Pensions Institute

DISCUSSION PAPER PI-1404

New Evidence on Mutual Fund Performance: A Comparison of Alternative Bootstrap Methods

David Blake, Tristan Caulfield, Christos Ioannidis and Ian Tonks

June 2014

ISSN 1367-580X

The Pensions Institute
Cass Business School
City University London
106 Bunhill Row
London EC1Y 8TZ
UNITED KINGDOM

http://www.pensions-institute.org/

New Evidence on Mutual Fund Performance:
A Comparison of Alternative Bootstrap Methods

David Blake*

Tristan Caulfield**

Christos Ioannidis***

and

Ian Tonks****

June 2014

Abstract

This paper compares the two bootstrap methods of Kosowski *et al.* (2006) and Fama and French (2010) using a new dataset on equity mutual funds in the UK. We find that: the average equity mutual fund manager is unable to deliver outperformance from stock selection or market timing, once allowance is made for fund manager fees and for a set of common risk factors that are known to influence returns; 95% of fund managers on the basis of the first bootstrap and almost all fund managers on the basis of the second bootstrap fail to outperform the zero-skill distribution net of fees; and both bootstraps show that there are a small group of "star" fund managers who are able to generate superior performance (in excess of operating and trading costs), but they extract the whole of this superior performance for themselves via their fees, leaving nothing for investors.

Keywords: mutual funds, unit trusts, open ended investment companies, performance measurement, factor benchmark models, bootstrap methods, stochastic dominance

JEL: C15, C58, G11, G23

* Pensions Institute, Cass Business School, City University London; ** University College London; ***Department of Economics, University of Bath; **** School of Management, University of Bath

The dataset used in this paper was constructed while Ian Tonks was an ESRC Business Fellow at the UK's Financial Services Authority in 2009 (RES-186-27-0014), and he is grateful to the FSA's Economics of Regulation Unit for hosting this visit.

1. Introduction

Evidence collected over an extended period on the performance of (open-ended) mutual funds in the US (Jensen, 1968; Malkiel, 1995; Wermers et al., 2010) and unit trusts and open-ended investment companies (OEICs)[1] in the UK (Blake and Timmermann, 1998; Lunde et al., 1999) has found that on average a fund manager cannot outperform the market benchmark and that any outperformance is more likely to be due to luck rather than skill.

More recently, Kosowski et al. (2006, hereafter KTWW) reported that the time series returns of individual mutual funds typically exhibit non-normal distributions.[2] They argued that this finding has important implications for the luck versus skill debate and that there was a need to re-examine the statistical significance of mutual fund manager performance using bootstrap techniques. They applied a bootstrap methodology (Efron and Tibshirani (1993), Politis and Romano (1994)) that creates a sample of monthly pseudo excess returns by randomly re-sampling residuals from a factor benchmark model and imposing a null of zero abnormal performance. Following the bootstrap exercise, KTWW determine how many funds from a large group one would expect to observe having large alphas by luck and how many are actually observed. Using data on 1,788 US mutual funds over the period January 1975–December 2002, they show that, by luck alone, 9 funds would be expect to achieve an annual alpha of 10% over a five-year period. In fact, 29 funds achieve this hurdle:

> "*this is sufficient, statistically, to provide overwhelming evidence that some fund managers have superior talent in picking stocks. Overall, our results provide compelling evidence that, net of all expenses and costs (except load charges and taxes), the superior alphas of star mutual fund managers survive and are not an artifact of luck*" (p. 2553).

Applying the same bootstrap method to 935 UK equity unit trusts and OEICs between April 1975–December 2002, Cuthbertson et al. (2008) find similar evidence of significant stock picking ability amongst a small number of top-performing fund

---

[1] These are, respectively, the UK and EU terms for open-ended mutual funds. There are differences, however, the principal one being that unit trusts have dual pricing (a bid and an offer price), while OEICs have single pricing.

[2] KTWW (p.2559) put this down to the possibilities that (1) the residuals of fund returns are not drawn from a multivariate normal distribution, (2) correlations in these residuals are non-zero, (3) funds have different risk levels, and (4) parameter estimation error results in the standard critical values of the normal distribution being inappropriate in the cross section.

managers. Blake et al. (2013) show that fund manager performance improves if the degree of decentralization – in the form of increasing specialization – is increased.

However, these results have been challenged by Fama and French (2010, hereafter FF) who suggest an alternative bootstrap method which preserves any correlated movements in the volatilities of the explanatory factors in the benchmark model and the residuals. They calculate the Jensen alpha for each fund, and then compute pseudo returns by deducting the Jensen alpha from the actual returns to obtain benchmark-adjusted (zero-alpha) returns, thereby maintaining the cross-sectional relationship between the factor and residual volatilities (i.e., between the explained and unexplained components of returns). Their sample consists of 5,238 US mutual funds over the period January 1984–September 2006, and following their bootstrap calculations, they conclude that there is little evidence of mutual fund manager skills.

There are three differences between the KTWW and FF studies. First, while both studies use data for US domestic equity mutual funds, KTWW use data from 1975-2002, whereas the dataset in FF is from 1984-2006. Second, the studies use different fund inclusion criteria: KTWW restrict their sample to funds that have a minimum of 60 monthly observations, whereas FF restrict theirs to funds that have a minimum of 8 monthly observations Third, with respect to the bootstrap method used, the former simulate fund returns and factor returns independently of each other, whereas the latter simulate these returns jointly.

It is important therefore to identify whether the different results from the two studies are due to the different time period analyzed, different inclusion criteria or the different bootstrap methods used. We will use a dataset of UK domestic equity mutual funds returns from January 1998–September 2008 to assess the performance of mutual fund managers. We will also compare the two different bootstrap methods using the *same* sample of funds over the *same* time period and with the *same* fund inclusion criterion.

It is well known that the Jensen alpha measure of performance is biased in the presence of fund manager market timing skills (Treynor and Mazuy, 1966; Merton and Henriksson, 1981). Grinblatt and Titman (1994) have suggested a total performance measure which is the sum of the Jensen alpha and market timing coefficients in an extended factor benchmark model. Allowing for market timing exacerbates the non-normality of standard significance tests and an additional contribution of this paper is to assess the significance of the total performance measure in the KTWW and FF bootstrapped distributions.

The structure of the paper is as follows. Section 2 reviews the approach to measuring mutual fund performance and shows how this approach has recently been augmented through the use of bootstraps. Section 3 discusses the dataset we will be using. The results are presented in Section 4, while Section 5 concludes.

2. Measuring mutual fund performance

2.1 Measuring performance using factor benchmark models

Building on Jensen's original approach, we use a four-factor benchmark model to assess the performance or excess return over the riskless rate ($R_{it} - rf_t$) of the manager of mutual fund *i* obtained in period *t*:

$$R_{it} - rf_t = \alpha_i + \beta_i \left(R_{mt} - rf_t\right) + \gamma_i SMB_t + \delta_i HML_t + \lambda_i MOM_t + \varepsilon_{it} \tag{2.1}$$

where the four common factors are the excess return on the market index $(R_{mt} - rf_t)$, the returns on a size factor, SMB_t, and a book-to-market factor HML_t (Fama-French, 1993), and the return on a momentum factor, MOM_t (Carhart, 1997). The genuine skill of the fund manager, controlling for these common risk factors, is measured by alpha (α_i) which is also known as the selectivity skill.[3]

Under the null hypothesis of no abnormal performance (i.e., no selectivity skill), the $\hat{\alpha}_i$ coefficient should be equal to zero. For each fund, we could test the significance of each $\hat{\alpha}_i$ as a measure of that fund's abnormal performance relative to its standard error. We might also test the significance of the average value of the alpha across the *N* funds in the sample (Malkiel, 1995). Alternatively, we could follow Blake and Timmermann (1998) (and also Fama and French, 2010, Table II) and regress an equal-weighted (or a value-weighted) portfolio *p* of the excess returns $(R_{pt} - rf_t)$ on the *N* funds on the four factors in (2.1) and test the significance of the estimated $\hat{\alpha}_p$ in this regression.

The original Jensen approach made no allowance for the market timing abilities of fund managers when fund managers take an aggressive position in a bull

[3] Ferson and Schadt (1996) suggest a conditional version of this four-factor benchmark model that controls for time-varying factor loadings. However Kosowski et al. (2006) report that the results from estimating the conditional and unconditional models are very similar, and in the remainder of this paper we follow them and only consider the unconditional version of (2.1).

market (by holding high-beta stocks) and a defensive position in a bear market (by holding low beta stocks). Treynor and Mazuy (1966) tested for market timing by adding a quadratic term in the market excess return in the benchmark model to capture the "curvature" in the fund manager's performance as the market rises and falls. To test jointly for selectivity and market timing skills, we estimate a five-factor benchmark model:

$$R_{it} - rf_t = \alpha_i + \beta_i \left(R_{mt} - rf_t\right) + \gamma_i\, SMB_t + \delta_i\, HML_t + \lambda_i MOM_t + \eta_i \left(R_{mt} - rf_t\right)^2 + \varepsilon_{it} \qquad (2.2)$$

Market timing ability is measured by the sign and significance of $\hat{\eta}_i$. To capture both selectivity and timing skills simultaneously, we use the Treynor-Mazuy total performance measure (TM_i) derived in Grinblatt and Titman (1994):

$$TM_i = \alpha_i + \eta_i Var\left(R_m - rf\right) \qquad (2.3)$$

who show that the significance of TM_i can be assessed with respect to its standard error defined in Grinblatt and Titman (1994, Appendix B, p. 441).

2.2 Measuring performance using bootstrap methods

On account of non-normalities in returns, bootstrap methods can be applied to both of the factor benchmark models (2.1) and (2.2) to assess performance. To apply the KTWW bootstrap in (2.1), we first obtain OLS-estimated alphas, factor loadings and residuals using a time series of monthly excess returns for fund *i* in equation (2.1). We then construct a sample of pseudo excess returns by randomly re-sampling residuals with replacement from $\{\hat{\varepsilon}_{it}, t = T_{i0}, \ldots, T_{i1}\}$ and impose the null of zero abnormal performance ($\alpha_i = 0$):

$$(R_{it} - rf_t)^b \equiv \hat{\beta}_i \left(R_{mt} - rf_t\right) + \hat{\gamma}_i\, SMB_t + \hat{\delta}_i\, HML_t + \hat{\lambda}_i MOM_t + \hat{\varepsilon}_{it}^b \qquad (2.4)$$

where b is the b^{th} bootstrap and $\hat{\varepsilon}_{it}^b$ is a drawing from $\{\hat{\varepsilon}_{it}, t = T_{i0}, \ldots, T_{i1}\}$. By construction, this pseudo excess return series has zero alpha. For bootstrap b = 1, we regress the pseudo excess returns on the factors:

$$(R_{it} - rf_t)^b = \alpha_i + \beta_i \left(R_{mt} - rf_t\right) + \gamma_i\, SMB_t + \delta_i\, HML_t + \lambda_i MOM_t + \tilde{\varepsilon}_{it} \qquad (2.5)$$

and save the estimated alpha. We repeat for each fund, $i = 1, \ldots, N$, to arrive at the first draw from the cross-section of bootstrapped alphas $\{\tilde{\alpha}_i^b, i = 1, ..., N; b = 1\}$ and the corresponding t-statistics $\{t(\tilde{\alpha}_i^b), i = 1, ..., N; b = 1\}$. We then repeat for all bootstrap iterations b = 1,…, 10,000. Note that the common risk factors are not re-sampled in the KTWW bootstrap: their historical ordering is not varied across simulation runs.

We now have the cross-sectional distribution of alphas from all the bootstrap simulations $\{\tilde{\alpha}_i^b, i=1,\ldots,N; b=1,...,10,000\}$ that result from the sampling variation under the null that the true alpha is zero. The bootstrapped alphas can be ranked from smallest to largest to produce the "luck" (i.e., pure chance or zero-skill) cumulative distribution function (CDF) of the alphas. We have a similar cross-sectional distribution of bootstrapped t-statistics $\{t(\tilde{\alpha}_i^b), i=1,\ldots,N; b=1,...,10,000\}$ which can be compared with the distribution of actual $\{t(\hat{\alpha}_i), i=1,\ldots,N\}$ values once both sets of t-statistics have been re-ordered from smallest to largest. We follow KTWW who prefer to work with the t-statistics rather than the alphas, since the use of the t-statistic "*controls for differences in risk-taking across funds*" (p. 2555).[4]

FF employ an alternative bootstrap method. They calculate alpha for each fund using the time series regression (2.1) as in KTWW. But FF do not re-sample the residuals of each individual fund as in KTWW, rather they re-sample with replacement over the full cross section of returns, thereby producing a common time ordering across all funds in each bootstrap. In our study, we re-sample from all 129 monthly observations in the dataset and we impose the null hypothesis as in FF by subtracting the estimate of alpha from each re-sampled month's returns.[5] For each fund and each bootstrap, we regress the pseudo excess returns on the factors:

$$\left[(R_{it}-rf_t)-\hat{\alpha}_i\right]^b = \alpha_i+\beta_i\left(R_{mt}-rf_t\right)+\gamma_i SMB_t+\delta_i HML_t+\lambda_i MOM_t+\tilde{\varepsilon}_{it} \tag{2.6}$$

and save the estimated bootstrapped alphas $\{\tilde{\alpha}_i^b, i=1,\ldots,N; b=1,...,10,000\}$ and t-statistics $\{t(\tilde{\alpha}_i^b), i=1,\ldots,N; b=1,...,10,000\}$. We then rank the alphas and t-statistics from lowest to highest to form the two "luck" distributions under the null hypothesis.

---

[4] KTWW (p. 2559) note that the t-statistic also provides a correction for spurious outliers by dividing the estimated alpha by a high estimated standard error when the fund has a short life or undertakes risky strategies.

[5] To illustrate, for bootstrap b = 1, suppose that the first time-series drawing is month $t=37$, then the first set of pseudo returns incorporating zero abnormal performance for this bootstrap is found by deducting α_i from $(R_{i,37}-rf_{37})$ for every fund i that is in the sample for month $t=37$. Suppose that the second time-series drawing is month $t=92$, then the second set of pseudo returns is found by deducting α_i from $(R_{i,92}-rf_{92})$ for every fund i that is in the sample for month $t=92$. After T drawings, the first bootstrap is completed.

The most important difference between the two bootstrap methods is that the KTWW bootstrap assumes independence between the residuals across different funds and that the influence of the common risk factors is fixed historically. In other words, the KTWW bootstrap assesses fund manager skill controlling only for the effect of non-systematic risk. By contrast, the FF bootstrap preserves the cross-correlation of returns across both funds and common risk factors. This implies that the FF bootstrap assesses fund manager skill controlling for both systematic and non-systematic risk. This is because, with the FF bootstrap, the factor loadings representing systematic risk are re-estimated with every bootstrap, whilst preserving the cross-correlation of fund returns.[6]

There are two other differences between the two bootstrap methods as implemented in the two studies. KTWW include funds in their analysis with more than 60 monthly observations in the dataset, whereas the fund inclusion criterion with FF is 8 months. The KTWW bootstrap uses 1,000 simulations, whereas the FF bootstrap uses 10,000 simulations.

FF report that the distribution of actual $t(\hat{\alpha}_i)$ values is to the left of that of the "luck" distribution of the bootstrapped $t(\tilde{\alpha}_i^b)$ values, particularly for funds with negative alphas, but also for most funds with positive alphas. FF conclude that there is little evidence of mutual fund manager skills. This contrasts with KTWW who conclude that there are a small number of genuinely skilled "star" fund managers.

FF point out a common problem with both methods. By randomly sampling across individual fund residuals in the first method and across individual time periods in the second, any effects of auto-correlation in returns is lost. KTWW (p. 2582) performed a sensitivity analysis of this issue by re-sampling in time series blocks up to 10 months in length. They found that the results changed very little.

3. Data

The data used in this study combines information from data providers Lipper, Morningstar and Defaqto and consists of the monthly returns on 561 UK domestic equity (open-ended) mutual funds (unit trusts and OIECS) over the period January 1998–September 2008, a total of 129 months. The dataset also includes information

[6] FF argue that the KTWW bootstrap's "*failure to account for the joint distribution of joint returns, and of fund and explanatory returns, biases the inferences of KTWW towards positive performance*" (p. 1940).

on annual management fees, fund size, fund family and relevant Investment Management Association (IMA) sectors. We include in our sample the primary sector classes for UK domestic equity funds with the IMA definitions: UK All Companies, UK Equity Growth, UK Equity Income, UK Equity & Growth, and UK Smaller Companies. The sample is free from survivor bias (see, e.g., Elton et al., 1996; Carpenter and Lynch, 1999) and includes funds that both were created during the sample period and exited due to liquidation or merger. We impose the restriction that funds in the sample must have at least 20 consecutive monthly returns. These criteria result in a final sample of 516 funds which will be used in our bootstrap analysis.

"Gross" returns are calculated from bid-to-bid prices and include reinvested dividends. These are reported net of on-going operating and trading costs, but before the fund management fee has been deducted. [7] We also compute "net" returns for each fund by deducting the monthly equivalent of the annual fund management fee. We have complete information on these fees for 451 funds. For each of the remaining 65 funds, each month we subtract the median monthly fund management fee for the relevant sector class and size quintile from the fund's gross monthly return. As in KTWW and FF, we exclude initial and exit fees from our definition of returns.

Table 3.1 provides some descriptive statistics on the returns to and the size of the mutual funds in our dataset. The average monthly gross return across all funds (equally weighted) and months in the dataset is 0.45% (45 basis points), compared with an average monthly return over the same period of 0.36% for the FT-All Share Index.[8] The overall standard deviation of these returns is 4.82%, and the distribution of returns also emphasises that there is some variability in these returns. In the subsequent regression analysis, we require a minimum number of observations to undertake a meaningful statistical analysis, and we imposed the requirement that time series fund parameters are only estimated when there were 20 or more monthly gross returns for that mutual fund. We also report the distribution of gross returns for the sub-sample of 516 mutual funds with a minimum of 20 time-series observations, and this can be compared with the distribution of returns across the whole sample to

[7] Operating costs include administration, record-keeping, research, custody, accounting, auditing, valuation, legal costs, regulatory costs, distribution, marketing and advertising. Trading costs include commissions, spreads and taxes

[8] Note that the FT-All Share Index return is gross of any costs and fees.

confirm that the sub-sample is indeed representative.[9] Overall, these results indicate that survivorship bias is very low in this dataset. The mean monthly net return is 0.35%, implying that the monthly fund management fee is 0.11%. The mean return is now very close to the mean return of 0.36% for the FT-All Share Index. This provides initial confirmation that the average mutual fund manager cannot "beat the market" (i.e., cannot beat a buy-and-hold strategy invested in the market index), once all costs and fees have been taken into account. The final column shows that the distribution of scheme size is skewed: with the median fund value in September 2008 being £64 million and the mean value £240 million. It can be seen that 10% of the funds have values above £527 million.

4. Results

We now turn to assessing the performance of UK equity mutual funds over the period 1998-2008. The results are divided into four sections. The first section looks at the performance of equal- and value-weighted portfolios of all funds in the sample against the four- and five-factor benchmark models over the whole sample period. The second section compares the alpha performance of all the funds based on the actual *t*-statistics $t(\hat{\alpha}_i)$ from the factor model with the simulated *t*-statistics $t(\tilde{\alpha}_i^b)$ generated by the bootstrap methods of KTWW and FF discussed above. We report the results for both gross and net returns. The third section conducts a total performance comparison based on the actual and simulated *t*-statistics, $t(TM_i)$ and $t(TM_i^b)$, for the two bootstraps, again using both gross and net returns. The fourth section performs a series of stochastic dominance tests involving pairwise comparisons between the distributions (of the *t*-statistics) under the two bootstrap methods and between each of the bootstrapped distributions and the actual distribution of both gross and net returns.

4.1 Performance against the factor benchmark models

Following Blake and Timmermann (1998), Table 4.1 reports the results from estimating the four- and five-factor models (2.1) and (2.2) across all T = 129 time-series observations, where the dependent variable is, first, the excess return on an

[9] We chose 20 observations as a compromise between the 8 observations that FF use – which we believe involves too few degrees of freedom in the regression equations – and the 60 observations used by KTWW – which could result in survivor bias. Panel A confirms that the distribution of net returns with 20 or more observations is very similar to the distribution with the full sample of funds.

equal-weighted portfolio *p* of all funds in existence at time *t*, and, second, the excess return on a value-weighted portfolio *p* of all funds in existence at time *t*, using starting market values as weights.[10] For each portfolio, the first two columns report the loadings on each of the factors when the dependent variable is based on gross returns, while the second two columns report the corresponding results using net returns. The loadings on the market portfolio and on the SMB_t factor are positive and significant, while the loadings are negative but insignificant on the HML_t factor. The factor loadings are positive but insignificant on the MOM_t factor.[11]

The alphas based on gross returns differ from the corresponding alphas based on net returns by the average level of fund management fees. However, the most important point is that the alpha (α_p) is not significant in the four-factor model and the total performance measure ($TM_p = \alpha_p + \eta_p Var\left(R_m - rf\right)$) is not significant in the five-factor model. In the latter case, while α_p can be significant – as in the case of the equal-weighted portfolio using gross returns at the 10% level – this is more than compensated for by the significantly negative loading on $\left(R_{mt} - rf_t\right)^2$. This holds whether the portfolio is equal-weighted or value-weighted,[12] or whether we use gross returns or net returns. A particularly interesting finding in Table 4.1 is that the estimate for α_p in the four-factor model is very similar in size to the estimate of TM_p in the corresponding five-factor model, even though both estimates are not statistically significant.[13] Again this is true whether we compare on the basis of gross or net returns, or an equal- or value-weighted portfolio. This can only happen, of

[10] We use the monthly FTSE All-Share Index as the market benchmark for all UK equities. We take the excess return of this index over the UK Treasury bill rate. $SMB_t, HML_t,$ and MOM_t are UK versions of the other factor benchmarks as defined in Gregory et al. (2013).

[11] Note that the estimated factor loadings for the models where the dependent variable is based on gross returns are very similar to those in the corresponding models where the dependent variable is based on net returns. This is because the fund management fee is fairly constant over time. While this will lead to different estimates of the intercept (α_p) in a regression equation, it will not lead to significant changes in the estimates of the slope coefficients.

[12] The lower values of α_p and TM_p in the value-weighted regressions compared with the corresponding equal-weighted regressions indicates diseconomies-of-scalein fund management performance.

[13] Grinblatt and Titman (1994, p. 438) report the same result in their dataset and argue that "*the measures are similar because very few funds successfully time the market. In fact, the measures are significantly different for those funds that appear to have successfully timed the market*".

course, if the estimate of α_p in the five-factor model is lower than the estimate of α_p in the corresponding four-factor model by an amount approximately equal to the size of $\eta_p Var\left(R_m - rf\right)$.

The implication of these results is that the average equity mutual fund manager in the UK is unskilled in the sense of being unable to deliver outperformance (i.e., unable to add value from the two key active strategies of stock selection and market timing), once allowance is made for fund manager fees and for a set of common risk factors that are known to influence returns, thereby reinforcing our findings from our examination of raw returns in Table 3.1. But what about the performance of the best and worst fund managers? To assess their performance, we need to turn to the bootstrap analysis.

4.2 Alpha performance of returns using the KTWW and FF bootstraps

We estimate the four-factor benchmark model (2.1) across *N* = 516 mutual funds with at least 20 monthly time series observations between 1998 and 2008. We now have a cross section of *t*-statistics on alpha which can be ranked from lowest to highest to form a cumulative distribution function (CDF) of the $\{t(\hat{\alpha}_i), i = 1, \ldots, N\}$ statistics for the actual fund alphas. We also generate 10,000 KTWW and FF bootstrap simulations for each fund as described in Section 2.2 above. For each bootstrap, this will generate a cross section of *t*-statistics on alpha, assuming no abnormal performance. The 5.16 million *t*-statistics can also be ranked from lowest to highest to create a CDF of bootstrapped "luck" $\{t(\tilde{\alpha}_i^b), i = 1, \ldots, N; b = 1, \ldots, 10,000\}$ statistics for each bootstrap. We compare the averaged values in selected percentile ranges of the CDF of the *t*-statistics on the actual alphas ($t(\hat{\alpha})$) with the averaged values of the *t*-statistics derived from the KTWW and FF bootstrap simulations ($t(\tilde{\alpha}^b)$) in the same percentile ranges. We report the results of the analysis first using gross and then net returns.

4.2.1 Alpha performance based on gross returns

Panel A of Table 4.2 looks at alpha performance based on gross returns and Figure 4.1 illustrates the results graphically. The left tail of the CDF of the actual *t*-statistics lies to the left of that of both bootstraps. For example, in the percentile range 4-4.99, the actual *t*-statistic is -1.85, while the KTWW *t*-statistic is -1.65 and the FF *t*-statistic is -

1.71. This suggests that those funds in the bottom of the distribution are there as a result of poor skill rather than bad luck. This holds for most of the distribution of returns. Only for percentiles of the CDF above about 70% is it the case that the actual *t*-statistics begin to exceed those from either simulation method. For example, in the percentile range 95-95.99, the actual *t*-statistic is 2.32, while the KTWW *t*-statistic is 1.72 and the FF *t*-statistic is 1.80. This means that those funds above the 70th percentile outperform their luck distribution providing evidence of skill in terms of gross returns.

4.2.2 Alpha performance based on net returns

Assessing alpha performance using net returns rather than gross returns raises the performance hurdle, since we are now assessing whether fund managers are able to add value for their investors after covering their operating and trading costs *and* their own fee. Subtracting fees from gross returns to derive net returns will reduce the values of both the actual alphas and their *t*-statistics. Figure 4.2 shows the consequences of this graphically: the CDF of the actual *t*-statistics of the alphas shifts significantly to the left.[14] This is confirmed by Panel B of Table 4.2. For example, in the percentile range 4-4.99, the actual *t*-statistic is -2.59, down from -1.85 in Panel A. By contrast, there is little change in either the KTWW *t*-statistic at -1.65 or the FF *t*-statistic at -1.70. This pattern holds for most of the distribution of returns. Funds need to be ranked above the 95th percentile before they generate *t*-statistics for actual returns above those of the KTWW bootstrap. They never exceed those of the FF bootstrap. In other words, above the 95th percentile, the actual distribution lies between the two bootstrap distributions.

[14] However, the CDFs for the averaged values of both the KTWW and FF bootstrap simulations do not move significantly at all when there is a switch from gross to net returns. In the case of the KTWW bootstrap, this can be seen if we set $\alpha_i = 0$ in (2.7) for both gross and net returns and no other variable on the right-hand side of (2.7) changes when we make an allowance for fund manager fees. In the case of the FF bootstrap, the influence of fees is broadly cancelled out in the dependent variable $\left[(R_{it} - rf_t) - \hat{\alpha}_i\right]^b$ in (2.9), since R_{it} will be lower by the i^{th} manager's fee and $\hat{\alpha}_i$ will be lower by the average fee across the sample which will be of similar size. Figures 1 and 2 in FF have the same result.

4.3 TM performance of returns using the KTWW and FF bootstraps

We now repeat the analysis of the previous sub-section but use the five-factor benchmark model (2.2) and focus on the *TM* total performance measure instead of alpha.[15] We report the results of the analysis first using gross and then net returns.

4.3.1 TM performance based on gross returns

Panel A of Table 4.3 looks at *TM* performance based on gross returns. A comparison of the *Act* column in this table with that in Panel A in Table 4.2 shows a remarkable similarity in the values of the *t*-statistics for the *TM* and alpha gross return performance measures at the same percentiles.[16] Both tables demonstrate that it is only for percentiles of the CDF above about 70% that it is the case that the actual *t*-statistics exceed those from either simulation method. For example, in the percentile range 95-95.99, the actual *t*-statistic is 2.35 (compared with 2.32 when the performance measure is alpha), while the KTWW *t*-statistic is 1.71 (compared with 1.72) and the FF *t*-statistic is 1.79 (compared with 1.80). The regression analysis in section 4.1 produced a similar finding. We therefore have the same interpretation of this performance, namely that only a minority of funds are able to generate returns from stock selection and market timing that are more than sufficient to cover their operating and trading costs.

4.3.2 TM performance based on net returns

Panel B of Table 4.3 examines *TM* performance based on net returns. A comparison of the *Act* column in this table with that in Panel B of Table 4.2 shows the same similarity between the values of the *TM* and alpha net return performance measures that the previous sub-section found when looking at gross returns. Funds need to be ranked above the 95th percentile before they are generating *t*-statistics for the actual returns above those of the KTWW bootstrap. They only beat the FF bootstrap above the 96th percentile.

[15] In the case of the FF bootstrap, the dependent variable in (2.9) becomes

$$\left[(R_{it}-rf_t)-\hat{\alpha}_i-\hat{\eta}_i\left(R_{mt}-rf_t\right)^2\right]^b.$$

[16] For the same reason given by Grinblatt and Titman (1994, p. 438) in footnote 13 above.

4.4 Stochastic dominance tests

In this sub-section, we provide a formal comparison of the actual and bootstrap distributions using stochastic dominance techniques. We compare the two bootstraps with each other and compare each bootstrap with the actual distribution from the relevant factor model of $t(\alpha)$ and $t(TM)$ for both gross and net returns. Tables 4.2-4.3 and Figures 4.1-4.2 revealed a consistent pattern. Although the bootstrap distributions for both $t(\alpha)$ and $t(TM)$ are very similar, at low percentiles, the KTWW bootstrapped CDF lies slightly to the right of the FF bootstrapped CDF. At high percentiles, the relationship switches around. The cross-over happens between the 30th and 40th percentiles in the case of alpha (for both gross and net returns) and around the 60th percentile in the case of *TM* (again for both gross and net returns).

The implication is that the KTWW bootstrap sets a higher hurdle than the FF bootstrap at the bottom of the distribution, but a lower hurdle at the top. This means that the KTWW bootstrap will identify marginally more funds in the left tail of the distribution as being *unskilled* compared with the FF bootstrap. It also means that the KTWW bootstrap will identify marginally more funds in the right tail of the distribution as being *skilled* compared with the FF bootstrap, i.e., the KTWW bootstrap will identify more fund manager "stars" than the FF bootstrap. FF suggested this result in their paper, but did not formally test it.

The moments of the actual $t(\alpha)$ and $t(TM)$ distributions (constructed from the 516 factor models) together with the corresponding KTWW and FF bootstrap distributions (constructed from the 5.16 million simulations) are shown in Table 4.4. All the distributions have zero mean by construction. The factor models generate similar distributions for both gross and net $t(\alpha)$ and $t(TM)$: with standard deviations in the range 1.5-1.7, modest positive skewness in the range 0.5 and kurtosis in the range 8-9. The KTWW bootstrap also generates similar distributions for both $t(\alpha)$ and $t(TM)$. The distributions have (approximately) unit variance. They are also fairly symmetric and have a modest degree of excess kurtosis compared with the normal distribution.[17] By contrast, the FF bootstrap distribution has a larger variance and much fatter tails (especially in the case of $t(TM)$, where the left-skew is also more prominent). In order to assess whether these distributions are statistically different

[17] Nevertheless, all the distributions described in Table 4.4 fail a Jarque–Bera test for normality (results not reported).

from each other, we will test for stochastic dominance between: a) the bootstrap distributions, and b) the actual and bootstrap distributions for gross and net returns.

Several methods have been proposed for testing for stochastic dominance and these can be classified into two main groups. The first group of tests relies on the comparison of both distributions at a finite number of grid points (e.g., Anderson, 1996; and Davidson and Duclos, 2000). The use of an $\inf$ or $\sup$ statistic over the support of the distributions is advocated under the second approach by McFadden (1989) and Kaur, Rao and Singh (1994, hereafter KRS). The power of such tests has been studied by a number of authors. Subsequently, Tse and Zhang (2004), Linton et al (2010) and Heathcote et al (2010) have suggested that the performance of such tests, especially in small samples, can be improved by the adoption of bootstrap methods. Given that our sample is of considerable size, we adopt the methodology of Davidson and Duclos (2000, henceforth DD) based upon the findings of Tse and Zhang (2004, p. 377) who report that the DD test dominates the others.

Following the testing procedure described in DD, Heathcote et al (2010) and Tse and Zhang (2004), we may test any order *s* of stochastic dominance for any two random variables (which need not be independent), $Y_1,...,Y_M$ and $Z_1,...,Z_M$, with empirical distribution functions, $\hat{F}_Y(x)$ and $\hat{F}_Z(x)$, by considering the following test statistic:

$$T^s(x) = \frac{\hat{D}_Y^s(x) - \hat{D}_Z^s(x)}{\sqrt{\hat{V}^s(x)}} \tag{4.1}$$

where[18]

$$\hat{D}_Y^s(x) = \int_0^x \hat{D}_Y^{(s-1)}(Y)dY = \frac{1}{(s-1)!}\int_0^x \left(x-Y\right)^{s-1} d\hat{F}_Y(Y) = \frac{1}{N(s-1)!}\sum_{i=1}^{N}\left(x-Y_i\right)_+^{s-1}$$

is a natural estimator of $D_Y^s(x)$ for $s \geq 2$ with $\hat{D}_Y^1(x) = \hat{F}_Y(x)$ (and analogously for $\hat{D}_Z^s(x)$), where (since the distributions are independent) $\hat{V}^s(x) = \hat{V}_Y^s(x) + \hat{V}_Z^s(x)$ is the variance of $\left(\hat{D}_Y^s(x) - \hat{D}_Z^s(x)\right)$ with

[18] $\left(x-Y_i\right)_+$ is defined as $\max\left(\left(x-Y_i\right),0\right)$.

$$\hat{V}_Y^s(x) = \frac{1}{N}\left[\frac{1}{N((s-1)!)}\sum_{i=1}^{N}\left(x - Y_i\right)_+^{2(s-1)} - \hat{D}_Y^s(x)^2\right]$$

(and analogously for $\hat{V}_Z^s(x)$), and where *N* is the sample size.

DD show that, under the null hypothesis:

$$H_0 : D_Y^s(x) = D_Z^s(x) \iff Y = Z .$$

$T^s(x)$ is distributed asymptotically as a standard normal variate against the alternative hypotheses:

$H_A : D_Y^s(x) \neq D_Z^s(x)$ ($\Leftrightarrow Y \neq Z$), but $Y \nsucc_s Z$ and $Z \nsucc_s Y$

$H_{A1} : Y \succ_s Z$

$H_{A2} : Z \succ_s Y$.

Y dominates *Z* stochastically at order *s* if $D_Y^s(x) \leq D_Z^s(x)$ for all $x \geq 0$, $D_Y^s(x) < D_Z^s(x)$ for some *x*, and $\mu(Y) \geq \mu(Z)$, where $\mu(.)$ is the mean of the distribution. We denote this above as $Y \succ_s Z$.

To control for the size of the test, it is conventional to use the studentized maximum modulus (SMM) whose critical values are reported in Stoline and Ury (1979). The significance of the test is determined asymptotically by the critical value of the SMM distribution with *k* and ∞ degrees of freedom, where *k* is the number of values of *x* in the relevant distribution at which the test is performed. We employ the following decision rules:

- If $\left|T^s(x_j)\right| < M_{\infty,\alpha}^k$ for all *j* = 1, …, *k*, accept H_0,
- If $T^s(x_j) < M_{\infty,\alpha}^k$ for all *j* and $-T^s(x_j) > M_{\infty,\alpha}^k$ for some *j*, accept H_{A1},
- If $-T^s(x_j) < M_{\infty,\alpha}^k$ for all *j* and $T^s(x_j) > M_{\infty,\alpha}^k$ for some *j*, accept H_{A2},
- If $T^s(x_j) > M_{\infty,\alpha}^k$ for some *j* and $-T^s(x_j) > M_{\infty,\alpha}^k$ for some *j*, accept H_A.

Panel A of Table 4.5 presents the summary results of our hypothesis tests at the 5% significance level for stochastic dominance between the distributions generated by the two bootstrap methods. H_0 is tested at *k* (= 19) values of *x* (i.e., in 20 equally spaced grids or bins) in each of the distributions.[19] The panel shows that the KTWW

[19] The critical values for *k* = 19 and ∞ degrees of freedom from the SMM distribution are as follows: 4.043 (1% significance level), 3.643 (5%), and 3.453 (10%).

distribution stochastically dominates the FF distribution in terms of $t(\alpha)$ and $t(TM)$ for both gross- and net-returns.[20] FF put forward a number of reasons to explain the differences in the distributions. One reason was different sample periods. Another was different inclusion rules: FF argued that the KTWW rule of only including funds with at least 60 months of returns "*produces more survival bias*" than their rule of including all funds with at least 8 months data (p. 1939). Our study uses the same dataset and the same inclusion rule of at least 20 months in the dataset, so these two explanations are not applicable. We are therefore left with FF's third explanation for the difference between the results of the two bootstraps, namely, the re-sampling assumptions used in the simulations: KTWW independently re-sample only the residuals in each fund's factor benchmark model, while FF simultaneously re-sample across all the funds' returns and factors.

We now turn to the stochastic dominance tests between the actual distributions and those generated by the bootstraps for both gross- and net-return $t(\alpha)$ and $t(TM)$. If we look across the whole of these distributions at the 5% level, we find (in unreported results) that, for the gross-return $t(\alpha)$ and $t(TM)$, the actual distribution appears to be equivalent to *both* the KTWW and FF distributions obtained under the null hypothesis that managers possess neither selectivity skills nor timing ability. These results therefore appear to be in line, qualitatively at least, with the conclusions reached by FF who documented absence of specific skills and attribute the incidents of exceptional performance to "luck". We also find that, for the net-return $t(\alpha)$ and $t(TM)$, *both* bootstrap distributions stochastically dominate the actual distributions. This would appear to indicate that the fund managers were not just "unlucky", they demonstrate no skill net of their fee.

However, the truth is more subtle than this if we take a closer look at the funds in the upper tail of the distributions. Following Lean et al. (2008. p.32), we divided the top 10 percent of the distributions into 20 equally spaced minor grids. The results are presented in Panels B and C of Table 4.5 at the 1% significance level. We find that, while the same result as above holds for the net-return $t(\alpha)$ and $t(TM)$, for the gross-return $t(\alpha)$ and $t(TM)$, the actual distributions now stochastically dominates *both*

[20] In most cases, the order of stochastic dominance is $s = 2$, but for the net-return $t(\alpha)$, it is $s = 3$. Since the KTWW distribution stochastically dominates the FF distribution, it provides a higher overall hurdle across the distribution of funds as a whole. It is only for top 5% of funds that the FF hurdle is higher as shown in Tables 4.2 and 4.3.

the KTWW and FF distributions. This supports the KTWW finding that there are a small number of "star" fund managers who have sufficient skills to generate returns enough to cover their operating and trading costs. But – and here is one of our key findings – they extract the full rent from their skills in the fees that they charge.

Finally, we note that the stochastic dominance tests indicate that, although the two bootstraps are statistically different from each other, the quantitative difference between them is small. The results in this sub-section indicate that, depending on the specific hypothesis being tested and the specific significance level of the test, the actual distribution either stochastically dominates *both* bootstrap distributions or is stochastically dominated by *both* bootstrap distributions. Figures 4.1-4.2 reveal how visually close the bootstraps are. This contrasts with the results in Panel B of Tables 4.2 and 4.3 which showed that, above the 95th percentile, the actual distribution lies *between* the two bootstrap distributions.

5. Conclusions

Our paper contributes to the literature in three ways. First, we use a new dataset of UK equity mutual funds to assess the Treynor-Mazuy measure of total performance (*TM*) skills of mutual fund managers using factor benchmark models. *TM* is superior to an assessment based on alpha alone, since it includes market timing skills as well as selectivity skills; most existing studies, including KTWW and FF, only examine selectivity. Second, we compare directly the KTWW and FF bootstrap methods for assessing mutual fund manager performance (both alpha and *TM*) using the *same* funds selected using the *same* inclusion criteria over the *same* sample period.[21] Third, we employ stochastic dominance to test formally whether a) the two bootstrap distributions of the *t*-statistics of the performance measures (alpha and *TM*) and b) the distributions of these *t*-statistics from the factor benchmark models and each of the corresponding bootstraps are statistically different from each other. We conduct the analysis for both gross and net (of fund manager fee) returns. On the basis of a dataset of equity mutual funds in the UK over the period 1998-2008, we draw the following conclusions.

First, the average equity mutual fund manager in the UK is unable to deliver outperformance from either stock selection or market timing, once allowance is made

[21] FF did not reproduce the KTWW bootstrap method on their dataset. They just used their bootstrap method with the KTWW inclusion criterion and sample period to assess the KTWW method.

for fund manager fees and for the set of common risk factors known to influence returns. Second, 95% of fund managers on the basis of the KTWW bootstrap and almost all fund managers on the basis of the FF bootstrap failed to outperform the bootstrap simulations of the net-return $t(\alpha)$ and $t(TM)$ statistics. The *TM* results, in particular, indicate that the vast majority of fund managers are very poor at market timing. Any selectivity skills that fund managers might possess – and at best only a very small number of them do – are wiped out by their attempts to time the market. We note that, on the basis of our UK dataset, the FF bootstrap sets a marginally higher hurdle than the KTWW bootstrap for fund managers to jump over before they can be considered to be "stars". This is because the FF bootstrap controls for the systematic relationship between a fund's returns and the factor benchmarks, while the KTWW bootstrap ignores this relationship and controls only for the non-systematic risk contained in the residuals of the factor benchmark models. Third, and focusing on the upper tail of the distributions, the stochastic dominance tests indicate that the distributions of the actual gross-return $t(\alpha)$ and $t(TM)$ statistics stochastically dominate *both* of the bootstrap distributions, but *both* bootstrap distributions for the net-return $t(\alpha)$ and $t(TM)$ stochastically dominate the actual distributions.

Taken together, the above results prove that the vast majority of fund managers in our dataset were not simply unlucky, they were genuinely unskilled. However, a small group of "star" fund managers are genuinely skilled and hence able to generate superior performance (in excess of operating and trading costs), but they extract the whole of this superior performance for themselves via their fees, leaving nothing for investors.

Our final conclusion is that, while "star" fund managers do exist, all the empirical evidence – including that presented here – indicates that they are incredibly hard to identify. Furthermore, it takes a very long time to do so: Blake and Timmermann (2002) showed that it takes 8 years of performance data for a test of a fund manager's skill to have 50% power and 22 years of data for the test to have 90% power. For most investors, our results show that it is simply not worth paying the vast majority of fund managers to actively manage their assets.

Table 3.1: Descriptive statistics on UK equity mutual funds 1998-2008

	Gross returns	*Gross returns (≥20 months)*	*Net returns (≥20 months)*	*Fund management fee (≥20 months)*	*Size at 30 Sep 2008 (£ millions)*
Mean	0.0045	0.0047	0.0035	0.0011	240.86
Std. dev.	0.0482	0.0480	0.0480	0.0002	656.56
Between std. dev.	0.0113				
Within std. dev.	0.0480				
10%	-0.0587	-0.0582	-0.0593	0.0008	8.45
25%	-0.0186	-0.0184	-0.0195	0.0010	26.6
50%	0.0128	0.0130	0.0118	0.0012	64.27
75%	0.0330	0.0330	0.0319	0.0012	202.25
90%	0.0525	0.0526	0.0514	0.0012	527.1
Obs.	48,061	47,492	47,492	47,492	
No. of funds	561	516	516	516	287

Note: the table reports average monthly returns from February 1998 to September 2008 (129 months). It also reports average monthly fund management fees over the same period, as well as the size of funds at the end of the sample period.

Table 4.1: Estimates of the four-factor and five-factor models for an equal-weighted and a value-weighted portfolio of UK equity mutual funds 1998-2008

	Equal-weighted				*Value-weighted*			
	Gross returns	*Gross returns with market timing*	*Net returns*	*Net returns with market timing*	*Gross returns*	*Gross returns with market timing*	*Net returns*	*Net returns with market timing*
α_p	0.0002	0.0016*	-0.0010	0.0005	-0.0002	0.0014	-0.0013	0.0003
	(0.20)	(1.71)	(-1.27)	(0.49)	(-0.21)	(1.414)	(-1.62)	(0.27)
$(R_{mt}-rf_t)$	0.9490***	0.9167***	0.9485***	0.9168***	0.9380***	0.9037***	0.9379***	0.9038***
	(41.53)	(40.36)	(41.46)	(40.3)	(41.39)	(43.29)	(41.39)	(43.29)
SMB_t	0.2526***	0.2522***	0.2528***	0.2524***	0.1832***	0.1828***	0.1834***	0.1829***
	(9.96)	(10.88)	(9.96)	(10.88)	(7.35)	(8.33)	(7.36)	(8.34)
HML_t	-0.0298	-0.0318	-0.0298	-0.0318	-0.0068	-0.0090	-0.0068	-0.0089
	(-1.27)	(-1.40)	(-1.26)	(-1.40)	(-0.30)	(-0.42)	(-0.30)	(-0.42)
MOM_t	0.0178	0.0136	0.0178	0.0135	0.0031	-0.0015	0.0031	-0.0015
	(0.99)	(0.78)	(0.98)	(0.78)	(0.17)	(-0.09)	(0.17)	(-0.09)
$(R_{mt}-rf_t)^2$		-0.8117**		-0.8102*		-0.8725**		-0.8694*
		(-2.16)		(-2.16)		(-2.15)		(-2.15)
TM_p		0.0002		-0.0010		-0.0001		-0.0012
		(0.24)		(1.28)		(-0.18)		(1.63)
R^2	0.964	0.966	0.964	0.966	0.958	0.961	0.958	0.961
Obs.	129	129	129	129	129	129	129	129

Note: The results are based on (2.1) without market timing and (2.2) with market timing. The dependent variable, $(R_{pt}-rf_t)$, is either the excess return on an equal-weighted portfolio or on a value-weighted portfolio *p* of all funds in existence at time *t*. The dependent variable is measured both gross and net of fund management fees. The total performance measure ($TM_p = \alpha_p + \eta_p Var(R_m - rf)$) is also reported. Relevant *t*-statistics estimated from White (1980)'s robust standard errors are reported in brackets underneath each parameter estimate: ***, ** and * denotes significance at the 1%, 5% and 10% levels.

Table 4.2: Percentiles of the actual and average KTWW and FF bootstrap cumulative density functions of $t(\alpha)$ in the four-factor model for both gross and net returns of UK equity mutual funds 1998-2008

	Panel A: Gross returns			Panel B: Net returns		
Pct	*Act*	*Sim* (*KTWW*)	*Sim* (*FF*)	*Act*	*Sim* (*KTWW*)	*Sim* (*FF*)
0-0.99	-2.6623	-2.2823	-2.4002	-3.4334	-2.2824	-2.3944
1-1.99	-2.3242	-2.0402	-2.1350	-3.1367	-2.0411	-2.1279
2-2.99	-2.1203	-1.8764	-1.9560	-2.9894	-1.8777	-1.9499
3-3.99	-2.0227	-1.7492	-1.8192	-2.8019	-1.7516	-1.8130
4-4.99	-1.8544	-1.6453	-1.7063	-2.5927	-1.6473	-1.7005
10-10.99	-1.5273	-1.2817	-1.3196	-2.2752	-1.2826	-1.3142
20-20.99	-0.9955	-0.8387	-0.8556	-1.7638	-0.8378	-0.8511
30-30.99	-0.7069	-0.5187	-0.5246	-1.3492	-0.5184	-0.5209
40-40.99	-0.3207	-0.2457	-0.2445	-0.9690	-0.2454	-0.2407
50-50.99	-0.0291	0.0094	0.0168	-0.6964	0.0092	0.0203
60-60.99	0.2453	0.2643	0.2784	-0.3662	0.2645	0.2818
70-70.99	0.5860	0.5395	0.5603	-0.0147	0.5388	0.5640
80-80.99	0.9750	0.8632	0.8954	0.3946	0.8622	0.8978
90-90.99	1.5603	1.3215	1.3752	0.9655	1.3197	1.3761
95-95.99	2.3221	1.7190	1.7987	1.7569	1.7173	1.7983
96-96.99	2.4773	1.8417	1.9323	1.9018	1.8399	1.9299
97-97.99	2.6135	1.9987	2.1058	2.0240	1.9988	2.1021
98-98.99	2.7891	2.2296	2.3637	2.2873	2.2308	2.3591
99-100	3.6649	2.7961	3.0522	3.0392	2.7964	3.0425

Note: The results are based on the four-factor model $R_{it} - rf_t = \alpha_i + \beta_i \left(R_{mt} - rf_t\right) + \gamma_i SMB_t + \delta_i HML_t + \lambda_i MOM_t + \varepsilon_{it} \quad (i = 1, \ldots, 516)$ where the dependent variable is excess gross returns in Panel A and excess net returns in Panel B. The table shows the averaged values in selected percentile ranges (*Pct*) of the cumulative distribution function of the actual $t(\alpha)$ statistics for the estimated alphas (*Act*) in this regression. The table also shows for the same percentile ranges the averaged values of $t(\alpha)$ from 5.16 million simulations of the KTTW and FF bootstraps (*Sim*(*KTWW*) and *Sim*(*FF*)).

Table 4.3: Percentiles of the actual and average KTWW and FF bootstrap cumulative density functions of *t*(*TM*) in the five-factor model for both gross and net returns of UK equity mutual funds 1998-2008

	Panel A: Gross returns			**Panel B: Net returns**		
Pct	*Act*	*Sim* (*KTWW*)	*Sim* (*FF*)	*Act*	*Sim* (*KTWW*)	*Sim* (*FF*)
0-0.99	-2.6736	-2.2908	-2.4820	-3.4453	-2.2910	-2.4760
1-1.99	-2.3809	-2.0483	-2.2020	-3.2097	-2.0488	-2.1966
2-2.99	-2.1268	-1.8835	-2.0162	-2.9931	-1.8846	-2.0116
3-3.99	-2.0142	-1.7563	-1.8755	-2.8065	-1.7575	-1.8694
4-4.99	-1.8733	-1.6513	-1.7606	-2.6565	-1.6523	-1.7535
10-10.99	-1.5304	-1.2847	-1.3633	-2.3031	-1.2857	-1.3567
20-20.99	-1.0018	-0.8394	-0.8898	-1.7824	-0.8408	-0.8836
30-30.99	-0.7093	-0.5187	-0.5529	-1.3865	-0.5207	-0.5477
40-40.99	-0.3200	-0.2457	-0.2684	-0.9742	-0.2467	-0.2639
50-50.99	-0.0233	0.0095	-0.0041	-0.6835	0.0080	0.0006
60-60.99	0.2381	0.2642	0.2612	-0.3644	0.2628	0.2645
70-70.99	0.6093	0.5378	0.5460	-0.0124	0.5361	0.5480
80-80.99	0.9705	0.8603	0.8836	0.4205	0.8590	0.8844
90-90.99	1.5998	1.3174	1.3667	1.0066	1.3167	1.3646
95-95.99	2.3484	1.7132	1.7919	1.7822	1.7135	1.7889
96-96.99	2.4797	1.8350	1.9256	1.9346	1.8355	1.9218
97-97.99	2.6474	1.9927	2.0998	2.0420	1.9923	2.0946
98-98.99	2.8579	2.2229	2.3595	2.3404	2.2205	2.3525
99-100	3.6563	2.7836	3.0643	3.0221	2.7809	3.0569

Note: The results are based on the five-factor model $R_{it} - rf_t = \alpha_i + \beta_i (R_{mt} - rf_t) + \gamma_i SMB_t + \delta_i HML_t + \lambda_i MOM_t + \eta_i (R_{mt} - rf_t)^2 + \varepsilon_{it}$ $(i = 1,...,516)$ where the dependent variable is excess gross returns in Panel A and excess net returns in Panel B. The table shows the averaged values in selected percentile ranges (*Pct*) of the cumulative distribution function of the actual *t*(*TM*) statistics for the estimated *TM*s (*Act*) in this regression. The table also shows for the same percentile ranges the averaged values of *t*(*TM*) from 5.16 million simulations of the KTTW and FF bootstraps (*Sim*(*KTWW*) and *Sim*(*FF*)).

Table 4.4: Moments of the cumulative density functions of $t(\alpha)$ and $t(TM)$ from the actual and bootstrap simulations based on both the gross and net returns of UK equity mutual funds 1998-2008

Method	*Moments*	$t(\alpha)$		$t(TM)$	
		Gross returns	*Net returns*	*Gross returns*	*Net returns*
Actual	Mean	0	0	0	0
	St. dev.	1.511	1.682	1.555	1.727
	Skewness	0.511	0.459	0.490	0.436
	Kurtosis	7.675	8.818	8.050	9.278
KTWW	Mean	0	0	0	0
	St. dev.	1.056	1.056	1.055	1.055
	Skewness	-0.043	-0.044	-0.056	-0.053
	Kurtosis	3.658	3.658	3.658	3.644
FF	Mean	0	0	0	0
	St. dev.	1.160	1.156	1.201	1.194
	Skewness	-0.058	-0.060	-0.278	-0.211
	Kurtosis	5.079	4.953	17.675	11.525

Note: The table shows key moments of the distribution of the $t(\alpha)$ and $t(TM)$ statistics from the actual factor model, the KTWW bootstrap and the FF bootstrap for both gross and net excess returns.

Table 4.5: Stochastic dominance tests: Summary of results

Panel A: Tests of KTWW vs. FF bootstrap cumulative density functions

1. Y = KTWW gross-return t(α) vs. Z = FF gross-return t(α): Accept $H_{A1}: Y \succ_2 Z$
2. Y = KTWW gross-return $t(TM)$ vs. Z = FF gross-return $t(TM)$: Accept $H_{A1}: Y \succ_2 Z$

3a. Y = KTWW net-return t(α) vs. Z = FF net-return t(α) (second-order test): Accept $H_A: Y \neq Z$, but $Y \not\succ_2 Z$ and $Z \not\succ_2 Y$

3b. Y = KTWW net-return t(α) vs. Z = FF net-return t(α) (third-order test): Accept $H_{A1}: Y \succ_3 Z$

4. Y = KTWW net-return $t(TM)$ vs. Z = FF net-return $t(TM)$: Accept $H_{A1}: Y \succ_2 Z$

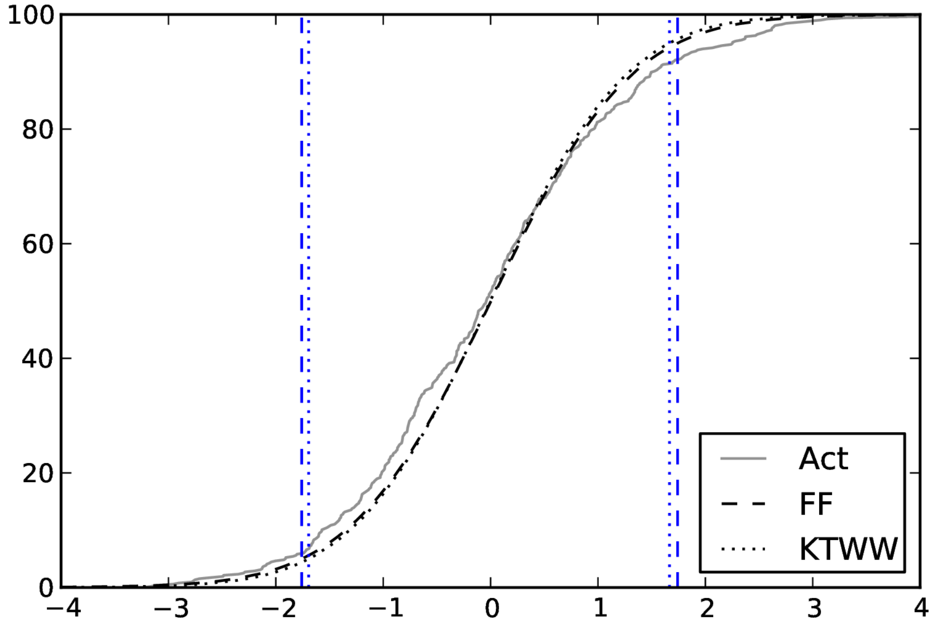
Panel B: Upper tail tests of actual four-factor model cumulative density functions vs. bootstrap cumulative density functions for *t*(α)

1. Y = Actual gross-return t(α) vs. Z = KTWW gross-return t(α): Accept $H_{A1}: Y \succ_2 Z$
2. Y = Actual gross-return t(α) vs. Z = FF gross-return t(α): Accept $H_{A1}: Y \succ_2 Z$
3. Y = Actual net-return t(α) vs. Z = KTWW net-return t(α): Accept $H_{A2}: Z \succ_2 Y$
4. Y = Actual net-return t(α) vs. Z = FF net-return t(α): Accept $H_{A2}: Z \succ_2 Y$

Panel C: Upper tail tests of actual five-factor model cumulative density functions vs. bootstrap cumulative density functions for *t*(*TM*)

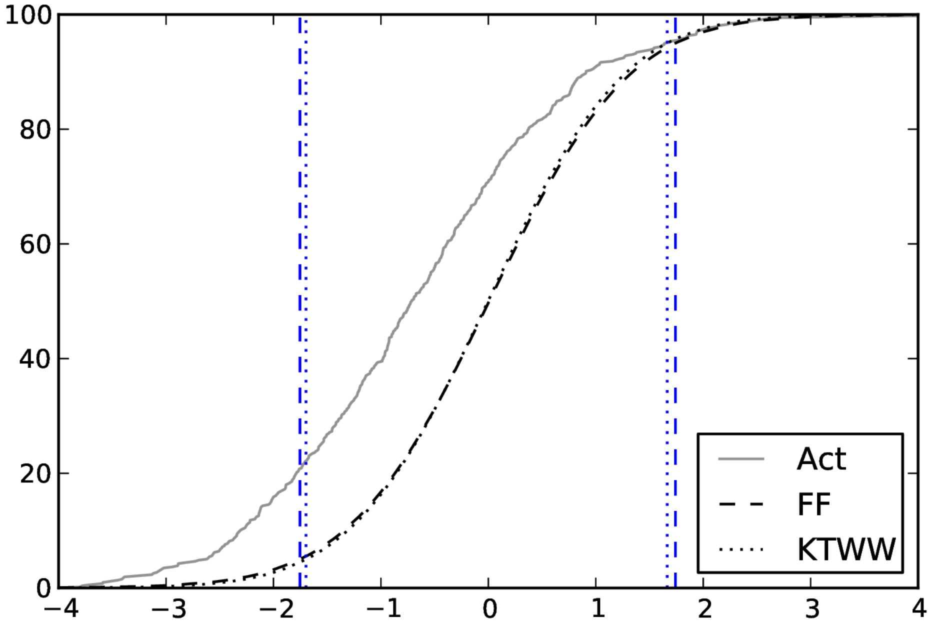
1. Y = Actual gross-return $t(TM)$ vs. Z = KTWW gross-return $t(TM)$: Accept $H_{A1}: Y \succ_2 Z$
2. Y = Actual gross-return $t(TM)$ vs. Z = FF gross-return $t(TM)$: Accept $H_{A1}: Y \succ_2 Z$
3. Y = Actual net-return $t(TM)$ vs. Z = KTWW net-return $t(TM)$: Accept $H_{A2}: Z \succ_2 Y$
4. Y = Actual net-return $t(TM)$ vs. Z = FF net-return $t(TM)$: Accept $H_{A2}: Z \succ_2 Y$

Figure 4.1: Gross-return $t(\alpha)$ – Actual and average KTWW and FF bootstrap cumulative density functions of $t(\alpha)$ in the four-factor model based on the gross returns of UK equity mutual funds 1998-2008



Note: The vertical lines indicate the 5th and 95th percentiles of the distribution of the bootstrapped *t*-statistics.

Figure 4.2: Net-return $t(\alpha)$ – Actual and average KTWW and FF bootstrap cumulative density functions of $t(\alpha)$ in the four-factor model based on the net returns of UK equity mutual funds 1998-2008



Note: The vertical lines indicate the 5th and 95th percentiles of the distribution of the bootstrapped *t*-statistics.

References

Anderson G., 1996. Nonparametric tests of stochastic dominance in income distributions. Econometrica 64, 1183-1193.

Blake, D., Rossi, A., Timmermann, A., Tonks, I., Wermers, R., 2013. Decentralized investment management: evidence from the pension fund industry. Journal of Finance, 68(3), 1133–1178.

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

Blake, D., Timmermann, A., 2002. Performance benchmarks for institutional investors: measuring, monitoring and modifying investment behaviour. In Knight, J., Satchell, S., (eds) Performance Measurement in Finance, Butterworth-Heinemann, Oxford, 108-141.

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

Cuthbertson, K., Nitzche, D., O'Sullivan, N., 2008. UK mutual fund performance: skill or luck. Journal of Empirical Finance 15, 613-634.

Davidson, R., Duclos, J.Y.,2000. Statistical inference for stochastic dominance and the measurement of poverty and inequality. Econometrica 68, 1435-1464.

Efron, B., Tibshirani, R. J., 1993. An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability. Chapman and Hall, New York.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E.F., French, K.R., 2010. Luck versus skill in the cross-section of mutual fund returns. Journal of Finance 65, 607-636.

Ferson, W.E., Schadt, R.W., 1996. Measuring fund strategy and performance in changing economic conditions. Journal of Finance 51, 425–461

Gregory, A., Tharyan, R., Huang, A., 2013. Constructing and testing alternative versions of the Fama-French and Carhart models in the UK, Journal of Business Finance and Accounting, 40(1-2), 172-214.

Grinblatt, M., Titman, S., 1994. A study of monthly mutual fund returns and performance evaluation techniques. Journal of Financial and Quantitative Analysis 29, 419-444.

Heathcote, A., Brown, S., Wagenmakers, E.J, Eidels, A., 2010. Distribution-free test of stochastic dominance in small samples, Journal of Mathematical Psychology 54, 454-463.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. Journal of Finance 23, 389-416.

Kaur, A., Rao, B.L.S.P., Singh, H., 1994. Testing for second-order stochastic dominance of two distributions. Econometric Theory 10, 849-866.

Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. Journal of Finance 59, 2551-2595.

Lean, H. H., Wong, W. K., Zhang, X., 2008. The sizes and powers of some stochastic dominance tests: a Monte Carlo study for correlated and heteroskedastic distributions. Mathematics and Computers in Simulation 79(1), 30-48.

Linton, O., Song, K., Whang Y.-J., 2010. An improved bootstrap test for stochastic dominance. Journal of Econometrics 154, 186-202.

Lunde, A., Timmermann, A., Blake, D., 1999. The hazards of mutual fund underperformance: a Cox regression analysis. Journal of Empirical Finance 6, 121-52.

Malkiel, B.G., 1995. Returns from investing in equity mutual funds 1971 to 1991. Journal of Finance 50, 549–72.

McFadden, D., 1989, Testing for stochastic dominance. In Fomby, T.B. Seo, T.K. (eds.) Studies in the Economics of Uncertainty, Springer Verlag, New York.

Merton, R. C., Henriksson, R. D., 1981. On market timing and investment performance II: Statistical procedures for evaluating forecasting skills. Journal of Business 54, 513–534.

Politis, D. N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303-1313.

Stoline, M. R., Ury H. K, 1979. Tables for the studentized maximum modulus distribution and an application to multiple comparisons among means. Technometrics 21(1), 87-93.

Treynor, J., Mazuy, K., 1966. Can mutual funds outguess the market? Harvard Business Review 44, 131-136.

Tse Y. K., Zhang, X., 2004. A Monte Carlo investigation of some tests for stochastic dominance. Journal of Statistical Computation and Simulation 74 (5), 361–378.

Wermers, R., Barras, L., Scaillet, O., 2010. False discoveries in mutual fund peformance: measuring luck in estimated alphas. Journal of Finance 65, 179-216

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.