

Cuthbertson, K. & Nitzsche, D. (2013). Winners and losers: German equity mutual funds. *European Journal of Finance*, 19(10), pp. 951-968. doi: 10.1080/1351847X.2012.684098



**CITY UNIVERSITY
LONDON**

[City Research Online](#)

Original citation: Cuthbertson, K. & Nitzsche, D. (2013). Winners and losers: German equity mutual funds. *European Journal of Finance*, 19(10), pp. 951-968. doi: 10.1080/1351847X.2012.684098

Permanent City Research Online URL: <http://openaccess.city.ac.uk/16983/>

Copyright & reuse

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

Versions of research

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

Enquiries

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at publications@city.ac.uk.

Winners and Losers: German Equity Mutual Funds

Keith Cuthbertson* and Dirk Nitzsche*

This version: 7th March 2012

Abstract:

We investigate the performance of winners and losers for German equity mutual funds (1990-2009) using empirical order statistics. When using gross returns and the Fama-French three factor (3F) model, the number of statistically significant positive-alpha funds is zero but increases markedly when market timing variables are added. However, when using a “total performance” measure (which incorporates both alpha and the contribution of market timing), the number of statistically significant winner funds falls to zero. The latter is consistent with bias in estimated alphas in the presence of market timing. We also find that many poorly performing funds are unskilled rather than unlucky.

Keyword : Mutual fund performance, order statistics.

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 two anonymous referees whose suggestions substantially improved the paper. We also thank David Barr, Don Bredin, Andrew Clare, Steven Thomas, Anil Keswani, Ian Marsh, Michael Moore, Mark Taylor, Lorenzo Trapani, Giovanni Urga and seminar participants at Barclays Global Investors, the Investment Managers Association Conference: “Challenges for Fund Management”, the Western Economic Association International, WEAI 2011 and the World Finance Conference 2011, for discussions and comments.

Winners and Losers: German Equity Mutual Funds

1. Introduction

The performance of individual US and UK mutual funds has been extensively analyzed. On balance, studies find relatively weak evidence of positive alpha performance (“selectivity”) in the extreme right tail of the cross-section performance distribution and relatively strong evidence of negative alpha performance throughout the left tail (e.g. Malkiel 1995, Kosowski et al 2006, Cuthbertson et al 2008, 2010a, Fama and French 2010). Successful market timing involves fund managers increasing (decreasing) their fund’s market-beta in anticipation of an increase (decrease) in market returns. Evidence for successful market timing in both parametric and non-parametric studies has been found (for the US see for example, Treynor and Mazuy 1966, Henriksson and Merton 1981, Ferson and Schadt 1996, Busse 1999, Becker et al 1999, Wermers 2000, Bollen and Busse 2001, Jiang 2003, Swinkels and Tjong-A-Tjoe 2007, Jiang et al 2007, Chen and Liang 2007 and for the UK see Chen et al 1992, Fletcher 1995, Leger 1997, Byrne et al 2006, Cuthbertson et al 2010b).

There has been little work done on analysing the performance of the German mutual fund industry, in particular regarding the market timing ability of fund managers. Although the German mutual fund industry is small compared to the US, it has seen substantial growth over the last 15 to 20 years and its assets under management peaked in 2007 at \$372bn, but then dropped 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.

Krahnert et al (2006) analysed 13 German domestic equity funds (1987-1998), Stehle and Grewe (2001) analysed 18 equity funds (1973 to 1998), Grieser and Kempf (2003) use 118 equity funds (1980 to 2000) and all find no positive statistically significant alphas. 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 an (unconditional and conditional) CAPM, a 3 factor Fama-French model and an SDF model on 50 German domestic equity funds (1994-2003) and find statistically significant underperformance and no outperformance, particularly for the SDF model. None of these studies examines market timing.

Many newspapers and trade journals present performance results in the form of league tables, so they too emphasize funds in the tails of the cross-section distribution. This gives rise to two major problems. First, because we are dealing with ordered/ranked funds, the performance distribution for a particular ranked fund (e.g. the best, or 2nd best, etc) differs from that of the parent distribution. For example, if the cross-section of funds' "true" alphas are normally distributed with a mean of zero, and a sample of n -funds is drawn from this distribution, then the distribution of the fund with the largest alpha (i.e. the best performing fund) will be non-normal with a positive mean. Second, if the performance statistic (alpha) for different funds have unknown and different underlying distributions, then the performance distribution of a particular percentile fund (e.g. the best fund) has to be obtained empirically¹. The contribution of this paper is to derive the empirical distributions for *individual* funds in the tails of the performance distribution, for a large number of German equity mutual funds using 20 years (1990-2009) of monthly data, for alternative factor models (including market timing effects) and bootstrap procedures. We use both alpha ("selectivity") as our performance measure and a measure of "total performance" ($perf_i$) which combines both the fund's alpha and the contribution of market timing to fund returns.

When using gross returns and the Fama-French three factor (3F) model, the number of statistically significant positive alpha funds is zero but increases markedly when market timing variables are added. However, when using a "total performance" measure (which incorporates alpha and the contribution of market timing), the number of statistically significant winner funds falls to zero. The latter is consistent with bias in estimated alphas in the presence of market timing. Our results therefore suggest that extreme caution should be used when assessing skill purely in terms of "selectivity" when market timing is present and that in these circumstances a better and more robust metric is "total performance". We also find that many poorly performing funds are genuinely unskilled (rather than unlucky) when using either selectivity (alpha, in the 3F-model) or a total performance measure (in a market timing model).

Results for winner funds are consistent with the Berk and Green (2004) competitive equilibrium model (with decreasing returns to scale based on *fund size*) - but the presence of many statistically significant unskilled funds is not. However, the latter result may be partly rationalized in the theoretical model of Pastor and Stambaugh (2010) where decreasing returns to scale apply to the size of the active fund *industry* as a whole. They find that even when industry average performance is negative, the size of the active mutual fund industry may remain large

¹ The theory of order statistics deals with the relationship between the underlying distribution of possible outcomes (e.g. normal) and the distribution of say the maximum value from a sample of size n . Analytic results are only available if the underlying distributions are known.

and also far from its equilibrium. In such circumstances it may be possible to observe large statistically significant negative performance statistics over considerable time periods because learning about the true parameters governing decreasing returns to scale and hence future performance is slow – in short, inertia in information processing results in investors continuing to hold poorly performing funds².

2. Methodology and Performance

A cross-section bootstrap procedure is used to separate ‘skill’ from ‘luck’ for *individual ranked* funds, when idiosyncratic risks are highly non-normal (Kosowski et al 2006). Consider an estimated model of equilibrium returns of the form: $r_{i,t} = \hat{\alpha}_i + \hat{\beta}_i' X_t + e_{i,t}$ for $i = \{1, 2, \dots, n$ funds $\}$, where T_i = number of observations on fund- i , $r_{i,t}$ = excess return on fund- i , X_t = vector of risk factors and $e_{i,t}$ are the residuals.

Our ‘basic bootstrap’ under the null of no outperformance is as follows. First, estimate the chosen model for each fund (separately) and save the vectors $\{X_t, \hat{\beta}_i, e_{i,t}\}$. Next, for each fund- i , draw a random sample (with replacement) of length T_i from the residuals $e_{i,t}$. Use these *re-sampled* bootstrap residuals $\tilde{e}_{i,t}$ (and their corresponding X_t values), to generate a simulated excess return series $\tilde{r}_{i,t}$ for fund- i , under the null hypothesis ($\alpha_i = 0$) that is, $\tilde{r}_{i,t} = 0 + \hat{\beta}_i' X_t + \tilde{e}_{i,t}$. This is repeated for all funds in our sample. This gives simulated returns for all funds under the null of zero alphas, for the first run of the bootstrap (B=1).

Next, using the simulated returns $\tilde{r}_{i,t}$ for each fund, the performance model is estimated fund-by-fund and the resulting estimates of alpha (say) $\tilde{\alpha}_i^{(1)}$ ($i = 1, 2, 3, \dots, n$) are obtained (for the first bootstrap, B=1). The $\tilde{\alpha}_i^{(1)}$ estimates for each of the n -funds represent sampling variation around a true value of zero (by construction) and are entirely due to ‘luck’. The $\tilde{\alpha}_i^{(1)}$ $\{i = 1, 2, \dots,$

² Pastor and Stambaugh’s (2010) model focuses on the relationship between the size of the mutual fund industry and *industry average* performance, rather than on individual fund performance. It is therefore likely that persistently poor performance by some funds also requires an assumption of “frictions” in the ability to switch from poorly performing to potential winner funds. Other explanations of long-term negative performance include pure inertia, influences of broker and manager advertising and tax considerations (Gruber, 1996) or negative performance as a payoff for countercyclical performance (Glode, 2009) or that investors buy and sell *actual* index funds at the wrong time, and passive indices do not reflect this (Savov 2009).

n) are then ordered from highest to lowest, $\tilde{\alpha}_{\max}^{(1)}$ to $\tilde{\alpha}_{\min}^{(1)}$ - these are the n -values of alpha from the 1st run of the bootstrap. The above process is repeated $B = 1,000$ times for each of the n -funds which gives a separate ‘luck distribution’ for each of the ordered funds $f(\tilde{\alpha}_i)$ in the performance distribution - from the best alpha-performer to the worst alpha-performer, all of which are solely due to luck.

For example, the 1,000 values for $\tilde{\alpha}_{\max}^{(B)}$ (for $B = 1, 2, 3, \dots, 1000$) represent the values of $\tilde{\alpha}_{\max}$ which occur by chance under the null that *all funds* have zero alphas - this “empirical” null distribution $f(\tilde{\alpha}_{\max})$ can be represented in a histogram. We can then compare the estimated value of alpha $\hat{\alpha}_{\max}$ for our “top ranked fund” using *actual* returns data, with its appropriate ‘luck distribution’. If $\hat{\alpha}_{\max}$ is greater than the 5% upper tail cut-off point of $f(\tilde{\alpha}_{\max})$, we reject the null that its performance is due to luck (at 95% confidence) and infer that the fund has skill.

This above procedure can be applied to a fund at any percentile of the performance distribution, right down to the ex-post worst performing fund. A key element of the approach is that under the null of zero alpha, we do *not assume* the distribution of the estimated alpha for each fund is normal. Each fund’s alpha can follow any distribution (depending on the fund’s residuals) and this distribution can be different for each fund. Hence the distribution under the null $f(\tilde{\alpha}_{\max})$, encapsulates all of the different individual fund’s empirical ‘luck distributions’ (and in a multivariate context this cannot be derived analytically from the theory of order statistics). We can also repeat the above bootstrap analysis for the *t-statistic* of alpha $t_{\tilde{\alpha}_i}$ which gives more robust inference in the extreme tails (Kosowski et al 2006)³.

Our alternative performance models are well known ‘factor models’. The Fama and French (1993) 3F-model is:

$$[1] \quad r_{i,t} = \alpha_i + \beta_{1i} r_{m,t} + \beta_{2i} \text{SMB}_t + \beta_{3i} \text{HML}_t + \varepsilon_{i,t}$$

³ t_{α_i} is a “pivotal statistic” and has better sampling properties than α_i - the obvious reason being that the former ‘corrects for’ high risk-taking funds (i.e. σ_{ε_i} large), which are likely to be in the tails. If different funds have different distributions of idiosyncratic risk (e.g. different skewness and kurtosis) then we cannot say a priori what the distribution of $f(t_{\alpha_i})$ will be – this is why we use the cross-section bootstrap. Fama and French (2010) bootstrap on $(r_{it} - \hat{\alpha}_i)$ across all funds- i with the same time subscript and therefore incorporate any contemporaneous correlations in the residuals across

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 value factors. The Fama and French (1993) 3F-model has mainly been applied to UK funds (e.g. Blake and Timmermann 1998, Quigley and Siquefield 2000, Tonks 2005) and German funds (e.g. Bessler et al 2009, Otten and Bams 2002) whereas for US funds the momentum factor (Carhart 1997) is usually found to be statistically significant. 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, $r_t = \alpha + \beta_t r_{m,t} + e_t$ with $\beta_t = \beta_0 + \delta r_{m,t} + v_t$, which results in the TM estimation equation:

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

The Hendricksson-Merton (HM, 1981) model assumes the market beta depends on the directional response of the market, $\beta_t = \beta_0 + \delta(I_t^+) + v_t$ where $I_t^+ = 1$ when $r_{m,t} > 0$ and zero otherwise, which results in the HM estimation equation:

$$[3] \quad r_t = \alpha + \beta_0 r_{m,t} + \delta f[r_{m,t}] + \varepsilon_t \quad \text{where} \quad f[r_{m,t}] = I_t^+ r_{m,t}$$

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*. Biases in estimating selectivity (alpha) and market timing δ , when the HM (TM) model is true but the TM (HM) model is estimated, are 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.

$$[4] \quad \text{perf}_i = (1/T) \sum_{t=1}^T (\alpha_i + \delta_i f[r_{m,t}]) = \alpha_i + \delta_i \overline{f[r_{m,t}]}$$

Total performance perf_i measures the average abnormal return from both security selection (α_i) and the ability to successfully time the market $\delta_i > 0$ - since the average abnormal return $\bar{r}_i - \beta_i \bar{X}_t = \alpha_i + \delta_i \overline{f[r_{m,t}]}$. Measuring security selection (alpha) without simultaneously

_____ funds – our reported results are invariant to this alternative bootstrap procedure as the contemporaneous correlations in our data are small.

considering the effect on fund performance of market timing effects, can give a misleading picture of overall performance. Clearly, good security selection $\alpha_i > 0$ together with negative market timing $\delta_i < 0$ (or vice versa), may not be beneficial for investors (relative to investing in a passive portfolio). Inclusion of market timing in the 3F model is straightforward.

We test $H_0: \text{perf}_i = 0$ for each ranked fund using our cross-section bootstrap procedure and a joint hypothesis test on (α_i, δ_i) . For the 3F-market timing model, we generate simulated returns $\tilde{r}_{i,t}$ for each fund, by bootstrapping on the residuals under the restriction $\alpha_i + \delta_i \overline{f[r_{m,t}]} = 0$ for all funds. The simulated returns $\tilde{r}_{i,t}$ under the null, are then used to re-estimate the 3F versions of equation [2] or [3] for all n-funds, to obtain values of $\text{perf}_i^{H_0} = \tilde{\alpha}_i + \tilde{\delta}_i \overline{f[r_{m,t}]}$ for each fund- i ($i = 1, 2, 3, \dots, n$). The values of $\text{perf}_i^{H_0}$ for all n-funds are then ranked. For example, for the best performing fund we take largest value $\text{perf}_{\max}^{H_0}$ as our first bootstrap value ($B=1$). We repeat the above procedure $B = 1,000$ times and obtain 1,000 values for $\text{perf}_{\max}^{H_0}$ which are solely due to random variation around the null of zero total performance for all funds - this gives us the null distribution $f(\text{perf}_{\max}^{H_0})$ for the best ranked fund. Using actual fund returns we estimate $\text{perf}_i^{\text{data}} = \hat{\alpha}_i + \hat{\delta}_i \overline{f[r_{m,t}]}$ for each fund and find the largest value $\text{perf}_{\max}^{\text{data}}$, which is then compared to the 5% cut-off point of the ordered null distribution, $f(\text{perf}_{\max}^{H_0})$.

3. Data and Empirical Results

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 at least 24 monthly observations)⁴ of which 85 invest solely in German equities, with the remainder investing outside Germany (“European” and “Global”). Gross returns are returns to the fund (i.e. before deduction of expenses) while net returns are (before-tax) returns to investors (i.e. after deduction of management fees).

⁴ The data set is from Bloomberg and consists of over 600 funds. This is reduced to just 555 after stripping out second units and funds with less than 2 years of data history.

Our factors are measured in the standard way. For funds with German, European and Global geographic mandates we have used the appropriate MSCI total return indices. For each geographical mandate, the SMB variables have been calculated by subtracting the total return index of the small cap MSCI index from the relevant market index. 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 region⁵. 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 Euro).

[Table 1 - here]

Table 1 (Panel A) shows that by limiting our analysis to funds with $T \geq 24$ observations we discard about 85 funds in our complete sample of 619 funds, of which about 45 of the funds discarded are “live” and 40 are “dead” funds and those discarded are mainly from funds invested with a European and Global mandate, rather than funds which invest in German stocks. Our sample of 555 funds (with $T \geq 24$) consists of 364 “surviving funds” and 119 “dead funds”. Average management fees and the spread of fees across funds and fund styles are similar and are also reasonably constant over time (Table 1, Panel B).

Table 2 reports summary statistics using net returns, for the Fama and French 3-factor model and the 3F model augmented with either the TM or HM market timing variables. For each model, cross-sectional (across funds) average statistics are calculated for all funds. 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*. 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 (bottom half of table 2) – thus motivating the use of bootstrap procedures⁶. Around 545 funds (from our 555) have statistically significant positive market betas (10% significance level). For the SMB factor around 420 funds are significantly positive while 17 funds have negative and statistically significant SMB-betas. The number of significant positive HML-betas is 103, with 247 having significant negative betas. Hence many more German funds invest in small rather than large stocks and are “growth orientated” rather than value orientated. For the 3F+TM model, we have 60 (158) funds with a

⁵ We do not have data on the Carhart (1997) momentum factor for those German domiciled funds who invest only in Germany and the much larger number who invest outside Germany.

⁶ The above results also apply when gross returns (i.e. returns before deduction of fees) are used. This is because fees are relatively constant over time.

statistically significant positive (negative) market timing coefficient δ_i^7 . We concentrate on results from the 3F model and the 3F+TM *market timing* model. (Results from the 3F+HM model are qualitatively similar and are not reported).

[Table 2 - here]

Winner Funds

The average management fee is 1.22% p.a. with a standard deviation (across funds) of 0.46 and is fairly constant over time. Below, we report results using gross returns (i.e. before deduction of management fees) – if a fund cannot achieve a statistically significant positive performance or has a negative performance in terms of gross returns, then such funds provide an even worse performance in terms of net returns to investors. After applying the cross-section bootstrap there are about 250 (out of 555) funds with positive alpha-performance statistics, across the different specifications. Table 3 (Panel A) reports alpha and t-alpha statistics (together with their bootstrap p-values) for funds at chosen percentiles, after ranking funds by each of these performance statistics, using the unconditional 3F model. “Alpha sort” is the value of alpha for a specific fund at a chosen percentile, after all funds’ alphas have been sorted from highest to lowest. “t-alpha sort” is defined analogously for the t-statistic of alpha. The 3F model gives no *statistically significant* positive-alpha funds (at a 5% significance level) – whether we use alpha or t-alpha as our performance statistic (Table 3, Panel A).

We now examine alpha-performance in the 3F model with the addition of the TM market timing variable, $r_{m,t}^2$. First, there is a dramatic increase in the number of statistically significant positive-alpha funds to around 200-240, and in the size of the alphas. This is illustrated in figure 1 where the kernel density for the estimated (non-ordered) alphas in the 3F+TM model (dashed line) lies to the right of the alpha estimates in the 3F-model (solid line).

We can compare the alpha and t-alpha performance of individual funds at particular percentiles of the performance distribution using the 3F model (Table 3, Panel A) and the 3F+TM market timing model (Table 3, Panel B). For example, using the 3F model (Table 3, Panel A), the fund ranked by alpha at the 10th percentile (“top 10%”) the value of alpha is 3.8% p.a. ($p = 0.97$) but this increases to 6.9% ($p = 0.0007$) in the 3F+TM model (Panel B). Using t-alpha as the performance measure a

⁷ 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 (Warther, 1995, Ferson and Warther 1996 and Edelen 1999). In addition, spurious timing effects can arise from option-like characteristics (Jagannathan and Korajczyk 1986), and interim trading (Goetzmann et al 2000, Ferson and Khang 2002), while “artificial timing bias” can arise even in “synthetic passive portfolios” (Bollen and Busse 2001). Hence we cannot rule out the possibility that some of our timing coefficients may be spurious.

similar increase occurs, as we add the TM market timing variable to the 3F model. For example, taking the fund ranked by t-alpha at the 10th percentile, the value of t-alpha is 1.16 ($p = 0.99$) for the 3F model (Panel A) whereas for the market timing model t-alpha increases to 1.70 ($p < 0.0001$). This pattern of results occurs for all positive alpha funds reported in Table 3. It would appear that the market timing model (Panel B) provides much stronger evidence of successful security selection skills than the 3F model (Panel A).

[Figure 1 – here]

[Table 3 - here]

However, using the 3F+TM model, and our measure of total performance - which combines the effect on returns of both security selection and market timing - we again find no (statistically significant) skilled funds whether we rank funds by perf_i or t_{perf} . Table 3 for the 3F+TM model, shows that for funds at specific percentiles, positive alpha-performance is prevalent (Panel B) but positive performance based on our total performance statistic perf_i is not (Panel C). For example, for the fund ranked at the 10th percentile (“Top 10%”) we have $\text{perf} = 0.32\%$ p.m. ($p = 0.99$) and $t_{\text{perf}} = 1.18$ ($p = 1.0$) which suggests no skill. Figure 2, reinforces this result by comparing the ranked performance of *all funds* with positive values of alpha or positive values of perf , for the 3F+TM model. Figure 2 shows for each ranked fund, the estimated performance statistic (solid line) using actual fund data and the bootstrap 5% critical value under the null of zero performance (dashed line) – for alpha (Panel A) and for total performance perf_i (Panel B), both using the 3F+TM model. Panel A clearly shows a substantial number of statistically significant positive-alpha funds ranked from the 30th best to the 250th best fund – as for these funds the estimated alphas using actual fund data (solid line) exceed their 5% critical values. But when we use perf_i (Panel B), the estimated values of perf_i using actual fund data (solid line) lie to the left of the bootstrap 5% critical values, for all of the ranked funds. Hence, we conclude that “total performance” for all actively managed German equity mutual funds is solely due to luck, not skill.

[Figure 2 - here]

Overall, we conclude that there are no statistically significant “winner funds”, even before deduction of management fees. This is because we discount the statistically significant results on

selectivity (alpha) in the 3F+TM model which are subject to potential bias and because perf_i is a sensible performance metric in the presence of market timing⁸.

Loser Funds

Using gross returns we have a large number of funds that are unskilled whatever performance metric or factor model we use. At a 5% significance level, we find 307 unskilled funds based on the ordered bootstrap t-alpha of the 3F-model, 43 based on the order bootstrap t-alpha of the 3F+TM model and 203 based on the bootstrap t-statistic of perf_i - out of a total of 555 funds. These performance results are model dependent but potential biases in alpha when market timing is present, suggests concentrating on the negative alphas from the 3F model (Table 3, Panel A) or negative values for our total performance measure, perf_i (Table 3, Panel C).

For the 3F-model, the ranked t_α 's for nearly all the percentile funds reported in the right-hand side of Panel A are negative and statistically significant (based on bootstrap p-values). However, for the (3F+MT) model, funds located at specific percentiles reported for the worst performers (Table 3, Panel C) are not representative of the statistical significance of funds with negative values of t_{perf} , across all funds in our sample. As can be seen in figure 3, for the 3F+MT model there are a substantial number of statistically significant negative t_{perf} funds in the left tail up to 438th ranked fund - as the actual t_{perf} values (solid line) for these funds lie below their 5% bootstrap critical values (dashed line)⁹.

[Figure 3 here]

For alpha-performance in the 3F model and total performance in the 3F+TM model both performance measures clearly show that many loser funds are unskilled rather than unlucky.

⁸ Using net returns (i.e. after deduction of management fees) this result applies a fortiori for alpha in the 3F model and for perf_i in the 3F+TM model. Hence potential "winner" funds with either positive alpha or positive perf_i are merely lucky rather than exhibiting true (statistically significant) outperformance, for all ordered funds – these results are available on request.

⁹ Figure 3 also shows that funds in the *very extreme left tail* with negative t_{perf} experience bad luck rather than being unskilled, as their t_{perf} statistic is not statistically different from zero (at a 5% significance level). Using net returns to investors, the equivalent results for negative performers for all three models (not reported) have p-values less than 0.01 – hence not surprisingly, for investors, there are a substantial number of loser funds which are genuinely unskilled rather than unlucky.

Despite the existence of low cost passive funds (either constructed from sector index mutual funds or ETFs), German investors continue to hold a large number of active funds which deliver a statistically significant negative abnormal performance (either in terms of alpha or total performance) – competition for investment funds does not appear to remove poorly performing funds from the marketplace. This may have serious consequences as Germany moves from a predominantly state provided pensions system to pensions based (in part) on stock market performance.

5. Conclusions

At a methodological level, our results suggest that one should not assess “skill” purely in terms of “selectivity” (alpha) when market timing is present, since estimation bias is likely to be substantial. In the presence of market timing a more useful and robust metric is “total performance” (which incorporates selectivity and the contribution to returns of successful market timing).

Comparing results on security selection (alpha) in the 3F model (i.e. excluding market timing) with results using our measure of total performance $perf_i$ in the 3F+TM model, we find that even in the tails of the performance distribution both measures give a consistent picture for German equity mutual fund investors. For funds with an estimated positive net return performance (to investors), either due solely to alpha or in terms of total performance (in the presence of market timing) - all funds are merely lucky rather than exhibiting skill. But for nearly all funds with negative estimated alphas (in the 3F model) or total performance (in the market timing model) these funds are genuinely unskilled, rather than unlucky.

When we add back fees, then for alpha in the 3F model and total performance in the 3F+TM model there are still no skilled funds even in the extreme right tail of the performance distribution but there are a substantial number of funds that are unskilled, delivering statistically significant negative abnormal performance (either alpha or total performance) before fees, and these funds are spread throughout much of the left tail of the performance distribution.

The message for German investors is clear – avoid active equity mutual funds and diversify using index tracker funds or ETFs.

References

- 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.
- 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, 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, 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.
- 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, 2010a, Mutual Fund Performance – Measurement and Evidence, *Journal of Financial Markets, Institutions and Instruments*, Vol. 19, No. 2, pp. 95-187.
- Cuthbertson, Keith, Dirk Nitzsche and Niall O'Sullivan, 2010b, The Market Timing Ability of UK Mutual Funds, *Journal of Business, Finance and Accounting*, Vol. 37 (1&2),

pp. 270-289

Edelen, Roger M., 1999, Investor Flows and the Assessed Performance of Open-End Mutual Funds, *Journal of Financial Economics*, 53, 439-466.

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 Alpha Estimates, *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.

Ferson, Wayne E. and Vincent A. Warther, 1996, Evaluating Fund Performance in a Dynamic Market, *Financial Analysts Journal*, Vol. 52 (6), pp. 20-28.

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.

Glode, Vincent, 2009, Why Mutual Funds "Underperform", Working Paper, Wharton.

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, 2003, Lohnt Aktives Fondsmanagement Anglengersicht? Ein Vergleich von Anlagestrategien in Activ und Passiv Verwalteten Aktienfonds, *Zeitschrift fur Betriebswirtschaft*, Vol. 73, pp. 201-224.

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.

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.

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.

Krahnhen, 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.

- Leger, L., 1997, UK Investment Trusts : Performance, Timing and Selectivity, *Applied Economics Letters*, Vol. 4, pp. 207-210.
- Malkiel, G., 1995, Returns from Investing in Equity Mutual Funds 1971 to 1991, *Journal of Finance*, Vol. 50, pp. 549-572.
- Otten, Roger and D. Bams, 2002, European Mutual Fund Performance, *European Financial Management*, Vol. 8(1), pp 75-101.
- Pastor, Lubos and Robert F. Stambaugh, 2010, On the Size of the Active Management Industry, August 26th, SSRN Working Paper.
- Quigley, Garrett, and Rex A. Siquel, 2000, Performance of UK Equity Unit Trusts, *Journal of Asset Management*, Vol. 1, pp. 72-92
- Savov, Alexi., 2009, For a Fee: The Hidden Cost of Index Fund Investing, Working Paper, University of Chicago.
- Stehle, Richard and Olaf Grewe, 2001, The Long-Run Performance of German Stock Mutual Funds, Humboldt University, Berlin, Discussion Paper.
- Swinkels, L. and L. Tjong-A-Tjoe, 2007, Can Mutual Funds Time Investment Styles?, *Journal of Asset Management*, Vol. 8, pp. 123–32.
- Tonks, Ian, 2005, Performance Persistence of Pension Fund Managers, *Journal of Business*, Vol. 78, pp. 1917-1942.
- 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.

Figure 1: Kernel Density: Alpha in 3F and 3F+TM Models

This figure shows the kernel density for the estimated (non-ordered) alphas in the in the 3F-model (solid line) and the 3F+TM model (dashed line).

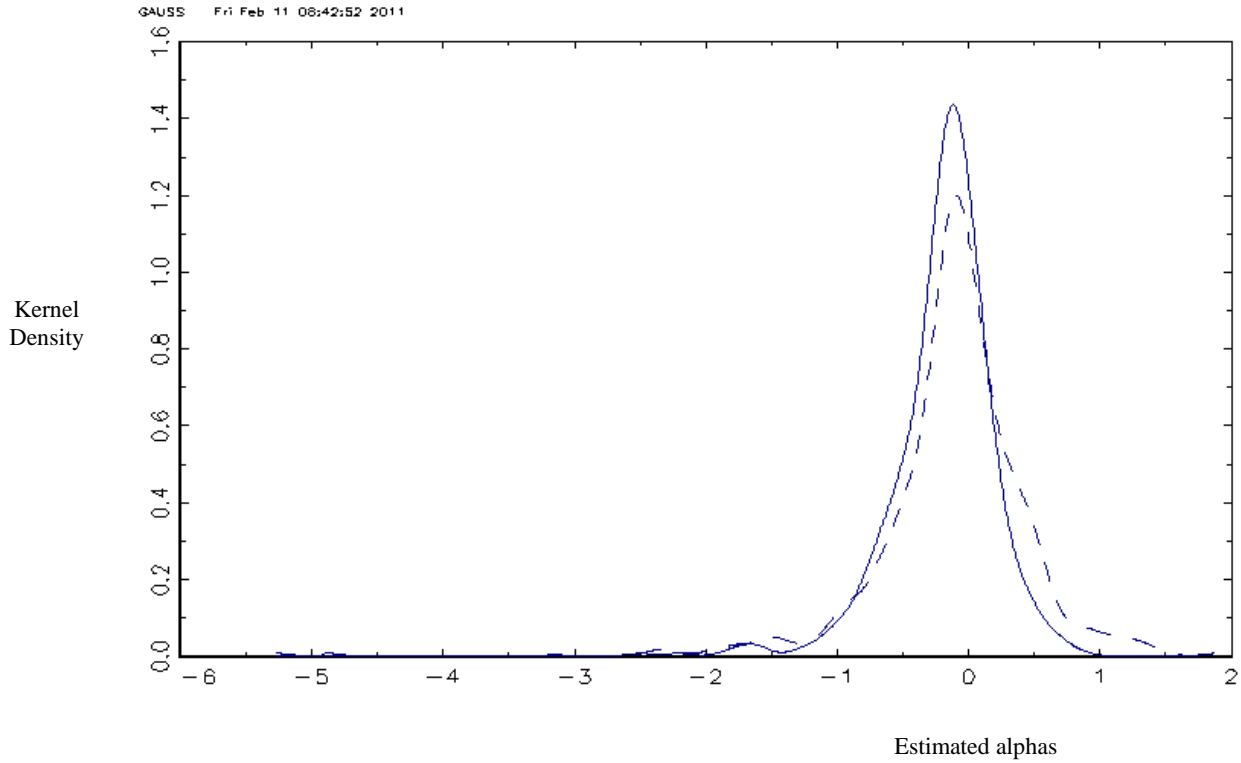


Figure 2. Winner Funds: Alpha versus Total Performance (3F+TM Model)

The figures show each of the two performance measures (Panel A = alpha, Panel B = total performance, $perf_i$) plotted against the ordered funds – both performance measures use the 3F+TM model. The solid lines are the estimated performance measures using actual fund data and the dashed lines are the bootstrap 5% critical values of the null distributions for the ordered funds.

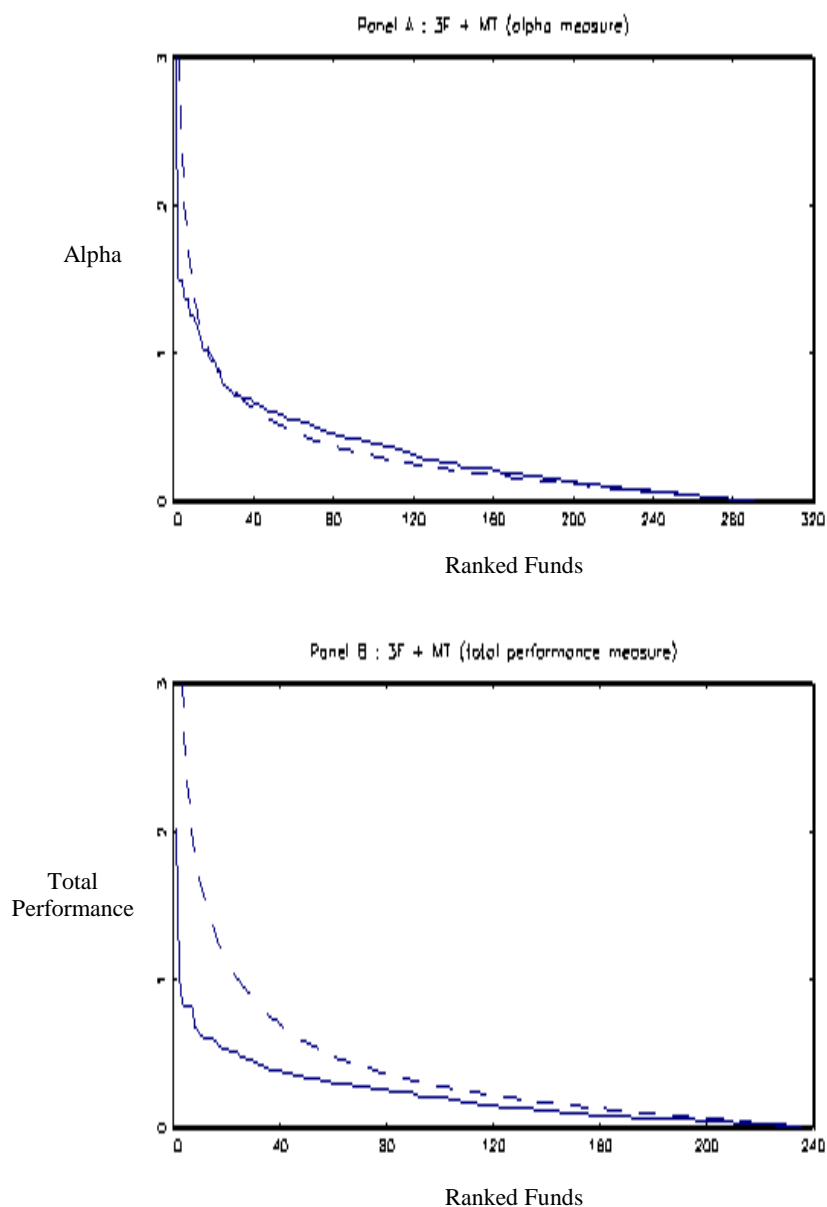


Figure 3. Loser Funds: Total Performance (3F+TM Model)

The figure shows the performance measure t_{perf} plotted against the ordered funds for the 3F+TM model. The solid line is the estimated performance measure using actual fund data and the dashed line is the bootstrap 5% critical value of the null distribution for the ordered funds.

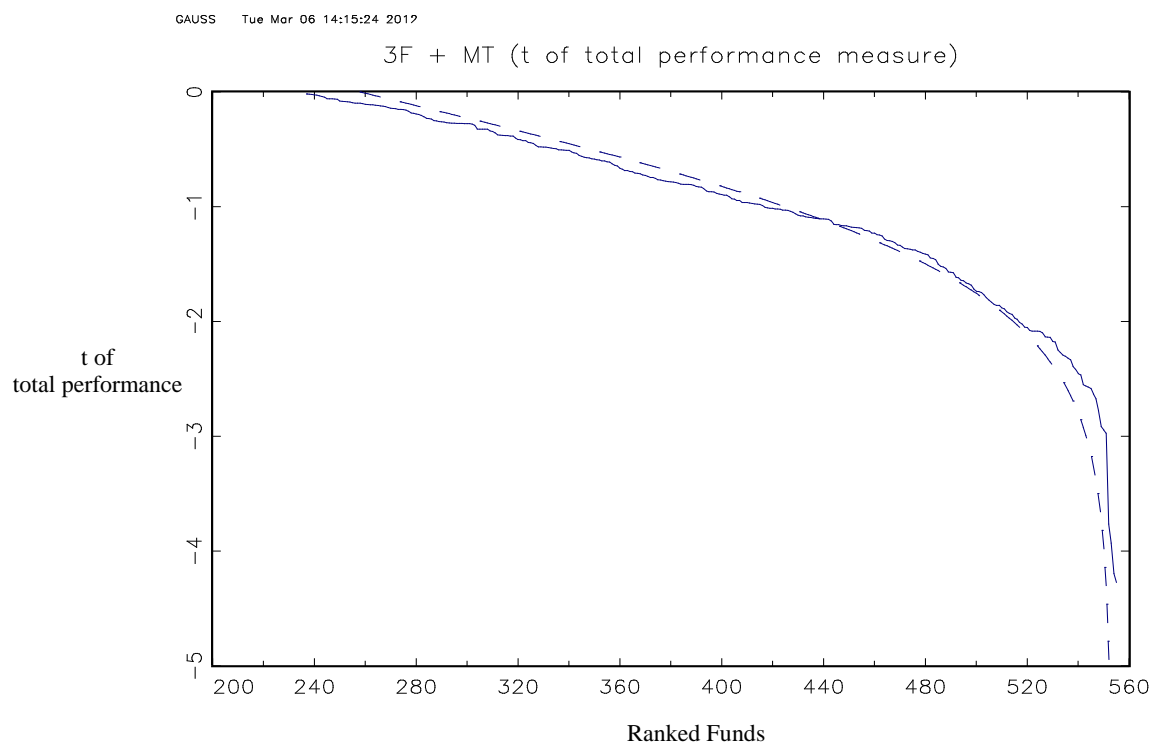


Table 1: Summary Statistics: Funds and Management Fees

Panel A shows the total number of funds, the number of 'live' and 'dead' funds and details of the geographical investment objective. Panel B shows average annual management fees (percentage) and their standard deviation.

Panel A : Number of Funds				
	All Funds	German Funds	European Funds	Global Funds
All Funds				
# funds	619	88	255	276
# funds with at least 24 obs.	555	85	226	244
'Live' Funds Only				
# funds	400	58	159	183
# funds with at least 24obs.	364	57	143	164
'Dead' Funds Only				
# of funds	219	30	96	93
# of funds with more than 23 obs.	191	28	83	80
Panel B : Management Fees (% p.a.) - Mean and Standard Deviation				
	All Funds	German Funds	European Funds	Global Funds
All Funds	1.22 (0.46)	1.22 (0.38)	1.13 (0.46)	1.29 (0.47)
Funds with at least 24 obs.	1.22 (0.46)	1.22 (0.37)	1.13 (0.46)	1.30 (0.43)
'Live' Funds Only	1.22 (0.42)	1.27 (0.38)	1.15 (0.43)	1.26 (0.42)
'Live' funds with at least 24 obs.	1.22 (0.41)	1.26 (0.38)	1.15 (0.43)	1.28 (0.39)
'Dead' Funds Only	1.21 (0.56)	1.03 (0.31)	1.09 (0.52)	1.38 (0.60)
'Dead' funds at least 24 obs.	1.21 (0.53)	1.07 (0.28)	1.11 (0.53)	1.35 (0.55)

Table 2: Factor Models and Market Timing

This table reports summary statistics of all funds used in the analysis. The sample period is January 1990 to December 2009 (monthly data) and includes 555 German domiciled mutual funds which have at least 24 observations. Returns are net of management fees. The average number of observations used is 111 months. We report averages of the individual fund statistics for three different models (3F, and the 3F+TM and 3F+HM market timing models). The first factor is the corresponding excess market return r_m , the second factor is the size factor SMB, and the third factor is the book-to-market factor, HML. t-statistics are based on Newey-West adjusted standard errors. BJ is the Bera-Jarque statistic for normality of residuals.

Panel A : Mean Values of Coefficients and t-statistics			
	3F Model (r_m , SMB, HML)	3F+TM (r_m , SMB, HML, r_m^2)	3F+HM (r_m , SMB, HML, r_m^+)
Alpha (% p.m.) (t-stat)	-0.1955 (-0.7299)	-0.1129 (-0.3009)	-0.0684 (-0.1141)
r_m (t-stat)	0.9668 (18.55)	0.9509 (17.91)	0.9929 (12.52)
SMB (t-stat)	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			
	3F Model (r_m , SMB, HML)	3F+TM (r_m , SMB, HML, r_m^2)	3F+HM (r_m , SMB, HML, r_m^+)
Mean R^2	0.7812	0.7896	0.7879
Skewness	-0.1586	-0.1173	-0.1252
Kurtosis	8.14	7.97	7.99
BJ – statistic	3123.34	3071.70	3077.62
% (Number) funds non-normal residuals	47.38% (263 funds)	50.45% (280 funds)	50.09% (278 funds)

Table3: Ordered Bootstrap Performance Measures: Gross Returns

This table reports performance measures for the full sample of ordered funds using gross returns (i.e. before deduction of management fees). Panel A reports alpha and t-alpha statistics from the unconditional 3F model for various percentiles of the performance distribution, together with their bootstrap p-values. Panel B repeats this for alpha and t-alpha in the 3F+TM market timing model, while Panel C does so for the total performance measure perf and t_{perf} statistics (also using the 3F+TM model). “Fund’s rank” is the numerical position of the fund (out of a total of 555 funds) when a particular performance statistic (e.g. alpha) is used to rank funds from highest to lowest. For example, the fund at the 1st percentile (“Top 1%”) is the fund ranked 6th out of 555 funds. “Alpha sort” is the value of alpha for a specific fund at a chosen percentile, after all funds’ alphas have been sorted from highest to lowest. “t-alpha sort” is defined analogously for the t-statistic of alpha. “Perf sort” and “t-perf sort” are defined analogously to “alpha sort” and “t-alpha sort” - but funds are ranked using the total performance measures perf or t_{perf} . Both actual (ex-post) and bootstrap t-statistics are based on Newey-West adjusted standard errors. The sample period is January 1990 to December 2009 (monthly data) and includes 555 German domiciled mutual funds which have at least 24 observations.

Panel A : Alpha, 3F model								
	Top 1%	Top 5%	Top 10%	Top 20%	Bottom 20%	Bottom 10%	Bottom 5%	Bottom 1%
Fund’s Rank	6	28	56	112	445	500	528	550
Alpha sort (% p.m.)	0.7956	0.4746	0.3159	0.1653	-0.3298	-0.5582	-0.7260	-1.5815
Bootstrap p-value	0.9965	0.9663	0.9723	0.9956	< 0.0001	< 0.0001	0.0040	0.0465
t-alpha sort	2.5179	1.7956	1.1643	0.7188	-1.1398	-1.6404	-2.1103	-2.6126
Bootstrap p-value	0.8872	0.5211	0.9944	0.9956	< 0.0001	< 0.0001	0.0001	0.3974
Panel B : Alpha, 3F plus TM (r_m^2) model								
	Top 1%	Top 5%	Top 10%	Top 20%	Bottom 20%	Bottom 10%	Bottom 5%	Bottom 1%
Fund’s Rank	6	28	56	112	445	500	528	550
Alpha sort (% p.m.)	1.3671	0.7568	0.5727	0.3609	-0.3247	-0.5909	-0.9328	-1.9162
Bootstrap p-value	0.5747	0.0989	0.0007	< 0.0001	0.0003	0.0004	0.0029	0.1120
t-alpha sort	3.1126	2.3586	1.7012	1.1310	-0.9424	-1.3714	-1.7245	-2.4619
Bootstrap p-value	0.0837	< 0.0001	< 0.0001	< 0.0001	0.0459	0.1440	0.3782	0.6984
Panel C : Total Performance, 3F plus TM (r_m^2) model								
	Top 1%	Top 5%	Top 10%	Top 20%	Bottom 20%	Bottom 10%	Bottom 5%	Bottom 1%
Fund’s Rank	6	28	56	112	445	500	528	550
Perf. sort (% p.m.)	0.8115	0.4592	0.3167	0.1598	-0.3237	-0.5599	-0.7377	-1.5379
Bootstrap p-value	1.00	1.00	0.9998	0.9915	0.9915	0.9704	0.9982	0.8493
t-perf. sort	2.5437	1.8393	1.1832	0.6799	-1.1564	-1.7367	-2.1355	-2.9461
Bootstrap p-value	0.9992	0.9914	1.00	1.00	0.9470	0.9279	0.5935	0.1389