Cuthbertson, K. & Nitzsche, D. (2013). Performance, Stock Selection and Market Timing of the German Equity Mutual Fund Industry. Journal of Empirical Finance, 21, pp. 86-101. doi: 10.1016/j.jempfin.2012.12.002

Performance, Stock Selection and Market Timing of the German Equity Mutual Fund Industry

Keith Cuthbertson* **and Dirk Nitzsche***

This Version: 29th August 2012

Abstract:

We investigate the performance of the German equity mutual fund industry over 20 years (monthly data 1990-2009) using the false discovery rate (FDR) to examine both model selection and performance measurement. When using the Fama-French three factor (3F) model (with no market timing) we find at most 0.5% of funds have truly positive alpha-performance and about 27% have truly negative-alpha performance. However, use of the FDR in model selection implies inclusion of market timing variables and this results in a large increase in truly positive alpha funds. However, when we use a measure of "total" performance, which includes the contribution of both security selection (alpha) and market timing, we obtain results similar to the 3F model. These results are largely invariant to different sample periods, alternative factor models and to the performance of funds investing in German and non-German firms – the latter casts doubt on the 'home-bias' hypothesis of superior performance in 'local' markets.

Keyword : Mutual fund performance, false discovery rate.
JEL Classification : C15, G11, C14

* Cass Business School, City University, London, UK

Corresponding Author:
Professor Keith Cuthbertson, Cass Business School, 106 Bunhill Row, London, EC1Y 8TZ.
Tel. : +44-(0)-20-7040-5070,
Fax : +44-(0)-20-7040-8881,
E-mail : K.Cuthbertson@city.ac.uk

We thank 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 and the FMA/EDHEC conference on fund performance, for discussions and comments.

Performance, Stock Selection and Market Timing of the German Equity Mutual Fund Industry

1. Introduction

This paper addresses a key generic issue namely, how to take account of false discoveries in empirical work. This problem arises in many different areas, in fact whenever we ask the question: "How many of our statistically significant results are likely to be "truly null" – that is "false discoveries". There are a number of possible approaches to multiple hypothesis testing which attempt to isolate truly null "entities" from the set of "statistically significant" entities - this includes the Bonferroni test and the Family Wise Error Rate. In this paper we use the False Discovery Rate (FDR) which measures the proportion of lucky funds among a group of funds, whose "performance" has been found to be statistically significant.

There have been no previous studies that use the FDR in model selection. Here we apply the FDR to assess the prevalence of market timing and to investigate the joint contribution to fund performance of security selection (alpha) and market timing. We therefore provide some additional methodological applications of the FDR. In particular we apply these techniques to investigate the performance of the German equity mutual fund industry.

There has been little work done on analysing the performance of the German mutual fund industry despite its substantial growth over the last 15 to 20 years. Although the German mutual fund industry is small compared to the US, its assets under management peaked in 2007 at $372bn and fell to $237bn at end of 2008. However, it is expected that the German mutual fund industry will become more important in future years as reforms to private pension provision place greater emphasis on defined contribution pensions (i.e. 'Riester Rente') and reforms result in a less generous state pension.

We assess the overall success of the German mutual fund *industry* over the period 1990-2009 using a monthly data set, free of survivorship bias. I If we simply count the number of funds which are found to have a "statistically significant" performance measure, we run the risk of including funds which are truly null (i.e. Type I errors). For example, suppose the FDR amongst 20 statistically significant best/winner funds (e.g. those with positive alphas) is 80%, then this implies that only 4 funds (out of the 20) have truly significant alphas[1] - this is clearly useful information for investors. A key issue is whether this correction gives different inferences from the standard approach of simply "counting" the number of significant funds with non-zero abnormal performance.

As robustness tests we also examine performance over different time periods, different factor models (including market timing models) as well as the performance of domiciled German funds which invest in Germany and outside Germany – the latter provides evidence on the 'home-bias' issue (Coval and Moskowitz 1999, Hong, Kubik and Stein 2005). 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. Part of the explanation for this may be the "dilution effect" whereby funds experience an increase in investor cash flows during periods when the *market* return is relatively high hence increasing the fund's cash position, leading to a concurrent lower overall portfolio return (Warther 1995, Edelen and Warner 2001, Bollen and Busse 2001, Bessler, Blake, Luckoff and Tonks 2010).

From a methodological perspective we also show how the FDR can be used in aiding model selection in an area where parametric tests of fund performance (e.g. alpha) suffer from low power and potential bias (Lehmann and Timmermann 2007). For example, in factor models we usually include variables based solely on their statistical significance - but this ignores possible false discoveries. We show how the FDR, informs our choice of the appropriate performance model. We also use this approach in assessing the dual activities of "security selection" (i.e. fund's alpha) and "market timing" - a distinction referred to as "performance attribution" in the literature. Clearly it is possible for a fund to simultaneously pursue both security section and market timing and previous studies have attempted to independently measure these two effects (e.g. Admati et al 1986). We argue from a theoretical perspective that the conditions required to successfully isolate performance attribution are unlikely to be met. Rather than use alpha as our performance measure we use an alternative which combines both the fund's alpha and the contribution of market timing to fund returns. We then adapt the FDR approach to infer the importance of this "total performance" measure for the mutual fund industry as a whole. , Funds in the tails of the cross-section performance distribution are often found to have non-normal specific risk (Kosowski et al 2006, Fama and French 2010, Cuthbertson et al 2008) and hence we use a variety of bootstrap procedures in all hypothesis tests, including those that use the FDR.

---

[1] 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

The US and UK mutual fund industries have been extensively analyzed and although the German fund market is smaller, our sample of around 550 equity funds provides a large comprehensive independent data set, which with the use of the FDR, mitigates possible claims of data snooping bias if results are primarily based on UK and US data.

We find that around 80% of German equity funds neither statistically beat nor are inferior to their benchmarks and therefore appear to do no better than merely tracking their style indexes. Next, there is a much higher proportion of false discoveries among the best funds than amongst the worst performing 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". For example, from amongst all 555 funds the number of significantly positive alpha funds (at a 10% significance level) is 26 (4.7% of all funds) but the estimated FDR is around 80% implying that only around 3 funds (0.5%) have truly positive alphas and these skilled funds are concentrated in the extreme right tail of the performance distribution. This is consistent with the competitive model of Berk and Green (2004).

For negative alpha funds, around 175 are statistically significant (at a 10% significance level) and with an estimated FDR of about 13% the number of truly unskilled funds is around 150 – hence a substantial 27% of all funds are genuinely poor performers. The latter result is *not* consistent with the predictions of the Berk and Green (2004) model or the model of Lynch and Musto (2003) whereby cash outflows from poorly performing funds lead to a "change of strategy" and subsequent higher returns.

When market timing is present and the FDR is used, we are able to explain previous conflicting results on "performance". Use of the FDR indicates a substantial proportion of funds with truly non-zero market timing effects – implying these variables should be included in factor models. Also, after applying the FDR to the funds' alphas in our market timing models, we find a substantial increase in the number of truly positive alphas (compared to the 3F model without timing variables). So our "market timing models" indicate substantial skill in "security selection." However, when we assess "total performance" from both security selection and market timing, we again find a very high FDR amongst the best performing funds and the number of truly "successful" funds is near zero. Hence when market timing models are subject to a "total performance" measure and the FDR is applied, we obtain performance results for winner funds similar to those in the 3F model. Without simultaneously accounting for these two effects and applying the FDR, previous studies may overstate the number of truly outperforming funds.

---

to be zero, is still subject to estimation error. Also note that the FDR says nothing about the statistical significance of the

In terms of robustness, the above results on "performance" are qualitatively and quantitatively similar over different 5-year periods, for investment in different geographical regions and across different factor and market timing models.

The rest of the paper is organized 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 look at performance models, in section 4 we present our empirical results and section 5 concludes.

2. The False Discovery Rate, FDR

The standard approach to determining whether the alpha of a single 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 γ = 5% when testing the alphas for each of M-funds, the probability of finding *at least one* non-zero alpha-fund in 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 γ / 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

---

alpha of any particular *individual* fund - conceptually, the FDR only applies to a group of significant funds.

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- γ)^M = z_M , which for a relatively small number of M=50 funds and conventional γ = 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.

null (Benjamini and Hochberg 1995, Storey 2002 and Storey, Taylor and Siegmund 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'. Note that the FDR approach can be used to assess any hypothesis test across all funds and we extend its use in the mutual fund area to provide an indicative tool to assess alternative factor models, market timing effects and alternative performance statistics.

Storey (2002) and Barras, Scaillet and Wermers (BSW 2010) provide a detailed account of the FDR methodology, so we shall be brief. Suppose the null hypothesis is that fund-i has no skill in security selection (alpha), the alternative being that the fund delivers either positive or negative performance :

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

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 $\gamma / 2$ $(0 < \gamma \le 1)$. 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 :

[1] $$E(F_\gamma^+) = \pi_0 \ (\gamma / 2)$$

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

[2] $$E(T_\gamma^+) = E(S_\gamma^+) - E(F_\gamma^+) = E(S_\gamma^+) - \pi_0(\gamma / 2)$$

(Similar formulae apply for negative-alpha funds). Choosing different levels for γ allows us to see if the number of truly skillful 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 for investors 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:

[3] $\pi_A^+ = T_{\gamma^*}^+$ $\pi_A^- = T_{\gamma^*}^-$

where γ^* is a sufficiently high significance level which can be determined using a mean squared error criterion, although setting γ^* = 0.35-0.45 produces similar results (BSW 2010). The expected FDR amongst the *statistically significant* positive-alpha funds is:

[4] $$FDR_\gamma^+ = \frac{E(F_\gamma^+)}{E(S_\gamma^+)} = \frac{\pi_0(\gamma/2)}{E(S_\gamma^+)}$$

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

[5] $$E(T_\gamma^+) / E(S_\gamma^+) = 1 - FDR_\gamma^+$$

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:

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

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 2. For example if π_0 = 100% and we choose λ = 0.6 then $W(\lambda)/M$ = 40% of p-values lie to the right of λ = 0.6 and our estimate of π_0 = 40%/ (1-0.6) = 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 estimation). We can exploit the bias-variance trade-off and choose λ to minimize the mean-square error $E\{\pi_0(\lambda)-\pi_0\}^2$ - this we refer to as the MSE-bootstrap method of estimating π_0 (Storey 2002, BSW 2010)[4].

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, Timmermann, White and Wermers, 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, …, 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)[5]. First, estimate the chosen factor model for each fund and save the residuals $\hat{e}_{i,t}$. 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 $\tilde{e}_{i,t}$, together with $\hat{\beta}_i{}' X_t$, to generate a simulated excess return series $\tilde{r}_{i,t}$ under the null hypothesis (α_i = 0). Then, using $\tilde{r}_{i,t}$ the performance model is estimated and the resulting t-statistic for performance measure, 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:

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

---

[4] BSW (2010) 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).

where $I(.)$ is a (1,0) indicator variable. An analogous procedure is used for other simple hypothesis tests and joint hypothesis tests on several parameters[6].

3. Performance Models

Our alternative performance models are well known 'factor models' and therefore we only describe these briefly. Unconditional models have factor loadings that are time invariant and the Fama and French (1993) 3F-model is:

$$[8] \qquad r_{i,t} = \alpha_i + \beta_{1i} r_{m,t} + \beta_{2i} SMB_t + \beta_{3i} HML_t + \varepsilon_{i,t}$$

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 and HML_t are size and book-to-market factors.

Market timing

Market timing in the one-factor Treynor and Mazuy (TM, 1966) model has a time varying market beta which depends linearly on the market return:

$$r_t = \alpha_i + \beta_t r_{m,t} + e_t \qquad \text{with} \qquad \beta_t = \beta_0 + \delta r_{m,t} + v_t$$

which results in the TM estimation equation:

$$[9] \qquad r_t = \alpha_i + \beta_0 r_{m,t} + \delta f[r_{m,t}] + \varepsilon_t \qquad \text{where} \qquad f[r_{m,t}] = r_{m,t}^2$$

The Hendricksson-Merton (HM, 1981) model assumes the market beta depends on the directional response of the market:

---

[5] Alternative bootstrapping procedures such as simultaneously bootstrapping the residuals and the X_t variables, or allowing for serial correlation (block bootstrap) or contemporaneous bootstrap across all (existing) funds at time t, produced qualitatively similar results, hence we only report results for the 'residuals only' bootstrap.

[6] The FDR 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 assess the performance of alternative forecasting rules in foreign exchange (McCracken and Sapp 2005), stock returns (Bajgrowicz and Scaillett 2008), hedge funds (Criton and Scaillet 2009) and to analyze US equity mutual fund performance (Barras, Scaillet and Wermers, BSW 2010).

$$\beta_t = \beta_0 + \delta(I_t^+) + v_t$$

where I_t^+ = 1 when $r_{m,t} > 0$ and zero otherwise, which results in the HM estimation equation:

[10] $$r_t = \alpha_i + \beta_0 r_{m,t} + \delta f[r_{m,t}] + \varepsilon_t \qquad \text{where} \quad f[r_{m,t}] = I_t r_{m,t}$$

The above two models are easily extended to include linear additive "other factors" such as SMB and HML[7]. If $\delta > 0\ (\delta < 0)$ this indicates successful (unsuccessful) market timing and security selection is given by $\alpha \neq 0$. Separating out these two effects is known as *performance attribution.*

It is possible to have a non-linear relationship between fund returns and the market return for reasons other than market timing. Spurious timing effects can arise if funds hold stocks that are more or less option-like than the average stock in the market index (Jagannathan and Korajczyk 1986). Also,"interim trading" can lead to $\delta \neq 0$ in TM and HM specifications and hence to spurious market timing. If funds trade each period but returns are only observed (say) every two periods then the estimated TM timing coefficient will be positive (negative) even though there is no market timing skill (Ferson and Khang 2002Goetzmann et al (2000) demonstrate another "interim trading" effect whereby the TM and HM timing coefficients δ are biased downwards if funds *successfully* time the market over a series of single periods (that is beta today depends on market returns tomorrow) but returns are measured over two (or more periods). This results in an errors in variables problem with the resultant usual downward bias when applying OLS.

Biases in estimating selectively (alpha) and market timing when the HM (TM) model is true but the TM (HM) model is estimated, are also possible. However Coles et al (2006) show that although these individual biases are large, they are almost offsetting and they suggest using a measure of "total performance", when market timing is present. We use the Bollen and Busse (2004) measure of total performance[8].

$$perf_i = (1/T)\sum_{t=1}^{T}(\alpha_i + \delta_i f[r_{m,t}]) = \alpha_i + \delta_i \overline{f[r_{m,t}]}$$

---

7 We do not consider market timing of factors other than the market return.

The $perf_i$ statistic tests the ability of a mutual fund to simultaneously provide stock selection and market timing skills. Different funds may focus on either of these elements of performance or may switch strategies through time, but $perf$ provides a useful summary statistic to measure "total performance" from these two skills. We assess $H_0 : perf_i = 0$ for each fund by bootstrapping under the null using a joint hypothesis test on (α_i, δ_i) - we then use the FDR to inform our view of the validity of $H_0 : perf_i = 0$ for the whole of the mutual fund *industry.*

Previous Studies

The literature on US fund performance is voluminous with less work being done on UK funds – and most studies examine funds which invest domestically. It is well documented that the *average* US or UK equity mutual fund underperforms its benchmarks (Elton, Gruber, Das and Hlavka 1993, Wermers 2000, Fletcher 1997, Blake and Timmermann 1998, Quigley and Sinquefield 2000). However, the cross-section standard deviation of alphas for individual funds in both the UK and US is high, and some studies do find a few funds with statistically significant positive alphas and many more with negative alphas (Malkiel 1995, Kosowski et al 2006, Fama and French 2010, Cuthbertson, Nitzsche and O'Sullivan 2008).

Studies which investigate possible *sources* of skillful and unskillful funds are almost exclusively based on US data. Past winner funds attract additional fund flows (Ivkovic and Weisbenner 2009, Del Guercio and Tkac 2008, Keswani and Stolin 2008) and this may lead to diseconomies of scale (Chen et al 2004, Yan 2008), dilution effects (Edelen 1999), distorted trading decisions (Alexander and Cici 2007, Coval and Stafford 2007, Plooet and Wilson 2008) or manager changes (Khorana 1996, 2001, Bessler, Blake, Luckoff, and Tonks 2010) - which in turn may affect future performance of winner funds. Poorly performing funds are subject to "external governance" (fund outflows) and "internal governance" (manager changes) which also influence their future performance (Dangl and Zecher 2008, Bessler, Blake, Luckoff and Tonks 2010)[9].

---

[8] Note that Coles et al (2006) use a different measure of total performance than Bollen and Busse (2004). They also show that model misspecification (i.e. TM is true but you estimate HM or vice-versa) does not appreciably alter the power to detect security selection or market timing – it only affects the bias.

[9] Studies of funds which invest internationally generally also find very few positive alpha funds and a substantial number of funds with negative alphas (see for example, Gallo and Swanson 1996 and Patro 2001 for the US and Fletcher and Marshall 2005 for the UK).

Most US and UK studies using the TM and HM models find some evidence of positive market timing and somewhat stronger evidence of negative market timing, for the mutual fund industry as a whole. However, non of the US or UK studies on market timing, appear to correct this "count" of statistically significant timing effects for potential false discoveries.[10]

Studies investigating the performance of the Germany mutual fund industry are rather sparse. Griese and Kempf (2002) using 105 German funds (1980-2000) find no positive abnormal performance while Otten and Bams (2002) analyse the performance of 4 portfolios of German equity funds and find predominantly negative and statistically insignificant alphas. Bessler, Drobetz and Zimmermann (2009) use unconditional and conditional CAPM, 3 factor Fama-French model and an SDF model on 50 German domestic equity funds and find underperformance. None of these studies examines market timing or the possibility of false discoveries.

In this paper we analyse 555 *individual* German funds which invest both domestically and internationally, we assess performance and market timing effects and measure the overall performance of the fund industry. We therefore considerably expand our knowledge of the German fund industry – taking account of possible false discoveries.

4. Empirical Results

In this study we use a comprehensive, monthly data set (free of survivorship bias) over 20 years (January 1990 to December 2009) for 555 German domiciled equity mutual funds (each with more than 24 monthly observations)[11]. We have removed 'second units' and index/tracker funds leaving only actively managed funds. Of the 555 funds which at least existed for 2 years 85 invest solely in German equities, with the remainder investing outside Germany ("Europe" and "Global"). 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). Our factors are measured in the standard way. For funds with German, European and Global geographic mandates we have used the appropriate MSCI

[10] For the US see for example, Treynor and Mazuy 1966, Henriksson and Merton 1981, Hendriksson 1984, Lee and Rahman 1990, Ferson and Schadt 1996, Busse 1999, Becker, Ferson, Myers and Schill 1999, Wermers 2000, Bollen and Busse 2001, Jiang 2003, Swinkels and Tjong-A-Tjoe 2007, Jiang, Yao and Yu, 2007, Chen and Liang 2007 and for the UK see Chen, Lee, Rahman and Chan 1992, Fletcher 1995, Leger 1997, Byrne, Fletcher and Ntozi 2006, Cuthbertson et al 2010.

[11] The complete data set is obtained from Bloomberg and consists of over 1000 funds, was reduced to just 702 after stripping out second units and to 555 funds with at least 2 years of data history.

total return indices[12]. The SMB variables have been calculated by subtracting the total return index of the small cap MSCI index from the relevant market index for the specific geographic mandate. Similarly, HML is defined as the difference between the total return indices of the MSCI Value index less the MSCI growth index for the specific geographic region[13]. The risk free rate is the 1-month Frankfurt money market rate. All variables are measured in Euros (or German Marks prior to the introduction of the single currency in Europe).

We first provide a brief overview of alternative factor models before refining these results using the FDR. Table 1 reports summary statistics for the three different models, the one-factor CAPM model, the two factor model which includes the SMB factor and the Fama and French 3-factor model, which adds the HML factor[14]. The 3F model is then augmented with either the TM or HM market timing variables. For each model, cross-sectional (across funds) average statistics are calculated for all funds over the period January 1990-December 2009 based on 555 funds, all with a minimum of $T_{i,\min}$ = 24 observations.

[Table 1 - here]

The factor models give a similar but small number of positive and statistically significant alphas and a much larger number of statistically significant negative alphas (Table 1, Panel A). The market return is highly significant followed by the SMB factor, while the HML factor and the market timing variables are not statistically significant on average. However, we note a relatively large increase in the number of statistically significant positive alphas (from around 7 to 35) and a reduction in the number of statistically significant negative alphas (from around 75 to 50) when the market timing variables are included – the market timing specification changes our view of the alpha-performance of the industry and below, this is examined further using the FDR.

[Figure 1 - here]

The distribution of alpha estimates for the 3F model (figure 1) shows a wide range of values. Most alphas are in the minus to plus 1% p.a. range but there are funds with very high and (especially) very low alphas. This implies that the extreme tails of the distribution may contain funds with abnormally "good" or "bad" security selection. This is important, since investors are

---

12 These geographical mandates should largely be followed by funds, whereas style mandates (e.g. aggressive growth, income, balanced etc.) often result in style drift (Cooper, Gulen, and Rau 2005).

13 Use of the MSCI indices allows consistency across factor definitions for "German", "European" and "Global" mandates. Worldscope has greater coverage for our factors but only for "German funds".

14 We found no evidence for the inclusion of conditioning variables such as the one-month yield, the dividend yield of the market factor and the term spread (Ferson and Schadt 1996, Christopherson, Ferson and Glassman 1998).

more interested in holding funds in the right tail of the performance distribution and avoiding those in the extreme left tail, than they are in the average fund's performance. This emphasizes the importance of examining fund-by-fund performance (rather than the weighted average of all funds) and then correcting for false discoveries to provide an assessment of overall industry performance[15].

Turning now to diagnostics (bottom half of table 1), the adjusted-R^2 across all three models is around 0.75, while the average skewness and kurtosis of the residuals is around -0.2 and 8 respectively and about 45% of funds have non-normal errors – thus motivating the use of bootstrap procedures.

How Important are the Individual Factors?

We know from table 1 that without taking account of the FDR, the market factor and the SMB factor appear to be statistically significant across many of the 555 funds, whereas the average t-statistic (absolute value) across all funds for the HML factor is around -0.85. Table 2 re-examines these results when we take account of possible false discoveries[16]. Around 545 funds have statistically significant positive market betas with a FDR less than 0.1% (at 10% significance level), so not surprisingly nearly all funds have truly positive market betas (Panel A, Table 1). For the SMB factor around 420 funds are significantly positive and the FDR is very low at 1.6% while for the 17 funds with negative and statistically significant SMB-betas the FDR of 38% implies over 60% of these are truly significant. Overall therefore it appears as if most funds truly have positive weighting on small stocks and as this strategy is replicable, its contribution to fund returns should not be counted as skill.

[Table 2 - here]

In contrast to the rather weak results based on the average (absolute) values of the HML-beta and its t-statistic (table 1) the number of significant positive HML-betas (10% significance level) is 103 (FDR_{γ}^{+} = 11.7%), with 247 (FDR_{γ}^{-} = 4.9%) having significant negative betas (table 2, Panel A) – hence many more German funds are "growth orientated" rather than value orientated. Use of the FDR to provide an indicative measure of the overall importance of these three factors, suggests all three factors should be included in our factor model. Hence, we

---

15 The same wide range for the distribution of fund alphas is found for the two 3F *plus market timing* models. In addition the residuals of funds in the extreme tails of the cross-section distribution of the 3F and 3F plus market timing models are non-normal, hence motivating the use of bootstrap standard errors.

16 Estimation of the FDR when interpreting tests on the factor betas requires an estimate of π_0 (the proportion of truly null betas across all funds). The method of estimation for π_0 is discussed below.

concentrate on results from the 3F model and the two, 3F *plus market timing* models (3F+TM and 3F+HM).

We now proceed as follows. First we discuss estimation of the proportion of truly zero-alpha funds π_0, positively skilled alpha-funds, π_A^+ and unskilled funds π_A^- among our total of M-funds. Then we analyze the FDR for the positive-alpha and negative-alpha funds taken separately – this allows us to ascertain whether such funds are concentrated in the tails of the performance distribution. Next we use the FDR to examine performance attribution – that is, the importance of market timing and security selection in the mutual fund industry. This analysis is extended to measure "total performance" using the FDR approach. Finally we present some robustness tests by examining performance across different factor models, across non-overlapping 5-year periods and performance for fund investments both within and outside Germany. Finally, we examine the sensitivity of the proportion of skilled and unskilled funds across the different factor models used in our analysis.

Estimation of π_0

The histogram of p-values when testing $H_0 : \alpha_i = 0$ across funds is given in figure 2 for the 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 2 a reasonable "eyeball" estimate would be λ = 0.3 giving $\hat{\pi}_0(\lambda)$ = 0.8.

[Figure 2 here]

Security Selection: Skilled and Unskilled Funds

Taking the 3F model and our universe of all M-funds, the MSE-bootstrap estimator gives the percentage of truly zero alpha funds $\hat{\pi}_0(\lambda)$ = 83% (se = 3.24), the percentage of negative-alpha funds $\hat{\pi}_A^-$ = 17.1% (se = 3.2) and skilled funds $\hat{\pi}_A^+$ = -0.2% (se = 0.2) - Table 3, Panel A. It is the estimate of $\hat{\pi}_0(\lambda)$ which determines our calculations of the FDR (for alpha) and this is statistically well determined because the estimation uses data on a large number of null funds (see figure 2). (Standard errors are in parentheses and are given in Genovese and Wasserman 2004 and Appendix-A of BSW 2010). Hence in the whole population of M-funds, most have truly zero long-run alphas, probably very few have positive alphas and a sizable proportion have negative alphas.

[Table 3 - here]

The most striking feature about the alpha-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, Panel B). For example for γ = 0.10 (right tail area 0.05), only S_{γ}^{+} = 4.7% (26 funds) have significant positive alphas but given that FDR_{γ}^{+} = 88.8%, only T_{γ}^{+} = 0.5% (3 funds) have truly positive alphas - but this estimate is not statistically different from zero. So, the standard "count" indicates 26 funds are significant but nearly all of these are probably false discoveries. Both S_{γ}^{+} and FDR_{γ}^{+} increase with γ but the percentage of truly skilled funds T_{γ}^{+} is statistically insignificantly different from zero (for $\gamma \leq 0.20$) - Table 3, Panel B.

For negative alpha funds the FDR_{γ}^{-} (for γ = 0.10) is relatively small at 13.3% so of the S_{γ}^{-} = 31.3% (174) significant worst funds, T_{γ}^{-} = 27.2% (150 funds) are truly unskilled rather than having bad luck. 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 γ = [0, 0.2].

Market Timing Models

We now use the FDR to inform our analysis of the importance of our two market timing variables when added to the 3F model (Table 2, Panels B and C). For example (at 10% significance level) for the TM model, we have 60 funds (S_{γ}^{+}=13.3%) with a positive and statistically significant market timing coefficient δ_i which with an estimated FDR_{γ}^{+} of 34.9% gives 39 funds (T_{γ}^{+} = 7.0%) which have truly positive market timing, while the comparable figures for negative market timing are 158 statistically significant δ_i's, an FDR_{γ}^{-} = 13.3%, with 137 funds (T_{γ}^{-} = 24.7%) having truly negative market timing. Hence there are a total of 31.7% of funds which have either truly positive or negative market timing effects - most of which have negative market timing. For the HM model the latter figure is very similar at 29.4% of funds and the results for the HM and TM specifications are very similar. Hence, we cannot ignore market timing effects in our parametric 3F factor model.

However, some caveats are in order when considering market timing results. The market timing parameter δ_i may be biased downwards (but not upwards) because of cash-flow effects. When market returns are high, cash inflows into funds tend to be high which leads to temporarily large cash positions and lower fund betas (Warther,1995, Ferson and Warther 1996 and Edelen 1999). In addition, artificial fund returns generated from "synthetic passive portfolios"[17] which have no market timing ability by construction, when used in the HM and TM timing models can give spurious positive or negative values for δ_i. This is "artificial timing bias" and on US data is particularly evident for funds which hold a preponderance of small stocks, value stocks and past winners and empirically it results in statistically significant *negative* "artificial timing" (i.e. $\hat{\delta}_i$ < 0). Also for US funds Kon (1983) and Hendriksson (1984) find a negative correlation between α_i and δ_i.

Spurious timing effects can arise if funds hold stocks that are more or less option-like than the average stock in the market index (Jagannathan and Korajczyk 1986). For example, if the fund's stocks are *more* option-like than those of the market index, a rise in the latter will lead to a disproportionately large rise in the fund's return and this convex relationship will result in a positive δ, even though the fund is not undertaking any market timing. If delta is biased upwards then alpha will be biased downwards and if this effect is pervasive, we expect a negative correlation between these two parameters, in the cross-section of funds.

We do not have data on stock holdings of German funds and hence cannot directly test for this spurious timing bias. But we do find a negative correlation of around -0.7 between α_i and δ_i in our cross-section of funds (see figures 3 and 4 for the TM and HM models, respectively).[18] Hence we cannot rule out the possibility that some of our positive timing coefficients may be spurious and hence biased.

[Figures 3 and 4 here]

Security Selection (Alpha) and "Total Performance" in Market Timing Models

What are the implications of security selection ('alpha') when we add market timing variables? Compared to the 3F model (i.e. excluding timing variables) there is a substantial increase (at a 10% significance level) in the number of statistically significant positive-alpha

[17] Synthetic passive portfolios" of stocks which mimic the stock holdings of funds are based on the fund's proportionate holdings of high and low book-to-market stocks, small and large stocks, momentum stocks, etc. – Bollen and Busse 2001.

[18] Also for US funds Kon (1983) and Hendriksson (1984) find a negative correlation between α_i and δ_i.

funds, a much lower FDR_{γ}^{+} and an increase in the number of truly positive alpha funds from 3 (0.5%) in the 3F model to 64 (7.4%) in the 3F+TM model and 96 (13.4%) in the 3F+HM model (Table 4, Panels A and B, respectively). Hence it would appear that market timing models provide much stronger evidence of successful security selection skills than the 3F model. It is also the case that the market timing models indicate less negative alpha performance than the 3F model since in the TM (HM) model 126 (109) funds have truly negative alphas, while for the 3F model the figure is 150 funds. Hence, market timing models indicate a substantially improved view of the overall level of skill in security selection (alphas) for the actively managed fund industry, than does the 3F model.

[Table 4 here]

Even though a number of researchers present results on market timing as described above (but without added information from the FDR) there are two acute problems. First is the well documented bias in estimation of the separate security selection and market timing effects. Second, measuring security selection (alpha) without simultaneously considering the effect on fund performance of any market timing effects, can give a misleading picture of overall performance. Clearly, good security selection together with negative market timing (or vice versa) may not be beneficial for investors (relative to investing in index funds or Exchange Traded Funds, ETFs).

Our “total performance” measure, which takes account of security selection and market timing effects on fund returns is $perf_i = \alpha_i + \delta_i \overline{f(r_{mt})}$. For the 3F+TM model (Table 5, Panel A) we reject (at a 10% significance level, for example) the null of $perf_i = 0$ against the alternative $perf_i > 0$ for 23 funds (out of 555) but the estimated FDR is 98% implying that no funds have truly positive total performance.[19] There are 158 funds with statistically significant negative values of $perf_i$ and with a relative low FDR of 14.3% this implies a substantial 135 funds (24.4%) have truly negative overall performance. Results are very similar for the 3F+HM model (Table 5, Panel B).

Comparing results on security selection (alpha) in the 3F model of table 3 with the results using our measure of total performance $perf_i$ in the 3F+MT models (Table 5), both give a

[19] The finding of a statistically significant value for $\hat{\pi}_A^{+}$ >0 when testing α_i = 0 but a statistically insignificant value of $\hat{\pi}_A^{+}$ > 0 when testing $perf_i$ = 0, is also consistent with these results.

consistent picture of the "performance" of German equity mutual funds. Whether performance is measured using 3F-alpha or "total performance" there are virtually no funds with superior performance, around 25% with truly poor performance and around 75% who have zero performance.

[Table 5 here]

Robustness Tests

The 'home-bias' mutual fund literature suggests that physical proximity may facilitate relevant information transmission, which results in a concentration of fund assets geographically (e.g. within a particular country, particular cities or concentrated in particular sectors) and this "superior information" leads to superior performance (Coval and Moskowitz 1999, Hong, Kubik and Stein 2005, Kacperczyk, Sialm and Zheng, 2005). For the 3F model the home-bias hypothesis does not appear to hold for investing in Germany versus investing in firms outside Germany. Table 6 shows that results from investing in these two geographical regions are very similar with a FDR_{γ}^{+} broadly in the 75-95% range (for significance levels 0.05 to 0.20), with only a very small proportion of truly positive alpha funds (around 0.1% to 2%) but a much higher proportion of truly negative alpha funds of around 20-35%[20].

[Table 6 here]

When either the 3F-alphas or the $perf_i$ statistic (for the two, 3F+MT models) are estimated over successive 5-year "short-term" periods January 1995 - December 1999, January 2000 - December 2004 and January 2005 - December 2009, the overall picture remains largely unchanged from the whole sample period results (reported in Tables 3 and 5) and therefore we do not report these results here. Hence in contrast to results for US equity funds where "short-term" truly positive alpha-performance declines from around 5% of all funds up to 2002 to zero percent by 2006 (BSW 2010), the positive performance of the German equity funds industry is zero over both the short-run and the whole life of the funds (for either alpha in the 3F-model or the $perf$ statistic for the two, 3F+MT models).

Above we have reported results based on the 3F and the two, 3F+MT models. Now we assess the sensitivity of our results on alpha and $perf_i$ when we exclude the SMB and HML factors and apply the FDR to the relevant performance measure. In Panel A of table 7 we present results for alpha for the 1F and 2F models and in Panel B for $perf_i$ for the two, 1F+MT and 2F+MT models. We find that the results are qualitatively unchanged from those reported above for the 3F and 3F+MT models and hence for brevity we only report results at the γ = 10% significance level[21].

When we add a momentum factor to the 81 funds which have a German only mandate our results for alpha and $perf_i$ are qualitatively similar[22]. For example, in moving from the 3F to the 4F model we find 5 statistically significant positive alpha funds (10% significance level) with an FDR_γ^+ of 57% in both cases. For negative alphas, the 3F and 4F models give 29 and 32 statistically significant alpha respectively, with an FDR_γ^- of 9% in both cases. The invariance of our results to the momentum factor may be due to its low correlation with the other factors (the maximum correlation of -0.25 is with the market return) and hence any omitted variables bias may be small[23].

[Table 7 here]

5. Conclusions

We use the FDR in model selection and performance measurement to assess the overall performance from both market timing and security selection of the German equity mutual fund industry. When using the Fama-French three factor (3F) model (with no market timing) we find less than 1% of funds (i.e. 6 out of 555) have truly positive alpha-performance, about 27% (150 funds) have truly negative-alpha performance and the majority have zero-alpha performance. These results using the FDR (but excluding market timing variables) are broadly similar to those found for US and UK funds (Kosowski et al 2006, Fama and French 2010, BSW 2010, Cuthbertson et al 2012)- namely, very few statistically significant alpha funds and substantially more negative alpha funds.

---

[20] Qualitatively similar results on the geographical performance are found when using the total performance measure in the two 3F+MT models, hence we do not report these results.

[21] As further tests on these models we have looked at the average Rbar-squared, Akaike (AIC) and Schwarz Bayesian Criterion (BIC) statistics for a) all funds, b) German domestic equity and c) German funds that invest internationally. The Rbar-squared and AIC support the inclusion of the market timing variables and the BIC criterion suggests little to choose between the 3-factor model and the 3-factor plus market timing models. Tests of higher order terms (e.g. the market return cubed) are not suggested by theory but we found this term to be statistically insignificant for nearly all funds. These results are available on request.

[22] The momentum factor for the domestic market is from the Centre for Financial Research, University of Cologne (see Artman et al 2010). If we also use the CFR market return, SMB and HML factors over this period our results remain broadly unchanged.

[23] We also constructed an international momentum variable as outlined in Fletcher and Marshall (2005). As in their table 5 for UK funds which invest internationally, we found no qualitative difference in our performance measures for this change in our factor model.

Use of the FDR in model selection, implies inclusion of the TM or HM market timing variables with results similar to those on UK and US funds – namely, some evidence of positive timing and stronger evidence for negative timing. When we examine the 3F+MT models this results in a large increase in the proportion of truly positive-alpha funds from around 1% to 7-13% (40 to 75 funds) and a reduction in the proportion of truly negative-alpha funds from around 27% (150 funds) to about 17% (95 funds). We also find evidence consistent with "spurious timing" which may bias downward, estimates of security selection (alpha). However, when we attempt to mitigate these problems by using a measure of "total performance", which includes the contribution of both security selection (alpha) and market timing, we obtain performance results similar to the 3F model (with no market timing). This demonstrates the importance of using the FDR to inform model selection and in using a measure of total performance when market timing variables are included in a factor model. The above results are largely invariant to the inclusion of different factors (except for the market factor), for different sample periods and to the performance of funds investing in German and non-German stocks – the latter casts some doubt on the "home-bias" hypothesis of superior performance due to comparative advantage in information about 'local' markets.

References

Admati, A.R., S. Bhattacharya, Stephen A. Ross, and P. Pfleiderer, 1986, On Timing and Selectivity, *Journal of Finance,* 41, 715-730.

Alexander, Gordon J, Gjergji Cici and Scott Gibson, 2007, Does Motivation Matter When Assessing Trade Performance? An Analysis of Mutual Funds, *Review of Financial Studies,* 20,125-150.

Artmann, S., Finter, P., Kempf, A, Koch, S. and E. Theissen 2010, The Cross-Section of German Stock Returns: New Data and New Evidence, University of Cologne, CFR Working Paper 10-12.

Bajgrowicz, Pierre and Olivier Scaillet, 2008, Technical Trading Rules Revisited: Persistence Tests, Transaction Costs and False Discoveries, HEC Geneva, Working Paper.

Barras, Laurent, Olivier Scaillet, and Russ Wermers, 2010, False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, *Journal of Finance*, Vol. 65, No. 1, pp. 179-216

Becker, C., W. Ferson, D.H. Myers and M.J. Schill, 1999, Conditional Market Timing with Benchmark Investors, *Journal of Financial Economics,* Vol. 52, pp. 47-78.

Benjamini Y. and Y. Hochberg, 1995, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *Journal of Royal Statistical Society,* Vol. 57 (1), pp. 289-300.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual Fund Flows and Performance in Rational Markets, *Journal of Political Economy,* Vol. 112, pp. 1269-95.

Bessler, Wolfgang, David Blake, Peter Luckoff and Ian Tonks, 2010, Why Does Mutual Fund Performance Not Persist? The Impact and Interaction of Fund Flows and Manager Changes, Pensions Institute, Cass Business School, WP PI-1009.

Bessler, Wolfgang, Wolfgang Drobetz and Heinz Zimmermann, 2009, Conditional Performance Evaluation for German Equity Mutual Funds, *European Journal of Finance,* Vol 15 (3), 287-316.

Blake, David, and Allan Timmermann, 1998, Mutual Fund Performance: Evidence from the UK, *European Finance Review,* 2, 57-77.

Bollen, Nicolas P.B. and Jeffrey A. Busse, 2001, On the Timing Ability of Mutual Fund Managers, *Journal of Finance*, LVI (3), pp. 1075-1094.

Bollen, Nicolas P.B., and Jeffrey A. Busse, 2004, Short-Term Persistence in Mutual Fund Performance, *Review of Financial Studies,* Vol. 18 (2), pp. 569-597

Busse, J. (1999), 'Volatility Timing in Mutual Funds: Evidence from Daily Returns,' *Review of Financial Studies*, Vol. 12, pp. 1009–41.

Byrne, A, J. Fletcher and P. Ntozi 2006, 'An Exploration of the Conditional Timing Performance of UK Unit Trusts,' *Journal of Business Finance and Accounting*, Vol. 33, pp. 816-38.

Carhart, Mark M, 1997, On Persistence in Mutual Fund Performance, *Journal of Finance Vol.* 52, pp. 57-82

Chen, C., C. F. Lee, S. Rahman and A. Chan 1992, ‘A Cross-Sectional Analysis of Mutual Funds’ Market Timing and Security Selection Skill,’ *Journal of Business Finance and Accounting*, Vol. 19, pp. 659–75.

Chen, Joseph, Harrison Hong, Ming Huang and Jeffrey D. Kubik 2004, Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization, *American Economic Review,* 94, 1276-1302.

Chen, Y. and B. Liang (2007), ‘Do Market Timing Hedge Funds Time the Market?’ *Journal of Financial and Quantitative Analysis*, Vol. 42, pp. 827–56.

Christopherson, Jon A., Wayne E. Ferson, and Debra A. Glassman, 1998, Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance, *Review of Financial Studies,* 11, 111-142

Coles, Jeffrey, L., Naveen D. Daniel and Frederico Nardari, 2006, Does the Choice of Model or Benchmark Affect Inference in Measuring Mutual Fund Performance?, Working Paper, Arizona State University, January.

Cooper, Michael J., Huseyin Gulen and P. Raghavendra Rau, 2005, Changing Names With Style: Mutual Fund Name Changes and Their Effects on Fund Flows, *Journal of Finance,* Vol. 60, pp. 2825-2858.

Coval, Joshua D., and Tobias J.Moskowitz, 1999, Home Bias at Home: Local Equity Preference in Domestic Portfolios, *Journal of Finance,* Vol. 54, pp. 2045-2074.

Coval, JoshuaD. And Erik Stafford 2007, Asset Fire Sales (and Purchases) in Equity Markets, *Journal of Financial Economics*, 86, 479-512.

Criton, Gilles and Olivier Scaillet, 2009, Time-Varying Coefficient Model for Hedge Funds, SSRN Working Paper, March.

Cuthbertson, Keith, Dirk Nitzsche and Niall O’Sullivan, 2008, ‘Performance of UK Mutual Funds: Luck or Skill ? “, *Journal of Empirical Finance*, Vol. 15(4), pp. 613-634.

Cuthbertson, Keith, Dirk Nitzsche and Niall O’Sullivan, 2010,The Market Timing Ability of UK Mutual Funds, *Journal of Business, Finance and Accounting*, Vol. 37 (1&2), pp. 270-289

Cuthbertson, Keith, Dirk Nitzsche and Niall O’Sullivan, 2012, False Discoveries in UK Mutual Fund Performance, *European Financial Management*, forthcoming

Dangl, Thomas, Youchang Wu and Josef Zechner, 2008, Market Discipline and Internal Governance in the Mutual Fund Industry, *Review of Financial Studies,* 21, 2307-2343.

Del Guercio, Diane and Paula A. Tkac, 2008, Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow, *Journal of Financial and Quantitative Analysis,* Vol. 43, No. 4, pp. 907-936.

Edelen, Roger M. and Jerold B. Warner, 2001, Aggregate Price Effects of Institutional

Trading: A Study of Mutual Fund Flow and Market Returns, *Journal of Financial Economics*, Vol. 59, pp. 196-220.

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

Elton, Edwin J., Martin J. Gruber, Das, S. and Hlavka, M. 1993, Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios, *Review of Financial Studies*, Vol. 6, pp. 1-21.

Fama, Eugene F. and Kenneth R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics,* Vol. 33, pp. 3-56.

Fama, Eugene F. and Kenneth R. French, 2010, Luck versus Skill in the Cross Section of Mutual Fund Returns*, Journal of Finance*, Vol. 65, No. 5, pp. 1915-1947

Ferson, Wayne E. and Khang, K 2002, 'Conditional Performance Measurement Using Portfolio Weights: Evidence for Pension Funds', *Journal of Financial Economics,* Vol. 65 (2), pp. 249-282.

Ferson, Wayne E. and Rudi W. Schadt, 1996, Measuring Fund Strategy and Performance in Changing Economic Conditions, *Journal of Finance,* Vol. 51, pp. 425-62.

Fletcher, Jonathan and Andrew Marshall, 2005, *Journal of Financial Services Research,* Vol 27(2), pp.183-206.

Fletcher, Jonathan, 1995, An Examination of the Selectivity and Market Timing Performance of UK Unit Trusts, *Journal of Business Finance and Accounting* Vol. 22, pp. 143-156.

Fletcher, Jonathan, 1997, An Examination of UK Unit Trust Performance Within the Arbitrage Pricing Framework, *Review of Quantitative Finance and Accounting*, Vol. 8, pp. 91-107.

Gallo , J.G and P.E. Swanson, 1996, Comparative Measures of Performance for US Based International Equity Funds, *Journal of Banking and Finance, Vol* 20, pp.1635-1650.

Genovese, Christopher and Larry Wasserman, 2004, A Stochastic Process Approach to False Discovery Control, *Annals of Statistics*, Vol. 32, pp. 1035-1061.

Goetzmann, William N., Ingersoll Jr., J., and Ivkovich, Z., 2000, Monthly Measurement of Daily Timers, *Journal of Financial and Quantitative Analysis*, Vol. 35, pp 257-290.

Griese, Knut and Alexander Kempf, 2002, Lohnt Aktives Fondsmanagement Anglengersicht? Ein Vergleich von Anlagestrategien in Activ und Passiv Verwaltenten Aktienfonds, *Zeitschrift fur Betriebswirtschraft*, 73, 201-224.

Hendriksson, R.D., 1984, Market Timing and Mutual Fund Performance : An Empirical Investigation, *Journal of Business*, Vol. 57 (1), pp. 73-96.

Henriksson, R.D. and Robert C. Merton, 1981, On Market Timing and Investment Performance : Statistical Procedures for Evaluating Forecasting Skills, *Journal of Business,* Vol. 54, pp. 513-533.

Holm, S., 1979, A Simple Sequentially Rejective Multiple Test Procedure, *Scandinavia Journal of Statistics*, Vol. 6, pp. 65-70.

Hong, Harrison, Jeffrey D. Kubik and Jeremy Stein, 2005, Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers, *Journal of Finance,* LX(6), pp. 2801-24.

Ivkovic, Zoran and Scott Weisbenner, 2009, Individual Investor Mutual Fund Flows, *Journal of Financial Economics,* 92, 223-237*.*

Jagannathan, R. and R. Korajczyk, 1986, Assessing the Market Timing Performance of Managed Portfolios, *Journal of Business*, 59, 217-235.

Jiang, George J., Tong Yao and Tong Yu, 2007, Do Mutual Funds Time the Market? Evidence from Portfolio Holdings, *Journal of Financial Economics,* 86 (3), 724-758.

Jiang, Wei, 2003, A Non-Parametric Test of Market Timing, *Journal of Empirical Finance,*10, 399-425.

Kacperczyk, Marcin, Clemens Sialm and Lu Zheng, 2005, On the Industry Concentration of Actively Managed Mutual Funds, *Journal of Finance,* LX(4), 1983-2011.

Khorana, Ajay, 1996, Top Management Turnover: An Empirical Investigation of Mutual Fund Managers, *Journal of Financial Economics,* 40 (3), pp. 403-427.

Khorana, Ajay, 2001, Performance Changes Following Top Management Turnover: Evidence From Open-End Mutual Funds, *Journal of Financial and Quantitative Analysis,* 36, 371-393.

Kon, S.J., 1983, The Market-Timing Performance of Mutual Fund Managers, *Journal of Business*, Vol. 56 (3), pp. 323-347.

Kosowski, R., A. Timmermann, H. White and R. Wermers, 2006, Can Mutual Fund 'Stars' Really Pick Stocks? New Evidence from a Bootstrapping Analysis, *Journal of Finance*, Vol. 61, No. 6, pp. 2551-2595.

Krahner, Jan Pieter, Frank Schmid and Erik Theissen, 2006, Investment Performance and Market Share : A Study of the German Mutual Fund Industry, in Wolfgang Bessler : Boersen, Banken and Kapitalmaerkte, Duncker&Humblot, Berlin, pp. 471-491.

Lee, C. and S. Rahman (1990), 'Market Timing, Selectivity, and Mutual Fund Performance: An Empirical Investigation,' *Journal of Business*, Vol. 63, pp. 261–78.

Leger, L., 1997, UK Investment Trusts : Performance, Timing and Selectivity, *Applied Economics Letters,* Vol. 4, pp. 207-210.

Lehmann, Bruce and Allan Timmermann, 2007, Performance Measurement and Evaluation, *Handbook of Financial Intermediation and Banking*, edited by Arnoud Boot and Anjan Thakor, Elsevier.

Lynch, Anthony W. and David K. Musto, 2003, How Investors Interpret Past Fund Returns, *Journal of Finance,* LVIII(5), pp. 2033-2058.

Malkiel, G., 1995, Returns from Investing in Equity Mutual Funds 1971 to 1991, *Journal of Finance*, Vol. 50, pp. 549-572.

McCracken, Michael. W. and Stephen G. Sapp, 2005, 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), pp. 473-494.

Otten, Roger and D. Bams, 2002, European Mutual Fund Performance, *European Financial Management*, Vol. 8(1), pp 75-101.

Patro, D.K., 2001, Performance of Closed-End International Mutual Funds, *Journal of Banking and Finance,* Vol 25, pp.1741-1767.

Politis, D.N. and J.P. Romano, 1994, The Stationary Bootstrap, *Journal of the American Statistical Association,* Vol. 89, pp. 1303-1313.

Pollit, Joshua M. and Mungo Wilson 2008, How Does Size Affect Mutual Fund Behavior?, *Journal of Finance,* 63, 2941-2969.

Quigley, Garrett, and Rex A. Sinquefield, 2000, Performance of UK Equity Unit Trusts, *Journal of Asset Management,* Vol. 1, pp. 72-92

Sirri Erik R. and Peter Tufano, 1998, Costly Search and Mutual Fund Flows, *Journal of Finance,* 53(5), pp.1589-1622.

Stehle, Richard and Olaf Grewe, 2001, The Long-Run Performance of German Stock Mutual Funds, Humboldt University, Berlin, Discussion Paper.

Storey J. D., 2002, A Direct Approach to False Discovery Rates, *Journal of Royal Statistical Society B,* Vol. 64, pp. 497-498.

Storey, J. D., J.E. Taylor and D. Siegmund, 2004, Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach, *Journal of Royal Statistical Society,* Vol. 66, pp. 187-205.

Swinkels, L. and L. Tjong-A-Tjoe (2007), 'Can Mutual Funds Time Investment Styles?' *Journal of Asset Management*, Vol. 8, pp. 123–32.

Treynor, Jack, and K. Mazuy, 1966, Can Mutual Funds Outguess the Market, *Harvard Business Review,* Vol. 44, pp. 66-86.

Warther, Vincent A., 1995, Aggregate Mutual Fund Flows and Security Returns, *Journal of Financial Economics*, Vol. 39, pp. 209-235.

Wermers, R., 2000, 'Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transaction Costs, and Expenses', *Journal of Finance*, Vol. 55, pp. 1655–95.

Yan, Xuemin (Sterling), 2008, Liquidity, Investment Style and the Effect of Fund Size on Fund Performance, *Journal of Financial and Quantitative Analysis, 43, 741-768.*

Table 1 Summary Statistics German Equity Mutual Funds

This table reports summary statistics of all the funds used in the analysis. The sample period is from January 1990 to December 2009 (monthly data) and includes 555 German domiciled mutual funds which have at least 24 observations. The average number of observations for the funds is 111 months. We report averages of the individual fund statistics for five different models (1F, 2F, 3F, and the 3F+TM and 3F+HM market timing models. The first factor is the corresponding excess market return, the second factor is the size factor and the third factor is the book-to-market factor. The t-statistics are based on Newey-West heteroscedastic and autocorrelation adjusted standard errors. Statistical significance is at the 5% significance level (two-tail test). BJ is the Bera-Jarque statistic for normality of residuals.

	1F Model (r_m)	2F Model (r_m ,SMB)	3F Model (r_m , SMB,HML)	3F+TM r_m^2	3F+HM r_m^+
Panel A : Average Coefficient Results					
Number (#) of Positive and Negative Alphas					
Positive (# significant)	165 (7)	146 (4)	159 (7)	212 (30)	259 (38)
Negative(#significant)	390 (65)	409 (85)	396 (79)	343 (54)	296 (43)
Mean Values of Coefficients and t-statistics					
Alpha (t-stat)	-0.1761 (-0.6391)	-0.2274 (-0.7783)	-0.1955 (-0.7299)	-0.1129 (-0.3009)	-0.0684 (-0.1141)
r_m (t-stat)	0.9764 (17.21)	0.9590 (17.88)	0.9668 (18.55)	0.9509 (17.91)	0.9929 (12.52)
SMB (t-stat)	-	0.3207 (2.31)	0.3326 (2.62)	0.3365 (2.70)	0.3360 (2.69)
HML (t-stat)	-	-	-0.2068 (-0.9530)	-0.1839 (-0.8597)	-0.1876 (-0.8669)
TM-Timing variable r_m^2	-	-	-	-0.0042 (-0.4529)	-
HM-Timing variable : r_m^+	-	-	-	-	-0.0756 (-0.3683)
Panel B : Diagnostics					
Mean R^2	0.7266	0.7583	0.7812	0.7896	0.7879
Skewness	-0.1334	-0.1460	-0.1586	-0.1173	-0.1252
Kurtosis	8.77	8.51	8.14	7.97	7.99
BJ – statistic	3279.14	3298.54	3123.34	3071.70	3077.62
% (Number) funds non-normal residuals	36.58% (203 funds)	40.36 (224 funds)	47.38% (263 funds)	50.45% (280 funds)	50.09% (278 funds)

Table 2 FDR: Different Independent Variables

This table reports parameters and the FDR (at various significance levels) when testing the null that a particular parameter is zero (against the alternative that it is either positive or negative). The sample period is from January 1990 to December 2009 (monthly data) and includes 555 German domiciled mutual funds which have at least 24 observations. We report the number (#) of statistically significant coefficients, the FDR, the proportion of statistically significant positive (S^+) and negative (S^-) alpha-funds, the proportion of truly positive (T^+) and negative (T^-) alpha-funds and the proportion of false positives (F^+) and false negative (F^-) alpha-funds, at various significance levels. Panel A reports the statistics on the r_m, SMB and HML coefficients and Panel C for the TM and HM market timing coefficients.

Panel A : Fama-French Factors									
Market Return r_m									
Positive Coefficient (552 funds)					Negative Coefficient (3 funds)				
Sign. level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign.funds	544	544	546	547	Sign. funds	0	0	0	1
FDR^+	0.0006	0.0011	0.0017	0.0023	FDR^-	N/A	N/A	N/A	1.2308
SMB variable									
Positive Coefficient (500 funds)					Negative Coefficient (55 funds)				
Sign. Level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign.Funds	384	419	439	451	#Sign.Funds	9	17	20	26
FDR^+	0.0085	0.0155	0.0222	0.0288	FDR^-	0.3611	0.3824	0.4875	0.5000
HML variable									
Positive Coefficient (176 funds)					Negative Coefficient (379 funds)				
Sign. Level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign.Funds	80	103	112	122	#Sign. funds	219	247	271	297
FDR^+	0.0754	0.1172	0.1617	0.1979	FDR^-	0.0276	0.0489	0.0668	0.0813
Panel B : Test on TM , r_m^2									
Positive Coefficients (199 funds)					Negative Coefficients (356 funds)				
Sign. level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign. funds	43	60	80	97	#Sign. funds	116	158	182	217
FDR^+	0.2434	0.3488	0.3924	0.4315	FDR^-	0.0902	0.1325	0.1725	0.1929
S^+	0.0902	0.1325	0.1725	0.1929	S^-	0.2090	0.2847	0.3279	0.3910
T^+	0.0586	0.0704	0.0876	0.0994	T^-	0.1902	0.2470	0.2714	0.3156
F^+	0.0189	0.0377	0.0566	0.0754	F^-	0.0189	0.0377	0.0566	0.0754
Panel C : Test on HM, r_m^+									
Positive Coefficients (204 funds)					Negative Coefficients (351 funds)				
Sign. level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign. funds	44	68	87	105	#Sign. funds	96	138	178	210
FDR^+	0.2419	0.3130	0.3670	0.4054	FDR^-	0.1109	0.1542	0.1794	0.2027
S^+	0.0793	0.1225	0.1568	0.1892	S^-	0.1730	0.2486	0.3207	0.3784
T^+	0.0601	0.0842	0.0992	0.1125	T^-	0.1538	0.2103	0.2632	0.3017
F^+	0.0192	0.0384	0.0575	0.0767	F^-	0.0192	0.0384	0.0575	0.0767

Table 3 Security Selection (Alpha): Fama-French 3F Model

This table reports statistics to test for security selection (alpha) for the 3F model. The sample period is from January 1990 to December 2009 (monthly data) and includes 555 German domiciled mutual funds which have at least 24 observations. We report the number (#) of statistically significant funds at various significance levels. Panel A reports the estimated proportions of truly null, skilled and unskilled funds. In panel B for various significance levels we report the FDR for positive and negative alpha funds, the proportion of statistically significant positive (S^{+}) and negative (S^{-}) alpha-funds, the proportion of truly positive (T^{+}) and negative (T^{-}) alpha-funds and the proportion of false positives (F^{+}) and false negative (F^{-}) alpha-funds. Standard errors are in parentheses.

Panel A : Proportion of Truly Null, Skilled and Unskilled Funds	
Proportion of Truly Null Funds : $\hat{\pi}_0$ = 0.8316 (0.0325)	
Proportion, skilled funds : $\hat{\pi}_A^{+}$ = -0.0024 (0.0020)	Proportion, unskilled funds : $\hat{\pi}_A^{-}$ = 0.1708 (0.0324)

Panel B : Calculation of FDR Statistics

Sign. level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign. funds	14	26	37	53	#Sign. funds	121	174	218	253
FDR^{+}	0.8242	0.8876	0.9356	0.8708	FDR^{-}	0.0954	0.1326	0.1588	0.1824
S^{+}	0.0252 (0.0067)	0.0468 (0.0090)	0.0667 (0.0106)	0.0955 (0.0125)	S^{-}	0.2180 (0.0175)	0.3135 (0.0197)	0.3928 (0.0207)	0.4559 (0.0211)
T^{+}	0.0044 (0.0082)	0.0053 (0.0130)	0.0043 (0.0176)	0.0123 (0.0226)	T^{-}	0.1972 (0.0185)	0.2719 (0.0227)	0.3304 (0.0265)	0.3727 (0.0301)
F^{+}	0.0208 (0.0008)	0.0416 (0.0016)	0.0624 (0.0024)	0.0832 (0.0033)	F^{-}	0.0208 (0.0008)	0.0416 (0.0016)	0.0624 (0.0024)	0.0832 (0.0033)

Table 4 Alpha Estimates: 3F+MT Models

This table reports statistics to test for security selection (alpha) for the two, 3F+MT models. Panel A reports results for the 3F+TM model and Panel B for the 3F+HM model. The sample period is from January 1990 to December 2009 (monthly data) and includes 555 German domiciled mutual funds which have at least 24 observations. We report the number (#) of statistically significant funds at various significance levels and the estimate of π_0 used to calculate the FDR. For various significance levels we report the FDR for positive and negative alpha funds, the proportion of statistically significant positive (S^+) and negative (S^-) alpha-funds, the proportion of truly positive (T^+) and negative (T^-) alpha-funds and the proportion of false positives (F^+) and false negative (F^-) alpha-funds. Standard errors are in parentheses.

Panel A : TM-Model: $\hat{\pi}_0$ = 0.8263 (0.0299)									
Positive Alpha (212 funds) Proportion, skilled funds : $\hat{\pi}_A^+$ = 0.0229 (0.0070)					Negative Alpha (343 funds) Proportion, unskilled funds : $\hat{\pi}_A^-$ = 0.1508 (0.0293)				
Sign. level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign. funds	50	64	91	115	#Sign. funds	85	126	162	197
FDR^+	0.2293	0.3583	0.3779	0.3988	FDR^-	0.1349	0.1820	0.2123	0.2328
S^+	0.0901 (0.0122)	0.1153 (0.0136)	0.1640 (0.0157)	0.2072 (0.0172)	S^-	0.1532 (0.0153)	0.2270 (0.0178)	0.2919 (0.0193)	0.3550 (0.0203)
T^+	0.0694 (0.0132)	0.0740 (0.0166)	0.1020 (0.0213)	0.1246 (0.0256)	T^-	0.1325 (0.0162)	0.1857 (0.0206)	0.2299 (0.0246)	0.2723 (0.0286)
F^+	0.0207 (0.0007)	0.0413 (0.0015)	0.0620 (0.0022)	0.0826 (0.0030)	F^-	0.0207 (0.0007)	0.0413 (0.0015)	0.0620 (0.0022)	0.0826 (0.0030)
Panel B : HM-Model: $\hat{\pi}_0$ = 0.7872 (0.0326)									
Positive Alpha (259 funds) Proportion, skilled funds : $\hat{\pi}_A^+$ = 0.0357 (0.0084)					Negative Alpha (296 funds) Proportion, unskilled funds : $\hat{\pi}_A^-$ = 0.1770 (0.0318)				
Sign. level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign. funds	63	96	115	137	#Sign. funds	72	109	143	178
FDR^+	0.1734	0.2276	0.2849	0.3189	FDR^-	0.1517	0.2004	0.2292	0.2455
S^+	0.1135 (0.0135)	0.1730 (0.0161)	0.2072 (0.0172)	0.2468 (0.0183)	S^-	0.1297 (0.0143)	0.1964 (0.0169)	0.2577 (0.0186)	0.3207 (0.0198)
T^+	0.0938 (0.0145)	0.1336 (0.0191)	0.1482 (0.0229)	0.1681 (0.0270)	T^-	0.1100 (0.0153)	0.1570 (0.0198)	0.1986 (0.0241)	0.2420 (0.0285)
F^+	0.0197 (0.0008)	0.0394 (0.0016)	0.0590 (0.0024)	0.0787 (0.0033)	F^-	0.0197 (0.0008)	0.0394 (0.0016)	0.0590 (0.0024)	0.0787 (0.0033)

Table 5 Total Performance $(perf)$: 3F+MT Models

This table reports statistics to test for "total performance" $(perf)$ for the two, 3F+MT models. Panel A reports results for the 3F+TM model and Panel B for the 3F+HM model. The sample period is from January 1990 to December 2009 (monthly data) and includes 555 German domiciled mutual funds which have at least 24 observations. We report the number (#) of statistically significant funds at various significance levels and the estimate of π_0 used to calculate the FDR. For various significance levels we report the FDR for positive and negative total performance $(perf)$ funds, the proportion of statistically significant positive (S^+) and negative (S^-) $perf$ funds, the proportion of truly positive (T^+) and negative (T^-) $perf$ funds and the proportion of false positives (F^+) and false negative (F^-) $perf$ funds. Standard errors are in parentheses.

Panel A: TM-Model: $\hat{\pi}_0$ = 0.8150 (0.0326)									
Positive $perf$ (155 funds) Proportion, skilled funds : $\hat{\pi}_A^+$ = -0.0023 (0.0020)					Negative $perf$ (400 funds) Proportion, unskilled funds : $\hat{\pi}_A^-$ = 0.1873 (0.0324)				
Sign. level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign funds	17	23	35	50	#Sign funds	104	158	211	246
FDR^+	0.6652	0.9833	0.9692	0.9046	FDR^-	0.1087	0.1431	0.1608	0.1839
S^+	0.0306 (0.0073)	0.0414 (0.0085)	0.0631 (0.0103)	0.0901 (0.0122)	S^-	0.1874 (0.0166)	0.2847 (0.0192)	0.3802 (0.0206)	0.4432 (0.0211)
T^+	0.0103 (0.0087)	0.0007 (0.0126)	0.0019 (0.0174)	0.0086 (0.0224)	T^-	0.1670 (0.0176)	0.2439 (0.0221)	0.3191 (0.0263)	0.3617 (0.0300)
F^+	0.0204 (0.0008)	0.0407 (0.0016)	0.0611 (0.0024)	0.0815 (0.0033)	F^-	0.0204 (0.0008)	0.0407 (0.0016)	0.0611 (0.0024)	0.0815 (0.0033)
Panel B: HM-Model: $\hat{\pi}_0$ = 0.8094 (0.0326)									
Positive $perf$ (156 funds) Proportion, skilled funds : $\hat{\pi}_A^+$ = -0.0022 (0.0020)					Negative $perf$ (399 funds) Proportion, unskilled funds : $\hat{\pi}_A^-$ = 0.1928 (0.0324)				
Sign. level	0.05	0.10	0.15	0.20	Sign. level	0.05	0.10	0.15	0.20
#Sign funds	15	24	36	53	#Sign funds	96	159	207	235
FDR^+	0.7487	0.9359	0.9359	0.8476	FDR^-	0.1170	0.1413	0.1628	0.1912
S^+	0.0270 (0.0069)	0.0432 (0.0086)	0.0649 (0.0105)	0.0955 (0.0125)	S^-	0.1730 (0.0161)	0.2865 (0.0192)	0.3730 (0.0205)	0.4234 (0.0210)

T^+	0.0068 (0.0084)	0.0028 (0.0127)	0.0042 (0.0175)	0.0146 (0.0226)	T^-	0.1527 (0.0171)	0.2460 (0.0222)	0.3123 (0.0262)	0.3425 (0.0298)
F^+	0.0202 (0.0008)	0.0405 (0.0016)	0.0607 (0.0024)	0.0809 (0.0033)	F^-	0.0202 (0.0008)	0.0405 (0.0016)	0.0607 (0.0024)	0.0809 (0.0033)

Table 6 Security Selection (Alpha): 3F Model, Different Geographic Regions

This table reports statistics to test for security selection (alpha) for the 3F model. The sample period is from January 1990 to December 2009 (monthly data) and includes 555 German domiciled mutual funds which have at least 24 observations. We report the number (#) of statistically significant funds at various significance levels. Panel A (Panel B) reports results for funds investing in only German companies (non-German companies). For various significance levels we report the FDR for positive and negative alpha funds, the proportion of statistically significant positive (S^+) and negative (S^-) alpha-funds, the proportion of truly positive (T^+) and negative (T^-) alpha-funds and the proportion of false positives (F^+) and false negative (F^-) alpha-funds. Standard errors are in parentheses.

Panel A : German Companies (85 funds)									
Positive Alpha (22 funds)					Negative Alpha (63 funds)				
Sign level	0.05	0.10	0.15	0.20	Sign level	0.05	0.10	0.15	0.20
#Sign Funds	2	4	7	9	#Sign Funds	17	30	32	39
FDR^+	0.8836	0.8836	0.7574	0.7854	FDR^-	0.1040	0.1178	0.1657	0.1812
S^+	0.0235 (0.0164)	0.0471 (0.0230)	0.0824 (0.0298)	0.1059 (0.0334)	S^-	0.2000 (0.0434)	0.3529 (0.0518)	0.3765 (0.0526)	0.4588 (0.0540)
T^+	0.0027 (0.0179)	0.0055 (0.0270)	0.0200 (0.0367)	0.0227 (0.0439)	T^-	0.1792 (0.0444)	0.3114 (0.0550)	0.3141 (0.0585)	0.3757 (0.0636)
F^+	0.0208 (0.0018)	0.0416 (0.0036)	0.0624 (0.0054)	0.0832 (0.0072)	F^-	0.0208 (0.0018)	0.0416 (0.0036)	0.0624 (0.0054)	0.0832 (0.0072)
Panel B : Non-German Companies (470 funds)									
Positive Alpha (137 funds)					Negative Alpha (333 funds)				
Significance level	0.05	0.10	0.15	0.20	Significance level	0.05	0.10	0.15	0.20
# Sign Funds	12	22	30	44	Number of Significant Funds	104	144	186	214
FDR^+	0.8143	0.8883	0.9771	0.8883	FDR^-	0.0940	0.1357	0.1576	0.1826
S^+	0.0255	0.0468	0.0638	0.0936	S^-	0.2213	0.3064	0.3957	0.4553

	(0.0073)	(0.0097)	(0.0113)	(0.0134)		(0.0191)	(0.0213)	(0.0226)	(0.0230)
T^+	0.0047 (0.0088)	0.0052 (0.0138)	0.0015 (0.0185)	0.0105 (0.0238)	T^-	0.2005 (0.0202)	0.2648 (0.0245)	0.3334 (0.0286)	0.3722 (0.0325)
F^+	0.0208 (0.0009)	0.0416 (0.0017)	0.0624 (0.0026)	0.0832 (0.0035)	F^-	0.0208 (0.0009)	0.0416 (0.0017)	0.0624 (0.0026)	0.0832 (0.0035)

Table 7 Performance: Alternative Models

This table reports performance measures for different models. Panel A reports statistics to test for security selection (alpha) in the 1F (market return) and 2F model (market return and SMB factor). Panel B reports statistics to test for "total performance" $(perf)$ for the 1F+TM and 2F+TM timing models while Panel C repeats the latter for the HM timing model. The sample period is from January 1990 to December 2009 (monthly data) and includes 555 German domiciled mutual funds which have at least 24 observations. We report the FDR for positive and negative alpha funds, the proportion of statistically significant positive (S^+) and negative (S^-) alpha-funds, the proportion of truly positive (T^+) and negative (T^-) alpha-funds and the proportion of false positives (F^+) and false negative (F^-) alpha-funds. Standard errors are in parentheses. All test results are reported for a significance level of 10% (two tail test).

Panel A : Security Selection (alpha): 1F and 2F Models							
1F Model $\hat{\pi}_0$ = 0.8700 (0.0296)				**2F Model** $\hat{\pi}_0$ = 0.7875 (0.0326)			
Positive Alpha $\hat{\pi}^+$ = -0.0044 (0.0009)		Negative Alpha $\hat{\pi}^-$ = 0.1343 (0.0295)		Positive Alpha $\hat{\pi}^+$ = -0.0021 (0.0020)		Negative Alpha $\hat{\pi}^-$ = 0.2149 (0.0325)	
#Funds	165	#Funds	390	#Funds	146	# Funds	409
#Sign funds	25	#Sign funds	161	#Sign funds	25	#Sign funds	184
FDR^+	0.9657	FDR^-	0.1500	FDR^+	0.8738	FDR^-	0.1187
S^+	0.0450 (0.0088)	S^-	0.2901 (0.0193)	S^+	0.0450 (0.0088)	S^-	0.3315 (0.0200)
T^+	0.0015 (0.0126)	T^-	0.2466 (0.0231)	T^+	0.0057 (0.0129)	T^-	0.2922 (0.0230)
F^+	0.0435 (0.0015)	F^-	0.0435 (0.0015)	F^+	0.0394 (0.0016)	F^-	0.0394 (0.0016)
Panel B: $perf$ **, 1F+TM and 2F+TM Timing Models**							
1F+TM Model $\hat{\pi}_0$ = 0.8676 (0.0323)				**2F+TM Model** $\hat{\pi}_0$ = 0.8082 (0.0301)			

Positive $perf$ $\hat{\pi}^{+}$ = -0.0007 (0.0027)		Negative $perf$ $\hat{\pi}^{-}$ = 0.1331 (0.0322)		Positive $perf$ $\hat{\pi}^{+}$ = -0.0022 (0.0020)		Negative $perf$ $\hat{\pi}^{-}$ = 0.1940 (0.0299)	
#Funds	165	#Funds	390	# Funds	144	#Funds	411
#Sign funds	23	#Sign funds	144	#Sign funds	26	#Sign funds	172
FDR^{+}	1.0468	FDR^{-}	0.1672	FDR^{+}	0.8626	FDR^{-}	0.1304
S^{+}	0.0414 (0.0085)	S^{-}	0.2595 (0.0186)	S^{+}	0.0468 (0.0090)	S^{-}	0.3099 (0.0196)
T^{+}	-0.0019 (0.0126)	T^{-}	0.2161 (0.0216)	T^{+}	0.0064 (0.0127)	T^{-}	0.2695 (0.0225)
F^{+}	0.0434 (0.0016)	F^{-}	0.0434 (0.0016)	F^{+}	0.0404 (0.0015)	F^{-}	0.0404 (0.0015)
Panel C: $perf$, 1F+HM and 2F+HM Timing Models							
1F+HM Model $\hat{\pi}_0$ = 0.8700 (0.0296)				**2F+HM Model** $\hat{\pi}_0$ = 0.7902 (0.0301)			
Positive $perf$ $\hat{\pi}^{+}$ = -0.0007 (0.0027)		Negative $perf$ $\hat{\pi}^{-}$ = 0.1307 (0.0294)		Positive $perf$ $\hat{\pi}^{+}$ = 0.0015 (0.0032)		Negative $perf$ $\hat{\pi}^{-}$ = 0.2083 (0.0299)	
#Funds	166	#Funds	398	#Funds	152	#Funds	403
#Sign funds	25	#Sign funds	140	#Sign funds	25	#Sign funds	166
FDR^{+}	0.9657	FDR^{-}	0.1724	FDR^{+}	0.8771	FDR^{-}	0.1321
S^{+}	0.0450 (0.0088)	S^{-}	0.2523 (0.0184)	S^{+}	0.0450 (0.0088)	S^{-}	0.2991 (0.0194)
T^{+}	0.0015 (0.0126)	T^{-}	0.2088 (0.0213)	T^{+}	0.0055 (0.0126)	T^{-}	0.2596 (0.0223)
F^{+}	0.0435 (0.0015)	F^{-}	0.0435 (0.0015)	F^{+}	0.0395 (0.0015)	F^{-}	0.0395 (0.0015)

Figure 1 Cross-Section Fund Alphas: 3F Model

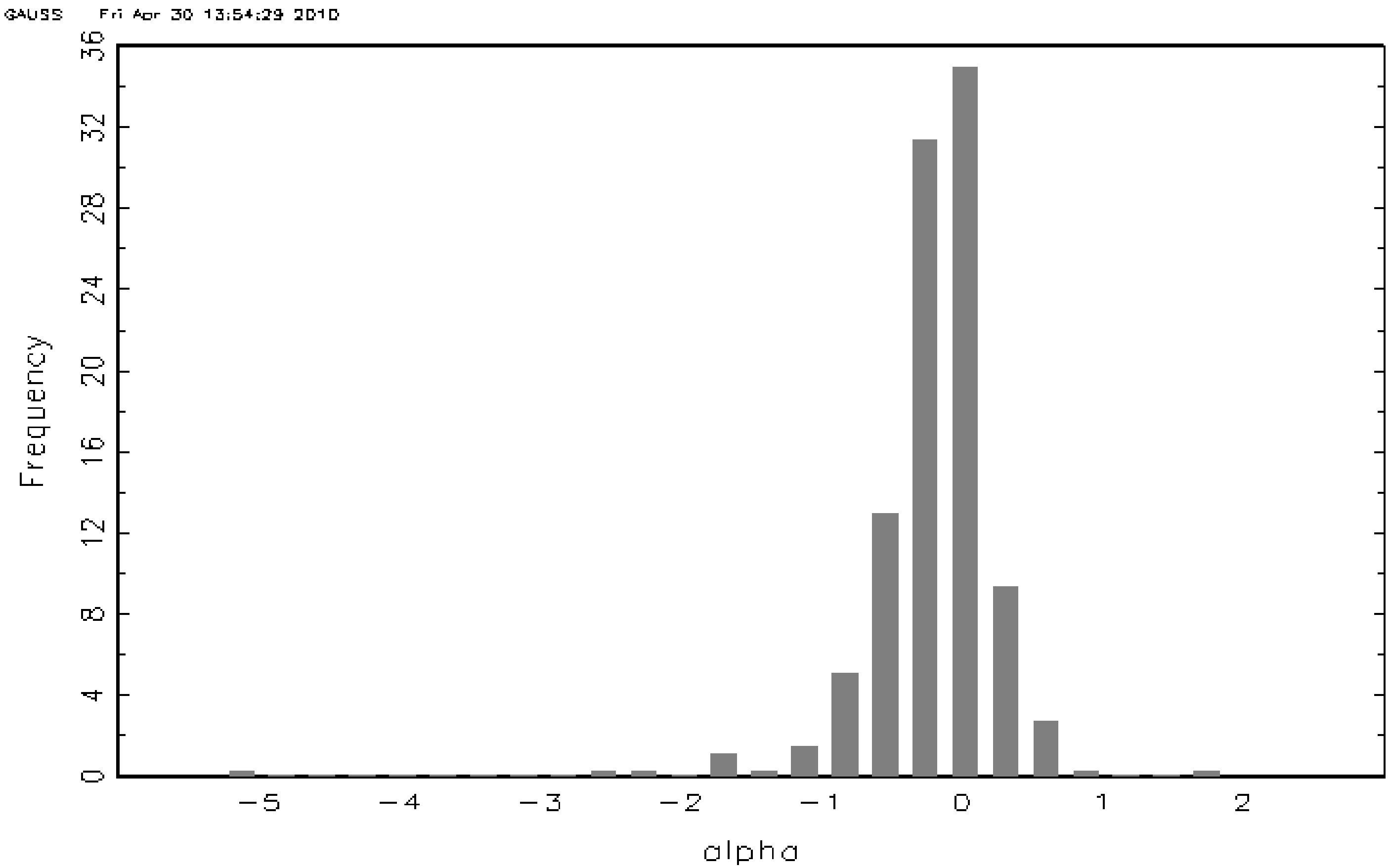

Figure 2 Calculation of π_0: p-values from 3F Model

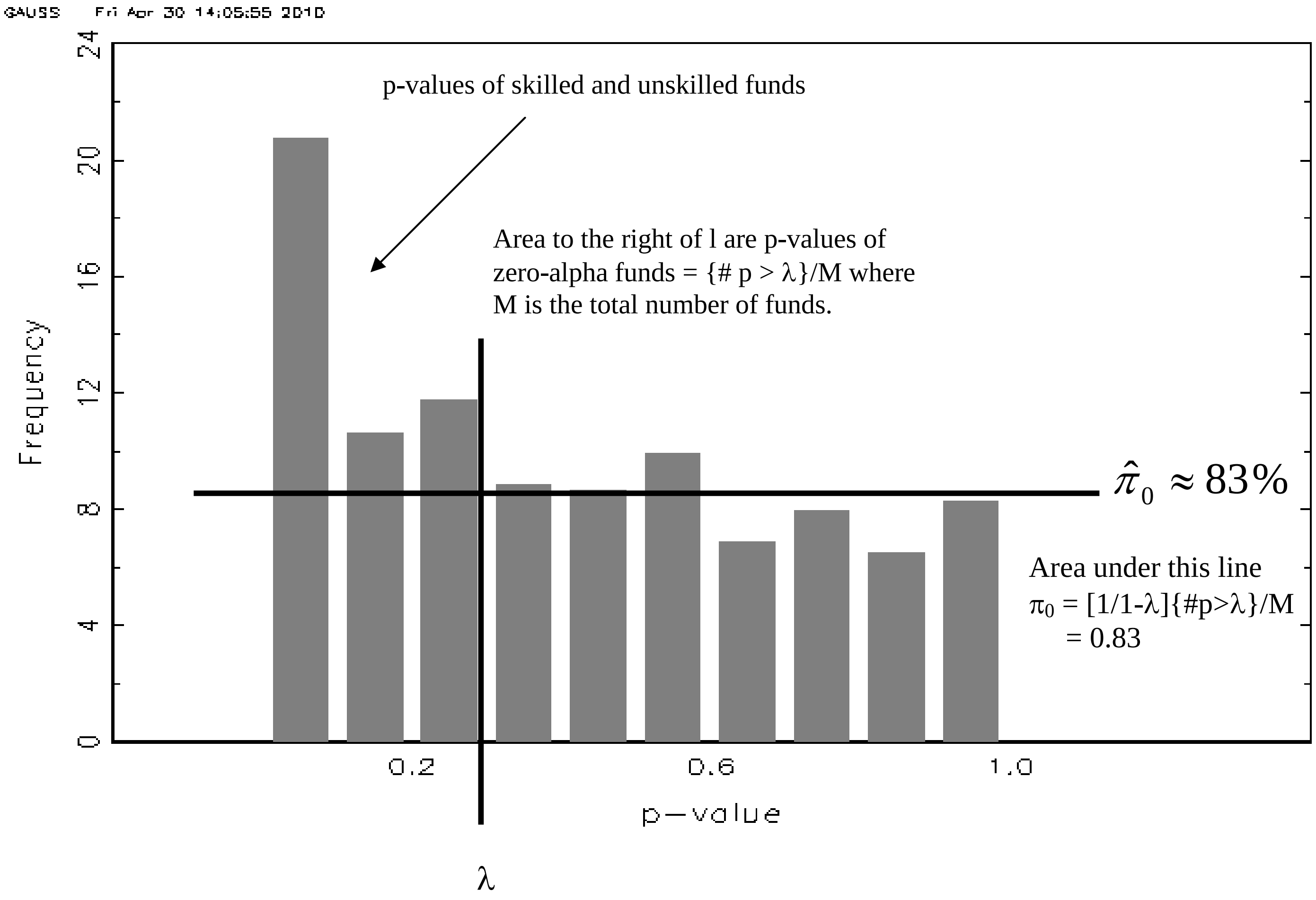

Figure 3 Correlation coefficient: alphas and TM timing coefficients (Rolling Window)

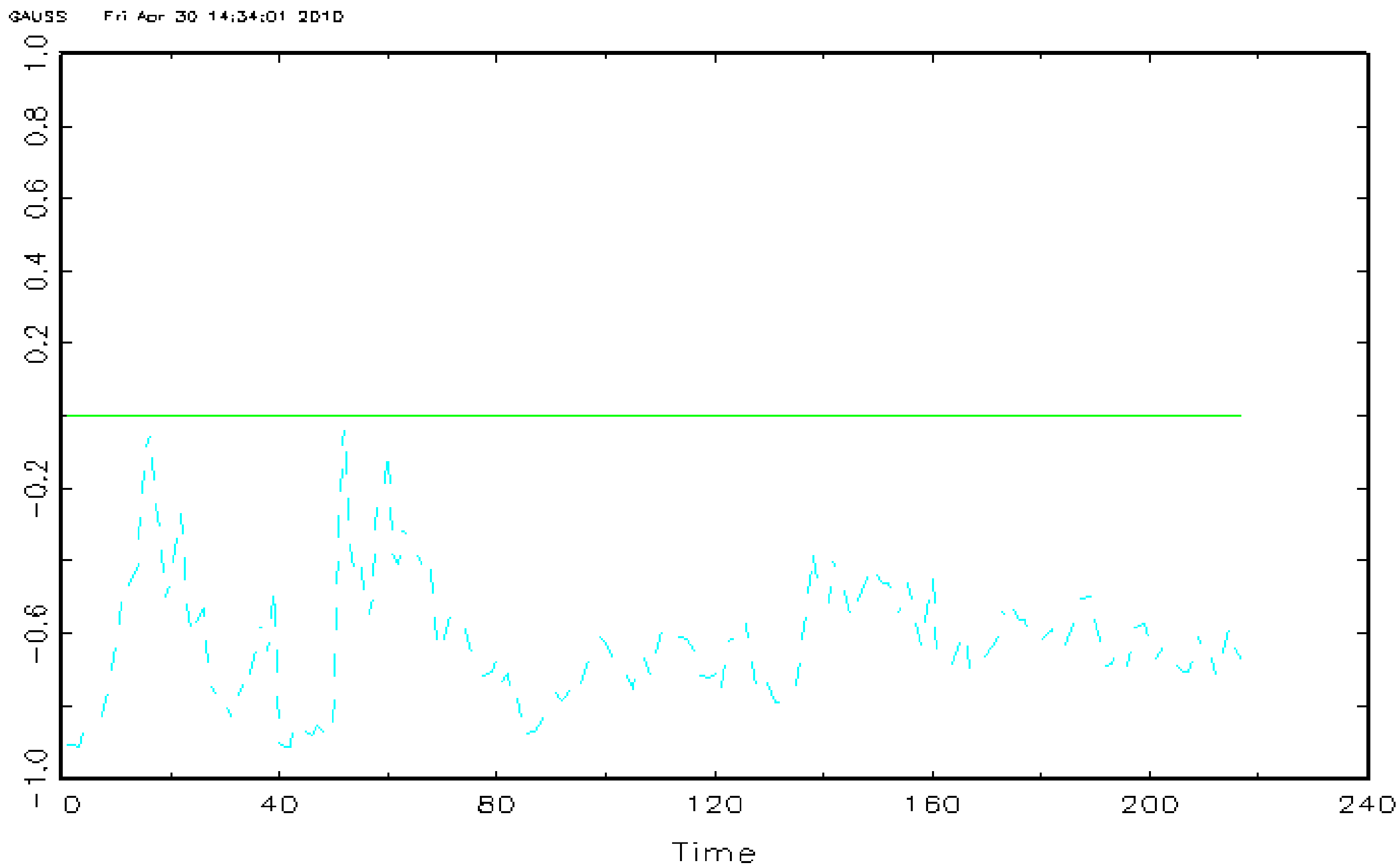

Figure 4 Correlation Coefficient: alphas and HM timing coefficients (Rolling Window)