UK Mutual Fund Performance: Genuine Stock-Picking Ability or Luck

Keith Cuthbertson[1]
Tanaka Business School
Imperial College London

Dirk Nitzsche[1]
Tanaka Business School
Imperial College London

Niall O' Sullivan[1,2]
Department of Economics
University College Cork, Ireland
(email:niall.osullivan@ucc.ie)

May 2004

Abstract

We use a bootstrap technique to construct a distribution of abnormal performance among UK equity mutual funds under a null hypothesis of zero abnormal performance. Such a distribution of random sampling variation around no abnormal performance is employed as an estimate of, or proxy for, luck in mutual fund performance. Actual performance is then compared against this luck distribution. Using a number of alternative risk adjustment performance models, we find that a small proportion of funds in the positive tail of a cross-sectional performance distribution produce a level of performance in excess of that which may be explained by good luck. Poor performance is generally found to be worse than bad luck.

Please do not quote or reproduce without author permission.

1. The authors are grateful to Micropal for providing the mutual fund data set used in the analysis.
2. The author wishes to gratefully acknowledge financial support from the Arts Faculty Research Fund at University College Cork, Ireland.

Section 1. Introduction

This study examines the performance of open-end mutual funds investing in UK equity (Unit Trusts and Open Ended Investment Companies (OEICs)) during the period April 1975 to December 2002. A data set of 1,596 funds is examined. This represents almost the entire UK equity mutual fund industry at the end of the sample period. In contrast to the US mutual fund industry, there have been comparatively few studies of the performance of UK unit trusts. Studies of UK unit trusts have, for the most part, examined issues such as overall fund performance relative to a benchmark market index, survivor bias and performance persistence. A discussion of the literature on both the UK and US mutual fund industries is provided in section 2.

This study advances the literature on UK mutual funds by explicitly controlling for random sampling variability in the performance measure using a bootstrap procedure. By constructing a distribution of sampling variability under a null hypothesis of zero abnormal performance one can estimate the distribution of performance which is simply due to random chance or 'luck'. This provides a means of determining whether the performance of funds with the best records is simply due to good luck or whether there is genuine stock picking talent on the part of the manager(s). Likewise, it is possible to evaluate whether the performance of the worst funds lies within the boundaries of random chance.

Many studies of UK equity mutual fund performance[1] rank fund performance and examine whether there is persistence in this performance among the top and bottom funds in subsequent periods throughout the sample period. Performance may be based on raw returns or on a risk adjusted measure which controls for the return premia accruing to the risk characteristics of the stockholdings within the fund. However, while these methods correct for common variation in fund returns, they do not correct for idiosyncratic

[1] A UK equity mutual fund is a fund in which at least 80% of the fund's capital is invested in UK equity, as defined by the Investment Management Association (IMA), formerly the Association of Unit Trusts and Investment Funds (AUTIF). The fund is not necessarily operated from within the UK. Of the 1,596 UK equity funds examined in this study 305 are operated from outside the UK.

variation. This is important because with such a large number of mutual funds in existence one would expect that a number of funds will exhibit strong performance simply due to chance. However, the extant literature on UK fund performance does not explicitly model the role of luck in performance.

The role of luck in mutual fund performance among US equity fund managers was first directly addressed by Kosowski, Timmermann, Wermers and White (2003). Kosowski et al apply a bootstrap technique to establish the sampling variation in the performance measures under a null hypothesis of zero abnormal (risk adjusted) performance and compare the actual distribution of US fund performance against this bootstrapped distribution.

A common difficulty in examining fund performance is that of survivor bias. Excluding funds which have failed to remain in existence throughout the sample period and drawing inferences about overall mutual fund performance based only on surviving funds can induce a potentially serious bias in such findings. This study controls for survivor bias by including 450 nonsurviving funds among the 1,596 funds which are examined.

This study also comprehensively examines UK equity unit trusts by evaluating their performance using a greater number of alternative models of performance measurement than identified in the extant literature. Performance measurement models are extended to include conditional risk factor loadings and conditional abnormal performance as well as conditional market timing models. The momentum effect in stock returns is also examined as a source of cross-sectional variation in unit trust performance. In addition, the sample period under investigation in this study is the longest among similar studies. Examining such a wide range of performance measurement methods over a relatively long sample period reduces the risk that findings could be model or sample period specific.

Fund performance may also be influenced by the investment objective of the fund. In this study funds are classified by their self-declared investment objective. These include growth stock funds, income stock funds, general equity funds (income and growth) or small company stock funds. One cannot be certain that these investment style characteristics of the fund are adequately controlled by standard risk adjustment measures. Therefore, in order to investigate whether stock picking skills vary across funds with different investment objectives, this study also carries out the bootstrap analysis separately among funds with these four different investment styles. Kosowski et al (2003) find that many of the US funds with apparent stock picking ability, or fund "stars", are those with growth oriented investment strategies. This study will identify whether such findings transfer to the UK mutual fund industry. In addition, by examining the stock picking skills of funds which specialize in small company stocks, this study investigates the claim that the market for small company stocks is less efficient and is therefore more easily exploited by small company mutual funds.

This study proceeds as follows: Section 2 describes the literature on performance measurement and persistence in performance among international studies, the vast majority of which are studies of the UK and US mutual fund industry. Section 3 describes the bootstrap methodology used to provide an estimate of luck in performance. Section 4 describes models of performance measurement and applies these models to the sample of UK equity mutual funds in this study from which a number of 'best-fit' models are selected for the bootstrap analysis. Section 5 provides a description of the data set of mutual funds and other variables used to measure performance. In section 6 the findings from the bootstrap analysis are reported while section 7 concludes.

Section 2: Literature Review

Available on Request.

Section 3. Methodology

Many approaches to estimating mutual fund performance rely on estimating hypothesised models of equilibrium security returns in order to measure abnormal (risk-adjusted) performance. In turn, inferences regarding the statistical significance of abnormal performance are often based on standard statistical tests of measures such as alpha (Jensen's alpha, Carhart's alpha etc). There are two central difficulties with these approaches.

First, for their statistical validity these tests require that the alpha performance measure be normally distributed. However, as will be seen in section 4 the residuals from Jensen, Carhart and other equilibrium model regressions are highly non-normal for around 70% of the mutual funds in the sample 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, it is also found that high variance non-normal residuals are far more prevalent in the top and bottom performing funds relative to the middle ranking funds and it is the former group of funds which are of most interest.

Second, with such a large number of UK equity mutual funds in existence, 1,596 in this study, one would expect that some funds will appear to exhibit abnormal performance simply due to chance alone. Therefore, the question arises as to how genuine stock picking ability may be distinguished from simple 'good luck'. Likewise, how may true inferior performance be distinguished from bad luck? Following from Kosowski et al (2003), the bootstrap procedure in this study is an attempt to establish the boundaries of performance (good and bad) that is explicable by chance. Observed performance in excess of this is deemed to be superior/inferior.

Adopting the Kosowski et al (2003) methodology, this study bootstraps the abnormal performance measure, (alpha or the t-statistic of alpha), under a null hypothesis of abnormal performance equal to zero. This allows a sampling distribution of fund performance to be constructed where 'true' abnormal performance is not present among funds. The procedure is to simulate the fund return under the null hypothesis (say 1,000 times) for each fund in the sample. In each simulation the performance model is re-estimated and the cross-section of performance measures are ranked from highest to lowest. Over 1,000 simulations this provides 1,000 best alphas, 1,000 second best alphas etc to 1,000 worst alphas, ie a distribution of performance is constructed under the null hypothesis at *each* point/percentile in the ranked cross-sectional distribution of performance. These bootstrap distributions represent random sampling variability in the performance measure at each point in the performance distribution around a 'true' value of zero, ie they are estimates of random chance or 'luck'.

We then compare the cross-sectional ranked measures of actual fund abnormal performance against the empirical bootstrap distribution of fund alphas under the null hypothesis at each point in the performance distribution. For example, we compare the highest ranked actual fund alpha against the distribution of performance under the null hypothesis at the extreme top end of the performance distribution. Similarly, we compare the second highest ranked fund alpha against the bootstrap distribution of performance at the second highest point in the performance distribution etc. The bootstrap p values indicate the probability of observing the actual observed level of performance simply due to random sampling variability in the performance measure around a true value of zero, ie the probability of observing this actual performance due to random chance or 'luck'. The estimate of luck is based on the (large) sample of peer group funds. Alternatively, for any given level of performance, good or bad, one can identify how may funds in the sample one would expect to achieve this level of performance by chance and compare this to how many funds actually achieve or exceed this performance.

Equally, one can evaluate 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 observed t-statistic of alpha. Using t-statistics has more reliable statistical properties. Funds with fewer observations may be estimated with higher variance and less precision and will in consequence tend to generate outlier alphas. There is a risk therefore 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. The distribution of bootstrapped t-statistics for extreme values of the unmodified return t-statistics is likely to have fewer problems with high variance relative to the bootstrap distribution of alpha at that percentile in the performance distribution.

For this reason in this study, the t-statistic of alpha is employed as the measure of abnormal performance and the bootstrap methodology described above is implemented for the t-statistic of alpha. All t-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors.

The bootstrap procedure has the advantage that it provides a nonparametric approach to statistical inference about performance as it makes no assumptions about the shape of the true distribution of performance measures. As such the bootstrap technique provides an improved picture of the 'empirical' distribution of the performance measures such as alpha.

To further improve the precision of performance estimates one can impose a minimum number of observations requirement for a fund to be included in the analysis. As indicated, an insufficient number of observations in the estimation is likely to increase the sampling variability of the resulting estimates which affects the tails of the actual and bootstrap performance distribution. In this study a minimum of 60 monthly observations is set as the requirement for the inclusion of funds in the analysis. The disadvantage with this approach, however, is that it may impose a certain survivor bias by restricting the examination to funds which have been skilled or lucky enough to survive for five years.

To examine the significance of this issue the sensitivity of the bootstrap results can be tested for a number of alternative minimum observations specifications.

In order to provide a comprehensive study of performance and to test the robustness of results, the bootstrap test is applied to alternative performance measurement models, ie both single and multi-factor models with unconditional and conditional factor loadings and alphas. Performance measurement models commonly applied in the literature are now described in the next section.

Section 4. Performance Measures and Model Selection

4.1 Performance Measurement

An appropriate method of adjusting for risk is required when examining mutual fund performance. The performance measure depends on the asset pricing model chosen to represent the cross section of expected returns. The most common measures that appear in the literature and the measures that will be bootstrapped in this study are presented in this section. First, the theoretical basis for the performance measurement models is discussed. All models are then applied to the data set of UK mutual funds with a view to selecting subsets of appropriate models with which to perform the computationally intensive bootstrap analysis.

4.1.1 Jensen's Alpha Measure

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

$$(R_{it}-R_f) = \alpha_i + \beta_i(R_{mt}-R_f) + \varepsilon_{it} \tag{4.1}$$

where R_{it} is the expected return on fund *i* in period *t*, R_{mt} is the expected return on a market proxy 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 Measure

The Carhart (1997) measure is the alpha estimate from a four-factor model which is an extension of (4.1) and includes additional risk factors for fund exposure to size, book-to-market value and momentum strategies to model expected fund returns:

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

where SMB_t, HML_t and $PR1YR_t$ are risk factor mimicking portfolios for size, book-to-market value and one-year momentum effects respectively in the stock holdings of the mutual funds. Carhart's alpha may be estimated empirically from (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[2]. The economic rational underpinning the specification of a size risk factor is related to relative prospects. The earnings prospects of small firms may

[2] The calculation of SMB and the other risk factors in (4.2) is described in Section 5.

be more sensitive to economic conditions with a resulting higher probability of distress during economic downturns. There is also the concern that small firms embody greater informational asymmetry for investors than large firms. Both these factors imply a risk loading for size and a higher required return by investors.

Fama and French (1992) also report a strong positive relationship between stock returns and the book-to-market value ratio: stocks with high book-to-market ratios have higher average returns than low book-to-market value stocks (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 low earnings on assets relative to low book-to-market firms. 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.

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 between a portfolio of previously high performing stocks and previously 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. Carhart finds that the momentum variable explains almost half of the spread in returns between the top and bottom decile portfolios of funds ranked by raw return. In a sense, the momentum factor is specified due to an *ex-poste* hypothesis that it 'must be' providing a proxy for a risk factor that explains a significant amount of common variation in fund returns. However, Chan, Jegadeesh and Lakonishok (1996) suggest that the momentum anomaly is a market inefficiency caused by slow reaction to information.

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 the scenario where fund managers identify changing market information about the expected returns and risk of individual securities, change the composition of the fund in response and thus possibly alter the risk of the portfolio. 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 (193), 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.

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)).

4.1.4 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

$$\beta_i = b_{0i} + B'_i(z_t) \tag{4.3}$$

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

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

As $E[z_t{*}r_{b,t+1}] = E(z_t).E(r_{b,t+1}) + Cov(z_t;r_{b,t+1})$, the specification in (4.4) captures the covariance between the market timing variables, z_t, and the conditional expected excess return of the benchmark portfolio. As before under the null hypothesis of zero abnormal performance $\alpha_i = 0$. 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 the economic information set Z_t. For *L* instruments in Z_t the conditional four-factor model involves (L+1)4 + 1 regressors. 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.

4.1.5 Conditional Alpha and 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 betas and alphas. 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 matrix 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) $$r_{i,t+1} = \sum_{i}^{N} W_{it+1} * R_{it+1} \text{, for N assets in the fund}$$

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 and 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.[3] Of course, all conditional models may

[3] A more detailed description of the conditioning variables as adopted in this study is provided in section 5

also be examined by applying subsets of the information set, Z_t. The findings from such tests are outlined later in this section.

4.1.6 Models of Market Timing

In addition to stock selection skills, models of portfolio performance should also attempt to identify whether fund managers have the ability to market-time or predict aggregate market movements. This is, can fund managers successfully assess the future direction of the market in aggregate and either increase or decrease the portfolio sensitivity (Beta) accordingly? Treynor and Mazuy (1966) and Merton and Henriksson (1981) are two commonly applied market timing models in the literature while Ferson and Schadt (1996) also estimate conditional versions of both these models.

4.1.7 The Treynor-Mazuy Model

The Treynor and Mazuy (1966) models is a quadratic extension of the single factor CAPM in (4.1). The model assumes that β_i in (4.1) at time *t* may 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) embodies the situation where during a market upswing a successful market timer has a higher than normal fund Beta and the fund performs better than it would otherwise, and *vice-versa*.

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, z_t. Substituting for β_i in (4.1) with this linear function yields a model of the form

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

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.

4.1.8 The Merton-Henriksson Model

Merton and Henriksson (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. A positive (negative) forecast causes the manager to target a higher (lower) fund Beta. This is, from (4.1), β_i is assumed to be 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 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. (Merton and Henriksson (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 Merton-Henriksson model to a conditional setting, suppose β_i in (4.1) is written as

(4.14) $$\beta_i = b_d + \gamma_{ic}D + (B'_d + \Delta'D)*z_t$$

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

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 $\alpha_i = 0$.

The broad range of models above describe the approaches commonly applied in the literature to measure risk adjusted (abnormal) performance among equity portfolio managers. In testing a hypothesis of fund abnormal performance, a researcher faces the joint hypothesis problem of whether the underlying model is in fact the correct model of equilibrium security returns. As a test of robustness, in this study each of the above models, and many variants of same, are estimated for each mutual fund. Results are averaged across funds and compared between models in order to select a number of 'best fit' or most appropriate risk adjusted performance measurement models to apply in the subsequent bootstrap analysis. Rather than select a single model, in this study representative models are selected from among the classes of models above, ie unconditional, conditional beta and conditional alpha and beta. The empirical findings from these models are now described.

4.2 Model Selection

In this section, the equilibrium models of security returns described above are estimated for all the mutual funds. 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, the normality and serial correlation characteristics of the funds, the significance of factor loadings and model selection diagnostics. Based on these statistics a subset of representative models from the classes of (i) unconditional models, (ii) conditional beta models and (iii) conditional alpha and beta models are selected for the bootstrap analysis in section 6. A subset of representative models is selected because the bootstrap methodology is computationally intensive and yields a large volume of results for discussion. These model estimation results are reported Table 4.1 where again, in the interests of parsimony, a selection of findings are presented.

4.2.1 Unconditional Models of Performance

Panel A of Table 4.1 shows the estimation results of the unconditional models including the CAPM (model 1), Fama and French (model 2) and Carhart (model 3) along with the unconditional Treynor-Mazuy (model 4) and Merton-Henriksson (model 5) market timing models. For example, the CAPM indicates that the cross-sectional average alpha was negative at –0.029% per month (-0.35% annually) indicating that the average mutual fund manager underperformed the market by this amount. However, this abnormal performance is not statistically significant at 5%. All t-statistics presented are averages of absolute values as otherwise average t-statistics may centre on zero. (This is particularly the case for the t-statistics of alpha). 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 results where the cross-sectional average alpha in each case is slightly lower at around –0.07% per month compared to the CAPM. A lower value of alpha is to be expected as fund performance is adjusted for additional risk factors in the multi-factor models. In all the unconditional model estimations, around 20% of funds yield a statistically significant value of alpha by a conventional t-test.

In terms of the factor loadings, the t-statistics across all unconditional models are consistent in showing the market risk factor and the size risk factor as statistically significant determinants of the cross-sectional variation in equity returns. For example, in the case of model 3, the Cahart four-factor performance measurement model, the cross-sectional average t-statistic attached to the market risk factor is 27.93 while for the size risk factor the average t-statistic is 5.437. In fact, in results not shown for model 3 in Table 4.1, 100% and 79% of the mutual funds indicated a statistically significant t-statistic on the market and size risk factors respectively. The value risk factor, with an average t-statistic of 1.326 in model 3 does not appear to be a significant influence in explaining equity returns. Only 24% of the sample of mutual funds produced a significant t-statistic on this risk factor. The one-year momentum factor from the Carhart specification (model 3) also appears to be relatively unimportant where only 21% of funds registered the momentum effect as an important determinant of returns.

4.2.2 Conditional Beta Models of Performance

Panel B of Table 4.1 presents the estimation results of the conditional beta models described above. Following Ferson and Schadt (1996) and Kosowski et al (2003), the public economic information variables used to model predictable variation in conditional betas include (i) the yield on a UK one-month Tbill, (ii) the slope of the term structure defined as the yield on the UK 20 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. 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.

The findings in relation to alpha and the distribution of alpha among these conditional beta models are remarkably similar to those in Panel A for the unconditional factor models. According to all conditional beta models the average mutual fund manager underperformed the market except, as previously, when a market timing factor is specified in the Treynor-Mazuy model (model 10). Although, from all models, alpha is not significantly different from zero, on average. 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.

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 across all conditional beta models while the value and momentum factors are not. (The only exception to this is in model 11 where the size risk factor is not significant. However, this is likely to be the result of a collinearity issue as the product of the size factor and the dividend yield is also specified in this model). It is also noteworthy that among this class of conditional beta models, the public economic information instruments employed to control for dynamically modified risk factor loadings are unanimously insignificant for the average fund at the 5% significance level. In results not shown, generally over this class of models the conditioning instruments proved to be statistically insignificant in more than 75% of the sample of mutual of funds.

4.2.3 Conditional Alpha and Beta Models of Performance

Panel C of Table 4.1 describes the estimation results of the conditional alpha and beta models described previously. Briefly, this class of models permits a conditional specification of alpha as well as of beta. Panel C also reveals, once again, that the full set of conditioning public economic information variables are not found to be significant (on average) in any of the conditional alpha and beta models by a t-test at 5%. This is also a robust finding from alternative conditional model specifications among results not shown where the conditional alpha coefficients prove to be insignificant by a t-statistic for more than 90% of the sample of funds.

The unanimous and unambiguous insignificance of the conditioning variables in the conditional beta and conditional alpha and beta models provides strong evidence against conditional models as the 'true' models of equilibrium security returns. Conditional factor model specifications permit dynamically adjusted portfolio sensitivities or generally embody market timing activities on the part of fund managers. 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. This finding is consistent with evidence from market timing tests among UK unit trusts in the literature (Fletcher (1995), Leger (1997)). While parametric tests inherently involve a joint hypothesis, Jiang (2003) also finds against superior market timing activity from nonparametric tests on US equity mutual funds. In contrast, the unconditional models appears to be a superior model of security returns.

Notwithstanding this caveat regarding conditional factor models as the 'true' model of equilibrium security returns, in this study factor models from each of the three classes of models (unconditional, conditional beta and conditional alpha and beta factor models), are selected for the bootstrap analysis. This is done as a means of examining the robustness of findings from the bootstrap procedure across alternative models of equilibrium returns. However, owing to the above caveat, bootstrap results from the conditional models should be treated with caution and perhaps given less weight relative to the better fit unconditional models.

4.2.4 Selecting Representative Models of Performance

Panels A, B and C in Table 4.1 present model selection statistics including the (cross-sectional average) Schwartz Information Criterion (SIC). Panel A indicates that from among the class of unconditional models the three-factor Fama and French specification, with an SIC measure of 1.299, provides the best fit (on average over all mutual funds). Indeed this model provides the best fit from among all classes of models estimated, a finding also reported in Kosowski et al (2003). 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 1.309. Finally, among the conditional alpha and beta models presented in Panel C, model 14 is suggested for selection by the SIC with the lowest measure of 1.331. However, this last model is very similar to model 8 above 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 disparity of specifications model 15 is instead selected from among the conditional alpha and beta models. Model 15, which also has a relatively low SIC value of 1.359, 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 section 6.

4.2.5 Non-normality and Serial Correlation

Also shown in Table 4.1 are statistics describing the percentage of mutual funds within each performance measurement model for which the null hypothesis of normally distributed regression residuals is rejected by a Jarque-Bera test at the 5% significance level. In addition, for each model the percentage of funds which reject the null hypothesis of no serial correlation among the estimated residuals by a Lagrange Multiplier (LM) test for up to 6 lags is also shown. In the case of all performance measurement models the normality assumption is rejected for around 70% 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 statistical tests such as the t-test and F-test. The finding of widespread non-normally distributed fund residuals also questions the reliability of past research which draws inferences from t-tests and F-tests regarding mutual fund abnormal performance. This strongly motivates the need to bootstrap performance estimates to determine whether significant outperformance (and underperformance) exists in the mutual fund industry.

The LM test statistics in each panel of Table 4.1 suggest that in the case of all performance measurement models a sizeable proportion (around 45%) of mutual fund estimations exhibit serial correlation of order one. This has implications for the implementation of the bootstrap methodology. Firstly, it is important to modify 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. In the next section the bootstrap methodology is implemented and results discussed.

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

Table 4.1 presents results from the estimation of the performance models described in Section 4 using all mutual funds. Panel A relates to unconditional models, Panel B relates to conditional Beta models while Panel C relates to conditional Alpha and Beta models. T-statistics are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors. (t-statistics shown 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%. Also shown is the Schwartz Information Criterion, a model selection criterion which trades off goodness of fit against degrees of freedom. This table also shows the alpha and its t-statistic for an equal weighted portfolio of all mutual funds. All figures shown are cross-sectional averages.

Panel A: Unconditional Factor Models							
			1	**2**	**3**	**4**	**5**
Model			**CAPM**	**FF**	**Carhart**	**TM**	**MH**
Regression Coefficients							
Average Alpha (percent per month)			-0.29	-0.069	-0.072	0.086	0.228
t-statistic			1.022	1.251	1.283	1.135	1.212
Standard Deviation of Alpha			0.223	0.219	0.222	0.300	0.418
Unconditional Betas							
(t-statistics in parentheses)	Rmrf		0.889 (24.318)	0.918 (27.485)	0.921 (27.93)	0.878 (24.08)	0.956 (16.76)
	SMB			0.286 (5.366)	0.286 (5.437)		
	HML			-0.008 (1.476)	-0.003 (1.326)		
	PR1YR				0.010 (1.240)		
Market Timing Measures							
Treynor - Mazuy						-0.005 (1.527)	
Merton-Henriksson							-0.149 (1.272)
Model Selection Criteria							
Adjusted R-square			0.743	0.821	0.824	0.747	0.746
SIC			1.521	1.299	1.309	1.534	1.537
Rejection of Normality (% of funds)			71	72	70	71	71
LM(1) statistic (% of Funds)			45	42	42	44	44
LM(6) statistic (% of Funds)			38	38	39	40	38
Equal Weighted Portfolio							
Alpha			0.027	-0.141	-0.146	0.174	0.370
t-statistic			0.380	-3.110	-3.183	2.291	3.499

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)		0.015	-0.042	-0.042	-0.045	0.106	-0.045
t-statistic		1.053	1.210	1.204	1.234	1.147	1.215
Standard Deviation of Alpha		0.251	0.226	0.221	0.224	0.305	0.216
Unconditional Betas							
(t-statistics in parentheses)	Rmrf	1.125 (6.524)	1.067 (7.134)	1.067 (7.584)	1.077 (7.625)	1.062 (5.772)	1.077 (7.649)
	SMB		0.284 (5.389)	0.283 (5.315)	0.284 (5.473)		0.228 (1.708)
	HML		-0.008 (1.467)	-0.007 (1.476)	-0.004 (1.363)		-0.116 (1.027)
	PR1YR				0.009 (1.194)		
Market Timing Measures							
Treynor - Mazuy	$(Rmrf)^2$					-0.005 (1.165)	
Conditioning Variables, Z_{t-1}							
$Z1_{t-1}*Rmrf_t$, Z1: One month rate		0.031 (1.232)	0.027 (1.336)		-0.024 (1.305)	-0.093 (1.143)	
$Z2_{t-1}*Rmrf_t$, Z2: Term Spread		-0.009 (1.157)	-0.078 (1.281)		-0.126 (1.239)	-0.139 (1.163)	
$Z3_{t-1}*Rmrf_t$, Z3: Dividend Yield		-0.074 (1.285)	-0.044 (1.356)	-0.041 (1.420)	-0.038 (1.236)	-0.037 (1.031)	-0.045 (1.451)
$Z3_{t-1}*SMB_t$							0.009 (1.546)
$Z3_{t-1}*HML_t$							0.032 (1.018)
Model Selection Criteria							
Adjusted R-square		0.751	0.828	0.824	0.831	0.753	0.828
SIC		1.568	1.341	1.309	1.353	1.587	1.338
Rejection of Normality (% of funds)		71	71	71	69	70	71
LM(1) statistic (% of Funds)		44	43	41	43	43	44
LM(6) statistic (% of Funds)		40	42	39	44	42	41
Equal Weighted Portfolio							
Alpha		0.076	-0.107	-0.110	-0.116	0.162	-0.110
t-statistic		1.079	-2.417	-2.465	-2.580	2.177	-2.524

Table 4.1 Continued, Panel C: Conditional Alpha and 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)		0.361	0.028	0.161	0.134	-0.175	-0.097	0.028
t-statistic		0.951	0.978	0.957	1.026	0.948	0.935	1.001
Standard Deviation of Alpha		1.506	1.438	1.205	1.247	1.217	0.808	1.422
Unconditional Betas								
(t-statistics in parentheses)	Rmrf	1.131 (6.616)	1.075 (7.237)	1.065 (7.622)	1.075 (7.673)	1.003 (7.451)	0.934 (9.488)	1.084 (7.363)
	SMB		0.283 (5.385)	0.283 (5.335)	0.222 (1.726)	0.286 (5.395)	0.225 (1.914)	0.283 (5.479)
	HML		-0.010 (1.442)	-0.008 (1.459)	-0.109 (1.038)	-0.011 (1.486)	-0.065 (1.068)	-0.006 (1.342)
	PR1YR							0.010 (1.917)
Conditioning Variables, Z_{t-1}								
$Z1_{t-1}$*$Rmrf_t$, Z1: One month rate		-0.016 (1.280)	-0.015 (1.374)			-0.120 (1.360)	-0.017 (1.216)	-0.065 (1.345)
$Z2_{t-1}$*$Rmrf_t$, Z2: Term Spread		-0.096 (1.217)	-0.145 (1.314)			-0.248 (1.411)		-0.192 (1.287)
$Z3_{t-1}$*$Rmrf_t$, Z3: Dividend Yield		-0.066 (1309)	-0.039 (1.374)	-0.040 (1.428)	-0.044 (1.465)			-0.033 (1.247)
$Z3_{t-1}$*SMB_t					0.010 (1.571)			
$Z3_{t-1}$*HML_t					0.030 (1.016)			
$Z1_{t-1}$*SMB_t							0.064 (1.547)	
$Z1_{t-1}$*HML_t							0.125 (1.005)	
$Z1_{t-1}$		1.259 (0.961)	1.167 (0.973)			0.285 (0.968)	0.120 (0.945)	1.158 (0.969)
$Z2_{t-1}$		2.340 (1.110)	1.161 (0.944)			0.304 (0.898)		1.134 (0.932)
$Z3_{t-1}$		-0.317 (1.020)	-0.203 (0.975)	-0.049 (0.965)	-0.042 (1.026)			-0.204 (0.982)
Model Selection Criteria								
Adjusted R-square		0.753	0.830	0.825	0.829	0.826	0.826	0.832
SIC		1.638	1.413	1.331	1.359	1.378	1.375	1.424
Rejection of Normality (% of funds)		72	70	71	70	72	73	68
LM(1) statistic (% of Funds)		45	49	44	48	46	46	50
LM(6) statistic (% of Funds)		42	46	38	42	45	43	47
Equal Weighted Portfolio								
Alpha		0.498	0.262	0.283	0.350	0.191	0.049	0.260
t-statistic		1.633	1.560	1.632	2.026	1.197	0.384	1.555

Section 5. Data Description

Available on Request.

Section 6. Empirical Results from the Bootstrap Analysis

6.1 Bootstrap Analysis of Mutual Funds – All Investment Objectives

In Section 4, performance measurement models were described and a representative model was selected from each of the three classes of unconditional alpha models, conditional beta models and conditional alpha and beta models. The findings from the application of the bootstrap methodology, as described in section 3, to these selected models are presented in Table 6.1 and Table 6.2 below.

Table 6.1 presents bootstrap statistics for the full sample of mutual funds, ie including funds of all investment objectives. Panel A reports findings for the unconditional Fama and French three-factor model, Panel B presents results for the conditional Fama and French model where the time varying market risk factor loading is specified as conditional on the market dividend yield and Panel C relates to the Fama and French conditional alpha and beta performance model where each risk factor loading is assumed to be modified in response to the market dividend yield. 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 where the t-statistic here is employed as a second measure of fund performance. Row 4 ("p-tstat") reports the bootstrap p values of the t-statistic in row 3. As explained in section 3 on methodology, the t-statistic measure has the advantage that it scales the alpha measure by its estimation error and is likely to have superior statistical properties in the extreme tails of the performance distribution. For this reason, throughout section 6 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. For further statistical reliability the analysis is restricted to funds with a minimum of 60 observations, unless as otherwise stated. This leaves 724 funds in this analysis. All t-statistics of alpha discussed in this study, ie both actual (unmodified) t-statistics of alpha and bootstrap t-statistics of alpha are based on Newey-West adjusted heteroscedasticity and autocorrelation adjusted standard errors. All bootstrap results reported throughout this section are based on 1,000 simulations.

In Table 6.1, Panel A reveals that the best fund ranked by alpha from the unconditional model achieved abnormal performance of 0.745% per month. This fund alpha has a t-statistic of 2.546. However, the highest ranked fund by the t-statistic of alpha has a t-statistic of 4.023. The bootstrap p value (of the t-statistic) equal to 0.056 indicates that from among the 1,000 bootstrap simulations across each and all of the funds under the null hypothesis of zero abnormal performance, 5.6% of the highest bootstrap t-statistics were greater than 4.023. Operating strictly at 95% confidence, the p value of 0.056 fails to reject the hypothesis that the performance of the best fund (as ranked by the t-statistic of alpha) is within the boundaries of performance that may be explained by random chance or luck at that point in the performance distribution. The p value of 0.056 fails to reject the hypothesis that the top ranked fund does not possess genuine stock picking ability. However, at 90% confidence (or even 94% confidence in this case) the hypothesis that the performance of the top ranked fund is merely due to luck is rejected, ie the fund possess genuine stock picking talent.

Looking across the entire right tail of the performance distribution, the evidence regarding outperformance is mixed. The 2nd, 3rd, 5th and 7th ranked funds do not exceed performance which could be explained by random sampling variability in the t-statistic measure at 95% confidence at each of these points in the performance distribution. However, there is strong evidence to indicate that lower ranked funds (10th, 12th, 15th and 20th) are sufficiently skilled in selecting stocks to cover their costs (annual charges imposed) and produce genuine abnormal performance for their investors. As one moves closer towards the centre of the performance distribution there is no evidence in support of stock picking ability among funds at each these points in the distribution.

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 –7.414. The bootstrap p value of the t-statistic of alpha of 0.000 at this point in the performance distribution strongly rejects the hypothesis that this fund's performance may be explained by bad luck alone. This fund has produced 'truly' inferior performance. The bootstrap p value 0.000 means that from among the 1,000 bootstrap simulations across each of the funds under the null hypothesis of zero abnormal performance, none of lowest bootstrap t-statistics were lower than –7.414.

It is clear from the left tail of the distribution in Panel A that all observed (unmodified) performance levels at these selected points in the distribution are worse than may be explained by sampling variation in the performance measure around zero actual abnormal performance, ie this poor performance is worse than may be attributable to bad luck.

As an alternative interpretation, the bootstrap procedure may be used to estimate how many funds from the sample one might expect to achieve a given level of alpha performance by random chance alone. This number can then be compared to the number of funds which actually achieve this level of alpha. For example, based on the bootstrap estimates from the unconditional model one would expect 6 funds to achieve an alpha estimate of 0.5% per month or higher based on random chance alone. In fact, 13 funds in the sample exhibit this level of performance (or higher). However, alphas of 0.1% or higher are expected to be achieved by 171 funds solely based on chance while in fact only 133 funds are observed to have reached this level of performance. This interpretation is, of course, consistent with the discussion of the p values above at selected points in performance distribution: there is greater evidence of genuine outperformance at higher ranked points in the performance distribution relative to lower ranked points.

Figure 6.1 offers an alternative insight into mutual fund performance relative to luck. The Figure plots Kernel density estimates of the distributions of both the actual (unmodified) t-statistics of alpha and the bootstrap distribution of the t-statistics of alpha. Both distributions are estimated from the unconditional model for all investment objective funds. 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 as this distribution is constructed under the null hypothesis of no abnormal performance. Therefore this bootstrap distribution provides a picture of the range in performance that may be expected simply due to chance or luck. Comparison between this distribution and the cross-sectional distribution of the actual t-statistics of alpha 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. The exception to this is in the extreme right tail, ie the top end of the performance distribution. This again indicates that there are a number of high ranking funds which achieve performance which is superior to that explicable by chance alone. However, comparing the left tails of the actual and bootstrap distributions of the t-statistics, poor performing funds cannot attribute performance to bad luck.

Figure 6.2a and Figure 6.3a show the bootstrap histogram of alpha at selected points of the performance distribution. The upper left panel of Figure 6.2a shows the histogram of the best alpha across funds from 1,000 bootstrap resamples under the null hypothesis of no outperformance while the upper right panel shows the 1,000 fifth best alphas and so on. It is quite evident from the four panels of Figure 6.2a that the best bootstrap alphas are highly non-normal and have a relatively high variance but that the histogram more closely approximates normality and exhibits a lesser variance as we move even slightly closer to the centre of the performance distribution. As can be seen from Figure 6.2b, which presents histograms of fund regression residuals at various points in the performance distribution, this finding follows closely from the fact that the residuals from the best unmodified fund regressions exhibit higher variance and a greater degree of non-normality than the residuals of funds closer to the centre of the performance distribution. It is this high variance 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 alpha distributions and generate a wide dispersion among these top alphas 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 1,000 worst alphas across funds generated from the bootstrap procedure under the null hypothesis of zero abnormal performance, the upper right panel shows the histogram of the fifth worst 1,000 alphas across funds etc. Once again it is evident that the histograms of performance at the worst point in the performance distribution exhibits a higher variance and a greater non-normality than the histograms of performance 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 from the worst fund alpha regressions exhibit higher variance and greater non-normality than the residuals of funds closer to the centre of the performance distribution. Again, it is this high variance among the worst funds' regression residuals (and the existence of large negative residuals) that causes these funds to populate the lower end of the bootstrap alpha distributions and generate a wide dispersion among these alphas in the bootstrap procedure.

This non-normality and high variance among the residuals and alpha estimates of the top and bottom fund regressions motivates the use of the bootstrap procedure to more correctly identify the distribution of performance at the extreme ends of the performance spectrum and therefore to more accurately draw inferences regarding the statistical significance of individual fund performance. Identifying the funds at the upper and lower end of the performance spectrum is of greater interest to the investor and researcher. The above investigation of the distributions of the residuals and of the bootstrapped alphas clearly demonstrates the non-normal nature of performance at these points in the performance distribution. This in turn highlights the potential for error in drawing inferences regarding top and bottom fund performance based on statistical tests which rely on the normality assumption.

The central limit theorem indicates that the distribution of actual and bootstrapped fund alphas is approximated by normality. However, the bootstrap procedure significantly improves upon this approximation. (See Bickel and Freedman (1981) and Hall (1986)). Furthermore, this improvement is particularly important at the extreme ends of the performance distribution where, as was seen above, deviations from normality are most acute. In addition, the extreme ends of the performance distribution are the areas of greater interest to investors and researchers.

In Panel B, 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 quite similar to that found for the unconditional model in Panel A. In the positive performance tail of the distribution, p values of the t-statistic of alpha less than 0.05 again point to genuine stock picking ability on the part of fund managers at these points in the performance distribution (adjusted for, or net of, annual charges imposed by the fund). The conditional beta model suggests the existence of genuine stock picking ability among many (although not all) funds ranked at the 95th percentile or higher.

In the left tail, poor performance is again found to be worse than that which may be explained by random sampling variation in the t-statistic performance measure around a true value of zero by construction. The hypothesis that poor performance is attributable to bad luck therefore is again strongly rejected at 95% confidence.

Panel C of Table 6.1 presents bootstrap findings for the conditional alpha and beta performance model. Consistent with the previous two classes of model, the conditional alpha and beta model points to true superior performance in the upper end of the performance distribution. Indeed this class of model points more strongly to more widespread genuine outperformance than either the unconditional model or the conditional beta model previously. The conditional alpha and beta model in Panel C suggests that funds ranked at the 60th percentile and higher perform better than could be accounted for by mere good luck.

However, in the left tail of the performance distribution in Panel C, the evidence regarding whether funds truly underperform bad luck is mixed and generally differs from the results of the previous two classes of model. With the exception of some selected points in the performance distribution around the 5th worst, 5th percentile and 10th percentile, the results from this class of conditional alpha and beta model generally indicate that one cannot reject the hypothesis that poor performance is within the boundaries of random chance or bad luck.

In order to investigate these contrasting results further among the p values of the t-statistic of alpha and to test the robustness of the bootstrap findings, bootstrap estimates of the t-statistics of alpha under a null hypothesis of no abnormal performance were constructed for several alternative performance measurement models within each of the three classes of model. The selected results presented here prove to be remarkably robust to selecting alternative models within each class.

The bootstrap methodology in this study has been applied here to the three classes of model as test of robustness in the findings. In summary, the unconditional model and the conditional beta model are generally found to yield similar conclusions regarding genuine stock picking ability among UK equity mutual funds. However, the model estimation results and bootstrap findings are sensitive to the specification of a conditional alpha measure. However, while such a specification is of economic significance it is not statistically significant. Recall from section 4 (Table 4.1, Panel C) that the set of public economic information variables applied to condition alpha were all shown to be statistically insignificant (on average across funds) in all alternative conditional alpha and beta model specifications shown (and in many further specifications not shown). Indeed, as also reported in section 4, within any conditional alpha and beta model examined, the set of variables used to condition alpha were statistically insignificant for more than 90% of the sample of mutual funds. This evidence from almost the entire population of UK equity mutual funds over a long sample time period would appear to be quite compelling evidence that the conditional factor models simply do not represent the 'true' model of equilibrium security returns. (At least this is the case given the chosen set of public information variables selected in this study to condition alpha and beta. This was the same set of conditioning public information variables adopted by Kosowski et al (2003), Ferson and Schadt (1996) and Christopherson, Ferson and Glassman (1998)).

Accordingly, this study finds that the unconditional three-factor Fama and French model is a more robust and reliable model of equilibrium security returns. Consequently, the estimation results and bootstrap p values above from the conditional models should be interpreted with caution. Therefore, in order to be more confident in the findings from the bootstrap analyses to follow, the unconditional three-factor Fama and French model is selected as the 'benchmark' model upon which conclusions are based.

6.2 Performance and Investment Styles

It is of also interest to investors to identify whether stock picking talent is related to the investment objective of the fund. From the mutual fund performance and persistence literature, particularly among studies of the US fund industry, there is some evidence that mutual funds with a growth stock investment style tend to be among the top performing funds (Chen, Jegadeesh Wermers (2000)). In this study, when the performance analysis is restricted to funds with a minimum of 60 observations, for improved statistical reliability, there are 724 funds remaining. These consist of 192 income funds (27%), 82 growth funds (11%), 326 general equity (income and growth) funds (45%) and 124 small stock funds (17%). The top 10 performing funds, ranked by the t-statistic of alpha from the three-factor unconditional model, are comprised of 4 income stock funds, 3 growth stock funds, 2 general stock funds and 1 small stock fund. The corresponding breakdown of the top 50 funds is 23, 7, 13 and 7 respectively while the breakdown of the top 40% of the performance distribution (top 290 funds) is: 110, 28, 113 and 39 respectively. At the opposite end of the performance scale, the worst 10 funds consist of 3 income stock funds, 1 growth stock fund, 5 general equity funds and 1 small stock fund. The corresponding breakdown of the bottom 50 funds is 5, 7, 29 and 9 respectively while the composition of the bottom 40% of the performance distribution is: 41, 38, 152 and 59 respectively. In this simple analysis income stock funds and growth funds perform relatively well while general equity funds and small stock funds compare poorly. For example, income stock funds comprise 27% of the total sample of funds but comprise 38% of the top 40% of the performance distribution and 46% of the top 50 ranked funds. However, income stock funds make up only 14% of the bottom 40% of the distribution and only 10% of the bottom 50 funds. In general, in relative terms income stock funds disproportionately occupy the upper end of the performance distribution and are disproportionately absent from the bottom end. Although less pronounced, growth stock funds similarly outperform. In contrast, general equity funds comprise 45% of the total sample of funds but make up only 39% of the top 40% of the performance distribution and only 26% of the top 50 funds while this class of funds comprises 53% of the bottom 40% of funds and 58% of the bottom 50 funds. Here, general equity funds appear to disproportionately occupy the bottom end of the performance distribution and are disproportionately absent from the top end. Similarly, although again slightly less pronounced, small stock funds also underperform in relative terms.

To further address the question of relative performance of mutual funds of different investment styles, this study implements the bootstrap procedure separately within each subgroup of investment styles or objectives. These investment styles are declared by the funds themselves but certified initially and subsequently monitored monthly by the Investment Management Association in the UK. Examining the performance and skills of managers separately within each fund classification has the added advantage that in each case one is examining a more homogenous risk group. This helps to control for possible unknown cross-sectional risk characteristics which may be unspecified by the equilibrium model of returns.

In Table 6.2, Panels A, B, C and D present the ranked alpha performance measures, its associated t-statistic, the ranked t-statistics of alpha and the bootstrapped p values of the t-statistics of alpha for the four mutual fund investment objectives: equity income, equity growth, general equity (income and growth) and small company stocks respectively. Following the discussion in Section 6.1, these bootstrap findings are estimated by the unconditional three-factor model. Looking at all four Panels in Table 6.2 it is clear that performance is not evenly divided between investment sectors. Based on the bootstrap p values of the t-statistics, it is evident from Panel A and Panel B that many high ranking equity income funds and equity growth funds respectively, again as ranked by the t-statistic of alpha, achieve levels of performance which cannot be accounted for by random sampling variation in the t-statistic where its true (assigned) value of zero. In particular, within the equity income investment style funds ranked at the selected points and percentiles from the 5th highest to the 90th percentile beat the bootstrap estimate of luck (estimated from among the peer group of funds) at 95% confidence. The 3rd highest ranked fund beats luck at 90% confidence. However, the 1st and 2nd highest ranked funds achieve abnormal returns which are not beyond the boundaries of that which may be explained by sampling variation in the performance measure at their respective points in the performance distribution. Similarly, among equity growth funds, in the right tail of the performance distribution p values of 0.05 or less indicate genuine stock picking ability at 95% confidence. While one cannot reject the hypothesis that the performance at the extreme top end could be attributed to good luck, the performance of slightly lower ranked funds cannot be so attributed at these points in the performance distribution.

In contrast, the separate bootstrap applications to the investment classes of general equity funds and small stock funds in Table 6.2 Panel C and Panel D respectively indicates comparatively poor performance. Among general equity funds, only the extreme highest ranked fund beats luck while all lower ranked funds are well within the boundaries of performance that may be accounted for simply by chance, based on the peer group of funds. Among small stock funds, there is no evidence of genuine outperformance or stock picking ability in the right tail of the performance distribution. The evidence in support of stock picking ability among income stock funds and growth stock funds but the apparent absence of such talent among general equity funds and small stock funds is consistent with the conclusions from the more simple analysis provided above.

In the case of small stock funds, in order to examine the possible sensitivity of performance findings to the choice of the benchmark risk factor for size, in this study the bootstrap procedure was repeated for an alternative size benchmark. In the Fama and French three-factor unconditional model the size risk factor, SMB_t, (ie returns on small cap stock minus returns on large cap stocks - see sections 4,5) was substituted by a small cap index. However, the conclusions discussed above were unaltered.

From a different perspective, in the case of income stock funds, for arbitrarily selected level of performance of say 0.1% per month, the bootstrap procedure indicates that 47 funds would be expected to achieve or exceed this level of performance simply by chance. In fact 52 funds achieve or exceed this performance. In contrast among general equity funds, the same ratio is 73:47.

The finding of outperformance among growth stock funds is consistent with similar findings from among US studies. Indeed using a similar bootstrap methodology Kosowski et al (2003) also report that growth stock funds perform well in relative terms. However, the evidence in this study against the existence of stock picking ability among small stock mutual funds is clearly at variance with the school of thought that the market for small stocks is less efficient. If this hypothesis is correct it has not been exploited by UK small stock fund managers.

From Table 6.2 Panels A, B, C and D, the performance 'picture' in the left tail of the distribution is similar across all four investment classes of funds. Among all four investment styles, in particular at the lower end of the performance distributions, the bootstrap p values of the t-statistic of alpha of less than 0.05 indicate that one can reject the hypothesis (at 95% confidence) that these funds were merely unlucky in their performance. This is consistent with the results of the procedure when all funds were examined together.

From Table 6.2 a caveat should be noted when comparing the relative stock picking abilities in excess of good luck of funds of different investment objectives. By estimating the bootstrap p values separately within each investment class, the bootstrap distribution of the t-statistics of alpha under the null hypothesis of no abnormal performance, ie the estimate of luck, is based only on the peer group (same investment style) of funds in each case. The estimate of, or proxy for, luck is not necessarily the same for each investment style and 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.

In conclusion, by the three factor unconditional model, which is found to be the best fit and most reliable model of performance measurement, we find evidence in support of genuine stock picking ability, controlling for sampling variability, among a number of top ranking UK equity mutual funds. Furthermore, our findings suggest is more prevalent among mutual funds with an income stock investment style and a growth stock investment style but is absent from the subclass of funds with an investment objective of investing in (i) small stocks and (ii) general equity, ie income and growth stocks together. In contrast, we find that funds ranking at the low end of the cross-sectional performance distribution cannot attribute their poor performance merely to bad luck.

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 section 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 and beta model. The first row in each panel reports alpha in percent per month at various points and percentiles in the performance distribution ranging from worst fund (min) to best fund (max). The second row reports the associated t-statistic of these alpha measures. Row 3 contains the t-statistics of alpha ranked from lowest (min) to highest (max). Row 4 reports the bootstrap p values of these t-statistics, (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 in this study are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors.

Panel A: Unconditional Model

	min	5.min	min5%	min10%	min40%	max40%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5.max	3.max	2.max	max
Alpha	-0.901	-0.663	-0.427	-0.330	-0.108	-0.027	0.181	0.301	0.411	0.439	0.478	0.507	0.530	0.588	0.598	0.686	0.723	0.745
t-alpha	-2.532	-4.445	-1.563	-2.406	-1.591	-0.235	1.757	1.920	3.389	0.806	2.777	2.776	1.698	4.023	2.556	2.991	1.322	2.546
t-stat	-7.414	-4.724	-3.077	-2.537	-0.958	-0.216	1.242	1.670	2.023	2.196	2.403	2.522	2.544	2.671	2.776	2.991	3.389	4.023
p-tstat	0.000	0.000	0.000	0.000	0.000	1.000	0.824	0.508	0.169	0.016	0.004	0.006	0.022	0.066	0.132	0.159	0.059	0.056

Panel B: Conditional Beta Model

	min	5.min	min5%	min10%	min40%	max40%	max10%	max5%	max3%	20max	15max	12max	10max	7max	5.max	3.max	2.max	max
Alpha	-0.662	-0.614	-0.376	-0.315	-0.095	-0.004	0.195	0.336	0.444	0.486	0.557	0.580	0.602	0.639	0.731	0.804	0.812	1.123
t-alpha	-2.566	-6.884	-3.353	-2.985	-0.969	-0.044	1.454	1.234	2.265	2.289	0.767	3.144	1.855	2.799	1.072	2.776	3.701	1.803
t-stat	-6.884	-4.149	-3.001	-2.371	-0.798	-0.027	1.362	1.837	2.161	2.289	2.341	2.556	2.650	2.776	2.799	3.701	4.088	4.368
p-tstat	0.000	0.000	0.000	0.000	0.000	1.000	0.193	0.022	0.009	0.003	0.028	0.003	0.002	0.019	0.111	0.001	0.002	0.022

Table 6.1 continued.

Panel C: Conditional Alpha and Beta Model

	min	5.min	min5%	min10%	min20%	min40%	max40%	max10%	max5%	max3%	20max	12max	10max	7max	5.max	3.max	2.max	max
Alpha	-4.871	-3.708	-1.609	-1.125	-0.586	-0.106	0.292	1.389	1.964	2.638	2.781	3.353	3.573	4.200	4.379	5.402	7.028	11.273
t-alpha	-2.070	-3.214	-2.210	-1.846	-0.648	-0.269	0.470	3.527	3.466	2.629	1.900	2.750	1.463	1.927	2.217	1.616	1.977	4.583
t-stat	-4.053	-3.091	-1.995	-1.509	-0.868	-0.197	0.466	1.816	2.217	2.750	2.768	3.181	3.371	3.463	3.527	3.858	4.583	4.816
p-tstat	0.129	0.038	0.009	0.036	0.745	0.927	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.014

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. Panel A reports results for the class of funds investing in income stocks. Panel B reports results for growth funds. Panel C reports results for general equity (income and growth) funds. Panel D reports results for small stock funds. 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 ranging from worst fund (min) to best fund (max). The second row reports the associated t-statistic of these alpha measures. Row 3 contains the t-statistics of alpha ranked from lowest (min) to highest (max). Row 4 reports the bootstrap p values of these t-statistics. p values are based on 1,000 bootstrap resamples. Both actual (unmodified) and bootstrap t-statistics in this study are based on Newey-West heteroscedasticity and autocorrelation adjusted standard errors.

Panel A: Equity Income

	min	5.min	min5%	min10%	min20%	min40%	max40%	max20%	max10%	max5%	max7%	max3%	5.max	3.max	2.max	max
Alpha	-0.666	-0.315	-0.227	-0.164	-0.106	-0.0131	0.027	0.121	0.208	0.284	0.240	0.365	0.382	0.431	0.459	0.478
t-alpha	-7.414	-3.148	-2.313	-1.178	-0.873	-0.240	0.202	1.225	1.623	2.673	2.522	2.196	2.403	2.544	1.726	2.777
t-stat	-7.414	-3.148	-2.469	-1.648	-0.981	-0.318	0.281	0.954	1.623	1.927	1.757	2.403	2.522	2.673	2.777	3.389
p-tstat	0.000	0.000	0.000	0.008	0.148	0.289	0.413	0.154	0.005	0.046	0.045	0.003	0.004	0.058	0.122	0.106

Panel B: Equity Growth

	min	2.min	3.min	5.min	min7%	min10%	min40%	max40%	max10%	max7%	5.max	3.max	2.max	max
Alpha	-0.393	-0.384	-0.345	-0.305	-0.299	-0.247	-0.141	-0.057	0.161	0.301	0.377	0.543	0.598	0.745
t-alpha	-3.562	-1.865	-3.318	-3.146	-2.859	-4.121	-1.511	-0.243	1.466	1.920	2.529	2.671	2.556	2.546
t-stat	-4.121	-3.687	-3.562	-3.146	-2.859	-2.236	-1.129	-0.427	1.463	1.920	2.282	2.546	2.556	2.671
p-tstat	0.011	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.175	0.042	0.004	0.014	0.102	0.333

Table 6.2 continued

Panel C: General Equity

	min	2.min	3.min	5.min	min3%	min5%	min10%	min40%	max40%	max10%	max5%	max3%	5.max	3.max	2.max	max
Alpha	-0.740	-0.673	-0.661	-0.650	-0.490	-0.413	-0.324	-0.130	-0.044	0.135	0.280	0.439	0.514	0.593	0.606	0.723
t-alpha	-3.043	-2.050	-3.199	-2.970	-2.574	-3.751	-1.937	-0.951	-0.624	0.889	1.529	0.806	2.127	1.450	2.292	1.322
t-stat	-5.166	-4.777	-4.190	-4.118	-3.625	-3.160	-2.706	-1.190	-0.418	1.053	1.427	1.627	2.127	2.500	2.776	4.023
p-tstat	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.997	0.995	0.985	0.746	0.452	0.325	0.025

Panel D: Smaller Companies

	min	2.min	3.min	5.min	min5%	min10%	min40%	max40%	max10%	max5%	5.max	3.max	2.max	max
Alpha	-0.901	-0.663	-0.589	-0.528	-0.520	-0.469	-0.250	-0.105	0.215	0.317	0.466	0.530	0.546	0.686
t-alpha	-2.532	-4.445	-2.728	-2.234	-3.077	-2.680	-1.135	-0.489	1.176	1.610	1.788	1.698	1.578	2.991
t-stat	-4.445	-3.306	-3.116	-3.095	-2.915	-2.739	-1.506	-0.470	0.792	1.570	1.610	1.788	2.409	2.991
p-tstat	0.003	0.012	0.002	0.000	0.000	0.000	0.000	1.000	1.000	0.654	0.850	0.877	0.338	0.225

Figure 6.1: Kernel Density Estimates of the Actual and Bootstrap distribution.

Figure 6.1 presents the Kernel density estimates of the actual and bootstrap distributions of the t-statistics of alpha from the unconditional three-factor performance measurement model over the full sample of mutual funds. t-statistics are Newey-West adjusted. Funds with a minimum of 60 observations are used. The plots are generated using a Gaussian Kernel function.

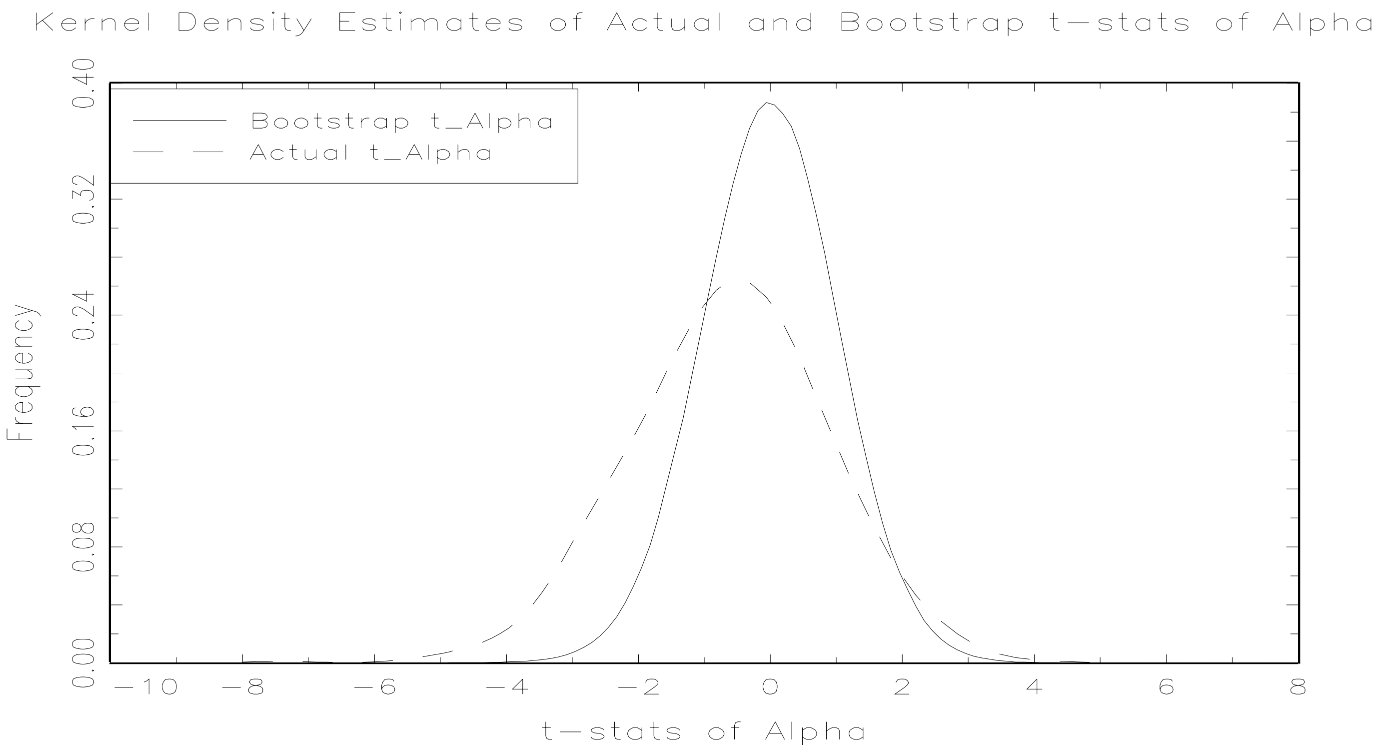


Figure 6.2a: Histograms of Bootstrap Alpha Estimates (Upper End of the Distribution)

Figure 6.2a presents histograms of the bootstrap alpha estimates at various points in the upper end of the performance distribution. These are estimated by the unconditional three-factor performance measurement model over the full sample of mutual funds.

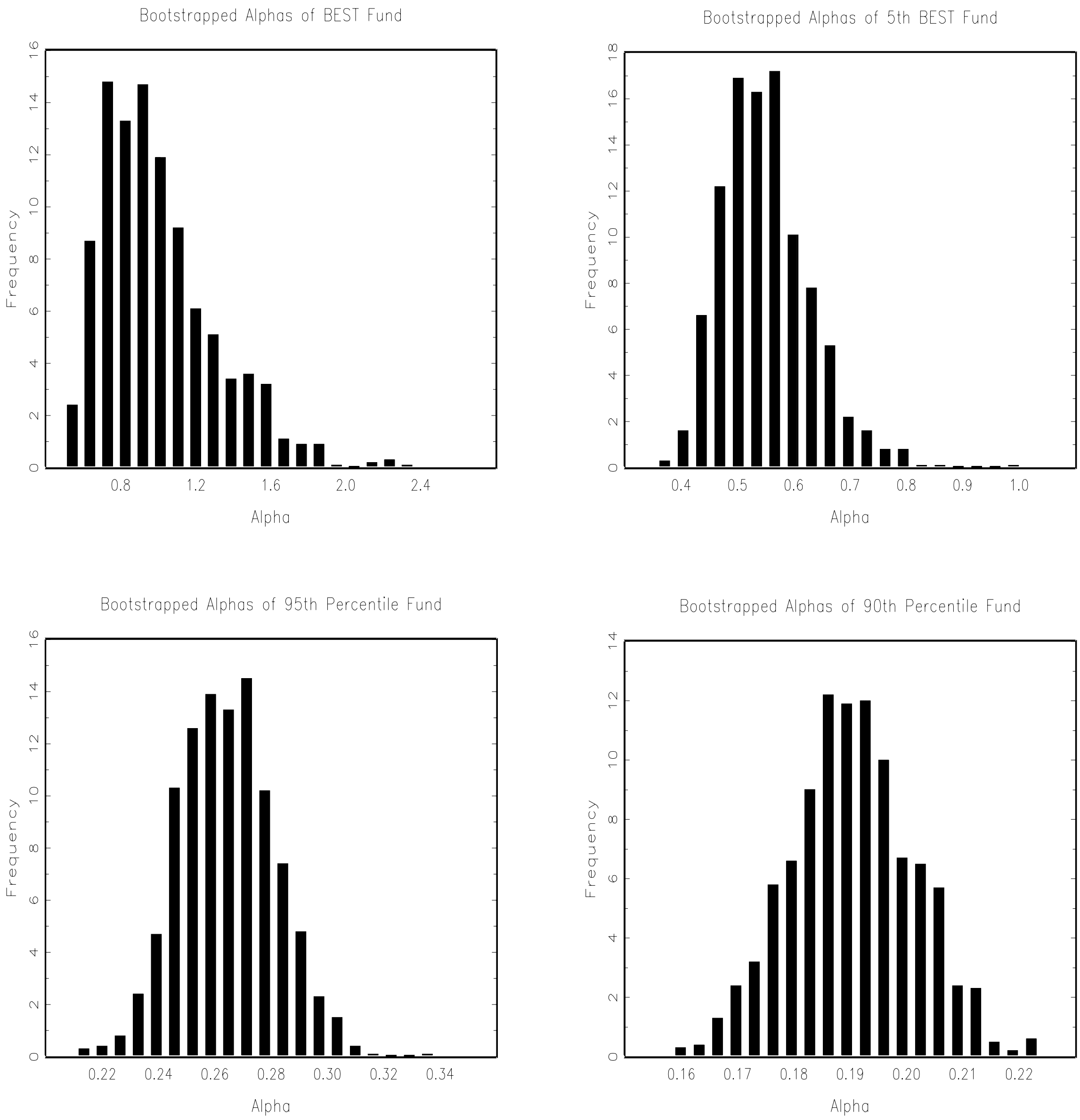


Figure 6.2b: Histograms of Residuals (Upper End of the Distribution)

Figure 6.2b presents histograms of the residuals from the estimation of alpha at various points in the upper end of the performance distribution. These are estimated by the unconditional three-factor performance measurement model over the full sample of mutual funds.

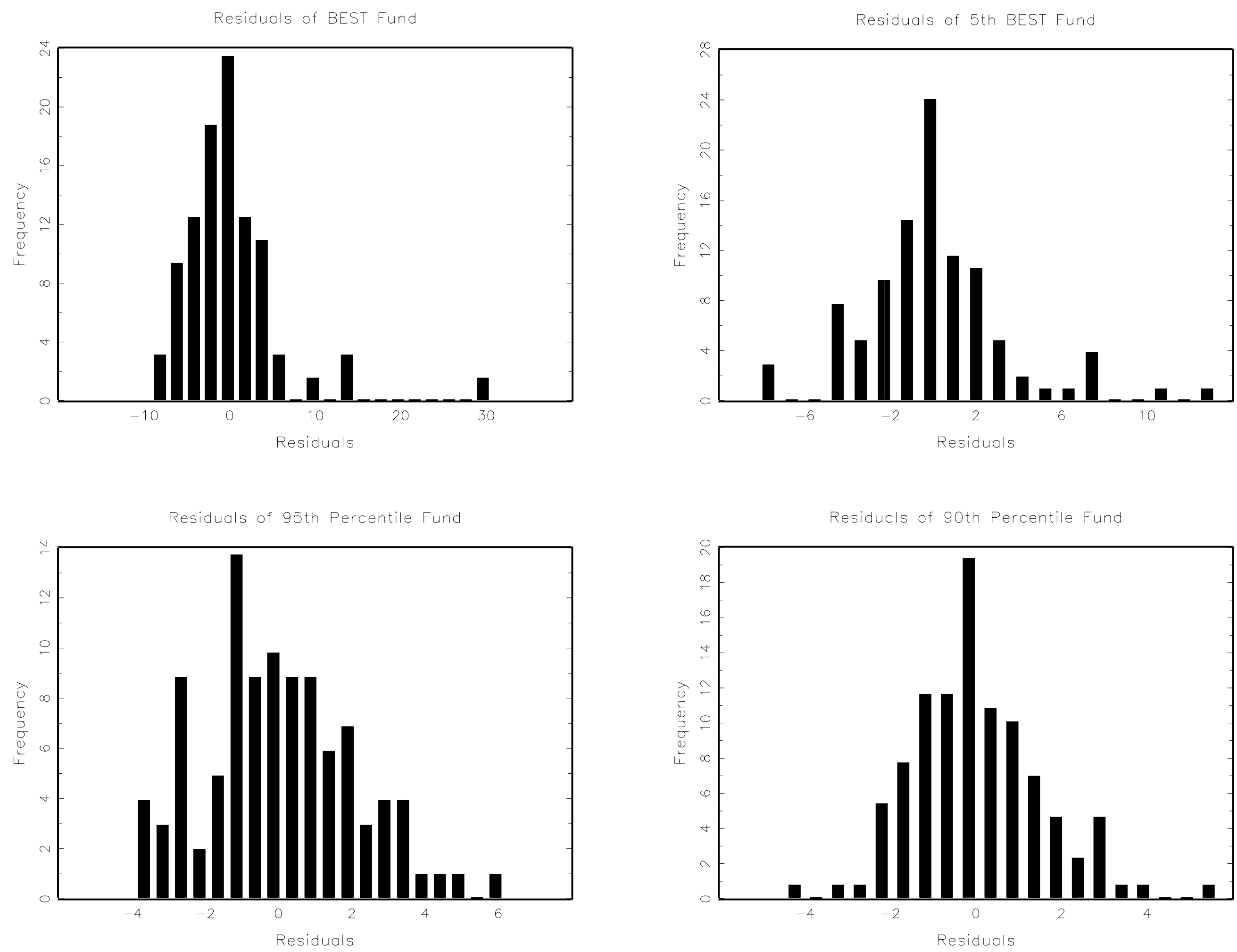


Figure 6.3a: Histograms of Bootstrap Alpha Estimates
(Lower End of the Distribution)

Figure 6.3a presents histograms of the bootstrap alpha estimates at various points in the lower end of the performance distribution. These are estimated by the unconditional three-factor performance measurement model over the full sample of mutual funds.

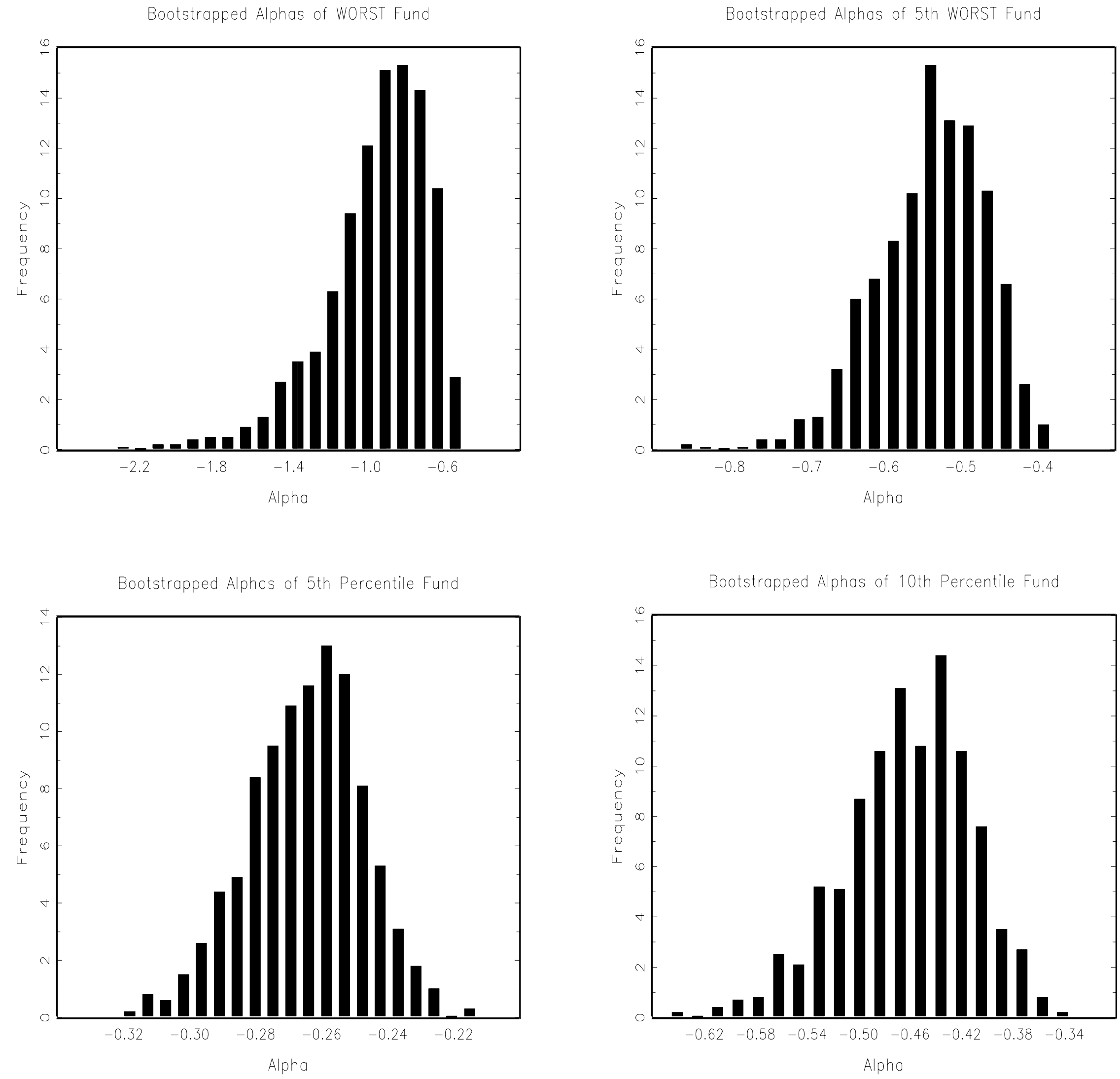
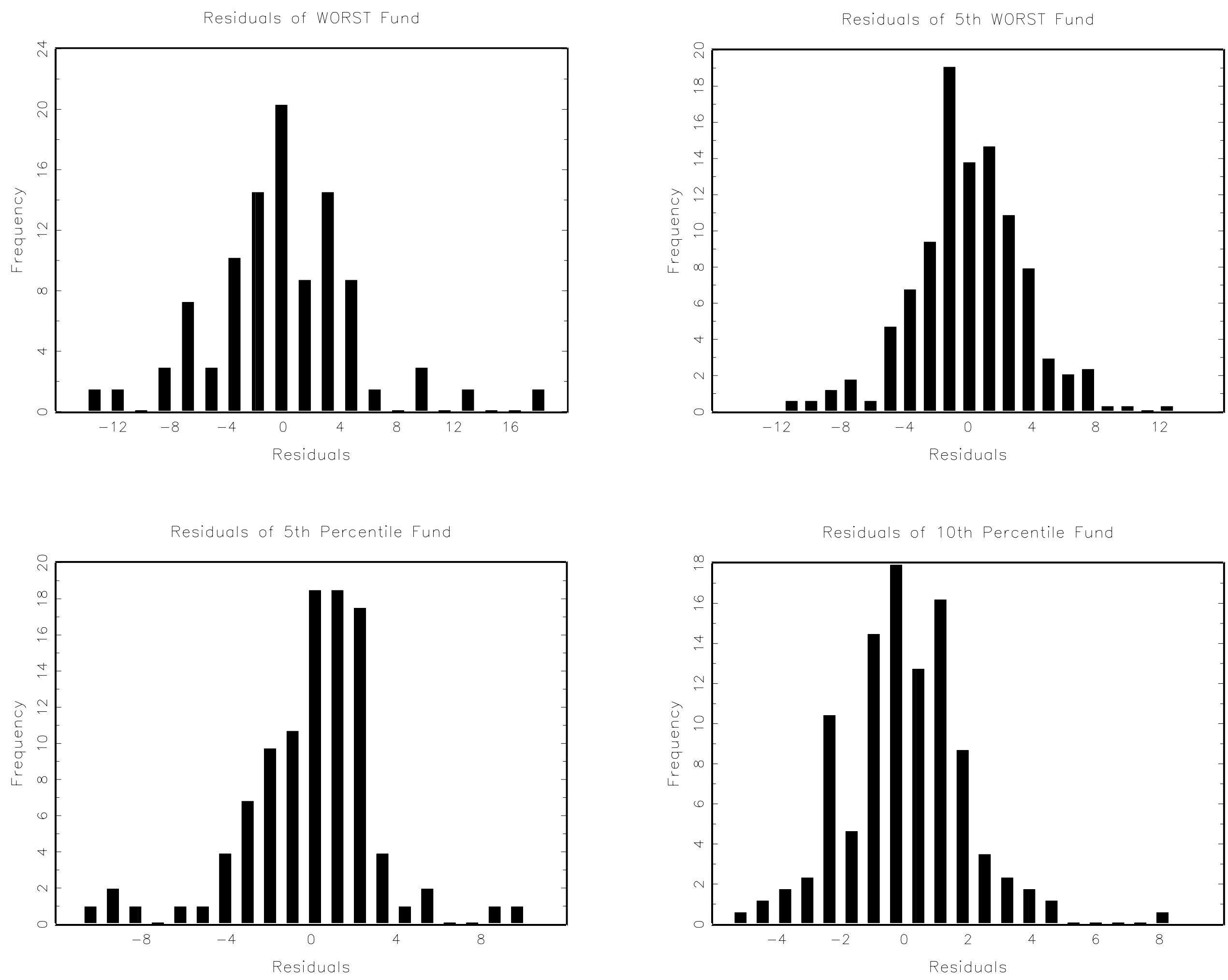


Figure 6.3b: Histograms of Residuals
(Lower End of the Distribution)

Figure 6.3b presents histograms of the residuals from the estimation of alpha at various points in the lower end of the performance distribution. These are estimated by the unconditional three-factor performance measurement model over the full sample of mutual funds.



Section 7. Conclusion

This study evaluates the performance of the UK equity unit trust sector using a large sample of 1,596 trusts over the period April 1975 – December 2002. Fund abnormal performance is first examined by applying a wide range of alternative models of equilibrium security returns, ie of risk adjusted (abnormal) performance measurement. These include models which incorporate dynamically adjusted risk factor loadings conditional on public information signals (conditional beta models), models which allow for conditional abnormal performance based on public information (conditional alpha models) as well as market timing models and unconditional factor models.

Using a number of selected best-fit models a bootstrap procedure is applied to a large sample of funds. The procedure bootstraps fund abnormal performance under a null hypothesis of zero abnormal performance. The bootstrap procedure provides a means of constructing a distribution of performance which is simply due to random sampling variation but where in fact there is no abnormal performance. In turn, this provides an estimate of chance performance or 'luck'. This is, a large number of bootstrap simulations enables the construction of a nonparametric distribution of performance at each point or percentile in the performance distribution. Therefore, given the sample of funds it is possible to establish the distribution of best fund performance, second best performance through all percentiles of performance to the worst fund performance under a null hypothesis of zero abnormal performance. From such a distribution it is possible to assess whether the best fund exhibits genuine stock picking talent or whether in fact we cannot reject the hypothesis that the fund's risk adjusted performance was simply due to random chance alone (luck). Similarly, it is possible to test whether poor fund performance is due to a lack of stock picking ability or whether it is within the boundaries of bad luck. In addition, from the bootstrap distribution it is possible estimate how many funds in the sample one would expect to achieve a given level of performance simply de to chance and compare this to how many funds in fact achieve this level.

Applying a number of alternative models of performance measurement, the results indicate that an unconditional three-factor Fama and French type model with market, size and value risk factors fits well as a model of equilibrium returns. On the other hand there is no advantage in model specifications which conditionalise factor loadings and abnormal performance on public information or specify market timing features.

Applying the bootstrap procedure to the three-factor unconditional model this study finds strong evidence in support of genuine stock picking ability on the part of many top ranked UK equity unit trusts. A number of fund managers in the positive tails of the performance distribution exhibit levels of performance which cannot be attributable to mere random chance. On the other hand for a large number of fund managers in the negative tail of the distribution, the hypothesis that their poor performance is attributable to bad luck is strongly rejected at 95% confidence.

To examine whether stock picking talent among mutual funds may be dependent on the investment style of the fund, the bootstrap methodology is also applied separately to funds classified by four investment styles of objectives. These include objectives of investing in growth stocks, income stocks, general equity (income and growth) and smaller companies. The existence of stock picking talent is more strongly supported among income funds and among growth funds but is less evident among general equity funds and among small stock funds.

References

Affleck-Graves, J. and McDonald, B. (1989). Non-normalities and tests of asset pricing theories, *Journal of Finance*, vol. 44, pp 889-908.

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.

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, 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.

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, S.J. 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.

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

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.

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

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

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. (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.

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

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.

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.

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

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.

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.

Goetzmann, W. and Ibbotson, R. (1994). Do winners repeat? Patterns in mutual fund performance, *Journal of Portfolio Management*, 20, pp 9-18.

Grinblatt, M. and Titman, S. (1989). Mutual fund performance: an analysis of quarterly portfolio holdings, *Journal of Business*, 62, pp 393-416.

Grinblatt, M. and Titman, S. (1992). 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.

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.

Gruber, M. (1996). Another puzzle: the growth in actively managed mutual funds, *Journal of Finance*, Vol. 51, No. 3, pp 783-810.

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.

Henrikson, R. (1984). Market timing and mutual fund performance: an empirical investigation, *Journal of Business*, 57, pp 73-96.

Hochman, S. (1983). The beta coefficient: an instrumental variable approach, *Research in Finance*, 4, pp 392-407.

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.

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.

Jensen, M. (1968). The performance of mutual funds in the period 1945 – 1964, *Journal of Finance*, 23, pp 389-416.

Kmenta, J. (1986). *Elements of Econometrics*, second edition, MacMillan, New York.

Kosowski, R., Timmermann, A., Wermers, R. and White, H. (2001). Can mutual fund stars really pick stocks? New evidence from a bootstrap analysis, Discussion Paper, University of California San Diego.

Kothari, S.P. and Warner, J.B. (2001). Evaluating mutual fund performance, Journal of Finance, Vol. 56, No. 5, pp 1985-2010.

Leger, L. (1997). UK investment trusts: performance, timing and selectivity, *Applied Economics Letters*, 4, pp 207-210.

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.

Malkiel, G. (1995). Returns from investing in equity mutual funds 1971 to 1991, *Journal of Finance*, 50, pp 549-572.

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.

MacKinlay, A.C. and Richardson, M.P. (1991). Using generalized methods of moments to test mean-variance efficiency, *Journal of Finance*, vol. 46, pp 511-527.

Merton, R. (1980). On estimating the expected return on the market: an exploratory investigation, *Journal of Financial Economics*, 8, pp 323-362.

Merton, R.C. and Henriksson, R.D. (1981). On market timing and investment performance II: statistical procedures for evaluating forecasting skills, *Journal of Business*, 54, pp513-33.

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.

Otten, R. and Bams, D. (2000). European mutual fund performance, working paper series, Limburg Institute of Financial Economics, Maastricht University.

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.

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, pp115-135.

Sirri, E.R. and Tufano, P. (1993). Buying and selling mutual funds: flows, performance, fees and services, working paper, Harvard Business School, Cambridge, MA.

Treynor, J. and Mazuy, K. (1966). Can mutual funds outguess the market?, *Harvard Business Review*, 44, pp131-136.

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.

WM Company, (1999). Comparison of Active and Passive Management of Unit Trusts, for Virgin Direct Financial Services.

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.

.