European Financial Management, 2010
doi: 10.1111/j.1468-036X.2009.00536.x

False Discoveries in UK Mutual Fund Performance

Keith Cuthbertson and Dirk Nitzsche

Cass Business School, City University, 106 Bunhill Row, London EC1Y 8TZ, UK
E-mail: K.Cuthbertson@city.ac.uk; d.nitzsche@city.ac.uk

Niall O'Sullivan

Department of Economics, University College Cork, Ireland
E-mail: niallmosullican@eircom.net

Abstract

We use a multiple hypothesis testing framework to estimate the false discovery rate (FDR) amongst UK equity mutual funds. Using all funds, we find a relatively high FDR for the best funds of 32.8% (at a 5% significance level), which implies that only around 3.7% of all funds truly outperform their benchmarks. For the worst funds the FDR is relatively small at 7.6% which results in 22% of funds which truly underperform their benchmarks. For different investment styles, this pattern of very few genuine winner funds is repeated for all companies, small companies and equity income funds. Forming portfolios of funds recursively for which the FDR is controlled at a 'acceptable' value, produces no performance persistence for positive alpha funds and weak evidence of persistence for negative alpha funds.

Keywords: *Mutual fund performance*, *false discovery rate*

JEL classification: $C15$, $G11$, $C14$

1. Introduction

In the USA and UK about 70% of institutional funds are actively managed and this rises to over 90% for retail funds. Tests of the performance of active mutual funds are important for investors choosing between active and index funds and for the broader question of the validity of the Efficient Markets Hypothesis EMH, given that the mutual

We thank the anonymous referees and editor, David Barr, Don Bredin, Ales Cerny, Anil Keswani, Ian Marsh, Michael Moore, Mark Taylor, Lorenzo Trapani, Giovanni Urga and seminar participants at Barclays Global Investors, Investment Managers Association Conference: 'Challenges for Fund Management', European Financial Management Association, the FMA/EDHEC conference on fund performance, for discussions and comments.

We gratefully acknowledge financial support from the Irish Research Council for the Humanities and Social Sciences. Correspondence: Keith Cuthbertson.

fund industry appears to be highly competitive with low barriers to entry and plentiful information available at relatively low cost. It is well documented that the average US or UK equity mutual fund underperforms its benchmarks (Elton *et al*., 1993; Wermers, 2000; Quigley and Sinquefield, 2000; Fletcher, 1997). However, the cross-section standard deviation of the alphas for individual funds in both the UK and USA is high, indicating the possibility that some funds are performing very well and others very badly (Malkiel, 1995; Kosowski *et al*., 2006; Cuthbertson *et al*., 2008). In the USA and UK the latter results are not overturned by the addition of market timing variables (Admati *et al*., 1986) or the use of conditional alpha-beta models (Treynor and Mazuy, 1966; Henriksson and Merton, 1981; Ferson and Schadt, 1996; Fletcher, 1995; Leger, 1997).[1]

In this paper we examine the performance of individual funds and address the question of *how many* actively managed UK funds truly have an abnormal net return performance (after adjustments for risk) which is positive, negative or zero.

The standard approach to determining whether the performance of a single fund (or a single portfolio such as the average fund) demonstrates skill or luck is to choose a rejection region and associated significance level γ and to reject the null of 'no outperformance' if the test statistic lies in the rejection region - 'luck' is interpreted as the significance level chosen. However, using $\gamma = 5\%$ when testing the alphas for each of M-funds, the probability of finding *at least one* non-zero fund from a sample of M-funds is much higher than 5% (even if all funds have true alphas of zero).[2] Put another way, if we find 20 out of 200 funds (i.e. 10% of funds) with significant positive estimated alphas when using a 5% significance level then some of these will merely be lucky. One method of dealing with the possibility of false discoveries is to test each of the M-funds independently but use a very conservative estimate for the significance level of each test - for example the Bonferroni test would use $\gamma/M = 0.000125$. This would ensure that the overall error rate in testing M-funds (known as the Family Wise Error Rate) is controlled at γ - but the danger here is in excluding funds that may truly outperform.[3]

In testing the performance of many funds a balanced approach is needed – one which is not too conservative but allows a reasonable chance of identifying those funds with truly differential performance. An approach known as the false discovery rate (FDR) attempts to strike this balance by classifying funds as 'significant' (at a chosen significance level γ) and then asks the question, 'What proportion of these significant funds are false discoveries?' – that is, are truly null (Benjamini and Hochberg, 1995; Storey, 2002; Storey *et al*., 2004). The FDR measures the proportion of lucky funds among a group of funds which have been found to have significant (individual) alphas and hence 'measures' luck among the pool of 'significant funds'. For example, suppose the FDR amongst 20 significant best/winner funds (e.g. those with positive alphas) is 80%

[1] Results for European based funds can be found in Otten and Bams (2002, 2007).

[2] This probability is the compound type-I error. For example, if the M tests are independent then Pr(at least 1 false discovery) $= 1 - (1 - \gamma)^{M} = z_{M}$, which for a relatively small number of $M = 50$ funds and conventional $\gamma = 0.05$ gives $z_{M} = 0.92$ – a high probability of observing at least one false discovery.

[3] Holm (1979) uses a step down method which uses significance level γ/m for the lowest p-value fund and higher significance levels for subsequent ordered p-values, but this also produces conservative inference.

then this implies that only 4 funds (out of the 20) have truly significant alphas[4] - this is clearly useful information for investors. So, investors when forming portfolios need to know both the size of the significant alphas of individual funds and also the FDR amongst these positive-alpha funds.

The competitive model of Berk and Green (2004) suggests that entry and exit of funds should ensure that in equilibrium there are neither funds with long-run positive nor negative abnormal performance. The US mutual fund industry has been extensively analysed and although the UK fund market is smaller, our sample of 675 UK equity funds provides a large comprehensive independent data set, thus mitigating possible claims of data snooping bias if results are only based on analysis of US data.[5]

In this paper we estimate the FDR for all UK equity mutual funds and for different style categories. The change in the FDR as the level of significance changes also allows us to determine whether the truly best and worse performing funds are concentrated or dispersed in the tails of the distribution.

The FDR approach seems to have been used first in testing the difference between genes in particular cancer cells (Storey, 2002) and has recently been used in the economics literature to asses the performance of multiple forecasting rules for foreign exchange (McCracken and Sapp, 2005), stock returns (Bajgrowicz and Scaillett, 2008), hedge funds (Criton and Scaillet, 2009) and to analyse US equity mutual fund performance (Barras *et al*. hereafter BSW, 2009). Here we follow the methodology in BSW but use UK data to examine the performance of UK equity funds over the 'long-run' (whole life of each fund) and 'short-run' (non-overlapping 5-year periods). Persistence in performance is examined by controlling the FDR at the portfolio formation date, in order to limit the proportion of 'lucky funds' in our *ex ante* portfolios. For example, as the significance level is increased, we will obtain more 'significant funds' in our portfolio but if this is accompanied by an increase in the FDR, many of these significant funds may be merely lucky – in forming portfolios of funds it may therefore be prudent to include a small number of significant funds which have a low FDR, rather than form a larger portfolio of significant funds having a higher FDR. This allows us to recursively identify a subset of funds for which the FDR is less than some chosen value, (say 10%) and provides the investor with a subset of significant funds to include in a fund-of-funds portfolio, for which she has set the FDR at an acceptable level. We then estimate the 'forward looking' alpha for this changing portfolio of funds. This would appear to be an intuitively better way to form *ex ante* portfolios than the standard method of including funds in a given fractile many of which could be false discoveries (i.e. lucky funds).

In summary, this paper applies the FDR methodology of BSW (2009) to assess the overall performance of the UK mutual fund sector and to see whether it gives different

---

[4] We use the usual language and terminology found in the statistical literature on false discoveries and error rates. The use of the word 'truly' (sometimes 'genuine' is used) should not be taken to mean that we are 100% certain that a proportion of funds among a particular group of significant funds have non-zero alphas – the FDR even if it is found to be zero, is still subject to estimation error. Also note that the FDR says nothing about the statistical significance of the alpha of any particular *individual* fund - conceptually, the FDR only applies to a group of significant funds.

[5] In other developed countries the mutual fund sector is generally less mature and smaller than in the USA and UK – indeed many countries have little reliable mutual fund returns data and auxiliary variables to capture risk factors or performance attribution are less readily available. Hence the USA has to-date provided most evidence on mutual fund performance.

inferences from the standard approach of 'counting' the number of significant funds. We also apply the FDR approach to performance persistence, where portfolio formation is conditional on a maximum predetermined FDR. Our key results are that there is a much higher proportion of false discoveries among the best funds than among the worst funds – so the standard method of simply counting the number of funds with 'significant' test statistics can be far more misleading for 'winners' than for 'losers'. We find few funds which truly outperform their benchmarks and these skilled funds are concentrated in the extreme right tail of the performance distribution, whereas there is a far greater proportion of genuinely poor performing funds, which are spread throughout the left tail. Around 75% of funds neither statistically beat nor are inferior to their benchmarks and therefore appear to do no better than merely tracking their style indexes.[6] For different investment styles, there are few skilled funds in any of our style categories but the proportion of equity income funds that are truly unskilled is much smaller than the proportion of unskilled funds in either the 'all Companies' or 'small company' sectors. Forming recursive portfolios conditional on a maximum level for the FDR, we find no evidence of persistence for positive alpha funds and weak evidence for persistence amongst negative alpha funds in the 1980s.

The rest of this article is organised as follows. In section 2 we briefly discuss the methodology behind the FDR and other methods of controlling for false positives in a multiple testing framework. In section 3 we outline our data set and in section 4 we evaluate the evidence on UK equity mutual fund *ex post* performance and performance persistence. Section 5 concludes.

2. The False Discovery Rate, FDR

When assessing the overall performance of the mutual fund sector most previous work uses the standard procedure of independently testing each fund's performance and stating the number of funds with significant alphas (hence assuming the FDR is zero). BSW (2009) provide a detailed account of the FDR methodology, so we shall be brief. The null hypothesis is that fund-i has no abnormal performance, the alternative being that the fund delivers either positive or negative performance:

$$H_0 : \alpha_i = 0 \quad H_A : \alpha_i > 0 \text{ or } \alpha_i < 0$$

The issues that arise in multiple testing of M-funds involve choosing a significance level γ and denoting a 'significant fund' as one for which the p-value for the test statistic (e.g. t-statistic on alpha) is less than or equal to some threshold $\gamma/2$ ($0 < \gamma \leq 1$). At a given significance level γ, the probability that a zero-alpha fund exhibits 'good luck' is $\gamma/2$. Hence, if the proportion of truly zero-alpha funds in the population of M-funds is π_0 then the expected proportion of false positives (sometimes referred to as lucky funds) is:

$$E\left(F_\gamma^+\right) = \pi_0(\gamma/2) \tag{1}$$

[6] Using US data Kosowski *et al*. (2006) measure the role of luck in mutual fund performance using p-values of the *ordered individual* funds – however, a simple count of funds with 'significant' p-values ignores the possibility of some significant funds being 'false discoveries'.

If $E(S_\gamma^+)$ is the proportion of significant positive-alpha funds, then the expected proportion of truly skilled funds (at a significance level γ) is:

$$E\left(T_\gamma^+\right) = E\left(S_\gamma^+\right) - E\left(F_\gamma^+\right) = E\left(S_\gamma^+\right) - \pi_0(\gamma/2) \tag{2}$$

(Similar formulae apply for negative-alpha funds). Choosing different levels for γ allows us to see if the number of truly skilful funds rises appreciably with γ or not, which tells us whether skilled funds are concentrated or dispersed in the right tail of the cross-sectional distribution – this information may be helpful in choosing an *ex ante* portfolio of skilled funds. An estimate of the true proportion of skilled (unskilled) funds π_A^+ (π_A^-) in the *population of M-funds* is:

$$\pi_A^+ = T_{\gamma^*}^+ \quad \pi_A^- = T_{\gamma^*}^- \tag{3}$$

where γ^* is a sufficiently high significance level which can be determined using a mean squared error criterion, although setting $\gamma^* = 0.35$-0.45 produces similar results (BSW 2009). The expected FDR amongst the *statistically significant* positive-alpha funds is:

$$FDR_\gamma^+ = \frac{E\left(F_\gamma^+\right)}{E\left(S_\gamma^+\right)} = \frac{\pi_0(\gamma/2)}{E\left(S_\gamma^+\right)} \tag{4}$$

It follows that the proportion of truly positive-alpha skilled funds amongst the *statistically significant* positive-alpha funds is:

$$E\left(T_\gamma^+\right)/E\left(S_\gamma^+\right) = 1 - FDR_\gamma^+ \tag{5}$$

An estimate of $E(S_\gamma^+)$ is the observed number of significant funds S_γ^+. To calculate all the above statistics we now only require an estimate of π_0 the proportion of truly null funds in the population of M-funds. To provide an estimate of π_0 we use the result that truly alternative features have p-values clustered around zero, whereas truly null p-values are uniformly distributed, [0,1]. The simplest method to estimate $\hat{\pi}_0(\lambda)$ is to choose a value λ for which the histogram of p-values becomes flat and to calculate π_0 using:

$$\hat{\pi}_0(\lambda) = \frac{W(\lambda)}{M(1-\lambda)} = \frac{\#\{p_i > \lambda\}}{M(1-\lambda)} \tag{6}$$

where $W(\lambda)/M$ is the area of the histogram to the right of the chosen value of λ (on the x-axis of the histogram) – see figure 1. For example if $\pi_0 = 100\%$ and we choose $\lambda = 0.6$ then $W(\lambda)/M = 40\%$ of p-values lie to the right of $\lambda = 0.6$ and our estimate of $\pi_0 = 40\%/(1\text{-}0.6) = 100\%$ as expected. If there are some truly alternative funds (i.e. $\alpha_i \neq 0$) then the histogram of p-values will have a 'spike' near zero. But if the histogram of p-values is perfectly flat to the right of λ then our estimate of π_0 is independent of the choice of λ. So, if we were able to count only truly null p-values then (6) would give an unbiased estimate of π_0. However, if we erroneously include a few alternative p-values then (6) provides a conservative estimate of π_0 and hence of the FDR.

For finite M, it can be shown that the bias in the estimate of $\hat{\pi}_0(\lambda)$ is decreasing in λ (as the chances of including non-zero alpha-funds diminishes) but its variance increases with λ (as we include fewer p-values in our estimate). An alternative method of estimating π_0 is to plot $\hat{\pi}_0(\lambda)$ against λ, fit a cubic spline to this data and take our estimate to be $\hat{\pi}_0(\lambda = 1)$- this is known as the smoothing method (Storey, 2002). A third method of estimating π_0 is to exploit the bias-variance trade-off and choose λ to minimise the mean-square error $E\{\pi_0(\lambda) - \pi_0\}^2$ - this we refer to as the MSE-bootstrap method (Storey 2002, BSW 2009). In fact, in practice if a sufficiently high value for

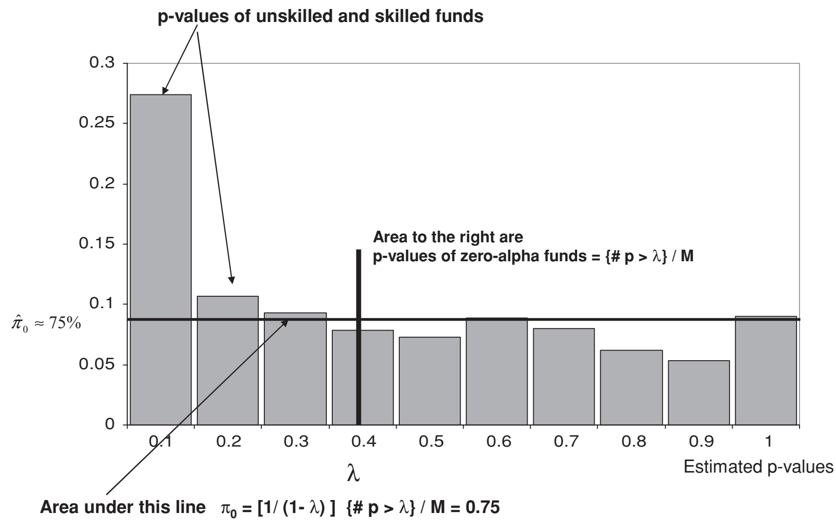

Fig. 1. Calculation of π_0 (3-factor model).

This figure shows the histogram of p-values of alpha from the Fama-French 3 factor model. This distribution provides an estimate of π_0 equal to (0.75) 75%.

$\lambda \approx$ 0.5-0.6 is chosen, the resulting estimate of π_0 is close to that from the other two methods (BSW 2009).

BSW (2009) use a Monte Carlo study to show that the estimators outlined above are accurate, are not sensitive either to *the method* used to estimate π_0 or to the chosen significance level γ and that the estimators are robust to the typical cross-sectional dependence in fund residuals (which tend to be low in monthly data).

Calculation of the FDR depends on correct estimation of individual p-values. Because of non-normality in regression residuals we use a bootstrap approach to calculate p-values of estimated t-statistics (Politis and Romano, 1994; Kosowski *et al*. hereafter KTTW, 2006). Consider an estimated model of equilibrium returns of the form: $r_{i,t} = \hat{\alpha}_i + \hat{\beta}_i' X_t + \hat{e}_{i,t}$ for i $= 1, 2, \ldots,$ M funds, where $T_i =$ number of observations on fund-i, $r_{i,t} =$ excess return on fund-i, $X_t =$ vector of risk factors, $\hat{e}_{i,t}$ are the residuals and $\hat{t}_i$ is the (Newey-West) t-statistic for alpha. For our 'basic bootstrap' we use residual-only resampling, under the null of no outperformance (Efron and Tibshirani, 1993). First, estimate the chosen factor model for each fund and save the vectors $\{\hat{\beta}_i, \hat{e}_{i,t}\}$. Next, draw a random sample (with replacement) of length T_i from the residuals $\hat{e}_{i,t}$ and use these *re-sampled* bootstrap residuals $\tilde{e}_{i,t}$ to generate a simulated excess return series $\tilde{r}_{i,t}$ under the null hypothesis ($\alpha_i = 0$). Then, using $\tilde{r}_{i,t}$ the performance model is estimated and the resulting t-statistic for alpha, t_i^b is obtained. This is repeated B = 1,000 times and for a two-sided, equal-tailed test the bootstrap p-value for fund-i is:

$$p_i = 2.\min\left[B^{-1}\sum_{b=1}^{B} I\left(t_i^b > \hat{t}_i\right), \quad B^{-1}\sum_{b=1}^{B} I\left(t_i^b < \hat{t}_i\right)\right] \tag{7}$$

where $I(.)$ is a (1,0) indicator variable.

3. Performance Models and Data

Our alternative performance models are well known 'factor models' and therefore we only describe these briefly. Each model can be represented in its unconditional, conditional-beta and conditional alpha-beta form. Unconditional models have factor loadings that are time invariant. Carhart's (1997) four-factor (4F) model is:

$$r_{i,t} = \alpha_i + \beta_{1i} r_{m,t} + \beta_{2i} SMB_t + \beta_{3i} HML_t + \beta_{4i} MOM_t + \varepsilon_{i,t} \quad (8)$$

where $r_{i,t}$ is the excess return on fund-i (over the risk-free rate), $r_{m,t}$ is the excess return on the market portfolio while SMB_t, HML_t and MOM_t are factor mimicking portfolios for size, book-to-market value and momentum effects, respectively. The Fama and French (1993) 3F-model includes only $\{r_{m,t}, SMB_t, HML_t\}$ and has mainly been applied to UK funds (e.g. Blake and Timmermann, 1998; Quigley and Sinquefield, 2000; Tonks, 2005) but for US funds Carhart (1997) finds that momentum is also statistically significant.

In conditional alpha-beta models it is assumed that alpha and the factor betas may depend linearly on lagged public information variables Z_t and for the CAPM this gives:

$$r_{i,t+1} = \alpha_i + A_i' \, z_t + b_i \, r_{b,t+1} + B_i'(z_t * r_{b,t+1}) + \varepsilon_{i,t+1} \quad (9)$$

where $r_{b,t+1}$ is the excess return on a benchmark portfolio (i.e. market portfolio in this case) z_t is the vector of deviations of Z_t from its unconditional mean and (A_i', B_i') are suitably dimensioned parameter vectors. Conditional-beta models (Ferson and Schadt, 1996) set $A_i' = 0$. Following earlier studies (Ferson and Schadt, 1996; Christopherson *et al*., 1998) our Z_t variables include permutations of: the one-month T-Bill rate, the dividend yield of the market factor and the term spread.

Our mutual fund data set comprises UK equity Unit Trusts and Open Ended Investment Companies (OEICs) and represent almost the entire set of UK equity funds which have existed at any point during the sample period under consideration, April 1975 to December 2002.[7] By restricting funds to those investing in UK equity, more accurate benchmark factor portfolios may be used in estimating abnormal performance. We have removed 'second units' and index/tracker funds leaving only actively managed funds. The equity funds are categorised by the investment objectives of the funds which include: 'equity income' (162 funds), 'all companies' (i.e. formerly general equity and equity growth, 553 funds) and 'small companies' (127 funds). The data set includes both surviving funds and non surviving funds.

All fund returns are measured gross of taxes on dividends and capital gains and net of management fees. Hence, we follow the usual convention in using net returns (bid-price to bid-price, with gross income reinvested). The market factor used is the FT All Share Index of total returns (i.e. including reinvested dividends). Excess returns are calculated using the one-month UK T-bill rate. The factor mimicking portfolio for the size effect, SMB, is the difference between the monthly returns on the Hoare Govett Small Companies (HGSC) Index and the returns on the FT 100 index.[8] The value premium, HML, is the difference between the monthly returns of the Morgan Stanley Capital International (MSCI) UK value index and the returns on the MSCI

[7] Mutual fund monthly returns data have been obtained from Fenchurch Corporate Services using Standard & Poor's Analytical Software and Data. The data base has been extensively checked for multiple entries and style classifications.

[8] The HGSC index measures the performance of the lowest 10% of stocks by market capitalisation, of the main UK equity market. Both indices are total return measures.

UK growth index.[9] The factor mimicking portfolio's momentum behaviour, MOM, has been constructed using the constituents of the London Share Price Database.[10] Other variables used in conditional and market timing models include the one-month UK T-bill rate, the dividend yield on the FT-All Share index and the slope of the term structure (i.e. the yield on the UK 20 year gilt minus the yield on the UK three-month T-bill).

4. Empirical Results

We begin with a discussion of our preferred factor models. Next we discuss alternative estimation methods for the proportion of truly zero alpha funds π_0, skilled funds, π_A^+ and unskilled funds π_A^- among our total of M-funds. Then we analyse the FDR for the skilled and unskilled funds taken separately – this allows us to ascertain whether such funds are concentrated in the tails of the performance distribution. Next we discuss these metrics for our three investment styles. Persistence in performance is examined by forming portfolios of funds conditional on a pre-specified maximum false discovery rate for potentially skilled (or unskilled) funds. Finally, we examine the sensitivity of the proportion of skilled and unskilled funds across the four different factor models used in our analysis.

4.1 *Preferred models*

In this section, alternative performance models are examined. All tests are conducted at a 5% significance level unless stated otherwise and results presented relate to all UK equity mutual funds over the period April 1975 to December 2002 and are based on 675 funds, all with a minimum of $T_{i,\min} = 36$ observations. For each model, cross-sectional (across funds) average statistics are calculated. A single 'best model' is chosen from each of the 3 model classes; (i) unconditional, (ii) conditional-beta and (iii) conditional alpha-beta, using the Schwartz Information Criterion (SIC) and these results are reported in Table 1.

In the best three models (top half of Table 1), the cross-sectional average alpha takes on a small and statistically insignificant negative value (consistent with Blake and Timmermann, 1998). However, of key importance for this study (and for investors) is the relatively large cross-sectional standard deviations of the alpha estimates which is around 0.26% p.m. (3.1% p.a.), for the unconditional and conditional-beta models and somewhat larger at 0.75% p.m. for the conditional alpha-beta model. This implies that the extreme tails of the distribution may contain funds with abnormally 'good' or 'bad' performance. This is important, since investors are more interested in holding funds in

[9] These indices are constructed by Morgan Stanley who rank all the stocks in their UK national index by their book-to-market ratio. Starting with the highest book-to-market ratio stocks, these are attributed to the value index until 50% of the market capitalisation of the national index is reached. The remaining stocks are attributed to the growth index. The MSCI national indices have a market coverage of at least 60% (more recently this has been increased to 85%). Total return indices are used for the construction of the HML variable.

[10] For each month, the equally weighted average returns of stocks with the highest and lowest 30% returns, over the previous eleven months are calculated. The MOM variable is constructed by taking the difference between these two variables. The universe of stocks is the London Share Price Data Base.

Table 1

Summary Statistics of the UK.

The table provides summary statistics for various factor models including alpha, factor betas, R-squared, Schwartz Information criterion (SIC) as well as measures of skewness and kurtosis. We report mean values using all 675 funds with at least 36 monthly observations. Estimation for each fund uses all available observations and the overall sample period is from April 1975 to December 2002.

	FF 3 Factor		Carhart 4F		Conditional Beta 3F		Conditional Alpha-Beta, 3F	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Alpha	−0.0570	−0.58	−0.0548	−0.51	−0.0319	−0.43	−0.1090	−0.42
$(Rm\text{-}rf)_t$	0.9123	25.19	0.9168	25.85	0.8639	21.18	0.8494	21.02
SMB_t	0.2886	4.58	0.2834	4.65	0.2854	4.51	0.2579	3.63
HML_t	−0.0246	−0.009	−0.0209	−0.14	−0.0236	−0.09	0.0169	0.38
Mom_t	-		0.0087	−0.09	-		-	
z_{t-1} $(Rm\text{-}rf)_t$	-		-		−0.0483	−0.90	−0.0560	−0.82
z_{t-1} SMB_t	-		-		-		−0.0025	0.40
z_{t-1} HML_t	-		-		-		0.0332	0.37
z_{t-1}	-		-		-		−0.0733	
Adj.Rsqd	0.8108		0.8227		0.8147		0.8209	
SIC	1.35		1.32		1.37		1.43	
Skewness	0.19		0.18		0.21		0.19	
Kurtosis	6.21		5.83		6.15		6.04	
# positive alpha	20		34		26		31	
# negative alpha	121		117		98		93	

the right tail of the performance distribution and avoiding those in the extreme left tail, than they are in the average fund's performance.

The excess market return, $r_{m,t}$, and the *SMB* factor betas are consistently found to be statistically significant across all three classes of model, whereas the *HML* factor beta is often not statistically significant, even at a 10% significance level. We find that the momentum factor (*MOM*) is generally not statistically significant at the individual UK fund level (e.g. Blake and Timmermann, 1998; Tonks, 2005), in contrast to US studies (Carhart, 1997; Kosowski *et al*, 2006; BSW, 2009). For the conditional-beta model only the dividend yield variable produces near statistically significant results. In the conditional alpha-beta model we find that none of the conditional alphas has a t-statistic greater than 1.1 but some of the conditional betas are bordering on statistical significance and our best model is shown in the final column of Table 1.

The above results suggest that the unconditional Fama-French 3F-model explains UK equity mutual fund returns data reasonably well. These findings are consistent with existing UK studies (Quigley and Sinqefield, 2000; Fletcher, 1995). Turning now to diagnostics (bottom half of Table 1), the adjusted R^2 across all three models is around 0.8, while the average skewness and kurtosis of the residuals is around 0.2 and 6 respectively and more than 60% of funds have non-normal errors (Bera-Jarque statistic – not reported here). The Schwartz Information Criterion (SIC) is lowest for the

Table 2

The proportion of null funds.

This table provides estimates of the values of the proportion of null funds ($\hat{\pi}_0$) for 4 different factor models and 3 different estimation techniques. The factor models used include the Fama-French 3 factor model, Carhart 4 factor model, as well as a conditional beta Fama-French 3 factor model and conditional alpha-beta Fama-French 3 factor model. To estimate $\hat{\pi}_0$ all 675 funds are used.

Model	FF 3 Factor	Carhart 4F	Conditional Beta 3F	Conditional Alpha-Beta 3F
Histogram $\lambda = 0.5$	0.72	0.63	0.78	0.78
Smoothing technique	0.84	0.62	0.89	0.82
Bootstrapping	0.72	0.64	0.74	0.73

unconditional 3F-model. The Fama-French 3F-model was selected as the 'best model' for all three categories: unconditional, conditional beta and conditional alpha-beta model but because the 4F model is widely used for US equity mutual funds, we also report some variants using this model.[11]

4.2 *Estimation of* π_0

The histogram of p-values is given in figure 1 for the unconditional 3F-model. Exploiting the fact that truly null p-values are uniformly distributed [0, 1], the height of the flat portion of the histogram gives an estimate of π_0. From figure 1 a reasonable 'eyeball' estimate would be $\lambda = 0.5$ giving $\hat{\pi}_0(\lambda) = 0.72$. Alternative estimates given by the smoothing and the bootstrap techniques for the four different factor models are given in Table 2.

Looking down the four columns in Table 2 we see that for any given factor model the estimate $\hat{\pi}_0(\lambda)$ is reasonably constant across the three different estimation methods. For three of the four factor models, the three alternative estimation methods give reasonably similar estimates of $\hat{\pi}_0(\lambda)$ of around 75-85% but for the Carhart 4F model $\hat{\pi}_0(\lambda)$ is somewhat lower and in the range of 62-64%. However, the FDR also depends on the number of significant funds S_γ which will vary across each factor model, so the different estimates of $\hat{\pi}_0(\lambda)$ need not translate into different estimates of the FDR. The results in Table 2 indicate a large proportion of null funds in our sample - overall, around 75% of active funds yield truly zero alphas.[12] For reasons of brevity and clarity we first report detailed results for the unconditional 3F model (which is the best 'in sample' model) and use $\hat{\pi}_0(\lambda) = 75\%$. We report results for other models in an appendix.

[11] The market timing models of Treynor-Mazuy (1966) and Henriksson–Merton (1981) are not as good as the 3F and 4F models according to the Schwartz Information Criterion and are not reported here.

[12] Barras *et al*. (2009) using data on US funds (and the unconditional 4F model) also estimate $\hat{\pi}_0(\lambda)$ to be about 75% when eyeballing the histogram of p-values and using $\lambda = 0.5 - 0.6$ or when using the minimum MSE-bootstrap estimator.

Table 3

Luck and long-term performance.

This table reports the performance measures for the Fama-French 3 factor model using the whole sample period from April 1974 to December 2002. 675 funds with at least 36 monthly observations are included in this analysis. Panel A reports estimates of the proportion of zero alpha, skilled and unskilled funds. Panel B reports the false discovery rate and the proportion of significant, lucky and skilled funds for 5 different significance levels, namely 2.5%, 5%, 10%, 15% and 20%. The same statistics for the negative (left) tail of the distribution are reported in Panel C. Standard errors (Genovese and Wasserman 2004, BSW 2009) are reported in parentheses.

Panel A.: Proportion of unskilled and skilled funds

	Zero alpha ($\hat{\pi}_0$)	Non-zero alpha	Skilled ($\hat{\pi}_A^+$)	Unskilled ($\hat{\pi}_A^-$)
Proportions (%)	75.1 (2.8)	24.9	1.6 (0.5)	23.3 (2.7)
Number of funds	507	168	10	158

Panel B. Impact of luck in the positive tail

Significance level (γ)		0.025	0.05	0.10	0.15	0.20
Sign. alpha	$\hat{S}^+$ (%)	3.0 (0.7)	5.5 (0.9)	9.0 (1.1)	12.6 (1.3)	15.9 (1.4)
	FDR$^+$	30.4	32.8	39.8	42.9	45.4
Lucky	$\hat{F}^+$ (%)	0.9 (0.03)	1.8 (0.1)	3.6 (0.1)	5.4 (0.2)	7.2 (0.3)
Skilled	$\hat{T}^+$ (%)	2.1 (0.7)	3.7 (1.0)	5.4 (1.5)	7.2 (1.8)	8.7 (2.3)

Panel C.: Impact of luck in the negative tail

Sign. alpha	$\hat{S}^-$ (%)	17.9 (1.5)	23.7 (1.6)	31.0 (1.8)	36.3 (1.9)	40.4 (1.9)
	FDR$^-$	5.0	7.6	11.6	14.9	17.8
Unlucky	$\hat{F}^-$ (%)	0.9 (0.03)	1.8 (0.1)	3.6 (0.1)	5.4 (0.2)	7.2 (0.3)
Unskilled	$\hat{T}^-$ (%)	17.0 (1.5)	21.9 (1.7)	27.4 (2.1)	30.9 (2.4)	33.2 (2.7)

4.3 *Skilled and unskilled funds*

Taking our universe of all M-funds, Table 3 (Panel A) gives an estimate of the percentage of zero alpha funds $\hat{\pi}_0(\lambda) = 75\%$, the percentage of skilled funds $\hat{\pi}_A^+ = 1.6\%$ and unskilled funds $\hat{\pi}_A^- = 23.4\%$ for the whole life of each fund. These 'long-run' values are all statistically well determined. (Standard errors are in parentheses and are given in Genovese and Wasserman 2004 and Appendix-A of BSW 2009). Hence in the whole population of M-funds, most have truly zero long-run alphas, very few have positive alphas and a sizable proportion have negative alphas. We now examine whether the skilled and unskilled funds are dispersed or concentrated in the right and left tails, respectively.

The most striking feature about the performance of the best and worst funds revealed by our analysis of the unconditional 3F model is the relatively high FDR_γ^+ for the best funds and low FDR_γ^- for the worst funds – this is true for any significance level chosen (Table 3, Panels B and C). For example for $\gamma = 0.05$, only $S_\gamma^+ = 5.5\%$ (37 funds) have significant positive alphas but given that $FDR_\gamma^+ = 32.8\%$ is much higher than $FDR_\gamma^- = 7.6\%$, only $T_\gamma^+ = 3.7\%$ (25 funds) have truly positive alphas (Panel B). So, the standard approach indicates 37 funds are significant but this 'simple count' does not incorporate false discoveries, which results in only 25 truly skilled funds. Although very few skilled funds truly outperform their 3F-benchmarks these funds appear to be concentrated in the extreme right tail of the performance distribution – both S_γ^+ and FDR_γ^+ increase with γ but the percentage of truly skilled funds T_γ^+ remains small for $\gamma \leq 0.1$ (Table 3, Panel B).

The standard approach gives a relatively more accurate picture of the performance of the worst funds. For example, for $\gamma = 0.05$, the FDR_γ^- is relatively small at 7.6% so of the $S_\gamma^- = 23.7\%$ significant worst funds, $T_\gamma^- = 21.9\%$ (148 funds) are truly unskilled (Panel C). In contrast to the location of the skilled funds, the proportion of truly unskilled funds T_γ^- increases with γ, indicating that the poorly performing funds are fairly evenly spread throughout the left tail of the performance distribution in the interval $\gamma = [0, 0.2]$ (Panel C).[13]

Recursive estimates of $\hat{\pi}_A^+$ are fairly small and constant. $\hat{\pi}_A^+$ shows a slight rise from 1% at the beginning of the 1990s to 3% in 1999 before falling back to 1.6% at end-2002, after the stock market crash. Similarly, recursive estimates of $\hat{\pi}_A^-$ remain high over the 1990-2002 period at around 25% (peaking at 30% in 1996). This suggests that the *ex post* performance of winner and loser funds over short-term horizons may be similar to that over the long-term as reported in Table 3. Results (not reported) show that this is indeed the case with little difference in the statistics reported in Table 3 when we repeat the exercise for non-overlapping 'short-term' 5-year periods (see BSW 2009, Table III).

Comparing our UK results with those for the USA in BSW (2009) – who have a longer sample period ending in 2006 - they find zero skilled funds ($T_\gamma^+ \approx 0$) over the whole sample period but evidence of a small proportion of skilled funds ($T_\gamma^+ \approx 2 - 2.5\%$) over 'short-term' (non-overlapping 5-year) horizons, concentrated in the extreme right tail. On UK data we find a small proportion of both 'long-term' and 'short-term' skilled funds of around $T_\gamma^+ = 4\%$. For unskilled funds, US results in BSW (2009) and our results are very similar - both studies find strong evidence of a sizeable proportion of unskilled funds in excess of $T_\gamma^- = 15\%$, spread throughout the left tail.

4.4 *Style categories*

It is useful for investors to know if different style categories give different results for the performance of the best and worst funds after taking account of the FDR.[14] It turns out that although there are some minor differences, the broad qualitative results found when analysing all mutual funds apply to the separate style categories. For each of the three styles we find a high FDR_γ^+ for the best funds, a low FDR_γ^- for the worse funds (for all significance levels). Therefore in Table 4 we only report results for the three style categories using $\gamma = 0.05$ (full results are available on request).

The 'all companies' sector has 423 funds in total with $S_\gamma^+ = 3.8\%$ statistically significant positive alphas (Table 4, Panel A). But with an $FDR_\gamma^+ = 47.6\%$ only 2%

---

[13] These results for the unconditional 3F model are robust across our 4 different factor models and these results are reported in the appendix.

[14] In order to calculate the FDR we used our estimate $\pi_0 = 75\%$ based on all funds, as M needs to be 'large'. The results however do not change much if π_0 is estimated using only funds who belong to the specific style category.

Table 4

Fama-French 3 factor model (different fund categories).

This table reports statistics for all funds and for income funds, all companies funds and small company funds. Panel A reports the false discovery rate and the percentage (and number) of significant funds for different fund styles and different significant levels for the positive tail of the distribution. Panel B reports these statistics for the negative tail of the distribution. The model used is the Fama-French 3 factor model. The whole sample period is April 1974 to December 2002 and includes funds with at least 36 monthly observations. The estimate π_0 used is 75%.

	All Funds	Income Funds	All Companies Funds	Small Company Funds
Number of funds	675	143	423	109
# Positive alpha funds (%)	236 (35%)	76 (53.1%)	127 (30%)	33 (30.3%)
# Negative alpha funds (%)	439 (65%)	67 (46.9%)	296 (70%)	76 (69.7%)

Panel A. Calculated statistics (positive tail of distribution)

Level of significance (γ)	FDR^+ (percent)	Percentage (Number) of significant funds (S^+)	FDR^+ (percent)	Percentage (Number) of significant funds (S^+)	FDR^+ (percent)	Percentage (Number) of significant funds (S^+)	FDR^+ (percent)	Percentage (Number) of significant funds (S^+)
0.025	30.37	2.96 (20)	18.39	4.90 (7)	38.07	2.36 (10)	32.70	2.75 (3)
0.05	32.84	5.48 (37)	17.16	10.49 (15)	47.59	3.78 (16)	32.70	5.50 (6)
0.1	39.84	9.04 (61)	27.09	13.29 (19)	49.12	7.33 (31)	35.67	10.09 (11)
0.15	42.88	12.59 (85)	32.17	16.78 (24)	50.76	10.64 (45)	36.79	14.68 (16)
0.2	45.42	15.85 (107)	31.20	23.08 (33)	53.43	13.48 (57)	46.16	15.60 (17)

Panel B. Calculated statistics (negative tail of distribution)

Level of significance (γ)	FDR^- (percent)	Percentage (Number) of significant funds (S^-)	FDR^- (percent)	Percentage (Number) of significant funds (S^-)	FDR^- (percent)	Percentage (Number) of significant funds (S^-)	FDR^- (percent)	Percentage (Number) of significant funds (S^-)
0.025	5.02	17.93 (121)	21.45	4.20 (6)	4.38	20.57 (87)	3.50	25.69 (28)
0.05	7.59	23.70 (160)	23.40	7.69 (11)	6.74	26.71 (113)	5.45	33.03 (36)
0.1	11.63	30.96 (209)	32.17	11.19 (16)	10.08	35.70 (151)	9.34	38.53 (42)
0.15	14.88	36.30 (245)	30.89	17.48 (25)	13.05	41.37 (175)	13.08	41.28 (45)
0.2	17.80	40.44 (273)	34.32	20.98 (30)	15.78	45.63 (193)	15.70	45.87 (50)

(about 8 funds) truly outperform their benchmarks. In the left tail of the performance distribution $S_\gamma^- = 26.7\%$ but the low $FDR_\gamma^- = 6.7\%$ implies that 24.9% (105) of all companies funds truly underperform their benchmarks. This pattern is broadly repeated for the 109 small company funds with $FDR_\gamma^+ = 32.7\%$ and $FDR_\gamma^- = 5.5\%$ which implies that (for $\gamma = 0.05$) only 3.7% of funds have truly positive alphas while 31.2% have truly negative alphas. For equity income funds the situation is a little different because $FDR_\gamma^+ = 17.2\%$ is not too dissimilar to $FDR_\gamma^- = 23.4\%$, and the percentage of truly positive and negative-alpha funds are approximately equal ($T_\gamma^+ = 8.5\%$ and 5.9%, respectively). There are only 143 equity income funds so the number of genuine outperformers and underperformers is around 10. For investors in both the US and UK funds, use of the FDR demonstrates that there are far fewer skilled funds than would be indicated by the standard approach but unskilled funds are plentiful and easy to discover.[15]

4.5 *Persistence: controlling the overall FDR*

Instead of simply estimating the false discovery rate amongst our funds, we can instead choose a 'threshold' FDR_γ^+ (say 10%) and find a sub-set of funds which have an overall $FDR_\gamma^+ < 10\%$. Limiting funds to those with FDR_γ^+less than 10% ensures that there are relatively few lucky funds (false discoveries) included in the portfolio. Of course, increasing the target FDR_γ^+ level introduces proportionately more lucky funds into a portfolio of skilled funds and we might expect this portfolio to perform less well in the future. However, there may be an additional diversification benefit when adding lucky funds, so although alpha may fall the information ratio $IR = \alpha/\sigma_\varepsilon$ may increase –but given that most funds are well diversified this effect may not be large.

In tests of persistence standard portfolio formation uses fractile portfolios based on past performance (e.g. returns, alphas, t-alphas) regardless of the possible changing FDR_γ^+ within this set of funds – hence there is a danger that many lucky funds are included along with genuinely skilled funds. Portfolio formation based on a low FDR_γ^+ mitigates this problem and may result in a higher probability of detecting persistence, if it truly exists. For example for $\gamma = 5\%$ there are only $S_\gamma^+ = 5.5\%$ significant positive-alpha funds (amongst the population of 675 funds) located in the extreme right tail but as $FDR_\gamma^+ = 32.8\%$, then 67.2% of these *significant funds* have truly positive alphas (Table 3). Although no methodology can isolate *individual* funds that are truly significant, the FDR approach seems preferable to the usual method of only considering 'significant funds', without any allowance for false discoveries. Portfolio formation follows the approach first set out in BSW (2009). We estimate

$$FDR_\gamma^+ = \frac{F_\gamma^+}{S_\gamma^+} = \frac{\pi_0(\gamma/2)}{S_\gamma^+} \tag{10}$$

for a range of significance levels ($\gamma = 0.01, 0.02, \ldots, 0.6$) using bootstrapped t-alpha p-values of all funds estimated over the previous 5 years. The change in the estimated FDR_γ^+ as γ is altered depends on two factors. The FDR_γ^+ tends to fall as the significance level γ is reduced but as the latter also produces a smaller S_γ^+ the impact of changing

[15] Style categories in the USA are different from those in the UK but BSW (2009, Appendix Table AIII)) find that skill resides in Aggressive Growth funds whereas a sizable percentage of unskilled funds are found in all style categories.

γ on FDR^+_γ can only be determined empirically. We use (10) to find the value of γ^+ ($= 0.02$ say) which gives an estimated value for FDR^+_γ that is closest to our target value (e.g. 10%). We then include all funds in our portfolio (with equal weight) which have p-values less than γ^+. The portfolio is held for one year and then rebalanced. If an included fund 'dies' in any month over the one-year holding period, the remaining funds are equally weighted from that point, to mitigate survival bias. The first portfolio formation date is 1993 and the last is 2001. The returns from these 'forward-looking' portfolios are then used to estimate 'forward-looking' alphas.

Intuitively, portfolio formation should use a low 'desired' $FDR^+_\gamma = 10\%$ (say) for fund inclusion but in so doing we encounter a similar practical problem to BSW (2009). Because the percentage of significant positive-alpha funds S^+_γ is low, the *estimated* FDR^+_γ tends to be high even for very low values of γ. Table 5 shows for each rebalancing period, the empirical FDRs achieved and the number of funds in the portfolio for target FDR rates of 10% and 20%. For positive alpha funds (Panel A) the $FDR^+_\gamma = 10\%$ target is often exceeded and the number of funds included in the rebalanced portfolio varies between 3 and 24.[16] However, the $FDR^+_\gamma = 20\%$ target and the negative $FDR^-_\gamma = 10\%$ and 20% targets (Panel B) are achieved empirically and the negative-FDR portfolios contain a substantial number of funds. Table 5, Panel B shows the performance statistics for the forward looking positive and negative-alpha portfolios and indicates that past skilled (unskilled) funds retain their positive (negative) alphas, but both forward looking alphas are not statistically significant. There is an increase in the information ratio as the FDR target increases, as more included funds lower the specific risk of the portfolio.[17]

For the negative-alpha funds we can still achieve the $FDR^-_\gamma = 10\%$ (and 20%) targets when we include the 1980s data (even though the overall fund universe is smaller) and here we do find weak evidence of persistence for $FDR^-_\gamma = 10\%$ and 20% portfolios with alphas of -1.13% p.a. ($p = 0.045$) and -1.21% p.a. ($p = 0.018$).[18] Hence portfolio formation based on controlling the FDR gives UK investors some indication of which funds to avoid but does not provide them with positive-alpha portfolios in the future.

The number of funds included in our FDR^+_γ portfolio varies over time but is generally small and each included fund has a relatively low p-value (usually below 0.014). Hence controlling for FDR^+_γ results in much smaller portfolios than the standard decile method and is tantamount to including a small set of funds with very low p-values, based on the previous 5 years of data. However, it is probably worth noting that a successful investment strategy does not necessarily require a fund to exhibit persistence. A group of funds can be successful and produce positive long-run alphas if they *trade infrequently* when they have genuine superior information, while earning zero alphas at other times.

[16] BSW (2009, Table V) find that attempting to hit a target $FDR^+_\gamma = 10\%$ results in portfolios that for 6 years out of their 27 portfolio rebalancing periods has a minimum estimated FDR^+_γ of over 70%, for a further 6 years the FDR^+_γ is between 30% and 60% and for the remaining 14 years is between 10% and 30%.

[17] For US funds BSW (2009, Table V) find that as the FDR^+_γ target level is increased from 10% to 50%, the information ratio falls from 0.36 to 0.33.

[18] BSW (2009) for US equity funds find evidence of positive persistence for $FDR^+_\gamma = 10\%$ and $FDR^+_\gamma = 30\%$ portfolios of 1.45% p.a. ($p = 0.04$) and 1.15% p.a. ($p = 0.05$) – they do not report results on negative persistence.

Table 5

Performance persistence and FDR.

Funds selected for inclusion in the portfolio are based on a desired target FDR rate, estimated using 60 months of data prior to the portfolio formation date. The portfolio is held for one year before being rebalanced. Holding period returns are concatenated to give a time series of "forward-looking" returns which are then used to estimate forward-looking alphas. Portfolios are formed based on alternative target FDR rates of 10 and 20 percent (for both the best and worst performing funds). The first portfolio formation date is January 1989 to December 1993 which determines the funds held in 1994. Panel A reports the estimated FDRs for each portfolio formation date, which are closest to the target FDRs of 10% and 20%, together with the number of funds which are selected for the portfolios. Panel B reports estimates of a Fama-French 3 factor model including the forward-looking alpha - the estimation period is 9 years, from January 1994 to December 2002. We report the estimated parameters and their bootstrap p-values (based on 1,000 bootstraps) together with the Sharpe ratio and information ratio.

Panel A.: Target FDR

	Positive alpha funds				Negative alpha funds			
	Target $FDR^+ = 10\%$		Target $FDR^+ = 20\%$		Target $FDR^- = 10\%$		Target $FDR^- = 20\%$	
Forecasting Year	Empirical FDR+	Number of Funds	Empirical FDR+	Number of Funds	Empirical FDR–	Number of Funds	Empirical FDR–	Number of Funds
1994	17.9	4	17.9	4	9.9	181	19.9	267
1995	25.1	3	25.1	3	9.9	183	19.6	292
1996	9.7	8	19.5	16	9.7	136	19.8	275
1997	11.9	21	19.2	26	9.8	127	19.8	236
1998	7.6	24	19.4	71	9.7	66	19.8	157
1999	9.0	11	18.5	16	9.9	30	19.6	181
2000	8.6	12	19.7	21	9.6	86	19.8	219
2001	23.6	9	23.6	9	10.6	10	19.5	98
2002	19.6	4	19.6	4	9.8	24	19.8	119

Panel B. Performance analysis

	Positive alpha funds				Negative alpha funds			
	Target $FDR^{+} = 10\%$		Target $FDR^{+} = 20\%$		Target $FDR^{-} = 10\%$		Target $FDR^{-} = 20\%$	
	estimate	p-value	estimate	p-value	estimate	p-value	estimate	p-value
alpha	0.1312	0.183	0.1166	0.234	−0.0565	0.246	−0.0626	0.190
β_1 $(R_m - r_f)$	0.9206	<0.001	0.9330	<0.001	0.9266	<0.001	0.9198	<0.001
β_2 SMB	0.4917	<0.001	0.5036	<0.001	0.1691	<0.001	0.1602	<0.001
β_3 HML	−0.1260	0.003	−0.1630	<0.001	−0.0140	0.282	−0.0147	0.241
Mean Return (%p.m.)	0.1893		0.1757		−0.0472		−0.0546	
SD Return (%p.m.)	4.2703		4.3811		3.8728		3.8291	
Sharpe Ratio	0.0443		0.0401		−0.0122		−0.0143	
Information Ratio	0.0885		0.0727		−0.0727		−0.0889	

5. Conclusions

We use a multiple hypothesis testing framework to estimate the proportion of truly zero-alpha funds in the universe of UK funds and also estimate the false discovery rate (FDR) amongst the best and worst UK equity mutual funds. The majority (around 75%) of UK mutual funds neither underperform nor outperform their benchmarks. At 2.5% and 5% significance levels, using all funds (and the unconditional 3F model) the standard approach gives the proportion of significant positive alpha funds as 3% (20 funds) and 5.5% (37 funds), respectively. But we find a relatively high FDR_{γ}^{+} for these funds of 30.4% and 32.8% which indicates that only around $T_{\gamma}^{+} = 2.0\%$ to 3.7% of funds (i.e. around 13 to 25 funds out of 675 funds) exhibit long-run skill over the life of each fund – this is broadly consistent with the predictions of the Berk and Green (2004) model.

For the worst funds, at a 2.5% (5%) significance level, the FDR_{γ}^{-} is relatively small at 5.0% (7.6%). Hence the proportion of funds that are unskilled is fairly substantial with $T_{\gamma}^{-} = 17\%$ (21.9%) of all funds and this does not differ greatly from the standard approach, which gives 17.9% (23.7%) as unskilled. In addition, skilled funds tend to be concentrated in the extreme right tail of the performance distribution while the unskilled funds are dispersed throughout the left tail. It appears that funds which perform badly over a long time period do not go out of business or revert to an 'index tracker' strategy, as predicted in the Berk and Green (2004) model. Broadly similar results are found for the long-run performance of US equity funds by BSW (2009).

When we examine different investment styles the general pattern of very few genuine skilled funds is repeated for all companies, small company and equity income funds. Also, there are a substantial proportion of funds in the all companies and small company sectors that are truly unskilled but only a small proportion of equity income funds that are unskilled.

When we form *ex ante* portfolios based on a maximum FDR_{γ}^{+} of 10% and 20% these positive alpha funds do not exhibit persistence, as their forward looking alphas are not statistically different from zero. Portfolios of historically poorly performing funds formed after conditioning on a maximum FDR_{γ}^{-} show some negative persistence in the 1980s but this disappears in the 1990s.

Appendix: FDR for Different Factor Models

Table A1

False discoveries: different factor models ($\gamma = 0.05$).

This table provides a sensitivity analysis when applying different factor models to all funds in our sample. Results are reported for a significance level of $\gamma = 0.05$ and $\hat{\pi}_0$ is estimated using the bootstrap method. For each model we report the estimated FDR, the number of significant funds, the number of false discoveries and the number of truly significant funds, for both tails of the performance distribution.

Panel A. Carhart 4F ($\pi_0 = 0.6438$)

	False Discovery Rate, FDR(%)	# of significant funds, S	# of false discoveries, F	# of truly significant funds, T
Best Funds	$FDR^+ = 36.2\%$	30	10.86	19.14
Worst Funds	$FDR^- = 9.6\%$	113	10.86	102.14

Panel B. Conditional beta 3F ($\pi_0 = 0.7437$)

	False Discovery Rate, FDR(%)	# of significant funds, S	# of false discoveries, F	# of truly significant funds, T
Best Funds	$FDR^+ = 44.8\%$	28	12.55	15.45
Worst Funds	$FDR^- = 13.4\%$	94	12.55	81.45

Panel C. Conditional alpha-beta 3F ($\pi_0 = 0.7259$)

	False Discovery Rate, FDR(%)	# of significant funds, S	# of false discoveries, F	# of truly significant funds, T
Best Funds	$FDR^+ = 42.2\%$	29	12.25	16.75
Worst Funds	$FDR^- = 15.5\%$	79	12.25	66.75

References

Admati, A.R., Bhattacharya, S., Ross, S.A. and Pfleiderer, P., ‘On timing and selectivity’, *Journal of Finance*, Vol. 41, 1986, pp. 715–730.

Bajgrowicz, P. and Scaillet, O., ‘Technical trading rules revisited: persistence tests, transaction costs and false discoveries’, *Working Paper* (HEC Geneva, 2008).

Barras, L., Scaillet, O. and Wermers, R., ‘False discoveries in mutual fund performance: measuring luck in estimated alphas,’ *Journal of Finance* forthcoming 2009.

Benjamini, Y. and Hochberg, Y., ‘Controlling the false discovery rate: a practical and powerful approach to multiple testing,’ *Journal of Royal Statistical Society*, Vol. 57, 1, 1995, pp. 289–300.

Berk, J. B. and Green, R.C., ‘Mutual fund flows and performance in rational markets’, *Journal of Political Economy*, Vol. 112, 2004, pp. 1269–95.

Blake, D. and Timmermann, A., ‘Mutual fund performance: evidence from the UK’, *European Finance Review*, Vol. 2, 1998, pp. 57–77.

Carhart, M. M., ‘On persistence in mutual fund performance,’ *Journal of Finance*, Vol. 52, 1997, pp. 57–82.

Christopherson, J. A., Ferson, W.E. and Glassman, D.A., ‘Conditioning manager alphas on economic information: another look at the persistence of performance’, *Review of Financial Studies*, Vol. 11, 1998, pp. 111–42.

Criton, G. and Scaillet, O., 'Time-varying coefficient model for hedge funds', *Working Paper* (SSRN. March, 2009).

Cuthbertson, K., Nitzsche, D. and O'Sullivan, N., 'Performance of UK mutual funds: luck or skill?' *Journal of Empirical Finance*, Vol. 15(4), 2008, pp. 613–34.

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

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

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

Ferson, W. E. and Schadt, R.W., 'Measuring fund strategy and performance in changing economic conditions', *Journal of Finance*, Vol. 51, 1996, pp. 425–62.

Fletcher, J., 'An examination of the selectivity and market timing performance of UK unit trusts', *Journal of Business Finance and Accounting*, Vol. 22, 1995, pp. 143–56.

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

Genovese, C. and Wasserman, L., 'A stochastic process approach to false discovery control', *Annals of Statistics*, Vol. 32, 2004, pp. 1035–61.

Henriksson, R. and Merton, R.C., 'On market timing and investment performance: statistical procedures for evaluating forecasting skills', *Journal of Business*, Vol. 54, 1981, pp. 513–33.

Holm, S., 'A simple sequentially rejective multiple test procedure', *Scandinavian Journal of Statistics*, Vol. 6, 1979, pp. 65–70.

Kosowski, R., Timmermann, A., White, H. and Wermers, R., 'Can mutual fund 'stars' really pick stocks? New evidence from a bootstrap analysis', *Journal of Finance*, Vol. LXI (6), 2006, pp. 2551–95.

Leger, L., 'UK investment trusts: performance, timing and selectivity', *Applied Economics Letters*, Vol. 4, 1997, pp. 207–10.

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

McCracken, M. W. and Sapp, S.G., 'Evaluating the predictabililty of exchange rates using long-horizon regressions: mind your p's and q's', *Journal of Money Credit and Banking*, Vol. 37(3), 2005, pp. 473–94.

Newey, W. D. and West, K.D., 'A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix', *Econometrica*, Vol. 55, 1987, pp. 703–08.

Otten, R. and Bams, D., 'European mutual fund performance', *European Financial Management*, Vol. 8, 2002, pp. 75–101.

Otten, R. and Bams, D., 'The performance of local versus foreign mutual fund managers', *European Financial Management*, Vol. 13, 2007, pp. 702–20.

Politis, D.N. and Romano, J.P., 'The stationary bootstrap', *Journal of the American Statistical Association*, Vol. 89, 1994, pp. 1303–13.

Quigley, G. and Sinquefield, R.A., 'Performance of UK equity unit trusts', *Journal of Asset Management*, Vol. 1, 2000, pp. 72–92.

Storey, J. D., 'A direct approach to false discovery rates', *Journal of Royal Statistical Society B*, Vol. 64, 2002, pp. 497–98.

Storey, J. D., Taylor, J.E. and Siegmund, D., 'Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach', *Journal of Royal Statistical Society*, Vol. 66, 2004, pp. 187–205.

Tonks, I., 'Performance persistence of pension fund managers', *Journal of Business*, Vol. 78, 2005, pp. 1917–42.

Treynor, J. and Mazuy, K., 'Can mutual funds outguess the market', *Harvard Business Review*, Vol. 44, 1966, pp. 66–86.

Wermers, R., 'Mutual fund performance : an empirical decomposition into stock picking talent, style, transactions costs, and expenses', *Journal of Finance*, Vol. 55, 2000, pp. 1655–1703.