbers of funds and average fund performance, firstly by length of fund history and secondly over time throughout the sample period. The number of funds (excluding second units) in existence grew consistently over the sample period: In the month of December of 1975, 1985, 1995 and 2002, for example, there were

65, 233, 542 and 631 funds in existence respectively. Table 3.3 shows the numbers of funds, average returns and average standard deviation of returns by length of fund history. For example, 842 funds exist for at least one month and have an average return of –0.142% per month , 405 funds exist for at least 10 years (120 obs) while 61 funds exist for the full sample period (333 obs).

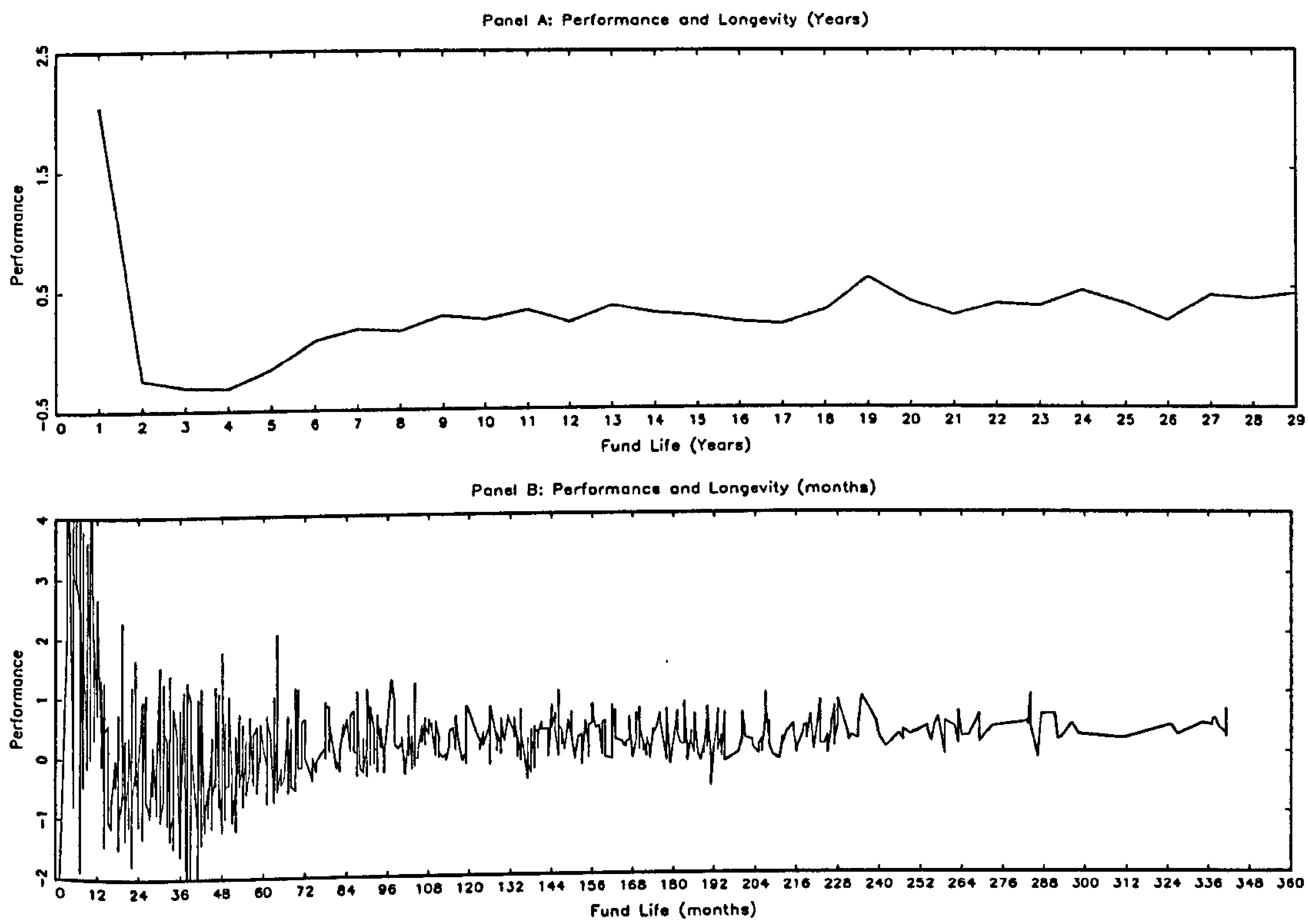
Table 3.3 Breakdown of Funds by Length of Fund Histories.

Minimum Observations	1	36	60	120	333
No. Funds	842	675	573	405	61
Mean Return	-0.142	0.172	0.279	0.360	0.522
Mean Stand. Dev.	4.916	4.824	4.817	4.967	5.217

Table 3.3 suggests that average performance rises with fund longevity. This might be partly explained by a higher concentration of poor performing nonsurviving funds among the shorter-lived funds. A more detailed picture of the relationship between performance and fund history is provided in Figure 3.1. Here funds are grouped by their number of observations and the average performance is plotted for each group. In Panel A, separate groups are formed of funds which existed for up to 1 year, between 1 year and 2 years, between 2 and 3 years and so on. In Panel B, similar groupings are formed by months rather than by years.

From Panel A, with the exception of very short-lived funds (up to 1 year), generally performance is relatively poor among shorter-lived funds up to approx 5 years but rises gradually in longer-lived funds. Again, this may be partly due to the higher concentration of some poor performing nonsurvivors in this category. However, short-lived funds here are not all nonsurvivors. They include funds which came into existence towards the end of the sample period and have relatively few observations but continue to exist at the end of the sample period. Therefore, some of these short-lived funds are recently launched and may be 'star' performers while others will be earlier shorter-lived poor performing nonsurvivor funds. This partly explains the higher cross sectional variation in returns among shorter versus longer-lived funds in Panel B of Figure 3.2. However, this relatively high cross-sectional variation is also explained by the fact that short-lived funds existed at different time periods within the full sample period, some of which were bull periods and some of which were bear periods, while longer-lived funds produce a smoother average return over time.

Figure 3.1 Performance and Fund Longevity



To further investigate whether poorer performance among short-lived funds is due to nonsurvivorship, Panel A of Figure 3.1 is plotted for both survivor and nonsurvivor funds separately in Figure 3.2. It is clear that nonsurviving funds with 2 to 3 years of observations or less underperform similar surviving funds. This tends to suggest that many of the nonsurvivng funds of this age which close do so because of poor performance. However, Figure 3.2 also demonstrates that longer-lived 'nonsurviving' funds outperform their age-counterpart surviving funds. This raises an important issue regarding funds labeled as "nonsurvivors": many such funds are not closed down because of poor performance. On the contrary, it may be their good performance and consequent attractiveness that causes them to be merged with, or taken over by, other funds.

In Figure 3.2 nonsurviving and surviving funds are grouped and compared according to the length of the fund history. Of course, in this comparison nonsurviving and surviving funds are not necessarily being compared at the same points in time. To provide further insight into the relative performance, Figure 3.3 charts the difference in returns between these two classes of funds through time. For each month the cross-sectional average return for nonsurvivors and survivors is calculated, this is then averaged within each year and the difference between the two (nonsurvivors minus survivors) is observed For the most part nonsurvivors outperform survivors through time further emphasizing the caveat above that it would be incorrect to conclude generally that nonsurvivors close due to poor performance.

Figure 3.2 Relative Performance by Number of Observations: Surviving and Nonsurviving Funds

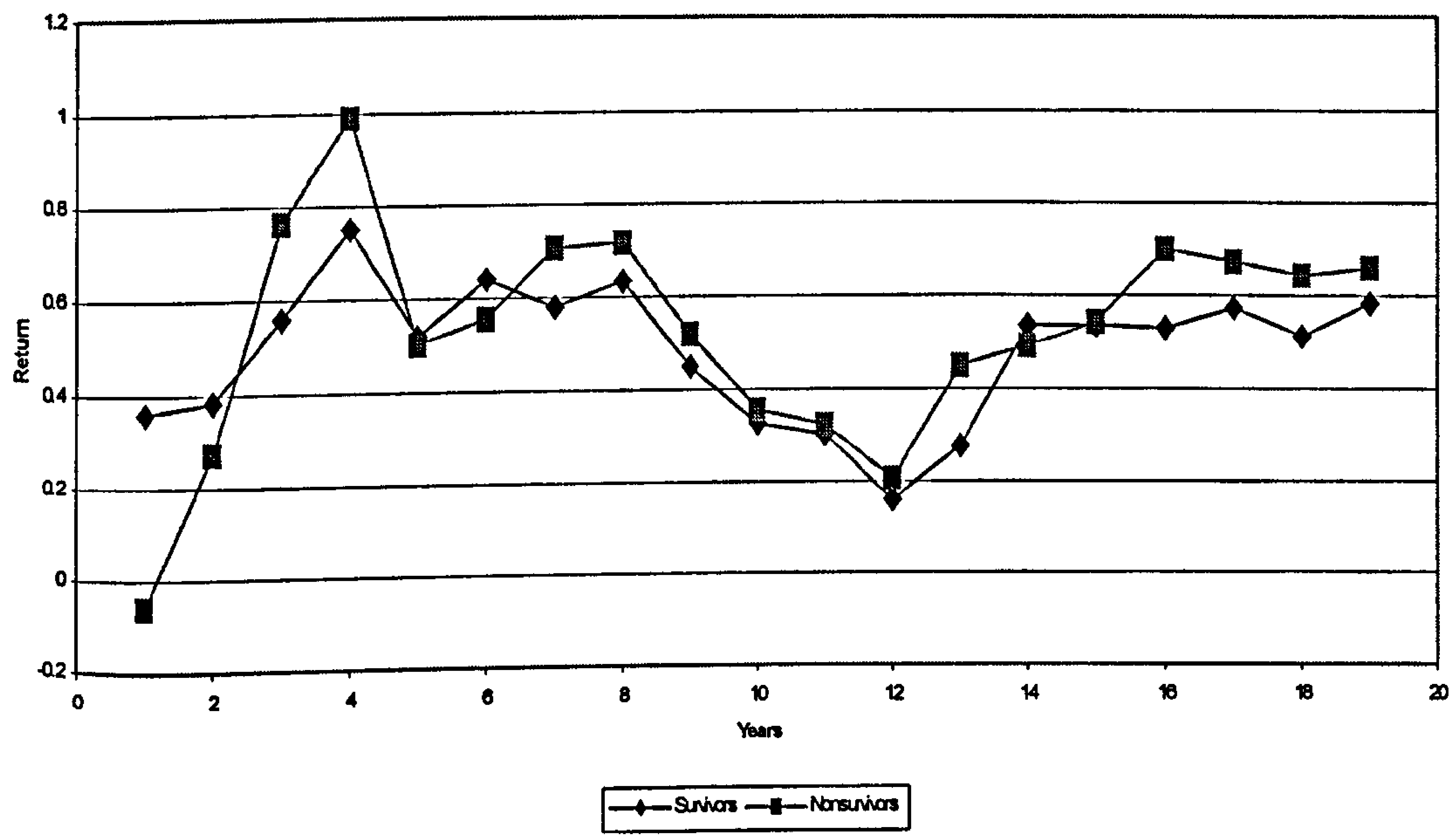
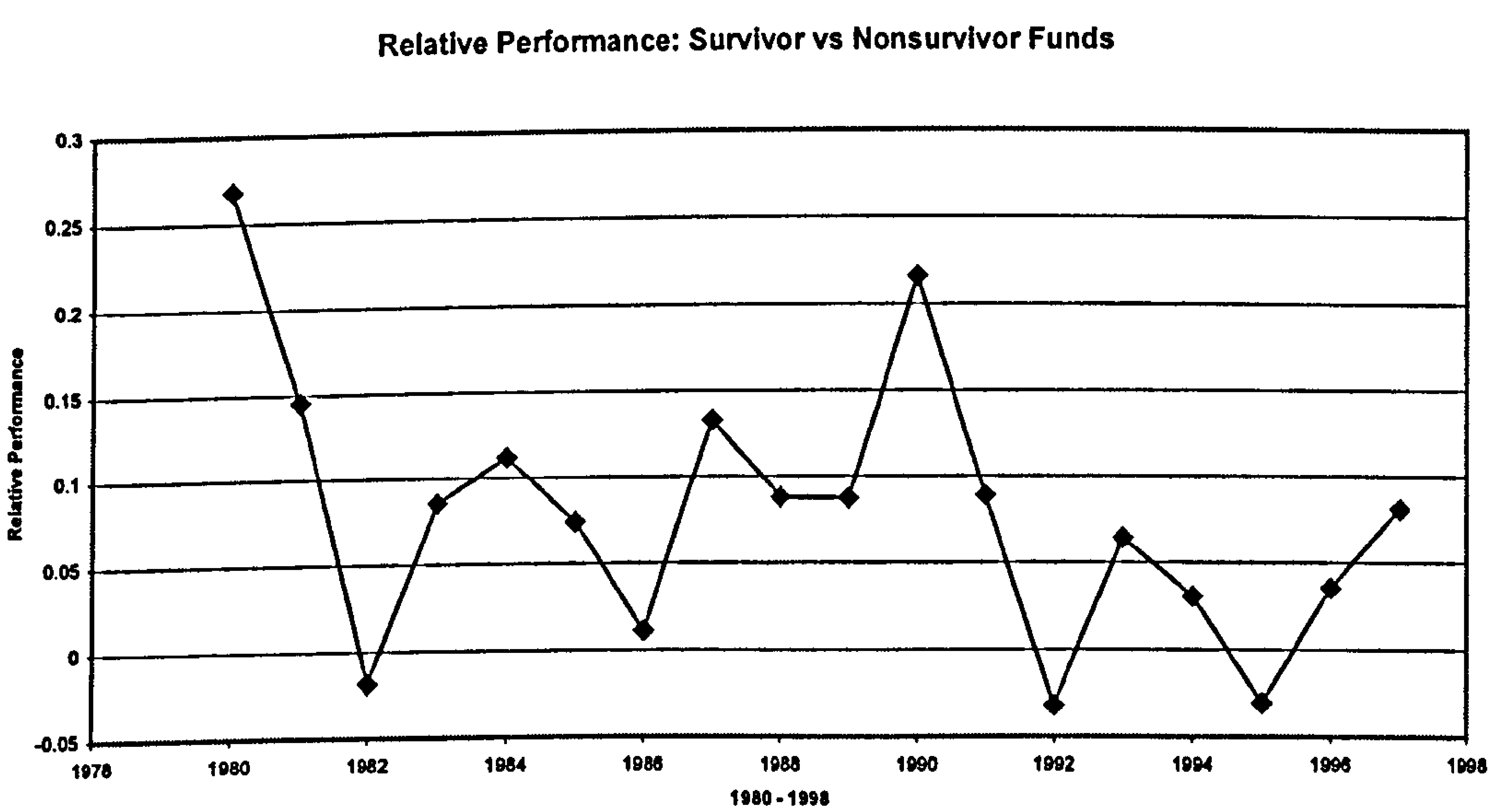


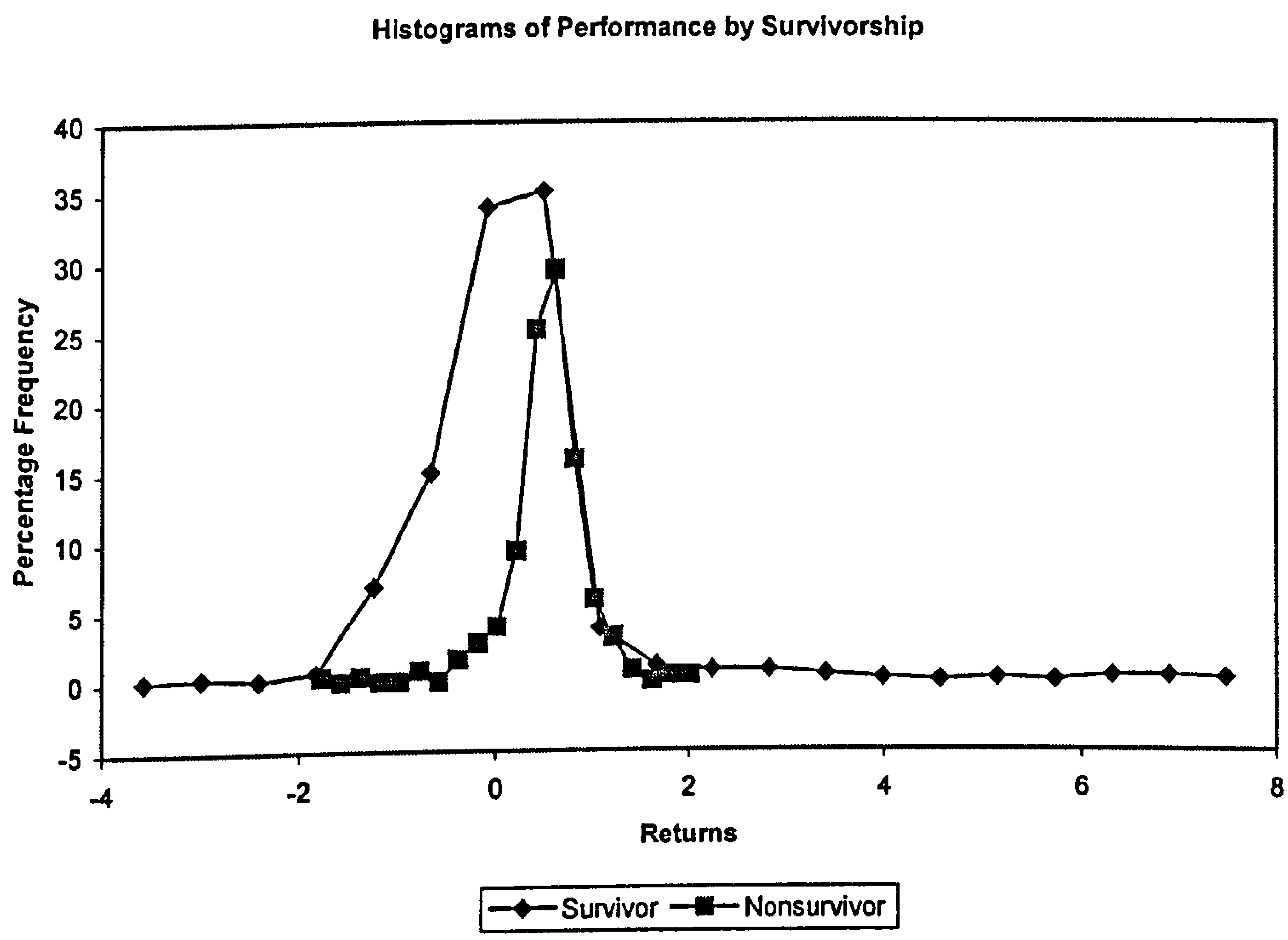
Figure 3.3 Relative Performance : Nonsurvivors - Survivors



The cross-sectional distribution of (time series) average fund returns within each group of surviving and nonsurviving funds is plotted in Figure 3.4. The graph reveals a considerably higher degree of variation in (time series) average returns among surviving funds. In addition, the relative underperformance of surviving funds as a group reported previously is concentrated largely in the negative tails of the distributions.

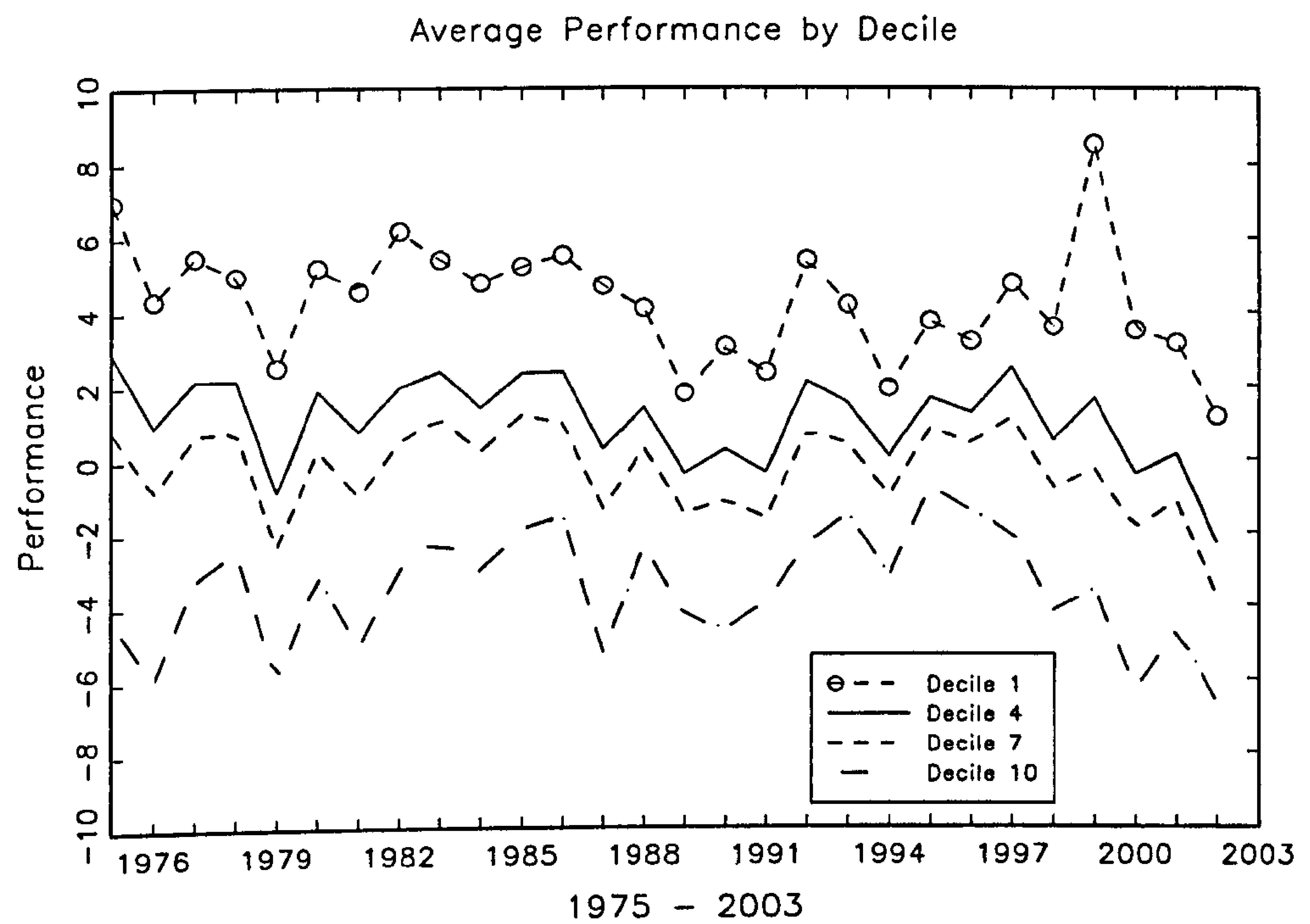
The sensitivity of fund performance to length of fund history is examined in more detail chapter 6 where a distinction between look-ahead bias and survivorship bias is discussed.

Figure 3.4 Distribution of Fund Average Returns by Survival



Further insight into the level of cross sectional variation in fund returns is provided by Figure 3.5 which charts selected decile performance over time. Each month over the full sample period funds were ranked and sorted into decile portfolios and average returns in each decile were calculated. Monthly averages for each year were then calculated. For ease of graphical illustration, time series of selected deciles 1,4,7 and 10 are plotted. The graph reveals substantial differences between the top and bottom decile performance which was often as wide as an average 10% per month. The difference between top and bottom decile performance peaks between 1998 and 2000 where stock sectors related to technology performed particularly well while more traditional stocks performed poorly in relative terms. Note, the gap between top and bottom decile performance narrows post-2000 in line with a revision in market views regarding 'technology stock' valuations. The breakdown in performance between funds of different investment objectives around this time is discussed in the next section.

Figure 3.5 Average Performance by Decile



3.3.3 Performance by Sector

This section further describes the relative performance of sub-classes of funds such as funds of different investment objectives and onshore versus offshore funds. Figure 3.6 charts the performance over time of the three investment objective categories. Within each class of funds, each month the cross-sectional average return is calculated and then monthly averages are taken for each year. The graph suggests that small stock funds outperform the other investment classes during a market-wide upswing. However, this class of funds also exhibits a higher degree of volatility through time and often yields a worse performance on average during a market-wide downswing. Indeed small stock funds have the highest average monthly return of 0.5% per month and the highest standard deviation of returns of 1.68%. For equity income funds and general equity funds the average return per month and standard deviation of returns are (0.47, 1.22) and (0.42, 1.21) respectively.

Note from Figure 3.6 that small stock funds outperform the other classes in the late 1990s while income stock funds underperform over the same period. Post 2000 these relative rankings are reversed. As alluded to previously, this reflects the rapid growth in the value of stock associated with information technology (IT) during the earlier period many of which were small and often new companies established to operate either directly in the IT industry or associated service industries. Well established income stocks were comparatively 'out of favour' during this period. Relative performances between these sub-classes of funds were reversed from around the first quarter of 2000 following a significant correction in market views regarding the valuation of many technology stocks.

Figure 3.7 provides further detail on the distribution of returns by investment objective. In each class, the time series average return of each fund is calculated. Each histogram shows the distribution of these averages. Again, the high variability in the returns of small stock funds is clear.

Figure 3.6 Performance by Investment Objective Over Time

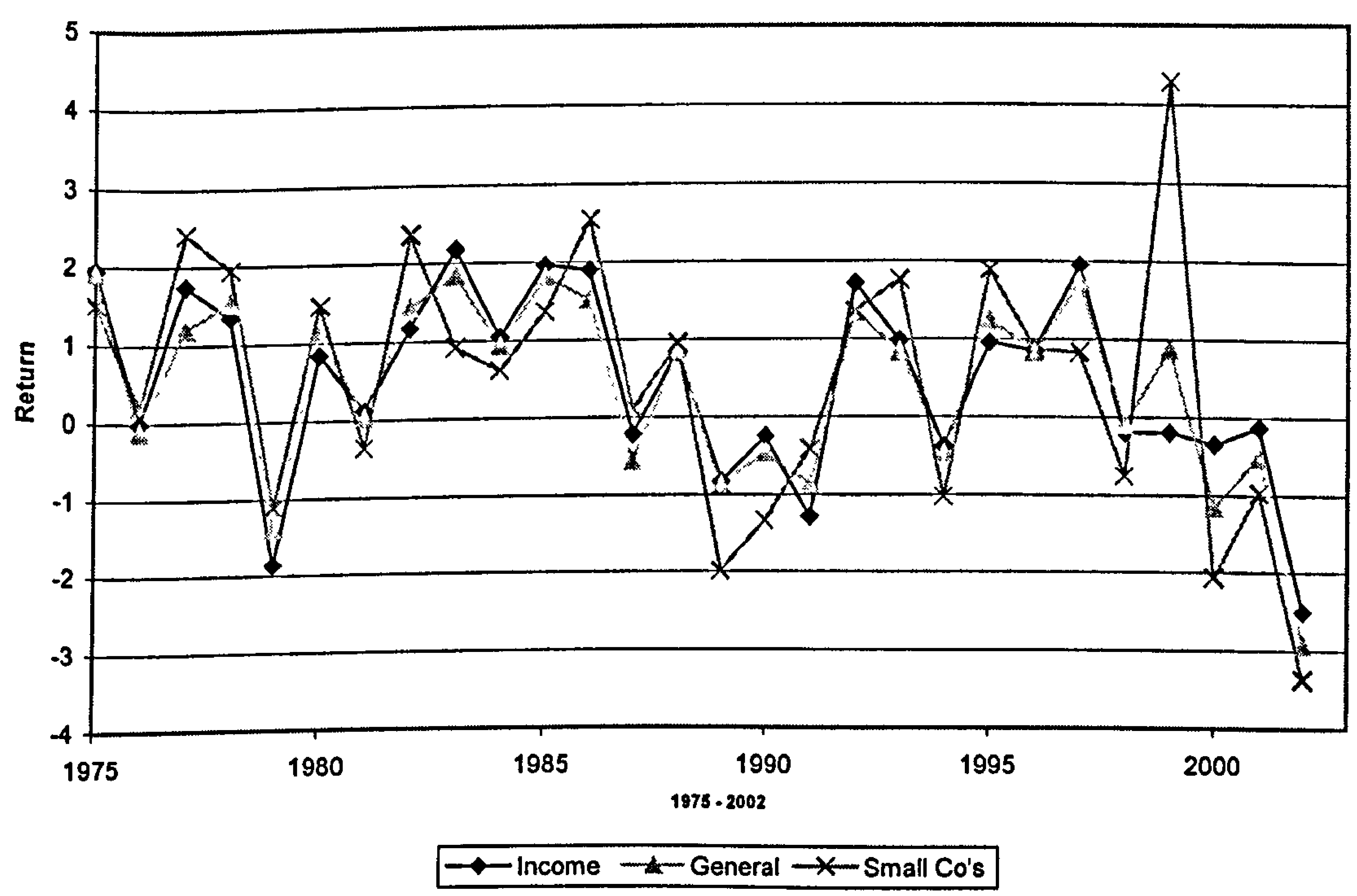
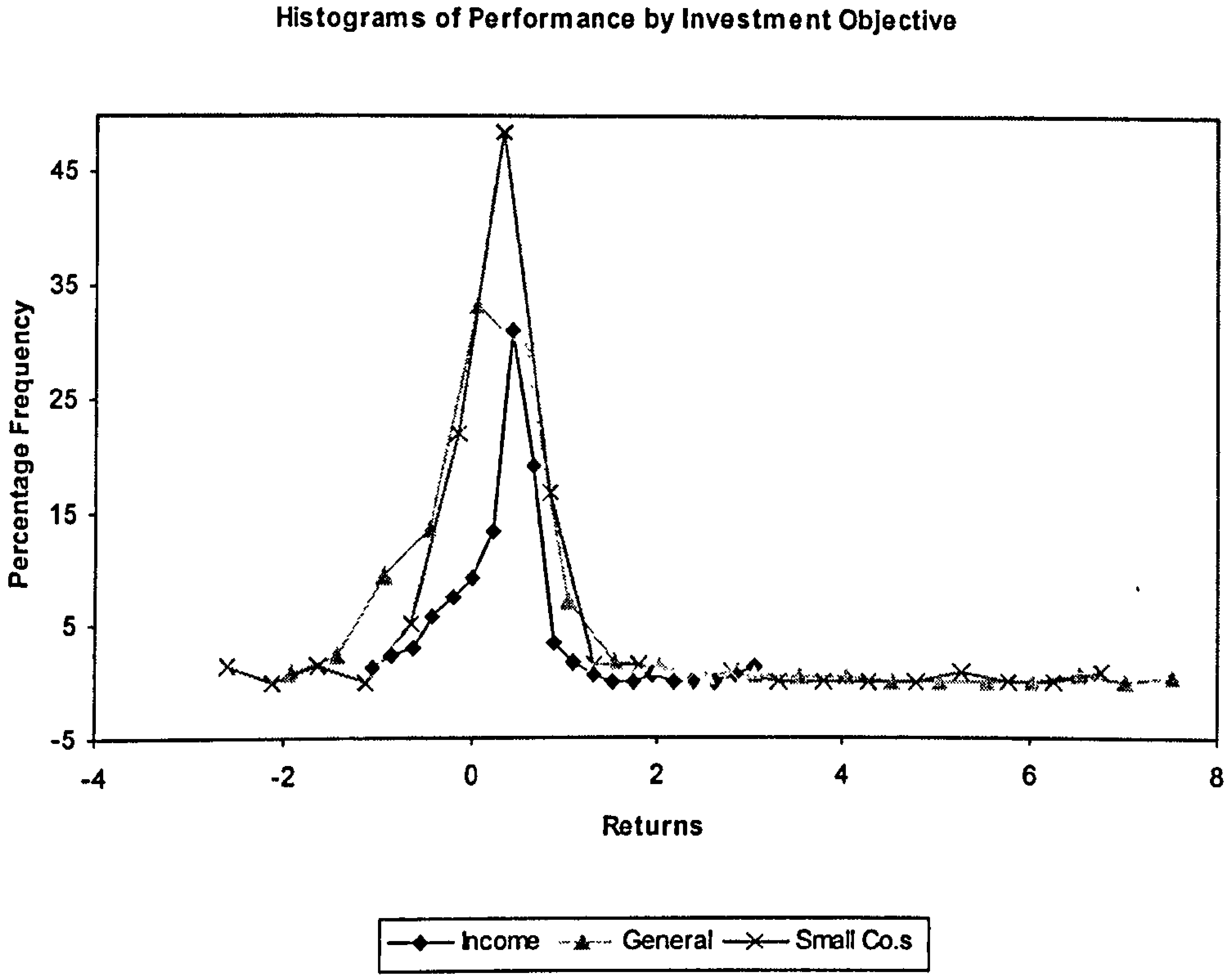


Figure 3.7 Distribution of Fund Averages by Investment Objective



It is also of interest to examine the relative performance of onshore and offshore funds over time. Figure 3.8 graphs these time series. In each sub-class, each month the cross-sectional average return is calculated. Monthly averages are taken for each year. Not surprisingly, both subgroups produce a similar pattern of returns over time as they select from the same population of UK equity. However, onshore funds, on average, do appear to yield a marginally higher return and seldom underperfom. Over the sample period as a whole, onshore funds yield a cross-sectional and time series average return of 0.45% per month against 0.28% for offshore funds.

Figure 3.9 provides a similar insight and also shows the cross-sectional variability in returns in both subgroups. The histograms show the distribution of the time series averages of fund returns. It is apparent that the distribution of onshore fund returns lies generally to the right of that of offshore funds and with wider tails it exhibits a higher degree of fund return variability.

Figure 3.8 Performance by Fund Location Over Time

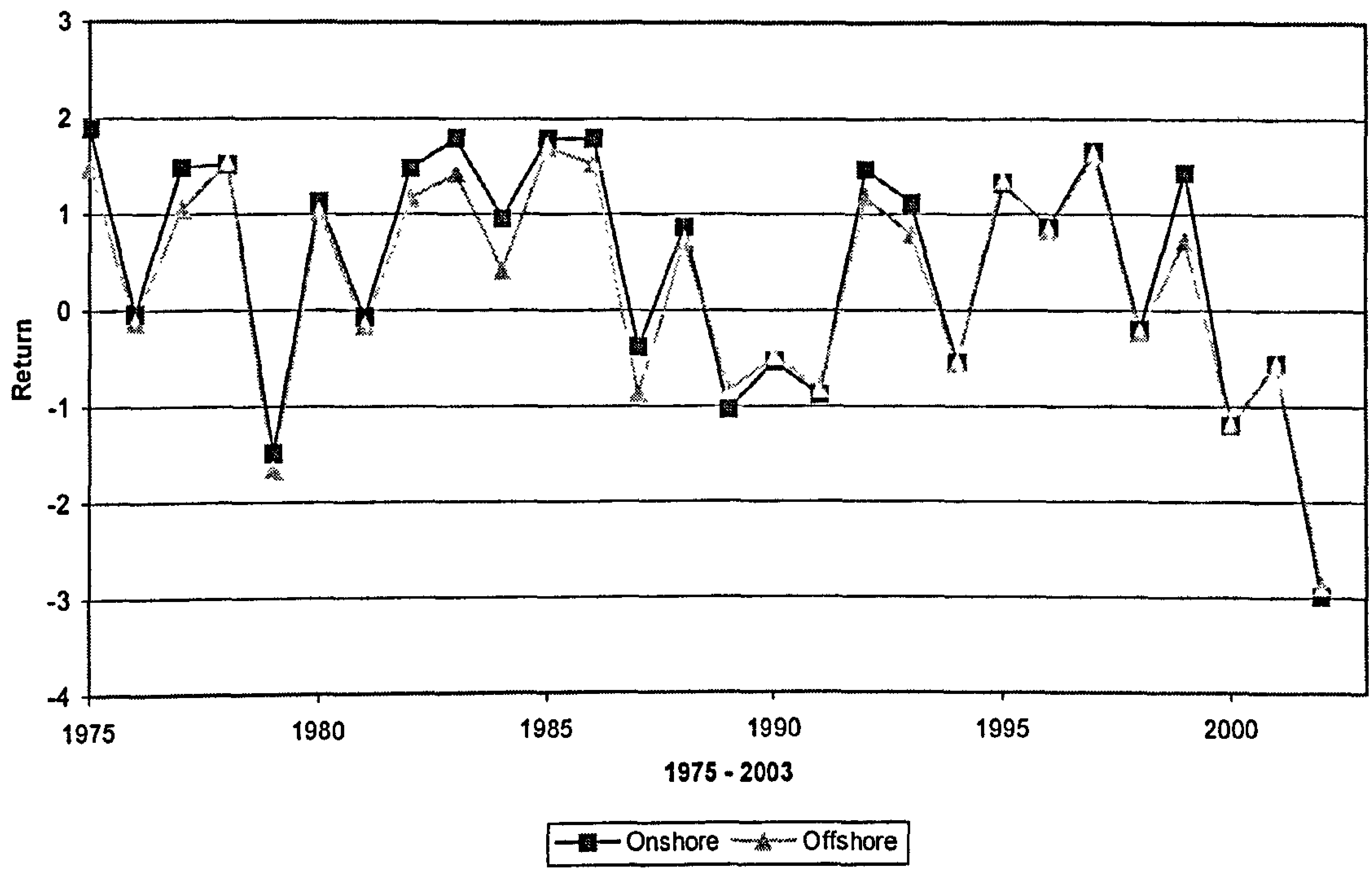
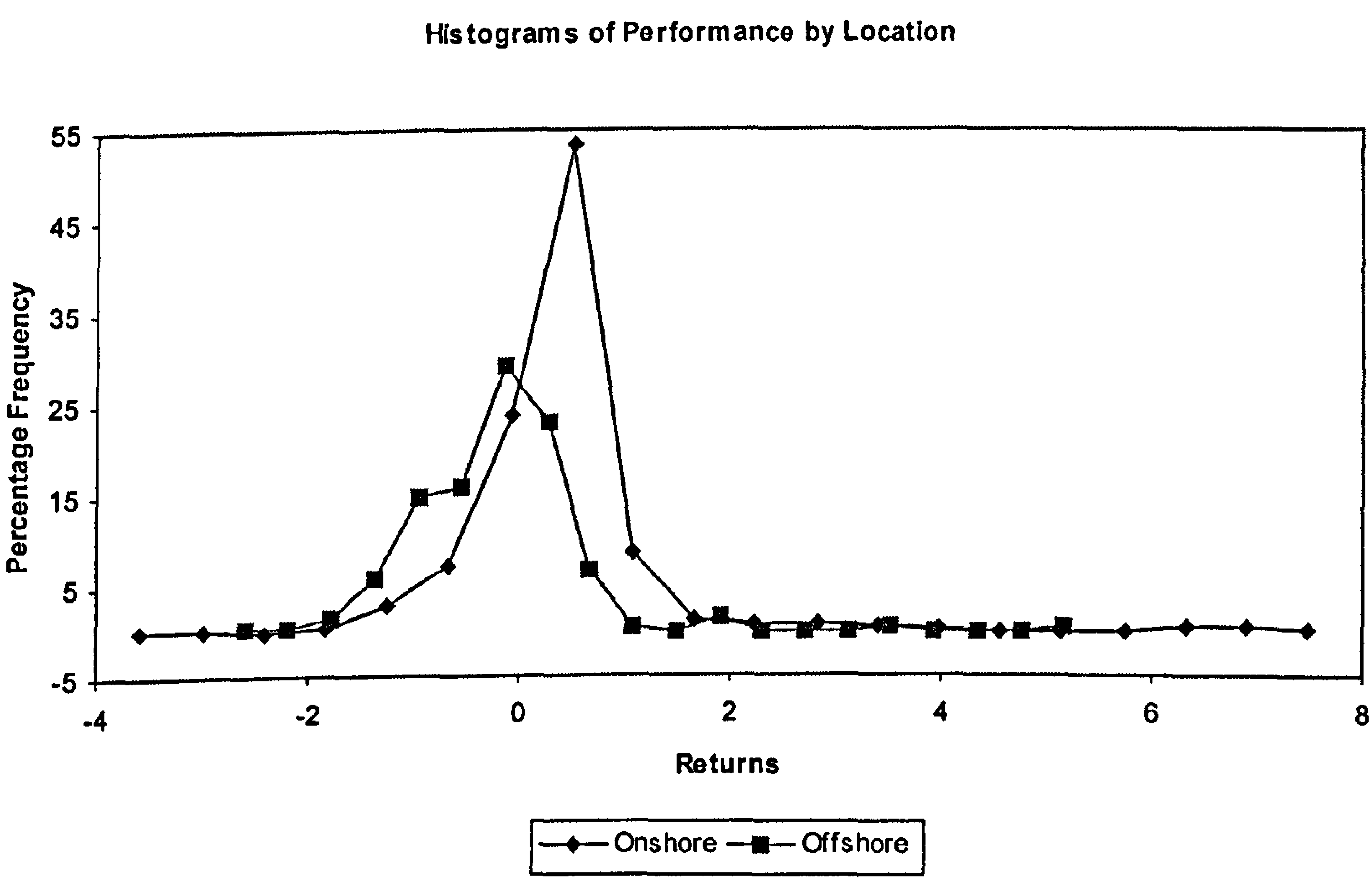


Figure 3.9 Distribution of Fund Averages by Fund Location



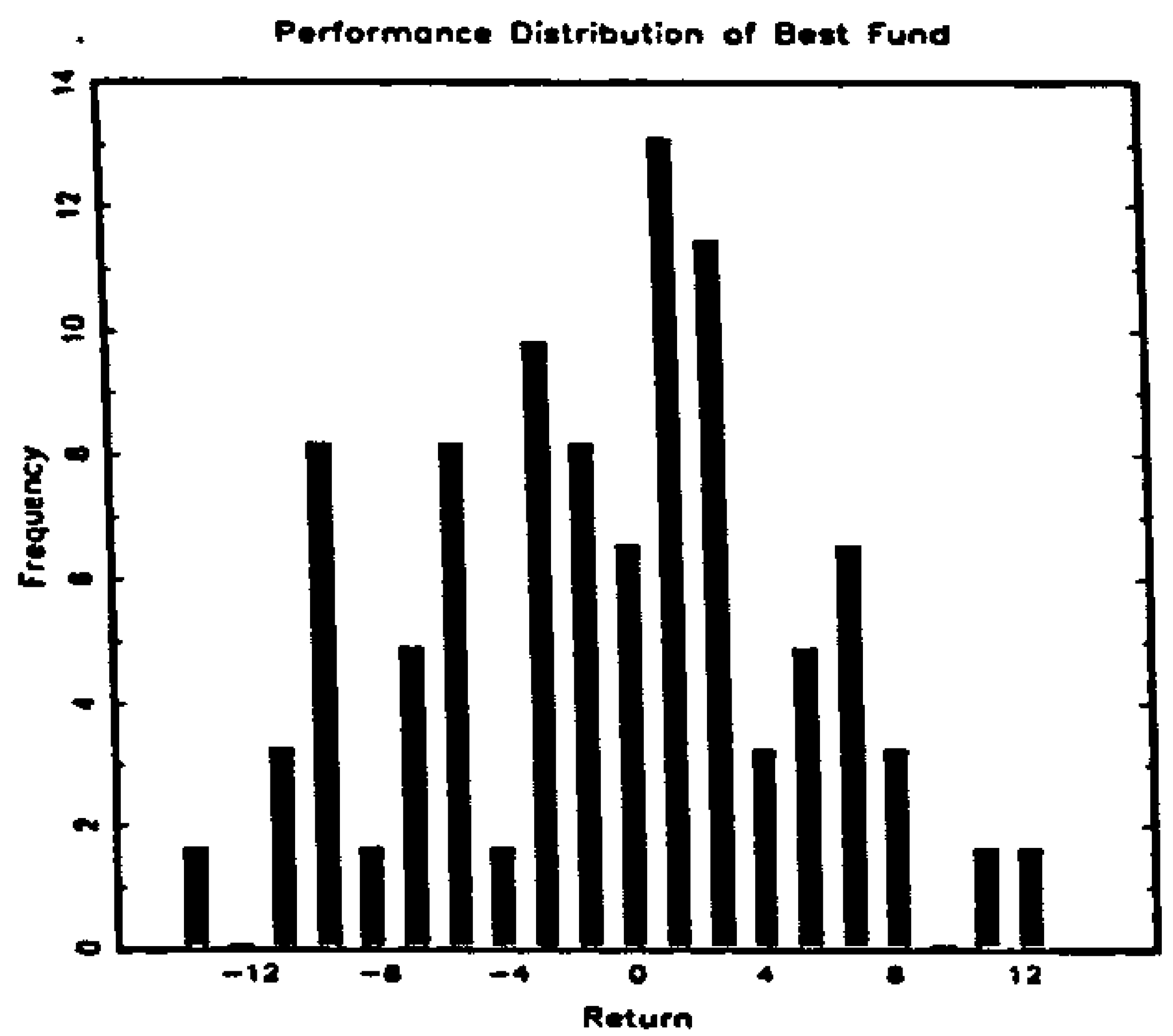
Figures 3.8 and 3.9 suggests that there may be differences in the stock selection ability between onshore versus offshore funds. This is also investigated in later chapters.

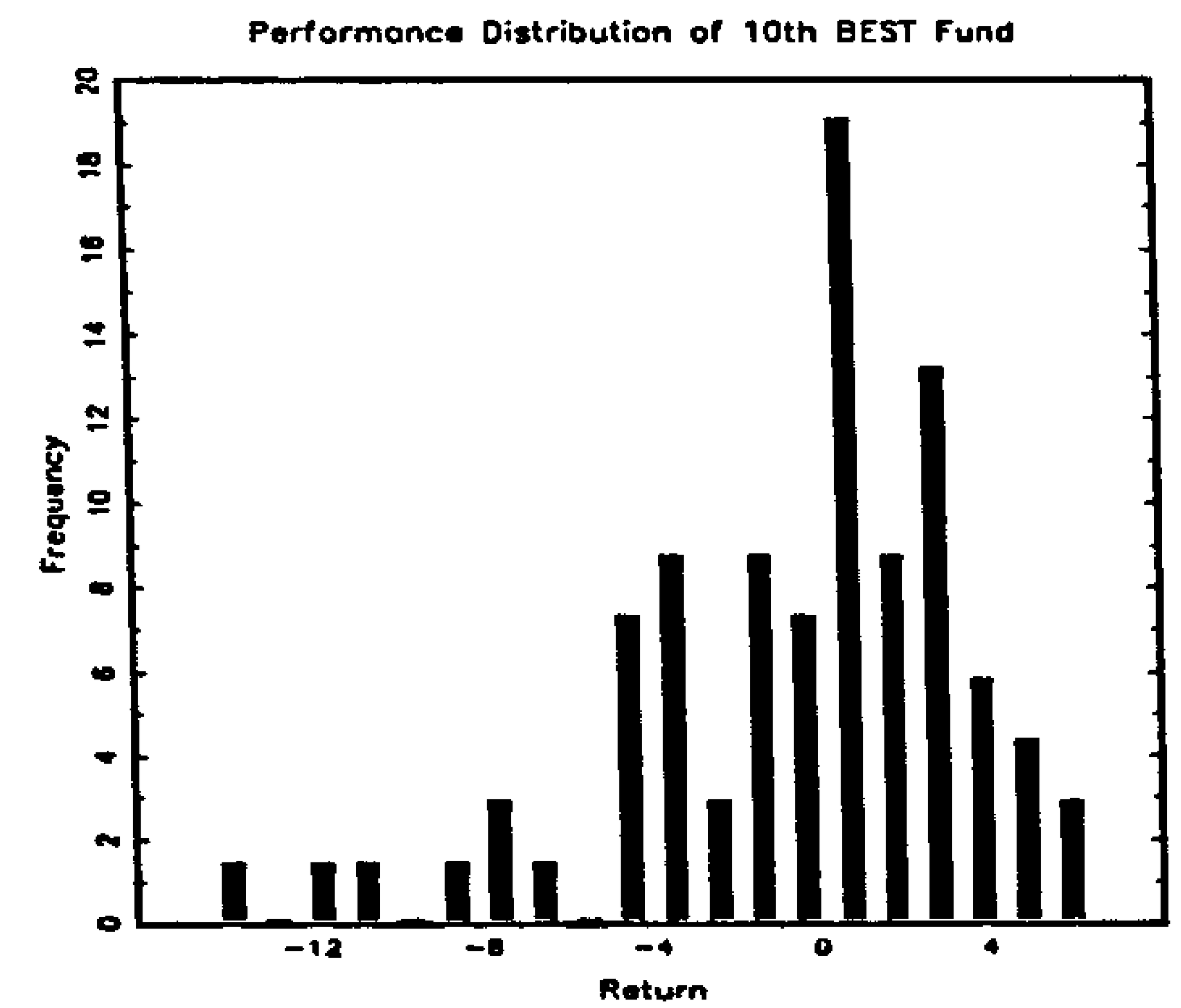
3.4 Normality of Fund Returns

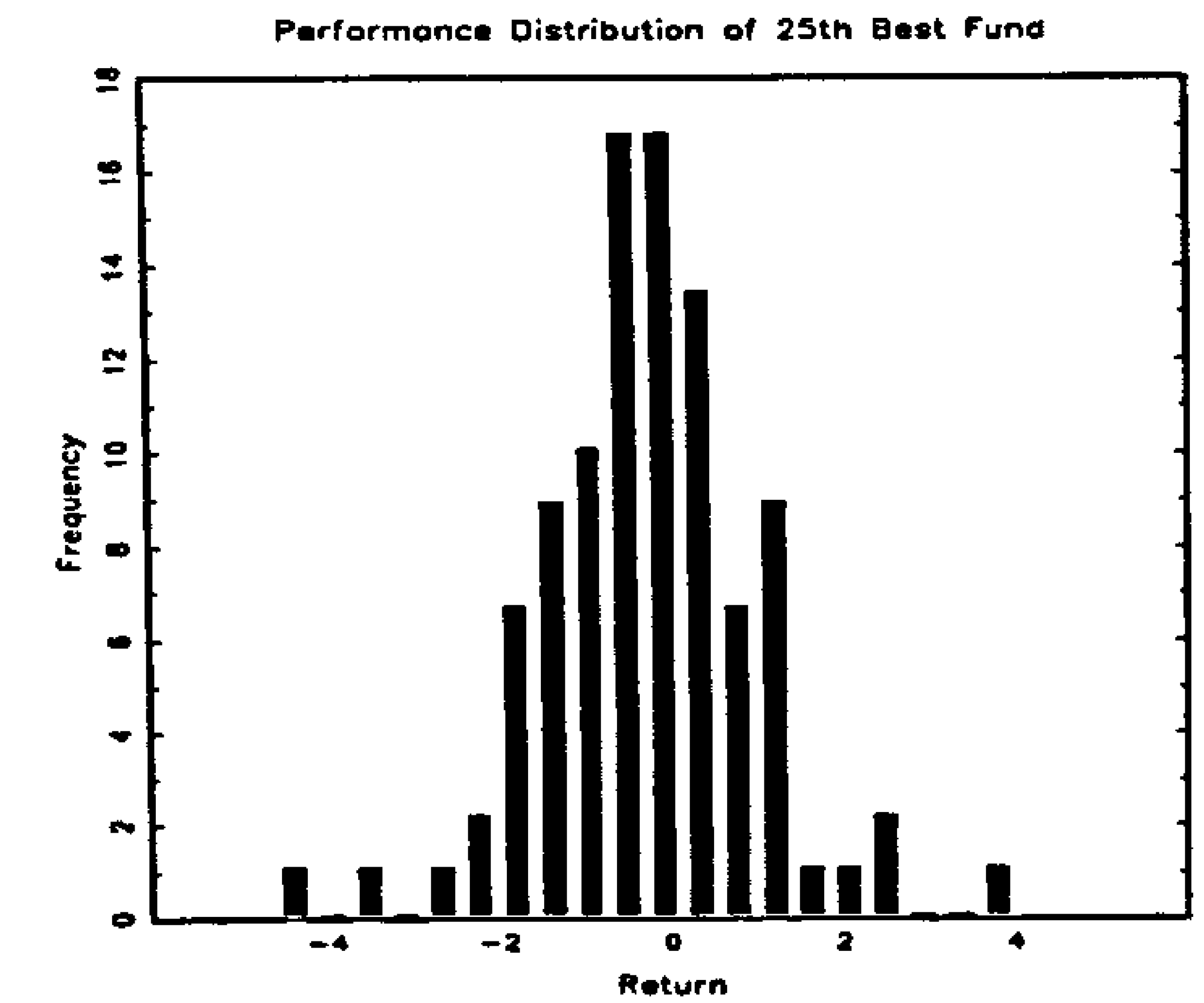
In chapter 4 many equilibrium models of security returns are described. In these models, the intercept term, alpha, is a measure of risk adjusted or abnormal fund return. Typically, its significance is determined by a standard t-test which in turn requires normally distributed fund returns for its validity. Chapter 4 provides evidence that this is not the case among UK equity mutual funds. Furthermore, there is evidence that the degree of non-normality is greater among top and bottom funds which are the funds of greatest interest to investors.

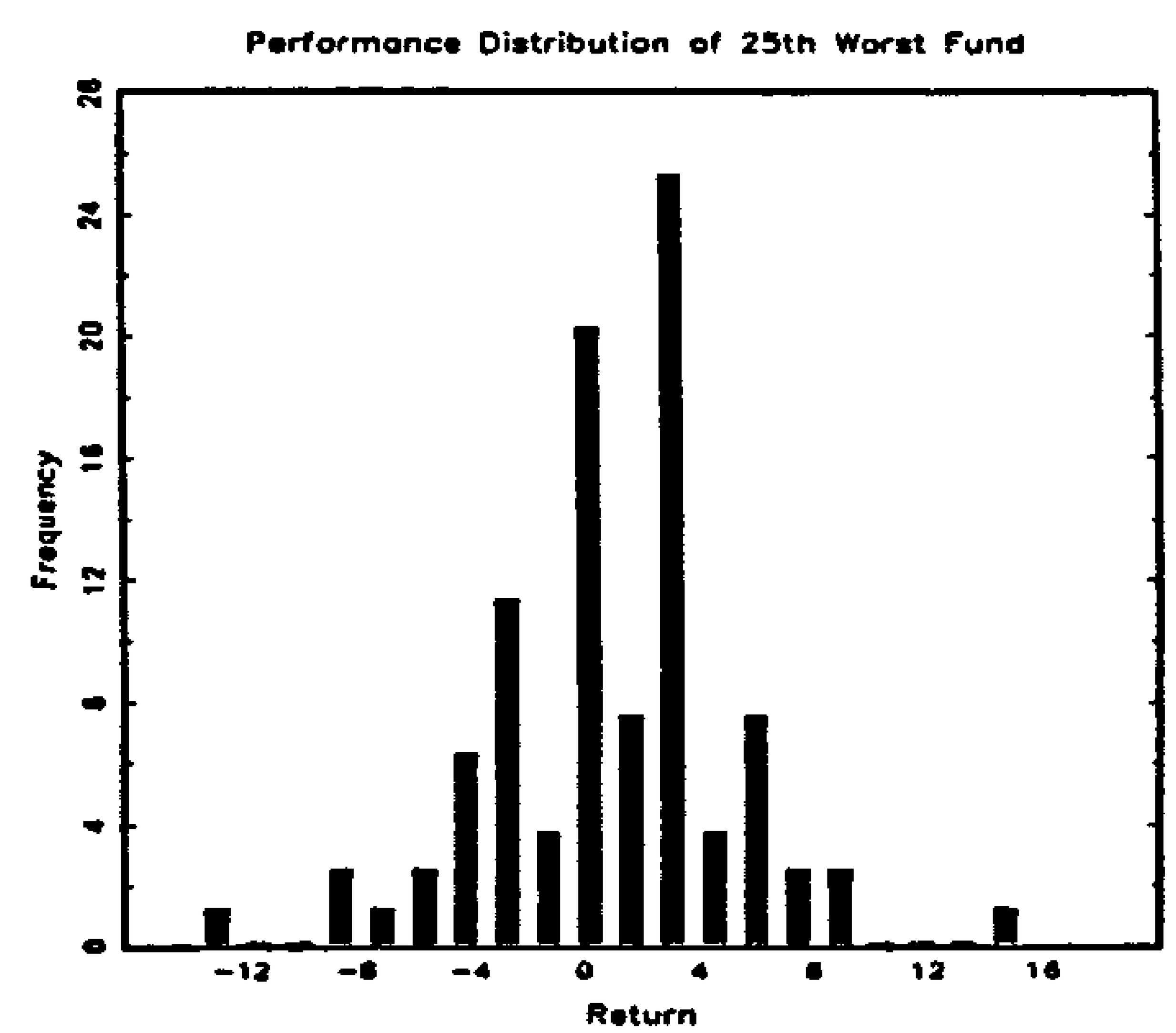
This pattern is illustrated in Figure 3.10. First, all funds were ranked by the time series average return. Figure 3.10 then plots the monthly returns of selected ranked funds. The upper left panel plots the distribution of the monthly return observations of the highest ranked fund. The upper middle panel plots the return observations of the 10^{th} best fund and so on as indicated. Similarly, the lower panels plot the distributions of returns of funds ranked at selected points at the lower end of the cross-sectional performance distribution. It is notable that funds at the top and bottom ends of the spectrum exhibit a higher variance and a greater non-normality in returns than funds ranked even slightly closer to the centre of the distribution. Therefore standard t-tests are particularly unreliable in assessing the performance of funds in the performance tails.

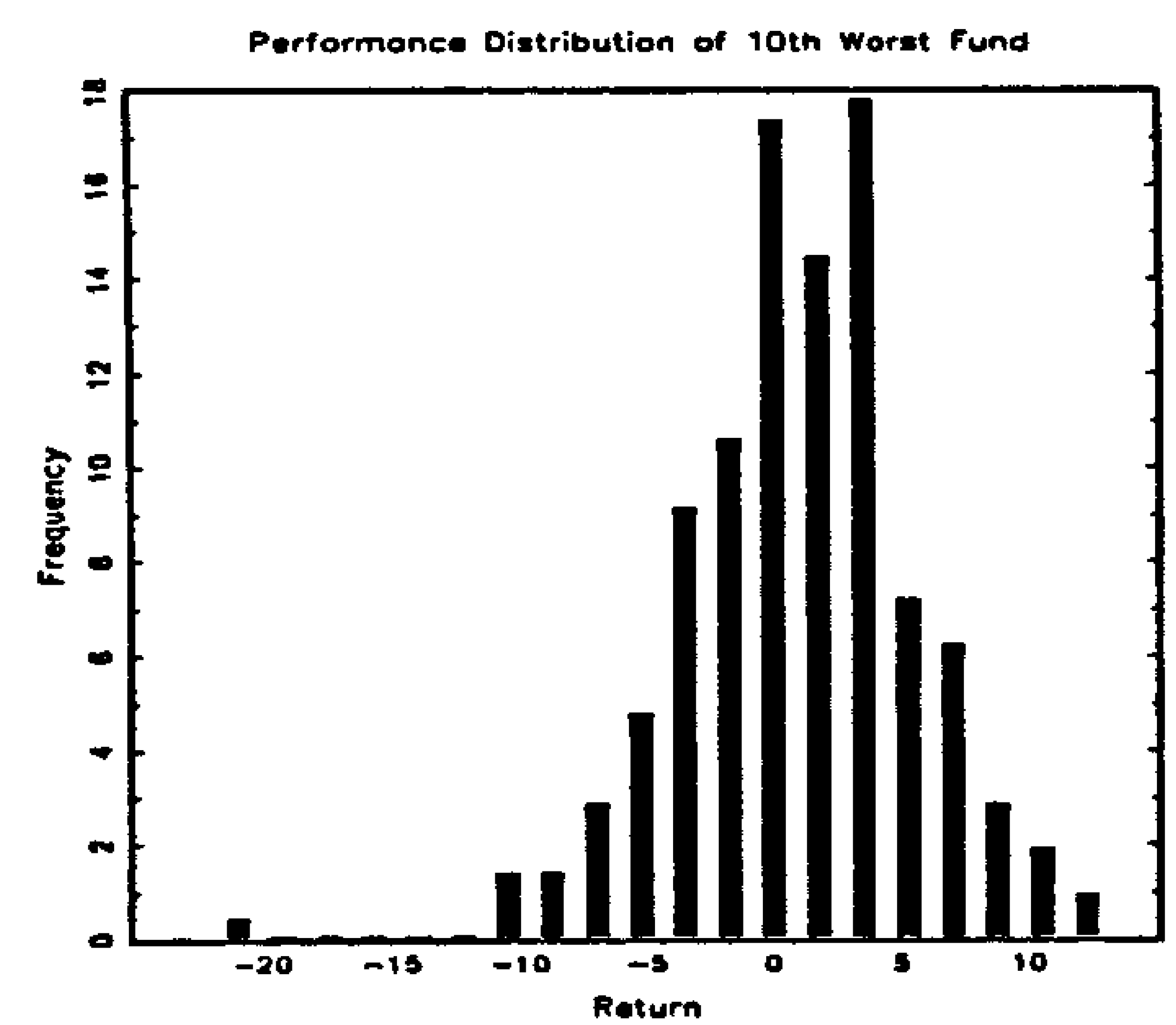
This finding largely motivates the use of the bootstrap methodology, described in chapter 5. This is a nonparametric procedure to construct 'empirical' distributions of abnormal (risk adjusted) return under the null hypothesis of zero abnormal return at all points in the performance distribution including in the extreme tails.

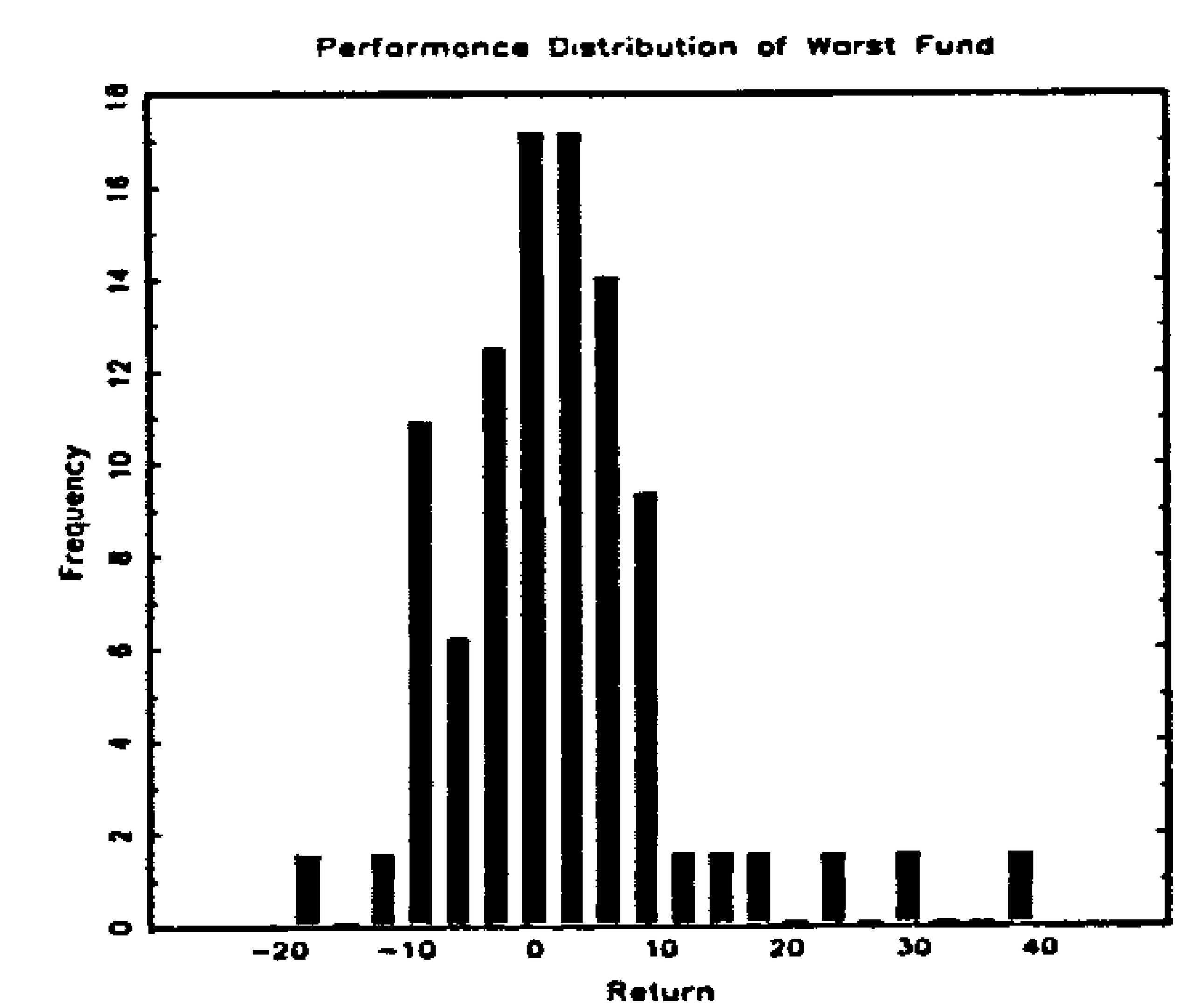

Performance Distribution of Best Fund
Frequency
Return


Performance Distribution of 10th BEST Fund
Frequency
Return


Performance Distribution of 25th Best Fund
Frequency
Return


Performance Distribution of 25th Worst Fund
Frequency
Return


Performance Distribution of 10th Worst Fund
Frequency
Return


Performance Distribution of Worst Fund
Frequency
Return

3.5 The Benchmark Factor Portfolios.

In this section the construction of the benchmark factor portfolios and conditioning variables as well as their data sources are described and summary statistics are presented. A detailed discussion of the theory underpinning these variables in performance modes is presented in chapter 4. In this study, excess mutual fund returns and excess market returns are calculated using the one-month UK TBILL rate taken from Datastream.

3.5.1 The Market Factor

The market index is the FT A All Share Index (The Financial Times Actuaries All Share Index) taken from Datastream. This is the most comprehensive UK stock market index. The index is an arithmetic mean of 800 shares and fixed income stocks which comprise more than 90% of the market capitalization of all listed companies on the London Stock Exchange. The FT A All Share Index is available as a price index and as an index of total returns. The latter also incorporates reinvested dividends and is more appropriate for use in this study because of its comparability with the mutual fund total returns. Unfortunately, however, the FT A All Share index of total returns is unavailable pre-January 1986. As a 'second-best' approach in this study in order to construct a consistent market index of total returns for the full sample period, one-twelfth of the monthly FT A All Share dividend yield is added to the FT A All Share Price index each month between April 1975 – December 2002. Using the FTSE A All Share Index as the benchmark market factor implicitly assumes that this is the universe of stocks from which fund managers select and it is the benchmark against which fund performance is measured. Recall that in this study only UK domestic equity mutual funds are examined.[3]

[3] This index excludes the Alternative Investment Market (AIM), which replaced the Unlisted Securities Market (which operated between 1980-1996), the Third Market (which operated between 1987-1990) and an earlier London Stock Exchange Market known as the section 4.2 market. The AIM includes shares not suitable for a listing such as some small cap stocks, some management buy-outs and closely-held shares (ie held by few people such as family members). (See Blake 2000 for further details). These are likely to form a very small proportion, if any, of most mutual funds' stock holdings. Furthermore, as these stocks are likely to be among the more risky assets, to include them in the benchmark market factor may be misleading.

3.5.2 The Size Factor

The factor mimicking portfolio for the size effect, SMB, is the difference between the monthly returns on the Hoare Govett Small Companies (HGSC) total return index and the returns on the FTSE 100 total return index. The HGSC index is an index of the total returns of the lowest 10% of stocks by market capitalization of the main UK equity market. It is a widely used index of small cap stocks. It is produced by Professors Elroy Dimson and Paul Marsh of London Business School. The FTSE 100 index is an index of the total returns of the UK's largest 100 companies by market capitalization. It is compiled by the *Financial Times*, the Institute and the Faculty of Actuaries and the London Stock Exchange. It covers approximately 75% by value of all UK shares. As total return indices, both the HGSC and FTSE indices incorporate a measure for reinvested dividends.

3.5.3 The Value Factor

The factor mimicking portfolio to model the value premium, HML, is the difference between the monthly returns of the Morgan Stanley Capital International (MSCI) UK value index and the returns on the MSCI UK growth index. Both indices are total return indices. To construct these indices Morgan Stanley ranks all the stocks in their UK national index by the book-to-market ratio. Starting with the highest book-to-market ratio stocks are attributed to the value index until 50% of the market capitalization of the national index is reached. The remaining stocks are attributed to the growth index. The MSCI national indices have a market representation of at least 60% (more recently this has been increased to 85%). The national indices are designed to represent a broad range of sectors rather than simply representing the highest market capitalization stocks.

3.5.4 The Momentum Factor

The factor mimicking portfolio to capture the momentum effect, PR1YR, is constructed from the London Share Price Database (LSPD) and this section should be read in conjunction with section 3.7 below. Each month all stocks in the LSPD are ranked based on their cumulative return over the previous 11 months (ranking period). Equal weighted portfolios of the top 30% of the stocks and the lowest 30% of the stocks are formed and held for one month (holding period). PR1YR is the difference in the holding period between the portfolio of past winners and past losers. The procedure is carried out in a rolling window moving the ranking and holding periods forward one month at-a-time.

This definition of momentum is as equivalent as possible to that in Carhart's (1997) US study. As a sensitivity analysis, in this study the momentum variable is constructed based only on (i) FTSE A All Share Index constituent stocks and (ii) FTSE 350 Index constituent stocks and in each case for permutations of six, three and one month ranking and holding periods. See also section 4.2.1, chapter 4.

In the construction of the momentum factor each month, the variable is based on the LSPD constituent stocks as they existed at *that* point in time *not* based on the historical returns of the LSPD constituent stocks at the end of the sample period (December 2002). The latter case would risk imposing a survivorship bias by limiting the analysis to stocks which had survived.

As the construction of the momentum factor relies on the LSPD, the variable can only be generated from December 1979 onwards. The LSPD Archive (see below) begins in January 1979. After the first holding period of 11 months the first observation of the momentum variable is then December 1979.

3.6 The Conditioning Variables

The instruments used in this study in the conditional beta and conditional alpha-beta models (see chapter 4) are sourced as follows: the one-month TBill rate is taken from Datastream. The dividend yield of the market factor is the dividend yield on the FT A All Share index taken from Datastream. The slope of the term structure is the yield on the UK 10 year gilt minus the yield on the UK one-month TBill, both taken from Datastream.

3.6.1 Descriptive Statistics: Benchmark Factors and Conditioning Variables

Table 3.4 reports summary statistics of the factor portfolios. The high standard deviations indicate that the factor portfolios may be able to explain considerable variation in returns. The t-statistics to test whether the means of the factors equal zero strongly reject the hypothesis for the size factor suggesting that of the three deviation variables (SMB, HML, PR1YR) size could account for much of the cross-sectional variation in the mean returns of funds. In addition, the low cross-correlations imply that multi-collinearity does not affect the estimated factor loadings.

Table 3.4 Summary Statistics of the Benchmark Factor Portfolios

Factor Portfolios	Means	Std. Dev.	t-stat for mean = 0	Cross-Correlations Rm-Rf	SMB	HML	PR1YR
Rm-Rf	0.228	5.149	0.817	1.000	-0.270	0.009	-0.101
SMB	0.518	3.469	2.756	-0.270	1.000	0.135	-0.088
HML	0.133	2.852	0.860	0.009	0.135	1.000	-0.296
PR1YR	0.122	4.147	0.543	-0.101	-0.088	-0.296	1.000

Figure 3.11 and Figure 3.12 chart the time series of the benchmark factor portfolios. Figure 3.11 charts the series in levels. Figure 3.12 charts the benchmark portfolios in index form. (This is provided as it may illustrate a clearer picture of the relative trends in the variables over time). Figure 3.12 indicates that small stocks have generally outperformed large stocks over the period but differences between value and growth stocks and between past winners and losers is much less discernable. Figure 3.13 charts the time series of the conditioning variables.

Figure 3.11 Time Series Plots of the Benchmark Portfolios (Levels).

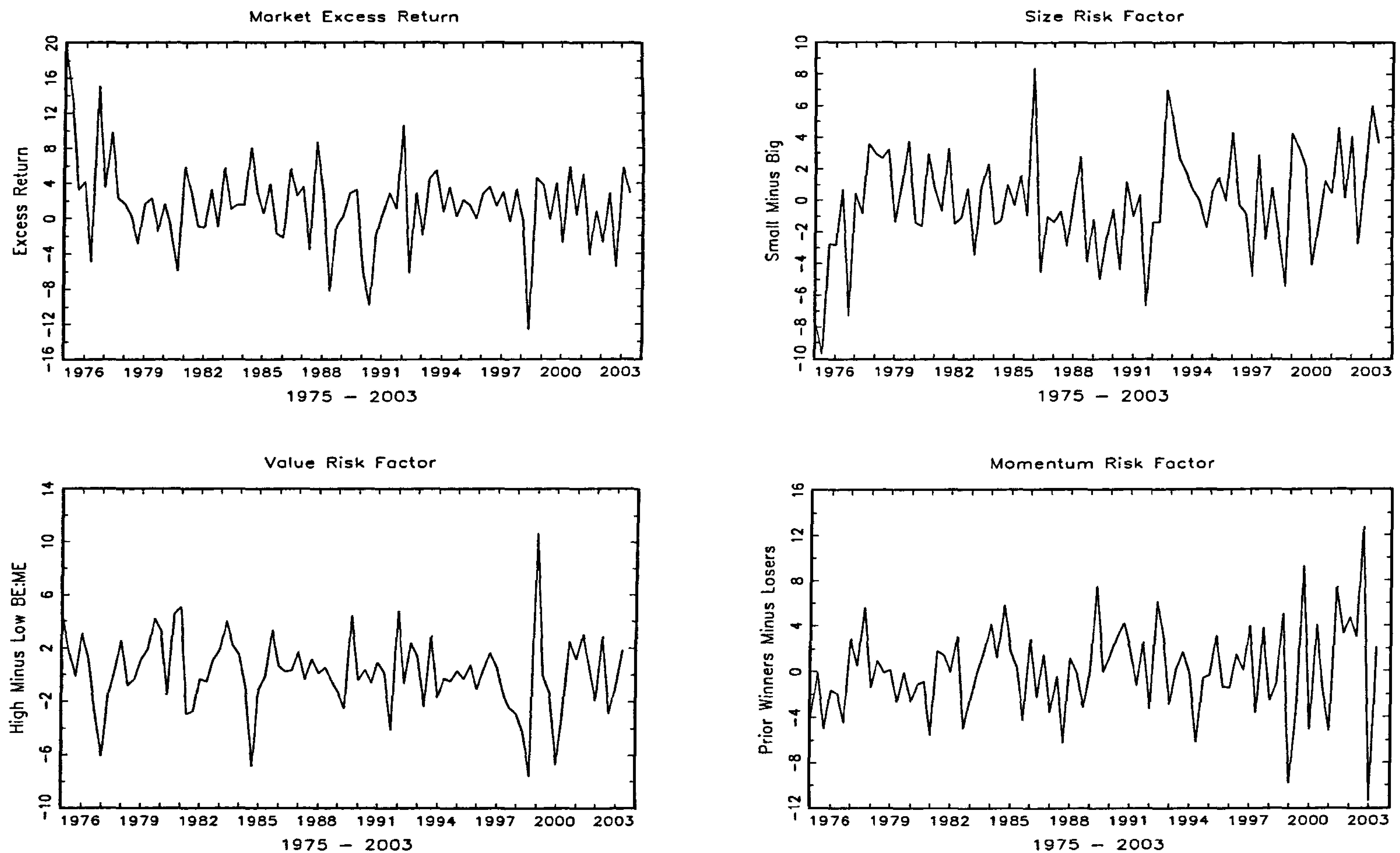


Figure 3.12 Time Series Plots of the Benchmark Portfolios. Index Base 1975 = 100

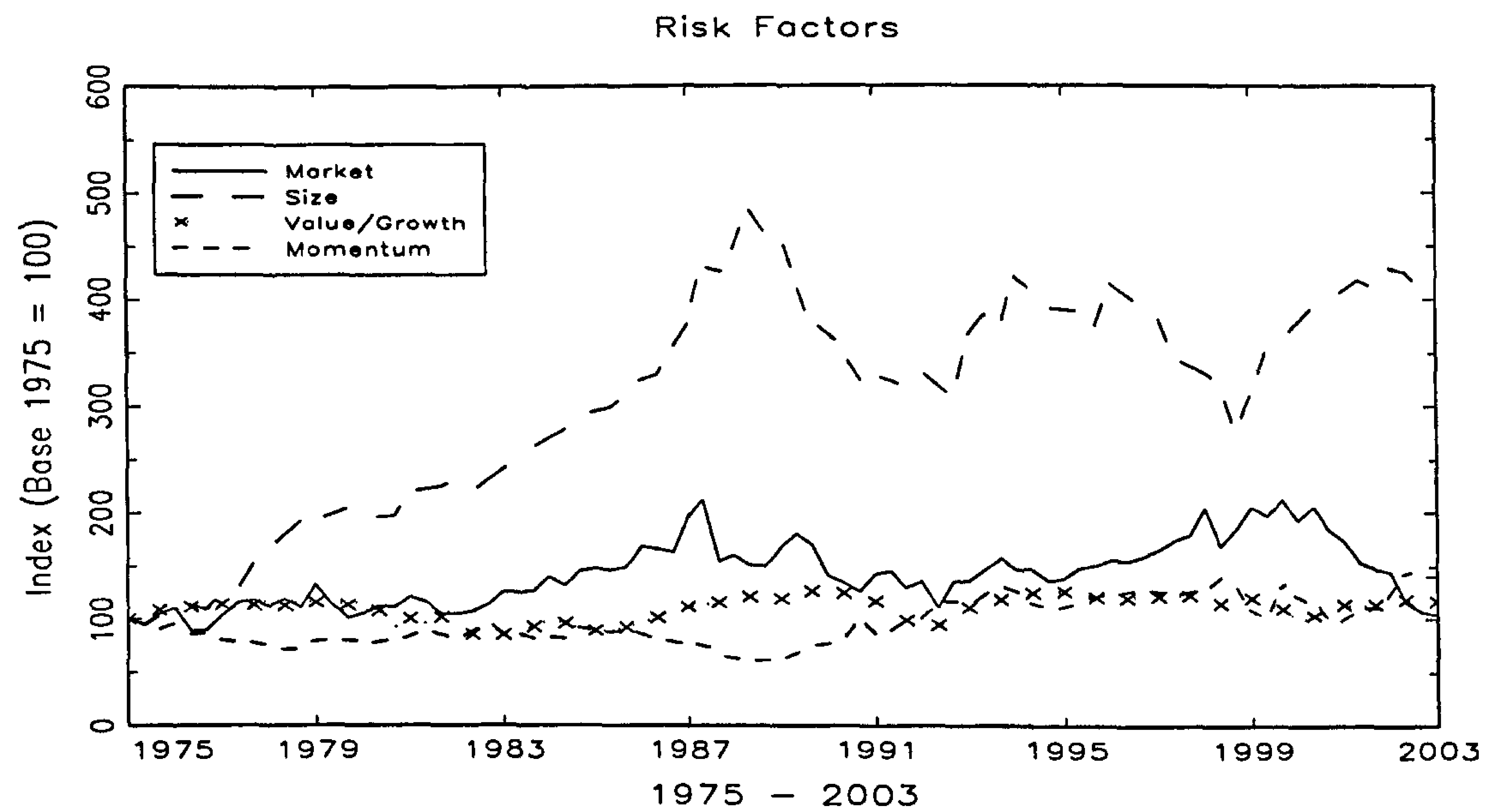
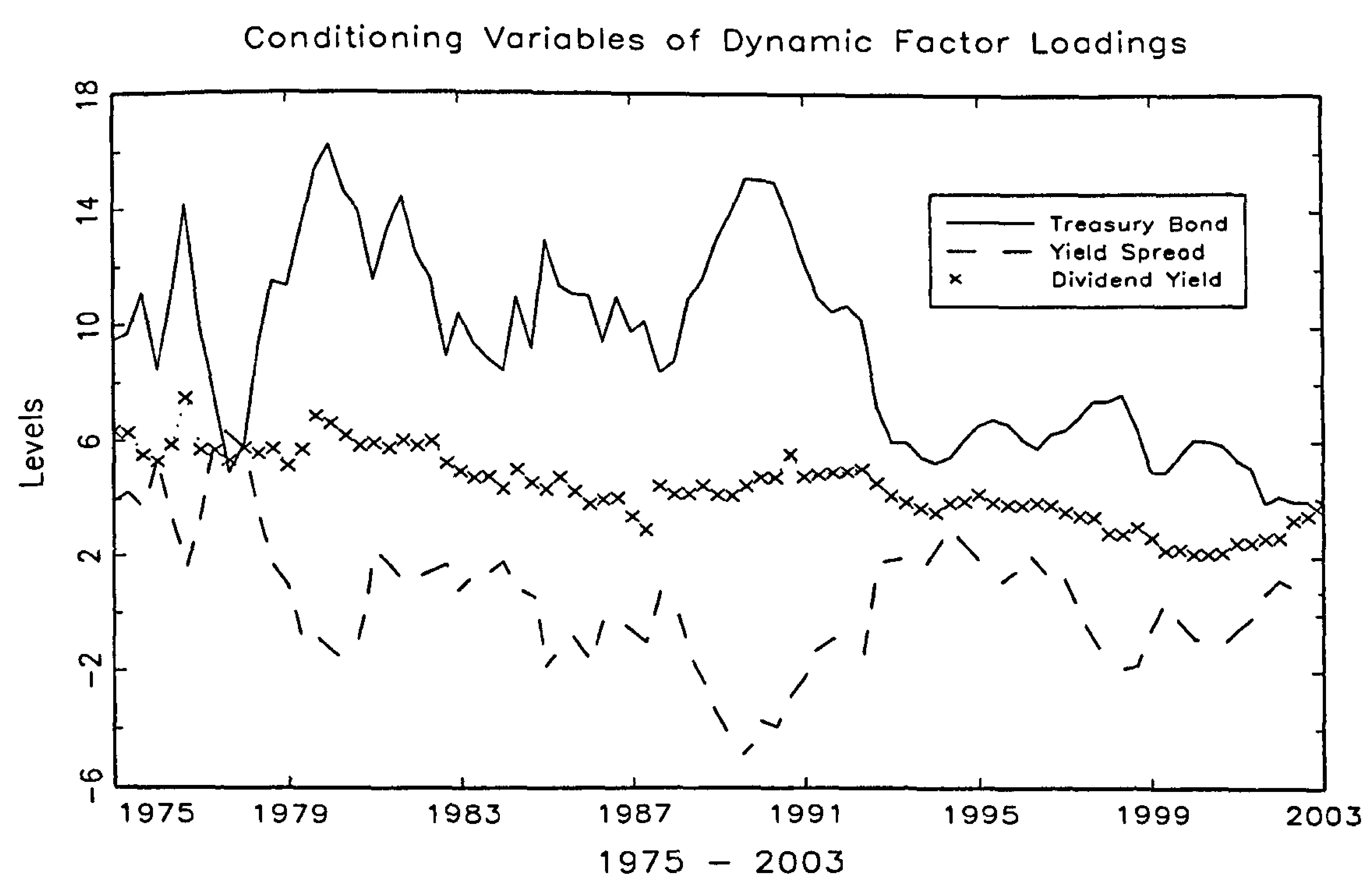


Figure 3.13 Time Series Plots of the Conditional Variables



3.7 The London Share Price Database

The London Share Price Database (LSPD) is operated by the Institute of Finance and Accounting at London Business School. The LSPD consists of a vast amount of company specific information on over 7,000 stocks with information dating from 1955. At the time of writing the database covers the period to the end of 2003. Briefly, it includes data on (i) general descriptives (number of capital changes, number of dividend payments, number of units of share types etc), (ii) capital change (share issues, repayments, types of share in issue etc), (iii) dividends (types of dividends payments, ex-dividend dates, tax credit details, announcement dates etc), (iv) prices (highs, lows, transaction prices, number of transactions etc), (v) returns (monthly returns, P/E ratios, stock exchange industrial classifications, trading velocity etc), (vi) industry indices and (vii) an archive (stock index constituents through time and other variables). It is not the intention to describe the LSPD in full here as much of its content is not directly relevant to this study. Instead, the relevant sections used in this study are described.

In this study, three LSPD files are used: the General Descriptive, Returns and Archive files. These are used in the construction of the momentum risk factor. The returns file contains monthly returns, *inter alia*, on all stocks which existed at any time since 1955. Each stock is uniquely identified by an ID number[4]. The General Descriptive file contains information on which stock indices the individual stocks were constituents of at the end of 2003. This information is coded as a binary field and it is possible for the user to indirectly determine (by a separate calculation) whether a stock was a constituent of a particular index, such as FTSE All Share index, FTSE 100 index etc, as of end – 2003. However, the Archive file contains similar index constituent information historically. That is, it is possible to ascertain from the Archive file whether a given stock was a constituent of a given index at a given point in time historically – not just at the end of 2003. However, the Archive file begins in 1979, not 1955. Furthermore, in the case of a number of stocks at a number of points in time information exists in the returns file but not in the Archive file and *vice-versa*.

[4] If a company is merged/acquired its return record ceases in the database at that point. The returns of the acquiring firm remain in the database from that point unless it's a foreign acquirer in which case the series is terminated.

In this study, a separate programme was written in *Gauss* in order to match each stock's return in the returns file with its index constituency in the Archive file for each month for over 7,000 stocks. In the LSPD the time series of returns and index constituencies on stocks are stacked in columns, ie the time series for each stock are stacked and listed in a single column. In this study, the data were reorganized in matrices where in a matrix of stock returns cell [t,i] is the return at time t on stock i, t = 1,2 ...333, i = 1 ,2 ...> 7000. A corresponding matrix of index constituent information was constructed where cell [t,i] = 1 if at time t stock i was a constituent of the index (say FTSE A All Share Index, or FTSE 350) and cell [t,i] = 0 otherwise. Three separate such matrices of index constituent information were constructed for the FTSE All Share index, FTSE 350 index and the FTSE 100 index. In the case of the FTSE 100 index the data cover the period from 1986.

In a separate programme these matrices then enable the user to limit the calculation of the momentum variable to constituent stocks of the FTSE A All Share Index or the FTSE 350 Index etc.

CHAPTER 4

PERFORMANCE MEASUREMENT AND MODEL SELECTION

4.1 Performance Measurement

Evaluating and comparing mutual fund performance based on raw returns fails to take account of the levels of risk borne by funds: high raw returns may be associated with higher risk. For this reason it is necessary to evaluate fund performance based on a risk adjusted or abnormal return measure. This performance measure depends on the asset pricing model chosen to explain the cross-section of expected returns. This chapter first describes the most common performance models to appear in the literature and the models that will be applied in this study. These models fall into three classes: unconditional, conditional beta and conditional alpha-beta models. Then all models are estimated for the entire set of UK equity mutual funds and estimation diagnostics are used to select one 'best-fit' model within each of the three classes. These selected models are used in the computationally intensive bootstrap procedure in chapter 6.

4.1.1 Jensen's Alpha

The Jensen (1968) measure represents abnormal performance based on a single risk factor model, ie the Capital Asset Pricing Model (CAPM) specification

$$(4.1) \qquad (R_i-R_f)_t = \alpha_i + \beta_i(R_m-R_f)_t + \varepsilon_{it}$$

where R_{it} is the expected return on fund i in period t, R_{mt} is the expected return on a market factor mimicking portfolio, R_f is a risk free rate, typically proxied in empirical work by the return on a treasury bill. If the CAPM is the correct model of equilibrium returns then the portfolio should lie on the Security Market Line and the value of alpha should be zero. Therefore, a positive and statistically significant value of alpha is hypothesised to indicate superior risk adjusted performance or stock picking skills (selectivity) on the part of the fund manager. That is, a positive alpha indicates that the portfolio has performed better than a random selection buy-and-hold strategy. Alpha may be estimated empirically from least squares regression of (4.1). Similarly, a statistically significant negative value of alpha is taken to indicate inferior risk adjusted performance.

4.1.2 Carhart's Alpha

The Carhart (1997) measure is the alpha estimate from a four-factor model which is an extension of (4.1) and controls for fund exposure to size risk, value risk as well as momentum strategies in modeling expected fund returns:

$$(4.2) \qquad (R_i\text{-}R_f)_t = \alpha_i + \beta_{1i}(R_m\text{-}R_f)_t + \beta_{2i}(SMB_t) + \beta_{3i}(HML_t) + \beta_{4i}(PR1YR_t) + \varepsilon_{it}$$

where SMB_t, HML_t and $PR1YR_t$ are risk factor mimicking portfolios for size, value versus growth and one-year momentum effects respectively in the stock holdings of the mutual funds. Carhart's alpha may be estimated empirically as the intercept term in (4.2).

The four-factor model is largely based on the empirical findings of Fama and French (1992 and 1993) and Carhart (1995). Fama and French (1992 and 1993) find that a three-factor model including market, size and book-to-market value risk factors provides significantly greater power than the CAPM alone in explaining common variation in stock returns. Fama and French (1992) report a strong negative relationship between stock returns and size: smaller firms tend to have higher average returns (the authors report a spread of 0.74% per month on average based on their size rankings). The size factor, SMB ('small minus big'), is a measure of the difference between the returns on small versus big stocks so β_{2i} measures a fund's sensitivity to the relative performance of small stocks[1]. The hypotheses underpinning the specification of a size risk factor are as follows: The earnings prospects of small firms may be more sensitive to economic conditions and small firms may be less likely to survive a period of financial distress during economic downturns. There is also the concern that small firms embody greater information asymmetry for investors than large firms. Finally, the ownership of small firm stock is generally less diluted making the stock less liquid and consequently subject to higher volatility arising from thin trading. All these factors imply a risk loading for size and a higher required return by investors.

[1] The calculation of SMB and the other risk factors in (4.2) in this study is described in chapter 3.

Fama and French (1992) also report a strong positive relationship between stock returns and the book-to-market value ratio: stocks with relatively high book-to-market ratios have higher average returns (the authors report a spread of 1.5% per month between the highest and lowest book-to market stocks in their study). The book-to-market value factor, HML ('high minus low'), is a measure of the difference between the returns on high versus low book-to-market stocks. As Fama and French outline, if stock prices are rational the book-to-market value ratio should reflect firms' relative prospects. A high book-to-market ratio firm indicates relatively low earnings on assets. Consequently, there is a high book-to-market or 'value' premium. Alternatively, if stock prices are irrational the cross-section of book-to-market ratios may be the result of market overreaction to the relative prospects of firms. High (low) book-to-market ratios represent firms whose prices have 'overshot' on the downside (upside) and therefore the ratio predicts the cross-section of stock returns.

Alternative explanations for the outperformance of value stocks, particularly in the U.S., have appeared in the literature. Chan and Lakonishok (2004) suggest that the value premium arises because investors extrapolate past information on company performance too far into the future. For example, value stocks tend to have a relatively poor history with respect to growth in earnings, cashflow and sales. Notwithstanding the lack of evidence of persistence in such relative growth rates, the authors maintain that analysts project past growth into the future leading to an under (over)–pricing of value (growth) stocks relative to their fundamentals. This misspricing generates higher returns for value stocks. (See also Chan et al 2000). LaPorta et al (1997) attribute the value premium to earnings surprises, which the authors find are systematically more positive for value stocks. These alternative hypotheses are based more on a market inefficiency rather than risk-based explanation of the value-growth return differential.

The fourth risk factor, PR1YR, in (4.2) is an additional factor capturing Jegadeesh and Titman's (1993) one year momentum anomaly. The PR1YR variable is the difference in returns in period t between a portfolio of past high performing stocks and past poor performing stocks. Its specification in (4.2) captures a fund's sensitivity to following a zero-investment strategy of investing in past strong performing 'momentum' stocks and short-selling stocks with low past returns. Carhart's main motivation for examining momentum effects is due to the inability of the Fama and French three-factor model to explain cross-sectional variation in ranked portfolio returns. The momentum factor is statistically well determined where Carhart finds that it explains almost half of the spread in returns between the top and bottom decile portfolios of funds ranked by raw return and therefore appears to mimic a risk factor which accounts for variation in fund returns. Chan, Jegadeesh and Lakonishok (1996) suggest that the momentum anomaly is a market inefficiency caused by slow reaction to information. If the intercept (alpha) is to represent skill, the performance model should control for momentum strategies which, even if they do contribute to higher returns, do not represent skill as such strategies are easily/mechanically implemented.

Carhart's four-factor model in (4.2) may be interpreted as a performance attribution model where the coefficients and premia on the risk factors indicate the proportion of mean returns attributable to four investment strategies: high versus low beta stocks, small versus large capitalisation stocks, value versus growth stocks and one-year momentum versus contrarian stocks.

4.1.3 Conditional Performance Measures

The Jensen and Carhart measures described above are unconditional measures of performance: fund alphas are calculated as the past average excess return minus a *fixed* factor loading(s) times the average excess return on a benchmark portfolio(s). However, unconditional performance measures do not incorporate changing market information about the expected returns and risk of individual securities. As an example of changing market information, Chan (1988) and Ball and Kothari (1989) highlight, using US data, that as the market corrects for the under-pricing (over-pricing) of 'loser' ('winner') shares a significant shift in the Beta of these shares can occur. Also a number of studies have shown that the risk of a share can change through time as the financial characteristics of the company change, ie gearing, earnings variability and dividend policy (Foster (1986), Mandelker and Rhee (1984), Hochman (1983), Bildersee (1975)). Therefore, even if the manager follows a buy-and-hold investment strategy, the risk of the portfolio may vary over time in line with the changing risk of the underlying securities. In addition, the weights in a passive buy-and-hold strategy will vary in line with the relative values of the underlying assets. Finally, in actively managed funds, the manager will manipulate portfolio weights and consequently the portfolio beta. These points taken together indicate that there may well be time variation in the portfolio beta(s).

Similarly, suppose as in Merton (1980), that a fund manager believes that expected market excess return and its volatility move together proportionately over time with economic conditions. Based on economic conditions a fund manager wishing to keep the fund volatility constant will lower the fund beta when market conditions are volatile and *vice-versa.* Because, as a result, the fund beta will be negatively correlated with the market premium, the average excess return of the fund will be less than the average beta of the fund applied to the average market premium. In this case the use of an unconditional beta would lead us to conclude that the fund has a negative alpha. In this example, this does not necessarily reflect poor stock-picking ability but the fact that in order to maintain constant volatility the fund reduces its risk when the premium for risk is high and *vice versa.* (see also Ferson and Schadt, 1996).

4.1.3.1 Conditional Beta Models

Ferson and Schadt (1996) extend the CAPM specification to a conditional performance measurement model by allowing the factor loading on the market risk factor at time *t* to be linearly related to a vector of instruments for the economic information set Z_t as follows

$$(4.3) \qquad \beta_{it} = b_{0i} + B'_i(z_t)$$

where z_t is the vector of deviations of Z_t from unconditional means. Therefore, b_{0i} is the unconditional mean of the conditional beta. Subbing (4.3) into (4.1) and generalising the notation to let $r_{b,t+1}$ denote the expected excess return on a benchmark portfolio (market portfolio in this case) the expected excess portfolio return in the conditional beta CAPM can be written as

$$(4.4) \qquad r_{i,t+1} = \alpha_i + b_{0i}(r_{b,t+1}) + B'_i(z_t * r_{b,t+1}) + \varepsilon_{i,t+1}$$

where $r_{i,t+1}$ is the expected excess return on fund *i*. As before under the null hypothesis of zero abnormal performance $\alpha_i = 0$. Note, z_t is in deviations from mean so that the mean value of this variable is zero. Therefore, its impact on α_i arises due to the time varying nature of existing risk factor loadings. To specify Z_t in levels would be to specify this variable as an additional risk factor. The model in (4.4) can be extended to the Carhart four-factor model where the additional factor loadings are each modeled as conditional betas and as linear functions of an economic information set Z_t. For *L* instruments in Z_t the conditional four-factor model involves estimating (L+1)4 + 1 parameters. This Ferson and Schadt performance measure computes the alpha of a managed portfolio controlling for investment strategies that use publicly available economic information, which it is hypothesized predicts factor returns, to dynamically adjust the portfolio's risk factor sensitivities.

4.1.3.2 Conditional Alpha-Beta Models

The model in (4.4) specifies the abnormal performance measure, α_i, as a constant. However, it may be the case that abnormal returns are also time varying. Christopherson, Ferson and Glassman (1998) extend the analysis of Ferson and Schadt (1996) to estimate conditional alphas as well as betas. They also assume a linear specification for the conditional alpha as a function of the instruments in Z_t as

(4.5) $$\alpha_i = \alpha_{0i} + A'_i(z_t)$$

Using (4.5) to modify (4.4) yields

(4.6) $$r_{i,t+1} = \alpha_{0i} + A'_i(z_t) + b_{0i}(r_{b,t+1}) + B'_i(z_t{*}r_{b,t+1}) + \varepsilon_{i,t+1}$$

The conditional alpha approach can also be applied to the Carhart four-factor model and is a simple extension of the four-factor model with conditional betas described above.

To further examine how conditional alpha and beta models arise, assume a general linear factor model of the form

(4.7) $$R_{it+1} = \alpha_i + \beta_i'(F_{t+1}) + \varepsilon_{it+1}$$

where F_{t+1} represents a vector of the expected values of risk factors and R_{it+1} is the expected excess return on *asset i.* A mutual fund expected excess return is then given by

(4.8) $$r_{i,t+1} = \sum_{i}^{N} W_{it+1} * R_{it+1},$$ for N assets in the fund

where W_{it+1} is time varying and is given by

(4.9) $$W_{it+1} = W_{i0} + W_i'(z_t)$$

where W_{i0}, W_i are constants. For example, W_{i0} may represent long run strategic asset allocation weights while W_i represents stock picking (or market timing) based on known information at time *t*.

Subbing (4.9) and (4.7) in (4.8) yields a model of the form in (4.6), which may have heteroscedastic errors.

The performance measure from this conditional alpha-beta model is the alpha of a managed portfolio, controlling for investment strategies that use publicly available economic information to (i) add stocks with abnormally high expected excess returns conditional on the information and (ii) dynamically adjust the portfolio risk factor sensitivities conditional on the information.

Ferson and Schadt (1996) and Christopherson, Ferson and Glassman (1998) use instruments for economic information, Z_t, that previous studies have shown are useful for predicting security returns and risk over time. These include: the lagged level of the one-month TBill yield, the lagged dividend yield of the market factor, a lagged measure of the slope of the term structure, a lagged quality spread in the corporate bond market and a dummy variable to capture the January effect.[2] Of course, all conditional models may also be examined by applying subsets of the information set, Z_t. The findings from such tests in this study are reported later in this chapter.

4.1.4 Models of Market Timing

In addition to stock selection skills, models of portfolio performance should also examine a fund's market timing skill. This is, can fund managers successfully predict the future direction of the market in aggregate and increase or decrease the portfolio sensitivity (Beta) accordingly? Treynor and Mazuy (1966) and Henriksson and Merton (1981) are two commonly applied market timing models in the literature while Ferson and Schadt (1996) also estimate conditional versions of both these models.

---

[2] A more detailed description of the conditioning variables as adopted in this study is provided in chapter 3.

4.1.4.1 The Treynor-Mazuy Model

The Treynor and Mazuy (1966) models is a quadratic extension of the single factor CAPM in (4.1), see also Admati et al 1986. The model permits β_i in (4.1) at time t to be expressed as a linear function of the expected future market excess return:

$$\beta_{it} = \theta_i + \gamma_{iu}[r_{m,t+1}] \tag{4.10}$$

Replacing β_i in (4.1) with (4.10) yields a quadratic of the form

$$r_{i,t+1} = \alpha_i + \theta_i(r_{m,t+1}) + \gamma_{iu}[r_{m,t+1}]^2 + \varepsilon_{i,t+1} \tag{4.11}$$

where $r_{i,t+1}$ and $r_{m,t+1}$ measure expected excess returns over the risk free rate. γ_{iu} is the unconditional measure of market timing ability. The quadratic specification in (4.11) implies that during a up (down) market the market timer has a higher (lower) than normal fund Beta and the fund performs better than it would otherwise.

Ferson and Schadt (1996) conditionalise the Treynor and Mazuy (1966) model by specifying β_i in (4.1) at time t as a linear function of both the expected future market excess return and the public information set (in deviations), z_t. Substituting for β_i in (4.1) with this linear response function yields a model of the form

$$r_{i,t+1} = \alpha_i + \theta_i(r_{m,t+1}) + C_i'(z_t{*}r_{m,t+1}) + \gamma_{ic}[r_{m,t+1}]^2 + \varepsilon_{i,t+1} \tag{4.12}$$

where the coefficient γ_{ic} measures the sensitivity of the manager's Beta to a private market timing signal. The term $C_i'(z_t{*}r_{mt+1})$ in (4.12) controls for the public information effect, ie it captures the part of the quadratic term in (4.11) which is attributable to public information variables, Z_t. Therefore, in the class of conditional model in (4.12) the correlation between fund betas and future market excess returns which is attributable to public information variables is not considered to reflect market timing ability.

4.1.4.2 The Henriksson and Merton Model

Henriksson and Merton (1981) describe a similar model of market timing. In this model, fund managers forecast whether the future market excess return will be positive or negative. In this binary response function, a positive (negative) forecast causes the manager to target a higher (lower) beta. From (4.1), β_i may be expressed as a (linear) function of a constant plus a dummy variable which takes a value of one (zero) corresponding to a positive (negative) market forecast. Subbing such a linear response function in place of βi in (4.1) yields a model of the form

$$(4.13) \qquad r_{i,t+1} = \alpha_i + \theta_i(r_{m,t+1}) + \gamma_{iu}[r_{m,t+1}]^+ + \varepsilon_{i,t+1}$$

where $[r_{mt+1}]^+$ is defined as $\max(0, r_{mt+1})$. γ_{iu} is the unconditional measure of market timing ability. (Henriksson and Merton (1981) interpret $\max(0, r_{mt+1})$ as the payoff to an option on the market portfolio with a strike price equal to the risk free rate).

To extend the Henriksson-Merton model to a conditional setting, suppose β_i in (4.1) is written as

$$(4.14) \qquad \beta_{it} = b_d + \gamma_{ic}D + (B'_d + \Delta'D)^*z_t$$

where D = a dummy variable which equals one for a positive forecast of the future market excess return and equals zero otherwise. The specification in (4.14) is equivalent to the following: if the forecast is positive the manager selects $\beta_{up} = b_{up} + B'_{up}{}^*z_t$ while if the forecast is negative the manager selects $\beta_d = b_d + B'_d{}^*z_t$, where forecasts are made conditional on z_t. Subbing (4.14) in place of β_i in (4.1) yields

$$(4.15) \qquad r_{i,t+1} = \alpha_i + b_d(r_{m,t+1}) + B'_d[z_t{}^*r_{m,t+1}] + \gamma_{ic}[r_{m,t+1}]^+ + \Delta'[z_t{}^*(r_{m,t+1})^+] + \varepsilon_{i,t+1}$$

where $\gamma_{ic} = b_{up} - b_d$, $\Delta = B_{up} - B_d$. The null hypothesis of no market timing ability implies that γ_{ic} and Δ are zero. The null hypothesis of no selectivity implies $\alpha_i = 0$.

4.2 Model Selection

In testing the abnormal performance hypothesis a researcher faces the joint hypothesis problem of whether the performance model is the 'true' model of equilibrium security returns. In this section, many variants of the performance models described above are estimated for the data set of UK mutual funds. In all, over 50 models were estimated. Each model is estimated for each individual fund. For each model, cross-sectional (across funds) average statistics are presented. It is useful to examine the level and distribution of alpha, normality and serial correlation characteristics, the significance of factor loadings and model selection diagnostics. Based on these statistics a single best-fit model from within each of the classes of (i) unconditional models, (ii) conditional beta models and (iii) conditional alpha-beta models is selected for the later bootstrap analysis in chapter 6. One model is selected from each class because the bootstrap methodology is computationally intensive.

The model estimation results are reported in Table 4.1 where, in the interests of parsimony, a summary selection of findings is presented. Results relate to the full cross-section of UK equity mutual funds over the period April 1975 – December 2002, ie over all equity sectors including equity income, general equity and small stock funds. Results are based on funds with a minimum of 36 observations in order to ensure a high degree of precision in the estimation of alpha and other statistics. (This issue is discussed in greater detail in chapter 5 on bootstrap methodology).

4.2.1 Unconditional Models of Performance

Table 4.1 Panel A shows the estimation results of the unconditional models including the CAPM (model 1), Fama and French three-factor (model 2) and Carhart (model 3) along with the unconditional market timing models of Treynor-Mazuy (model 4) and Henriksson-Merton (model 5). The CAPM indicates that the cross-sectional average alpha was 0.001% per month (0.012% annually) indicating that the average mutual fund manager outperformed the market by this amount. However, this abnormal performance is not statistically significant at 5%. t-statistics presented are averages of absolute values as otherwise average t-statistics may centre on zero. In addition, all t-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors. The Fama and French and Carhart multi-factor models produce broadly similar performance findings where the cross-sectional average alpha in each case is negative at around –0.05% per month (0.6% annually) but is not statistically significant on average. A lower value of alpha may be expected as fund performance is adjusted for additional risk factors. In results not shown, the three-factor model indicates that around 3% (16%) of funds yield a statistically significant positive (negative) value of alpha by a conventional t-test. The finding of negative abnormal performance (on average) is consistent with the findings of Blake and Timmermann (1998) who report an even lower average alpha. The difference in findings may lie in the minimum 36 observations restriction imposed here which may remove some short-lived poor performing funds. This explanation is supported by looking at the alpha of a single equal weighted portfolio of all funds (see Table 4.1). In forming the equal weighted portfolio the minimum 36 observation restriction was dropped for comparison. The equal weighted portfolio alpha is -0.072% in the FF 3 factor model. Had the restriction been maintained, the average alpha across funds and the equal weighted portfolio alpha would be equal.

The cross sectional average alpha is positive when a market timing factor is specified (models 4,5). However, alpha remains statistically insignificant on average.

In terms of the factor loadings, the t-statistics across all unconditional models consistently show the market and size risk factors as statistically significant determinants of the cross-sectional variation in fund returns. In model 2, 100% and 74% of funds indicated a statistically significant t-statistic on these factors respectively. The value risk factor, with an average t-statistic of 1.472 is not significant: only 23% of the funds have a significant t-statistic on this risk factor. The one-year momentum factor from the Carhart specification (model 3) is also insignificant on average – significant for only 33% of funds.

In chapter 3, section 3.5.4, the construction of the momentum factor was described in detail where alternative definitions were discussed based on different stock indices and various ranking and holding periods. Here, the alternative definitions were estimated. Momentum based the FTSE A All Share Index using an 11 month ranking period and 1 month holding period was found to be the best determined. This is the same definition as used by Carhart (1997). The results presented in Table 4.1 are based on this momentum measure.

In terms of the market timing models (4 and 5), the evidence from Panel A indicates that the average fund manager (in fact more than 74% of managers) did not possess market timing ability over the period. In none of the models tested are the Treynor-Mazuy (1966) and Merton-Henriksson (1981) market timing variables significant.

4.2.2 Conditional Beta Models of Performance

Table 4.1 Panel B presents the estimation results of the conditional beta models. Following Ferson and Schadt (1996) and Kosowski et al (2004), the public economic information variables used to model conditional betas are (i) the yield on a UK one-month Tbill, (ii) the slope of the term structure defined as the yield on the UK 10 year gilt minus the yield on the one-month Tbill and (iii) the dividend yield on the FT A All Share index.

For example, model 6 is the Ferson and Schadt (1996) model with the market factor loading 'conditioned' on the full set of public information variables. Model 7 and model 8 are Fama and French three-factor specifications where the market factor loading is conditional on the full set of, and a subset of, the public economic information variables respectively as indicated. Model 11 is also a Fama and French based factor model but in this case all factor loadings are specified as conditional on the dividend yield. In results not shown, numerous alternative conditional models were estimated. In particular, these included hypotheses that the loadings on the market and size factors may be dynamically adjusted by fund managers based on the market dividend yield and/or the term spread.

The findings in relation to alpha and the distribution of alpha among the conditional beta models are very similar to those in Panel A for the unconditional models. According to all conditional beta Fama and French and Carhart models the average mutual fund manager underperformed. Although, from all models alpha is not significantly different from zero. Again, in results not shown, only about 20% of funds yield a statistically significant value of alpha at the 5% significance level by a standard t-test. Results in relation to an equal weighted portfolio of all funds applied to the multi-factor Fama and French and Carhart models show the value of alpha falling. Again, this is in line with findings from the unconditional models previously.

Once again, the conclusions regarding the significance of the factor loadings are very similar to those reached with the unconditional factor models, ie the market factor and the size factor are consistently statistically significant at 5% across all conditional beta models while the value and momentum factors are not, although momentum is significant at 10%. It is also noteworthy that among conditional beta models, the public economic information instruments are unanimously insignificant for the average fund at the 5% significance level and were generally insignificant for more than 75% of the funds.

4.2.3 Conditional Alpha-Beta Models of Performance

Table 4.1 Panel C describes the estimation results of the conditional alpha-beta models. This class of model reveals a more mixed picture where average alpha varies between positive and negative across models. However, in this class of models, the alpha performance measure is still not found to be significantly different from zero. Panel C strongly demonstrates, once again, that the full set of conditioning public economic information variables are not found to be significant (on average). This was generally found to be the case for more than 90% of the funds.

The unambiguous insignificance of the conditioning variables in the conditional beta and conditional alpha-beta models provides strong evidence against conditional models as the 'true' models of equilibrium security returns. The above tests provide evidence that fund managers collectively either (i) do not dynamically adjust the risk factor loadings of the portfolio, or at least do not do so successfully or (ii) do not adjust the factor loadings in response to the set of public economic information variables examined in this study. However, the chosen set of public information instruments here is typical of that in the literature.

Notwithstanding this caveat, in this study a model from within each of the three classes of models is selected for the bootstrap analysis. This is done as a means of examining the robustness of the bootstrap findings across alternative models of equilibrium returns. However, owing to the insignificance of the conditional public information variables, bootstrap results from the conditional models should be treated with caution and perhaps given less weight than the better fit unconditional models.

4.2.4 Selecting Representative Models of Performance

The key model selection metrics are the statistical significance of the individual parameters and the Schwartz Information Criterion (SIC). The SIC trades off a reduction in a model's residual sum of squares for a parsimonious best-fit model and indicates that the model with the lowest SIC should be selected. In Table 4.1, the cross-sectional average (across funds) SIC is shown for each model.

Among unconditional models in Panel A, the three-factor Fama and French and four-factor Carhart specifications, with SIC measures of 1.328 and 1.316 respectively, provide the best fit. Although the four-factor model yields a slightly lower SIC, the momentum factor is statistically significant for only 33% of the funds. Consequently, in order to avoid a possible bias arising from misspecification, in this study the three-factor model is selected as the baseline model for the bootstrap analysis. However, as a test of robustness the bootstrap results will also be described later for a two-factor (market and size risk) and four-factor unconditional model.

Panel B indicates that model 8, the three-factor Fama and French model with the market factor loading conditioned on the market dividend yield, generates the lowest SIC value of 1.365. Finally, among the conditional alpha-beta models in Panel C, model 14 with the lowest SIC of 1.392 is suggested. However, model 14 is very similar to model 8 where alpha is also specified as conditional on the dividend yield. Given the similarity between model 8 and model 14, in the interests of presenting results from a wider range of specifications model 15 is instead selected from among the conditional alpha-beta models in Panel C. Model 15, which also has a relatively low SIC value of 1.432, is also a three-factor Fama and French model where all three factor loadings are time varying. This conditional specification hypothesizes that the fund manager dynamically modifies the portfolio's sensitivity to the three risk factors based on a signal provided by the market dividend yield.

Therefore, in this study model 2, model 8 and model 15 are selected as representative models from within each of the three classes of models above for the bootstrap analysis to follow in chapter 6.

Figures 4.1, 4.2 and 4.3 show histograms of the cross-sectional distributions of alpha from models 2, 8 and 15 respectively. All models selected indicate a negative average alpha. However, of key importance for this study (and for investors) is the relatively large cross-sectional standard deviations of the alpha estimates which is around 0.26% p.m. (3.1% p.a.), for the unconditional and conditional-beta models and somewhat larger at 0.75% p.m. for the conditional alpha-beta model. (Refer to Panels A, B & C). This implies that the extreme tails of the distribution of abnormal performance may contain a substantial number of funds. This is important since investors are more interested in holding funds in the right tail of the performance distribution and avoiding those in the extreme left tail, than they are in the average fund's performance.

4.2.5 Non-normality and Serial Correlation

Table 4.1 also reports the percentage of funds which reject the null hypotheses of (i) normality and (ii) serial correlation in the regression residuals at 5% by a Bera-Jarque test and LM test up to 6 lags respectively. In the case of all models the normality assumption is rejected for around 64% of the mutual funds. It is this finding which largely motivates the use of the bootstrap technique as non-normal residuals suggests that the alpha estimates themselves are also non-normally distributed which in turn invalidates the use of standard t-tests and F-tests and questions the reliability of past research which draws inferences based on these tests.

The LM test statistics in each panel suggest that in the case of all models a sizeable proportion (around 42%) of mutual fund exhibit serial correlation of order one. This has implications for the bootstrap methodology. Firstly, it is important to examine a modification of the bootstrap procedure to preserve the information content in the serial correlation in order that the bootstrap simulations mimic the original fund return generating process as closely as possible. Secondly, the use of Newey-West autocorrelation adjusted t-statistics, as in this study, should incorporate the correct order of serial correlation.

This concludes the description of the model selection process. Chapter 5 describes the bootstrap methodology.

Table 4.1. Model Selection: Cross-Sectional Results of Model Estimations.

Table 4.1 presents results from the estimation of the performance models using all mutual funds. Panel A relates to unconditional models, Panel B relates to conditional Beta models while Panel C relates to conditional Alpha-Beta models. T-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors. (t-statistics are cross sectional averages of the absolute value t-statistics). Also shown are statistics on the percentage of funds which (i) reject normality among the residuals by a Jarque-Bera test at 5% and (ii) reject a null hypothesis of no serial correlation among residuals at lags 1 and 6 by a LM test at 5%. The Schwartz Information Criterion is also presented. Also shown is alpha and its t-statistic for an equal weighted portfolio of all mutual funds. All figures shown are cross-sectional averages. Funds with at least 36 observations are used.

Panel A: Unconditional Factor Models						
		1	2	3	4	5
Model		**CAPM**	**FF**	**Carhart**	**TM**	**MII**
Regression Coefficients						
Average Alpha (percent per month)		0.001	-0.050	-0.054	0.128	0.273
t-statistic		1.033	1.214	1.288	1.201	1.268
Standard Deviation of Alpha		0.289	0.261	0.268	0.400	0.528
Unconditional Betas						
(t-statistics in parentheses)	r_{mt}	0.886 (22.229)	0.914 (25.396)	0.916 (25.847)	0.870 (21.814)	0.954 (15.353)
	SMB		0.284 (4.887)	0.283 (5.043)		
	HML		-0.024 (1.472)	-0.021 (1.329)		
	PR1YR			0.008 (1.591)		
Market Timing Measures						
Treynor-Mazuy					-0.006 (1.453)	
Henriksson-Merton						-0.159 (1.245)
Model Selection Criteria						
Adjusted R-square		0.728	0.811	0.819	0.734	0.732
SIC		1.560	1.328	1.316	1.577	1.581
Rejection of Normality (% of funds)		64	64	59	64	64
LM(1) statistic (% of Funds)		44	39	40	43	43
LM(6) statistic (% of Funds)		34	33	36	36	35
Equal Weighted Portfolio						
Alpha		0.028	-0.072	-0.080	0.167	0.355
t-statistic		0.391	-1.661	-1.793	2.213	3.392

Table 4.1 continued, Panel B: Conditional Beta Models							
		6	7	8	9	10	11
Model		FS	FF(1)	FF(2)	Carhart	TM	FF(3)
Regression Coefficients							
Average Alpha (percent per month)		0.062	-0.022	-0.031	-0.014	0.147	-0.035
t-statistic		1.074	1.190	1.174	1.230	1.226	1.205
Standard Deviation of Alpha		0.333	0.281	0.261	0.282	0.405	0.267
Unconditional Betas							
(t-statistics in parentheses)	r_{mt}	0.374 (15.649)	0.855 (17.163)	0.863 (21.193)	0.844 (16.438)	0.401 (15.510)	0.855 (21.097)
	SMB		0.286 (4.935)	0.285 (4.905)	0.280 (4.982)		0.267 (4.038)
	HML		-0.023 (1.449)	-0.023 (1.451)	-0.020 (1.379)		0.010 (1.224)
	PR1YR				0.0007 (1.693)		
Market Timing Measures							
Treynor-Mazuy	r_{mt}^2					-0.004 (1.156)	
Conditioning Variables, Z_{t-1}							
$z1_{t-1} * r_{mt}$, **Z1: One month rate**		-0.782 (1.225)	0.078 (1.279)		0.047 (1.365)	-0.802 (1.139)	
$z2_{t-1} * r_{mt}$, **Z2: Term Spread**		-0.351 (1.146)	0.155 (1.220)		0.214 (1.262)	-0.450 (1.133)	
$z3_{t-1} * r_{mt}$, **Z3: Dividend Yield**		-0.168 (1.337)	-0.067 (1.290)	-0.048 (1.408)	-0.070 (1.294)	-0.137 (1.112)	-0.052 (1.476)
$z3_{t-1}$*SMB_t							0.001 (1.482)
$z3_{t-1}$*HML_t							0.030 (1.051)
Model Selection Criteria							
Adjusted R-square		0.738	0.816	0.812	0.829	0.741	0.816
Schwartz Information Criterion (SIC)		1.619	1.410	1.365	1.371	1.642	1.406
Rejection of Normality (% of funds)		64	63	64	57	64	63
LM(1) statistic (% of Funds)		44	42	40	41	45	42
LM(6) statistic (% of Funds)		37	39	35	39	39	37
Equal Weighted Portfolio							
Alpha		0.069	-0.111	-0.112	-0.048	0.153	-0.111
t-statistic		0.962	-2.455	-2.464	-1.113	2.050	-2.508

Table 4.1 continued, Panel C: Conditional Alpha-Beta Models								
		12	13	14	15	16	17	18
Model		FS	FF(1)	FF(2)	FF(3)	FF(4)	FF(5)	Carhart
Regression Coefficients								
Average Alpha (percent per month)		-0.227	0.458	-0.093	-0.109	0.315	0.027	0.446
t-statistic		0.979	1.149	1.180	1.187	1.148	1.103	1.156
Standard Deviation of Alpha		4.332	4.902	0.693	0.754	2.411	0.709	4.553
Unconditional Betas								
(t-statistics in parentheses)	r_{mt}	0.365 (15.761)	0.833 (17.145)	0.863 (21.230)	0.849 (21.068)	0.997 (18.174)	0.914 (21.387)	0.836 (16.407)
	SMB		0.282 (4.878)	0.284 (4.903)	0.257 (4.043)	0.290 (5.026)	0.253 (4.086)	0.277 (4.930)
	HML		-0.022 (1.449)	-0.024 (1.449)	0.016 (1.207)	-0.024 (1.452)	0.020 (1.288)	-0.020 (1.371)
	PR1YR							-0.001 (1.685)
Conditioning Variables, Z_{t-1}								
$z1_{t-1} * r_{mt}$, **Z1: One month rate**		-0.951 (1.261)	-0.008 (1.319)			0.169 (1.335)	-0.019 (1.231)	-0.004 (1.403)
$z2_{t-1} * r_{mt}$, **Z2: Term Spread**		-0.563 (1.185)	0.058 (1.265)			0.093 (1.354)		0.070 (1.307)
$z3_{t-1} * r_{mt}$, **Z3: Dividend Yield**		-0.151 (1.333)	-0.064 (1.306)	-0.048 (1.426)	-0.055 (1.496)			-0.066 (1.302)
$z3_{t-1}*SMB_t$					-0.002 (1.513)			
$z3_{t-1}*HML_t$					0.033 (1.044)			
$z1_{t-1}*SMB_t$							-0.031 (1.529)	
$z1_{t-1}*HML_t$							0.159 (1.001	
$z1_{t-1}$		2.964 (1.267)	2.271 (1.040)			1.009 (0.975)	0.198 (0.927)	2.193 (1.049)
$z2_{t-1}$		4.669 (1.410)	2.300 (1.001)			0.957 (0.953)		2.198 (0.988)
$z3_{t-1}$		-0.819 (1.416)	0.200 (1.042)	-0.071 (1.000)	-0.073 (1.037)			-0.184 (1.099)
Model Selection Criteria								
Adjusted R-square		0.742	0.818	0.813	0.818	0.815	0.814	0.830
Schwartz Information Criterion (SIC)		1.697	1.497	1.392	1.432	1.451	1.448	1.457
Rejection of Normality (% of funds)		61	61	63	62	63	64	54
LM(1) statistic (% of Funds)		45	47	41	44	45	44	47
LM(6) statistic (% of Funds)		40	45	35	39	42	41	45
Equal Weighted Portfolio								
Alpha		0.022	-0.122	-0.107	-0.107	-0.117	-0.116	-0.060
t-statistic		0.313	-2.652	-2.466	-2.517	-2.652	-2.589	-1.252

Figure 4.1: Cross-Sectional Alpha, Unconditional Model

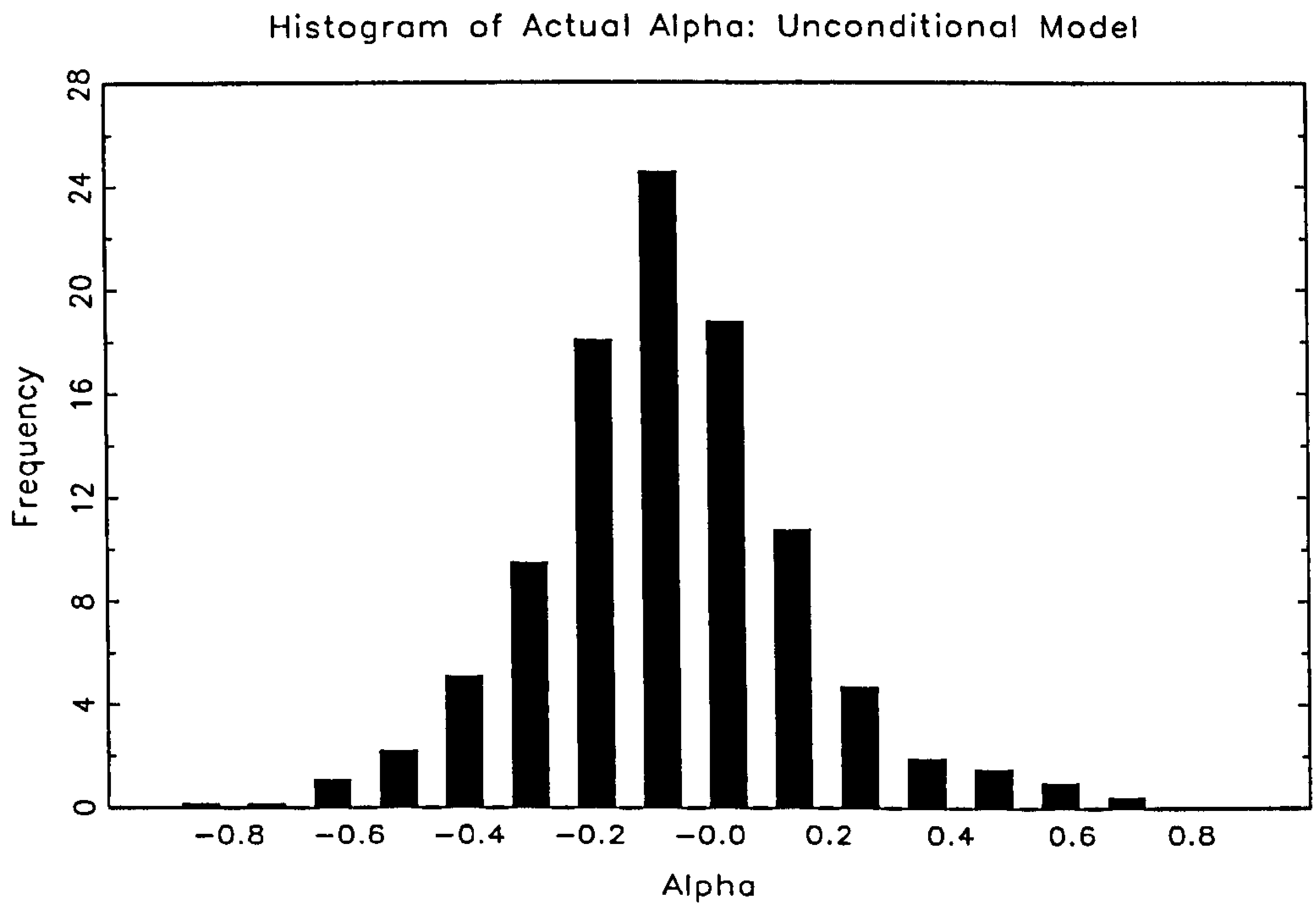


Figure 4.2: Cross-Sectional Alpha, Conditional Beta Model

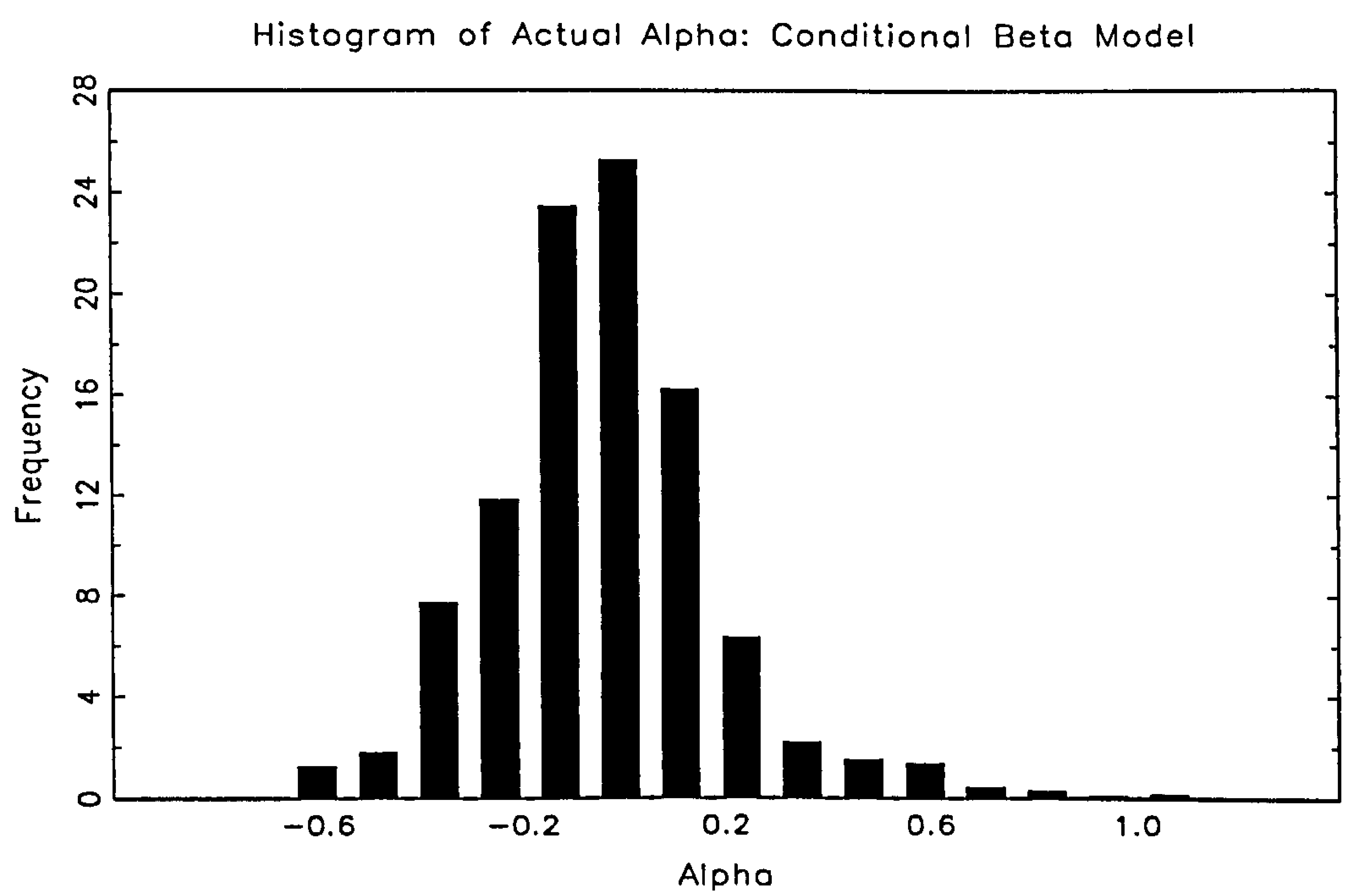
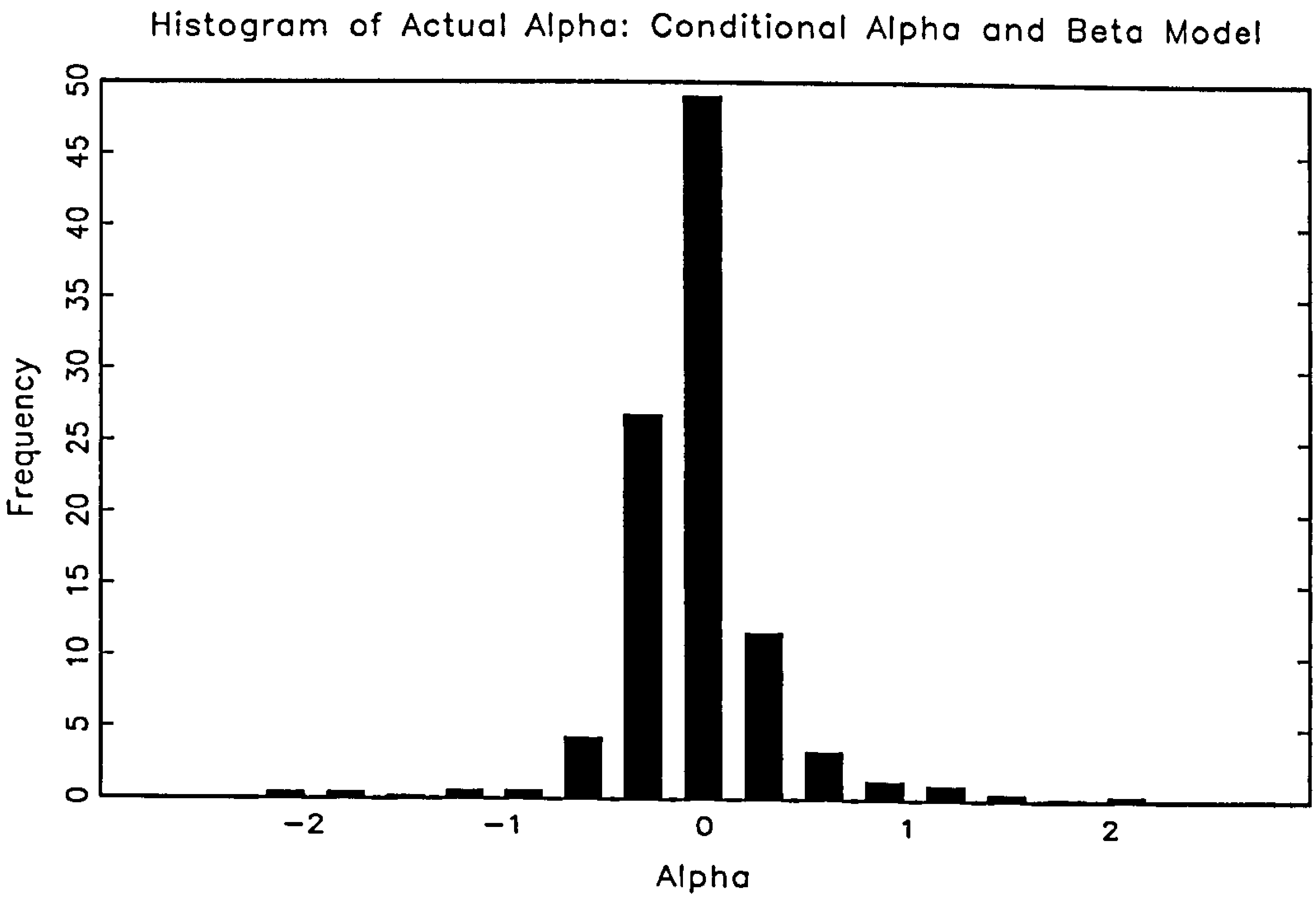


Figure 4.3 Cross-Sectional Alpha, Conditional Alpha-Beta Model



CHAPTER 5

THE BOOTSTRAP METHODOLOGY: SKILL VERSUS LUCK IN PERFORMANCE

This chapter describes the bootstrap methodology used in this study to distinguish skill from luck in fund performance. In chapter 4 many models of security returns and tests of fund abnormal performance were discussed. Inferences regarding fund abnormal performance are usually based on t-tests of the significance of measures such as alpha (Jensen's alpha, Carhart's alpha etc). There are two difficulties with this approach.

First, for their statistical validity these tests require that the alpha measure be normally distributed. However, as was seen in chapter 4 the residuals from numerous performance models applied to the funds are highly non-normal for around 60% of the mutual funds under investigation in this study. Hence, the vector of model random disturbances may be poorly approximated by multivariate normality and in turn the distribution of alpha may not in fact be normal as required. Furthermore, as will be seen in chapter 6, high variance non-normal residuals are particularly prevalent in the top and bottom performing funds and it is these funds which are of most interest to investors. There are a number of possible explanations as to why non-normal security returns can remain at the portfolio (mutual fund) level. As noted by Kosowski et al (2004), co-skewness of individual constituent non-normal security returns may not be diversified away in a fund[1]. Also, funds often hold derivatives to hedge return outcomes which may result in a non-normal return distribution. The bootstrap methodology in this study is used to control for this non-normality in fund returns.

Second, in a large universe of funds, 842 in this study, there are likely to be some funds that perform well (badly), simply due to good (bad) luck. Suppose that *all* funds have no stock picking ability (i.e. H_0: $\alpha_i = 0$ for i = 1, 2, ..., n), each fund's 'true' alpha is normally distributed and each fund has a different but *known* standard deviation σ_i. Suppose we are interested in the performance of the best fund. If we 'replay history' just

[1] The central limit theorem implies that a large, well diversified and equal weighted portfolio of non-normally distributed securities will approximate normality. However, many funds do not have these characteristics.

for the 'best fund', where *we impose* $\alpha_i = 0$ (here i = best fund) but 'luck' is represented by the normal distribution with known standard deviation σ_i, we would sample a different estimate of alpha. Of course there is a high probability that we sample a value of alpha close to zero, but 'luck' implies that we may sample a value for alpha which is in the extreme tails of the distribution. Similarly, when we resample the alpha for all the other n-1 funds, all with $\alpha_i = 0$ (but with different σ_i), it is quite conceivable that the second or third etc. ranked fund in the *ex-post* data, now has the highest alpha. This would hold *a fortiori* if the distributions of the second, or third, etc. ranked funds have relatively large values of σ_i.

From this single 'replay of history', with $\alpha_i = 0$ across all funds, we have $(\alpha_1^{(1)}, \alpha_2^{(1)}, \ldots \ \alpha_n^{(1)})$ from which we choose the largest value $\alpha_{\max}^{(1)}$. So, taking the 'luck distribution' across *all funds* into consideration (with different σ_i's), we now have one value $\alpha_{\max}^{(1)}$ for the best fund which arises purely due to sampling variability or luck. However, by repeating the above (B-times) and each time choosing $\alpha_{\max}^{(k)}$ (for k = 1, 2, ..., B trials) we can obtain the complete distribution of $\alpha_{\max}$ under the null of no outperformance, which is denoted as $f(\alpha_{\max})$ here.

Note that the distribution $f(\alpha_{\max})$ uses the information about 'luck' represented by all the funds and not just the 'luck' encountered by the 'best fund' in the ex-post ranking. This is a key difference between this study and many earlier studies. It is important to measure the performance distribution of the 'best fund' not just by re-sampling from the distribution of the best fund ex-post, since this is a single realization of 'luck' for one particular fund. Clearly, re-running history for just the ex-post best fund ignores the other possible distributions of luck (here just the different standard deviations) encountered by all other funds – these other 'luck distributions' provide highly valuable and relevant information.

Having obtained the 'luck distribution', it is possible to compare the best fund's actual *ex-post* performance given by its estimated $\hat{\alpha}_{\max}$ against the 'luck distribution' for the best fund, $f(\alpha_{\max})$. If $\hat{\alpha}_{\max}$ exceeds the 5% right tail cut off point in $f(\alpha_{\max})$, one can reject the null hypothesis that the performance of the best fund is attributable to luck.

Above, one could have chosen any fund (e.g. the 2^{nd} best fund) on which to base the 'luck distribution'. So it is possible to compare the actual *ex-post* ranking for any chosen fund against *its* luck distribution and separate luck from skill, for all *individual* funds in the sample.

Under the null of no out-performance, this procedure does *not assume* the distribution of alpha for each fund is normal and *each fund's alpha* can in principal take on any distribution. The distribution for each fund's 'luck' is represented by the empirical distribution observed in the historic data and this distribution can be different for each fund. Hence the distribution under the null $f(\alpha_{\max})$, encapsulates all of the different individual fund's luck distributions (and in a multivariate context this cannot be derived analytically from the theory of order statistics).

As alluded to above, investors are particularly interested in funds in the tails of the performance distribution, such as the best fund, the second best fund, and so on. This study finds below that the empirical 'luck distribution' of alpha for these funds are highly non-normal, thus invalidating the usual test statistics. This motivates the use of the cross-section bootstrap to ascertain whether the 'outstanding' or 'abysmal' performance of 'tail funds' is due to either, good or bad skill or good or bad luck, respectively.

5.1 Description of the Steps in the Bootstrap Procedure[2]

Consider in general an OLS estimated model of equilibrium returns of the form

$$r_{it} = [X_t]\hat{b}_i + \hat{e}_{it} \tag{5.1}$$

where r_{it} = a (T_i x 1) vector of excess returns over the risk free rate for fund i, X_t = a (T_i x k) matrix of observations on a constant and k-1 risk factors, $\hat{e}_{it}$ = a vector of OLS residuals while $\hat{b}_i$ = a (k x 1) vector of $\hat{\alpha}_i$and the estimated factor loadings. The bootstrap methodology is as follows: The performance measurement model is first estimated by OLS for each fund in the sample. Let N denote the number of funds in the sample. For

[2] For a general discussion of bootstrap methodologies and their properties see Efron and Tibshirani (1993).

each fund, the estimated factor loadings and OLS residuals, $\hat{b}_i$ and $\hat{e}_{it}$, are saved. In the next step, for each fund *i*, a random sample of residuals of size T_i is drawn (with replacement) from $\hat{e}_{it}$. For each fund, using the estimated factor loadings from step one and the original chronological ordering of X_t and setting $\hat{\alpha}_i = 0$ under the null hypothesis of no abnormal performance, bootstrap fitted returns, $\tilde{r}_{it}$, are constructed. By construction, for these bootstrapped simulated returns each fund has 'true' abnormal performance of zero. Using these bootstrap fitted returns, the performance measurement model is re-estimated and a bootstrap estimate of abnormal performance under the imposed null hypothesis is obtained, ie $\tilde{\alpha}_i$. For each fund, $\tilde{\alpha}_i$ represents random sampling variation around a true value of zero. This simulation process is repeated B times for each of the N funds, where B denotes the number of bootstrap simulations. (In this study, results are reported for B = 1,000, although higher values of B were tested). The bootstrap estimates of performance may be gathered in a matrix of dimension NxB as follows.

$$\begin{bmatrix} \tilde{\alpha}_1^{b,1}\tilde{\alpha}_1^{b,2}\ldots\ldots\tilde{\alpha}_1^{b,B} \\ \tilde{\alpha}_2^{b,1}\tilde{\alpha}_2^{b,2}\ldots\ldots\tilde{\alpha}_2^{b,B} \\ \vdots \\ \vdots \\ \tilde{\alpha}_N^{b,1}\tilde{\alpha}_N^{b,2}\ldots\ldots\tilde{\alpha}_N^{b,B} \end{bmatrix}$$

The first column of this bootstrap matrix contains the bootstrap estimates of abnormal performance under the null hypothesis of zero abnormal performance for each fund $i = 1, 2 \ldots N$ from the first bootstrap simulation. The second column of the matrix contains the estimates from the second bootstrap simulation for each fund and so on. The first row of this bootstrap matrix contains the B (or 1,000) bootstrap estimates of alpha under the null for the first fund. The second row contains the B bootstrap estimates for the second fund and so on.

The next step is to sort each column of this bootstrap matrix from highest to lowest. The first row of this sorted bootstrap matrix now contains the highest values of $\tilde{\alpha}_i$ from the B bootstrap simulations of each of the funds, under the null hypothesis where in all cases the true value of α_i is known to be zero. The second row contains the second highest values of $\tilde{\alpha}_i$ from the B bootstrap simulations of each of the funds. The last row contains the lowest values of $\tilde{\alpha}_i$ from the B bootstrap simulations of each of the funds.

The B elements of the first row of this sorted bootstrap matrix represent the distribution of the highest possible performance which is simply due to random sampling variation or chance as in fact, by construction, there is no true abnormal performance. This distribution provides an estimate of, or proxy for, luck for the extreme top performance. The second row of this sorted matrix represents the distribution of the second highest possible performance which is simply due to chance or luck. The last row of this matrix represents the distribution of worst possible performance which is due to random sampling variation in the performance measure or bad luck. Each row of this sorted bootstrap matrix provides a distribution of performance due to luck at *each* point or percentile in the performance distribution from extreme best performer to extreme worst performer.

Having already estimated the *actual* or unmodified alpha performance measures for each of the N funds in the sample, $\hat{\alpha}_i$, the next step is to rank these funds by $\hat{\alpha}_i$ from highest to lowest. One can then compare the actual highest ranked $\hat{\alpha}_i$ against the distribution of highest performance under the null hypothesis (the first row of the sorted bootstrap matrix above) to evaluate whether this fund's performance is superior to random sampling variation or luck at that point in the performance distribution, ie does this fund possess genuine stock picking ability to beat luck? Similarly, the second highest $\hat{\alpha}_i$ fund can be compared against the bootstrap luck distribution at the second highest point in the performance distribution (the second row of the sorted bootstrap matrix above) to evaluate whether it possesses genuine stock picking ability. The worst actual $\hat{\alpha}_i$ fund may be compared against the last row of the sorted bootstrap matrix above to assess whether its performance is worse than bad luck.

The sorted bootstrap matrix enables the calculation of p values in respect of whether funds perform better (or worse) than good (bad luck). The probability that the best actual fund alpha, ie highest $\hat{\alpha}_i$, represents performance better than mere good luck may be estimated as the percentage of bootstrap alphas, $\tilde{\alpha}_i$ at the highest point in the performance distribution under the null, (ie the first row of the sorted bootstrap matrix) which exceed the highest $\hat{\alpha}_i$. Similarly, the probability that the worst actual fund alpha performs worse than bad luck is the percentage of bootstrap alphas at the lowest point in the performance distribution (ie the last row of the sorted bootstrap matrix) which are lower than the lowest ranked actual fund alpha. Such comparisons between actual ranked fund alphas and bootstrap distributions of alpha under the null hypothesis of zero abnormal performance can be made at all points and percentiles of the performance distribution from the highest to lowest ranked funds.

As an alternative interpretation or use of the matrix of bootstrap alphas under the null, one can (arbitrarily) select a level of performance, good or bad, and then identify how many funds in the sample one would expect to achieve this level of performance by chance alone. This can then be compared with how many funds actually achieve or exceed this performance.

5.2 The t-statistics of Alpha as the Measure of Performance

Equally, we can employ the t-statistic of alpha, rather than alpha itself, as the measure of fund performance and repeat the bootstrap methodology exactly as above. This is, we can construct the bootstrap distribution of the t-statistic of alpha under the null hypothesis of zero abnormal performance, Ho: $\alpha_i = 0$, and compare this bootstrap distribution to the actual ranked t-statistics of alpha. Using t-statistics has an added advantage: the alphas of funds with few observations may be estimated with high standard error. This may generate outlier alphas in the sample. There is a risk that these funds will disproportionately occupy the extreme tails of the actual and bootstrapped alpha distributions. The t-statistic provides a correction by scaling alpha by its estimated precision, ie its standard error. Therefore, the distribution of bootstrapped t-statistics has superior statistical properties and more reliably identifies talented fund managers. This is especially the case in the extreme ends of the performance distribution – the areas of particular interest. In this study, the t-statistic is adopted as the fund performance measure. (see Mamaysky, Spiegel and Zhang 2004) for discussion of estimation error in alpha).

In general, if θ_i, a parameter of interest, is a pivotal statistic, ie one that is independent of 'nuisance' parameters such as regression error variance, an Edgeworth expansion provides a refined approximation to the asymptotic probability distribution function of $\sqrt{T_i}.[\ \theta_i(\hat{C}_i)\ -\ \theta_i(C_i)]$, where C_i is the population cumulative distribution function of the returns data of fund *i*, r_i, $\hat{C}_i$ is the empirical distribution of same and $\theta_i(\hat{C}_i)$

is an estimated performance statistic. In turn the bootstrap method provides a close approximation to the Edgeworth expansion. In this study, fund abnormal performance as measured by $\hat{\alpha}_i$ is not a pivotal statistic. However, the t-statistic of alpha is a pivotal statistic. (See Hall 1992, Kosowski et al 2004 for further discussion).

5.3 A Restriction to Funds With a Minimum Number of Observations

As indicated, funds with few observations may have higher sampling variability in the alpha estimate. This could widen the tails of the bootstrap distribution making it more difficult for funds to beat luck. While using the t-statistic of alpha mitigates this potential problem it may also be advisable to restrict the analysis to funds with a minimum number of observations. In this study a minimum fund history of 3 years (36 observations) is imposed for a fund to be included in the analysis. However, while statistically advantageous, this may induce a look-ahead bias by restricting the entire analysis to funds which have been skilled (or lucky) enough to survive for at least 3 years. Because using the t-statistic of alpha rather than alpha as the performance measure already mitigates the problem, in this study the restriction is set at 36 observations rather than say 60 or higher in order to minimize the probability of inducing a look-ahead bias. To examine the significance of this issue, the sensitivity of the bootstrap results in this study are tested for a number of alternative minimum observations restrictions.

On a point of clarity, note that survivorship bias arises from omitting nonsurviving funds and therefore is a property of the sample. Look-ahead bias arises by requiring funds to have existed for a minimum period of time and as such is a property of the methodology.

Accurate modeling of the tails of the performance distribution is the central purpose behind the bootstrap methodology here and is an important strength of this study. A further strength is the large number of funds in the sample. In particular, this allows for improved estimation of the sampling distributions of performance in the tails. For example, a study with relatively few *ex-post* high and low ranked funds risks having slimmer tails in the bootstrapped distribution of performance which may in turn lead to unreliably low p values. This point underpins the importance of including the nonsurvivor class of funds in the analysis for more accurate modeling of the left tail of performance. If nonsurvivor funds have closed due to poor performance, the increased number of poor funds allows for a more accurate picture of the left tail of the bootstrapped distribution.

5.4 Extensions of the Bootstrap Methodology

The methodology outlined above is the 'baseline' methodology in this study and is implemented in chapter 6. However, the bootstrap procedure can also be modified to incorporate additional fund return characteristics such as serial correlation, heteroscedasticity or cross-sectional (across funds) correlations among fund regression residuals. Where such features are present, refinements to the bootstrap procedure help to retain this information in the construction of the bootstrap distributions. This is important in order to mimic the underlying 'true' return generating process as closely as possible. Furthermore, alternative bootstrap procedures may be applied in which we randomly resample not only the residuals but also the risk factors. These alternative bootstrap procedures are important as the variability of the bootstrap estimates of alpha under the null hypothesis may be sensitive to the choice of procedure. These extensions of the bootstrap procedure are now discussed. As a sensitivity analysis these extensions to the bootstrap procedure are also implemented in chapter 6 and results are compared with the baseline procedure.

5.4.1 Serial Correlation

Residual serial correlation indicates that the underlying distributions of model random errors in the data generating process may not be independently (or identically) distributed. However, a bootstrap residual re-sampling procedure which randomly selects residuals 'one-at-a-time' makes this assumption. Hence such a bootstrap procedure fails to incorporate possibly valuable information content in the residuals as to the nature of the 'true' sampling mechanism from which returns are generated and as such may fail to mimic the return generating process as closely as possible. In this study it is found that around 40% of the sample of mutual funds exhibit serial correlation by a Lagrange Multiplier test (see chapter 4). To incorporate this residual serial dependence, the bootstrap procedure should draw residuals (with replacement) in block lengths corresponding to the suspected order of serial correlation. For example, for serial correlation of order 1, residuals should be randomly drawn in chronological pairs, for serial correlation of order 2, residuals should be randomly drawn in chronological triplets etc.

5.4.2 Newey-West Adjusted Standard Errors

Due to the evidence of serial correlation, throughout this study all t-statistics are calculated based on Newey-West autocorrelation (and heteroscedasticity) adjusted standard errors. These include the actual (unmodified) t-statistics of alpha and the t-statistics of alpha calculated under the null hypothesis in the bootstrap procedure. For conclusions to be robust it is necessary to examine whether the estimated bootstrap p values are sensitive to the order of autocorrelation in the Newey-West adjustment.

5.4.3 Cross-Sectional Dependence

Fund regression residuals may provide evidence of cross-sectional correlations, ie $cov(\varepsilon_i,\varepsilon j) \neq 0$ for funds *i* and *j*, where ε_i and εj are vectors of disturbances or the idiosyncratic components of fund returns. Such correlations may arise from a misspecification in the performance model which is common to many funds and could be even more pronounced among funds which have similar stock holdings due to herding behaviour. Again, these cross-fund correlations may contain significant information about the true return generating process among funds and this information should be retained by the bootstrap procedure. The bootstrap procedure described up to now assumes independence among residuals across funds. To refine the bootstrap procedure to preserve cross-sectional correlations among residuals involves ensuring that within each simulation the order in which residuals are randomly drawn is the same for each fund, ie the random time ordering of residuals varies across simulations but within a simulation it is constant across funds. This will be denoted as a 'cross-sectional bootstrap'. This bootstrap refinement addresses the estimation of cross-sectional correlations in regression residuals and avoids estimating a large covariance matrix of these residuals, which would otherwise be necessary in order to characterize the joint distribution of residuals and performance estimates.

However, this refinement presents a difficulty. The sample of funds do not exist contemporaneously, ie some funds have closed before others were opened. Thus, ensuring that the order in which residuals are randomly drawn is the same for each fund may involve selecting residuals for many funds at time periods when these funds do not exist. One solution is to select a sub-sample period during which a sub-sample of funds have a complete return history. Because this sub-sample of funds all exist at the same time, it is possible to implement a cross-sectional bootstrap procedure. A caveat applies with regards to this approach: (i) this restriction would severely reduce the number of funds which could be selected for the analysis, (ii) it would limit the analysis to a shorter sample period and (iii) it could potentially impose a survivorship bias by excluding possibly poorer performing funds which have died during the sub-period. However, of interest here is to determine whether there is consistency in findings *between* the baseline and cross-sectional bootstrap procedures and so, as a sensitivity analysis, both procedures are applied to a sub-sample of funds in a sub-period.

Model misspecification is a likely cause of any cross-fund residual correlation. However, in this study many alternative equilibrium models of performance are tested and hence results are unlikely to be model specific. Kosowski et al (2004) also implement a cross-sectional bootstrap technique but the authors report that this did not alter conclusions.

5.4.4 Alternative Bootstrap Procedures

The baseline bootstrap procedure is a 'residual-only' resampling, ie only the residuals are resampled in each of the bootstrap simulations. A 'factor-residual' bootstrap resampling procedure may also be applied in which the residual resampling procedure is augmented with (i) factor returns that are resampled independently of the residuals or (ii) factor returns and residuals that are resampled together in pairs thus observing the same time series pairings as observed in the actual data.

To illustrate the alternative bootstrap techniques, consider again (5.1) as follows:

$$r_{it} = [X_t]\hat{b}_i + \hat{e}_{it} \qquad (5.1)$$

In case 1, 'residual-only' resampling, residuals are randomly drawn (with replacement) from $\hat{e}_{it}$ while X_t is unaltered from its original chronological time ordering, ie X_{it} is nonstochastic in repeated sampling. In this case, if for a given fund i the variance of $\hat{e}_{it}$ is small (of course $\hat{e}_{it}$ always has a mean of zero) then the resulting variability in $\tilde{\alpha}_i$ over B = 1,000 bootstrap simulations will also be relatively small. As will be seen in the discussion of bootstrap results (chapter 6), the variance of fund regression residuals is larger for funds at both extreme ends of the performance distribution but this variance falls as one moves even slightly closer towards the centre of the distribution, i.e. a large proportion of funds have a relatively low variance of residuals and consequently these funds are likely to produce relatively low sampling variation in the bootstrap coefficient estimates (including alpha) over 1,000 simulations.

In case 2, an alternative resampling procedure is to augment the random resampling from $\hat{e}_{it}$ in (5.1) by also independently resampling from X_{it}, ie X_{it} is stochastic in repeated sampling. In this study this is denoted 'independent factor-residual' resampling. Resampling factor returns as well as residuals may allow for greater sampling variation in the bootstrap coefficient estimates (including alpha) over the 1,000 simulations. Furthermore, within this factor-residual resampling technique, consideration must also be given to a 'cross-sectional factor-residual' bootstrap. In this procedure, within each bootstrap simulation both factors and residuals are independently resampled but in addition, across funds, the same realized time index of factors and the same realized time index of residuals is maintained, i.e. the random time ordering of factors and residuals vary across simulations but within a simulation they are constant across funds. This cross-sectional factor-residual bootstrap controls for non-normality in residuals that may have a factor component, for example a market component, which gives rise to co-skewness in returns across funds.

In case 3, a further alternative is to randomly resample from $\hat{e}_{it}$ and X_{it} in pairs, ie maintaining the same random time sequence from both in each one of the 1,000 draws. This procedure captures possible heteroscedastic features in the underlying return generating process, ie heteroscedastic errors which are conditional on a factor. In this study, this procedure is denoted 'pairwise factor-residual' resampling. However, in such a procedure the only source of sampling variation in the bootstrap coefficient estimates over 1,000 simulations lies in the sampling with replacement, ie the possibility that the same time pairing from [$\hat{e}_{it}$, X_{it}] could be drawn more than once. (If sampling, of size T_i, was carried out without replacement then the bootstrap coefficients, including alpha, would be identical over all simulations). Therefore, it is possible that within this case 3 bootstrap procedure one may obtain a relatively low sampling variation in the bootstrap coefficient estimates, including alpha, over the 1,000 simulations. Similar to case 2 above, one may also implement a 'cross-sectional pairwise factor-residual' resampling procedure in which the realized time index of pairwise factor-residual draws is kept constant across funds within each of the 1,000 simulations.

It is important to consider each of these alternatives for carrying out the bootstrap methodology as the variability in the bootstrap alphas, and consequently the p values, may be sensitive to the chosen procedures.

5.5 The Power of the Bootstrap Methodology

In order to examine the power of the bootstrap methodology to identify true performance in the data, Kosowski et al (2004) carry out a Monte Carlo analysis of the bootstrap procedure.

The authors generate artificial return data according to a singe factor model:

(5.2)
$$r_{it} = \alpha_i + b_i[X_t] + \varepsilon_{it}, \quad i = 1, 2, \ldots N, \; t = 1, \ldots Ti$$
$$X_t \sim N(0,\sigma_x^2), \;\; \varepsilon_{it} \sim N(0,\sigma_\varepsilon^2).$$

where N and T_i correspond to the number of funds and life spans of funds as found in the actual data. The study carries out 1,000 Monte Carlo simulations which assume a normally distributed factor and fund random disturbance terms. In the Monte Carlo analysis residuals are assumed to be uncorrelated across funds. To examine the properties and power of the bootstrap procedure, Kosowski et al impose true performance levels on the artificially generated fund returns, in separate Monte Carlo simulations, of (i) $\alpha_i = -0.2$, (ii) $\alpha_i = -0.02$, (iii) $\alpha_i = 0$, (iv) $\alpha_i = +0.02$, (v) $\alpha_i = +0.2$.

Kosowski et al find that under each of the five performance scenarios, the bootstrap results are in accordance with the performance imposed on the data. For example, with the 'true' α_i set equal to 0.2 for all funds, the bootstrap p values of funds with high estimated alphas all strongly indicate that these funds beat luck while the p values of funds with low estimated alphas are shown to be unlucky. In the scenario where the true α_i is set equal to zero for all funds, the bootstrap p values for each point in the performance distribution are all approximately 0.5, as they should be under the set up. These Monte Carlo findings support the bootstrap methodology as a means of correctly identifying outperformance and underperformance in the data.

5.6 Bootstrap Analysis of Subgroups of Mutual Funds

Also of interest is to examine whether stock picking skills differ between subcategories of funds. For example, is it affected by (i) the investment objective of the fund, (ii) possible informational asymmetries arising from where the fund is domiciled or (iii) the length of fund history. This can be examined by applying the bootstrap procedure separately to the different subgroups of funds in each case. For example, in the case of fund investment objectives it is interesting to examine whether the evidence of outperformance among growth stock funds in the US (Chen et al, 2000) transfers to the UK or to test the hypothesis that the market for small stocks is less efficient.

When the bootstrap methodology is applied separately to subcategories of funds, the bootstrap distribution of performance, ie the estimate of luck, is based only on a more homogenous peer group or risk group of funds. This further helps to control for cross-sectional risk characteristics which may not be adequately captured by the performance model.

This concludes the discussion on the bootstrap methodology. In chapter 6 the methodology and the various refinements and extensions described above are implemented and results reported.

CHAPTER 6

EMPIRICAL RESULTS OF BOOTSTRAP ANALYSIS

In chapter 5 the 'baseline' bootstrap methodology of this study, along with numerous extensions of this methodology, were described in detail. In chapter 6 these methodologies are implemented and results discussed. All tables and graphs of results are presented collectively at the end of, rather than throughout, the chapter.

6.1 All Investment Objectives

Table 6.1 presents the bootstrap performance statistics for the full sample of UK equity mutual funds, ie including funds of all investment objectives and all locations. In chapter 4, three representative models were selected for the bootstrap analysis. These were a Fama and French based (i) unconditional model, (ii) conditional beta model and (iii) conditional alpha-beta model. In Table 6.1, Panels A, B and C report the bootstrap findings for these models respectively. For ease of presentation results are reported for selected points in the cross-sectional distribution of performance as indicated. The first row in each panel shows alpha, measured in percent per month. The second row in each panel presents "t-alpha", the corresponding t-statistic of the alpha in row 1. Row 3 ("t-stat") presents the t-statistics of alpha ranked from lowest to highest. Throughout this chapter the t-statistic of alpha is employed as the performance measure and the bootstrap findings are discussed in terms of the p values of the t-statistics. From chapter 5, as the t-statistic scales the alpha measure by its estimation error, it has superior statistical properties, particularly in the extreme tails of the performance distribution. Row 4 ("p-tstat") reports the bootstrap p values of the t-statistic in row 3. For further statistical reliability the analysis is restricted to funds with a minimum of 36 observations, unless as otherwise stated. This leaves 675 funds in this analysis. All t-statistics of alpha, both actual (unmodified) t-statistics and bootstrap t-statistics (under the null), are based on Newey-West adjusted heteroscedasticity and autocorrelation adjusted standard errors. All bootstrap results are based on 1,000 simulations. Where relevant, many tables also contain information on the investment style ("Style"), survival status ("Survival") and operation location ("Location") of the funds corresponding to the ranked t-statistics of alpha.

In Table 6.1, Panel A reveals that the best fund ranked by alpha from the unconditional model achieved abnormal performance of 2.235% per month. This alpha has a t-statistic of 2.226. However, ranking by the t-statistic of alpha the highest fund has a t-statistic of 3.389. The bootstrap p value (of the t-statistic) equal to 0.437 indicates that from among the 1,000 bootstrap simulations across each and all of the funds under the null hypothesis of zero abnormal performance, 43.7% of the highest bootstrap t-statistics were greater than 3.389. Operating at a 5% significance cut-off, the p value of 0.437 fails to reject the hypothesis that the performance of the best fund is within the boundaries of performance that may be explained by random sampling variation in the t-statistic around a 'true, known' value of zero. The p value indicates that the fund's performance is attributable to chance or luck at *that* point in the performance distribution. The p value of 0.437 fails to reject the hypothesis that the top ranked fund does not possess genuine stock picking ability/skill.

Looking across the entire right tail of the performance distribution, the evidence regarding skill versus luck is mixed. The 1st, 2nd, 3rd, 5th and 7th ranked funds do not exceed performance which could be explained by good luck at 5% significance at each of these points in the distribution. However, there is strong evidence to indicate that lower ranked funds (10th, 12th) possess genuine skill. Among the top 20 funds ranked *ex-post* (where the 20th ranked fund has a t-statistic of 2.023), 7 (12) funds are found to exhibit stock picking ability at 5% (10%) significance. There is no evidence of stock picking ability in any fund ranked lower than the top 20.

Table A6.1 in the appendix to this chapter shows the ranked t-statistics and p values of the top and bottom 50 funds for further detail. Critically, the top 7 ranked funds are simply lucky while only 7 funds ranked further inside the extreme right tail between 8th and 17th are skillful at 5% significance. Clearly therefore skill is not distinguished from luck simply by the fund's rank. This is important to note as fund ranking is often used in the marketing of funds. This underpins the value of the bootstrap methodology which uses an empirical nonparametric distribution at *each* point in the performance distribution to determine the statistical significance of performance rather than relying on a standard t-test. The latter, which relies on the normality assumption, would have concluded that all of the top 20 funds yielded significant abnormal performance.

In the left tail of the distribution, ie the left side of Panel A, the worst ranked fund by alpha yields a negative return of -0.901% per month with a t-statistic of –2.532. The lowest ranked fund by the t-statistic of alpha yields a t-statistic of –5.358. The bootstrap p value of the t-statistic of 0.009 means that from among the 1,000 bootstrap simulations across each of the funds under the null hypothesis of zero abnormal performance, less than 1% of the lowest bootstrap t-statistics were lower than –5.358. This strongly fails to reject the hypothesis that this fund's performance is worse than random sampling variation around zero. This fund has produced 'truly' inferior performance worse than bad luck. It is clear from the left tail of the distribution generally in Panel A that performance levels at these selected points in the distribution are worse than may be explained by bad luck.

As described in chapter 5, as an alternative interpretation, the bootstrap may be used to estimate how many funds one might expect to achieve a given level of performance by random chance alone. This can then be compared to the number of funds which actually do achieve this level of performance. For example, based on the unconditional model, by random chance one would expect 7 funds to achieve an alpha of 0.5% per month or higher. In fact, 19 funds achieve this level of performance (or higher). However, alphas of 0.1% are expected to be achieved by 171 funds based on chance while in fact only 142 actual funds reach this level. This interpretation is consistent with the discussion of the p values above: there is greater evidence of genuine outperformance towards the top end of the performance distribution (among some of the top 20) but not further inside the right tail.

Figure 6.1 offers further insight into mutual fund performance relative to luck. The figure plots Kernel density estimates of the distributions of both the actual and bootstrap t-statistics (unconditional model). The distribution of bootstrapped t-statistics (solid line) is a graphical illustration of the random variation or dispersion in the t-statistics of alpha around a 'true' value of zero – by construction. It provides a picture of the range in performance that may be expected simply due to luck. Comparison between this distribution and the distribution of the actual (unmodified) t-statistics puts actual performance in context relative to luck. It is clear from Figure 6.1 that the actual performance distribution (dashed line) lies largely to the left of the bootstrap distribution. There is some exception in the right tail which indicates that there are a number of high ranking funds which achieve performance superior to luck. However, the left tail of the actual distribution lies to the left of the bootstrap distribution: poor performing funds cannot attribute performance to bad luck[1].

Figure 6.2a shows histograms of the bootstrapped t-statistic of alpha at selected points in the upper end of the performance distribution. The upper left panel shows the histogram of the highest t-statistics across funds from each one of the 1,000 bootstrap resamples under the null hypothesis while the upper right panel shows the 1,000 99th percentile t-statistics and so on as indicated. Comparing the four histograms it is evident that the distribution of bootstrap t-statistics at the top end of the performance scale (BEST, 99th percentile) are highly non-normal and have a relatively high variance but upon moving even slightly closer to the centre of the performance scale (95th, 90th percentiles) the histograms more closely approximate normality and exhibit a lower variance. Further insight to explain this is provided in Figure 6.2b which presents histograms of the regression residuals of the funds ranked at the same points in the performance distribution. The set of residuals from the higher ranked funds have higher variance and greater non-normality than the residuals of the funds even slightly closer to the centre of the performance distribution. It is this high variance and non-normality among the top funds' regression residuals, and in particular the existence of large positive residuals, that causes these funds to populate the top end of the bootstrap t-statistic distributions and generate a wide non-normal dispersion among these top t-statistics in the bootstrap procedure.

In Figure 6.3a and Figure 6.3b an almost mirror image of this is presented for the lower end of the performance distribution. The upper left panel of Figure 6.3a shows the histogram of the lowest bootstrap t-statistics across funds from each one of the 1,000 bootstrap simulations under the null hypothesis, the upper right panel shows the histogram of the 1,000 t-statistics at the 1st percentile across funds and so on. Once again there is evidence that the histograms at the lower points in the performance distribution exhibit a higher variance and greater non-normality than the histograms closer to the centre of the performance distribution. Similar to above, as can be seen from Figure 6.3b, this reflects the fact that the residuals of the lower ranked funds generally exhibit higher variance and greater non-normality than the residuals of funds closer to the centre of the distribution. (the absolute worst fund is the exception here). Again, it is this high variance among the

[1] Note the Kernels compare the frequency of a given level of performance among actual performance against the frequency of this level of performance in the *entire* bootstrap matrix. The bootstrap p value is a more sophisticated measure and compares the actual performance measure against the bootstrap distribution of performance *at the same point* in the cross-sectional performance distribution.

lower funds' regression residuals (and the existence of large negative residuals) that causes these funds to populate the lower end of the bootstrap t-statistic distributions and generate a wide dispersion among these t-statistics in the bootstrap procedure.

This non-normality and high variance among the residuals and t-statistics of the *top* and *bottom* funds motivates the use of the bootstrap procedure to more correctly identify the distribution of performance at the *extreme* ends of the performance spectrum, rather than relying on a normality assumption. Therefore, we can more accurately draw inferences regarding the statistical significance of individual fund performance in the areas of performance of greater interest to investors.

All results presented above relate to the unconditional three-factor Fama and French model. In chapter 4 on model selection it was noted that the value factor in the three-factor model is not well specified on average across funds while the Schwartz Information Criterion (SIC) for the 4 factor model was also relatively low as it is for the 3 factor model. In results not shown here (but available on request), the bootstrap analysis was also implemented for two-factor (market and size factors) and four-factor models. The conclusions as described above for the three-factor model are unaltered: the two, three and four risk factor models are remarkably consistent in showing a small number of funds beating luck at the upper end of the distribution while poor performance is found to be worse than bad luck as above.

In Panel B of Table 6.1, the bootstrap findings from the conditional beta model are reported. The interpretation of results for both the left and right tail of the performance distribution is broadly similar to that of the unconditional model in Panel A. Fund performance in the extreme right tail of the distribution is attributable to luck. Within the top 5% of ranked funds, 'true' stock picking skill is slightly more prevalent than suggested by the unconditional model. However, again performance is not superior to luck at ranked performance even slightly closer to the center of the distribution. In the left tail, poor performance is again found to be worse than bad luck.

Panel C of Table 6.1 presents bootstrap findings for the conditional alpha-beta performance model. Again, results here are very similar to both previous classes of model. Here, there is superior performance among some, but not all, funds ranked within the top

3% in the upper end of the performance distribution but not at the extreme right tail. In the left tail of the distribution performance is found to be worse than bad luck.

The findings reported above for the three classes of model are quite consistent with those of the Kosowski et al (2004) US study. Using the t-statistic as the performance measure, these authors find in support of stock picking ability among many top performing funds while generally poor performance is found to be worse than bad luck. However, the evidence of skill in the Kosowski et al study is slightly more widespread than is found here among UK equity funds

In conclusion, the results suggest that there are a small number of UK equity mutual funds which can deliver genuine stock picking ability for their investors net of the annual expenses charged. Although returns are gross of the load fee. (see chapter 3 on data description).

The bootstrap methodology has been applied to the three classes of model as a test of robustness in findings. Findings are generally quite robust. However, from chapter 4 on model selection the conditional models are relatively poorly statistically determined. Accordingly, the unconditional model is selected here as the 'benchmark' model and, in the interest of parsimony, the discussion to follow is based on this model.

6.2 Performance and Investment Styles

It is of also interest to investors to identify whether stock picking talent is related to the investment style (objective) of the fund. From the mutual fund performance and persistence literature, specifically among US studies, there is some evidence of outperformance among growth stock funds, (Chen, Jegadeesh and Wermers 2000). When the analysis is restricted here to funds with a minimum of 36 observations, there are 675 funds remaining. These consist of 143 equity income funds (21%), 423 equity funds (63%) and 109 small stock funds (16%). The top 10 performing funds, (ranked by the t-statistic of alpha from the unconditional Fama-French 3 factor model) are comprised of 3 equity income funds, 5 equity funds and 2 small stock funds. The corresponding breakdown of the top 50 funds is 18, 24 and 8 respectively while the breakdown of the top 40% of the performance distribution (top 270 funds) is: 78, 154 and 38 respectively. At the opposite end of the performance scale, the worst 10 funds consist of 0 equity income funds, 8 equity funds and 2 small stock fund. The corresponding breakdown of the bottom 50 funds is 1, 37 and 12 respectively while the composition of the bottom 40% of funds is: 22, 198 and 50 respectively. In this simple analysis, equity income funds perform relatively well while (general) equity funds and to a lesser extent small stock funds compare poorly. For example, equity income funds comprise 21% of the total sample of funds but comprise 29% of the top 40% of the distribution and 36% of the top 50 ranked funds. However, equity income funds make up only 8% of the bottom 40% of funds and only 2% of the bottom 50 funds and are completely absent from the bottom 10 funds. In general, in relative terms equity income funds disproportionately occupy the upper end of the performance distribution and are disproportionately absent from the bottom end. In contrast, equity funds comprise 63% of the total sample of funds but make up only 57% of the top 40% of funds and only 48% of the top 50 funds while this class of funds comprises 73% of the bottom 40% of funds and 74% of the bottom 50 funds. Here, equity funds appear to disproportionately occupy the bottom end of the performance distribution and are disproportionately absent from the top end. Similarly, although slightly less pronounced, small stock funds also underperform in relative terms.

In Table 6.1 Panel A, the row denoted "Style" indicates the investment style of funds at various points in the performance distribution, where 1 = equity income fund, 2 = equity fund and 3 = small stock fund. From among these selected points in the distribution all funds which beat luck at 5% significance are equity income funds. Of the top 20 ranked funds, 12 beat luck at 10% significance and these are made up of 6 equity income funds, 5 equity funds and 1 small stock fund. Again, equity income funds perform relatively well.

To investigate this issue further the bootstrap procedure is implemented separately for each investment style. As noted in chapter 5 on methodology, this peer group evaluation has the advantage that in each case one is examining a more homogenous risk group which helps controls for investment style related risk characteristics which vary across funds but which may not be adequately captured by the performance model.

In Table 6.2, Panels A, B and C present the ranked alpha, its associated t-statistic, the ranked t-statistics and the bootstrapped p values of the ranked t-statistics for equity income, general equity and small stocks funds respectively. From the three panels it is clear that skill is not equal between the investment classes. Among equity income funds in Panel A the p values indicate that although the performance of the top 4 funds may be attributed to luck, many funds ranked between 5th highest and the 90th percentile 'beat' luck. In results not shown, 11 equity income funds beat luck at 5% significance. Therefore, as with the full set of mutual funds, among equity income funds skill exists inside the extreme right tail rather than at the extreme end of the tail. At the lower end of the performance distribution, the p values > 0.05 indicate that the performance of these funds is not worse than may be explained by bad luck at their points in the distribution.

In contrast, both Panel B and Panel C indicate comparatively poor performance. In this peer group analysis of equity funds, there is very little evidence of true stock picking talent even among high ranked funds while the worst funds perform worse than bad luck. Among small stock funds, only the 2nd highest ranked fund demonstrates skill at 5% significance while, again, poor performance may not be excused as bad luck. The evidence of some skill among equity income funds but its absence among equity funds is consistent with the more simple analysis above.

In the case of small stock funds, to examine whether findings may be sensitive to the choice of benchmark risk factor for size, the bootstrap procedure was repeated for an alternative size benchmark. In the three-factor model the size risk factor, SMB_t, (ie returns on small minus large capitalization stocks - see chapter 4) was substituted by a small capitalization index. However, the conclusions found above were unaltered.

In the case of equity income funds, for an arbitrarily selected level of performance of say 0.1% per month, the bootstrap procedure indicates that 25 funds would be expected to achieve this level of performance simply by chance. In fact 43 funds achieve this performance. In contrast among general equity funds, the same ratio is 103:75 while for small stock funds the ratio is 36:24.

The lack of supporting evidence in this study of stock picking ability among small stock mutual funds suggests that the market for small stocks is not less efficient, as has been asserted. If the inefficiency hypothesis is correct it has not been exploited by UK small stock fund managers. The evidence strongly suggests to investors that pursuing an income objective is the most reliable way to achieve genuine abnormal performance.

Figure 6.4 provides further performance comparison between the different investment styles showing the Kernel density estimates of both the actual and bootstrap distribution of the t-statistic of alpha in each case. A comparison of the figures reveals that for equity funds and small stock funds the distribution of actual t-statistics (dashed line) lies largely to the left of the bootstrap distribution (solid line). In the left tails, this indicates that actual performance lies outside the boundary of bad luck. This is consistent with the low bootstrap p values in the left tails of the performance distributions in Panels B and C in Table 6.2. In the right tail, the actual performance distribution lies, for the most part, within mere random sampling variation which is consistent with the high p values in the right hand side of Panels B and C. The actual and bootstrap Kernel densities of equity income funds reveal a different picture. Many low t-statistic funds lie within random sampling variation while many top t-statistic funds lie outside the boundaries of good luck. Once again this is similar to the evidence provided by the p values in Panel A, Table 6.2.

While the Kernel density plots provide a helpful graphical illustration of the comparison between actual and random performance, the plots will not necessarily yield an identical conclusion to that of the bootstrap p values. This can be seen in the extreme right tail of the distributions of equity funds and small stock funds in Figure 6.4 which suggest some skill while the bootstrap p values in Table 6.2 indicate that actual performance does not exceed luck. See footnote 1 for an explanation.

As noted, estimating the bootstrap p values separately within each investment class may further control for risk. Of course, however, the bootstrap estimate of luck is not necessarily the same for each investment style. Hence care should be taken when *comparing* across investment classes under separate bootstrap analyses. In contrast, in Table 6.1 the bootstrap estimate of luck is based on the full sample of all funds.

6.3 Performance and Fund Location

As discussed in chapter 3, while all mutual funds in this study invest only in UK equity, they may be domiciled either onshore UK or offshore. It is useful for investors to know whether there are differences in performance between these classes of funds.

Of the 675 funds in this study (after the minimum 36 observation restriction), there are 553 onshore funds and 122 offshore funds. Onshore funds comprise 82% of the total sample yet they form 89% of the top 40% of the distribution, 96% of the top 50 ranked funds and 100% of the top 10 funds. Offshore funds comprise 18% of the total sample yet at the lower end of the distribution, they make up 25% of the bottom 40% of funds, 30% of the lowest 50 ranked funds and 30% of the lowest 10 funds. Therefore, onshore funds appear to disproportionately occupy the top end of the performance distribution while offshore funds disproportionately occupy the bottom end.

In Table 6.1 Panel A, the row denoted "Location" indicates the fund's domicile where 1 = onshore, 0 = offshore. In this table and in results not shown all funds which beat luck at 5% significance are onshore funds.

For further insight the bootstrap procedure is applied separately to onshore and offshore funds. Again, this peer group evaluation may be based on a more homogenous risk group in each case and controls for possible unknown location related risk characteristics not captured by the performance model. Table 6.3 presents the bootstrap findings for onshore and offshore funds in Panel A and Panel B respectively. Among onshore funds, many (14) of the top 20 ranked funds beat luck at 5% significance. However, in Panel B no offshore fund is found to beat luck. From both panels it is clear from the bootstrap p values at the lower end of the performance distribution that poor performance is worse than bad luck.

Figure 6.5 plots the Kernel densities of the actual (dashed line) and bootstrap (solid line) distributions of the t-statistics of alpha for onshore and offshore funds. The inferences from the kernels are clearly consistent with those of the p values above. The upper panel depicting onshore funds shows that in the more extreme right tails of the distributions, actual performance lies to the right of the bootstrap distribution indicating some genuine skill. This is not the case in the right tails of offshore funds. Similarly, in both panels the left tails of the actual performance distributions lie outside the boundaries of bad luck.

There are a number of possible explanations for the relative underperformance of offshore funds. First, it may arise due to possible information asymmetries between onshore and offshore funds, ie it may be more difficult or expensive for offshore domiciled funds to obtain relevant UK equity information and thus such funds may incur an informational disadvantage. This is more likely in the past prior to the more sophisticated information technology enjoyed today. Second, it may be that offshore funds are genuinely less skillful given identical information. However, it should be noted that offshore funds are often (though not always) associated with higher charges, particularly during the earlier part of the sample period, and the return data here is net of annual charges. Higher fees arose because (i) there was less competition among UK equity funds in the offshore areas, (ii) offshore funds were often sold through intermediaries such as life assurance companies who added a layer of fees.

6.4 Performance and Survival

An important strength of this study is the inclusion of nonsurviving funds in order to control for possible survivorship bias in performance results. A commonly held assumption is that nonsurviving funds close due to poor performance. If this is true then one would expect nonsurvivng funds to underperform surviving funds.

Of the 675 funds in the bootstrap analysis, 482 (71%) are surviving funds while 193 (29%) are nonsurviving funds. From a simple ranking of funds by the t-statistic of alpha, nonsurvivors represent 31% of the top 40% of the performance distribution, 22% of the top 50 ranked funds and 30% of the top 10 funds. While nonsurvivors are slightly under-represented among the top 50 ranked funds, they are generally not disproportionately absent from the upper end of the performance distribution as might be expected if these group of funds close due to poor performance. Nonsurvivors comprise 27% of the bottom 40% of the performance distribution, only 22% of the bottom 50 ranked funds and only 10% of the lowest 10 ranked funds: oddly, nonsurvivors are disproportionately absent from the lower end of the performance distribution.

Furthermore, nonsurviving funds are well represented among the group of funds which have genuine stock picking talent by the bootstrap analysis. Among the top 20 ranked funds, 7 funds beat luck at 5% significance, 2 of which are nonsurvivors while 12 funds beat luck at 10% significance, 3 of which are nonsurvivors.

As a possible explanation for the better than expected performance of nonsurvivors, imposing the minimum 36 observations restriction here may induce a look-ahead bias and improve the apparent performance of nonsurviving funds as a group. To examine this the restriction was reduced to 18 observations (still maintaining reasonable degrees of freedom). However, this does not alter the performance comparison.

As a further possible explanation, it may be possible that nonsurviving funds yield particularly poor results towards the end of their lives (say in the final 3,6 or 12 months) and while this poor performance is responsible for their closure it is not adequately captured by the t-statistic which is an 'average' performance measure over the life of the fund. To investigate this, Figure 6.6 plots the relative performance of surviving and nonsurviving funds over the last 24 months of nonsurvivor fund lives. Specifically, for each of the last 24 months of each nonsurviving fund the difference between the nonsurviving fund's return and the cross-sectional average of the surviving funds' returns that month is calculated. For each of the last 24 months this is then averaged across nonsurviving funds in 'event time' and is plotted in Figure 6.6. If nonsurviving funds deteriorate relative to surviving funds towards the end of their lives, the plot in Figure 6.6 would trend downwards from left to right reflecting a growing gap in performance prior to termination. However, there is no evidence of this occurring.

A possible explanation for the better than expected performance of nonsurviving funds is that a large number of these funds do not close due to poor performance but are merged or are taken over, possibly even because of their strong performance and consequent attractiveness. If this is the case it clearly gives rise to an important caveat when interpreting 'nonsurvivor' data in mutual fund data sets (particularly the UK case). Indeed, Blake and Timmermann (1998), citing the UK Unit Trust Yearbook, point out that over their sample period 1972 – 1995, 89% of the funds reported as nonsurvivors were merged with other funds while only 11% were closed down over the period.

As a general point note that in making performance comparisons between classes of funds such as survivors/nonsurvivors, onshore/offshore, income/equity/small stock funds etc., the comparisons are based on funds which are not necessarily in existence at the same points in time. For example, if the concentration of onshore and offshore funds exist during different time periods, this may account for differences in performance if this is not otherwise adequately captured by the performance model. However, while this issue is noteworthy, it is highly unlikely to impact on results. First, the data set is sufficient to ensure that during all time intervals of 3 years or more (the minimum requirement) a large number of funds exist within each class of funds where a comparison is being made. Second, if the alpha measures of the different sub-classes were sensitive to the sub-time periods in which they are estimated one might expect evidence of time varying alphas. From chapter 4 the conditional alpha models found no evidence of this.

6.5 Sensitivity of Findings to Length of Fund Histories

In this section the potential for look-ahead bias arising from imposing the minimum 36 observation restriction is examined in further detail. This study conducts the bootstrap procedure for a range of alternative minimum fund histories. The results are reported in Table 6.4 where Panels A, B, C, and D present results for minimum fund histories of 18, 36, 60 and 120 observations respectively. Panel B repeats the results of Table 6.1 Panel A for ease of comparison. The final row of each panel, denoted "No. Obs." indicates the number of observations of the fund at that selected point in the performance distribution. Results relate to the unconditional Fama-French model applied to all funds.

The values of alpha, the t-statistics and the bootstrap p values across Panels A, B, C and D reveal that the minimum observation restriction does not bias performance findings upwards. On the contrary, it is evident from the upper end of the performance distributions that performance is generally higher as the minimum fund history requirement is reduced from 120 to 60 to 36 to 18. In each case, reducing the fund history restriction includes some high performing funds. Accordingly, one cannot say that the top end of the performance distribution is comprised of the same funds irrespective of the minimum observation restriction. The p values reveal that stock picking skill becomes more prevalent as minimum fund histories become shorter. In results not shown, when the restriction is set at 18 observations, there are 754 funds in the bootstrap analysis. 32 of the top 37 funds (as ranked by the t-statistic) show evidence of skill at 5% significance while 37 do so at 10%. Increasing the restriction to 36 observations leaves 675 funds in the analysis where there are 7 (12) skilled funds at 5% (10%) significance. Increasing the restriction to 60 observations leaves 573 funds where there is 1(5) skilled funds at 5% (10%) significance. Finally, when the restriction is further increased to a minimum of 120 observations there are 405 funds included but there is no evidence of stock picking talent at either 5% or 10%.

From Table 6.4, at the lower end of the performance distributions the interpretation of the bootstrap p values is unchanged across all minimum observation restrictions: poor performing funds cannot attribute their performance to chance or bad luck. The row denoted "No. Obs." reports the number of observations of the fund at the various points in the distribution. From Panel A, some of the poorest performing funds have survived for a relatively long time compared to many shorter lived funds in the right tail. In results not shown, when the minimum requirement is set at 18 observations, 6 of the 10 lowest ranked funds have more than 200 observations while 8 of the 10 funds have more than 150 observations. In contrast, of the top 10 ranked funds only 1 fund has more than 100 observations. Why any fund, particularly a long-lived fund, with such strong evidence of poor performance would be permitted to survive in a competitive market is puzzling. Kosowski et al (2004) suggest that performance measurement is a difficult task requiring, for precision, a long fund life-span and that sustained underperformance is consistent with consumers who have difficulty identifying the comparatively few funds which can beat their benchmarks. Hendricks et al (1993) suggest that sustained poor funds are those without skill which "churn" their portfolios too much, incurring high expenses which lowers net performance. Here in this study it is suggested that successful fund managers may be enticed away from their funds to manage other funds thus limiting long run superior performance in any given fund. It may be that shorter-lived funds are initially set up to exploit perceived investment opportunities but 'run out of ideas' in the longer term. Or it may be that long-lived funds may be less efficiently managed, possibly because they become quite large.

Figure 6.7 shows Kernel density plots of the actual and bootstrap distributions of the t-statistics for the alternative fund history requirements as indicated. In particular the left tails of the distributions clearly demonstrate this lack of skill relative to luck.

The findings in this section provide some indication that investors would do better selecting shorter-lived funds. These are, for the most part, funds which exist in the later part of the sample period. Therefore we might expect that if the bootstrap procedure is applied separately to earlier and later sub-sample periods we would find greater evidence of selectivity skill in the latter relative to the former. Indeed this is the case, in results not shown. Selecting periods of (i) 1975-1990 and (ii) 1991-2002 no evidence of skill was found in the first period. In the later 1991–2002 period the results are very similar to the results for the full period as reported in Table 6.1, Panel A. However, these findings raise the bigger issue for investors: is there *persistence* in performance over time among mutual funds. This is the discussion of chapter 7 and chapter 8.

However, in the bootstrap procedures that follow in this chapter the minimum 36 observation restriction is maintained. This is done confident in the evidence that this restriction does not impart a look-ahead bias.

6.6 Extensions of the Bootstrap Methodology

In chapter 5 on the bootstrap methodology, a number of extensions to the 'baseline' bootstrap procedure were discussed. These involve retaining valuable information on the fund return generating process and incorporating it into the bootstrap procedure. This includes residual serial correlation and cross-sectional (across funds) dependence among fund residuals. They also involve applying bootstrap techniques where both residuals and risk factors are resampled in the simulations. The results from implementing these extensions are now presented.

6.6.1 Serial Correlation in Fund Performance Regressions

Incorporating serial correlation involves randomly drawing residuals in the bootstrap simulations in block lengths corresponding to the suspected order of serial correlation. To determine whether the bootstrap findings presented above may be sensitive to this modified procedure, the bootstrap was repeated for a number of alternative block lengths. Findings are reported in Table 6.5. Row 4 to row 6 report the estimated p values of the t-statistic of alpha for blocks of 1, 2 and 4: hypotheses of serial correlation of orders 0, 1 and 3 respectively. From Lagrange multiplier tests, evidence of serial correlation diminishes quickly in the lag order and there is very little evidence across funds supporting serial correlation of lag order higher than 3.

Comparing these results reveals that the p values are not sensitive to the block length in the residual re-sampling procedure. At 5% significance, the p values consistently provide the same inferences for all residual sampling block lengths. (The only exceptions are for the 7^{th} and 15^{th} ranked funds when resampling in block length of size 2).

Therefore, this study finds that the conclusions drawn previously from the bootstrap methodology are not sensitive to whether the bootstrap procedure is modified to incorporate residual serial correlation of alternative relevant block lengths. This finding is consistent with that of the Kosowski et al (2004).

6.6.2 Newey-West Adjusted Standard Errors

Throughout this study all t-statistics are based on Newey-West autocorrelation and heteroscedasticity adjusted standard errors. In this section, we examine whether the conclusions presented previously are sensitive to the order of autocorrelation in the Newey-West adjustment.

Table 6.6 shows the ranked values of alpha, the associated t-statistic of these alphas, the ranked t-statistic of alpha and the bootstrap p values of the ranked t-statistics for autocorrelation orders of 1, 2 and 3 assumed for all funds. A bootstrap procedure was also implemented where a Lagrange Multiplier test was used to select the order of autocorrelation for each fund individually in the Newey-West adjustment. The results from this procedure are very similar to the result shown for autocorrelation of order 3.

The p values in the final three rows of Table 6.6 show that results are not sensitive to the autocorrelation order in the Newey-West adjustment. At 5% significance the p values at all selected points in the performance distribution are consistent in interpretation between all three autocorrelation orders.

Of course, as shown in Table 6.6, the t-statistics of alpha change with the alternative Newey-West autocorrelation adjustments. However, in results not shown, the top 20 funds ranked by the t-statistic are common in all three Newey-West adjustments.

Therefore, this study finds that the conclusions drawn previously from the bootstrap methodology are robust to the order of autocorrelation adjustment in the Newey-West corrected standard errors.

6.6.3 Alternative Bootstrap Procedures

From chapter 5 on methodology, the alternative bootstrap resampling methodologies were (i) 'residual-only' resampling, (ii) 'independent factor-residual' resampling and (iii) 'pairwise factor-residual' resampling. It was noted that the sampling variability of the bootstrap t-statisticss, and therefore the p values, may be sensitive to the choice of bootstrap procedure. The set of bootstrap findings presented so far are derived from a residual-only resampling. To examine whether these are indeed sensitive to the choice of bootstrap procedure, this study applies all three bootstrap procedures here. Results are reported in Table 6.7.

Row 4 of Table 6.7, denoted "p-tstat(i)", reports the p values of the t-statistic under a residual-only resampling procedure. Row 5, denoted "p-tstat(ii)", reports the equivalent p values under independent factor-residual resampling while row 6, denoted "p-tstat(iii)", presents the p values from a pairwise factor-residual resamplng procedure. In particular, this final procedure captures possible heteroscedastic features which may be present in the underlying data, where the heteroscedasticity is related to the factors.

Comparing the three sets of p values in Table 6.7 it is clear that the bootstrap results are not sensitive to the choice of resampling procedure. At 5% significance, at all selected points in the performance distribution the three sets of p values are consistent in interpretation, (the only exception is p-stat(iii) of the 10^{th} ranked fund). This implies that the variability of the bootstrap t-statistics is not substantially different between these three bootstrap procedures. This finding of robustness in the results lends support to the interpretation and conclusions already discussed.

In results not shown, this study also applied all three bootstrap resampling procedures in separate bootstrap analyses of mutual funds (i) by investment objective and (ii) by onshore/offshore location. The interpretation of the resulting p values was also unaltered from the results already reported for the residual-only resampling procedure in Table 6.2 and Table 6.3 respectively.

6.6.4 Cross-Sectional Dependence

In addition to time series dependence in residuals within funds, there may be a cross-sectional dependence between residuals across funds. In this study, the (absolute) average correlation coefficient between fund residuals (from the unconditional model) is 0.33[2]. This section implements a 'cross-sectional bootstrap' procedure in which within each one of the 1,000 simulations the realized time index of residuals drawn is kept constant across funds. It examines the sensitivity of findings to this refinement to the procedure.

[2] This figure is calculated for a subset of 311 funds which exist at the same time.

However, as discussed in detail in chapter 5 on methodology, implementing this procedure presents a difficulty because the funds do not all exist at the same time. To maintain the same time sequence of residuals across funds would be likely to involve selecting residuals for many funds at time periods when these funds do not exist. To surmount this difficulty, in this study a sub-sample period is selected (January 1988 to December 1998) in which as many funds as possible (311) have a complete return history.

The results of this cross-sectional bootstrap procedure are presented in Table 6.8. Panel A reports the results from a 'residual-only' resampling. Row 4, denoted "p-tstat(i)", shows the bootstrap p values of the t-statistic for the 'baseline' bootstrap procedure where in each of the 1,000 simulations the time sequence of residuals drawn from each fund is allowed to vary across funds. This is an identical bootstrap procedure as applied previously except in this case the results relate only to the subset of funds in the sub-sample period. Row 5 of Panel A records the bootstrap p values of the t-statistic where in each of the 1,000 simulations the time sequence of residuals drawn from each fund is kept constant across funds, ie a 'cross-sectional residual-only' bootstrap. Both sets of p values are consistent in indicating that over the sub-sample period funds at the upper end of the performance distribution do not beat luck. At the lower end of the distribution both sets of the p values indicate that the poor performance is worse than bad luck. The central point here is that both sets of p values yield the same interpretation. This is supporting evidence that the bootstrap results described previously are not sensitive to the choice between baseline residual-only or cross-sectional residual-only bootstrap procedures.

In Panel B of Table 6.8, bootstrap p values of the t-stat from an 'independent factor-residual' resampling procedure and 'cross-sectional independent factor-residual' resampling procedure are presented. Row 4 of Panel B reports p values where in each of the 1,000 simulations the realized time sequence of residuals and the realized time sequence of factors drawn from each fund are allowed to vary across funds. Row 5 reports p values where in each of the 1,000 simulations the realized time sequence of residuals and the realized time sequence of factors drawn from each fund are kept constant across funds[3]. Again without exception, at every selected point in the performance distribution both sets of p values provide the same conclusion with regards to how both good and bad funds perform relative to luck.

[3] In each of the 1,000 bootstrap simulations the realized time sequence of residuals is not the same as the realized time sequence of factors.

Finally, Panel C of Table 6.8 reports the p values from a 'pairwise factor-residual' resampling and a 'cross-sectional pairwise factor-residual' resampling bootstrap procedure in row 4 and row 5 respectively. In both procedures residuals and factors are drawn in pairs at identical time periods but in addition in the case of a cross-sectional factor-residual resampling, the time sequence of pairs drawn are kept constant across funds within each of the 1,000 bootstrap simulations. The results in Panel C reveal once again that the p values from both bootstrap procedures provide the same conclusion regarding fund actual performance relative to luck.

Therefore, from Panels A, B and C in Table 6.8, it is evident that the bootstrap results are not sensitive to cross-sectional correlations among the idiosyncratic component of returns in funds. This robustness supports the overall conclusions from the bootstrap methodology presented so far regarding fund performance and luck.

While it may be interesting to implement a cross-sectional bootstrap procedure separately for funds by investment objective, the requirement that funds have a complete return history means there are too few funds available within each investment style to produce reliable conclusions.

6.7 Conclusion

Overall, the findings in this study point to the existence of genuine stock picking ability among a small number of top performing UK equity mutual fund managers, ie performance which is not solely due to luck or cannot be explained by random sampling variation in the performance estimator. However, the skillful funds are not simply the highest funds ranked either by alpha or its t-statistic. At the opposite end of the performance scale, the analysis strongly rejects the hypothesis that poor performing funds are merely unlucky. Controlling for different investment objectives among funds, it is found that equity income funds show more evidence of stock picking ability while such ability is not found among general equity funds or small stock funds. Furthermore, relative to luck, onshore funds are found to outperform while offshore funds generally underperform. To improve the statistical reliability of the performance estimates, this study imposes a restriction that funds have existed for a minimum of 3 years in order to be included in the analysis. However, a sensitivity analysis reveals that this does not impart a look-ahead bias. Conclusions from the baseline bootstrap procedure are remarkably robust with respect to block bootstrapping, serial correlation adjustments, factor-residual resampling and cross-sectional residual resampling.

Table 6.1: Statistical Significance of Mutual Fund Performance

Table 6.1 presents bootstrap statistics for the full sample of mutual funds including all investment objectives for each of the three performance measurement models selected in chapter 4. Panel A reports bootstrap statistics from the unconditional Fama and French (three-factor) model. Panel B presents results from the conditional beta Fama and French model. Panel C relates to the Fama and French conditional alpha-beta model. $Z3_t$ is the market dividend yield. The first row in each panel reports alpha in percent per month at various points and percentiles in the performance distribution sorted from worst fund (min) to best fund (max). The second row reports the t-statistic of the alphas in row 1. Row 3 contains the t-statistics of alpha sorted from lowest (min) to highest (max). Row 4 reports the bootstrap p values of the t-statistics in row 3, (ie where the t-statistic of alpha is the performance measure). p values are based on 1,000 bootstrap resamples. Both actual (unmodified) and bootstrap t-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors. Row 5 describes the investment objective of the funds: 1 = equity income fund, 2 = general equity fund, 3 = small stock fund. Row 6 indicates whether the fund is a survivor or non-survivor fund: 1 = surviving fund, 0 = non-surviving fund. Row 7 describes the location from where the fund is operated: 1 = onshore mainland UK, 0 = offshore fund. The Style, Survival and Location identifiers correspond to funds as ranked by the t-statistic. Results are restricted to funds with a minimum of 36 observations.

Panel A : Unconditional Model

$$(R_i - r_f)_t = \alpha_i + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SMB_t + \beta_{3i}HML_t + \varepsilon_{it}$$

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5.max	3.max	2 max	max
Alpha	-0.901	-0.661	-0.427	-0.350	-0.122	0.029	0.236	0.395	0.487	0.490	0.543	0.593	0.617	0.686	0.716	0.757	1.447	2.235
t-alpha	-2.532	-3.199	-1.563	-4.952	-0.912	0.257	0.434	1.859	1.948	2 677	2.672	1.450	1.212	2.991	1.742	1 902	3.365	2.226
t-stat	-5.358	-4.278	-3.045	-2.509	-0 873	0.212	1.177	1.757	1.955	2.023	2.282	2.501	2.545	2.677	2.777	2.991	3.365	3.389
p-tstat	0.009	0	0	0	0	1	0.978	0.491	0.447	0.284	0.070	0.020	0.038	0.094	0.157	0 232	0.128	0.437
Style	2	2	2	3	1	2	3	1	2	2	2	1	1	2	1	3	3	2
Survival	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1	0	0	1
Location	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Panel B : Conditional-Beta Model

$$(R_i - r_f)_t = \alpha_i + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}[Z3_{t-1}(R_m - r_f)] + \varepsilon_{it}$$

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5 max	3.max	2 max	max
Alpha	-0.733	-0.622	-0.395	-0.331	-0.094	0.049	0 236	0.426	0.542	0.552	0.580	0.600	0.641	0.690	0.777	0.967	1.431	2.309
t-alpha	-2.104	-2.474	-2.072	-1.578	-0.749	0.381	1.834	0.599	0.956	2.268	3.161	0.924	1.088	1.590	1.849	1.602	3.036	2.305
t-stat	-5.333	-4.160	-2.928	-2.376	-0.722	0.354	1.394	1.856	2.089	2.189	2.305	2.456	2 622	2.808	3.036	3 161	3.352	4.036
p-tstat	0.012	0	0	0	0	1	0.147	0.034	0.093	0.022	0.048	0.045	0.017	0.017	0.022	0.093	0.153	0.100
Style	2	2	2	2	3	1	2	2	2	3	3	2	2	2	3	2	3	2
Survival	1	1	1	1	0	1	0	1	1	1	1	1	0	1	0	1	0	1
Location	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Panel C : Conditional Alpha-Beta Model

$$(R_i - r_f)_t = \alpha_{0i} + \alpha_{1i}Z3_{t-1} + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SMB_t + \beta_{3i}\,HML_t + \beta_{4i}[Z3_{t-1}\,(R_m - r_f)_t] + \beta_{5i}[[Z3_{t-1}\,(SMB)_t] + \beta_{6i}[Z3_{t-1}\,(HML)_t] + \varepsilon_{it}$$

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5.max	3.max	2.max	max
Alpha	-5.769	-4.514	-1.012	-0.497	-0.131	0.058	0.441	0.812	1.139	1.184	1.388	1.485	1.703	1.892	2.095	2.163	2.528	3.801
t-alpha	-1.299	-2.112	-0.720	-3.129	-1.332	0.450	3.351	0.342	0.672	1.265	0.461	1.411	2.580	3.784	1.589	2.818	3.732	5.081
t-stat	-9.942	-4.105	-2.870	-2.251	-0.784	0.307	1.313	1.894	2.459	2.499	2.725	2.869	3.063	3.229	3.537	3.732	3.784	5.081
p-tstat	0	0.001	0	0	0	1	0.881	0.211	0.001	0.002	0.002	0.002	0.002	0.008	0.006	0.055	0.169	0.069
Style	2	2	2	2	2	2	2	2	2	2	3	3	1	2	3	3	2	3
Survival	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	0
Location	0	0	0	0	0	1	1	0	1	1	1	1	1	1	0	1	1	1

Table 6.2: Statistical Significance of Mutual Fund Performance by Investment Objective

Table 6.2 presents bootstrap statistics for the separate bootstrap analyses of mutual funds categorized by investment objectives as indicated in each panel. All results pertain to the unconditional Fama and French three-factor model. The first row in each panel reports alpha in percent per month at various points and percentiles in the performance distribution sorted from worst fund (min) to best fund (max). The second row reports the t-statistic of the alphas in row 1. Row 3 contains the t-statistics of alpha sorted from lowest (min) to highest (max). Row 4 reports the bootstrap p values of the t-statistics in row 3. p values are based on 1,000 bootstrap resamples. Both actual (unmodified) and bootstrap t-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors. Row 5 indicates whether the fund is a survivor or non-survivor fund: 1 = surviving fund, 0 = non-surviving fund. Finally, row 6 describes the location from where the fund is operated: 1 = onshore mainland UK, 0 = offshore fund. Funds have a minimum of 36 observations.

Unconditional Three Factor Model : $(R_i - r_f)_t = \alpha_i + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SML_t + \beta_{3i}HML_t + \varepsilon_{it}$

Panel A : Equity Income

	min	5.min	min5%	min10%	min20%	min40%	max30%	max20%	max10%	10max	7max	5.max	3.max	2.max	max
Alpha	-0.358	-0.314	-0.223	-0.163	-0.119	-0.020	0.102	0.150	0.229	0.279	0.315	0.358	0.431	0.470	0.478
t-alpha	-2 679	-1.864	-1.049	-0.596	-2.010	-0.163	1.034	1.423	2.245	2.634	1.909	0.615	2.545	0.977	2.777
t-stat	-2.954	-2.166	-1.863	-1.488	-0.912	-0.197	0 685	0.977	1.667	1.740	2.245	2.501	2.545	2.634	2.777
p-tstat	0.337	0.165	0.179	0.141	0.398	0.721	0 111	0.171	0.007	0.107	0.008	0.005	0.067	0.181	0 415
Survival	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1
Location	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1

Panel B : General Equity

	min	5min	min5%	min10%	min20%	min40%	max30%	max20%	max10%	10max	7max	5max	3max	2max	max
Alpha	-0.741	-0 650	-0.420	-0.329	-0.245	-0.138	0.002	0.075	0.193	0.514	0.593	0.617	0.706	0.724	0.757
t-alpha	-3.043	-2.971	-2.773	-4.180	-1.455	-0.489	0.027	0.667	1.145	2.127	1.450	1.212	1.009	1.323	1.902
t-stat	-5.359	-4.190	-3.118	-2.575	-2.009	-1.084	0.011	0.509	1.045	2.023	2.282	2.672	2.776	2.965	3.389
p-tstat	0.014	0	0	0	0	0	1	1	1	0.643	0.422	0.083	0.285	0.296	0.301
Survival	1	1	1	0	1	1	1	0	0	1	0	0	1	1	1
Location	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1

Panel C : Smaller Companies

	min	5.min	min5%	min10%	min20%	min40%	max30%	max20%	max10%	10max	7max	5.max	3.max	2.max	max
Alpha	-0.901	-0.520	-0.518	-0.443	-0.380	-0.250	0.003	0.128	0.341	0.374	0.530	0.624	0.716	1.447	2.235
t-alpha	-2.532	-3.077	-2.181	-2.702	-2.112	-1.135	0.008	0.372	1.571	1.045	1.699	1.397	1.742	3.365	2.226
t-stat	-4.953	-3.117	-3.095	-2.772	-2.464	-1.128	0.008	0.397	1.286	1.356	1.610	1.742	2.226	2.991	3.365
p-tstat	0 001	0	0	0	0	0	1	1	0 647	0.624	0.490	0.579	0.256	0.040	0.096
Survival	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0
Location	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1

Table 6.3: Statistical Significance of Mutual Fund Performance by Fund Location

Table 6.3 presents bootstrap statistics for the separate bootstrap analyses of mutual funds by investment location as indicated in each panel. All results pertain to the unconditional Fama and French three-factor model and are estimated over all investment objectives. The first row in each panel reports alpha in percent per month at various points and percentiles in the performance distribution sorted from worst fund (min) to best fund (max). The second row reports the t-statistic of the alphas in row 1. Row 3 contains the t-statistics of alpha sorted from lowest (min) to highest (max). Row 4 reports the bootstrap p values of the t-statistics in row 3. p values are based on 1,000 bootstrap resamples. Both actual (unmodified) and bootstrap t-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors. Row 5 describes the investment style of the fund: 1 = equity income fund, 2 = general equity fund 3 = small stock fund. Row 6 indicates whether the fund is a survivor or non-survivor fund: 1 = surviving fund, 0 = non-surviving fund. Results are restricted to funds with a minimum of 36 observations.

Unconditional Three Factor Model : $(R_i - r_f)_t = \alpha_i + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SML_t + \beta_{3i}\,HML_t + \varepsilon_{it}$

Panel A : Onshore UK Funds

	min	5.min	min5%	min10%	min20%	min40%	max20%	max10%	20max	15max	10max	5.max	3 max	2.max	max
Alpha	-0.661	-0.559	-0.413	-0.328	-0.199	-0.100	0.114	0.236	0.471	0.504	0.561	0.716	0.757	1.447	2.235
t-alpha	-3.199	-2 881	-1.825	-2.474	-2.327	-0.952	0.947	2.041	1.078	1.284	1.013	1.742	1.902	3.365	2.226
t-stat	-5.359	-4 119	-2.915	-2.464	-1.681	-0.739	0.755	1.287	2.023	2.282	2.545	2.777	3.112	3.365	3.389
p-tstat	0.011	0	0	0	0	0	0.970	0 615	0.057	0.010	0 006	0.073	0.107	0 063	0.306
Style	2	2	3	3	2	1	1	2	2	2	1	1	3	3	2
Survival	1	1	0	1	1	1	0	1	1	0	1	1	0	0	1

Panel B : Offshore Funds

	min	5min	min5%	min10%	min20%	min40%	max20%	max10%	20max	15max	10max	5max	3max	2max	max
Alpha	-0.901	-0.609	-0.526	-0.420	-0.311	-0.241	0.003	0.172	0.024	0.124	0.284	0.487	0.593	0.632	0.706
t-alpha	-2.532	-2.415	-3.073	-2.773	-2.627	-2.061	0.008	0.364	0.063	0.356	1.390	1.948	1.450	1.040	1.009
t-stat	-4.278	-3.751	-3.521	-2.953	-2.330	-1.454	0.008	0.489	0.067	0.429	0.881	1.101	1.390	1.450	1.948
p-tstat	0.012	0	0	0	0	0	1	1	1	1	1	1	0.998	1	0.985
Style	2	3	2	2	2	2	3	2	2	2	2	3	2	2	2
Survival	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 6.4: Statistical Significance of Performance by Length of Fund Histories

Table 6.4 reports bootstrap statistics for alternative minimum number of observations restrictions as indicated in each panel. All results pertain to the unconditional Fama and French three-factor factor model and are estimated over all investment objectives. In each case fund histories are contiguous. The first row in each panel reports alpha in percent per month at various points and percentiles in the performance distribution sorted from worst fund (min) to best fund (max). The second row reports the t-statistic of the alphas in row 1. Row 3 contains the t-statistics of alpha sorted from lowest (min) to highest (max). Row 4 reports the bootstrap p values of the t-statistics in row 3. p values are based on 1,000 bootstrap resamples. Both actual (unmodified) and bootstrap t-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors. Row 5 describes the investment style or objective of the fund: 1 = equity income fund, 2 = general equity fund, 3 = small stock fund. Row 6 indicates whether the fund is a survivor or non-survivor fund: 1 = surviving fund, 0 = non-surviving fund. Row 7 describes the location from where the fund is operated: 1 = onshore mainland UK, 0 = offshore fund. Row 8 reports the number of monthly observations of the fund during the sample period. The Sample period is 1975M4:2002M12 (333 observations).

All Investment Objectives : Unconditional Three Factor Model : $(R_i - r_f)_t = \alpha_i + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SML_t + \beta_{3i}\,HML_t + \varepsilon_{it}$

Panel A : Minimum Fund History of 18 observations

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5 max	3.max	2.max	max
Alpha	-3.057	-1 451	-0 469	-0.363	-0.127	0.040	0.293	0.472	0 603	0.624	0.716	0.803	1.079	1.119	1.447	2.138	2.216	2.235
t-alpha	-1.991	-0 651	-2.681	-2.650	-1.453	0.335	1.628	1.356	1 085	1.397	1.742	2.270	4.283	1.013	3.365	4.530	2.800	2.223
t-stat	-5.359	-4.278	-3 033	-2.474	-0.788	0.303	1.326	1.815	2.282	2.458	2 634	2.776	2.800	2.965	3.083	3.389	4.283	4.530
p-tstat	0.043	0	0	0	0	1	0.543	0.136	0	0	0	0.001	0.007	0 012	0.039	0.063	0.012	0.074
Style	2	2	2	2	2	2	2	2	2	2	1	2	2	2	1	2	2	2
Survival	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0	1	1	1
Location	1	0	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1
No. Obs	255	171	52	197	84	21	159	21	144	33	255	159	33	53	33	212	31	23

Panel B : Minimum Fund History of 36 observations

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5.max	3.max	2 max	max
Alpha	-0.901	-0.661	-0.427	-0.350	-0.122	0.029	0.236	0.395	0.487	0.490	0.543	0.593	0.617	0.686	0.716	0.757	1.447	2.235
t-alpha	-2.532	-3.199	-1.563	-4.952	-0.912	0.257	0.434	1.859	1.948	2.677	2.672	1.450	1.212	2.991	1.742	1.902	3.365	2.226
t-stat	-5.358	-4.278	-3.045	-2.509	-0.873	0.212	1.177	1.757	1.955	2.023	2.282	2.501	2.545	2.677	2.777	2.991	3.365	3.389
p-tstat	0.009	0	0	0	0	1	0.978	0.491	0.447	0.284	0.070	0.020	0.038	0.094	0 157	0.232	0.128	0 437
Style	2	2	2	3	1	2	3	1	2	2	2	1	1	2	1	3	3	2
Survival	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1	0	0	1
Location	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
No. Obs	255	171	228	186	228	228	93	218	66	148	144	70	178	38	110	63	48	212

Panel C : Minimum Fund History of 60 observations

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5.max	3 max	2.max	max
Alpha	-0.901	-0.609	-0.420	-0.330	-0.123	0.009	0.195	0.310	0.431	0.396	0 456	0.472	0.504	0.530	0.593	0.686	0.716	0.724
t-alpha	-2.532	-2.415	-2.773	-2.406	-1.330	0.056	1.472	1.051	2.545	1 860	1.726	1.356	1.284	1.699	1.450	2.991	1.742	1.323
t-stat	-5.359	-4.190	-3.078	-2.573	-0.981	0.075	1.177	1.671	1.909	1 788	2.023	2.282	2.403	2.545	2 672	2.777	2.991	3.389
p-tstat	0.010	0	0	0	0	1	0.963	0.558	0.549	0.791	0.429	0.079	0.059	0.110	0.165	0.378	0.349	0.307
Style	2	2	3	2	1	1	3	1	1	3	2	2	1	1	2	1	3	2
Survival	1	1	0	1	0	1	0	1	1	1	1	0	1	1	0	1	0	1
Location	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
No. Obs.	255	176	137	182	228	330	93	279	166	111	148	144	185	178	70	110	63	212

Panel D : Minimum Fund History of 120 observations

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5.max	3.max	2.max	max
Alpha	-0.661	-0.520	-0.420	-0.349	-0.130	0.001	0.126	0.233	0.309	0.241	0.281	0.315	0.382	0 412	0 440	0.507	0.530	0 606
t-alpha	-3.199	-3.078	-2.773	-3.941	-0.952	0 005	0.951	1.671	1.031	2.522	1.530	1.909	2.403	3.389	0 806	2.776	1.699	2.292
t-stat	-5.359	-4.190	-3.147	-2.609	-1.084	0.005	0 956	1.429	1.758	1.472	1.682	1.860	2.023	2.292	2.522	2.634	2.776	3.389
p-tstat	0.013	0	0	0	0	1	1	0.987	0.866	0.981	0.873	0.716	0.479	0.220	0.136	0.307	0.360	0.185
Style	2	2	2	3	2	1	2	1	1	2	1	2	2	2	1	1	2	2
Survival	1	1	0	1	1	0	0	1	1	1	1	1	1	1	0	1	1	1
Location	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
No. Obs	255	176	128	140	331	169	228	210	218	177	177	192	148	227	207	255	159	212

Table 6.5: Sensitivity Analysis of the Bootstrap Results – Residual Resampling in Blocks

Table 6.5 reports bootstrap statistics for alternative block lengths in the residuals re-sampling. All results pertain to the unconditional Fama-French 3 factor model and are estimated over all investment objectives. Row 1 reports alpha in percent per month at various points and percentiles in the performance distribution sorted from worst fund (min) to best fund (max). Row 2 reports the t-statistic of the alphas in row 1. Row 3 contains the t-statistics of alpha sorted from lowest (min) to highest (max). Rows 4 to 6 report the bootstrap p values of the t-statistics in row 3 for the various block lengths as indicated in parentheses. Both actual (unmodified) and bootstrap t-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors. The bootstrap p values are based on 1,000 re-samples. Results are restricted to funds with a minimum of 36 observations.

All Investment Objectives : Unconditional Three Factor Model : $(R_i - r_f)_t = \alpha_i + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SML_t + \beta_{3i}\,HML_t + \varepsilon_{it}$

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5.max	3.max	2.max	max
Alpha	-0.901	-0.661	-0.427	-0.350	-0.122	0.029	0.236	0.395	0.487	0.490	0.543	0.593	0.617	0.686	0.716	0.757	1.447	2.235
t-alpha	-2.532	-3.199	-1.563	-4.952	-0.912	0.257	0.434	1.859	1.948	2.677	2.672	1.450	1.212	2.991	1.742	1.902	3.365	2.226
t-stat	-5.358	-4.278	-3.045	-2.509	-0.873	0.212	1.177	1.757	1.955	2.023	2.282	2.501	2.545	2.677	2.777	2.991	3.365	3.389
p-tstat(1)	0.009	0	0	0	0	1	0.978	0.491	0.447	0.284	0.070	0.020	0.038	0.094	0.157	0.232	0.128	0.437
p-tstat(2)	0.010	0	0	0	0	1	0.960	0.308	0.264	0.158	0.020	0.003	0.011	0.043	0.099	0.167	0.094	0.399
p-tstat(4)	0.032	0.002	0	0	0	1	0.973	0.426	0.394	0.241	0.059	0.016	0.039	0.104	0.198	0.284	0.177	0.534

Table 6.6: Sensitivity Analysis of the Bootstrap Results – Newey-West Autocorrelation Adjustment

Table 6.6 reports bootstrap p values for alternative autocorrelation lag orders in the Newey-West adjusted t-statistics. All results pertain to the unconditional Fama-French 3 factor model and are estimated over all investment objectives. Row 1 reports alpha in percent per month at various points and percentiles in the performance distribution sorted from worst fund (min) to best fund (max). Rows 2 to 4 report the t-statistic of the alphas in row 1 for various autocorrelation lag orders as indicated. Rows 5 to 7 report the t-statistics of alpha sorted from lowest (min) to highest (max) for the alternative lag orders. Rows 8 to 10 report the bootstrap p values of these sorted t-statistics. p values are based on 1,000 bootstrap resamples. Results are restricted to funds with a minimum of 36 observations.

All Investment Objectives : Unconditional Three Factor Model : $(R_i - r_f)_t = \alpha_i + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SML_t + \beta_{3i}HML_t + \varepsilon_{it}$

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5 max	3.max	2 max	max
Alpha	-0.901	-0.661	-0.427	-0.350	-0.122	0.029	0.236	0.395	0.487	0.490	0.543	0.593	0.617	0 686	0.716	0.757	1.447	2.235
t-alpha(1)	-2.532	-3.199	-1.563	-4 952	-0.912	0.257	0.434	1.859	1.948	2.677	2.672	1.450	1.212	2.991	1.742	1.902	3.365	2.226
t-alpha(2)	-2.349	-3.105	-1.511	-5.123	-0.915	0.258	0.465	1.796	2.035	2.923	2.628	1.442	1.377	2.781	1.623	2.037	3.005	2.024
t-alpha(3)	-2.291	-2.979	-1.476	-5.317	-0.910	0.268	0.468	1.717	1.937	3.147	2.709	1.422	1.552	2.646	1.538	1.992	2.803	1.958
t-stat(1)	-5.358	-4.278	-3.045	-2.509	-0.873	0.212	1.177	1.757	1.955	2.023	2.282	2.501	2.545	2.677	2.777	2.991	3.365	3.389
t-stat(2)	-5.636	-4.242	-3.076	-2.563	-0.908	0.217	1.196	1.650	2.037	2.060	2.258	2.482	2.581	2.640	2.781	3.005	3.081	3.490
t-stat(3)	-5.954	-4.399	-3.152	-2.628	-0.917	0.214	1.189	1.646	1.980	1.992	2.263	2.508	2.614	2.709	2.803	3.148	3.158	3.429
p-tstat(1)	0.009	0	0	0	0	1	0.978	0.491	0.447	0.284	0.070	0.020	0.038	0.094	0.157	0.232	0.128	0.437
p-tstat(2)	0.006	0	0	0	0	1	0.967	0.713	0.179	0.181	0.098	0.027	0.025	0.118	0.152	0.218	0.381	0.352
p-tstat(3)	0.002	0	0	0	0	1	0.970	0.730	0.356	0.393	0.086	0.018	0.021	0.062	0.131	0.086	0.280	0.401

Table 6.7: Sensitivity Analysis of the Bootstrap Results – Alternative Resampling Procedures

Table 6.7 reports bootstrap p values of the t-statistic of alpha for alternative bootstrap resampling procedures. Row 1 reports alpha in percent per month at various points and percentiles in the performance distribution sorted from worst fund (min) to best fund (max). Row 2 reports the t-statistic of the alphas in row 1. Row 3 contains the t-statistics of alpha sorted from lowest (min) to highest (max). Rows 4, 5 and 6 report the p values of the t-statistic in row 3 for the (i) residual-only, (ii) independent factor-residual and (iii) pairwise factor-residual resampling procedures respectively. All results pertain to the unconditional FF 3 factor model and are estimated over all investment objectives. The bootstrap p values are based on 1,000 re-samples. Both actual (unmodified) and bootstrap t-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors. Results are restricted to funds with a minimum of 36 observations.

All Investment Objectives : Unconditional Three Factor Model : $(R_i - r_f)_t = \alpha_i + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SML_t + \beta_{3i}HML_t + \varepsilon_{it}$

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5.max	3.max	2.max	max
Alpha	-0.901	-0.661	-0.427	-0.350	-0.122	0.029	0.236	0.395	0.487	0.490	0.543	0.593	0.617	0.686	0.716	0.757	1.447	2.235
t-alpha	-2.532	-3.199	-1.563	-4.952	-0.912	0.257	0.434	1.859	1.948	2.677	2.672	1.450	1.212	2.991	1.742	1.902	3.365	2.226
t-stat	-5.358	-4.278	-3.045	-2.509	-0.873	0.212	1.177	1.757	1.955	2.023	2.282	2.501	2.545	2.677	2.777	2.991	3.365	3.389
p-tstat(i)	0.009	0	0	0	0	1	0.978	0.491	0.447	0.284	0.070	0.020	0.038	0.094	0.157	0.232	0.128	0.437
p-tstat(ii)	0.005	0	0	0	0	1	0.981	0.511	0.456	0.306	0.076	0.011	0.042	0.092	0.166	0.230	0.121	0.437
p-tstat(iii)	0.022	0	0	0	0	0	0.986	0.578	0.537	0.370	0.110	0.028	0.068	0 131	0.264	0.352	0.206	0 540

Table 6.8: Sensitivity Analysis of the Bootstrap Results – Cross-Sectional Bootstrap

Table 6.8 reports bootstrap p values of the t-statistic of alpha for the (i) baseline and (ii) cross-sectional bootstrap resampling procedures, denoted "p-tstat(i)" and "p-tstat(ii) respectively. Panels A, B and C report both these sets of statistics for the residual-only, independent factor-residual and pairwise factor-residual resampling procedures as indicated. The bootstrap p values are based on 1,000 re-samples of 311 funds (all investment objectives) with full return histories over the sub-sample period 1988M1:1998M12. Both actual (unmodified) and bootstrap t-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors.

All Investment Objectives : Unconditional Three Factor Model : $(R_i - r_f)_t = \alpha_i + \beta_{1i}(R_m - r_f)_t + \beta_{2i}SML_t + \beta_{3i}\,HML_t + \varepsilon_{it}$

Panel A: Residual-Only Resampling

	min	5.min	min5%	min10%	min20%	min40%		max30%	max20%	max10%	max5%	10max	5 max	3.max	2 max	max
Alpha	-0.816	-0.508	-0.367	-0.319	-0.244	-0.172		-0.078	-0.028	0.037	0.103	0.155	0.329	0.400	0.406	0.609
t-alpha	-3.075	-2.836	-2.030	-1.296	-2.941	-1.743		-0.559	-0.333	0.238	0.872	0.986	1.368	2.490	2.328	0.691
t-stat	-4.942	-4.204	-3.269	-2.763	-2.216	-1.634		-0.660	-0.297	0.278	0.686	0.956	1.688	2 328	2.490	3.046
p-tstat(i)	0.288	0	0	0	0	0		1	1	1	1	1	1	0.780	0.762	0.440
p-tstat(ii)	0.279	0	0	0	0	0.004		0 995	0.993	0.982	0.975	0.952	0.718	0.458	0 415	0.280

Panel B: Independent Factor-Residual Resampling

	min	5.min	min5%	min10%	min20%	min40%		max30%	max20%	max10%	max5%	10max	5.max	3 max	2.max	max
Alpha	-0.816	-0.508	-0.367	-0.319	-0.244	-0.172		-0 078	-0.028	0.037	0.103	0.155	0.329	0.400	0.406	0.609
t-alpha	-3.075	-2.836	-2.030	-1.296	-2.941	-1.743		-0.559	-0.333	0.238	0.872	0.986	1.368	2.490	2.328	0.691
t-stat	-4.942	-4.204	-3.269	-2.763	-2.216	-1.634		-0.660	-0.297	0.278	0.686	0.956	1.688	2.328	2.490	3.046
p-tstat(i)	0.260	0	0	0	0	0		1	1	1	1	1	1	0.760	0.741	0.461
p-tstat(ii)	0.270	0	0	0	0.001	0.001		0.997	0.993	0.981	0.971	0.950	0.726	0.420	0.409	0.251

Panel C: Pairwise Factor-Residual Resampling

	min	5.min	min5%	min10%	min20%	min40%		max30%	max20%	max10%	max5%	10max	5.max	3.max	2 max	max
Alpha	-0.816	-0.508	-0.367	-0.319	-0.244	-0.172		-0.078	-0.028	0.037	0.103	0.155	0.329	0.400	0.406	0.609
t-alpha	-3.075	-2.836	-2 030	-1.296	-2.941	-1.743		-0.559	-0.333	0.238	0.872	0.986	1.368	2.490	2.328	0.691
t-stat	-4.942	-4.204	-3.269	-2.763	-2.216	-1.634		-0.660	-0.297	0.278	0.686	0 956	1.688	2.328	2.490	3.046
p-tstat(i)	0.379	0.001	0	0	0	0		1	1	1	1	1	1	0.781	0.751	0.466
p-tstat(ii)	0.342	0	0	0 001	0.001	0		0.995	0.994	0.984	0.978	0.959	0.725	0.411	0 422	0.262

Figure 6.1: Kernel Density Estimates of the Actual and Bootstrap distribution of t-statistics

Figure 6.1 presents Kernel densities of the actual and bootstrap distributions of the t-statistics of alpha from the unconditional Fama-French 3 factor model over the full sample of mutual funds. t-statistics are Newey-West adjusted. Funds with a minimum of 36 observations are used. The plots are generated using a Gaussian Kernel.

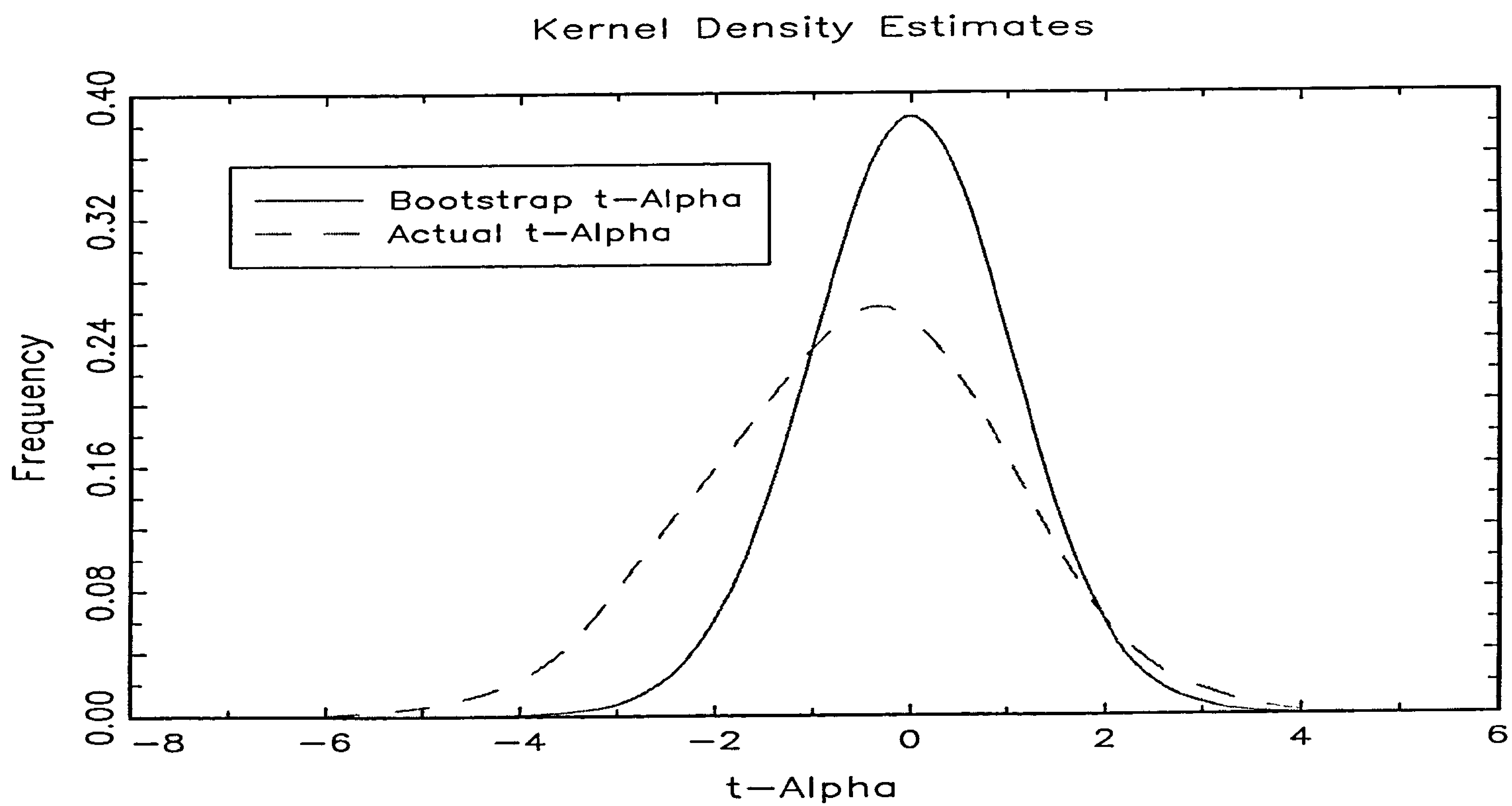


Figure 6.2a: Histograms of Bootstrap t-statistics (Upper End of the Distribution)

Figure 6.2a shows histograms of the bootstrap t-statistics of alpha from the unconditional model, at various points in the upper end of the performance distribution. The actual t-statistics at each ranking is indicated by the vertical dashed line.

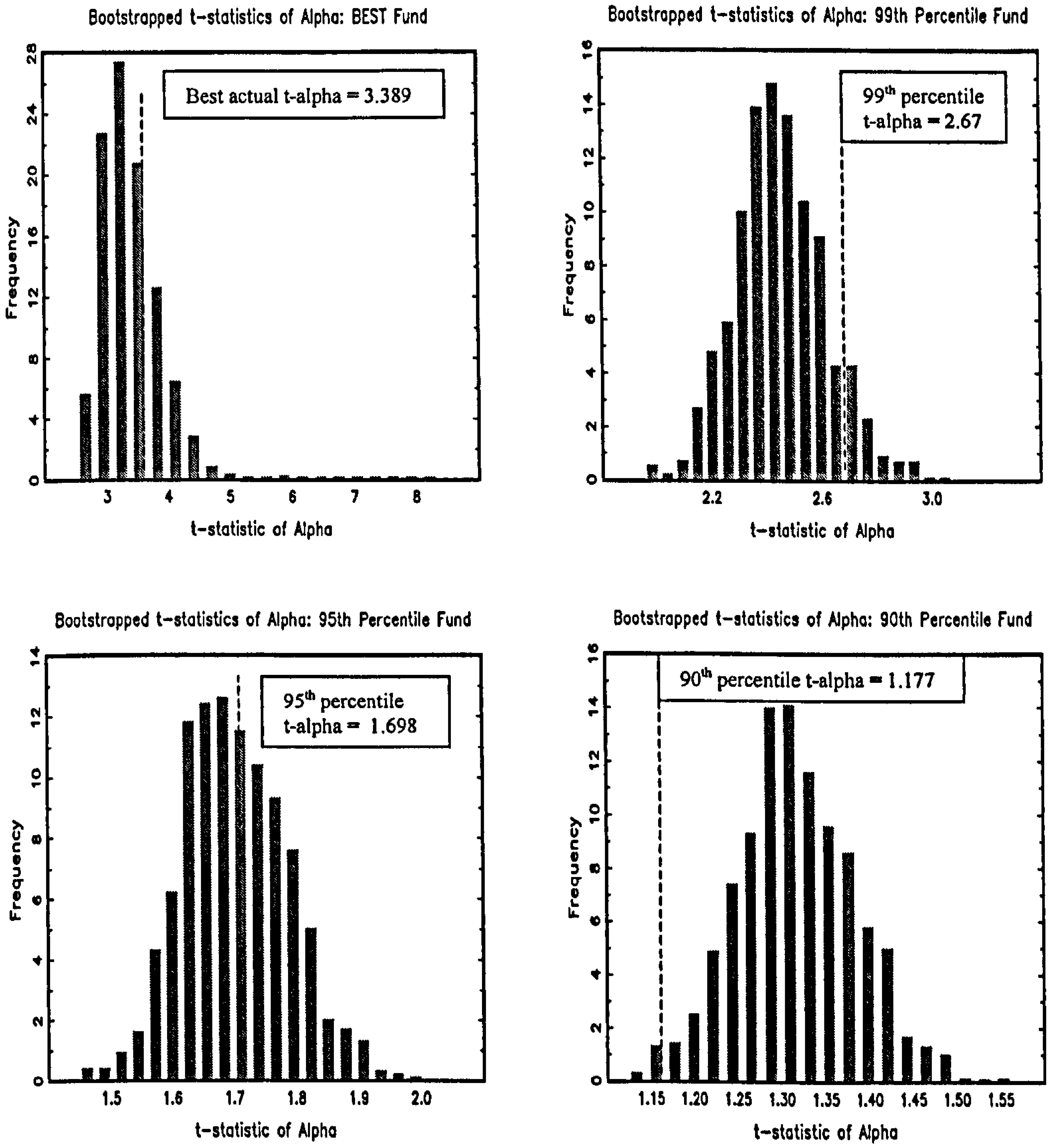


Figure 6.2b: Histograms of Residuals (Upper End of the Distribution)

Figure 6.2b shows histograms of the residuals of funds ranked at various points in the upper end of the cross-sectional performance distribution.

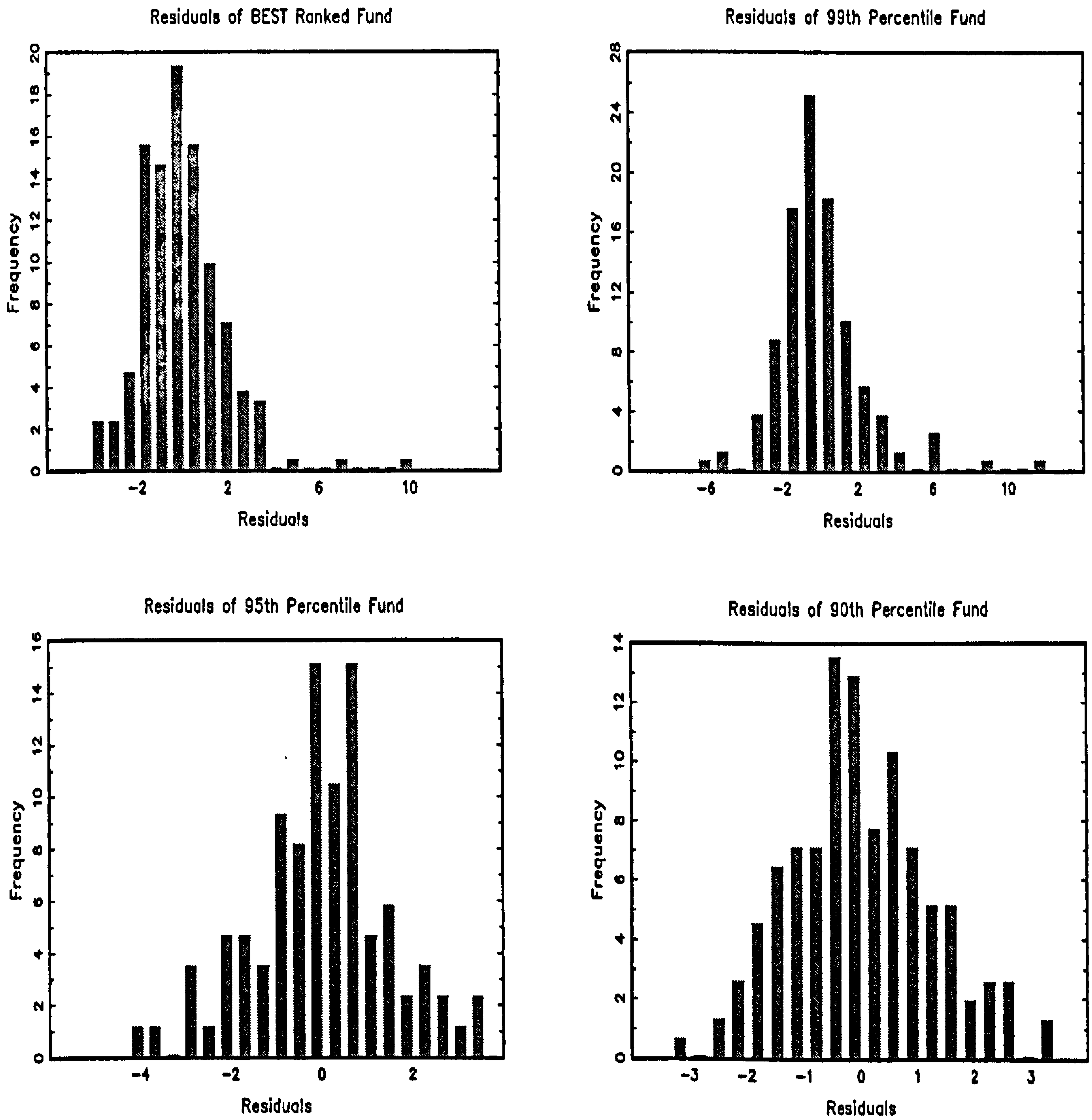


Figure 6.3a: Histograms of Bootstrap t-statistics (Lower End of the Distribution)

Figure 6.3a shows histograms of the bootstrap t-statistics of alpha from the unconditional model, at various points in the lower end of the performance distribution. The actual t-statistics at each ranking is indicated by the vertical dashed line.

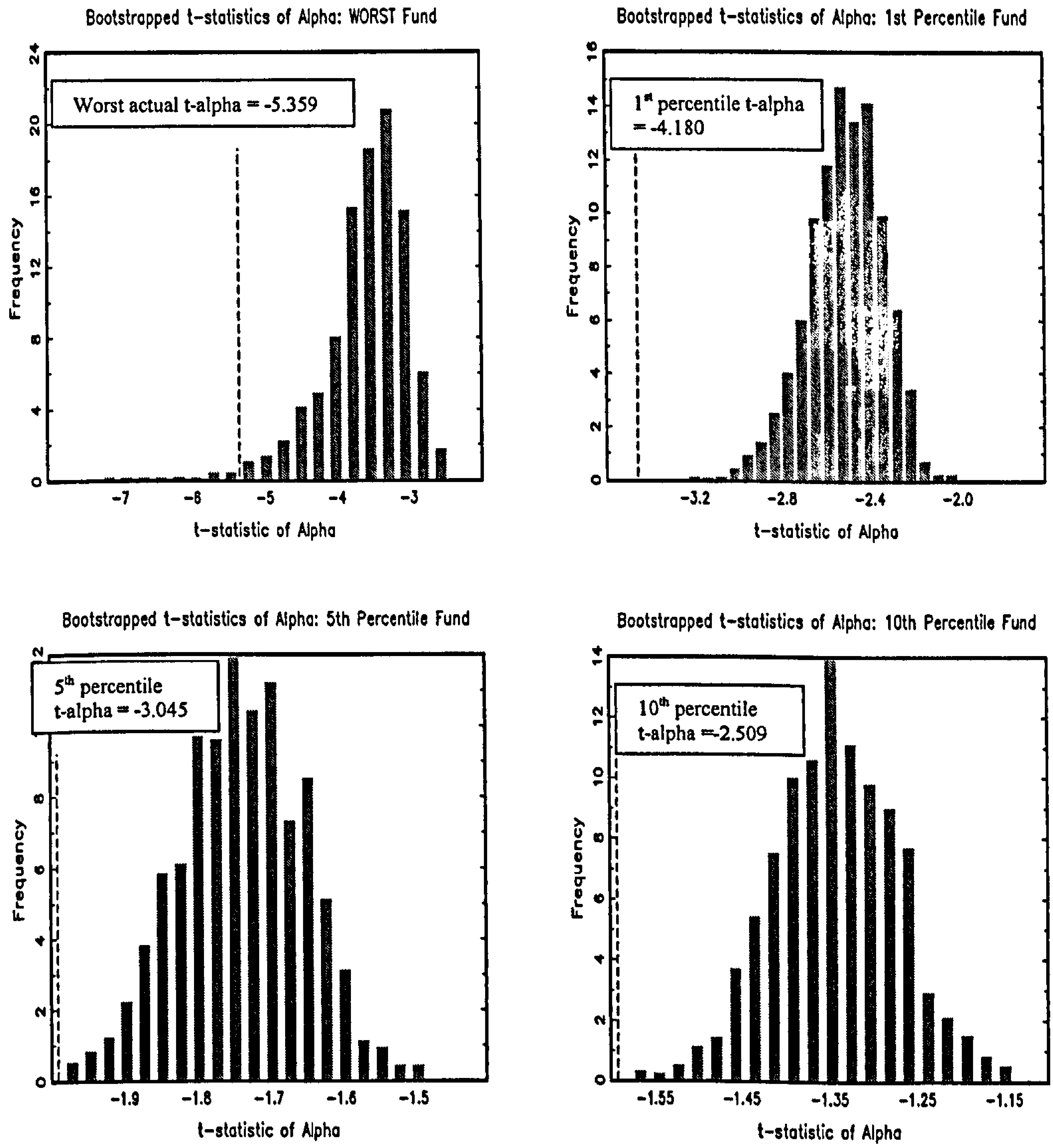


Figure 6.3b: Histograms of Residuals (Lower End of the Distribution)

Figure 6.3b shows histograms of the residuals of funds ranked at various points in the lower end of the cross-sectional performance distribution.

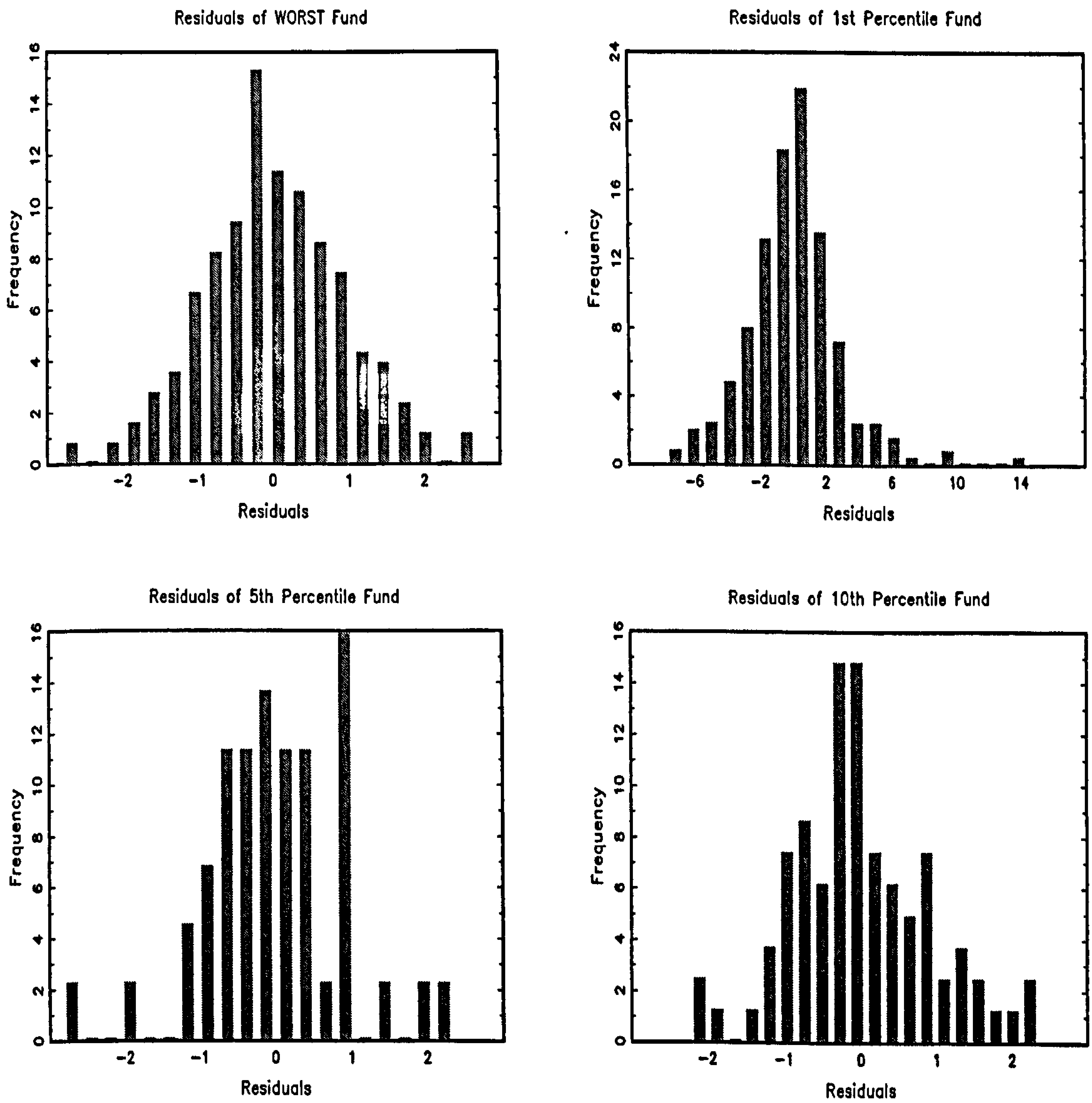
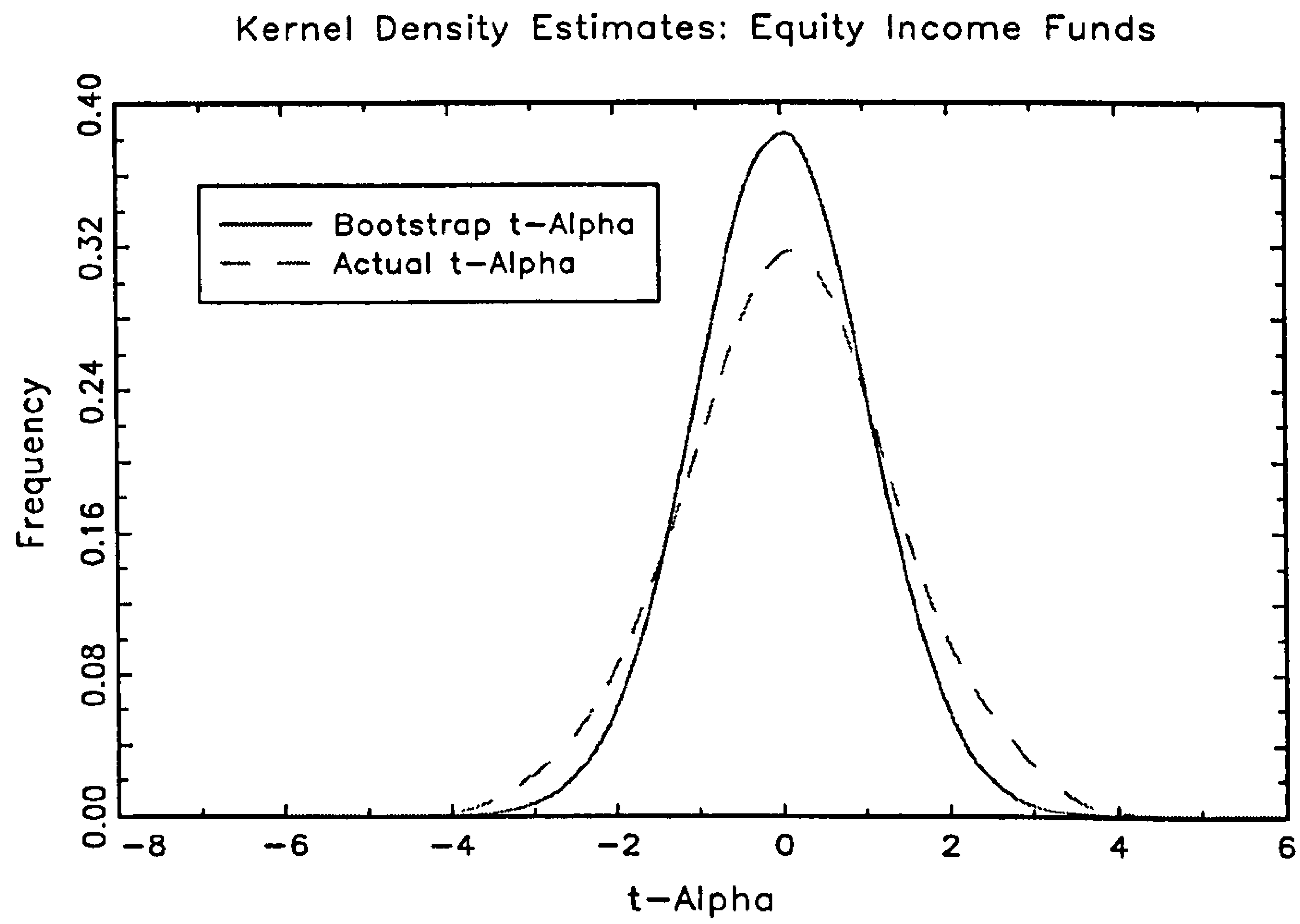


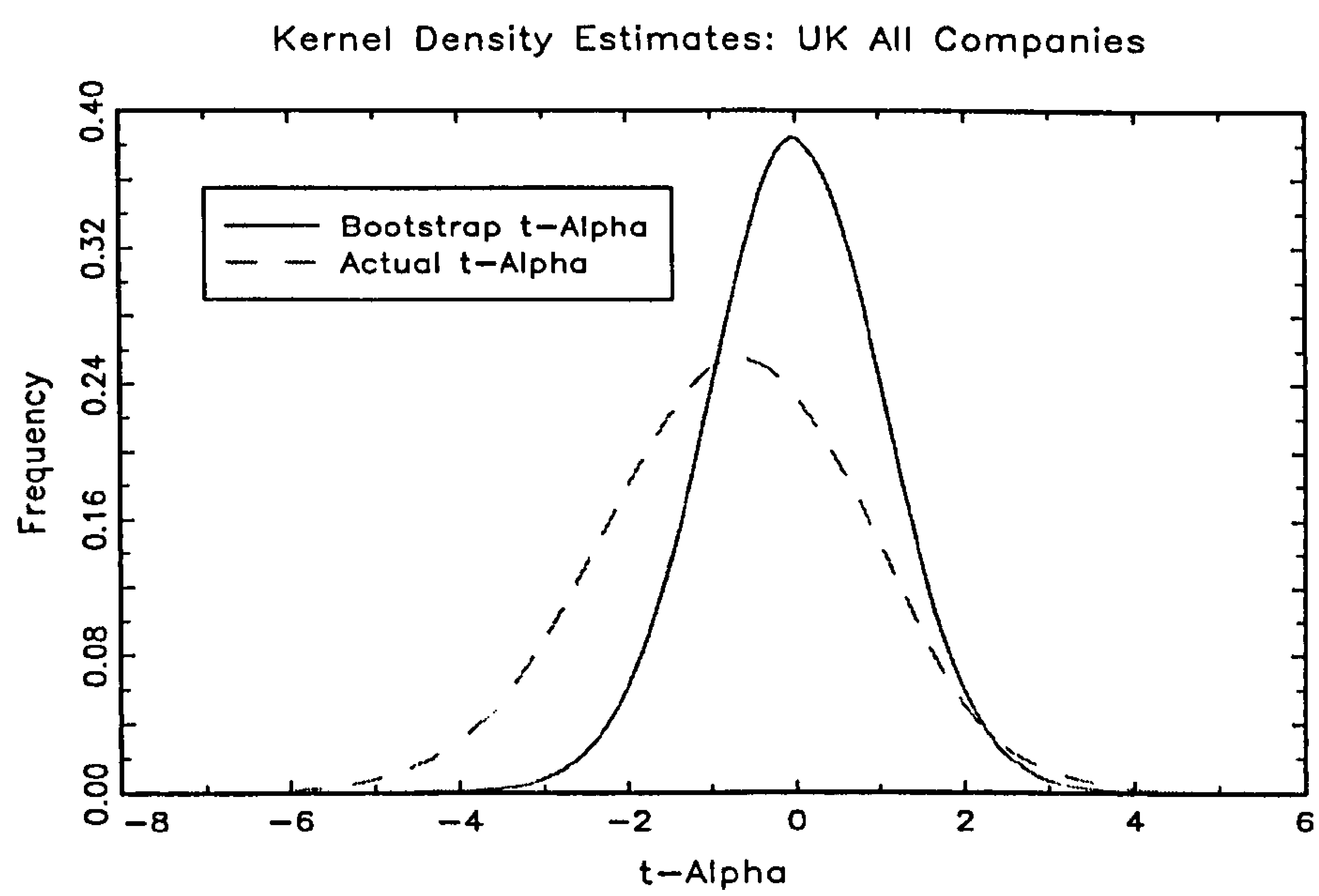
Figure 6.4: Kernel Density Estimates of the Actual and Bootstrap distribution of t-statistics - by Investment Style

Figure 6.4 shows Kernel densities of the actual and bootstrap distributions of the t-statistics of alpha applying the boorstrap procedure separately to funds of different investment styles. Results relate to the unconditional Fama-French model. t-statistics are Newey-West adjusted. A minimum of 36 observations are used. The plots use a Gaussian Kernel.

Equity Income



General Equity



Smaller Companies

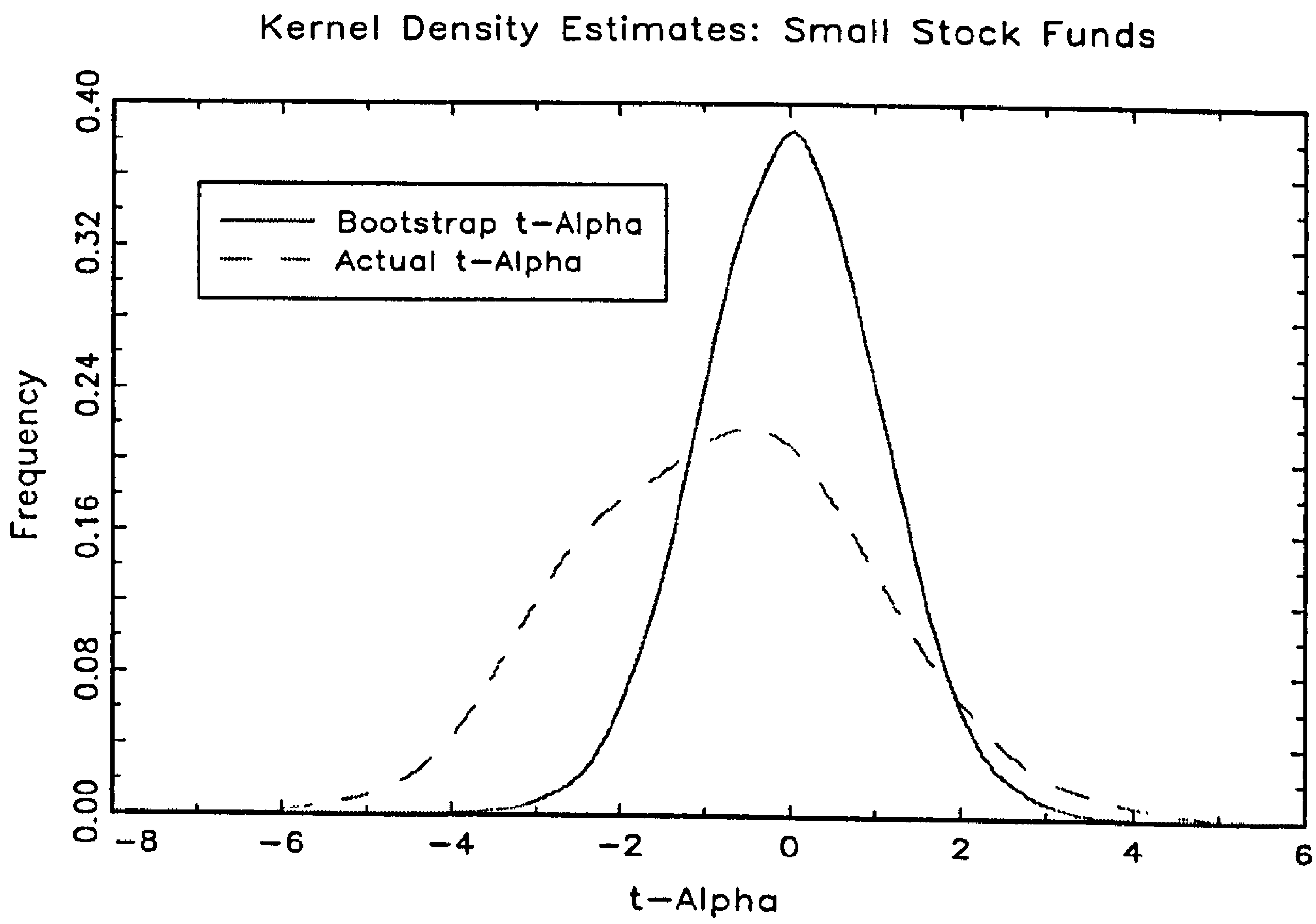
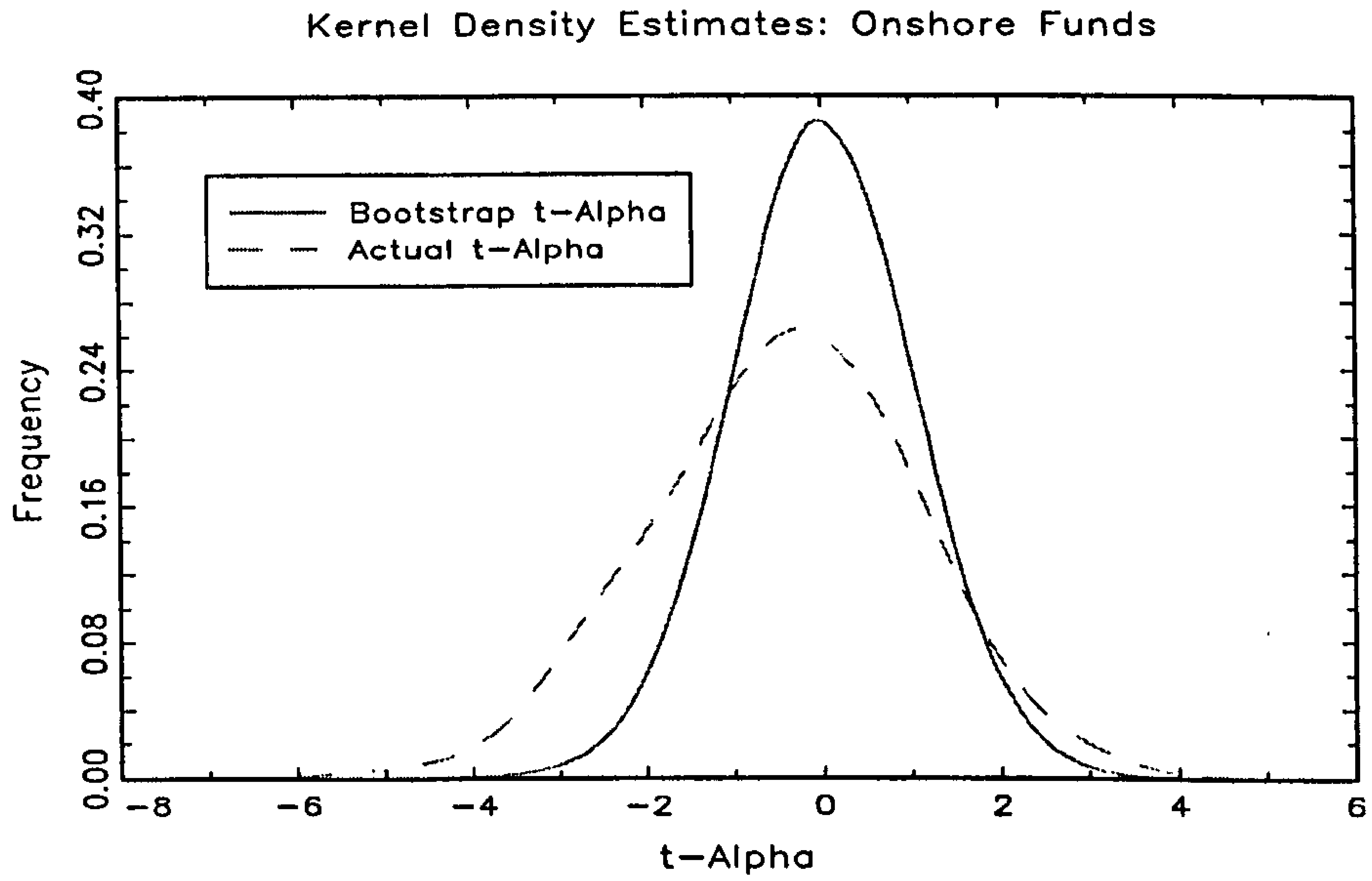


Figure 6.5: Kernel Density Estimates of the Actual and Bootstrap Distribution of t-statistics - by Location

Figure 6.5 shows Kernel densities of the actual and bootstrap distributions of the t-statistics of alpha applying the bootstrap procedure separately to the onshore and offshore funds. Estimates are from the unconditional Fama-French model, t-statistics are Newey-West adjusted and funds with a minimum of 36 observations are used. The plots are generated using a Gaussian Kernel.

Onshore Funds



Offshore Funds

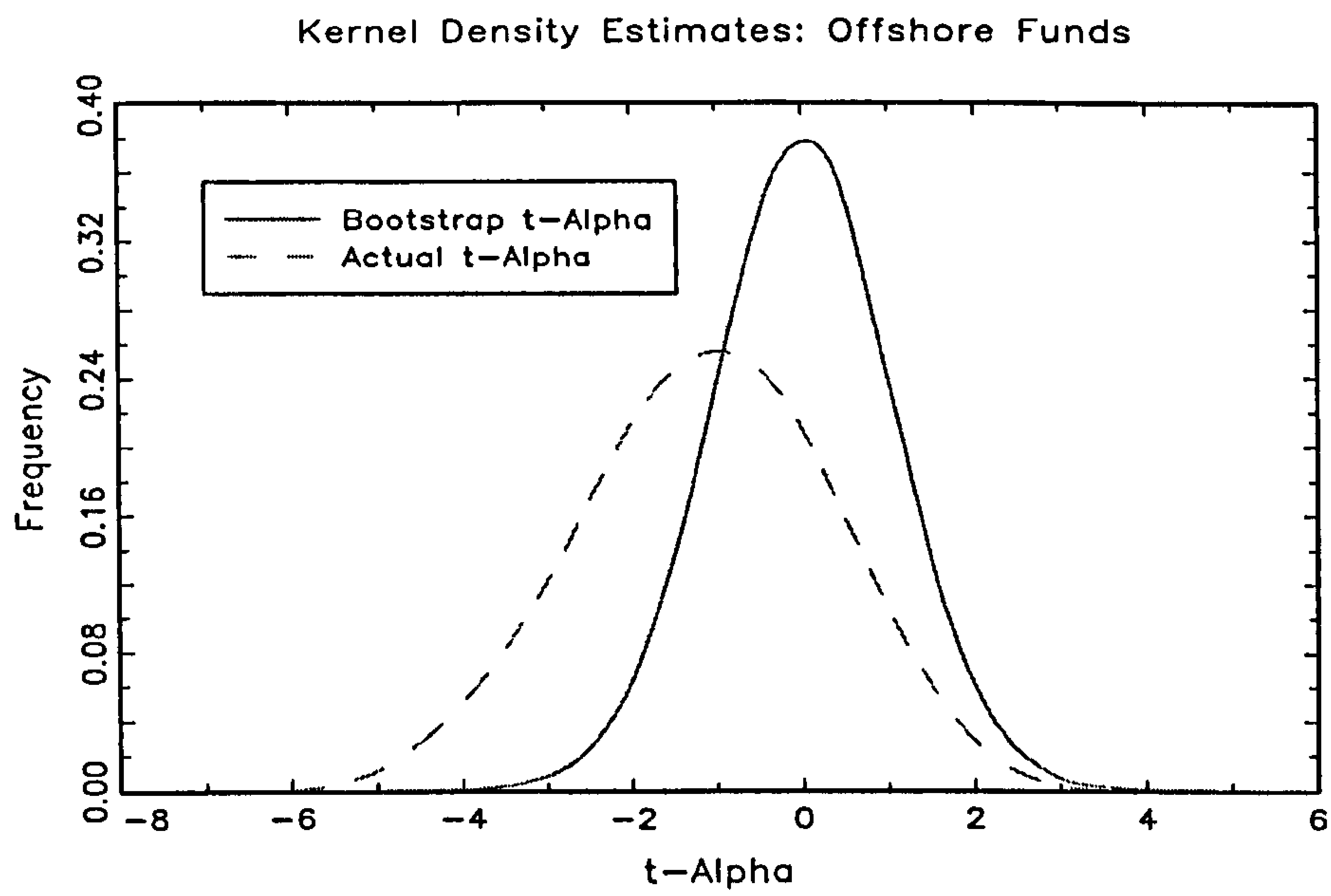


Figure 6.6: Relative Performance of Nonsurvivor Funds Prior to Closure

For each of the last 24 months of each nonsurviving fund's history the difference between the nonsurviving fund's return and the cross-sectional average of the surviving funds' returns that month is calculated. For each of the last 24 months this is then averaged across nonsurviving funds in event time.

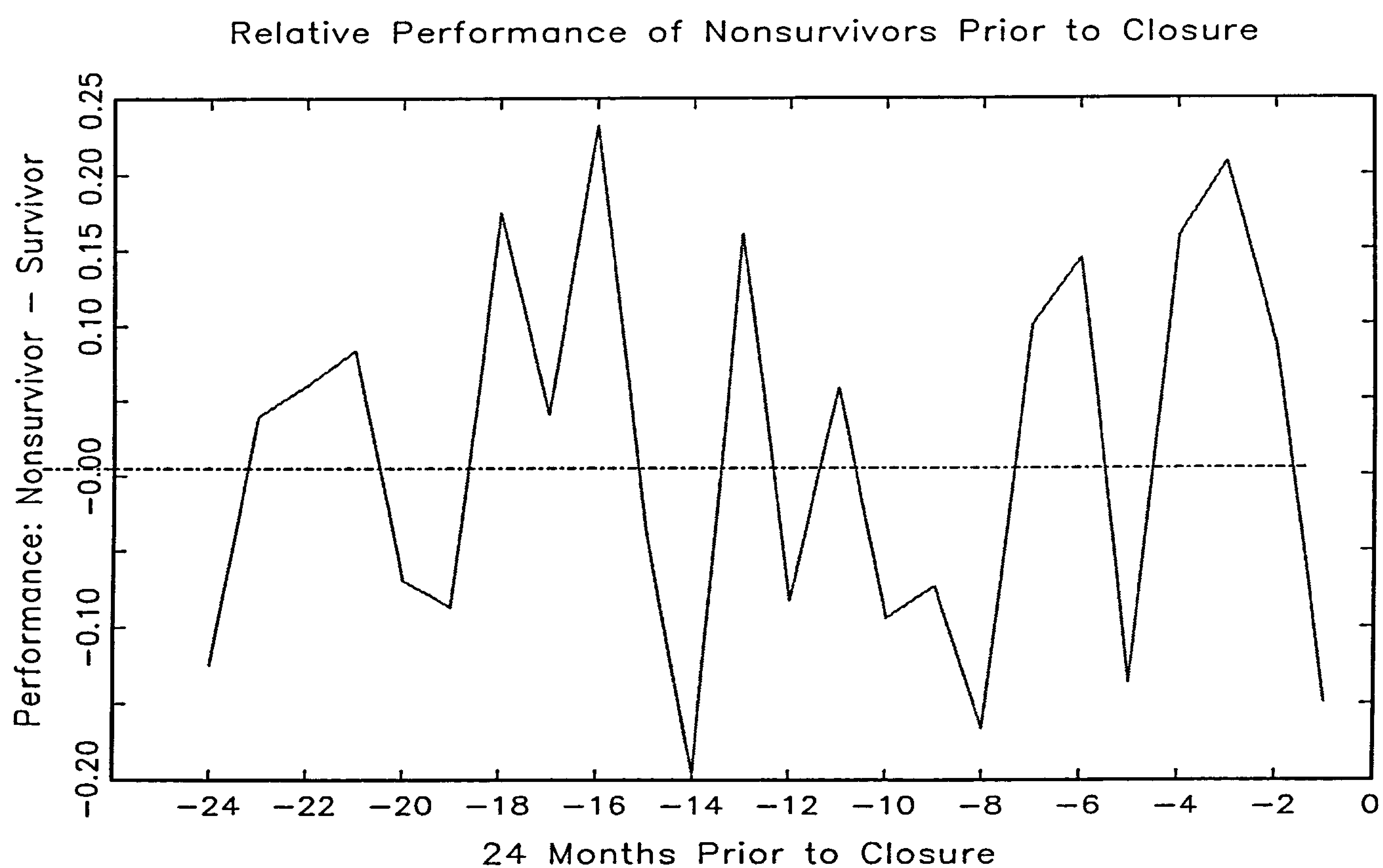
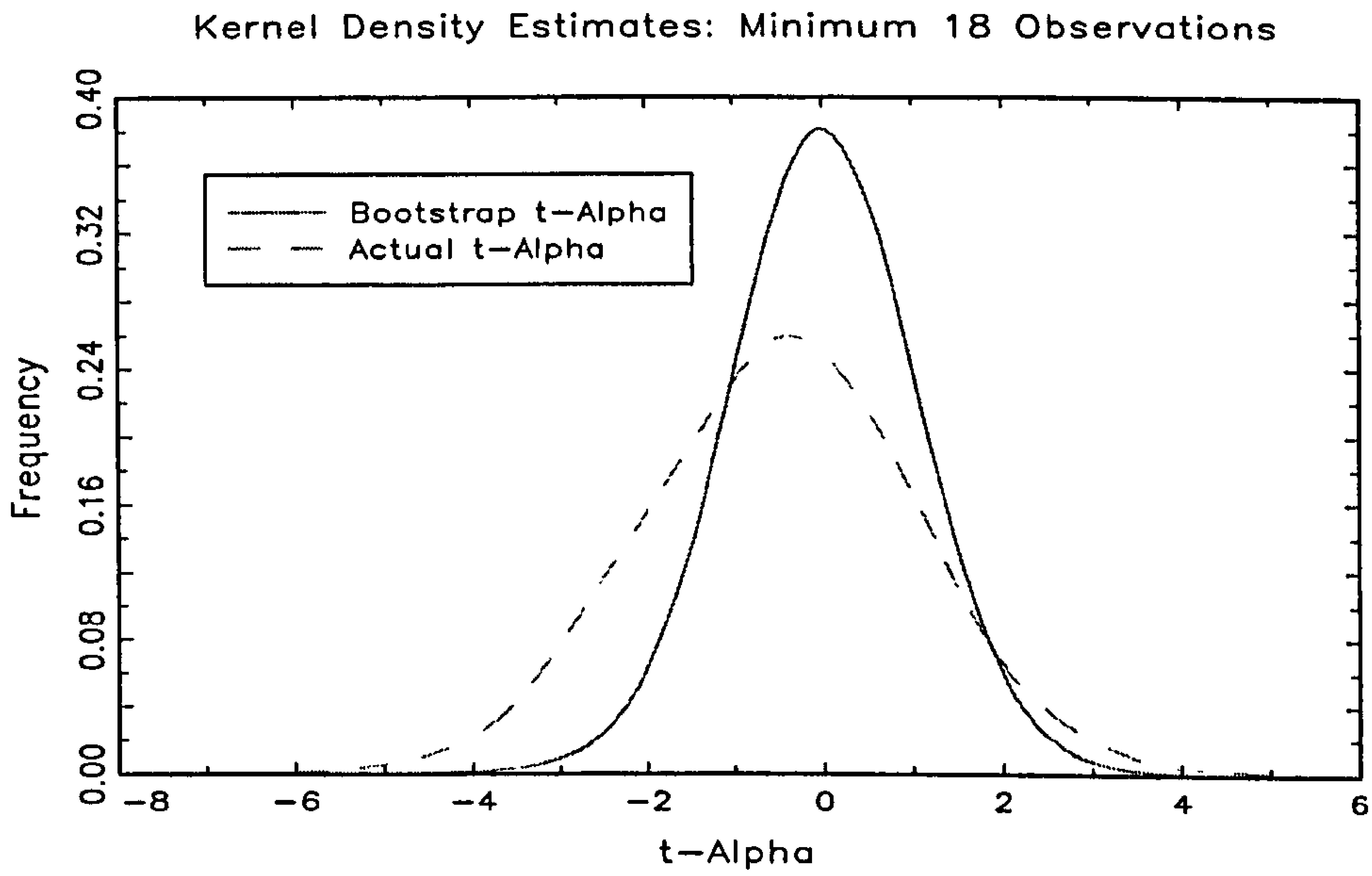


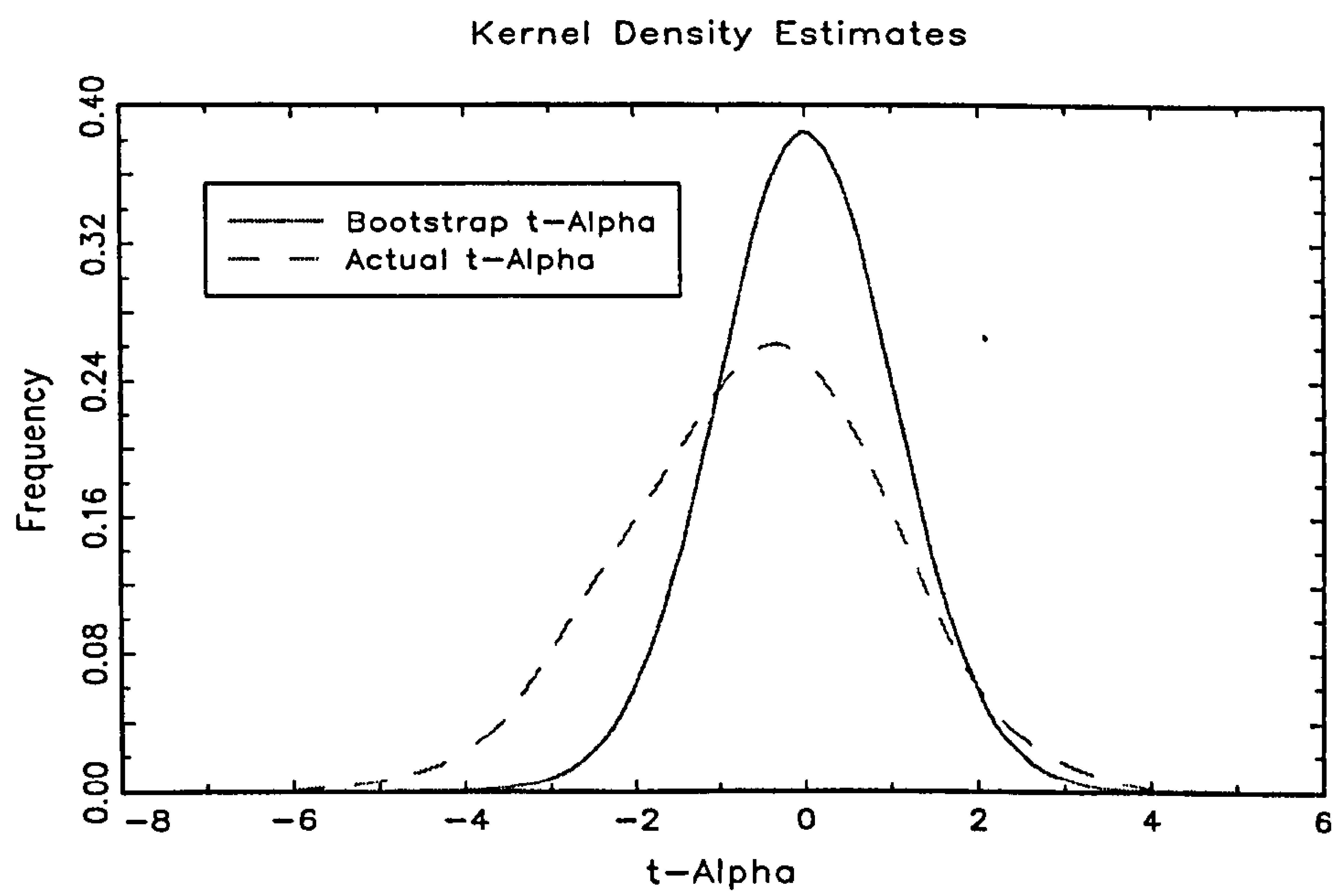
Figure 6.7 Kernel Density Estimates of the Actual and Bootstrap Distribution of t-statistics – Alternative Minimum Fund Histories

Figure 6.7 presents Kernel densities of the actual and bootstrap distributions of the t-statistics of alpha for funds with a minimum of 18, 36, 60 and 120 observations using the unconditional Fama-French 3 factor model. t-statistics are Newey-West adjusted. The plots are generated using a Gaussian Kernel function.

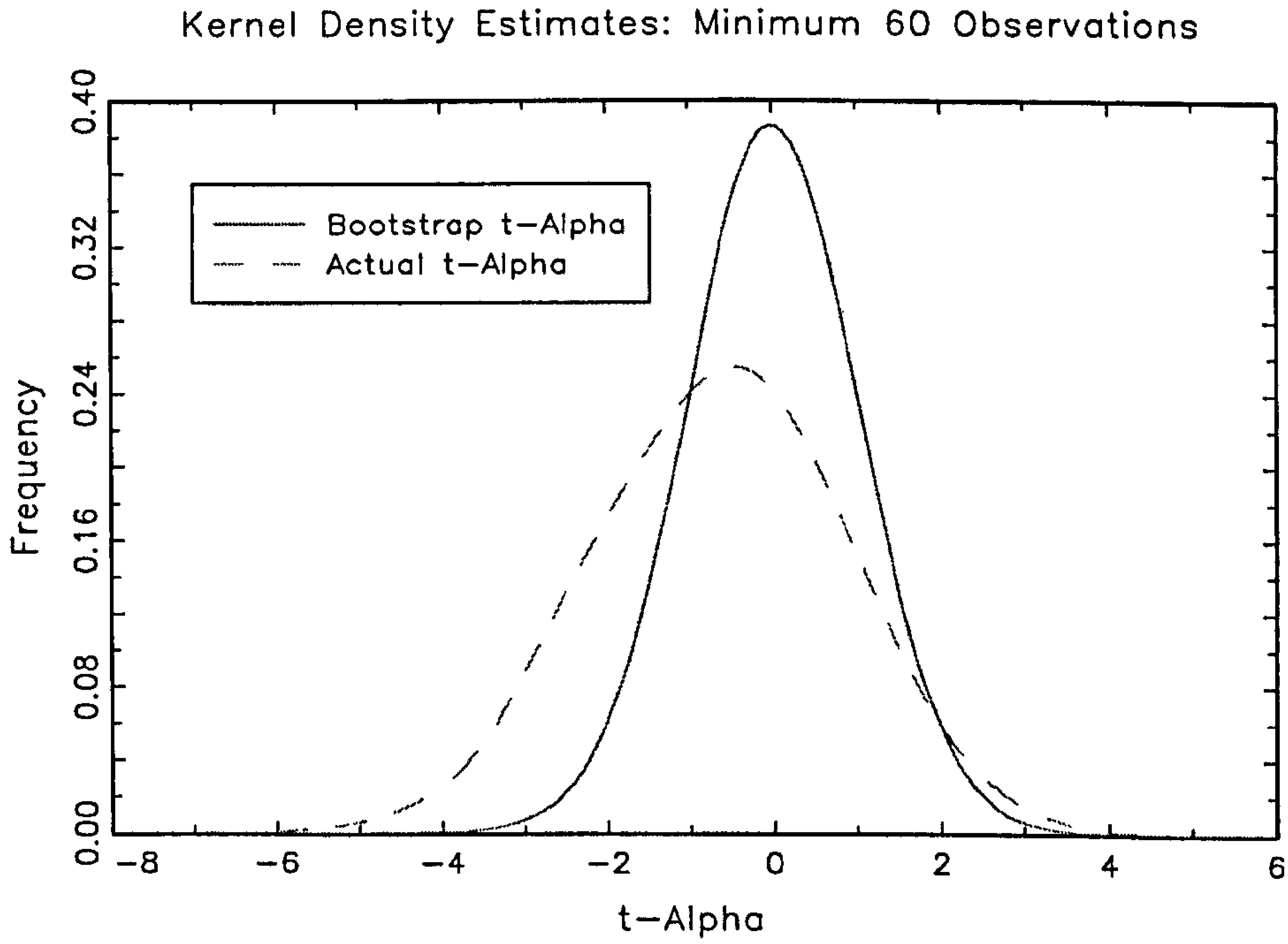
Minimum 18 Observations



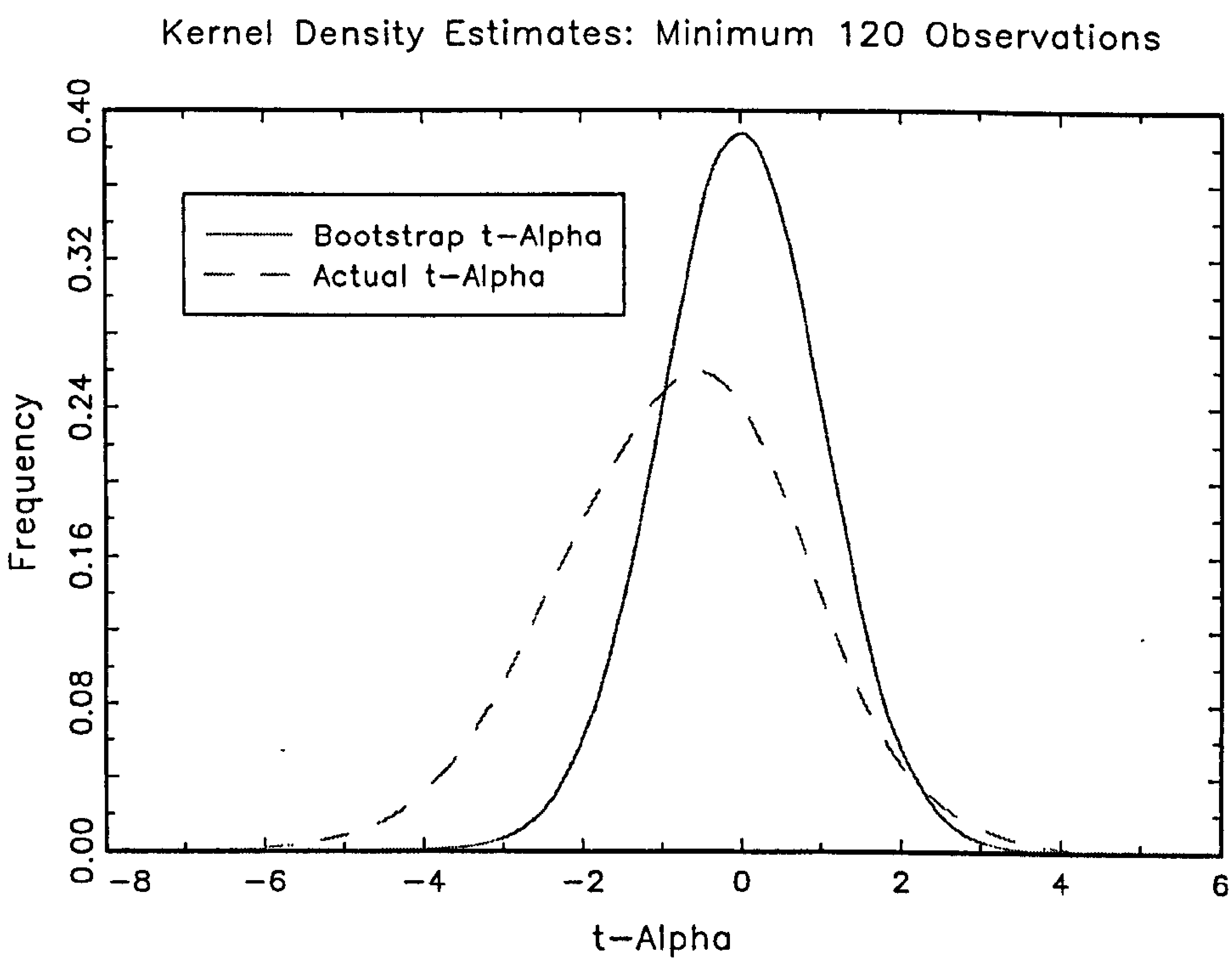
Minimum 36 Observations



Minimum 60 Observations



Minimum 120 Observations



APPENDIX of TABLES

Table A6.1
Unconditional Factor Model

Rank	Rank Alpha	t-Alpha	Rank t-stat	p value
1	2.2348	2.2259	3.3891	0.4370
2	1.4474	3.3650	3.3650	0.1280
3	0.7569	1.9019	2.9910	0.2320
4	0.7239	1.3227	2.9646	0.1090
5	0.7159	1.7423	2.7771	0.1570
6	0.7061	1.0093	2.7761	0.0770
7	0.6861	2.9910	2.6775	0.0940
8	0.6317	1.0404	2.6719	0.0460
9	0.6237	1.3975	2.6341	0.0320
10	0.6169	1.2124	2.5448	0.0380
11	0.6065	2.2920	2.5223	0.0280
12	0.5930	1.4504	2.5009	0.0200
13	0.5879	1.1009	2.4035	0.0360
14	0.5613	1.0131	2.2920	0.1000
15	0.5434	2.6719	2.2822	0.0700
16	0.5304	1.6986	2.2451	0.0710
17	0.5142	2.1271	2.2259	0.0480
18	0.5071	2.7761	2.1271	0.1450
19	0.5037	1.2840	2.0407	0.2990
20	0.4908	2.6775	2.0235	0.2840
21	0.4874	1.9482	1.9554	0.4470
22	0.4786	2.9646	1.9482	0.3830
23	0.4782	2.7771	1.9205	0.3920
24	0.4724	1.3563	1.9087	0.3670
25	0.4708	1.0779	1.9019	0.3200
26	0.4699	0.9771	1.8599	0.3900
27	0.4664	1.7886	1.7886	0.6150
28	0.4598	1.7264	1.7578	0.6760
29	0.4474	2.0235	1.7423	0.6590
30	0.4398	0.8060	1.7399	0.5880
31	0.4312	2.5448	1.7264	0.5810
32	0.4140	1.0455	1.7199	0.5420
33	0.4117	3.3891	1.7041	0.5350
34	0.3957	1.8599	1.6986	0.4910
35	0.3864	2.2822	1.6826	0.5000
36	0.3822	2.4035	1.6707	0.4830
37	0.3741	1.0455	1.6672	0.4430
38	0.3576	0.6150	1.6276	0.5440
39	0.3553	0.8807	1.6236	0.5040
40	0.3428	1.7199	1.6177	0.4770
41	0.3406	1.5707	1.6105	0.4480
42	0.3176	1.6105	1.5707	0.5720
43	0.3172	0.6415	1.5296	0.7120
44	0.3150	0.4067	1.4988	0.8040
45	0.3150	1.9087	1.4721	0.8680
46	0.3145	0.4736	1.4504	0.9010
47	0.3103	1.0508	1.4290	0.9140
48	0.3095	1.0314	1.4271	0.8860
49	0.3024	0.7221	1.4230	0.8700
50	0.3016	2.5009	1.4137	0.8600

Rank	Alpha	t-Alpha	t-stat	p value
625	-0.3915	-2.0702	-2.7069	1.0000
626	-0.3936	-3.5620	-2.7088	1.0000
627	-0.3985	-1.2944	-2.7133	1.0000
628	-0.3988	-2.0563	-2.7397	1.0000
629	-0.3998	-4.2783	-2.7423	1.0000
630	-0.4038	-2.7915	-2.7718	1.0000
631	-0.4049	-2.7718	-2.7735	1.0000
632	-0.4069	-1.6944	-2.7915	1.0000
633	-0.4130	-1.8248	-2.8812	1.0000
634	-0.4132	-1.3225	-2.9150	1.0000
635	-0.4133	-2.9716	-2.9256	1.0000
636	-0.4136	-3.7514	-2.9533	1.0000
637	-0.4192	-2.5966	-2.9543	1.0000
638	-0.4200	-2.7735	-2.9707	1.0000
639	-0.4216	-2.3563	-2.9716	1.0000
640	-0.4240	-2.4753	-3.0333	1.0000
641	-0.4266	-2.7088	-3.0432	1.0000
642	-0.4272	-1.5634	-3.0448	1.0000
643	-0.4295	-3.0768	-3.0732	1.0000
644	-0.4353	-2.5088	-3.0761	1.0000
645	-0.4403	-2.9533	-3.0768	1.0000
646	-0.4414	-2.6088	-3.0776	1.0000
647	-0.4418	-1.1358	-3.0951	1.0000
648	-0.4428	-2.7023	-3.1032	1.0000
649	-0.4514	-1.8264	-3.1169	1.0000
650	-0.4532	-2.1445	-3.1183	1.0000
651	-0.4663	-3.5220	-3.1244	1.0000
652	-0.4691	-2.6807	-3.1469	1.0000
653	-0.4791	-3.0951	-3.1609	1.0000
654	-0.4828	-2.0230	-3.1748	1.0000
655	-0.4907	-2.5747	-3.1913	1.0000
656	-0.5091	-1.7442	-3.1994	1.0000
657	-0.5104	-2.9150	-3.2499	1.0000
658	-0.5182	-2.1812	-3.5212	1.0000
659	-0.5204	-3.0776	-3.5220	1.0000
660	-0.5220	-2.4643	-3.5620	1.0000
661	-0.5231	-3.7656	-3.6163	1.0000
662	-0.5259	-3.0732	-3.6251	1.0000
663	-0.5523	-3.0333	-3.6973	1.0000
664	-0.5531	-0.7559	-3.7514	1.0000
665	-0.5586	-2.8812	-3.7656	1.0000
666	-0.5780	-1.6406	-3.8621	1.0000
667	-0.5947	-1.6795	-3.9414	1.0000
668	-0.5965	-4.4855	-4.1187	1.0000
669	-0.6092	-2.4154	-4.1804	1.0000
670	-0.6503	-2.9707	-4.1900	1.0000
671	-0.6614	-3.1994	-4.2783	1.0000
672	-0.7111	-2.3310	-4.4855	1.0000
673	-0.7269	-2.3871	-4.4985	1.0000
674	-0.7407	-3.0432	-4.9526	1.0000
675	-0.9015	-2.5321	-5.3589	0.9910

CHAPTER 7

Persistence Testing Methodology

A major issue in mutual fund performance is whether abnormal performance can be identified *ex-ante* and for how long it persists. Persistence in the sense of 'statistical predictability' is usually examined using (rank) correlations or regressions of future on past performance or using a contingency table approach. 'Economic predictability' is usually based on post-sort performance alphas or by observing actual trades of mutual funds (i.e. holdings and buy/sell data) and using a characteristic selectivity measure in an event study framework.

This study contributes to the debate on persistence in several ways. First a comprehensive survivorship bias free UK data base of over 900 equity mutual funds (which includes tracker funds in some of the analysis) over a long data period is used. This has not always been the case in earlier persistence studies. Second, persistence is examined using a wide variety of alternative sorting rules in evaluating repeat 'winner' (and 'loser') funds. Third, the study does not focus on statistical predictability as this does not necessarily result in explicit investment rules that produce future positive risk adjusted returns. Instead it concentrates on the economic value of persistence by focusing on the recursive portfolio approach. Here, if persistence is found it may represent an exploitable strategy for investors. Fourth, as well as analyzing the post-sort performance of quite large and possibly heterogeneous portfolios of funds (e.g. deciles), as done in earlier studies, this study also examines alternative smaller fund-of-funds portfolios which is probably of more practical interest to both professional and retail investors. Finally, the study examines not only the risk adjusted average performance of past winner/loser portfolios of funds but also the *distribution* of final wealth from this *ex-ante* strategy, taking account of 'luck' across *all* funds and transactions costs of rebalancing. In this study 'luck' is represented by the empirical distribution of all funds' idiosyncratic risks and hence picks up any non-normality and contemporaneous cross-correlations in idiosyncratic risks. As far as can be ascertained, this approach to an analysis of final wealth has not been developed in previous studies. Since most saving in mutual funds (as a whole) is long-term, investors are interested in the distribution of final wealth (e.g. mean, skewness, kurtosis) from an active strategy, relative to that from alternative strategies such as holding index trackers. These alternatives are examined with respect to the distribution of final wealth, taking account of 'luck' and transactions costs for both strategies. The methodology adopted in this study moves the debate on persistence closer to the practical issues surrounding the implementation of *ex-ante* investment strategies by fund investors.

7.1 Persistence Methodology

Tests for persistence/predictability fall into two broad categories. One can test for 'statistical' predictability or 'economically significant' predictability or both. Statistical measures of persistence rank funds over some past horizon and measure the association between past performance and future performance, where the performance metrics may be different in the ranking and post-ranking periods (e.g. ranking into decile portfolios based on past raw returns but the post-rank metric being future alpha performance). The statistical approach measures the average association between the *relative orderings* of funds in the pre-sort and post-sort periods using correlation, regression or contingency tables. However, although such tests may provide evidence of persistence it is often not clear how this may be exploited by investors. For example, rank correlations or a regression of pre-sort and post-sort alphas can be used to establish predictability. However, even though there may be a high correlation between the alphas of past decile ranked funds and their subsequent alphas, all of the post-sort decile alphas may be negative, indicating (relative) predictability but poor future abnormal performance for all decile portfolios. Therefore measured persistence could be due mainly to repeat losers rather than repeat winners. In addition (Spearman) rank correlations treat each point in the ranking equally and lack power against the hypothesis that predictability in performance is concentrated in the tails of fund performance. In this study this is taken up by examining the *ex-ante* performance of small portfolios of funds.

While the above approaches can be used to establish statistical predictability, investors are presumably more interested in the future absolute risk adjusted performance of both winners and losers *taken separately* so this study uses the 'recursive portfolio' approach which allows a direct assessment of the economic as well as statistical significance of persistence. The methodology is as follows. At any point in time, *t*, funds are ranked according to some measure of past performance over a particular horizon (e.g. raw returns over the past year). Fractile portfolios are then formed (e.g. deciles) and held over the next h months, which gives rise to a sequence of h monthly 'forward looking' returns. Rebalancing takes place every h months which results in a sequence of forward looking concatenated monthly returns for *each* fractile portfolio over the whole test period, $R_i^f(t,T)$ where $(t = t+1, t+2, ...T)$. $R_i^f(t,T)$ are then used to estimate alternative factor models (or other performance statistics) which give an economic measure of persistence such as post-sort alphas, α_i^f (which are referred to here as 'forward looking' alphas). Here, results are reported for the Carhart (1997) 4-factor model:

$$(7.1) \qquad r_{i,t}^f = \alpha_i^f + \beta_{1i} r_{m,t} + \beta_{2i} SMB_t + \beta_{3i} HML_t + \beta_{4i} MOM_t + \varepsilon_{i,t}$$

where $r_{i,t}^f = R_{i,t}^f - r$ is the excess return on the forward looking portfolio, $r_{m,t}$ is the excess return on the market portfolio, SMB_t, HML_t and MOM_t are the zero-investment factor mimicking portfolios for size, book-to-market value and momentum effects respectively, (see chapter 4). For example, a statistically significant positive (negative) 'forward looking' alpha of past winners (losers) indicates that a successful *ex-ante* strategy would involve 'smart investors' redirecting cash flows into past winner funds and away from past loser funds. In this study the forward looking returns $R_i^f(t,T)$ are based on either sorting on the fund's past 12 months (raw) return or on the t-alpha from the 4-factor model (using a 60 month estimation window)[1]. The portfolios are equally weighted and alternative rebalancing periods of 1, 3, 6 and 12 months are examined. Because of possible non-normality in a portfolio's idiosyncratic risk (particularly for the extreme tails of the performance distribution), this study reports bootstrap p-values of alpha for the different size forward looking portfolios. Size here refers to the number of funds held in this portfolio of funds strategy.

The forward looking returns $R_i^f(t,T)$ consist of a changing portfolio of different mutual funds, although not all funds will necessarily be 'switched' at each rebalancing date. The implications for transactions costs of this 'fund-of-funds' strategy is discussed below. Note, in this study when a fund closes sometime over the holding period it is included in the forward looking portfolio until it closes and the portfolio is then rebalanced among the remaining 'live' funds until the next rebalancing period. This removes any 'look-ahead bias' (Carhart 1997).

[1] Using t-alpha (rather than alpha) helps reduce sorting errors due to outlier residuals.

7.2 Calculating Final Wealth – Rebalancing with Transactions Costs .

Use of forward looking alphas provides one metric for judging the economic significance of persistence but of equal importance to investors is the level of terminal wealth from alternative investment strategies. Although figures for a fund's terminal wealth are regularly reported in the financial press, actual terminal wealth from a specific active strategy (over a specific investment horizon) may be a misleading measure of performance since it may be largely due to luck (from chapter 6). It would be useful therefore if we could measure the riskiness in terminal wealth from an active strategy. After all, investors are not simply interested in the risk adjusted return of their *ex-ante* investment strategy but also in the distribution of their final wealth, taking account of the cross-section of risk across all funds and any rebalancing costs of an active strategy.

In an active strategy, if rebalancing takes place m times per year (y) then the total number of rebalancing periods is $T = m.y$. If p_t = proportion of funds switched at each rebalancing period, c = percentage cost of rebalancing (assumed fixed) and $R^f_{p,t}$ = *annual* forward looking portfolio return, then (the log of) final wealth, $\ln W_T$, has expected value and variance:

(7.2)
$$\ln W_T = \ln W_0 + \sum_{t=1}^{T} \ln(1 + R^f_{p,t} / m) + \ln(1 - c.p_t)$$
$$\approx \ln W_0 + \sum_{t=1}^{T} (R^f_{p,t} / m) - c.p_t = \ln W_0 + \sum_{t=1}^{T} R^{f,net}_{p,t}$$

$$E(\ln W_T) \approx \ln W_0 + \sum_{t=1}^{T} ER^{f,net}_{p,t}$$
$$\sigma(\ln W_T) \approx e'\Omega_p\, e$$

where $R^{f,net}_{p,t} = (R^f_{p,t} / m) - c.p_t$ is the per period return net of transactions costs, e is the (Tx1) unit vector and Ω_p is the (TxT) covariance matrix of net returns for the forward looking portfolio. If expected returns and the expected proportion of funds switched are constant each rebalancing period then the active strategy gives:

(7.3)
$$E(\ln W_T) \approx \ln W_0 + y.ER^f_p - y.(m.Ep).c$$

A passive (index tracker) strategy with constant expected returns each period and zero transactions costs has $E(\ln W_{tr,T}) \approx \ln W_0 + y.ER^f_{tr}$ and the variance of the passive strategy is $\sigma(\ln W_{tr,T}) \approx e'\Omega_{tr} e$. The breakeven condition for the active strategy to give the same level of expected wealth as the passive strategy is:

$$ER^f_p - ER^f_{tr} = (m.Ep).c \tag{7.4}$$

Therefore the standard result is that the active strategy gives a higher expected wealth than the passive strategy when the excess (annual) return from the active strategy exceeds the annual cost of switching.

Consider what determines expected final wealth from the active and passive strategies. For the active strategy, both $R^f_{p,t}$ and p_t are likely to depend on the number of rebalancing periods per year, m, and the size of the portfolio, z. If the persistence of winners is short-lived then more frequent rebalancing (higher m) will tend to give a higher value for the expected annual return, ER^f_p, and with zero transactions costs, higher expected terminal wealth. However, the proportion of funds rebalanced *each year*, $(m.Ep)$, will probably rise with increased frequency of rebalancing which, with transaction costs, will detract from returns as transactions cost *per annum* rise proportionately with m. Hence with transactions costs, the resulting level of expected final wealth as rebalancing frequency increases depends on the interplay of a higher expected return and higher annual rebalancing costs. This cannot be determined *a priori* and forms a key element of the empirical analysis.

If persistence is concentrated in the extreme tails of the performance distribution then the expected annual return on the 'winner' portfolio, ER^f_p, is likely to be negatively related the size of the portfolio, z, for any given rebalancing frequency. Therefore portfolios containing a small number of winner funds are likely to have higher gross returns (i.e. ignoring transactions costs) than larger portfolios. However, it is difficult to say *a priori* how the proportion of funds rebalanced each period (Ep) will vary with the size of the portfolio and hence whether the active strategy will 'beat' the passive strategy (net of transaction costs), see (7.4) where m is fixed. For example, for a winner portfolio, how Ep changes with z, depends on the distribution of mean returns and standard deviation for those funds that are close to the bottom of the z-cohort. If there is no persistence then p will rise as the size of the fund gets smaller. In contrast, if there is persistence only in very small portfolios of winner funds, then we expect p to rise as the size of the fund increases (at least initially), since as the fund gets larger we introduce extra funds into the winner portfolio whose returns differ only randomly, resulting in more turnover at each rebalancing date. The above analysis suggests that if persistence of winners is concentrated in relatively small portfolios, this results in higher average gross returns and lower transactions costs due to less frequent rebalancing, giving the active strategy the best chance of beating the passive strategy. The empirical analysis provides evidence on the sensitivity of expected final wealth to different size portfolios.

Turning now to the distribution of final wealth for the active and passive strategies. For the active strategy $\sigma(\ln W_T)$ depends on the covariance matrix Ω_p of net returns $R_{p,t}^{f,net} = (R_{p,t}^{f} / m) - c.p_t$. If portfolio returns $R_{p,t}^{f,net}$ are $niid(\mu, \sigma^2)$ then the logarithm of final wealth is $\ln W_T \sim N(\ln W_0 + \mu T, \sigma^2 T)$ - so final wealth is lognormally distributed with known mean and standard deviation. However, when using alternative sorting rules there may be forward looking portfolio returns $R_p^f(t,T)$ which are persistent and hence one would not expect the independence assumption to hold and normality is unlikely to hold, particularly for small portfolios of funds (from chapter 6 and Kosowski et al 2004). In addition, we are interested in the distribution of final wealth taking account of the transactions costs of switching and the latter is stochastic depending in part on the serial correlation in past returns. So if there is persistence the covariances between future returns and the proportion of funds switched p_t (which depends on the autocorrelation in returns) are likely to be non-zero. Therefore, to obtain the distribution of final wealth this study uses bootstrap techniques.

To illustrate the cross-section bootstrap procedure (see Efron and Tibshirani 1993, Politis and Romano 1994) and the methodology used to assess the impact of uncertainty across all fund returns on final wealth, consider the simple case of sorting all funds on their past t-alphas and initially assume zero transactions costs. First, at each rebalancing date t, using the past 60 observations, the 4-factor model is estimated $r_{i,t} = \hat{\alpha}_i + \hat{\beta}_i ' F_t + e_{i,t}$, for i = {1, 2, ..., M) funds, where F_t = vector of risk factors and $e_{i,t}$ are the residuals of fund i. For each fund i, the procedure (contemporaneously across funds) draws a random sample (with replacement) from the residuals $e_{i,t}$ and uses these *re-sampled* bootstrap residuals $\tilde{e}_{i,t}$ to generate a simulated excess return series $\tilde{r}_{i,t} = \hat{\alpha}_i + \hat{\beta}_i{}' F_t + \tilde{e}_{i,t}$. Next the procedure re-estimates the factor model for all funds using the simulated data (and the factors) and then sorts funds by t-alpha and forms different fractile portfolios. The returns on these 'forward looking' fractile portfolios are measured over the period (t,t+h). This process is repeated for each fractile at each rebalancing period to give the simulated series $\tilde{R}_i^f(t,T)$, which are used to calculate final wealth (W_T) for each fractile portfolio. This constitutes the first run of the bootstrap. The above is then repeated B-times (B=1,000) which gives the bootstrap distribution of final wealth $f(W_T)$, for each fractile portfolio.

Given the randomness introduced by the cross-section bootstrap at each rebalancing date, one will generally obtain a different ordering of the funds, which is solely due to the distribution of idiosyncratic risk across all funds. This is what determines the distribution of final wealth for each fractile portfolio. Hence for each run of the bootstrap, different funds are 'switched' into and out of any given fractile portfolio at each rebalancing date.

Now consider transactions costs. At each rebalancing date the specific funds that are switched are noted and tracked so one can calculate portfolio turnover and hence transactions costs. This has been done here for alternative rebalancing costs (load fees and bid-ask spread) of 0%, 2.5% and 5% (of market value) for each fund that is switched and the final wealth is calculated for an investment of £1,000 over a 10 year horizon ending in December 2002.

When funds 'drop out' of the portfolio at a particular rebalancing date, they are sold and the whole of their value is then invested (in equal amounts) in the funds which replace them. When purchasing a new fund alternative transactions costs of 0%, 2.5% or 5% (of value) are assumed which are immediately deducted from the amount invested.

Also of interest is to compare the distribution of final wealth from the above active strategies to that from a passive strategy. The bootstrap distribution for the passive strategy is more simple. From among the index tracker funds in the data set, one procedure is to randomly choose (with equal probability) a single tracker fund at t = 0 and hold that fund over the investment horizon. Hence we incur no rebalancing costs and the chosen tracker's returns for the first run of the bootstrap can be used to calculate final wealth. The above is repeated B-times to give the distribution of final wealth for the passive investment strategy. Although one might expect a relatively narrower distribution of final wealth from the passive relative to an active strategy, there is still variation among tracker performance, not least because of different management fees (e.g. see Mahoney 2004 and Elton et al 2004). All mutual fund returns in the data set are net of management fees and in generating the bootstrap distribution of final wealth in this passive strategy the procedure assumes zero transactions costs for the single buy-sell transaction of the tracker investment. (For consistency transactions costs are not applied to the initial equally weighted investment in the active strategy above).

This completes the discussion of the persistence testing methodology. In chapter 8, this study implements these procedures to test for persistence among both fractiles and 'smaller' portfolios of funds, to estimate actual final wealth from persistence strategies and its (bootstrapped) distribution which embodies risk across all funds and also to compare final wealth from an active persistence strategy with that of a passive strategy.

CHAPTER 8

Empirical Results of the Persistence Tests

This chapter presents the main findings from the application of the persistence testing methodologies discussed in chapter 7 to the data set of UK mutual funds. Tables and Charts of results are presented at the end of the chapter.

8.1 Recursive Portfolio Formation

In this section, the study uses all 842 *actively* managed UK equity mutual funds which exist for some or all of the data period. When sorting on past raw returns, results for alternative 'ranking/rebalancing' horizons of 12/12, 12/6, 12/3 and 12/1 months are reported. When ranking using past 4-factor t-alphas, at each rebalancing date only funds which have at least 60 months of data are used and 'ranking/rebalancing' horizons of 60/12, 60/6, 60 /3 and 60/1 months are presented.[1,2]

Table 8.1 reports the 'forward looking' decile alphas from the 4-factor model where funds are ranked either by past raw returns (Panel A) or past t-alpha (Panel B). Results using the 4-factor model are qualitatively similar to results obtained using the Fama-French 3-factor model but as the momentum variable is statistically significant for most of the 'forward looking' portfolios, in this chapter the study reports results for the 4-factor model. In particular, the momentum variable is found to be statistically significant for the extreme winner and lose decile portfolios (but much less so for the middle ranked, 4^{th} to 6^{th} decile portfolios). The statistical significance of the momentum variable is also more evident the more frequent the rebalancing.

After ranking on past one-year returns (Panel A), most of the decile portfolios have negative forward looking alphas with those of past 'loser' funds being statistically

[1] The minimum 60 observation restriction is applied to reduce estimation error. There is a trade off between reduced estimation error and survivorship bias and using 60 months seems a reasonable compromise (see Kothari and Warner 2001 on size and power properties in estimating alpha using factor models and Mamaysky, Spiegel and Zhang 2005 on errors when sorting on alphas estimated over short horizons).

[2] Different ranking and holding periods have also been tested but this does not qualitatively affect the key results. For example, 6/12, 6/6, 6/3 and 6/1 periods for funds sorted by past returns and 36/12, 36/6, 36/3,36/1 periods for funds sorted by past t-alphas.

significant. Only the top decile portfolio, when rebalanced every 6 months (or more frequently), reveals a positive (but statistically insignificant) alpha. Broadly similar results are obtained when sorting by t-alpha (Table 8.1, Panel B). For example, generally when sorted by t-alphas, the negative forward looking alphas for the lower deciles are statistically significant and range between -0.14 (-1.68% p.a.) and -0.17 (-2.04% p.a.) and similarly when sorted by returns, range between -0.14 (-1.68% p.a.) and -0.24 (-2.88% p.a.). For the top decile when sorted by t-alpha, the forward looking alphas lie between -0.003 (-0.036% p.a.) and 0.012 (0.14% p.a.) and similarly when sorted by returns are between -0.046 (-0.55% p.a.) and 0.038 (0.45% p.a.) – but none is statistically significant. Therefore sorting on either past returns or past t-alphas gives qualitatively similar results for the forward looking alphas of the top and bottom deciles – past losers continue to perform badly, while the past winner decile shows little evidence of positive and economically significant persistence.

After sorting on past returns or past t-alphas the top decile does not show evidence of statistically significant forward-looking alphas, although some are positive. The latter suggests that if there is any positive persistence among UK mutual funds one would expect to find it *within* the top decile of past winners. Given the increase in the number of funds through the sample period, the top decile (or any fractile) portfolio tends to contain an increasing number of funds over time. The data set consists of 842 (active) funds which implies the top decile contains a maximum of 84 funds. This portfolio might be too heterogeneous to identify (a few) genuine repeat winners due to a very wide spread in past returns or t-alphas. Hence, this study repeats the above analysis for different 'smaller' fund-of-funds from among the past winner funds. This mitigates the heterogeneity problem but smaller portfolios may involve severe non-normality in idiosyncratic risks (from chapter 6). This necessitates the use of bootstrap standard errors on the forward looking alphas which are reported throughout. Smaller portfolios of a fixed size (throughout the whole investment horizon) are likely to be a more realistic *ex-ante* strategy for many investors.

Table 8.2 reports the alphas (and bootstrap p-values) for the 4-factor model when funds are sorted either by returns (Panel A) or by t-alphas (Panel B) using alternative 'smaller' portfolios of a fixed number of funds ranging from 1 to 50. Sorting by returns (Panel A) there are some statistically significant forward looking alphas for portfolios containing up to the 3 best funds when the rebalancing period is less than 12 months –

with the best results in terms of statistical significance for the 12/3, 'ranking/rebalancing' period. Sorting by t-alpha, portfolios containing between 3 and 6 funds show statistically significant persistence (at a 10% significance level) for monthly rebalancing and a portfolio of 3 or 6 funds is statistically significant for 3 monthly rebalancing. When funds are sorted by t-alpha, the statistically significant forward looking 4-factor alphas range from 0.14 (1.68% p.a.) to 0.21 (2.52% p.a.) and when sorted by returns the alphas are between 0.28 (3.36% p.a.) and 0.79 (8.4% p.a.). What is clear from Table 8.2 is that there is some statistically significant positive persistence for small size portfolios of past winners as long as rebalancing is quite frequent but this may involve high transactions costs, an issue examined later.

Given that the bottom two deciles of past loser funds tend to have statistically significant negative forward looking alphas from Table 8.1, it is not surprising that this carries over to smaller portfolios of past loser funds. For almost all small portfolios of past loser funds sorted by past returns and comprising between 1 and 50 of the worst funds (and for all alternative rebalancing periods), this study finds statistically significant negative forward looking alphas (Table 8.3, Panel A). When the worst performing funds are sorted by past t-alphas, the results are qualitatively similar with negative forward looking 4-factor alphas, very many of which are statistically significant. For example, the portfolio formed from the 5 worst past t-alpha funds gives a forward looking 4-factor alpha of around -0.2% p.m. (-2.4% p.a., $p < 0.01$) for the ranking/rebalancing periods, 60/12, 60/6 and 60/3. There is therefore clear evidence that poor performance persists.

8.2 Final Wealth: Passive versus Active Strategies with Transaction Costs

Comparing risk versus return in the distribution of final wealth from alternative strategies is problematic. While there may be agreement on the metric for the average level of final wealth (e.g. median or mean), the metric for 'risk' is much less clear cut. For example, variance, semi-variance, interquartile range, skewness, kurtosis, lower partial moment, value at risk, expected tail loss, worse case scenario are all potential candidates. In what follows this study therefore does not use a single metric for the relative importance of risk versus return but merely comments on various aspects of the alternative distributions of final wealth, from the alternative investment strategies.

The previous discussion of results indicated that some small portfolios of past winner funds yield statistically significant positive forward looking alphas if frequently rebalanced (every three months or less, Tables 8.1 and 8.2). The study therefore initially presents results for the distribution of final wealth for three month rebalancing periods, with alternative transactions costs of 0%, 2.5% and 5% and alternative size portfolios. Figure 8.1 shows results for funds sorted on past returns while Figure 8.2 shows results for funds sorted on past t-alpha. Each figure shows the median level of final wealth from the bootstrap distribution (as described in chapter 7, section 7.2) as well as the 5% and 95% percentiles for each size portfolio for the active strategy.

For the passive strategy the median level of final wealth from the bootstrap procedure is £1,800 (a geometric average return of 6% p.a.) with a relatively small standard deviation of £180 and a range of £360. The distribution is approximately normal. In this section final wealth is calculated for an investment of £1,000 over a 10 year horizon ending in December 2002.

In the active strategy, when sorting by past raw returns (Figure 8.1), average final wealth (irrespective of transactions costs) falls as the number of funds, z, rises from 1 fund to around 10 funds. From above, the analysis suggests that the very top funds have some persistence, i.e. there is persistence in small portfolios of funds, although from Figure 8.1, even in small portfolios higher average returns are offset as we add funds whose returns vary randomly. However, as long as the portfolio is in excess of about 10 funds, final wealth is not dramatically influenced by the size of the portfolio, particularly when transactions costs are accounted.

Average final wealth falls dramatically (for all size portfolios) when transactions costs are introduced. For example, for a 10-fund portfolio average final wealth falls by 36% for 2.5% transactions costs and by a further 39% when transactions costs are at 5%. However, when we measure risk as the 5^{th}-95^{th} range of wealth outcomes (i.e. $W_T(95\%) - W_T(5\%)$), there is no reduction in 'risk per unit of expect wealth' with higher transactions costs, as this risk-to-wealth ratio (for the 10-fund portfolio) is always around 65% for either zero, or 2.5% and 5% transactions costs.

From Figure 8.1, given that the median level of wealth from the passive strategy is £1,800, it is probably the case that most investors would choose the active strategy with zero transactions costs over the passive strategy. The simulations indicate a zero probability that the active strategy with zero transaction costs would produce a final wealth less than £1,800. When transactions costs are 2.5% (per fund transaction), the median level of final wealth in an active portfolio with less than 20 funds is greater than the median final wealth under the passive strategy, though this is reversed for portfolio sizes greater than 20 funds. For portfolio sizes up to about 10 funds there is a probability of up to 25% (the probability rises with portfolio size) that the active strategy will yield a lower final wealth than the passive strategy. For the larger portfolios this probability rises to 50%. There is much less downside risk with the passive strategy (with a standard deviation of only £180) but at a cost in terms of less upside potential. So the choice between them is relatively even but a loss averse investor would probably favour the passive strategy. With transactions costs at 5% there is only a 5% probability or less that the active strategy will 'beat' a passive strategy even for a very small portfolio of only two funds. For larger portfolios the probability is close to zero.

Turning now to the results based on the t-alpha sort in Figure 8.2. Again, the previous section (Table 8.2) points to some evidence of persistence in small portfolios of funds. Note, first from Figure 8.2, unlike in Figure 8.1, in small portfolios higher average returns are not offset as more funds are added up to around 5 or 10 funds. This is consistent with the results in Table 8.2 where evidence of persistence was found among slightly larger portfolios of funds (up to 6). Second, when sorting by past t-alpha here average final wealth falls less dramatically (for all size portfolios) after transactions costs are incorporated, than when sorting by past returns. For example, for a 10-fund portfolio average final wealth falls by about 26% for 2.5% transactions costs and by a further 26% when transactions costs are at 5%. This latter effect is due to lower proportionate rebalancing costs when sorting by t-alpha rather than past returns since in the former case any large alpha estimates due to (good or bad luck) are scaled by their standard errors, so less funds change their ranking at any rebalancing date (compared with sorting by returns which has no 'correction' for luck). As before, however, as long as the portfolio size is above around 10 funds, when sorting by either past raw returns or past t-alpha final wealth is not dramatically influenced by the size of the portfolio.

Similar to the returns sorted case, when looking at the t-alpha sorted results we find that there is no reduction in the risk-wealth ratio when transactions costs are incorporated. For example, for the 10-fund portfolio this metric is always around 12% for either zero, or 2.5% and 5% transactions costs. The latter compares with 65% for the returns sorted strategy – so the latter has more risk per unit of average wealth than the t-alpha ranking strategy. But as we argued above, looking at any *single* metric for the risk-return trade off is somewhat arbitrary.

Comparing the t-alpha (active) ranking strategy with the passive strategy it is probably the case that nearly all investors would choose the active strategy with zero transactions costs and also with 2.5% transactions costs, since even in the latter case the active strategy has a less than 5% probability of doing worse than the median outcome of £1,800 for the passive strategy (for almost all portfolios sizes). However, with transactions costs at 5%, the simulations indicate that the active strategy has a 0% chance of producing a higher final wealth outcome than the passive strategy for all size portfolios (and has considerably more downside risk) so the choice here seems to favour the passive strategy.

The above results show how important transactions cost are when deciding between a passive or active strategy. The analysis suggests that if transactions costs are less than 2.5% per fund round trip, an active strategy sorting on past t-alpha (with 3 month rebalancing) will with high probability yield a higher final wealth than the passive strategy. With transactions costs of 5% the passive strategy is most probably superior. So there is a *prima facie* case for active management in the UK based on *ex-ante* 'sorting rules', for 'small' fund-of-funds portfolios with rebalancing every 3 months and where transactions costs of rebalancing are at or below 2.5% per fund transaction.

8.3 Conclusion

Chapter 7 and chapter 8 examined the economic and statistical significance of persistence for a large sample of UK 'active' equity mutual funds with a data set from April 1975 to December 2002. Sorting funds into *deciles* based on past raw returns or on past 4-factor t-alphas the study finds strong evidence that past decile loser funds continue to perform badly in terms of their future 4-factor alphas and little evidence that past decile winner funds provide future positive risk adjusted performance – similar to earlier results in Quigley and Sinquefield (1999) using around 750 UK equity funds over the period 1978-

97. However, on investigating relatively small fund-of-funds portfolios of past winners, there is evidence that rebalancing every 3 months (or every 1 month), with portfolios of up to around 5 funds does yield some statistically and economically significant 4-factor alphas. So an active portfolio strategy of avoiding past loser funds and choosing a small portfolio of past winner funds may earn positive risk adjusted returns (after management fees but before the deduction of transactions costs of the active strategy).

Long-term, investors are interested in the distribution of final wealth (e.g. mean, skewness, kurtosis) from an active strategy, relative to that from alternative strategies such as holding index trackers - taking account of 'luck' across *all* funds and transactions costs of rebalancing. Using a cross-section bootstrap approach the study derived the empirical distribution of final wealth over a 10 year horizon for active strategies, incorporating transactions costs in each rebalancing period and compared this with the distribution from a passive strategy of (randomly) holding an index tracker fund over the whole investment horizon (thus avoiding transactions costs). The analysis suggests that if rebalancing is frequent (e.g. every 3 months) and if transactions costs are less than 2.5% per fund round trip, then an active strategy of sorting on past t-alpha seems at least as good as a passive strategy but with transactions costs of 5% the passive strategy is most probably superior. So there is a *prima facie* case for an active strategy in the UK based on *ex-ante* 'sorting rules', for 'small' fund portfolios as long as average switching costs of rebalancing are below 2.5%.

Table 8.1 : Persistence Results – Decile Portfolios

Table 8.1 Panel A presents persistence results for decile portfolios formed on past raw returns for alternative formation and holding periods as indicated. Alpha, t-alpha and the bootstrap p-values of alpha of the forward looking decile portfolios are reported.

PANEL A												
Decile Portfolios	Portfolios formation period and holding period											
	12 months formation 12 months holding			12 months formation 6 months holding			12 months formation 3 months holding			12 months formation 1 months holding		
	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values
1	-0.0458	-0.4469	0.3530	0.0024	0.0272	0.4820	0.0377	0.3822	0.3550	0.0028	0.0287	0.4980
2	-0.0443	-0.6183	0.2710	-0.0773	-1.4030	0.0740	-0.0305	-0.4556	0.2920	-0.0691	-1.0055	0.1490
3	-0.0269	-0.5196	0.3140	-0.0672	-1.3036	0.1080	-0.0759	-1.4636	0.0780	-0.1076	-1.8958	0.0380
4	-0.0654	-1.4608	0.0750	-0.0406	-1.0171	0.1620	-0.0715	-1.7581	0.0440	-0.0238	-0.5526	0.2940
5	-0.0869	-1.9514	0.0190	-0.0699	-1.6894	0.0470	-0.0538	-1.2934	0.1040	-0.0791	-2.0265	0.0170
6	-0.0779	-1.7528	0.0410	-0.0407	-1.0681	0.1530	-0.0660	-1.6694	0.0420	-0.0562	-1.3190	0.0910
7	-0.0749	-1.6718	0.0490	-0.0690	-1.7658	0.0460	-0.0764	-1.7958	0.0390	-0.0838	-1.8127	0.0530
8	-0.0995	-1.8534	0.0290	-0.0987	-1.9395	0.0250	-0.1057	-2.1311	0.0120	-0.1114	-2.3455	0.0120
9	-0.0941	-1.2934	0.1190	-0.1707	-2.7461	0.0000	-0.1448	-2.5164	0.0050	-0.1618	-2.2774	0.0080
10	-0.2426	-2.5894	0.0060	-0.2139	-2.2656	0.0190	-0.2176	-2.6263	0.0070	-0.1401	-1.5682	0.0520

Table 8.1Panel B presents persistence results for decile portfolios formed on past t-alphas for alternative formation and holding periods as indicated. Alpha, t-alpha and the bootstrap p-values of alpha of the forward looking decile portfolios are reported.

PANEL B												
Decile Portfolios	Portfolios formation period and holding period											
	60 months formation 12 months holding			60 months formation 6 months holding			60 months formation 3 months holding			60 months formation 1 months holding		
	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values
1	-0.0001	-0.0012	0.5120	0.0078	0.0971	0.4670	0.0120	0.1463	0.4250	-0.0029	-0.0357	0.5060
2	-0.0602	-0.8468	0.1770	-0.0609	-0.8609	0.1840	-0.0311	-0.4811	0.3120	-0.0597	-0.8702	0.1790
3	-0.0646	-1.0102	0.1650	-0.0460	-0.6996	0.2570	-0.0648	-1.1088	0.1490	-0.0096	-0.1490	0.4530
4	-0.0503	-1.1170	0.1200	-0.0484	-1.0825	0.1290	0.0153	0.3386	0.6240	-0.0552	-1.3863	0.0080
5	-0.0809	-1.9056	0.0270	-0.0599	-1.3062	0.1000	-0.1202	-2.7443	0.0020	-0.0940	-1.9842	0.0260
6	-0.0368	-0.9522	0.1650	-0.0379	-0.9151	0.1850	-0.0582	-1.4618	0.0660	-0.0810	-1.9254	0.0300
7	-0.0887	-1.9603	0.0300	-0.0827	-1.9573	0.0270	-0.0815	-2.0234	0.0290	-0.0941	-2.5591	0.0080
8	-0.6617	-1.0217	0.1620	-0.0835	-1.4459	0.0800	-0.0903	-1.6570	0.0580	-0.0774	-1.6559	0.0610
9	-0.1280	-2.6934	0.0030	-0.1753	-3.7007	0.0010	-0.1530	-2.7343	0.0050	-0.1150	-1.8803	0.0390
10	-0.1426	-2.3797	0.0130	-0.1735	-2.5275	0.0070	-0.1654	-2.4729	0.0060	-0.1404	-2.0707	0.0240

Table 8.2 : Persistence Results – Alternative Size Portfolios (Past Winners)

Table 8.2 Panel A presents persistence results for 'smaller' portfolios of past winners formed on past raw returns for alternative formation and holding periods as indicated. Alpha, t-alpha and the bootstrap p-values of alpha of the forward looking portfolios are reported.

PANEL A												
Portfolio Size	Portfolios formation period and holding period											
	12 months formation 12 months holding			12 months formation 6 months holding			12 months formation 3 months holding			12 months formation 1 months holding		
	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values
1	0.1549	0.3472	0.3660	0.5290	1.6002	0.0260	0.7855	2.5452	0.0010	0.4754	1.4388	0.0970
2	0.1205	0.4498	0.3390	0.2820	1.5155	0.0500	0.4679	1.7646	0.0190	0.3635	1.6662	0.0500
3	0.1147	0.5547	0.3640	0.2911	2.0148	0.0180	0.4094	2.0439	0.0120	0.2132	1.1180	0.1400
4	0.0922	0.4960	0.2910	0.1572	1.2085	0.1190	0.2101	1.1821	0.1060	0.1405	0.7925	0.1950
5	0.0184	0.1110	0.4400	0.0913	0.7455	0.2230	0.1593	0.9734	0.1660	0.0989	0.6044	0.3000
6	-0.0195	-0.1238	0.5280	0.0865	0.7505	0.2150	0.1479	0.9560	0.1790	0.0490	0.3109	0.3820
7	-0.0416	-0.2814	0.6130	0.0437	0.3703	0.3410	0.1013	0.6880	0.2350	0.0519	0.3503	0.3670
8	-0.0623	-0.4280	0.6460	0.0225	0.1863	0.4180	0.1386	1.0016	0.1500	0.0776	0.5425	0.2910
9	-0.0347	-0.2555	0.6010	0.0448	0.3920	0.3610	0.1115	0.8171	0.2210	0.0400	0.2934	0.3970
10	-0.0349	-0.2656	0.5820	0.0274	0.2405	0.4170	0.0972	0.7339	0.2240	0.0481	0.3601	0.3710
20	-0.0433	-0.3603	0.6410	0.0118	0.1168	0.4510	0.0209	0.1885	0.4330	0.0252	0.2270	0.4240
25	-0.0451	-0.3977	0.6400	-0.0010	-0.0099	0.4900	0.0304	0.2865	0.3770	0.0168	0.1584	0.4360
30	-0.0488	-0.4451	0.6790	-0.0058	-0.0628	0.5440	0.0280	0.2771	0.4040	0.0051	0.0498	0.4940
50	-0.0609	-0.6279	0.7400	-0.0421	-0.5170	0.6860	-0.0023	-0.0255	0.4800	-0.0313	-0.3486	0.6340

Table 8.2 Panel B presents persistence results for 'smaller' portfolios of past winners formed on past t-alphas for alternative formation and holding periods as indicated. Alpha, t-alpha and the bootstrap p-values of alpha of the forward looking portfolios are reported.

PANEL B												
Portfolio Size	Portfolios formation period and holding period											
	60 months formation 12 months holding			60 months formation 6 months holding			60 months formation 3 months holding			60 months formation 1 months holding		
	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values
1	-0.0938	-0.6412	0.7430	0.0643	0.5044	0.3220	-0.0349	-0.2758	0.6220	-0.0066	-0.0508	0.5290
2	0.0859	0.8052	0.2250	0.1388	1.3841	0.0950	0.1226	1.2682	0.1130	0.1250	1.2346	0.1130
3	0.0289	0.2840	0.3800	0.1467	1.5337	0.0570	0.1817	1.8969	0.0260	0.2082	2.2587	0.0110
4	-0.0029	-0.0306	0.5240	0.0829	0.9273	0.1770	0.1207	1.3411	0.0900	0.1908	2.2304	0.0160
5	-0.0174	-0.1807	0.5470	0.0784	0.9315	0.1880	0.1359	1.4860	0.0840	0.1629	1.9729	0.0240
6	0.0237	0.2643	0.3690	0.1114	1.3928	0.0810	0.1449	1.7094	0.0410	0.1410	1.7393	0.0440
7	0.0146	0.1631	0.4190	0.0827	1.0141	0.1450	0.1315	1.5270	0.0590	0.1140	1.4068	0.0810
8	0.0342	0.3911	0.3570	0.0812	1.0080	0.1440	0.1089	1.2139	0.1120	0.0686	0.8324	0.2130
9	0.0564	0.6767	0.2420	0.0737	0.9491	0.1800	0.1001	1.1940	0.1240	0.0766	0.9505	0.1690
10	0.0434	0.5073	0.2940	0.0563	0.7041	0.2350	0.0682	0.7835	0.2130	0.0224	0.2666	0.3970
20	-0.0193	-0.2308	0.5780	-0.0382	-0.4487	0.6540	0.0079	0.0947	0.4330	0.0009	0.0107	0.4800
25	0.0020	0.0240	0.4760	-0.0052	-0.0631	0.5180	0.0369	0.4480	0.3220	0.0065	0.0814	0.4590
30	-0.0115	-0.1444	0.5400	0.0023	0.0286	0.4620	0.0213	0.2686	0.3710	-0.0113	-0.1393	0.5300
50	-0.0506	-0.6762	0.7690	-0.0373	-0.5101	0.7080	-0.0208	-0.2813	0.6300	-0.0261	-0.3504	0.6560

Table 8.3 : Persistence Results – Alternative Size Portfolios (Past Losers)

Table 8.3 Panel A presents persistence results for 'smaller' portfolios of past losers formed on past raw returns for alternative formation and holding periods as indicated. Alpha, t-alpha and the bootstrap p-values of alpha of the forward looking portfolios are reported.

PANEL A												
Portfolio Size	Portfolios formation period and holding period											
	12 months formation 12 months holding			12 months formation 6 months holding			12 months formation 3 months holding			12 months formation 1 months holding		
	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values
1	-0.3542	-1.0611	0.8440	-0.2737	-0.8798	0.7850	-0.5387	-1.9534	0.9610	-0.3443	-0.8209	0.7520
2	-0.4528	-2.0255	0.9760	-0.5395	-2.4543	0.9830	-0.7413	-3.6554	1.0000	-0.4869	-1.8073	0.9460
3	-0.4344	-2.1585	0.9830	-0.5740	-2.7708	0.9980	-0.6561	-3.5549	0.9990	-0.3397	-1.4324	0.8960
4	-0.3832	-2.2185	0.9790	-0.5051	-2.8644	0.9950	-0.6349	-3.8060	1.0000	-0.4596	-2.3073	0.9810
5	-0.4528	-2.8375	0.9930	-0.5067	-3.0101	0.9990	-0.6325	-3.9342	1.0000	-0.4233	-2.4150	0.9890
6	-0.4395	-3.0643	0.9970	-0.4657	-3.0649	0.9980	-0.5725	-3.7349	1.0000	-0.4028	-2.4907	0.9930
7	-0.4345	-3.0568	0.9990	-0.4552	-3.1581	1.0000	-0.5730	-3.8661	1.0000	-0.3679	-2.7326	0.9910
8	-0.4120	-3.0637	0.9970	-0.4305	-3.1706	0.9990	-0.5163	-3.6971	0.9990	-0.3258	-2.2162	0.9810
9	-0.4021	-3.0796	1.0000	-0.4190	-3.2481	1.0000	-0.5066	-3.8015	1.0000	-0.3314	-2.3637	0.9880
10	-0.4010	-3.2288	0.9970	-0.3721	-2.8860	0.9960	-0.4622	-3.5945	1.0000	-0.3108	-2.2883	0.9900
20	-0.3585	-3.2760	1.0000	-0.3114	-2.9357	0.9960	-0.3608	-3.4267	1.0000	-0.2374	-2.1762	0.9850
25	-0.3273	-3.1387	1.0000	-0.2978	-2.8993	0.9980	-0.3319	-3.3132	0.9990	-0.2271	-2.1931	0.9880
30	-0.2991	-3.0143	0.9990	-0.2754	-2.7437	0.9930	-0.2965	-3.0542	0.9980	-0.2036	-2.0387	0.9840
50	-0.2158	-2.4384	0.9880	-0.2116	-2.3386	0.9820	-0.2243	-3.0712	0.9970	-0.1863	-2.4030	0.9910

Table 8.3 Panel B presents persistence results for 'smaller' portfolios of past losers formed on past t-alphas for alternative formation and holding periods as indicated. Alpha, t-alpha and the bootstrap p-values of alpha of the forward looking portfolios are reported.

PANEL B												
Portfolio Size	Portfolios formation period and holding period											
	60 months formation 12 months holding			60 months formation 6 months holding			60 months formation 3 months holding			60 months formation 1 months holding		
	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values	alpha	t-alpha	p-values
1	-0.2177	-2.3512	0.9930	-0.0876	-1.2245	0.8920	-0.0989	-0.9051	0.8040	-0.1408	-1.3265	0.9000
2	-0.2745	-3.1706	1.0000	-0.1550	-1.9071	0.9670	-0.1515	-1.7085	0.9620	-0.1497	-1.5887	0.9450
3	-0.1701	-2.3542	0.9880	-0.1742	-1.9504	0.9670	-0.2007	-2.5767	0.9950	-0.1697	-2.0130	0.9690
4	-0.2052	-2.9432	0.9960	-0.2119	-2.5791	0.9910	-0.2105	-2.7892	0.9960	-0.1609	-2.0064	0.9750
5	-0.2013	-2.9144	0.9950	-0.2395	-3.5669	1.0000	-0.2460	-3.6790	1.0000	-0.0810	-0.5785	0.6070
6	-0.1136	-0.9428	0.7070	-0.1457	-1.2050	0.7650	-0.2452	-3.6153	0.9990	-0.1082	-0.8911	0.6930
7	-0.1175	-1.1066	0.7780	-0.1589	-1.4521	0.8370	-0.2340	-3.5022	0.9990	-0.1175	-1.0634	0.7750
8	-0.1434	-1.4830	0.8560	-0.1663	-1.6396	0.8880	-0.1562	-1.5801	0.8700	-0.1127	-1.0854	0.7810
9	-0.1411	-1.5789	0.8680	-0.2053	-2.1689	0.9480	-0.1637	-1.7930	0.9100	-0.0987	-0.9964	0.7690
10	-0.1355	-1.6483	0.9160	-0.1887	-2.1851	0.9570	-0.1609	-1.8588	0.9320	-0.1162	-1.2616	0.8590
20	-0.1360	-1.9096	0.9680	-0.1596	-2.3086	0.9820	-0.1645	-2.4719	0.9900	-0.1371	-1.9582	0.9590
25	-0.1007	-1.7431	0.9560	-0.1504	-2.2423	0.9970	-0.1502	-2.3896	0.9870	-0.1300	-2.0153	0.9760
30	-0.1080	-1.8790	0.9630	-0.1691	-2.6189	0.9980	-0.1569	-2.5832	0.9970	-0.1387	-2.2327	0.9890
50	-0.1201	-2.3189	0.9890	-0.1546	-2.5915	0.9940	-0.1379	-2.3237	0.9910	-0.1168	-1.8956	0.9760

Figure 8.1: Average Final Wealth – Active Strategy Forming on Past Raw Returns

Figure 8.1 shows the median (solid lines) from the bootstrapped distributions of final wealth in the case of each portfolio size (1 to 50) as indicated. The three lines show median wealth for 0%, 2.5% and 5% transaction costs as indicated. In the case of each transaction cost and each portfolio size the figure also shows the 5^{th} and 95^{th} percentile of the bootstrap distribution. The figures relate to final wealth from an active strategy of sorting on past 12 month raw returns and rebalancing every 3 months.

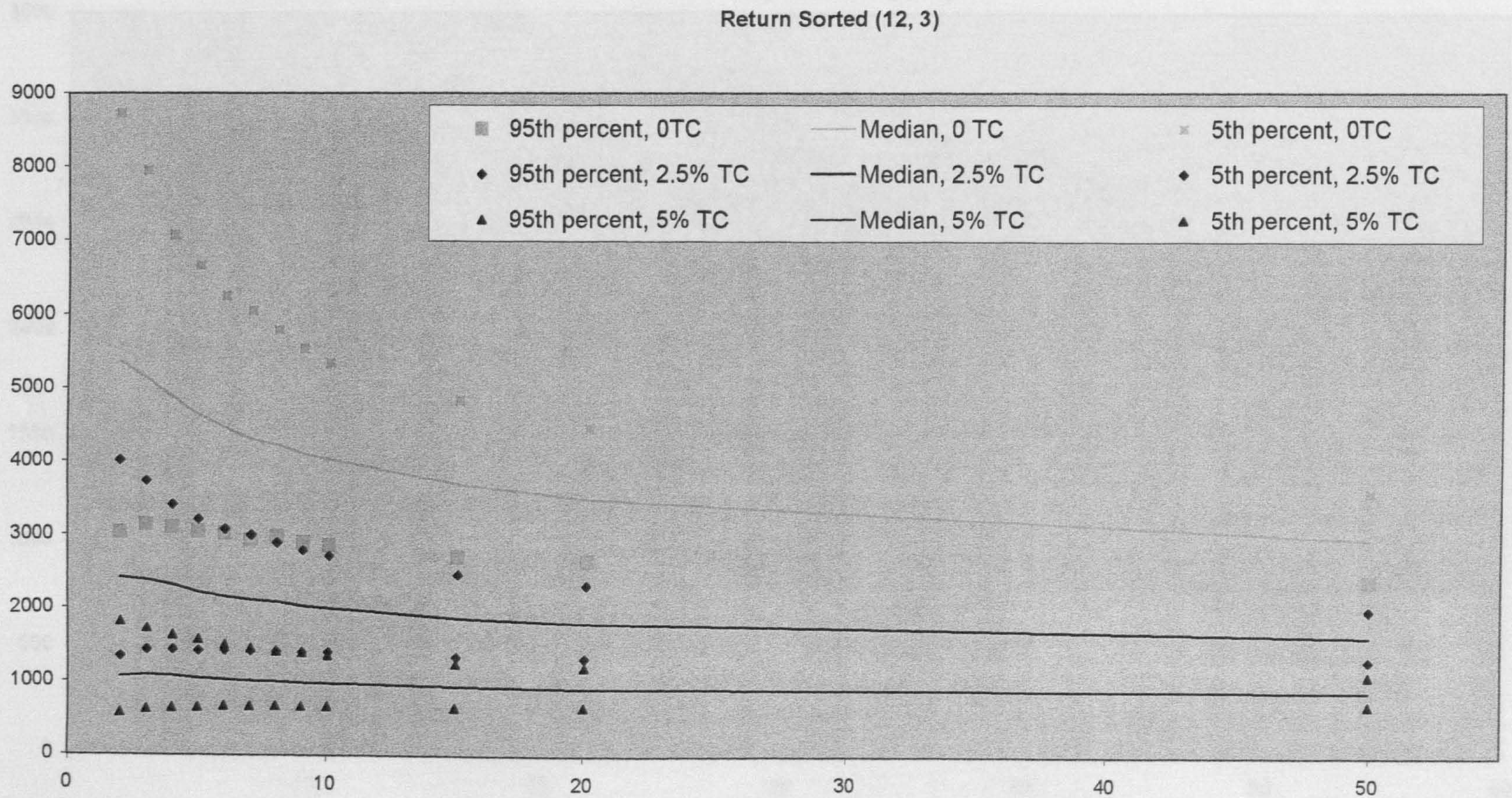
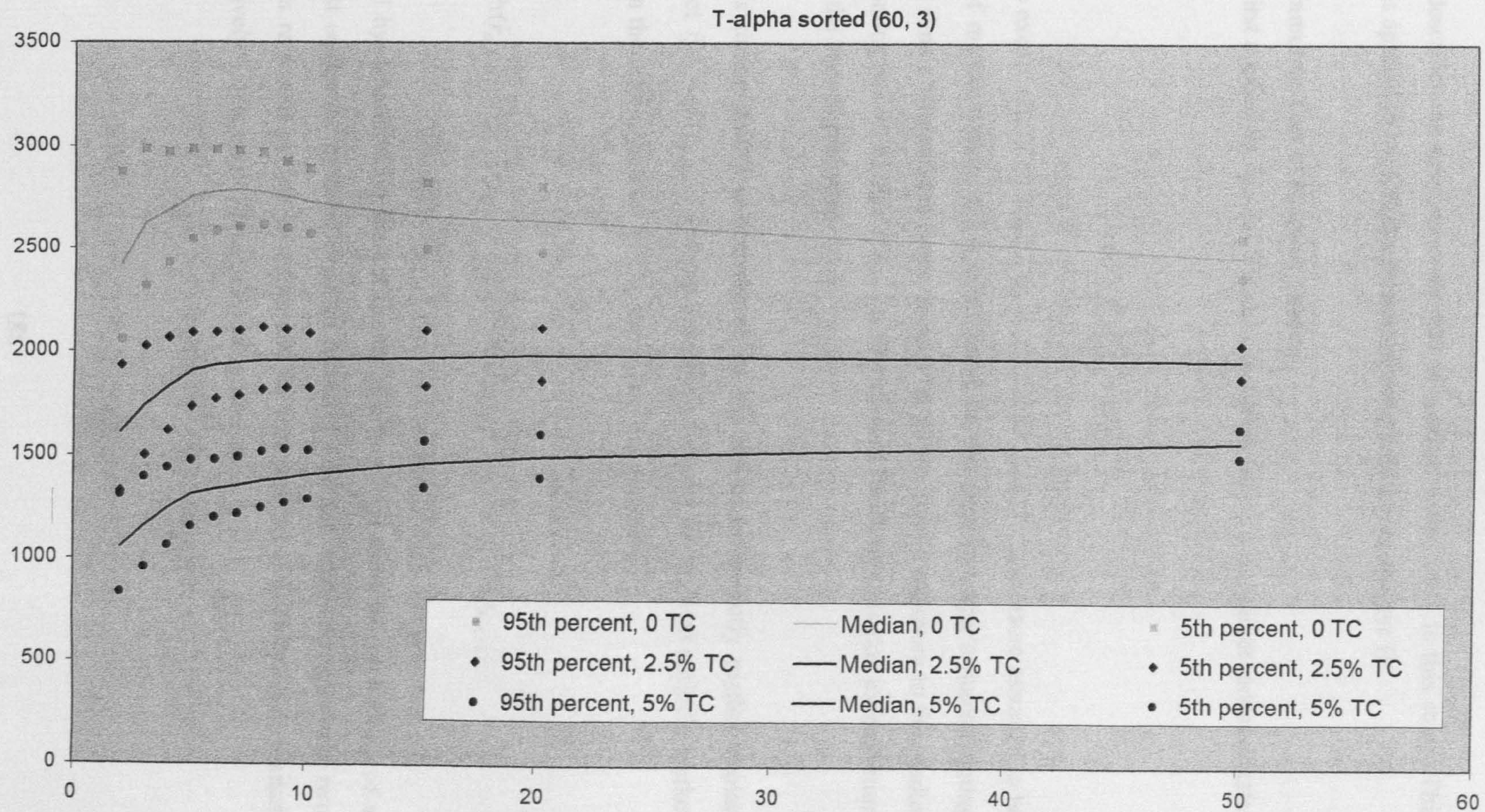


Figure 8.2: Average Final Wealth – Active Strategy Forming on Past t-alpha

Figure 8.2 shows the median (solid lines) from the bootstrapped distributions of final wealth in the case of each portfolio size (1 to 50) as indicated. The three lines show median wealth for 0%, 2.5% and 5% transaction costs as indicated. In the case of each transaction cost and each portfolio size the figure also shows the 5th and 95th percentile of the bootstrap distribution. The figures relate to final wealth from an active strategy of sorting on past 60 month t-alpha and rebalancing every 3months.



CHAPTER 9

NONPARAMETRIC MARKET TIMING METHODOLOGY

This chapter describes the nonparametric test of market timing used in this study. This methodology is applied to the UK equity mutual fund industry in chapter 10.

9.1 A Nonparametric Test of Market Timing

The test was first applied by Jiang (2003) to US mutual funds. The market model dictates that

$$r_{i,t+1} = \alpha_i + \beta_{i,t} r_{m,t+1} + \varepsilon_{i,t+1} \tag{9.1}$$

where r_i is the excess return on fund i, α_i is a security selectivity measure assumed to be independent of market timing, $\beta_{i,t}$ is conditional on the fund manager's market timing information at time t. The market timer decides on β_t, the fund's exposure to the market for the forthcoming period, at time t. $r_{m,t+1}$ is the relevant benchmark market excess return against which the fund is evaluated.

The manager's timing ability is determined by his ability to correctly predict market movements. Let $\hat{r}_{m,t+1} = E(r_{m,t+1} \mid I_t)$ be the manager's forecast for the next period's market return based on the information set I_t. The parameter ν is defined as

$$\nu = \Pr(\hat{r}_{m,t_2+1} > \hat{r}_{m,t_1+1} \mid r_{m,t_2+1} > r_{m,t_1+1}) - \Pr(\hat{r}_{m,t_2+1} < \hat{r}_{m,t_1+1} \mid r_{m,t_2+1} > r_{m,t_1+1}) \tag{9.2}$$

Under the null hypothesis of no market timing ability $\nu = 0$ since the probability of a correct forecast equals the probability of an incorrect forecast. $\nu \in [-1,1]$ where the two extreme values represent perfect negative and perfect positive (i.e. successful) market timing respectively. Equation (9.2) may also be written as:

$$\nu = 2\Pr(\hat{r}_{m,t_2+1} > \hat{r}_{m,t_1+1} \mid r_{m,t_2+1} > r_{m,t_1+1}) - 1 \tag{9.3}$$

The next step is to link the manager's forecast of the market with his response in adjusting $\beta_{i,t}$ in (9.1). For any triplet of market return observations $\{r_{m,t_1}, r_{m,t_2}, r_{m,t_3}\}$ sampled from *any* three time periods (not necessarily in consecutive chronological order) with $\{r_{m,t_1} < r_{m,t_2} < r_{m,t_3}\}$ an informed timer will maintain a higher exposure to the market over the $[r_{m,t_2}, r_{m,t_3}]$ range than in the $[r_{m,t_1}, r_{m,t_2}]$ range. Simple (nonparametric) beta estimates for both time ranges are $\beta_{t1} = (r_{i,t_2} - r_{i,t_1})/(r_{m,t_2} - r_{m,t_1})$ and $\beta_{t2} = (r_{i,t_3} - r_{i,t_2})/(r_{m,t_3} - r_{m,t_2})$. Beta embodies both the precision of the market return forecast and the aggressiveness of the manager's response. The latter is affected by risk aversion. Grinblatt and Titman (1989) show that for a fund *i* with non-increasing absolute risk aversion and independent timing and selectivity information $\frac{\partial \beta_t}{\partial \hat{r}_{m,t+1}} > 0$ yielding a convex fund return/market return relationship

(9.4) $$\frac{r_{i,t_3} - r_{i,t_2}}{r_{m,t_3} - r_{m,t_2}} > \frac{r_{i,t_2} - r_{i,t_1}}{r_{m,t_2} - r_{m,t_1}}$$

which allows (9.3) to be written as

(9.5) $$v = 2\Pr(\beta_{t_2} > \beta_{t_1} \mid r_{m,t_2+1} > r_{m,t_1+1}) - 1$$

A sample statistic of a fund's timing ability may be constructed as follows: assigning a sign function that assumes a value of 1(-1) if the argument is positive (negative) and equals zero if the argument equals zero, it is a *U*-statistic with kernel of order three.

(9.6) $$\hat{\theta}_n = \binom{n}{3}^{-1} \sum_{r_{m,t_1} < r_{m,t_2} < r_{m,t_3}} sign\left(\frac{r_{i,t_3} - r_{i,t_2}}{r_{m,t_3} - r_{m,t_2}} > \frac{r_{i,t_2} - r_{i,t_1}}{r_{m,t_2} - r_{m,t_1}} \right)$$

$\hat{\theta}_n$ is the average sign across all triplets taken from *n* observations. $\hat{\theta}_n$ is a √n-consistent and asymptotically normal estimator (Abrevaya and Jiang, 2001, Serfling, 1980).

Hence $\sqrt{n}(\hat{\theta}_n - \theta) \to N(0, \sigma^2_{\hat{\theta}_n})$

where

$$(9.7) \qquad \hat{\sigma}^2_{\hat{\theta}_n} = \frac{9}{n}\sum_{t_1=1}^{n}\left(\binom{n}{2}^{-1}\sum_{t_2<t_3, t_1\neq t_2, t_1\neq t_3} h(z_{t_1}, z_{t_2}, z_{t_3}) - \hat{\theta}_n\right)^2$$

and

$$(9.8) \qquad h(z_{t_1}, z_{t_2}, z_{t_3}) = sign\left(\frac{r_{i,t_3} - r_{i,t_2}}{r_{m,t_3} - r_{m,t_2}} > \frac{r_{i,t_2} - r_{i,t_1}}{r_{m,t_2} - r_{m,t_1}} \mid r_{m,t_1} < r_{m,t_2} < r_{m,t_3}\right)$$

Under the null hypothesis of no market timing $z = \sqrt{n}.\hat{\theta}_n / \hat{\sigma}_{\hat{\theta}_n}$ is asymptotically N(0,1). Note, the calculation in (9.7) includes triplets $h(z_{t_1}, z_{t_2}, z_{t_3}), h(z_{t_2}, z_{t_1}, z_{t_3}), h(z_{t_3}, z_{t_1}, z_{t_2})$, i.e. the same three market return observations drawn in different combinations. However, the sign in (9.8) is equal in all three cases as from (9.8) it is conditional on $r_{m,t_1} < r_{m,t_2} < r_{m,t_3}$, i.e. irrespective of the order in which the market return observations are drawn they are first sorted in ascending order and there can only be one such sorting.

One difficulty in examining a fund's market timing skill is decomposing the quality of the manager's information concerning the future market return and the aggressiveness with which he responds to this information. A rational investor is more concerned with the former as the investor can control the latter himself by choosing the proportion of their wealth to be invested with the mutual fund manager rather than in cash. The market timing tests of Treynor-Mazuy (TM) and Henriksson-Merton (HM), as outlined in the literature review section of this study, do not decompose the quality of market timing signals from the aggressiveness of response in market timing.

Recall from the literature review that the TM test specifies a quadratic regression of the form

(9.9) $$r_{i,t+1} = \alpha_i + \theta_i(r_{m,t+1}) + \gamma_{iu}[r_{m,t+1}]^2 + \varepsilon_{i,t+1}$$

where the coefficient γ_{iu} measures market timing ability. $r_{i,t+1}$ and $r_{m,t+1}$ are the fund and market excess returns respectively. Admati et al (1986) demonstrate that the model is consistent with a manager with constant absolute risk aversion who adjusts the portfolio beta at time *t* according to a private linear signal of the form

(9.10) $$\beta_{it} = \theta_i + \gamma_{iu}[r_{m,t+1}] + \eta_t$$

where η_t is random noise. The hypothesis of no abnormal timing performance implies $\gamma_{iu} = 0$.

Henriksson and Merton (1981) propose a model in which the conditional portfolio beta has two target values in a binary response function depending on the manager's forecast of whether market return will exceed the risk free rate. The authors show that if the manager can successfully time the market then the coefficient γ_{iu} in the following regression will be positive.

(9.11) $$r_{i,t+1} = \alpha_i + \theta_i(r_{m,t+1}) + \gamma_{iu}[r_{m,t+1}]^+ + \varepsilon_{i,t+1}$$

where $[r_{mt+1}]^+$ is defined as $\max(0, r_{mt+1})$.

In the TM measure the estimated market timing coefficient in (9.9) clearly picks up on the response (aggressiveness) function, γ_{iu}, in (9.10). Consequently more aggressive response may show up as a higher timing coefficient in (9.9). Similarly, Henrikkson-Merton (1981) show that $(p_1 + p_2 - 1).(\eta_2 - \eta_1)$ is a consistent estimate of γ_{iu} in 9.11, where p_1 and p_2 are the conditional probabilities of the manager correctly forecasting negative and positive market excess returns respectively in period *t+1* and η_1 and η_2 are the fund target betas in each case. Hence the estimated HM timing measure in 9.11, $\hat{\gamma}_{iu}$, incorporates both the quality of manger information, $p_1 + p_2 - 1$, and the aggressiveness of response, $\eta_2 - \eta_1$. The nonparametric measure of Jiang (2003) on the other hand simply measures how often a manager correctly forecasts a market movement and acts on it, irrespective of how aggressively he acts on it. This is reflected in the fact that the sign function in (9.6) assigns a value of 1(-1) if the argument is positive (negative) regardless of the size of the argument.

A further advantage of the nonparametric measure is that it is more robust in testing for timing skill among managers whose timing frequency may differ from the frequency of the sample data and/or whose timing frequency may not be uniform. The timing statistic in (9.6) investigates timing over *all* triplets of fund returns rather than just consecutive observations and consequently uses more information than the regression based tests. Furthermore, therefore, the nonparametric measure permits the cross-section of fund managers to have different timing strategies. The model based approaches of TM and HM, however, are more restrictive in that they assume knowledge of the timing frequency of the manager and that this is the same across managers.

As discussed in the earlier section on the literature review, the HM regression may exhibit conditional heteroscedasticty. Breen et al (1986) show that the test that ignores heteroscedasticity falsely rejects the true null hypothesis of no market timing too often while the probability of failing to reject the null hypothesis when it is not true also increases significantly. This can significantly affect the conclusions of the HM test. However, as noted by Jiang (2003) the asymptotic distribution of the nonparametric timing measure in (9.6) is unaffected by heteroscedasticity in fund returns.

The nonparametric test is mildly restrictive in behaviour but less restrictive than standard parametric tests: The nonparametric test requires β_t be a non-decreasing function of $\hat{r}_{m,t+1}$. Grinblatt and Titman (1989) demonstrate that this requires (i) non-increasing absolute risk aversion, (ii) independently and identically distributed (*iid*) market returns and (iii) independent selectivity and timing information.

First, the requirement of non-increasing absolute risk aversion is less stringent than that of the TM and HM measures which require more specific linear and binary response functions respectively.

Second, the *iid* assumption rules out heteroscedasticity in market returns and hence volatility timing by fund managers. As noted earlier in the discussion on market timing literature, a manager may reduce the fund's exposure during periods of expected high volatility and consequently if market volatility and market return are positively correlated, the market timing measure may under-estimate the quality of the manager's timing information. If market volatility and return are uncorrelated then the timing measure remains consistent in the presence of volatility timing. US evidence suggests that the relation between market return and volatility is weak (Breen et al 1989, Glosten et al 1993 and Busse 1999).

Third, it is a common assumption among market timing studies that the manager's market timing and selectivity information are independent (Admati et al, 1986, Grinblatt and Titman, 1989). The nonparametric procedure makes a similar assumption. However, as discussed in the earlier literature review, Jagannathan and Korajczyk (1986) question this assumption, particularly when applied to portfolios containing options and option-like securities (such as the common stock of highly levered firms) with nonlinear pay-off structures. The Jiang measure, like that of TM and HM, cannot distinguish between market timing and spurious option related effects. However, as described in the section on data description, all funds examined in this study are composed of at least 80% UK domestic equity so any distortion due to holding options is likely to be relatively small.

9.2 Public versus Private Information: Conditional Market Timing

It is interesting to examine whether mutual fund managers are able to add market timing value to investors by the quality of the manager's private market timing information (timing signals) in excess of the information quality contained in publicly available information. The latter information is already available to the investor and as such may reduce the need for the investor to avail of mutual fund services.

The nonparametric test can be applied as a conditional statistic after allowing for market timing skill attributable to public information. This conditional measure involves first calculating both sets of residuals from regressions of the mutual fund returns and market returns on the set of public information variables. Clearly, these residuals represent the variation in the fund and market returns not explained by the public information. Denote the pairwise fund and market residuals as $\tilde{r}_{i,tj}$ and $\tilde{r}_{m,tj}$, j = 1,2,3...T, respectively. The procedure described above in (9.6) may then be applied to the residuals to yield a conditional timing measure

$$(9.12) \qquad \tilde{\theta}_n = \binom{n}{3}^{-1} \sum_{\tilde{r}_{m,t_1} < \tilde{r}_{m,t_2} < \tilde{r}_{m,t_3}} sign\left(\frac{\tilde{r}_{i,t_3} - \tilde{r}_{i,t_2}}{\tilde{r}_{m,t_3} - \tilde{r}_{m,t_2}} > \frac{\tilde{r}_{i,t_2} - \tilde{r}_{i,t_1}}{\tilde{r}_{m,t_2} - \tilde{r}_{m,t_1}} \right)$$

Note, $\hat{\theta}_n$ in (9.6) and $\tilde{\theta}_n$ in (9.12) can clearly be of different magnitudes but may also be of different sign. For example, $\hat{\theta}_n$>0 but $\tilde{\theta}_n \leq 0$ may indicate a successful market timing manager whose skill is attributable to public information.

This study examines conditional market timing using a set of public information variables commonly applied in this area of the literature, (see Ferson and Schadt, 1996). These variables may have market predictability. They include (i) the one month UK Tbill rate, (ii) the market divided yield and (iii) the term spread (10 year – 1 month yield spread) and (iv) the gilt/equity yield ratio. The gilt/equity yield ratio is the ratio of the coupon yield on a long term government bond to the market dividend yield. The coupon yield should be that of a long term bond to more closely resembles the expected maturity of equity. The ratio captures the relative attractiveness of bonds versus equity and as such predicts capital flows and price movements in both markets. (See Clare, Wickens and Thomas 1994). In this study the yield on a 30 year UK government bond is used.

Finally, in this study, the nonparametric tests (both conditional and unconditional) as described above are implemented for the sample of UK equity mutual funds. However, for these tests the sample period is shortened to 15 years and runs from January 1988 to December 2002, ie 180 monthly observations. This is necessary in order to reduce the computational intensity involved in calculating the cross-sectional set of $\hat{\theta}_n$ in (9.6) which rises exponentially with the number of fund observations. As described above, the calculation of $\hat{\theta}_n$ involves assessing market timing over all triplets of n return observations. If the longer full sample period was used, a fund with $n = 333$ observations involves examining $\binom{333}{3}$ = 6,099,006 triplets. Even a high speed computer takes a substantial length of time to calculate the market timing statistic of such long history funds. Because market timing is examined for all triplets of return observations, 15 years (180 obs.) is regarded here as a sufficient length of time to reliably gauge whether timing skill exists. Including funds of longer life spans would be unlikely to substantially alter overall conclusions. However, a minimum observation requirement of 12 monthly observations is imposed in order to improve the statistical reliability of the results.

CHAPTER 10

EMPIRICAL RESULTS OF MARKET TIMING TESTS

This chapter presents the results from the implementation of the nonparametric market timing tests described in chapter 9. Results from both unconditional and conditional tests are reported while possible differences in market timing skills between different fund investment styles and fund locations are also examined. The sensitivity of results to the chosen market benchmark is also discussed. Results relate to the period 1988M1:2002M12 unless otherwise stated (see chapter 9 on methodology). All tables and figures of results are presented together at the end of the chapter.

10.1 Market Timing Performance – All Investment Objectives

The findings from the unconditional market timing tests are presented in Table 10.1, that is tests which are not conditional on public information. Table 10.1 presents market timing statistics at various points in the cross-section of performance ranging from the best to the worst performer. Fund performance is sorted by the test statistic, $z = \sqrt{n}.\hat{\theta}_n / \hat{\sigma}_{\hat{\theta}_n}$, in row 1. To further improve statistical reliability results are reported for funds with a minimum of 12 observations which leaves 791 funds in the analysis. Row 2 in Table 10.1 displays the market timing coefficient, $\hat{\theta}_n$.

From row 1, it is evident that there are a relatively small number of skilled market timers as only the top 12 ranked funds demonstrate statistically significant positive market timing ability at the 5% significance level (one-tail test), around 1.5% of the sample of funds. In fact the cross-sectional average timing test statistic is –0.738. More specifically, 77% of funds demonstrate negative (perverse) market timing. 12% of funds show statistically significant negative market timing. Positive market timing is found for 23% of funds but this is significant for only 1.5% of funds. Figure 10.1 plots a histogram of the cross-sectional distribution of the market timing test statistics and it is clear that the distribution is centred on a value less than zero (indicating negative market timing ability on average).

Overall, the nonparametric market timing test procedure fails to find evidence of timing ability among more than a 'handful' of actual UK equity mutual funds. For comparison, in Table 10.1 row 3 and row 4 present the t-statistics of the market timing coefficients of the Treynor-Mazuy (TM) and Henriksson-Merton (HM) tests for the funds as ranked in row 1. Interestingly, 10 of the top 12 funds which are found to have statistically significant market timing ability by the nonparametric test are also found to be successful market timers using the TM procedure while 11 yield statistically significant timing coefficients by the HM procedures (at the 5% significance level). The overall picture regarding the market timing ability of UK equity unit trusts differs slightly across the different testing methods where 31 (22) funds are found to have significant timing skill by the TM (HM) procedures. A matrix of correlation coefficients between the market timing test statistics of the three procedures reveals a higher coefficient of 0.95 between the TM and HM tests than the nonparametric/TM correlation coefficient of 0.81 or the nonparametric/HM correlation coefficient of 0.86. Jiang (2003) reports similar findings and suggests that the higher correlation between the TM and HM measures may arise because these methods capture not only the quality of the fund manager's timing information but also the aggressiveness of his response to this information (see discussion on methodology). The nonparametric timing coefficient, on the other hand, ignores the aggressiveness of response in timing. This methodological difference may also account for the slightly higher prevalence of timing found by the TM and HM methods relative to the nonparametric procedure.

Comparing the nonparametric market timing test results here with the bootstrap stock selection results in chapter 6, there is no overlap between the best performing funds in each case, i.e. funds which are found to have stock picking skill are not the highest performing market timers. In fact across the full sample of funds the cross-sectional correlation coefficient between the funds' nonparametric market timing statistic and t-statistics of alpha is only 0.05. This may suggest that funds generally do not pursue dual strategies of stock selection and market timing. However, as noted by Jiang (2003) and discussed previously in chapter 9 on methodology, such a conclusion is complicated by the fact that a fund manager's stock selection and market timing information may be correlated and both performance measures may bias one another.

Clearly, evaluating the timing skill of fund managers may depend on the market benchmark against which funds are assessed. The market timing results discussed above are based on the FTSE A All Share (total) returns index. The market index that actual fund managers attempt to time (if any) is generally unknown in the case of the UK[1]. To examine a possible bias in results from selecting an incorrect market benchmark and to examine the robustness of findings, in this study the market timing tests above are repeated employing an alternative market benchmark: the (total) returns on the FTSE 100 index. The results of this procedure are reported in row 5, Table 10.1. The results between the two benchmark indices are broadly similar. Using the FTSE 100 index, 6 funds (around 1%) demonstrate statistically significant market timing skill by the nonparametric test. The same 6 funds were also identified as successful timers against the FT A All Share index. In the case of the FTSE 100 index, a higher 24% of funds display significant negative timing. Generally, however, the timing tests are robust to the choice of market index.

Similar to the earlier chapters in this study on the bootstrap evaluation of security selection skill, this section on market timing also mitigates against survivorship bias by including nonsurviving funds in the analysis[2]. Of the 791 funds examined, 208 are nonsurvivors. In Table 10.1, the row denoted 'Survival' indicates whether the sorted funds were survivors or nonsurvivors: 1 = a survivor, 0 = a nonsurvivor. None of the funds which demonstrate statistically significant positive timing ability are nonsurvivors. This under-representation of nonsurviving funds among the top performing market timers differs from the disproportionately high number of nonsurvivors among funds found to have security selection skill by the bootstrap analysis. However, as hypothesised previously, it may well be such 'selectivity' skill which motivates the acquisition/merger of a number of these funds causing them to appear as 'nonsurvivors' in the first instance.

These market timing results for the UK mutual fund industry are broadly in line with those of Jiang (2003) for the US industry. Jiang reports unconditional test results for a number of alternative benchmark market indices but consistently finds that less than 5% of the sample of US funds possess statistically significant positive timing skill. For some selected indices, this figure is less than 2%. Similarly, Jiang also finds that the average US fund manager displays negative timing ability. One difference in findings between the two studies is in the relationship between fund age and fund market timing performance.

[1] Except in the case of index tracker funds but these are excluded in this study of market timing.
[2] However, there is a possible look-ahead bias. This arises because of the restriction that a fund must possess a minimum of 12 monthly observations to be included in the analysis. This restriction is imposed to improve the statistical reliability of the market timing estimates.

The final row of Table 10.1 reports the number of (monthly) observations of the sorted funds[3]. It is evident that better performing market timers are generally shorter-lived. In results not shown, the average market timing test statistic among funds of between 1 and 5 years maturity is –0.493 while among funds of greater than 10 years maturity is –0.936, although note both figures are negative and statistically insignificant. Jiang (2003) similarly reports negative and insignificant market timing (on average) among these age categories but the finding is reversed in that the author finds that market timing performance improves, rather than worsens with age. Interestingly, the poorer timing performance among longer-lived UK funds is consistent with the relatively poor selectivity skill from the bootstrap analysis previously. A number of possible explanations for this finding were posited in chapter 6.

10.2 Market Timing and the Sample Period

In chapter 9 on methodology, it was explained that for computational reasons the sample period is restricted to January 1988 – December 2002 (15 years). Although this is a sufficiently long sample period from which to draw overall inferences, it nevertheless raises a question regarding whether market timing results could be sample period specific. More particularly, might timing ability vary according to market conditions, i.e. a bull or bear market, a sharp versus moderate upward or downward market trend etc?

Figure 10.2 shows a time series plot of the market index used in this study, the FTSE A All Share Index over the full sample period January 1975 – December 2002. A casual inspection suggests that while the market is generally rising over the period, the upward trend is stronger in the latter half of the period (aside from a levelling off post 1999). To examine whether the market timing skills of funds may depend on the strength of the market trend, this study repeats the nonparametric market timing tests for the earlier 15 year period: 1975 – 1989. Results are presented in row 2 of Table 10.2 where row 1 repeats the results of Table 10.1 for comparison. At the upper end of the performance distributions there are the same number of successful market timers in both subperiods, i.e. 12 funds at 5% significance. (However, the successful funds in the two sub-periods are not the same funds). Interestingly, however, the percentage of funds showing positive (negative) timing statistics is 55% (45%) in the earlier period compared to 77% (23%) in the later period. This indicates that the degree of market mis-timing (on average across

[3] This is examined to address the question of whether the age of the fund leads to better market timing.

funds) was higher during the stronger upward market trend. One possible explanation for this is the volatility timing described in chapter 9 on methodology: market 'mis-timing' may be rational if funds are reducing the fund's beta during a high volatility/high return period - if these conditions are positively correlated. (see Breen et al 1989, Glosten et al 1993 and Busse 1999).

10.3 Market Timing Performance and Investment Style

As before with security selection skill, it is of interest to examine whether there are differences in market timing skills between funds of different investment objectives, i.e. income funds, general equity funds, small stock funds etc. Cleary, this would be a valuable input in an investor's asset allocation decisions. However, there is some potential for spurious timing inferences across fund investment styles. One difficulty is the assumed independence between selectivity and market timing. A manager's information in both these areas may be correlated and consequently selectivity and market timing inferences may bias one another. For example, it has been argued that small stock funds may exhibit spurious timing against a market benchmark comprised of large stocks, (Jagannathan and Korajczyk, 1986). This is because small stocks have characteristics resembling that of a call option on the market: a bull market leads to a high pay-off, a bear market is more likely to mean financial distress for a small stock and no pay-off. This may lead to the appearance that small stock funds are market timers. Alternatively, it may be argued that general equity funds select from the broadest universe of stocks which make up the benchmark market portfolio, again creating an overlap between selectivity and timing decisions.

To examine this question further, Table 10.3 reports the nonparametric market timing test statistics of funds by investment style as indicated in each panel. (Results here relate to the period 1988M1:2002M12). Comparing row 1 of each panel it is clear that there is some evidence of superior market timing ability for both equity income funds and general equity funds in the extreme right tails of the distributions. However, small stock funds perform very poorly. The average market timing test statistic among the small stock funds is –1.55 compared to –0.62 and –0.57 among the equity income and general equity funds respectively. This comparatively poor performance is also evident in Figure 10.3 which shows the three performance distributions. The results of the TM and HM procedures point to a similar conclusion. This would appear to contradict the argument above that small stock funds may demonstrate spurious market timing ability because of possible option-like characteristics.

However, interestingly there is evidence that a number of small stock funds attempt instead to time a small capitalisation market benchmark rather than a broader market benchmark. In Panel C, the row denoted 'HGSC' reports the nonparametric market timing test statistics of small stock funds measured against the Hoare Govett Small Capitalisation index of UK small stocks. The cross-sectional distribution reported in this row lies considerably further to the right of the distribution presented in row 1 using the broader FTSE A All Share market returns. The extent of this difference provides evidence that UK small stock funds do attempt to market time in the small stock market and a small number do so successfully. Jiang (2003) fails to report significant differences in timing ability between funds of various investment objectives where all sectors except a specialist technology sector are shown, on average, to mistime the market.

10.4 Market Timing Performance and Operation Location

Table 10.4 presents the nonparametric market timing test statistics of UK equity mutual funds categorised by the fund location. Panel A presents results for the 623 onshore UK funds while Panel B reports results for the 168 offshore funds. Earlier in this study the bootstrap investigation of security selection ability revealed substantial differences between onshore and offshore funds. Informational asymmetry and/or genuine skill differentials were suggested as possible explanations for these differences. This finding motivates a similar inquiry here regarding market timing skill.

From Table 10.4, see also Figure 10.4, it is evident that a small number of both onshore and offshore funds exhibit statistically significant market timing ability at 5% significance by the nonparametric test. Similarly, in results not shown, both subgroups of funds show negative market timing ability on average. In aggregate, differences in selectivity do not appear to transfer to significant differences in market timing skill between onshore and offshore UK equity mutual funds.

10.5 Conditional Market Timing

From chapter 9 on methodology, conditional market timing tests control for the part of fund managers' timing which may be attributed to publicly available information. The conditional market timing literature suggests that this part of performance does not add value to the investor who could employ such information himself (Ferson and Schadt, 1996; Becker et al, 1999). In this section, conditional market timing tests are applied to determine whether the findings from the unconditional tests above of a small number of successful market timers may be attributed to public information or whether it represents genuine (private) skill in timing the market. In chapter 9 the set of public information variables used in this study was described.

Table 10.5 reports the results from a selection of conditional tests. (In results not shown, conditional tests using a number of alternative combinations of the public information variables were applied and results are similar to those presented). Row 1 repeats of the *unconditional* test results in Table 10.1 for ease of comparison. The remaining rows present the nonparametric *conditional* market timing test statistics using the public information variables: Z1 = 1 month UK Tbill rate, Z2 = term spread, Z3 = market dividend yield and Z4 = gilt/equity yield ratio. Note, the conditional test statistics are for the funds as sorted in row 1. The results show that in the extreme right tail of the distribution the skilful market timing performance of the top 3 sorted funds cannot be attributed to public information (there is one slight exception for the highest sorted fund conditioned on Z1, Z2 & Z3 but this is significant at 10% significance). However, as we move further inside the right tail of the distribution, for example for the 10^{th} and 12^{th} highest sorted funds, there is evidence that their statistically significant market timing ability by the *unconditional* test, may in fact be explained by the public information variables where the *conditional* test statistics are generally insignificant at the 5% level.

Overall, the evidence indicates that we cannot reject the hypothesis that a very small number of funds skillfully time the market, even controlling for public information. However, Figure 10.5 shows distributions of the nonparametric tests statistics from various conditional tests as indicated where it is clear in all cases that the distributions of the timing statistics are centred on negative values.

10.6 Conclusion

This study finds that a relatively small number (around 1.5%) of UK equity *mutual funds* possess significant positive market timing skill. A sizeable majority of funds are shown to mistime the market. This finding is robust with respect to the choice of benchmark market returns. A smaller number of funds are found to be significant positive market timers after controlling for publicly available information.

Such predominantly poor market timing performance by managers may have a number of mitigating factors. First, managers may be attempting to time market volatility rather than expected market returns thus lowering the fund's market exposure during a high volatility, but possibly also a high return, period. Second, note that the equity funds in this study are open ended. This implies that when the equity market generally is performing well (rising market) the funds may experience higher investor capital inflows, a relatively high (short term) cash position, lower market beta and lower return. Conversely, a falling market may be associated with higher redemptions, causing the fund to liquidate its cash position leading to a higher market beta[4]. The latter may offer some 'defence' to funds who are not successfully timing the market.

[4] The funds in this study are UK domestic equity unit trusts. As outlined in chapter 3 on data description, this means (by definition) that the funds are invested in *at least* 80% domestic UK equity. The remaining up to 20% is likely to involve a variable cash position in the short term.

Table 10.1: Mutual Fund Market Timing Performance – Unconditional Tests

Table 10.1 presents results for the unconditional market timing tests. Row 1 reports the nonparametric test statistic, $z = \sqrt{n}.\hat{\theta}_n / \hat{\sigma}_{\hat{\theta}_n}$, and the funds are sorted from worst to best based on this statistic. Row 2 reports $\hat{\theta}_n$, the market timing coefficient, for funds in row 1. Row 3 and row 4 show the t-statistics of the TM and HM timing coefficients respectively. Row 5 reports the nonparametric test statistic, z, using the FT100, rather than the FTSE A All Share index, as the market benchmark. Row 6 describes the investment objective of the sorted funds: 1 = equity income fund, 2 = general equity fund, 3 = small stock fund. Row 7 indicates whether the fund is a survivor or non-survivor fund: 1 = surviving fund, 0 = non-surviving fund. Row 8 describes the fund location: 1 = onshore, 0 = offshore fund. Row 9 displays the number of fund observations. Results relate to the period 1988M1:2002M12 and are restricted to funds with a minimum of 12 observations, leaving 791 funds in the analysis.

Unconditional Market Timing Results

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5 max	3.max	2.max	max
Test Stat, z.	-4.927	-3 054	-2.398	-2.071	-1.026	-0.174	0.563	0.956	1.343	1 407	1.549	1 668	1 812	1.952	2 028	2.801	2 861	3.868
$\hat{\theta}_n$	-0.472	-0.093	-0.077	-0.063	-0.030	-0.007	0.020	0.052	0.127	0.133	0.117	0.101	0.116	0.226	0.128	0.152	0.190	0.231
TM	-6.438	-3.512	-2.179	-1.792	-1.811	-0.194	-0.330	3.280	1.957	2.394	1.796	1.032	1.119	5.996	3.128	3.025	2.848	4.318
HM	-6.873	-3 569	-2.367	-1.747	-1.469	-0.580	-0.041	2 676	1.532	2 010	1 919	1.887	1.476	4 322	3.004	2 784	3 088	3 957
z (FT100)	-6.120	-3.509	-2.855	-2.508	-1.570	-0.808	0.044	0.566	0.817	0.929	1.177	1.285	1.300	1.542	1.814	2.563	3.078	3.092
Style	2	2	3	3	2	2	2	2	2	2	2	2	2	2	2	2	1	1
Survival	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Location	1	1	1	1	0	0	1	1	0	1	1	0	1	1	0	1	1	1
No. Obs.	15	180	132	180	147	143	157	105	30	25	41	25	36	17	79	55	44	39

Figure 10.1: Distribution of the Unconditional Market Timing Test Statistic

Figure 10.1 shows a histogram of the unconditional market timing test statistic, z. The figure is based on 791 funds with a minimum of 12 monthly observations.

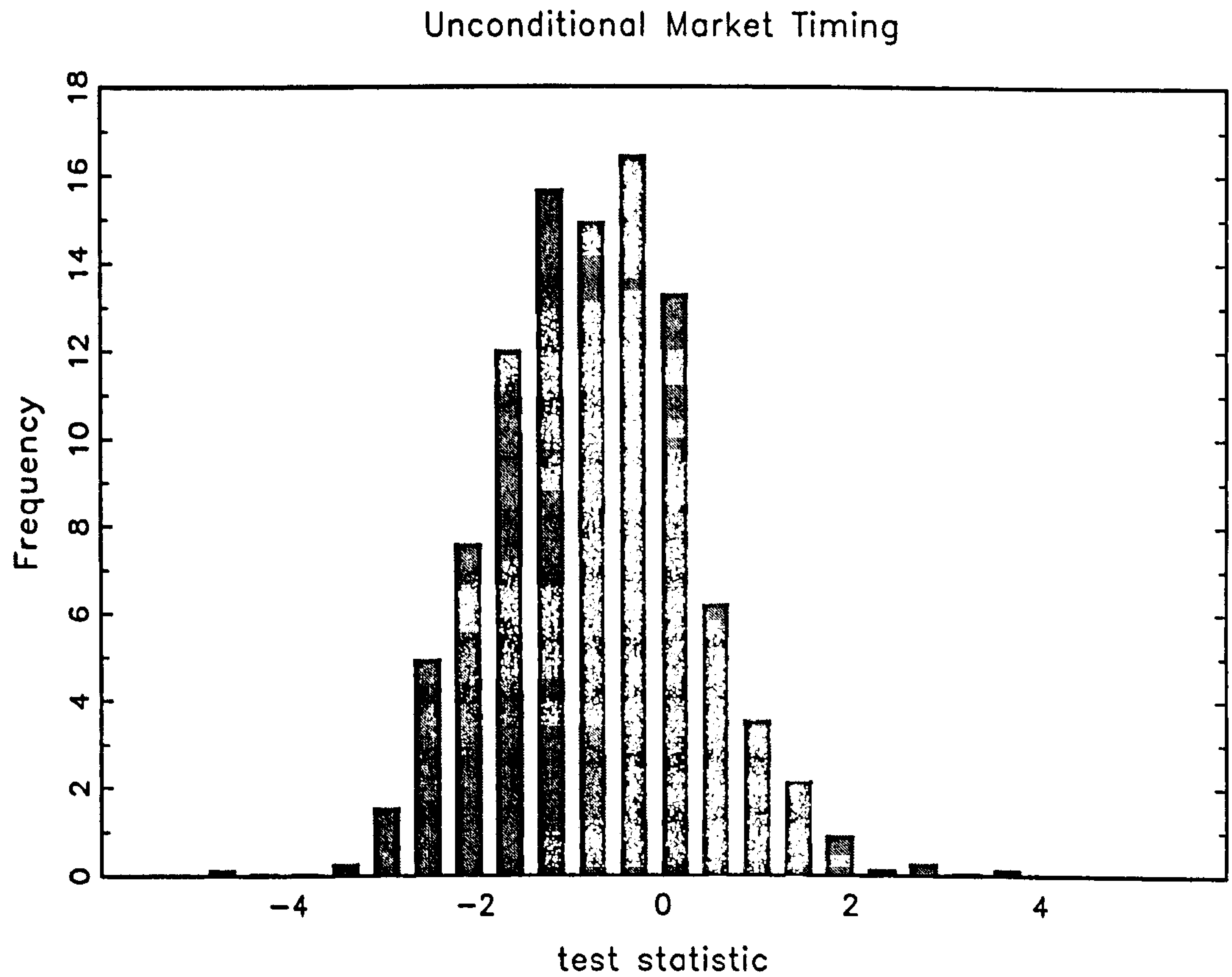


Figure 10.2: FTSE A All Share Index 1975-2003 Total Returns (Index)

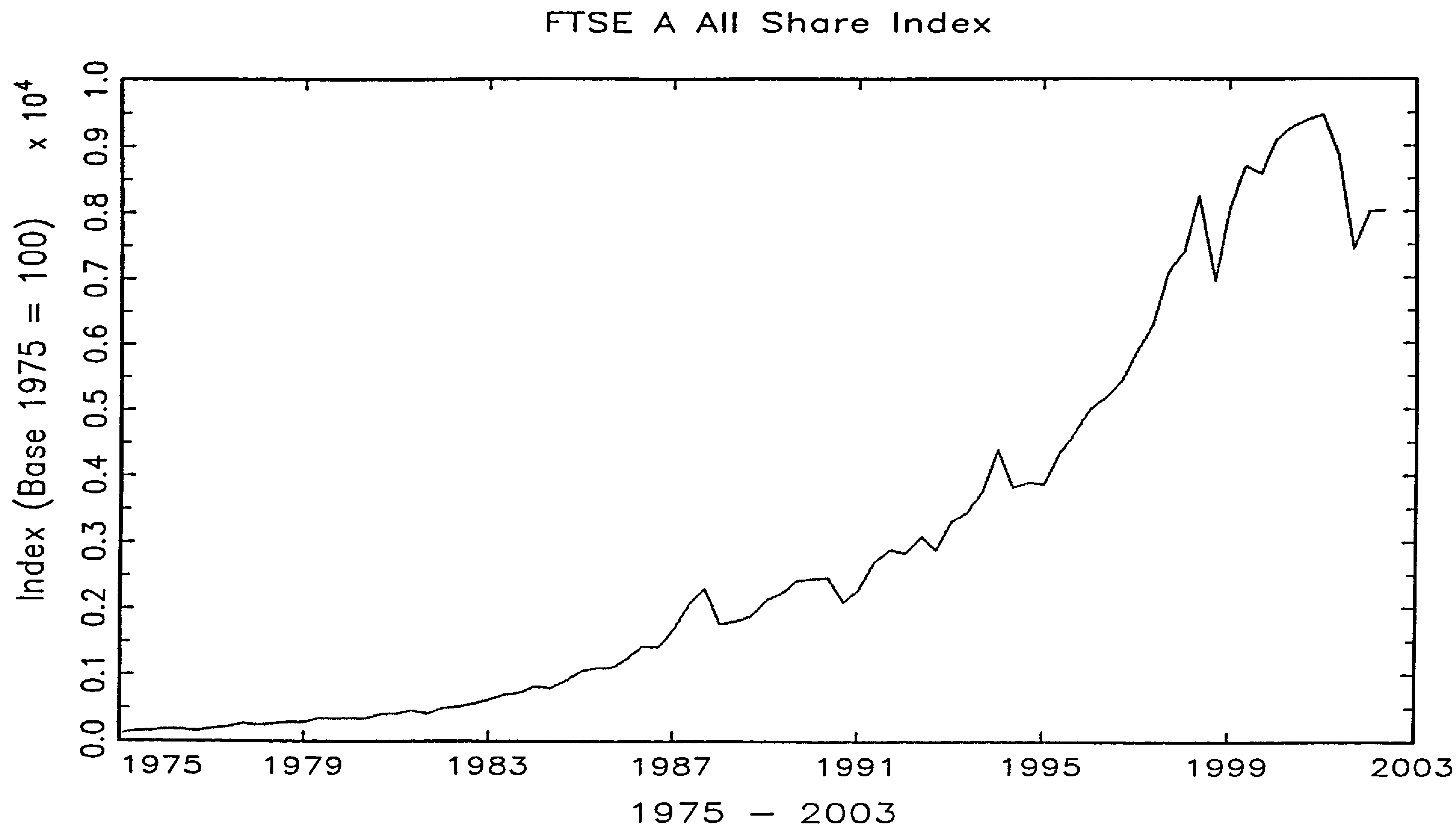


Table 10.2: Market Timing Performance – Alternative Sample Periods

Table 10.2 presents results for the unconditional market timing tests for different sample periods as indicated. Each row shows the nonparametric test statistic, $z = \sqrt{n}.\hat{\theta}_n / \hat{\sigma}_{\hat{\theta}_n}$, and the funds are sorted from worst to best based on this statistic. Results are restricted to funds with a minimum of 12 observations.

Unconditional Market Timing - Alternative Sample Periods

	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5 max	3.max	2 max	max
1988-2002	-4.927	-3.054	-2.398	-2.071	-1.026	-0.174	0.563	0.956	1.343	1 407	1.549	1.668	1.812	1.952	2.028	2.801	2.861	3.868
1975-1989	-2.653	-2.144	-1.774	-1.427	-0.393	0 266	1.045	1.363	1.674	1.299	1.452	1.605	1.674	1.851	1.935	2.390	2.933	3.053

Table 10.3: Mutual Fund Market Timing Performance – By Investment Style

Table 10.3 presents results for the unconditional market timing tests by investment style. In each panel, Row 1 reports the nonparametric test statistic, $z = \sqrt{n}.\hat{\theta}_n / \hat{\sigma}_{\hat{\theta}_n}$, and the funds are sorted from worst to best based on this statistic. Row 2 reports $\hat{\theta}_n$, the market timing coefficient for funds in row 1. Row 3 and row 4 show the t-statistics of the TM and HM timing coefficients respectively. In Panel A, row 5 reports the nonparametric test statistic, z, using the FT100, rather than the FTSE A All Share index, as the market benchmark. In Panel C, row 5 reports the test statistic, z, using the Hoare Govett Small Cap (HGSC) index as the market benchmark. In all panels, rows denoted 'survival' indicate whether the fund is a survivor or non-survivor fund: 1 = surviving fund, 0 = non-surviving fund. Rows denoted 'Location' indicates fund location: 1 = onshore, 0 = offshore fund. The final row in each panel displays the number of fund observations. Results relate to the period 1988M1:2002M12 and are restricted to funds with a minimum of 12 observations, leaving 791 funds in the analysis.

Unconditional Market Timing – By Investment Style

Panel A : Equity Income

	min	5.min	min5%	min10%	min20%	min40%	max30%	max20%	max10%	10max	7max	5.max	3 max	2.max	max
Test Stat, z	-3.137	-2.275	-1.969	-1.838	-1.520	-0.898	-0.202	0.081	0.517	0.762	1.066	1.401	2.179	2.861	3.868
$\hat{\theta}_n$	-0.082	-0.296	-0.062	-0.051	-0.048	-0.032	-0.007	0.002	0.022	0.023	0.062	0.044	0.229	0.190	0.232
TM	-3.445	-2.943	-1.832	-0.820	-3.615	-0.611	-0.383	-0.109	-0.499	-0.373	3.062	0.121	2 084	2.848	4.318
HM	-3.701	-3.821	-1.892	-1.250	-3.144	-0.556	-0.168	0.018	-0.513	0.056	2 669	0.481	2.090	3.088	3.957
z (FT100)	-3.717	-2.855	-2.576	-2.268	-1.907	-1.528	-0.863	-0 613	-0 031	0.346	0.711	0.914	2.563	3 078	3.092
Survival	1	1	1	1	1	0	0	1	1	0	1	0	1	1	1
Location	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1
No. Obs.	180	14	180	180	180	132	132	177	105	132	83	129	32	44	39

Panel B : General Equity

	min	5min	min5%	min10%	min20%	min40%	max30%	max20%	max10%	10max	7max	5max	3max	2max	max
Test Stat, z	-4.927	-2.875	-2.331	-1.925	-1.394	-0.821	-0.055	0.221	0.744	1.617	1.812	1.922	2.019	2.028	2.801
$\hat{\theta}_n$	-0.472	-0.192	-0.144	-0.156	-0.043	-0.026	-0.002	0.007	0.080	0.169	0.116	0.121	0.216	0.128	0.152
TM	-6.438	-3.065	-3.486	-1.727	-2.791	-1.757	-0.232	1.183	1.249	0.395	1.119	3.766	1.742	3.128	3.025
HM	-6.873	-3.195	-2.473	-2.596	-2.281	-1.437	-0.204	1.134	0.904	0.549	1.476	3.049	1.924	3.004	2.784
Survival	1	1	0	1	1	1	0	0	1	1	1	1	1	1	1
Location	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1
No. Obs.	15	39	36	34	178	180	132	132	25	30	36	73	13	79	55

Panel C : Smaller Companies

	min	5 min	min5%	min10%	min20%	min40%	max30%	max20%	max10%	10max	7max	5 max	3.max	2.max	max
Test Stat, z	-3.243	-2.752	-2 658	-2.439	-2.218	-1.806	-1.258	-1.079	-0.513	-0.248	-0.055	0.072	0.573	0.683	0 983
$\hat{\theta}_n$	-0.094	-0.085	-0.081	-0.132	-0.066	-0.053	-0.049	-0.062	-0.042	-0.022	-0.002	0.003	0.116	0.069	0.116
TM	-3 178	-2 005	-2 289	-2.262	-3.530	-1.973	-1.299	-1.018	-0.495	-0 812	-1.062	-0.641	-0.728	0.859	0.389
HM	-3 198	-2 491	-2.343	-1.778	-3.478	-2.130	-1.268	-1.336	-0 621	-0.723	-0.526	-0.372	-0.511	0 425	0.433
z (HGSC)	-1.781	-1.445	-1.195	-0.972	-0.567	0.309	0.510	0 787	1.186	1.232	1.698	1.722	1.966	2.058	2.167
Survival	0	0	1	0	1	1	1	1	1	1	0	1	1	1	1
Location	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1
No. Obs.	132	132	180	54	180	180	115	71	50	46	71	116	15	28	30

Table 10.4: Mutual Fund Market Timing Performance – By Fund Location

Table 10.4 presents results for the unconditional market timing tests by fund location. Row 1 reports the nonparametric test statistic, $z = \sqrt{n}.\hat{\theta}_n / \hat{\sigma}_{\hat{\theta}_n}$, and funds are sorted from worst to best based on this statistic. Row 2 reports $\hat{\theta}_n$, the market timing coefficient for funds in row 1. Row 3 and row 4 show the t-statistics of the TM and HM timing coefficients respectively. Row 5 indicates whether the fund is a survivor or non-survivor fund: 1 = surviving fund, 0 = non-surviving fund. Row 6 describes the investment objective of the sorted funds: 1 = equity income fund, 2 = general equity fund, 3 = small stock fund. Row 7 displays the number of fund observations. Results relate to the period 1988M1:2002M12 and are restricted to funds with a minimum of 12 observations, leaving 791 funds in the analysis.

Unconditional Market Timing – By Investment Location

Panel A : Onshore UK Funds

	min	5.min	min5%	min10%	min20%	min40%	max20%	max10%	20max	15max	10max	5.max	3 max	2.max	max
Test Stat, z	-4.927	-3.054	-2.430	-2.11	-1.693	-1.104	0.072	0.544	1.237	1.407	1.574	1 952	2.801	2.861	3.868
$\hat{\theta}_n$	-0.472	-0.093	-0.072	-0.065	-0.056	-0.032	0.002	0.022	0.066	0.133	0.075	0.226	0.152	0.190	0.232
TM	-6.438	-3.512	-1.596	-2.548	-2.144	-1.415	0.420	0.014	2 691	2.394	1.271	5.996	3 025	2.848	4.318
HM	-6.873	-3.659	-2.021	-2.721	-2.088	-1.396	0.026	0.329	2.574	2.010	1.259	4 322	2.784	3.088	3.957
Survival	2	2	2	3	2	2	2	2	2	2	2	2	2	1	1
Style	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
No. Obs.	15	180	180	180	180	180	180	17	83	25	73	17	55	44	39

Panel B : Offshore Funds

	min	5min	min5%	min10%	min20%	min40%	max20%	max10%	20max	15max	10max	5max	3max	2max	max
Test Stat, z	-2.675	-2.505	-2.332	-1.919	-1.357	-0.754	0.217	0.741	0.525	0.804	1.115	1.473	1 893	2 028	2.179
$\hat{\theta}_n$	-0.127	-0.170	-0.179	-0.064	-0.059	-0.038	0.046	0.417	0.047	0.028	0.514	0.089	0.108	0.128	0.229
TM	-2.392	-2.578	-1.926	-2.143	-1.018	0.412	0.612	-0.326	-0.188	0.982	0.389	2.978	1.876	3.128	2.084
HM	-2.136	-2.434	-2.545	-1.845	-1.308	-0.338	0.047	0.026	-0.106	0.887	0.509	2.833	2.016	3.004	2.090
Survival	2	2	2	2	2	1	1	2	2	2	2	2	2	2	1
Style	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
No. Obs.	64	47	35	180	143	87	15	56	33	156	96	81	46	79	32

Table 10.5: Mutual Fund Market Timing Performance – Conditional Tests

Table 10.5 presents results for the conditional market timing tests. Rows report the nonparametric test statistic, $z = \sqrt{n}.\hat{\theta}_n / \hat{\sigma}_{\hat{\theta}_n}$, and funds are sorted from worst to best based on this statistic. For ease of comparison, row 1 shows the *unconditional* timing results as in Table 10.1. Row 2 to row 6 report the nonparametric test statistics of the *conditional* market timing tests for the funds as sorted in row 1. Public information instruments are: Z1 = 1 month UK Tbill rate, Z2 = term spread, Z3 = market dividend yield and Z4 = gilt/equity yield ratio. Results relate to the period 1988M1:2002M12 and are restricted to funds with a minimum of 12 observations, leaving 791 funds in the analysis.

Conditional Market Timing																		
	min	5.min	min5%	min10%	min40%	max30%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5 max	3 max	2 max	max
Test Stat, z	-4.927	-3.054	-2.398	-2.071	-1.026	-0.174	0.563	0.956	1.343	1.407	1.549	1.668	1.812	1.952	2.028	2.801	2.861	3.868
Z1	-3.833	-2.791	-2.015	-1.528	-0.318	0.121	0.528	1.182	1.323	1.133	1 897	1 616	2.068	1.846	2.448	2.083	2.482	2.353
Z2	-1.742	-2.673	-1.702	-1.503	-0.735	-0.117	0.086	1.012	1.253	1.301	1 874	1.522	1.962	2.325	2.753	2.305	2.426	2.114
Z3	-5.416	-2.103	-2.203	-1.512	-0 288	0.019	0.955	0.422	1.428	1.443	1.578	0.635	1 328	1.970	1.593	2.523	2 567	3.715
Z1,Z2,Z3	-1.646	-1.928	-1.665	-1 256	0.035	0.267	0.445	0.827	1.330	0 200	1 378	1.363	0 934	2.790	1.531	1.968	2.121	1.612
Z4	-3.103	-3.105	-1.860	-2.355	-0.865	-0.278	0.325	0 935	1.461	0.551	1.150	1.069	0 931	1.830	1.916	2 637	2.986	3.602

Figure 10.3: Distribution of the Unconditional Market Timing Test Statistic – By Investment Style

Figure 10.3 shows a histogram of the unconditional market timing test statistics, z, by investment style as indicated. The figures are based on 154 equity income, 514 equity and 122 small stock funds with at least 12 monthly observations.

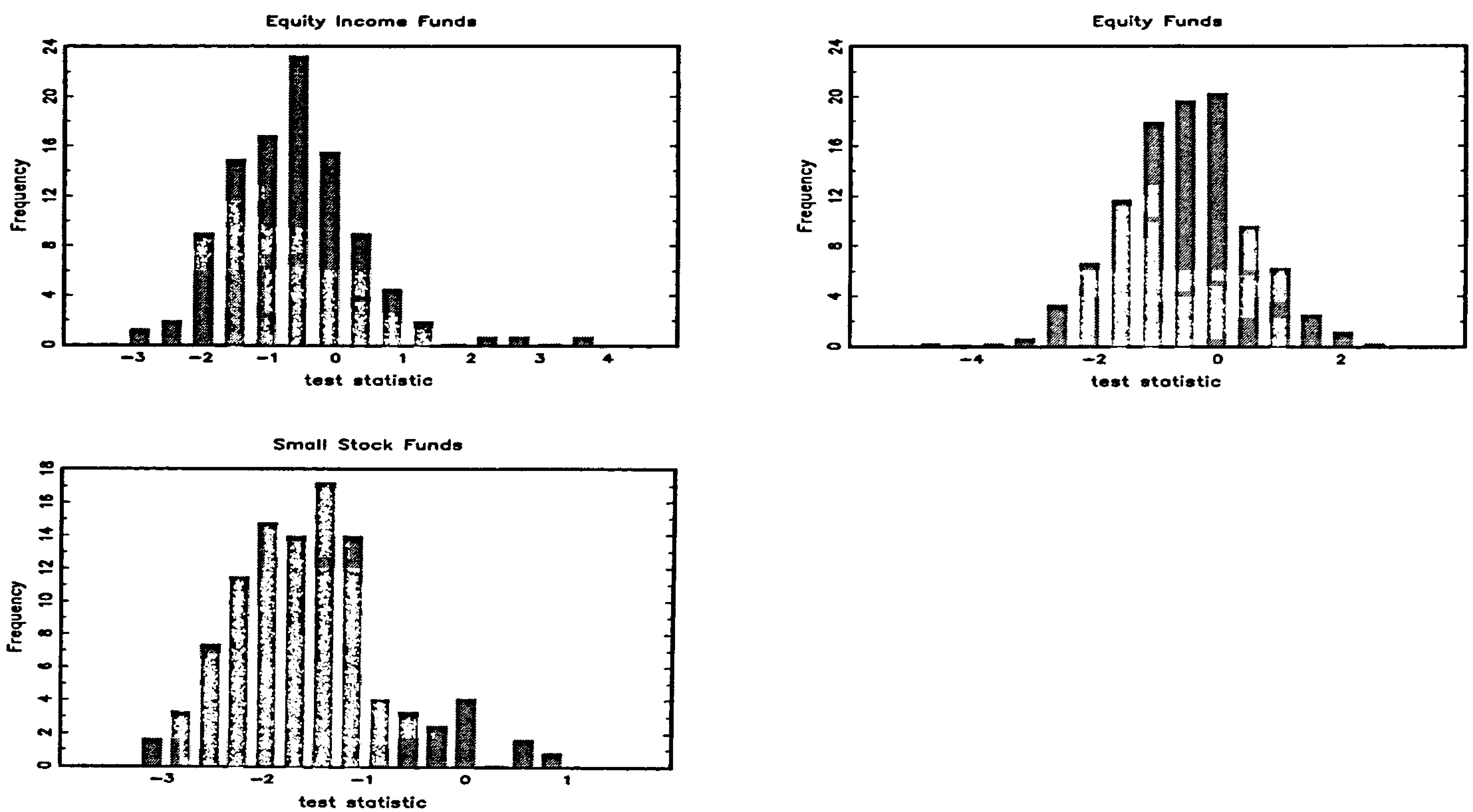


Figure 10.4: Distribution of the Unconditional Market Timing Test Statistic – By Investment Location

Figure 10.4 shows a histogram of the unconditional market timing test statistics, z, by investment location as indicated. The figures are based on 623 onshore funds and 168 offshore funds with at least 12 monthly observations.

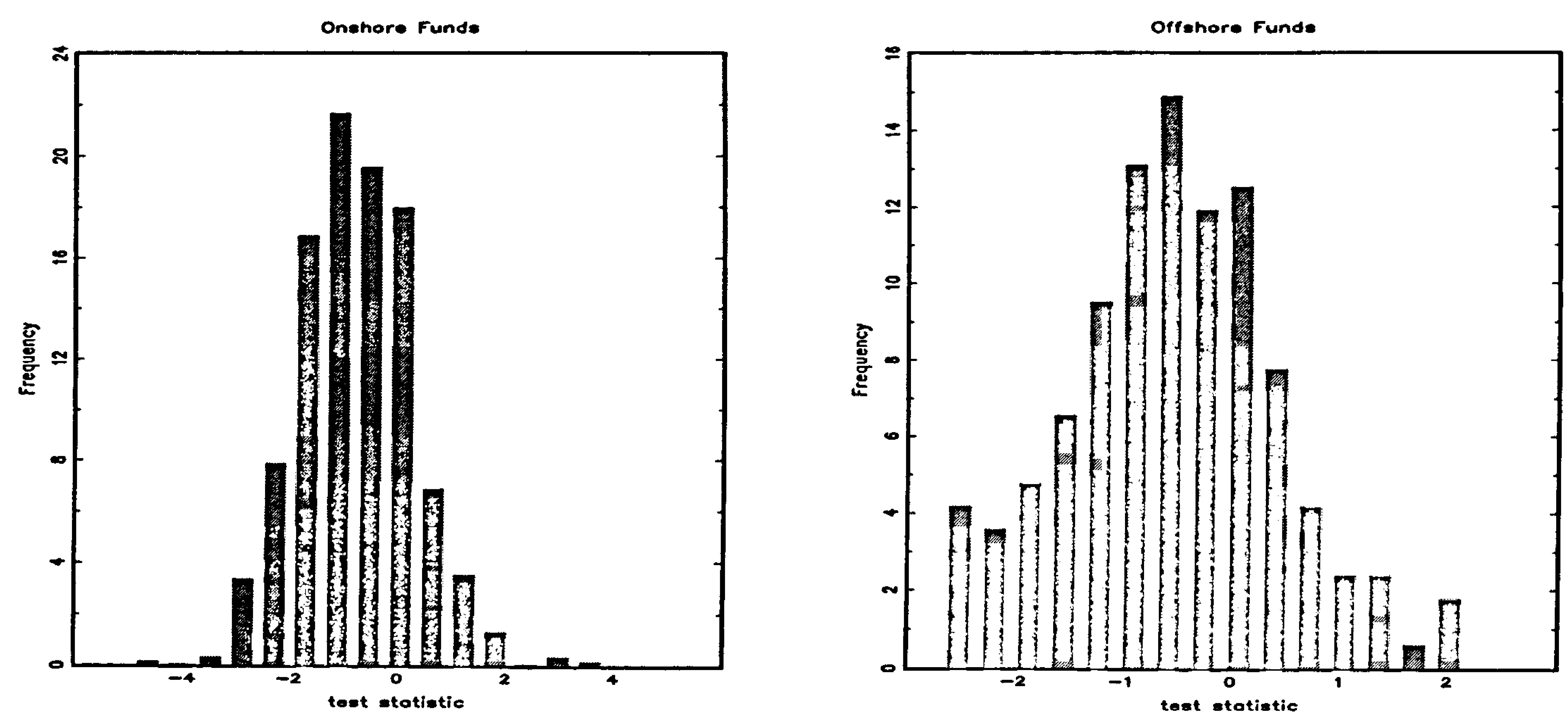
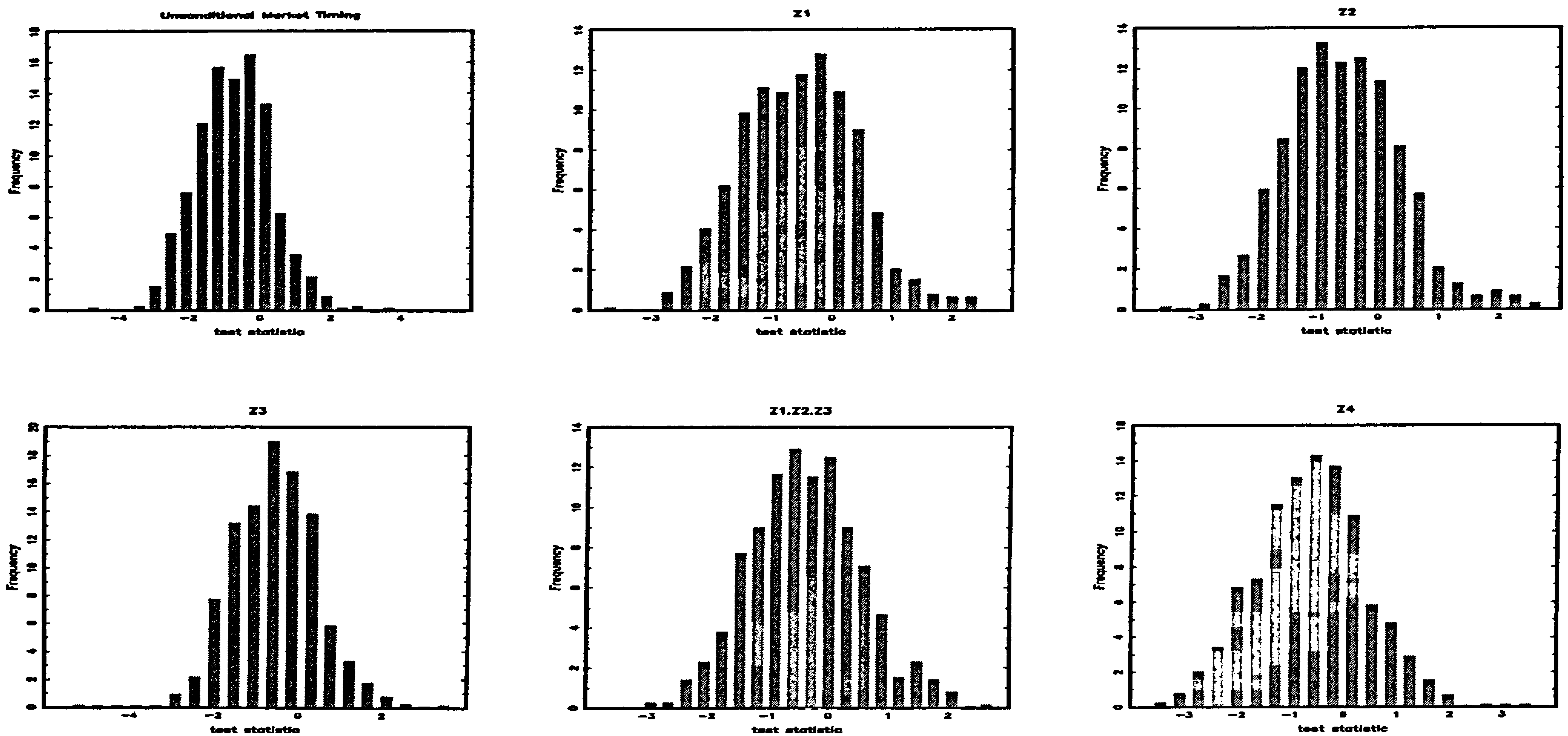


Figure 10.5: Distributions of Conditional Market Timing Test Statistics

Figure 10.5 shows a histogram of the conditional market timing test statistics using public information variables as indicated: Z1= the 1 month UK Tbill rate, Z2 = the term spread, Z3 = the market dividend yield and Z4 = the gilt/equity yield ratio.



CHAPTER 11

CONCLUSION

This study evaluates the performance of the UK equity unit trust industry using a large data base of 1,620 trusts over the period April 1975 – December 2002. Excluding second units of funds as well as market trackers leaves 842 funds, 216 of which are nonsurvivors. Fund abnormal performance is first examined by applying different classes of performance measurement models. These include unconditional single and multi-factor models, conditional-beta models which dynamically adjust risk factor loadings conditional on public information signals, conditional alpha-beta models which allow for conditional abnormal performance and models of market timing.

Using best-fit models from within each class, a bootstrap procedure is applied to the funds. The bootstrap simulation procedure under the null hypothesis of zero abnormal performance provides a means of constructing a distribution of performance which is simply due to random sampling variability (chance or luck) in the performance measure. The technique uses the *empirical* distribution of fund idiosyncratic risk *across all funds* and not just fund by fund. The procedure enables the construction of a nonparametric distribution of performance at *each* point or percentile in the cross-sectional (across funds) distribution of performance including in the extreme tails – the funds of greatest to investors. By comparing the bootstrap distribution against the actual distribution of fund performance it is possible to determine whether high performing funds exceed random sampling variability in the performance measure, or 'luck'. Similarly, it is possible to evaluate whether poor performance is worse than bad luck.

The model selection process indicates that an unconditional three-factor Fama and French type model with market, size and value risk factors fits the data well while the evidence in support of conditional models is very weak. The bootstrap results from the unconditional model find in support of genuine stock picking ability on the part of a small number of top ranking UK equity unit trusts - even when returns are measured net of annual charges imposed by the fund. Of the top 20 ranked funds, 7 (12) funds exceed performance attributable to good luck at 5% (10%) significance. However, these funds are not the extreme top 7 (12) performers but lie slightly inside the extreme right tail of the distribution. In the negative tail of the distribution, the hypothesis that poor performance is attributable to bad luck is strongly rejected at 5% significance. These results are broadly similar to those of a US study by Kosowski et al (2004), although these authors report a greater prevalence of 'star' performers relative to the UK.

Adjustments to the bootstrap simulation procedure are also applied which retain characteristics of the underlying data in order to mimic the 'true' return generating processes of the funds as closely as possible. These include adjustments for serial correlation, heteroscedasticity and cross-sectional (across funds) correlations among fund idiosyncratic risk. Furthermore, alternative bootstrap procedures are applied which resample (i) residuals only, (ii) residuals and factors independently and (iii) residuals and factors in 'time-ordered pairs'. Each procedure examines the variability of the bootstrap performance estimates under the null hypothesis. Conclusions are remarkably robust with respect to these alternative procedures.

In separate applications of the bootstrap procedure the study examines whether security selection skill varies across (i) funds of different investment objectives (income stock funds, general equity funds and small stocks funds) and (ii) funds domiciled onshore versus offshore. These separate subcategories represent a more homogenous risk group of funds and this controls for risk characteristics in funds which may not be adequately captured by the performance model. It is found that some of the top ranked equity income funds show genuine stock picking skills whereas such ability is generally not found among small stock funds and general equity funds. Some highly ranked positive performance among onshore funds is found to be due to genuine skill, whereas for offshore funds, positive performance is attributable to luck.

The study controls for survivorship bias in its results by including nonsurviving funds in the analysis. However, it is found that funds labeled 'nonsurvivors' are well represented among the high performers. This may suggest that many *such funds do not* cease to exist because of poor performance but rather were attractive funds that may have merged or were taken over due to good performance.

As a means of improving statistical precision, baseline results are based on funds with a minimum of 36 observations. While this has many statistical advantages it may introduce some survivorship bias. To examine this the bootstrap procedure is re-applied to the entire sample of funds separately for a number of alternative minimum fund history restrictions ranging from 18 to 120 observations. The inclusion of shorter lived funds slightly increases the prevalence of skill found in fund performance. Similarly, longer lived funds are found to dominate among the poor performing funds which are found to perform worse than bad luck. This result is not consistent with the competitive model. The continued existence over long time periods, of a large number of funds which have a truly inferior performance (which cannot be attributed to bad luck), indicates that many investors either cannot correctly evaluate fund performance or find it 'costly' to switch between funds or suffer from a disposition effect. Hendricks et al (1993) suggest that sustained poor funds are those without skill which "churn" their portfolios too much, incurring high expenses which lowers net performance. Here in this study it is suggested that successful fund managers may be enticed away from their funds to manage other funds thus limiting long run superior performance in any given fund. Alternatively, it may be that shorter-lived funds are initially set up to exploit perceived investment opportunities but 'run out of ideas' in the longer term.

Overall, the results indicate that the majority of funds with positive abnormal performance may attribute this to 'good luck'. It is also shown that 'genuine' top performers are not necessarily those with an *ex-post* ranking right at the top. This makes it extremely difficult for the 'average investor' to pinpoint *individual* active funds which demonstrate genuine skill, based on their track records. The above results suggest that many UK equity investors may be better off holding index/tracker funds, with their lower transactions costs. This question is investigated further in an examination of fund persistence.

The study also examines a second major aspect of mutual fund performance, i.e. persistence. In particular, the study concentrates on the economic as well as statistical significance of persistence using the recursive portfolio formation approach of Carhart (1997) and others and sorts funds by both past raw returns and past four-factor t-alphas. Sorting funds into *deciles* based on either performance metric, the study finds strong evidence that decile portfolios of past loser funds continue to perform badly in terms of their future four-factor alphas and little evidence that past winner funds provide future positive risk adjusted performance. However, unlike past studies this study also investigates smaller portfolios of funds which are likely to be of more practical relevance to investors. Here, it is found that among relatively small 'fund-of-fund' portfolios of past winners, there is evidence that rebalancing every 3 months (or every 1 month), with portfolios of up to around 5 funds does yield some statistically and economically significant four-factor alphas. Therefore, an active portfolio strategy of avoiding past loser funds and choosing a small portfolio of past winner funds may earn positive risk adjusted returns. These returns are net of management fees but gross of the transactions involved in the active strategy. The study then examines the impact of transaction costs involved in such an active strategy by constructing the empirical distribution of final wealth.

Long-term, investors are also interested in the distribution of final wealth (e.g. mean, skewness, kurtosis) from an active strategy, relative to that from alternative strategies such as holding index trackers - taking account of 'luck' across *all* funds and transactions costs of rebalancing. Using a cross-section bootstrap approach the study derives the empirical distribution of final wealth over a 10 year horizon for active strategies, incorporating transactions costs in each rebalancing period and compares this with the outcome from a passive strategy of (randomly) holding an index tracker fund over the whole investment horizon (thus avoiding transactions costs). The analysis suggests that if rebalancing is frequent (e.g. every 3 months) and if transactions costs are less than 2.5% per fund round trip, then an active strategy seems at least as good as a passive strategy but with transactions costs of 5% the passive strategy is most probably superior. Therefore, there is a *prima facie* case for an active strategy in the UK based on *ex-ante* 'sorting rules', for 'small' fund portfolios as long as average switching costs of rebalancing are below 2.5%.

The study also examines a third important aspect of UK equity unit trust performance, i.e. market timing. This study implements a nonparametric timing test and is the first to apply this test to the UK market. The nonparametric procedure has several advantages over the widely used regression based tests of Terynor-Mazuy (1966) and Henriksson-Merton (1981). The nonparametric test is used to examine both unconditional and conditional market timing where the latter considers whether mutual fund managers possess *private* market timing information (timing signals) in excess of the information quality contained in publicly available information. Market timing skill is assessed against a number of alternative benchmark market indices.

This study finds that a relatively small number (around 1%) of UK equity mutual funds possess significant positive market timing skill at 5% significance. A sizeable majority of funds are shown to mistime the market. This finding is robust with respect to the choice of benchmark market returns against which funds are evaluated and with respect to whether timing performance is measured unconditionally or conditionally upon public information.

Apparent mistiming of the market by funds may not be irrational, however. It may be that managers are attempting to time market volatility rather than market returns thus lowering the fund's market exposure during high volatility. This appears as negative market timing if returns and volatility are positively correlated, (Breen et al 1989, Glosten et al 1993 and Busse 1999). Furthermore, in the funds' 'defense', as the equity funds are open ended a rising market is likely to be associated with higher capital inflows to the fund, a relatively high cash position in the short term, lower market beta and lower return. Conversely, a falling market may lead to higher redemptions, causing the fund to liquidate its cash position leading to a higher market beta.

On the policy side the UK government wishes to encourage long term saving via mutual (and pension) funds (Turner 2004). The type of performance findings in this study are central to such policy discussions. The evidence suggests that large scale investment (saving) by the public in the majority of actively managed equity funds, in pursuit of abnormal performance, may well represent a misallocation of resources relative to passive investment strategies.

REFERENCES

Abrevaya, J. and Jiang, W. (2001). Pairwise slope statistics for testing curvature, working paper, University of Chicago Graduate School of Business and Columbia Business School.

Admati, A., Bhattacharya, S., Ross, S. and Pfleiderer, P. (1986). On timing and selectivity, *Journal of Finance*, 41, pp 715-730.

Admati, A. and Pfleiderer, P. (1997). Does it all add up? Benchmarks and the compensation of active portfolio managers, *Journal of Business*, 70, pp 323-350.

Allen, D. and Tan, M. (1999). A test of the persistence in the performance of UK managed funds, *Journal of Business Finance and Accounting*, 26(5) & (6), pp 559-593.

Ball, R. and Kothari, S.P. (1989). Nonstationary expected returns: implications for tests of market efficiency and serial correlations in returns, *Journal of Financial Economics*, 25, pp 51-74.

Barberis, N., 2000. Investing for the Long Run When Returns are Predictable, *Journal of Finance* 55, pp 225-264.

Barberis, N., Shleifer, A. and Vishny, R. (1998). A model of investor sentiment, *Journal of Financial Economics*, 49, pp 307-343.

Barberis, N. and Thaler, R. (2003). A survey of behavioural finance, in Handbooks of the economics and finance, edited by G.M. Constantinidis, M. Harris and R. Stulz, Elsevier Science B.V.

Becker, C., Ferson, W., Myers, D. and Schill, M. (1999). Conditional market timing with benchmark investors, *Journal of Financial Economics*, 52, pp 119-148.

Benartzi, S. and Thaler, R.H., 2001. Naïve Diversification Strategies in Defined Contribution Savings Plans, *American Economic Review* 91, pp 79-88.

Berk, J. and Green, R. (2004). Mutual fund flows and performance in rational markets, *Journal of Political Economy*, 112(6), pp 1269-1295.

Bickel, P.J. and Freedman D.A. (1981). Some asymptotics on the bootstrap, *Annals of Statistics*, 9, pp 1196 – 1271.

Bildersee, J.S. (1975). The association between a market determined measure of risk and other measures of risk, *Accounting Review*, 50, pp 81-98.

Black, A., Fraser, P. and Power D. (1992). UK unit trust performance 1980-1989: a passive time-varying approach, *Journal of Banking and Finance*, 16, pp 1015-33.

Blake, C. and Morey, M. (2000). Morningstar ratings and mutual fund performance, *Journal of Financial and Quantitative Analysis*, 35(3), pp 451-483.

Blake, D. (2000). *Financial Market Analysis*, second edition, Wiley and Sons.

Blake, D., Lehman, B. and Timmermann, A. (1999). Asset allocation dynamics and pension fund performance, *Journal of Business*, 72, pp 429-461.

Blake, D. and Timmermann, A. (1998). Mutual fund performance: evidence from the UK, *European Finance Review*, 2, pp 57-77.

Blake, D. and Timmermann, A., 2003. Performance Persistence in Mutual Funds: An Independent Assessment of the Studies Prepared by Charles River Associates for the Investment Management Association, *Financial Services Authority (also available at www.pensions-institue.org/reports)*

Bodie, Zvi, R.C. Merton, and W. Samuelson. 1992. Labor Supply Flexibility and Portfolio Choice in a Lifecycle Model. *Journal of Economic Dynamics and Control*, vol. 16, nos. 3/4 (July–October), pp 427–449.

Bollen, N. and Busse, J. (2001). On the timing ability of mutual fund managers, *Journal of Finance*, 56, pp 1075-1094.

Bollen, N. and Busse, J. (2005). Short term persistence in mutual fund performance, *Review of Financial Studies*, 18(2), pp 569-597.

Breen, W., Jagannathan, R. and Ofer, A. (1986). Correcting for heteroscedasticity in tests for market timing ability, *Journal of Business*, 59, pp 585-598.

Brennan, M.J., Schwartz, E.S. and Lagnado, R., 1997. Strategic Asset Allocation, *Journal of Economic Dynamics and Control* 21, pp 1377-1403.

Bridgeland, S. (2000). SimIAn® - Example research results, in *ICBI Fund Forum Conference Proceedings, New Perspectives on Risk*, 6 July, Bacon and Woodrow, London.

Bridgeland, S. (2001). Process attribution – a new way to measure skill in portfolio construction, *Journal of Asset Management*, 2 (3), pp 247-259.

Brown, G., Draper, P. and Mckenzie, E. (1997). Consistency of UK pension fund investment performance, *Journal of Business Finance and Accounting*, 24, (2), pp 155-178.

Brown, K., Harlow, W.V. and Starks, L.T. (1996). Of tournaments and temptations: an analysis of managerial incentives in the mutual fund industry, *Journal of Finance*, 51, pp 85-110.

Brown, S. and Goetzmann, W.N. (1995). Performance persistence, *Journal of Finance*, 50, pp 679-698.

Brown, S., Goetzmann, W., Ibbotson, R. and Ross, S. (1992). Survivorship bias in performance studies, *Review of Financial Studies*, 5, pp 553-580.

Burns, P. (2004). Performance measurement via random portfolios, working paper, Burns Statistics. Available at http:// www.burns-stat.com

Busse, J. (1999). Volatility timing in mutual funds: evidence from daily returns, *Review of Financial Studies*, 12, pp 1009-1041.

Campbell, J.Y., Chan, Y.L. and Viceira, L.M., 2003, A Multivariate Model of Strategic Asset Allocation, *Journal of Financial Economics* 67, pp 41-80.

Campbell, J.Y. and Viceira, L.M. 1999, Consumption and Portfolio Decisions when Expected Returns are Time Varying, *Quarterly Journal of Economics* 114, pp 433-496.

Carhart, M. (1995). Survivor bias and mutual fund performance, working paper, School of Business Administration, University of Southern California, Los Angeles, California.

Carhart, M. (1997). On persistence in mutual fund performance, *Journal of Finance*, 52, pp 57-82.

Carhart, M., Carpenter, J., Lynch, A. and Musto, D. (2002). Mutual fund survivorship, *The Review of Financial Studies*, winter, vol. 15, No. 5, pp 1439-1463.

Carpenter, J. (2000). Does option compensation increase managerial risk appetite, *Journal of Finance*, 55, pp 2311-2331.

Carpenter, J. and Lynch, A. (1999). Survivorship bias and attrition effects in measures of performance persistence, *Journal of Financial Economics*, vol. 54, pp 337-374.

Chalmers, J., Edelen, R. and Kadlec, G. (1999). Transaction cost expenditures and the relative performance of mutual funds, working paper, University of Oregon.

Chan, L.K.C. (1988). On the contrarian investment strategy, *Journal of Business*, 61, pp 147-164.

Chan, L.K.C., Karceski, J. and Lakonishok, J. (2000). New paradigm or same old hype in equity investing?, *Financial Analysts Journal*, 56, July/August, pp 23-36.

Chan, L.K.C. and Lakonishok, J. (2004). Value and growth investing: review and update, *Financial Analysts Journal*, 60, January/February, pp 71-86.

Chang, E.C. and Lewellen, W.G. (1984). Market timing and mutual fund investment performance, *Journal of Business*, 57, pp 57-72.

Charles River Associates, 2002. Performance Persistence in UK Equity Funds – An Empirical Analysis, prepared by T. Giles, T. Wilsdon and T. Worboys, October.

Chen, C.R. and Stockum, S. (1985). Selectivity, market timing and random behaviour of mutual funds: a generalized model, *Journal of Financial Research*, pp 87-96.

Chen, L., Jegadeesh, N. and Lakonishok, J. (1996). Momentum strategies, *Journal of Finance*, 51, pp 1681-1713.

Chen, L., Jegadeesh, N. and Wermers, R. (2000). The value of active mutual fund management: an examination of the stock holdings and trades of fund managers, *Journal of Financial and Quantitative Analysis*, 35, pp 343-368.

Chevalier, J. and Ellison, G. (1997). Risk taking by mutual funds as a response to incentives, *Journal of Political Economy*, 105, pp 1167-1200.

Chevalier, J. and Ellison, G. (1999). Are some mutual fund managers better than others? Cross sectional patterns in behaviour and performance, *Journal of Finance*, Vol. 54, No. 3, pp 875-899.

Christopherson, J., Ferson, E. and Glassman, D. (1998). Conditioning manager alphas on economic information: another look at the persistence of performance, *Review of Financial Studies*, Vol. 11, No. 1, pp 111-142.

Clare, A.D., Thomas, S.H. and Wickens, M.R. (1994). Is the Gilt-Equity yield ratio useful for predicting UK stock returns, *The Economic Journal*, 104, pp 303-315.

Coggin, D., Fabozzi, F. and Rahman, S. (1993). The investment performance of US equity pension fund managers, *Journal of Finance*, 54, pp 875-899.

Connor, G. and Korajczyk, R. (1986). Performance measurement with the arbitrage pricing theory, *Journal of Financial Economics*, 15, pp 373-394.

Conrad, J. and Kaul, G. (1998). An anatomy of trading strategies, *Review of Financial Studies*, 11, pp 489-519.

Daniel, K., Grinblatt, M., Titman, S. and Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance*, 52, pp 1035-1058.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A. (1998). Investor psychology and security market under- and over-reactions, *Journal of Finance*, 53, pp 1839-1886.

Dawson, R. and Young, R. (2003). Near uniformly distributed, stochastically generated portfolios. In Satchell, S. and Scowcroft, A., editors, *Advances in Portfolio Construction and Implementation*, Butterworth Heinemann.

De Bondt, W.F. and Thaler, R.H. (1989). Anomolies: a mean reverting walk down Wall Street, *Journal of Economic Perspectives*, 3, pp 189-202.

Del Guercio, D. and Tkac, P.A. (2000). The determinants of flow of funds of managed portfolios: mutual funds versus pension funds, working paper, Lundquist College of Business, University of Oregon.

Droms, W. and Walker, D. (1994). Investment performance of international mutual funds, *The Journal of Financial Research*, Vol. 17, No. 1, Spring, pp 1-11.

Edelen, R.M. (1999). Investor flows and the assessed performance of open-end mutual funds, *Journal of Financial Economics*, 53, pp 439-466.

Efron, B. and Tibshirani R.J. (1993). *An introduction to the bootstrap*, Monographs on Statistics and Applied Probability, Chapman and Hall, New York.

Elton, E., Gruber M. and Blake, C. (1996a). Survivorship bias and mutual fund performance, *The Review of Financial Studies*, vol. 9, No. 4, pp 1097-1120.

Elton, E., Gruber M. and Blake, C. (1996b). The persistence of risk adjusted mutual fund performance, *Journal of Business*, 69, (2), pp 133-157.

Elton, E., Gruber, M., Das, S. and Hlavka, M. (1993). Efficiency with costly information: a reinterpretation of evidence from managed portfolios, *Review of Financial Studies*, 6, pp 1-21.

Fama, E. (1991). Efficient capital markets II, *Journal of Finance*, 46, pp 1575-1617.

Fama, E. and French, K. (1992). The cross section of expected stock returns, *Journal of Finance*, 47, pp 427-465.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, 33, pp 3-56.

Fama, E. and MacBeth, J. (1973). Risk, return and equilibrium: empirical tests, *Journal of Political Economy*, 81, pp 607-636.

Ferson, W. and Khang, K. (2001). Conditional performance measurement using portfolio weights: evidence for pension funds, *Journal of Financial Economics*, 65, pp 249-282.

Ferson, W. and Schadt, R. (1996). Measuring fund strategy and performance in changing economic conditions, *Journal of Finance*, 51, pp 425-462.

Ferson, W. and Warther, V. (1996). Evaluating fund performance in a dynamic market, *Financial Analysts Journal*, 52, pp 20-28.

Fletcher, J. (1995). An examination of the selectivity and market timing performance of UK unit trusts, *Journal of Business Finance and Accounting*, 22(1), pp 143-156.

Fletcher, J. (1997). An examination of UK unit trust performance within the arbitrage pricing framework, *Review of Quantitative Finance and Accounting*, 8, pp 91-107.

Fletcher, J. (1999). The evaluation of the performance of UK American unit trusts, *International Review of Economics and Finance*, Vol. 8, No. 4, November, pp 455-466.

Fletcher, J. and Forbes, D. (2002). An exploration of the persistence of UK unit trust performance, *Journal of Empirical Finance*, 9, pp 475-493.

Fortin, R. and Michelson, S.E. (1995). Are load mutual funds worth the price, *Journal of Investing*, 4, (3), pp 89-94.

Foster, G. (1986). *Financial Statement Analysis*, Prentice Hall, Englewood Cliffs.

Friend, I., Blume, M. and Crockett, J. (1970). Mutual funds and other institutional investors, McGraw Hill, New York.

Gibbons, M., Ross, S. and Shanken, J. (1989). Testing the efficiency of a given portfolio. *Econometrica*, 57, pp 1121 – 1152.

Glosten, L., Jagannathan, R. and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 48, pp 1779-1801.

Goetzmann, W. and Ibbotson, R. (1994). Do winners repeat? Patterns in mutual fund performance, *Journal of Portfolio Management*, 20, pp 9-18.

Goetzmann, W., Ingersoll Jr., J., and Ivkovich, Z. (2000). Monthly measurement of daily timers, *Journal of Financial and Quantitative Analysis*, 35, pp 257-290.

Goetzmann, W.N., Ivkovic, Z. and Rouwenhorst, G., 2001. Day Trading International Mutual Funds: Evidence and Policy Solutions, *Journal of Financial and Quantitative Analysis* 36, pp 287-310.

Grinblatt, M. and Titman, S. (1989a). Mutual fund performance: an analysis of quarterly portfolio holdings, *Journal of Business*, 62, pp 393-416.

Grinblatt, M. and Titman, S. (1989b). Portfolio performance evaluation: old issues and new insights, *Review of Financial Studies*, 2, pp 393-421.

Grinblatt, M. and Titman, S. (1992a). The persistence of mutual fund performance, *Journal of Finance*, 42, pp 1977-1984.

Grinblatt, M. and Titman, S. (1992b). Performance measurement without benchmarks: an examination of mutual fund returns, working paper, University of California, Los Angeles.

Grinblatt, M. and Titman, S. (1994). A study of monthly mutual fund returns and performance evaluation techniques, *Journal of Financial and Quantitative Analysis*, 29 (3), pp 419-444.

Grinblatt, M., Titman, S. and Wermers, R. (1995). Momentum investment strategies, portfolio performance and herding: a study of mutual fund behaviour, *American Economic Review*, Vol. 85, No. 5, December, pp 1088-1105.

Gruber, M. (1996). Another puzzle: the growth in actively managed mutual funds, *Journal of Finance*, Vol. 51, No. 3, pp 783-810.

Hall, P. (1986). On the bootstrap and confidence intervals, *Annals of Statistics*, 14, pp1431-1452.

Hall, P. (1992). The Bootstrap and Edgeworth Expansion, Springer Verlag.

Hendricks, D., Patel, J. and Zeckhauser, R. (1993). Hot hands in mutual funds: short run persistence of performance, 1974-88, *Journal of Finance*, 48, pp 93-130.

Henriksson, R. (1984). Market timing and mutual fund performance: an empirical investigation, *Journal of Business*, 57, pp 73-96.

Henriksson, R.D. and Merton, S.C. (1981). On market timing and investment performance II: statistical procedures for evaluating forecasting skills, *Journal of Business*, 54, pp 513-33.

Hochman, S. (1983). The beta coefficient: an instrumental variable approach, *Research in Finance*, 4, pp 392-407.

Hong, H. and Stein, J. (1999). A unified theory of under-reaction, momentum trading and over-reaction in asset markets, *Journal of Finance*, 54, pp 2143-2184.

Huberman, G. and Kandel, S. (1985). A size based stock returns model, working paper, University of Chicago.

Ippolito, R. (1989). Efficiency with costly information: a study of mutual fund performance 1965-1984, *The Quarterly Journal of Economics*, February, pp 1-24.

Jagannathan, R. and Korajczyk, R. (1986). Assessing the market timing performance of managed portfolios, *Journal of Business*, 59, pp 217-235.

Jegadeesh, N. (1990). Evidence of predictable behaviour in security returns, *Journal of Finance*, 45, pp 881-898.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency, *Journal of Finance*, Vol. 48, No. 1, pp 65-91.

Jegadeesh, N. and Titman, S. (2001). Profitability of momentum strategies: an evaluation of alternative explanations, *Journal of Finance*, Vol 56, No. 2, pp 699-720.

Jensen, M. (1968). The performance of mutual funds in the period 1945 – 1964, *Journal of Finance*, 23, pp 389-416.

Jiang, W. (2003). A nonparametric test of market timing, *Journal of Empirical Finance*, 10, pp 399-425.

Keswani, A. and Stolin, D. (2005). Mutual fund performance persistence and competition: a cross sectional analysis, (forthcoming), *Journal of Financial Research.*

Kmenta, J. (1986). *Elements of Econometrics*, second edition, MacMillan, New York.

Kon, S. (1983). The market timing performance of mutual fund managers, *Journal of Business*, 56, pp 323-47.

Kosowski, Robert, Allan Timmermann, Hai White, and Russ Wermers, 2004, Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis, *Discussion Paper,* INSEAD, April.

Kothari, S.P. and Warner, J.B. (2001). Evaluating mutual fund performance, *Journal of Finance*, Vol. 56, No. 5, pp 1985-2010.

Kritzman, M. and Page, S. (2002). Asset allocation versus security selection: evidence from global markets, *Journal of Asset Management*, 3(3), pp 203-212.

Kritzman, M. and Page, S. (2003). The hierarchy of investment choice, *Journal of Portfolio Management*, 29(4), pp 11-23.

Laplante, M. (2001). Conditional market timing with heteroscedasticity, working paper, University of Washington,

La Porta, R., Lakonishok, J., Shleifer, A. and Vishny, R. (1997). Good news for value stocks: further evidence on market efficiency, *Journal of Finance*, Vol. 52, No. 2, pp 859-874.

Leger, L. (1997). UK investment trusts: performance, timing and selectivity, *Applied Economics Letters*, 4, pp 207-210.

Lockwood, L.J. and Kadiyala, K.R. (1985). Utilization of market forecasts in portfolio management: an empirical analysis, working paper, University of Texas at Arlington, College of Business Administration.

Lunde, A., Timmermann, A. and Blake, D. (1999). The hazards of mutual fund underperformance: a Cox regression analysis, *Journal of Empirical Finance*, 6, pp 121-152.

Lynch, A. and Musto, D. (2002). How investors interpret past returns, *Journal of Finance*, 58(5), pp 2033-2055.

Malkiel, G. (1995). Returns from investing in equity mutual funds 1971 to 1991, *Journal of Finance*, 50, pp 549-572.

Mamaysky, H., Spiegel, M. and Zhang, H. (2005). Improved forecasting of mutual fund alphas and betas, Yale School of Management, ICF working paper 04-23.

Mandelker, G.N. and Rhee S.G. (1984). The impact of the degrees of operating and financial leverage on systematic risk of common stock, *Journal of Financial and Quantitative Analysis*, 19, pp 45-57.

Morey, M.R. (2003). Should you carry the load? A comprehensive analysis of load and no-load mutual fund out-of-sample performance, *Journal of Banking and Finance*, 27, pp 1245-1271.

Merton, R. (1980). On estimating the expected return on the market: an exploratory investigation, *Journal of Financial Economics*, 8, pp 323-362.

Myners, P., 2001. Institutional Investment in the United Kingdom: A Review, *Report prepared for the Chancellor of the Exchequer*, H.M Treasury, London.

OECD, 2003, Monitoring the Future Social Implication of Today's Pension Policies, *OECD*, Paris, unpublished.

Otten, R. and Bams, D. (2000). European mutual fund performance, working paper series, Limburg Institute of Financial Economics, Maastricht University.

Presidential Commission on Social Security Reform, 2001. Strengthening Social Security and Creating Personal Wealth for All Americans, *Report of the President's Commission*, Washington D.C. (available at www.csss.gov).

Quigley, G. and Sinquefield, R. (1999). The performance of UK equity unit trusts, Report by Dimensional Fund Advisors for Institute for Fiduciary Education.

Remolona, E.M., Kleiman, P. and Gruenstein, D. (1997). Market returns and mutual fund flows, *Economic Policy Review*, Federal Reserve Bank of New York, (July).

Rhodes, M. (2000). Past imperfect? The performance of UK equity managed funds, Occasional Paper Series No. 9, Financial Services Authority.

Roll, R. (1977). A critique of the asset pricing theory's tests: part 1. On the past and potential testability of the theory. *Journal of Financial Economics*, 4, pp 129-176.

Roll, R. (1978). Ambiguity when performance is measured by the security market line, *Journal of Finance*, 33, pp 1051-1069.

Serfling, R. (1980. Approximation theorems of mathematical statistics, Wiley, New York.

Sharpe, W.F. (1966). Mutual fund performance, *Journal of Business*, 39, pp 119-38.

Shukla, R. and Trzcinka, C. (1992). Persistent performance in the mutual fund market: tests with funds and investment advisors, *Review of Quantitative Finance and Accounting*, 4, pp 115-135.

Sirri, E.R. and Tufano, P. (1993). Buying and selling mutual funds: flows, performance, fees and services, working paper, Harvard Business School, Cambridge, MA.

State of the Union Address, 2005. Chamber of the U.S. House of Representatives, The United States Capitol, Washington, D.C. (available at www.whitehouse.gov).

Surz, J. (1998). Cyberclone peer groups, *Journal of Investing*, 7(4), pp 63-68.

Thomas, A. and Tonks, I. (2001). Equity performance of segregated pension funds in the UK, *Journal of Asset Management*, 1 (4), pp 321-343.

Tonks, I. (2004). Performance persistence of pension fund managers, working paper, Centre for Finance and Investment, University of Exeter.

Treynor, J. and Mazuy, K. (1966). Can mutual funds outguess the market?, *Harvard Business Review*, 44, pp 131-136.

Turner, Adair, 2004, Pensions : Challenges and Choices : The First Report of the Pensions Commission, The Pensions Commission, *The Stationary Office*, London.

Viceira, L.M., 2001. Optimal Portfolio Choice for Long-horizon Investors with Nontradable Labor Income, *Journal of Finance* 56, pp 433-470.

Volkman, D. and Wohar, M. (1995). Determinants of persistence in relative performance of mutual funds, The Journal of Financial Research, Vol. 18, No. 4, Winter, pp 415-430.

Warther, V.A., (1995). Aggregate mutual fund flows and security returns, *Journal of Financial Economics*, 39, pp 209-235.

Wermers, R. (2000). Mutual fund performance: an empirical decomposition into stock-picking talent, style, transactions costs and expenses, *Journal of Finance*, 55, pp 1655-1695.

Wermers, R. (2003). Is Money Really 'Smart'? New Evidence on the Relation Between Mutual Fund Flows, Manager Behavior, and Performance Persistence, http://ssrn.com/abstract=414420

WM Company, (1999). Comparison of Active and Passive Management of Unit Trusts, for Virgin Direct Financial Services.

Xia, Yihong, 2001. Learning About Predictability : The Effects of Parameter Uncertainty on Dynamic Asset Allocation, *Journal of Finance* 56, pp 205-246.

Zheng, L. (1999). Is money smart? A study of mutual fund investors' fund selection ability, *Journal of Finance*, Vol. 54, No. 3, pp 901 – 933.

BIBLIOGRAPHY

Asness, C. (1997). The interaction of value and momentum strategies, *Financial Analysts Journal*, March/April, pp 29-36.

Baks, K.P., Metrick, A. and Wachter, J. (2001). Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation, *Journal of Finance*, 56, 1, pp 45-85.

Bauman, W., Conover, C. and Miller, R. (1998). Growth versus value and large-cap versus small-cap stocks in international markets, *Financial Analysis Journal*, March/April, pp 75-88.

Bauman, W. and Miller, R. (1994). Can managed portfolio performance be predicted? *Journal of Portfolio Management*, Summer, pp 31-38.

Benning, C.J. (1997). Prediction skills of real world market timers, *Journal of Portfolio Management*, Winter 1997, pp 55-64.

Berk, J. (1997). Does size really matter?, *Financial Analysts Journal*, September/October, pp 12-18.

Blake, D. and Timmermann, A. (2000). Determinants of international portfolio performance, discussion paper, The Pensions Institute, Birkbeck College, University of London.

Blake, D., Lehmann, B. and Timmermann, A. (2002). Performance clustering and incentives in the UK pension fund industry, *Journal of Asset Management*, 3, 2, pp 173-194.

Brocato, J. and Chandy P.R. (1994). Does market timing really work in the real world, *Journal of Portfolio Management*, Winter 1994, pp 39-44.

Chen, Z. and Knez, P. (1996). Portfolio performance measurement: theory and applications, *Review of Financial Studies*, 9,2, pp 511-55.

Cheng-few, L. and Rahman, S. (1990). Market timing, selectivity and mutual fund performance: an empirical investigation, *Journal of Business*, 63, 2, pp 261-278.

Dennis, P, Perfect, S., Snow, K. and Wiles, K. (1995). The effects of rebalancing on size and book-to-market ratio portfolio returns, *Financial Analysts Journal*, May/June, pp 47-56.

Detzel, L. and Weigand, R. (1998). Explaining persistence in mutual fund performance, *Financial Services Review*, 7, 1, pp 45-55.

Goetzmann, W. and Massa, M. (1999). Daily momentum and contrarian behaviour of index fund investors, working paper, Yale School of Management.

Gorman, L. (2003). Conditional performance, portfolio rebalancing, and momentum of small cap mutual funds, *Review of Financial Economics*, 12, pp 287-300.

Groenewold, N. and Fraser, P. (2001). Tests of asset pricing models: how important is the iid-normal assumption?, *Journal of Empirical Finance*, 8, pp 427-449.

Halpern, P. (1991). Investment management fees and determinants of pricing structure in the industry, *Journal of Portfolio Management*, 17, 2, pp 74-79.

Ippolito, R. (1992). Consumer reaction to measures of poor quality: evidence from the mutual fund industry, *Journal of Law and Economics*, 35, pp 45-70.

Jensen, G., Johnson, R. and Mercer, M. (1997). New evidence on size and price-to-book effects in stock returns, *Financial Analysts Journal,* November/December, pp 35-39.

Jensen, G., Johnson, R. and Mercer, M. (1997). The inconsistency of small firm and value stock premiums, *Journal of Portfolio Management*, Winter, pp 27-36.

Keim, D. (1999). An analysis of mutual fund design: the case of investing in small cap stocks, *Journal of Financial Economics*, 51, pp 173-194.

Khorana, A. (1996). Top management turnover: an empirical investigation of mutual fund managers, *Journal of Financial Economics,* 40, pp 403-427.

Kihn, J. (1996). To load or not to load? A study of the marketing and distribution charges of mutual funds, *Financial Analysts Journal*, 52, 3, pp 28-36.

Larsen, G. and Wozniak, G.D. (1995). Market timing can work in the real world, *Journal of Portfolio Management*, Spring 1995, pp 75-81.

Lehman, B. and Modest, D. (1987). Mutual fund performance evaluation: a comparison of benchmarks and benchmark comparisons, *Journal of Finance*, 42, pp 233-265.

Levis, M. and Liodakis, M. (1999). The profitability of style rotation strategies in the United Kingdom, *Journal of Portfolio Management*, Winter, pp 73-85.

Lo, A. and MacKinlay, C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial Studies*, 3, pp 175-206.

MacKinlay, A.C. and Pastor, L. (2000). Asset pricing models: implications for expected returns and portfolio selection, *The Review of Financial Studies*, 13, 4, pp 883-916.

Odean, T. (1998). Are investors reluctant to realise their losses? *Journal of Finance*, 53, pp 1775-1789.

Odean, T. (1998). Do investors trade too much, working paper, UC Santa Cruz.

Pastor, L. and Stambaugh, R. (2002). Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics*, 63, pp 315-349.

Pastor, L. and Stambaugh, R. (2002). Investing in equity mutual funds, *Journal of Financial Economics*, 63, pp 351-380.

Perold, A.F. and Sharpe, W. (1995). Dynamic strategies for asset allocation, *Financial Analysts Journal*, January/February, 16-27.

Prather, L., Bertin, W. and Henker, T. (2004). Mutual fund characteristics, managerial attributes and fund performance, *Review of Financial Economics*, 13, 4, pp 305-326.

Shefrin, H. and Statman, M. (1985). The disposition to sell winners too early and ride losers too long:theory and evidence, *Journal of Finance*, 40, pp 777-790.

Shefrin, H. and Statman, M. (1995). Making sense of beta, size and book-to-market, *Journal of Portfolio Management*, winter, pp 26-34.

ter Horst, J., Nijman, T. and Verbeek, M. (2001). Estimating look-ahead bias in evaluating persistence in mutual fund performance, *Journal of Empirical Finance,* 8, pp 345-373.

Titman, S. (1989). Adverse risk incentives and the design of performance based contracts, *Management Science*, 35-7, pp 807-822.

Wermers, R. (1999). Mutual fund herding and the impact on stock prices, *Journal of Finance*, 53, pp 581-621.

Woerheide, W. (1982). Investor response to suggested criteria for the selection of mutual funds, *Journal of Financial and Quantitative Analysis*, 17, 1, pp 129-137.