Evaluating the Rating of Stiftung Warentest: How good are Mutual Fund Ratings and can they be Improved?

Sebastian Müller and Martin Weber*

Juli 26, 2011

Abstract

We test the abilities of the Stiftung Warentest fund rating system to predict future fund performance among German registered funds for six equity categories: Germany, Euro-Zone, Europe, North-America, Pacific, and World. Stiftung Warentest is a consumer protection agency and a major provider of fund ratings in Germany. Our empirical analysis documents predictive abilities of the rating system. The reason is that measures of past performance are positively related to future performance in several of these markets, even after controlling for momentum. Measures of fund activity are also helpful to predict performance, and in particular to identify likely future losers.

Keywords: mutual funds, performance evaluation, performance persistence, mutual fund ratings, active management

JEL Classification Code: G11, G12, G1

---

*Sebastian Müller is from the Lehrstuhl für Bankbetriebslehre, Universität Mannheim, L 5, 2, 68131 Mannheim. E-Mail: mueller@bank.BWL.uni-mannheim.de. Martin Weber is from the Lehrstuhl für Bankbetriebslehre, Universität Mannheim, L 5, 2, 68131 Mannheim and CEPR, London. E-Mail: weber@bank.BWL.uni-mannheim.de. The authors appreciate helpful comments and suggestions from an anonymous referee, Jieyan Fang, Stephan Jank, Stefan Ruenzi, seminar participants at the University of Mannheim and the 17th annual meeting of the German Finance Association (DGF) in Hamburg. We also thank Stiftung Warentest for providing data on the ratings and fund classification, Morningstar for data on mutual fund portfolio holdings and total expense ratios, and Nico Hemker for excellent research assistance. Special thanks goes to Andrew Patton for providing the code for the monotonic relationship (MR) tests in Patton and Timmermann (2010) on his web-site. Financial support from the Deutsche Forschungsgemeinschaft and Boerse Hamburg and Hannover is gratefully acknowledged.

1 Introduction

Focusing on the U.S. equity fund market a tremendous amount of academic research has examined whether measures of past portfolio performance are informative about future performance. A balanced reading of these papers suggests that there is little evidence of persistence in equity funds' risk-adjusted returns after controlling for survivorship bias and for momentum in stock returns (see e.g., Busse et al. (2010), Carhart (1997), Malkiel (1995), Jensen (1968)). Further, recent studies by Barras et al. (2010), Fama and French (2010), and Cuthbertson et al. (2010) highlight the problem of identifying truly skilled fund managers if one has to rely on a limited sample of historical returns. The results of these papers suggest that even most funds in the extreme right tail of the cross-sectional estimated alpha distribution have been rather lucky than skilled. Consequently, it is not surprising that studies analyzing the value of the Morningstar rating system, which is based on past portfolio performance, find the rating to be a rather poor predictor of future mutual fund performance for U.S. funds (see e.g., Blake and Morey (2000), Morey (2005) and Gerrans (2006)). Moreover, as shown by Kräussl and Sandelowsky (2007), the forecasting abilities of the Morningstar approach have declined over time. However, recent academic work by Amihud and Goyenko (2009), Cremers and Petajisto (2009), Kacperczyk et al. (2005) and Wermers (2003) shows that measures which quantify the degree of active portfolio management are associated with higher risk-adjusted fund returns. These results suggest that investors should consider the extent to which an open-end actively managed fund really pursues an active strategy in order to select well-performing funds.

In this paper, we provide evidence on the value of a mutual fund rating system and other measures of past performance as performance predictors in an international context. Specifically, we test whether the mutual fund rating system of "Stiftung Warentest" is able to differentiate between outperforming and underperforming German registered funds that invest in one of the following six equity categories: Germany, Euro-Zone, Europe, North America, Pacific, and World. In addition, we analyze whether measures of fund activity ($Active\ Share, Tracking\ Error$

and R^2) also predict future fund performance outside the U.S. fund market. Given the results of the above mentioned studies, the degree of active management appears to be one candidate to better differentiate between luck and skill in the mutual fund industry. If so, ratings which are based on historical portfolio returns could potentially be improved by taking these additional measures into account.

Analyzing the quality of the fund rating system of Stiftung Warentest is interesting for several reasons. First, Stiftung Warentest is a major fund rating provider in Germany and it covers the entire German fund market. This allows us to analyze the performance of funds that invest outside the local German market and even funds that invest worldwide, whereas previous studies on mutual fund performance and performance persistence outside the U.S. vastly focus on funds solely investing in their domestic market (see e.g., Otten and Bams (2002), Griese and Kempf (2003), Korkeamaki and Smythe (2004), Stotz (2007), and Cuthbertson et al. (2010)). Second, Stiftung Warentest is a consumer protection agency, which aims at providing independent information on products and services in a broad range of different fields in order to protect consumer interest. Since Stiftung Warentest receives financial support from the German government and its constitution prohibits any advertisements, its mutual fund recommendations should be free of any conflicts of interest which have been documented for other financial advisers (see Reuter and Zitzewitz (2006)). Third, information intermediaries like Stiftung Warentest exist in many countries all over the world. Examples include ”Consumers Union” in the U.S. or ”Which?” in the United Kingdom. It is evident that consumer protection agencies play an important role in generating and disseminating information to the public. This is in particular true for Stiftung Warentest, which enjoys a very high reputation among consumers in Germany. According to a recent survey, 96% of the population above 18 years know the organization, 81% consider the test results as highly reliable and roughly 30% use the recommendations as an orientation guide when buying consumer products or services.[1]

[1] The survey has been conducted by the commercial marketing research institute Forsa in 2007.

Fourth, independent, reliable and easy to understand information might be of particular relevance for private households which want to select a mutual fund. Alexander et al. (1998) show that many mutual fund customers often lack the financial expertise to assess a product's quality. Moreover, the mutual fund industry has seen a rapid growth over the last years in most developed countries.[2] This development has also led to an increased complexity in selecting a fund: As the BVI (Bundesverband Investment und Asset Management e.V.) notes, more than 9200 competing products were available for sale in Germany at the end of 2008.[3]

Our examination of the predictive abilities of Stiftung Warentest shows that the rating is positively correlated with future fund performance. Funds in the lowest rating quintile underperform funds in the highest rating quintile on average by roughly 10 to 20 basis points per month over the next twelve months, depending on the performance measure. The performance spreads between the highest and lowest rating quintiles are statistically and economically significant, and - although they decrease - they do not erode after we control for momentum. Furthermore, using the monotonic relationship test of Patton and Timmermann (2010), we are able to empirically verify a strictly increasing relation between the rating and future fund performance for many fund markets and performance measures under consideration.

The reason for the success of the Stiftung Warentest rating is that performance persists over a short time horizon for several fund categories which we analyze. However, even high rated funds do not deliver returns that are significantly above the returns of their benchmark. As a result, no feasible trading strategy can be build upon solely the rating of Stiftung Warentest in order to generate superior risk-adjusted returns. We test whether fund expenses might explain our results,

---

[2]According to the statistical releases of the Investment Company Institute (ICI), fund assets worldwide amounted to US$ 19.0 trillion at the end of 2008 coming from a little more than US$ 7.6 trillion at the beginning of this century, see Investment Company Institute (2009).

[3]See BVI (2008). The BVI is the central association of the German mutual fund industry and collects information about the German fund market. It is comparable to the ICI in the U.S.

but there are only minor differences in the total expense ratios of high and low rated funds. Finally, we show that measures of the degree of active management are also related to future fund performance. Hence, they may be used as additional predictors to select better performing funds outside the U.S. fund market as well. Our results indicate that taking into account fund activity is particularly useful to separate skill from luck among underperforming mutual funds.

The remainder of the study has the following structure. In section 2, we describe the methodology of the Stiftung Warentest rating, the fund sample and our empirical evaluation approach. Sections 3 and 4 present the major empirical results of this paper. We first analyze the rating's predictive abilities for future fund performance, compare it to alternative measures of past performance, and examine to which extent the different predictors are related to differences in expenses. In Section 4, we then test the potential value of quantifying the degree of active portfolio management in order to identify funds with superior future performance. Section 5 concludes.

2 Data and Methodology

2.1 Stiftung Warentest and Mutual Fund Ratings

The consumer protection agency Stiftung Warentest is a foundation under public-law which was launched by the German government in 1964. The exclusive goal of the foundation is to evaluate consumer products and services in an independent and objective manner and to disseminate information about the quality of different products to the public. By doing so, it aims at enabling consumers to make better purchasing decisions. Mutual fund ratings of Stiftung Warentest can be found in its financial magazine Finanztest, which has a monthly print run of 300,000.[4] Stiftung Warentest receives financial support from the Federal Ministry of Food, Agriculture and Consumer Protection in Germany and its magazines are free of advertisements.

[4] To put this into perspective, ”Der Spiegel”, which achieves the highest circulation among magazines of general interest in Germany, has a weekly circulation of slightly more than one million.

To construct its fund ratings, the rating system classifies funds into different categories based on the asset class and the regional focus of the fund. However, since many of the categories do not contain a meaningful number of funds, the classification scheme differentiates between major fund categories consisting of up to several hundred funds (e.g., equity funds Europe) and other non-major fund categories. In its monthly print issues, Finanztest publishes comprehensive rating results for the highest rated funds of its major fund categories. All other fund ratings are available on the website of Stiftung Warentest. The ratings are based on the net return history of the funds over the previous 5 years (assuming reinvested dividends) and are recomputed monthly. Funds with a return history of less than 5 years do not receive a rating. The agency covers all funds that are available for sale in Germany. During our sample period the overall Stiftung Warentest rating in a given month is the weighted sum of 2 variables and defined as follows:

$$Stiftung\ Warentest\ Score = Performance \cdot 75 + Stability \cdot 25 \qquad (1)$$

$Performance$ expresses the relative performance of a fund compared to its peer group over the last 60 months. More specifically, for fund i at the beginning of month t this variable is computed as:

$$Performance_{i,t} = \frac{\sum_{j=t-61}^{t-1} I_j \cdot |r_{i,j} - r_{group,j}|}{\sum_{j=t-61}^{t-1} |r_{i,j} - r_{group,j}|}, \qquad (2)$$

where $r_{i,j}$ and $r_{group,j}$ are the return of the fund and the average peer group return in month j. I_j is an indicator variable taking the value of 1 (0) whenever the fund return is above (below) the peer group return. Hence, $Performance$ relates the sum of all positive return deviations over the previous 60 months to the sum of all absolute return differences.

To calculate the variable $Stability$ Stiftung Warentest simply divides the number of months in which the fund outperformed its peer group to the total number of months, i.e. $Stability$ measures the fraction of months in which the fund had an above average return. The overall rating for a fund is bounded between 0 and 100. Obviously, the rating system is missing any theoretical foundation. In particular, it does not account for differences in systematic risk exposure. Whether or not it is able to predict future fund performance is at the end an empirical issue, though.

2.2 Mutual Fund Sample

Stiftung Warentest ratings are available for the period from December 2001 to June 2008. We receive on-disk data containing the following variables: *fund name, ISIN, Stiftung Warentest fund category*, *Stiftung Warentest rating*, and *reporting date*. These data are available for dead (liquidated or merged) as well as surviving funds and hence free of a survivorship-bias problem. The data set from Stiftung Warentest is merged with fund return data which is computed using the total return index from Thomson Reuters Datastream (code: *R*I). For mutual funds *R*I measures the hypothetical growth in the funds' net asset value (NAV) assuming reinvested dividends. Hence, returns are net of any ongoing fees which are automatically deducted from the funds' NAV but do not include sales loads, which may vary among investors for the same fund. We also obtain data on fund expenses (total expense ratios) from Morningstar Direct for the later part of our sample period (i.e. for funds with financial years ending in 2005 and later).[5] Following previous studies on Morningstar ratings (see e.g., Blake and Morey (2000), Kräussl and Sandelowsky (2007), and Del Guercio and Tkac (2008)) our focus is on fund share classes. We find that Stiftung Warentest ratings commonly differ between different share classes of the same fund. Moreover, to assess statistical significance in the empirical analysis, we solely rely on the time-series mean and standard deviation of monthly portfolio returns or coefficients (see subsection 2.3.1). Hence, we break any cross-sectional dependencies and our t-statistics are not inflated as a result of double-counting.[6]

We examine the predictive abilities of the Stiftung Warentest rating for the following major

[5] Funds registered for sale in Germany were not legally required to report data on total fund expenses prior to 2003. Moreover, there is only little fee coverage in Morningstar Direct prior to 2005.

[6] In a robustness test, we keep only the oldest available share class of a fund and repeat the analysis of section 3. The results are almost identical and not reported for the sake of brevity.

equity fund categories: Germany, Europe, Euro-Zone, North America, Pacific, and World.[7] The restriction of the sample size is due to three reasons. First, these are the largest equity fund categories. We exclude non-equity funds (like balanced or fixed income funds) because Stiftung Warentest refrained to assign a rating for those funds until 2003 as a result of the limited comparability of funds within these categories. Second, since ratings are only published for the major fund categories on a monthly basis, it is reasonable to assume that those fund groups receive most of the attention by mutual fund investors. Third, as mentioned above, many of the non-major fund categories do not consist of enough funds for an empirical investigation. Table 1 shows the total number of funds receiving a rating over the course of the sample period for the different fund categories. Consistent with the total growth in the industry, there is a sharp rise in the number of funds covered by Stiftung Warentest. The only exception is the category ”German equity” which comprises an almost stable fund universe over time. This highlights the increased internationality of the German mutual fund industry.

Insert Table 1 here

2.3 Empirical Methodology

2.3.1 Testing for Predictive Abilities

We employ two different methods to test whether the Stiftung Warentest rating accurately forecasts future fund performance: a dummy variable regression analysis, which is often used in studies analyzing the Morningstar rating (e.g., Blake and Morey (2000), Gerrans (2006), Kräussl and

[7] In 2004 Stiftung Warentest started to further differentiate between funds focusing on small cap and large cap stocks for several of these categories. Similarly, in 2004 the fundgroup "Pacific" was split into funds focusing solely on Japan and funds covering the whole Pacific region. In order to keep the tables clear and manageable we do not further split our categories into subgroups when presenting our results. Note however, that we control for exposure to small vs. large cap stocks by using a Carhart (1997) four factor alpha as performance measure (see section 2.3.2). Funds that invest solely in Japanese stocks receive the MSCI Japan as benchmark, instead of the MSCI Pacific. Our conclusions are not affected if we only use the MSCI Pacific as benchmark for these funds or exclude the fundgroup "Pacific" completely.

Sandelowsky (2007)), and a trading strategy analysis. In the dummy variable regression analysis, funds are sorted into quintiles based on their rating for every month from December 2001 to June 2008. The sorting is conducted separately for every fund category throughout the analysis. We then study the relationship between these quintiles and out-of-sample performance via multiple cross-sectional regressions, using the Fama and MacBeth (1973) procedure:

$$S_i = \gamma_1 + \gamma_2 \cdot D_{2,i} + \gamma_3 \cdot D_{3,i} + \gamma_4 \cdot D_{4,i} + \gamma_5 \cdot D_{5,i} + \varepsilon_i. \quad (3)$$

In equation 3, S_i is the out-of-sample performance metric for fund i and $D_{j,i}$ ($j = 2, 3, 4, 5$) are dummy variables taking the value 1 if fund i is sorted into quintile j. The coefficient $\gamma_1$equals the expected value of the out-of-sample performance metric if all dummy variables are 0, i.e. if the fund is in the first quintile. Hence, the quintile comprising the funds with the lowest Stiftung Warentest rating is used as a reference group. The other coefficients γ_j ($j = 2, 3, 4, 5$) represent the differences in performance between the respective quintiles and the reference group. If the predictor has perfect forecasting abilities, we should observe strictly increasing values for the coefficients γ_1 to γ_5.

In our baseline regressions, for which we present results in section 3, we investigate the relationship between rating quintiles and performance in the subsequent year (i.e. from month t+1 to month t+12). As we run the cross-sectional regressions for every month, fund returns are overlapping. To correct for the resulting serial correlation in the regression residuals, t-statistics are calculated using the Newey-West procedure with a lag of eleven months. Beyond calculating simple t-statistics, we also apply the recently proposed monotonicity test of Patton and Timmermann (2010) in order to test whether the coefficients from γ_1 to $\gamma_5$are indeed strictly increasing, as it should be expected under perfect forecasting abilities of the rating. When computing the test statistic we make use of Andrew Patton's code provided on his web-site.[8] As a robustness check, we

[8] See http://econ.duke.edu/~ap172/. When computing the p-values we apply the standard settings as suggested by Andrew Patton for monthly data, i.e. 1000 bootstrap replications and a block length equal to 6. We verified similar p-

also test the discriminatory power of the rating for longer out-of-sample evaluation periods (up to 36 months) and analyze whether sorting funds into deciles instead of quintiles affects our conclusions. We briefly comment on our findings for the additional tests.

Positive coefficients for a fund quintile in the dummy variable regression analysis signal that these funds are on average able to deliver a better performance than funds being assigned into the reference quintile. However, they do not necessarily imply positive risk-adjusted returns for an investor. In order to examine the potential profitability of a Stiftung Warentest-based investment strategy we therefore use the Jegadeesh and Titman (1993) methodology. For every month of the sample period funds are again divided into 5 equal-weighted portfolios based on their rating. Portfolio $Q1$ represents the fund portfolio having the lowest Stiftung Warentest rating in the particular month and portfolio $Q5$ consists of the funds with the highest rating. We then analyze the profitability of investing into these 5 portfolios. In addition, we also consider the returns of a hypothetical zero-cost strategy investing long (short) in the Q5 (Q1) portfolio. We investigate holding periods of one, three, six, twelve, 24, and 36 months. Like in Jegadeesh and Titman (1993), we construct overlapping portfolios. That is, for a holding period of T months the $Q1$ to $Q5$ portfolios consist of all quintile portfolios formed in the current month and the previous T - 1 months. Returns of the portfolios in a particular month are average returns of all T portfolios overlapping in that month. These overlapping portfolios are equivalent to a composite portfolio in which each month 1/T of the holdings are revised. Whenever a fund is liquidated within the evaluation period, we assume that fund shares can be sold at the fund's net asset value of the last trading day. In the following month, the proceeds will then be re-invested equally in the other funds of the particular portfolio.

---

values for a block length equal to 12. The reported p-values of the monotonic relationship test are studentised and based on all possible pair-wise comparisons.

2.3.2 Performance Measures

We apply three different metrics to evaluate the out-of-sample performance of mutual funds: the benchmark-adjusted return (BaR), the Jensen (1968) one factor alpha (α^{Jen}), and the Carhart (1997) four factor alpha (α^{Car}). All returns are measured in Euro. Formally, the performance measures for fund (or fund portfolio) i are calculated as follows:

$$BaR_i = \frac{1}{T} \cdot \left[\sum_{t=1}^{T} \left(r_{i,t} - r_{bm,t}\right)\right], \tag{4}$$

$$r_{i,t} - r_{f,t} = \alpha_i^{Jen} + \beta_i \cdot ERM_{i,t} + \varepsilon_{i,t}, \tag{5}$$

$$r_{i,t} - r_{f,t} = \alpha_i^{Car} + \beta_i \cdot ERM_{i,t} + \beta_{i,SMB} \cdot SMB_{i,t} + \beta_{i,HML} \cdot HML_{i,t} + \beta_{i,WML} \cdot WML_{i,t} + \varepsilon_{i,t}. \tag{6}$$

In the three equations, $r_{i,t}$, $r_{f,t}$, and $r_{bm,t}$ are the returns of fund i, the risk-free asset, and the benchmark of fund i in month t. $ERM_{i,t}$ is the excess return over the risk-free rate of the benchmark in t. We use the three-month Euribor as a proxy for the risk-free rate. The equivalent equity indices of Morgan Stanley Capital International (MSCI) are selected as benchmarks, i.e. the MSCI Germany, MSCI Europe, MSCI Euro-Zone, MSCI North America, MSCI Pacific, and MSCI World. In a robustness test we verify that we obtain similar results when using alternative appropriate benchmark indexes.[9] The expressions SMB (small minus big), HML (high minus low), and WML (winners minus losers) aim at capturing the size, value, and momentum effects documented in stock returns. We construct the factors using Datastream's stock universe and following the instructions outlined on Kenneth French's website. In order to compute the appropriate factors for these funds targeting regional stock markets like Europe we utilize the methodology of Griffin (2002). That is, the regional factors are market weighted averages of the

[9] Specifically, we use the following alternative market indexes: Composite DAX for Germany, DJ Stoxx 600 for Europe, DJ Euro Stoxx for Euro-Zone, S&P 500 for North America, Topix for Pacific, and FTSE All World for World.

country-specific components. Appendix A provides the reader with a detailed description of the construction of the size, value, and momentum factors.[10]

To compute the benchmark-adjusted return (BaR_i), we deduct the benchmark return from the return of the fund for every month of the evaluation period and then take the arithmetic average of the monthly excess returns. In the dummy variable regression analysis the one factor (α^{Jen}) respectively four factor alphas (α^{Car}) are calculated using all months of the out-of-sample evaluation period. For portfolios of funds (e.g. all funds belonging to quintile 5 according to their Stiftung Warentest rating), we compute benchmark-adjusted portfolio returns as equal-weighted average of individual benchmark-adjusted fund returns. Hence, we replicate a hypothetical trading strategy that for each fund sells the market index and invests the proceeds into the fund in order to capture the above (or below) expected market-adjusted return of that fund. To assess the risk-adjusted performance of the fund portfolios representing the various trading strategies, we use the complete sample period to calculate one factor and four factor alphas.

Since the rating of Stiftung Warentest is based on past performance, its success as a performance predictor depends on whether fund performance persists in the fund categories we investigate. To analyze this issue, we also use the benchmark-adjusted return, the Jensen (1968) one factor alpha, and the Carhart (1997) four factor alpha as alternative predictors to the Stiftung Warentest rating. We rely on the in-sample period of 60 months prior to fund selection to calculate the alternative predictors because Stiftung Warentest also uses the previous 60 months to calculate its ratings.

3 The Predictive Abilities of Stiftung Warentest: Empirical Results

3.1 Dummy Variable Regression Results

This section presents the results of the dummy variable regression analysis. Regression coefficients

---

[10] Factor realizations are available from the first author upon request.

are reported separately for the total fund sample (denoted as ”ALL”) and each fund group in Table 2.

Insert Table 2 here

For the total fund sample, Table 2 demonstrates that the rating system of Stiftung Warentest is able to predict future fund performance. For instance, Panel A documents that the average out-of-sample benchmark-adjusted return per month of funds being assigned to quintile 5 is 0.182% higher than the mean benchmark-adjusted return of funds in quintile 1. This amounts to an annualized difference of 2.18%. Similar patterns can be observed when considering the one factor alpha in Panel B. In this case, the coefficient for funds in the lowest rating quintile, γ_1, is -0.217%. In contrast, γ_5 is 0.180%, indicating an annualized difference of 2.16%. With respect to the four factor alpha, a performance difference between high and low rated funds can be observed as well, though it is less pronounced. Funds being assigned to quintile 5 generate an average four factor alpha that is 0.094% per month respectively 1.13% per year higher compared to quintile 1 funds. For the total fund sample, both coefficients, γ_1 as well as γ_5, are significantly different from zero for every performance measure. Moreover, the coefficients monotonically increase as we move from γ_1 to γ_5. Consequently, the last column shows that for the total fund sample, the monotonic relation (MR) tests always reject the null hypothesis with a high degree of statistic significance.

Table 2 further shows that the forecasting abilities of the Stiftung Warentest rating seem to depend on the fund group. For instance, the discriminatory power tends to be stronger for equity funds Europe and World, and to some extent for equity funds Germany. For equity funds North America, we do not find evidence on performance persistence, in particular with respect to the four factor alpha. Inspection of the p-values associated with the MR test leads to the same conclusion: For several fund markets and several performance measures we do not find evidence of a statistically significant increasing relation as we move from γ_1 to γ_5.

Interestingly, despite the positive and increasing values for the coefficients γ_2 to γ_5 in the total fund sample, there is no evidence that high rated funds are able to outperform their benchmark

MSCI index. Considering the one factor alpha, for instance, the performance spread of 0.180% between low and high rated funds still implies a negative alpha of -0.037% per month for funds belonging to quintile 5 given that the average one factor alpha of quintile 1 funds, the reference portfolio, is -0.217%. Also, while the γ_1-coefficients are highly negative in Panel C, the other coefficients γ_2 to γ_5 are very similar in size and increase only marginally. Obviously, although low rated funds realize a very low four factor alpha out-of-sample, the rating is not very well capable of discriminating in terms of the four factor alpha for the other fund quintiles. A natural question arising in this context is how much of the well-known size, value, and momentum factors are captured by the Stiftung Warentest rating. We explore this question in the next (sub-)sections.

Despite these potential problems, the results of the dummy variable regression analysis collectively support the notion of predictive abilities of the Stiftung Warentest rating system, which are statistically and economically significant. Funds in the highest quintile group outperform funds in the lowest quintile group up to 18 basis points per month in the next year. Moreover, in contrast to previous Morningstar-based studies, most of the coefficients do not only have their expected sign for low-rated but also for high-rated funds. Our robustness tests confirm these conclusions. The performance spread between high and low rated funds is of similar size and statistical significance when we extent the out-of-sample period to 24 or 36 months. Organizing funds into deciles instead of quintiles shows a slightly larger performance spread between the lowest and highest rating category for most fund groups. Still we do not see any evidence that funds in the highest rating category can generate a significant positive performance compared to their benchmark.

3.2 Stiftung Warentest-Based Trading Strategy

This section contains the results of the trading strategy analysis, which addresses the question of how profitable an investment into funds with a high Stiftung Warentest rating is in terms of benchmark-adjusted and risk-adjusted returns. Panel A of Table 3 shows the average benchmark-adjusted returns of the quintile portfolios *Q*1 to *Q*5 and the excess return of the zero-cost (*Q*5-*Q*1)-portfolio for the total fund sample. Panel B of this Table summarizes the returns of the zero-cost

($Q5$-$Q1$)-portfolios separately for each fund group. Panel C reports the results from a Carhart (1997) four factor regression which relates the returns of the ($Q5$-$Q1$)-portfolio on the market, size, value, and momentum factor. This allows us to examine whether common factors of stock returns are able to explain any forecast abilities of the rating. Regressions are carried out separately for every equity fund category. For the sake of brevity we focus on a twelve-month trading strategy in Panel C, but the results are similar for a shorter rebalancing frequency.

Insert Table 3 here

Inspection of Panel A shows that an investment in the $Q1$-portfolio generates significantly negative abnormal returns for all holding periods under consideration. The return of this portfolio equals -0.184% per month or -2.21% per year in the case of monthly rebalancing. For a holding period of 36 months the return of the $Q1$-portfolio is -0.141% per month, indicating a modest improvement for longer holding periods. Considering the results for the other portfolios, it is evident that returns generally increase with the Stiftung Warentest rating. However, even an investment into the $Q5$-portfolio with monthly portfolio rebalancing delivers only a marginally positive benchmark-adjusted return of 0.093% per month (1.12% per year), which is not statistically significantly different from zero. If we extent the holding period, returns of the $Q5$-portfolio tend to decrease. Hence, while the returns of the zero-cost portfolios are positive and significant for all holding periods analyzed and the MR tests of Patton and Timmermann (2010) again confirm a monotonic relation, most of the return difference stems from the underperformance of the $Q1$-portfolio. Panel B documents that in four out of six cases the statistical significant return of the ($Q5$-$Q1$)-portfolio can also be observed if equity fund categories are analyzed separately.

Even though the trading strategy analysis confirms the conclusion of predictive abilities for Stiftung Warentest drawn in subsection 3.1, it also shows the difficulties arising if one wants to use the ratings to establish a benchmark-outperforming strategy. Since mutual funds cannot be sold short, it is not possible to profit from the continued underperformance of low rated funds. This also implies that the returns generated from the long-short strategy are only hypothetical in nature.

Moreover, transaction costs (in particular front-end loads) are neglected in the calculations.[11]

The results displayed in Panel C of Table 3 show that after controlling for well-known return factors we observe a statistically positive alpha of the (*Q*5-*Q*1) zero cost strategy only for the fund categories European Monetary Union (*EFEMU*), Europe (*EFE*), Pacific (*EFP*), and World (*EFW*). The other alphas are positive but not statistically significant. The analysis confirms that the (*Q*5-*Q*1)-portfolios tend to load positively on return factors, in particular the market, size, and momentum factor. This supports the notion that some of the predictive abilities documented previously are simply due to the fact that the rating process of Stiftung Warentest ignores these additional factors.

3.3 Alternative Predictor Results

To compare the forecasting abilities of Stiftung Warentest with those of the alternative predictors, we repeat the dummy variable regression and the trading strategy analysis. To do so, funds are ranked based on their alternative predictor and then sorted into quintiles. We investigate the performance of the alternative predictors using the dummy variable regression approach in Table 4 and the trading strategy approach in Table 5. For the sake of brevity, we report results solely for the complete fund sample.

Insert Table 4 here

Insert Table 5 here

Inspection of both tables shows that the alternative predictors have about the same discriminatory power as the rating of Stiftung Warentest. Like Stiftung Warentest, funds being assigned to the lowest quintile strongly underperform their benchmark index. Moreover, all Fama-MacBeth regression coefficients for γ_5 are statistically significantly positive and economically

[11] It is tempting to test whether a trading strategy that is not based on quintiles but selects only the top ranked funds, say the top 5 or top 10 performers within a fund category, increases the profitability of a long-only fund investment. We find that the returns to such an investment rule are only slightly higher and they are associated with higher standard errors. As a result such a trading strategy does not yield a significant outperformance of the top-rated portfolio either.

meaningful which indicates a substantial performance difference between low and high rated funds. However, as revealed in the trading strategy analysis, even for funds in the highest quintile there is no statistical evidence of outperforming abilities with respect to the benchmark MSCI index. In unreported results we find that the degree of return predictability varies across fund categories. As expected, Stiftung Warentest ratings are better predictors in categories which display some level of performance persistence. This is in particular the case for funds investing in the European market or even globally.

Our results indicate that measures of historical fund performance are useful in predicting future fund performance at least for some fund categories. The forecasting power of the Stiftung Warentest rating system is broadly comparable to other performance measures which stem from the academic literature like the one factor or the four factor alpha.[12] The statistical evidence is to a large extent restricted to funds with an inferior past performance which continue to underperform in the near future, though. From an investor's perspective this is not a very useful feature since mutual funds cannot be sold short, as noted above. In contrast, all predictors seem to have problems in identifying funds that significantly outperform their benchmark.

3.4 Performance Predictors and Differences in Fees

We now turn to the question to what extent differences in investment expenses explain the persistence in risk-adjusted performance documented in the previous section. Our data source is Morningstar Direct which provides total expense ratios for a sufficiently large number of funds in our sample since 2005, i.e. for the second part of our sample period. Specifically, we have expense data for 37% of all funds at the end of 2005, 56% at the end of 2006, and 72% at the end of 2007. At the end of every of these three calendar years, funds are sorted into quintiles according to the different performance predictors (Stiftung Warentest rating, the benchmark-adjusted return, the

[12] In order to increase the forecasting power of the estimated alphas we have also employed the back test procedure developed in Mamaysky et al. (2007). We do not find that their methodology helps improving the predictive power of the models in our fund sample.

Jensen (1968) one factor alpha, and the Carhart (1997) four factor alpha). For every quintile, we calculate the arithmetic average of the total expense ratios of the funds' latest financial year. We test for significant differences in fees between the lowest and highest rated funds by using two-tailed t-tests. Results are displayed in Table 6.

Insert Table 6 here

The results are indicative of a slight negative relation between the various performance predictors and total fund expenses. For instance, funds in the highest Stiftung Warentest quintile charge fees that are approximately 0.2% lower per year than the fees charged by funds in the lowest quintile. A similar result can be found when considering the four factor alpha as performance predictor. However, the difference in fees is statistically insignificant if we consider benchmark-adjusted returns or the one factor alpha. Overall, these findings suggest that differences in fund expenses can only partly explain the predictability of mutual fund returns. For instance, as shown in subsection 3.1, funds in the highest Stiftung Warentest quintile outperform funds in the lowest quintile by 1.13% per year according to their four factor alpha which is substantially higher than the 0.2% difference in fees. The evidence of performance persistence after controlling for differences in total expense ratios is consistent with differences in managerial skill in some of the fund categories studied. However, other interpretations are well possible given that the expense data covers only a subsample of the fund universe and does not incorporate transaction costs due to turnover.

4 Is Fund Activity Related to Future Performance?

After all, a manager can only beat his benchmark index if he deviates from it. To do so, he can overweight and underweight certain stocks or industries. To the extent that managers who deviate more from their reference index are not overconfident but condition their allocation indeed on valuable information, public or private, a higher degree of active management signals the fund's potential of generating superior future returns. In this section we report the returns to investment strategies which follow this intuition and quantify the degree of active management before selecting mutual funds.

Recent evidence for U.S. domestic equity funds supports the idea that fund activity is positively related to future fund performance. Wermers (2003) finds that tracking error volatility or simply $Tracking\ Error$, i.e. the standard deviation of fund's benchmark-adjusted return, is positively related to fund performance. Amihud and Goyenko (2009) show that a mutual fund's R^2 obtained from a four-factor regression can be used to predict its future performance. Cremers and Petajisto (2009) and Kacperczyk et al. (2005) compare the holdings of a mutual fund with the holdings of a benchmark index. They find that a larger deviation from the benchmark by overweighting some stocks or a particular industry is associated with a higher risk-adjusted portfolio return. To measure the extent of the deviation from the benchmark, Cremers and Petajisto (2009) define a new measure labeled as $Active\ Share$: [13]

$$Active\ Share = \frac{1}{2} \cdot \sum_{i=1}^{N} \left| w_{i,fund} - w_{i,index} \right| , \quad (7)$$

where $w_{i,fund}$ and $w_{i,index}$ are the portfolio weights of stock i in the fund and in the benchmark index. The calculation of $Active\ Share$ requires information about the composition of the fund and benchmark portfolios whereas $Tracking\ Error$ and R^2 can be retrieved very easily from the mutual funds' and benchmarks' returns. However, Cremers and Petajisto (2009) find that $Tracking\ Error$ by itself is not related to future performance after controlling for $Active\ Share$. Hence, $Active\ Share$ may be a preferred predictor which captures a different dimension of active management than $Tracking\ Error$ or R^2.

We test the predictive abilities of all three measures of active management in our German mutual fund sample. We obtain the funds' $Tracking\ Error$ and R^2 from a standard four factor

[13] In contemporaneous work, Cremers et al. (2011) also investigate the degree of active management for non-US funds and obtain similar conclusions: While truly active funds are able to outperform their benchmark, so called "clostet indexers" (i.e. active funds that closely follow their benchmark) generally fail to add value for their customers.

regression where $Tracking\ Error$ is defined as the standard deviation of the residual. Like before the estimation is based on the latest 60 return observations and the equity indices of MSCI serve as benchmarks.

In order to compute the funds' $Active\ Share$ we receive data on monthly portfolio holdings from Morningstar for the equity fund categories Germany, Euro-Zone, and Europe. For these three categories 92% of the funds in the Stiftung Warentest sample are also covered by Morningstar. However, two problems are associated with the computation of $Active\ Share$ in our data set. First, Morningstar has no data on portfolio holdings for European funds prior to October 2002. Hence, our sample period is reduced by approximately one year. Additionally, not all fund families report the portfolio holdings of their funds to Morningstar, which further reduces our fund sample size and potentially creates a sample selection bias. We test whether funds which report their portfolio positions have a higher performance than funds which do not report. We find virtually no performance differences between both groups irrespective of the performance metric used. Hence, we believe it is unlikely that our results are affected by a sample selection bias.

A second problem arises because we do not have data on the composition of the funds' benchmarks. We therefore create an artificial benchmark portfolio by aggregating the portfolio positions of all funds within the same Morningstar category. Stocks which are reported by less than 10% of all funds in a certain month are excluded. To obtain the benchmark weights we value-weight the remaining stock positions. These difficulties may lead to inaccuracies in the calculation of $Active\ Share$ which should rather work against finding a significant relation.

Tables 7 and 8 display the results for the dummy variable regression and the trading strategy analysis for our three measures of fund activity. We report results for the complete fund sample. Note that we reverse the sorting for the R^2 to ease the interpretation. Funds with the highest R^2 relative to their benchmark are sorted into quintile 1 and funds with the lowest R^2 into quintile 5 because a higher R^2 indicates a closer replication of the benchmark, and hence a lower fund activity.

Insert Table 7 here

Insert Table 8 here

The findings suggest that investors should avoid active funds which closely track their benchmark. According to the results of the dummy variable regressions, the underperformance amounts to approximately 20 basis points per month over the next year. The underperformance is similar for every predictor and every out-of-sample performance measure and it is highly statistically significant with t-statistics ranging between 3.3 and 8.5. This high level of statistical significance indicates that the underperformance of these ”closet indexers” can be forecasted with a high degree of confidence because they do not have enough active positions to beat their benchmark and to cover their fees and transaction costs. There is also some evidence that more active funds have a better performance, but the relation is not always monotonically increasing and it is weaker if we consider the four factor alpha as performance measure. For instance, Table 7 reveals a difference in benchmark-adjusted returns of 0.330% per month (3.96% per year) between funds in the highest and lowest $Active\ Share$ quintile. In contrast, the difference is only 0.094% per month (1.13% per year) if we consider the four factor alpha. As a result, for the four factor alpha we are generally unable to reject the null hypothesis of no monotonic relation using the MR test.

The trading strategy analysis shows that an investment into the $Q5$-portfolio, which contains the funds with the highest level of fund activity, yields positive benchmark-adjusted portfolio returns. The returns are moderately positive if we consider the R^2 or $Tracking\ Error$ with an outperformance ranging between 5 to 8 basis points per month. Sorting funds on their $Active\ Share$ and investing into the $Q5$-portfolio appears to be more promising. In this case the benchmark-adjusted portfolio returns amount to approximately 25 basis points per month, irrespective of the rebalancing frequency. However, this outperformance is never significant in a statistical sense because of the high standard errors associated with the strategy. To a certain extent this lack of statistical significance might be caused by the reduced fund sample size and sample period for which we can calculate $Active\ Share$.

We obtain very similar results for the funds' R^2 and $Tracking\ Error$ as performance predictors because both measures are highly correlated with a rank-order correlation of -0.84. The correlation between $Active\ Share$ and the other two measures is considerably lower (-0.43 for R^2 and 0.32 for $Tracking\ Error$). Hence, $Active\ Share$ might indeed capture a different dimension of active management as argued by Cremers and Petajisto (2009).

Overall, our results give a first indication that the degree to which a fund pursues a truly active strategy is a valuable information for investors outside the U.S. market as well. Interestingly, measures of past performance (either the rating of Stiftung Warentest or others) and fund activity are only weakly correlated with each other in our sample.[14] Hence, they might be used in combination to select better performing funds.

To analyze the value of such a strategy, we consider a double sorting procedure which takes fund activity in addition to the rating into account. If fund activity helps to differentiate between future winners and losers, portfolios consisting of funds which score high in both dimensions should achieve a better performance than funds which only have a high Stiftung Warentest rating. In a similar vein, funds with a low score in both dimensions should have a worse performance than funds that only have a low rating. We test this proposition in the following way. In a given month, all funds which have a Stiftung Warentest rating and information about the specific fund activity measure are first sorted into quintiles based on their rating. Next, we further subdivide funds into five equal-sized groups based on their level of fund activity. Table 9 reports the benchmark-adjusted performance of the resulting 25 different fund portfolios with annual portfolio rebalancing. We report the results only for $Tracking\ Error$ and $Active\ Share$ as measures of fund activity, because the difference between using $Tracking\ Error$ and R^2 is small.

Insert Table 9 here

As it can be inferred from the ”High-Low”-portfolio returns in both panels of Table 9, portfolios consisting of funds which score high in both dimensions very often achieve the best

[14] For instance, the correlations between the rating of Stiftung Warentest and the three measures of fund activity are: -0.08 (R^2), 0.01 (*Tracking Error*), and 0.11 (*Active Share*).

performance, while funds with a low score in both dimensions tend to have the worst performance. For instance, Panel A shows that the portfolio containing funds with the highest Stiftung Warentest rating and the highest $Tracking\ Error$ has an average benchmark-adjusted return of 0.28% per month while the portfolio comprising funds with a low Stiftung Warentest rating and a low $Tracking\ Error$ has an average benchmark-adjusted return of -0.24% per month. The difference between both portfolios is highly statistically significant and economically meaningful. As it can be seen in Panel B, the results are similar when we condition on $Active\ Share$. Overall, our results indicate that taking into account fund activity is useful for differentiating between skill and luck, in particular for underperforming funds. As a caveat however, we note that we are not always able to statistically confirm a truly monotonic relation between the sorting variables and benchmark-adjusted returns, in particular for Panel B ($Active\ Share$).

5 Conclusion

This paper studies the degree to which the mutual fund rating system of Stiftung Warentest predicts future mutual fund performance in the German fund market. Stiftung Warentest is a consumer protection agency and mutual fund ratings are published in its financial magazine Finanztest on a monthly basis. The magazine is devoted to inform and educate its readers in order to make better financial decisions. In addition, we analyze whether the degree of active portfolio management is positively related to future fund performance outside the U.S. as well.

Our investigation leads to the following findings. Firstly, we find a positive and significant relation between the Stiftung Warentest rating of a fund and its future performance. Depending on the performance measure, the investment horizon, and the evaluation technique used, funds in the lowest rating quintile underperform funds in the highest rating quintile by 0.094% to 0.291% per month. These differences in risk-adjusted returns appear to be only slightly driven by differences in fees. Secondly, the forecasting power is not robust over different categories. For instance, the prediction power is rather high for the category equity funds World, but we find no forecasting abilities for the category equity funds North America. The underlying reason is that performance – according to the performance measures used in this study – is persistent in some fund categories, while it is not in others. Thirdly, high rated funds deliver a performance which is in the range of their benchmark but they fail to significantly beat their benchmark. Fourthly, measures of fund activity ($Active\ Share$, $Tracking\ Error$, and R^2) also predict future fund performance outside the U.S. fund market. Our results suggest that fund activity is one candidate to differentiate between skill and luck when drawing inferences about a limited sample of historical fund returns. This strategy is particularly successful among underperforming funds, but, at least in terms of benchmark-adjusted returns, it also helps to identify better performing active funds. Hence, rating agencies should provide information about the degree of fund activity.[15]

Our study reveals significant differences in the degree of performance persistence between different fund markets. As the question whether superior performance persists is central to our understanding of stock market efficiency, future research should investigate whether certain fund management or country characteristics can explain these differences across fund markets. For instance, Ramos (2009) shows that more developed fund markets provide higher returns for investors. Alternatively, differences in performance predictability may also result from differences in investor rationality across fund markets. According to the model of Berk and Green (2004), rational mutual fund investors chase past performance because high historical returns signal managerial talent. However, as managers are assumed to have only a limited capacity of generating a positive alpha, this behavior deteriorates expected future fund performance. Glode et al. (2009) argue that investor rationality in the mutual fund market is varying over time. Future research may investigate whether it is also varying across different fund markets. The German fund market might serve as an interesting laboratory in this context, since, as shown by Jank (2010), private households do not seem to entirely behave according to the predictions of the model of Berk and Green (2004).

[15] Recently, Stiftung Warentest started to report information on how close a fund replicates its underlying benchmark in its rating tables. However, this information is qualitative in nature and not incorporated in the rating construction.

Appendix: Factor Construction Details

A.1 Selection of the Stock Universe

Security data is extracted from Thomson Reuters Datastream. For each stock market of interest, we create a constituent list based on all securities which belong to that market and are coded as TYPE ”Equity” or ”Preference Share”. Securities are included independent of their status (”Active”, ”Dead” or ”Suspended”). We construct constituent lists for the following Datastream markets: Australia, Austria, Belgium, Brazil, Canada, China, Denmark, Finland, France, Germany, Greece, Hong Kong, India, Indonesia, Ireland, Italy, Japan, Mexico, Netherlands, New Zealand, Norway, Portugal, Russia, Singapore, South Africa, South Korea, Spain, Sweden, Switzerland, Turkey, and United Kingdom. These are all countries/markets that are either used for the construction of the MSCI regional benchmark indices MSCI Europe, MSCI Euro-Zone, MSCI North America, MSCI Japan respectively Pacific or comprise the G-20 group of the world’s largest economies. Exceptions are Saudi Arabia, for which no data is available from Datastream, and the United States, for which we can obtain factor returns from Kenneth French’s website. Our procedure results in a total sample of 47,130 unique securities. We then make use of the Datastream security identifier (DSCD) to download the following data for our sample: International Securities Identification Number (ISIN), geographical status (GEOGN), total return index (RI), unadjusted price (UP), Datastream total market value of equity (MV), Worldscope total market capitalization at fiscal year end (WC08001), and Worldscope book value of equity at fiscal year end (WC03501). All numerical values are converted in Euro.

A.2 Data Screens and Return Calculations

We use the variable GEOGN to identify and exclude firms which are assigned to the wrong stock market. All firms having either no stock data (no values for RI or UP) or no Worldscope data (no values for WC08001 or WC03501) are dropped from the sample. Security returns are calculated using the total return index (RI) which is adjusted for dividends (i.e. assumes that dividends are reinvested) and stock splits. To clean the return data, we apply the following screens advocated by Ince and Porter (2006). First, each month we identify firms that have been delisted previously. Second, firm observations are classified as penny stocks whenever their unadjusted price (UP) was in the lowest decile in more than 50% of the last 12 months. Third, we remove unrealistic returns from the data by setting any return above 300% that is reversed within one month to missing.

A.3 Factor Construction

Factors are constructed separately for each stock market. We follow the methodology outlined on Kenneth French's website. In particular, in order to construct the value and size factors, we form six value-weighted portfolios based on firm size and equity book-to-market ratio each year at the end of June. A firm's equity book-to-market ratio for June is defined as WC03501/WC08001 using values at the end of the firm's fiscal year ending anywhere in the previous calendar year. Firm size for June is the total market value of equity (MV) at the end of June. To be included in any of the portfolios, we impose the following requirements: 1) the firm's stock must have valid price data at the end of June (i.e. no previous delisting and no penny stock), and 2) neither WC03501 nor WC08001 nor MV must be negative.

We use the same breakpoints as Fama and French (1993) to sort stocks into the portfolios, i.e. the breakpoints for the book-to-market ratio are the 30th and 70th percentiles and the size breakpoint is the median market equity. Returns for the size factor (SMB) and value factor (HML) are then calculated as follows:

$$SMB = \frac{1}{3} \cdot (Small\,Value + Small\,Neutral + Small\,Growth) - \frac{1}{3} \cdot (Big\,Value + Big\,Neutral + Big\,Grwoth). \tag{8}$$

$$HML = \frac{1}{2} \cdot (Small\,Value + Big\,Value) - \frac{1}{2} \cdot (Small\,Growth + Big\,Growth). \tag{9}$$

The momentum factor is computed based on six value-weighted portfolios formed on total market equity (MV) at the end of the previous month and prior 1-year return (excluding the return of the most recent month). In contrast to the value and size factor-mimicking portfolios, the momentum portfolios are rebalanced monthly. Return breakpoints are the 30th and 70th percentiles and the size breakpoint is the median market equity. 1) Invalid price data, 2) negative MV data, or 3) missing prior 1-year return data results in exclusion of the firm's stock for the particular month concerned. Returns for the momentum factor (WML) are then calculated as follows:

$$WML = \frac{1}{2} \cdot (Small\ High + Big\ High) - \frac{1}{2} \cdot (Small\ Low + Big\ Low). \qquad (10)$$

A.4 Construction of the Regional Factors

In order to compute the factors for the different world regions comprising of several markets/countries, we utilize the methodology of Griffin (2002). That is, the regional factors are market weighted averages of the country-specific components. For example, the value factor for Europe in month t is calculated as: $HML_{t,Eur} = \sum_{k=1}^{N} I_{k,t,Eur} \cdot \omega_{k,t-1,Eur} \cdot HML_{k,t}$, where N is the total number of countries, $I_{k,t,Eur}$ is an indicator variable taking the value 1 (0) when country k is a part of Europe, $\omega_{k,t-1,Eur}$ is the fraction of the total dollar-denominated European market capitalization attributable to country k at the end of the previous month, and $HML_{k,t}$ is the factor-mimicking return for country k in month t. Data on the total dollar-market capitalization of each country is extracted from Datastream (e.g., code TOTMKBD(MV) for Germany). Our assignment of the countries to the regions Europe, European Monetary Union, North America, and Japan/Pacific follows the index country membership definition used by MSCI. To construct the world factors, we use a slightly different methodology: Each month, we sort countries in descending order based on their total dollar-denominated market capitalization and calculate the cumulative coverage of the world market capitalization at each country. After a total market coverage of 85% is achieved, we stop with this procedure and exclude all other countries from the factor return calculations. Market weights for the included countries are adjusted proportionally. With this approach, we account for the fact that errors in the database are more likely for firms in smaller, emerging economies, especially in the earlier parts of the sample period.

References

Alexander, G. J., Jones, J. D. and Nigro, P. J., ‘Mutual Fund Shareholders: Characteristics, Investor Knowledge, and Sources of Information’, *Financial Services Review*, Vol. 7, 1998, pp. 301–316.

Amihud, Y. and Goyenko, R., ‘Mutual Fund’s R^2 as Predictor of Performance’, *Working Paper* (NYU, 2009).

Barras, L., Scaillet, O. and Wermers, R., ‘False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas’, *Journal of Finance*, Vol. 65, 2010, pp. 179–215.

Berk, J. B. and. Green, R. C., ‘Mutual fund flows and performance in rational markets’, *Journal of Political Economy*, Vol. 112, 2004, pp. 1269–1295.

Blake, C. R. and Morey, M. R., ‘Morningstar Ratings and Mutual Fund Performance’, *Journal of Financial and Quantitative Analysis*, Vol. 35, 2000, pp. 451–483.

Bundesverband Investment und Asset Management e.V., 2008, *Jahrbuch 2008*.

Busse, J. A., Goyal, A. and Wahal, S., ‘Performance and Persistence in Institutional Investment Management’, *Journal of Finance*, Vol. 65, 2010, pp. 765–790.

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

Cremers, K. M. and Petajisto, A., ‘How Active Is Your Fund Manager? A New Measure That Predicts Performance’, *Review of Financial Studies*, Vol. 22, 2009, pp. 3329–3365.

Cremers, K. M., Ferreira, M., Matos, P. and Starks, L. ‘The Mutual Fund Industry Worldwide: Explicit and Closet Indexing, Fees, and Performance’, *Working Paper* (Yale School of Management, 2011).

Cuthbertson, K., Nitzsche, D. and O’Sullivan, N., ‘False Discoveries in UK Mutual Fund Performance’, 2010, *European Financial Management*, forthcoming.

Del Guercio, D. and Tkac, P. A., ‘Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow’, *Journal of Financial and Quantitative Analysis*, Vol. 43, 2008, pp. 907–936.

Fama, E. and French, K. R., ‘Luck versus Skill in the Cross Section of Mutual Fund Returns’, *Journal of Finance*, Vol. 65, 2010, pp. 1915-1947.

Fama, E. F. and French, K. R., ’Common Risk Factors in the Returns on Stocks and Bonds’, *Journal of Financial Economics*, Vol. 33, 1993, pp. 3–56.

Fama, E. F. and MacBeth, J. D., ‘Risk, Return and Equilibrium: Empirical Tests’, *Journal of Political Economy*, Vol. 71, 1973, pp. 607–636.

Gerrans, P., ‘Morningstar Ratings and Future Performance’, *Accounting & Finance*, Vol. 46, 2006, pp. 605–628.

Glode, V., Hollifield, B., Kacperczyk, M. and. Kogan, S., ‘Is Investor Rationality Time Varying? Evidence from the Mutual Fund Industry‘, *Working Paper* (NBER, 2009).

Griese, K. and Kempf, A., ‘Lohnt aktives Fondsmanagement aus Anlegersicht?‘, *Zeitschrift für Betriebswirtschaft*, Vol. 73, 2003, pp. 201–224.

Griffin, J. M., ‘Are the Fama and French Factors Global or Country Specific?’, *Review of Financial Studies*, Vol. 15, 2002, pp. 783–803.

Ince, O. and Porter, R. B., ‘Individual Equity Return Data from Thomson Datastream: Handle with Care!’, *Journal of Financial Research*, Vol. 29(4), 2006, pp. 463–479.

Investment Company Institute, *200*9 *Investment Company Fact Book*, Electronic copy available from www.ici.org, 2009.

Jank, S., ‘Are There Disadvantaged Clienteles in Mutual Funds? Evidence from German Mutual Fund Investors’, *Working Paper* (Deutsche Bundesbank, 2010).

Jegadeesh, N. and Titman, S., ‘Returns to Buying Winners and Selling Losers: Implication for Stock Market Efficiency’, *Journal of Finance*, Vol. 48, 1993, pp. 65–91.

Jensen, M. C., ‘The Performance of Mutual Funds in the Period 1955-1964’, *Journal of Finance*, Vol. 23, 1968, pp. 389–416.

Kacperczyk, M., Sialm, C. and Zheng, L., ‘On the Industry Concentration of Actively Managed Equity Mutual Funds’, *Journal of Finance*, Vol. 60, 2005, pp. 1983–2011.

Korkeamaki, T. and Smythe, T. Jr., ‘Effects of Market Segmentation and Bank Concentration on

Mutual Fund Expenses and Returns – Evidence from Finland', *European Financial Management*, Vol. 10, 2004, pp. 413-438.

Kräussl, R. and Sandelowsky, R. M., 'The Predictive Performance of Morningstar's Mutual Fund Ratings', *Working Paper* (VU University Amsterdam, 2007).

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

Mamaysky, H., Spiegel, M. and Zhang, H., 'Improved Forecasting of Mutual Fund Alphas and Betas', *Review of Finance,* Vol. 11, 2007, pp. 359–400.

Morey, M. R., 'The Kiss of Death: A 5-Star Morningstar Mutual Fund Rating?', *Journal of Investment Management*, Vol. 3, 2005, pp. 41–52.

Otten, R. and Bams, D., 'European Mutual Fund Performance', *European Financial Management,* Vol. 8, 2002, pp. 75–101.

Patton, A. J. and Timmermann, A., 'Monotonicity in Asset Returns: New Tests with Applications to the Term Structure, the CAPM and Portfolios Sorts', *Journal of Financial Economics,* Vol. 98, 2010, pp. 605-625.

Ramos, S. B., 'The Size and Structure of the World Mutual Fund Industry', *European Financial Management*, Vol. 15, 2009, pp. 145–180.

Reuter, J. and Zitzewitz, E., 'Do ads influence editors? Advertising and bias in the financial media', *Quarterly Journal of Economics*, Vol. 121, 2006, pp. 197–227.

Stotz, O., 'Selection, Market Timing and Style Timing of Equity Mutual Funds -Evidence from Germany', *Zeitschrift für Betriebswirtschaft*, Vol. 77, 2007, pp. 51–73.

Wermers, R., 'Are Mutual Fund Shareholders Compensated for Active Management "Bets"?', *Working Paper (*University of Maryland, 2003).

Table 1: Number of Funds by Investment Category

This table presents the number of funds with a Finanztest rating for a subset of reporting dates. We investigate the following equity mutual fund categories: Germany (EFG), Europe (EFE), Euro-Zone (EFEMU), North America (EFA), Japan/Pacific (EFP), World (EFW). Abbreviations are in parentheses.

	Mutual fund category (Abbreviation)					
Date	**Germany (EFG)**	**Europe (EFE)**	**Euro-Zone (EFEMU)**	**Pacific (EFP)**	**North Amerika (EFA)**	**World (EFW)**
31.12.2001	87	118	-	91	108	117
31.12.2002	89	159	-	105	122	161
31.12.2003	80	182	36	154	136	172
31.12.2004	76	226	60	187	163	227
31.12.2005	92	404	97	323	292	421
31.12.2006	92	498	118	371	382	542
31.12.2007	93	592	141	419	454	661
30.06.2008	89	616	130	405	454	657

Table 2: Dummy Variable Regression Results

This table presents results from monthly regressions of mutual fund performance in the subsequent year (from month t+1 to month t+12) on Stiftung Warentest rating quintile dummies for the following fund groups: Equity funds Germany (EFG), equity funds European Monetary Union (EFEMU), equity funds Europe (EFE), equity funds Pacific (EFP), equity funds North America (EFA), equity funds World (EFW). Mutual fund performance is assessed using the benchmark-adjusted fund return (BaR) in Panel A, the Jensen (1968) one factor alpha (α^{Jen}) in Panel B, and the Carhart (1997) four factor alpha (α^{Car}) in Panel C. Fama-MacBeth regression coefficients are the times-series average of the monthly cross-sectional regression coefficients. t-statistics (in parentheses) are calculated using the time-series standard error of the mean and adjusted for heteroskedasticity using the Newey-West procedure with a lag of eleven months. See section 2.3 for a description of the methodology. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level. The last column of this table reports studentised p-values from the monotonic relation (MR) test of Patton and Timmermann (2010) applied to the quintile portfolios based on all possible pair-wise comparisons (MR^{all}). To compute the p-values we use the Matlab code provided by Andrew Patton on his web-site.

Group	γ1	γ2	γ3	γ4	γ5	Adj. R2	F-Test	MR^{all} p-value
Panel A: BaR over next 12 months (in %)								
ALL	-0.162	0.016	0.064	0.134	0.182	0.028	6.433	0.000
(t-stat)	(-5.781)***	(1.212)	(3.137)***	(3.189)***	(3.861)***			
EFG	-0.057	0.001	-0.001	0.032	0.061	0.070	1.569	0.137
(t-stat)	(-0.571)	(0.049)	(-0.016)	(0.536)	(0.676)			
EFEMU	-0.179	0.028	0.034	0.031	0.097	0.171	4.082	0.110
(t-stat)	(-1.700)*	(0.856)	(1.148)	(0.517)	(0.788)			
EFE	-0.176	-0.015	0.054	0.146	0.242	0.077	5.531	0.000
(t-stat)	(-4.035)***	(-0.709)	(1.450)	(2.969)***	(4.686)***			
EFP	-0.122	0.064	0.131	0.166	0.152	0.041	1.550	0.112
(t-stat)	(-1.226)	(1.889)*	(2.537)**	(1.848)*	(2.315)**			
EFA	-0.197	0.015	0.000	0.052	0.074	0.040	1.888	0.354
(t-stat)	(-5.280)***	(0.583)	(0.005)	(0.995)	(1.799)*			
EFW	-0.189	0.024	0.096	0.198	0.263	0.067	4.880	0.000
(t-stat)	(-5.744)***	(1.064)	(5.538)***	(6.616)***	(4.393)***			
Panel B: Jensen alpha over next 12 months (in %)								
ALL	-0.217	0.018	0.068	0.132	0.180	0.025	5.549	0.000
(t-stat)	(-4.789)***	(1.597)	(4.225)***	(4.055)***	(5.246)***			
EFG	-0.190	0.019	0.035	0.095	0.162	0.072	1.552	0.000
(t-stat)	(-2.766)***	(1.127)	(1.909)	(4.087)***	(2.446)**			
EFEMU	-0.332	0.091	0.115	0.128	0.181	0.174	3.946	0.066
(t-stat)	(-5.278)***	(4.852)***	(4.434)***	(1.977)**	(1.647)			
EFE	-0.312	0.021	0.082	0.173	0.234	0.069	4.863	0.000
(t-stat)	(-6.131)***	(1.126)	(3.427)***	(4.527)***	(6.605)***			
EFP	-0.036	0.034	0.115	0.165	0.132	0.040	1.452	0.398
(t-stat)	(-0.313)	(1.082)	(1.980)**	(1.965)**	(2.337)**			
EFA	-0.254	0.014	0.004	0.037	0.070	0.040	1.998	0.243
(t-stat)	(-4.324)***	(0.637)	(0.184)	(0.832)	(1.574)			
EFW	-0.230	0.009	0.077	0.157	0.248	0.067	5.303	0.000
(t-stat)	(-5.474)***	(0.292)	(2.578)***	(3.377)***	(4.368)***			

Panel C: Carhart alpha over next 12 months (in %)

ALL	-0.220	0.020	0.032	0.064	0.094	0.008	3.049	0.000
(t-stat)	(-4.862)***	(1.894)*	(1.957)**	(3.781)***	(3.150)***			
EFG	-0.275	0.009	0.006	0.084	0.112	0.067	1.456	0.021
(t-stat)	(-1.670)*	(0.276)	(0.198)	(2.925)***	(1.971)**			
EFEMU	-0.255	0.097	0.129	0.094	0.119	0.146	3.470	0.331
(t-stat)	(-2.024)**	(2.782)***	(4.890)***	(1.945)*	(1.893)*			
EFE	-0.241	0.053	0.088	0.115	0.135	0.042	3.273	0.007
(t-stat)	(-1.926)*	(2.206)**	(2.527)**	(2.299)**	(2.098)**			
EFP	0.017	0.035	0.117	0.171	0.202	0.036	1.701	0.041
(t-stat)	(0.109)	(0.726)	(1.520)	(1.747)*	(2.290)**			
EFA	-0.241	0.021	-0.014	-0.023	-0.017	0.036	1.965	0.773
(t-stat)	(-3.675)***	(0.592)	(-0.461)	(-0.586)	-(0.325)			
EFW	-0.286	-0.034	-0.049	-0.001	0.055	0.038	3.762	0.418
(t-stat)	(-5.356)***	(-1.225)	(-1.443)	(-0.018)	(0.824)			

Table 3: Stiftung Warentest-based Trading Strategy Results

This table presents results of a Stiftung Warentest-based trading strategy using the Jegadeesh and Titman (1993) methodology. Panel A shows the benchmark-adjusted returns of five equally-sized fund portfolios, which are sorted on the basis of their Stiftung Warentest rating, for the total fund sample. For portfolios of funds (e.g. all funds belonging to quintile 5 according to their rating), we compute benchmark-adjusted portfolio returns as equal-weighted average of individual benchmark-adjusted fund returns. (Q5-Q1) reports the return to a hypothetical zero-cost strategy investing long (short) in the Q5 (Q1) portfolio. Panel B reports the excess returns of the (Q5-Q1)-strategy separately for each fund category. Panel C documents the results from regressing (Q5-Q1)-portfolio excess returns on the factor returns ERM, SMB, HML and WML separately for each fund category. We consider the following fund groups: Equity funds Germany (EFG), equity funds European Monetary Union (EFEMU), equity funds Europe (EFE), equity funds Pacific (EFP), equity funds North America (EFA), equity funds World (EFW). We investigate holding periods of one, three, six, twelve, 24 and 36 months in Panel A and B. In Panel C, portfolios are rebalanced on an annual basis. See section 2.3 for a description of the methodology. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level. In addition, the last row of Panel A reports studentised p-values from the monotonic relation (MR) test of Patton and Timmermann (2010) applied to the quintile portfolios based on all possible pair-wise comparisons (MR^{all}). To compute the p-values we use the Matlab code provided by Andrew Patton on his web-site.

Panel A: BaR of the fund portfolios for the total fund sample

Finanztest Quintile	**Rebalancing Frequency** T=1	T=3	T=6	T=12	T=24	T=36
Q1	-0.184	-0.198	-0.182	-0.168	-0.151	-0.141
(t-stat)	(-2.101)**	(-2.217)**	(-2.027)**	(-1.856)*	(-1.680)*	(-1.570)
Q2	-0.156	-0.163	-0.161	-0.152	-0.146	-0.141
(t-stat)	(-1.763)*	(-1.860)*	(-1.829)*	(-1.702)*	(-1.622)	(-1.565)
Q3	-0.106	-0.104	-0.099	-0.093	-0.088	-0.085
(t-stat)	(-1.154)	(-1.144)	(-1.118)	(-1.039)	(-0.984)	(-0.951)
Q4	-0.027	-0.017	-0.014	-0.008	-0.010	-0.006
(t-stat)	(-0.271)	(-0.174)	(-0.146)	(-0.086)	(-0.101)	(-0.065)
Q5	0.093	0.093	0.093	0.075	0.064	0.046
(t-stat)	(0.839)	(0.849)	(0.864)	(0.711)	(0.624)	(0.461)
Q5-Q1	0.277	0.291	0.276	0.243	0.215	0.187
(t-stat)	(4.115)***	(4.575)***	(4.626)***	(4.571)***	(4.364)***	(3.960)***
MR^{all} p-value	0.000	0.000	0.000	0.001	0.021	0.085

Panel B: (Q5-Q1)-portfolio returns for every fund group						
Group	**Rebalancing Frequency**					
	T=1	**T=3**	**T=6**	**T=12**	**T=24**	**T=36**
EFG	0.074	0.136	0.133	0.121	0.122	0.125
(t-stat)	(0.694)	(1.287)	(1.282)	(1.268)	(1.410)	(1.498)
EFEMU	0.176	0.197	0.196	0.217	0.224	0.220
(t-stat)	(1.862)*	(2.094)**	(2.152)**	(2.353)**	(2.396)**	(2.402)**
EFE	0.332	0.344	0.330	0.284	0.235	0.197
(t-stat)	(4.281)***	(4.842)***	(4.793)***	(4.332)***	(3.713)***	(3.286)***
EFP	0.150	0.158	0.169	0.151	0.146	0.129
(t-stat)	(1.623)	(1.835)*	(1.945)*	(1.758)*	(1.795)*	(1.606)
EFA	0.228	0.218	0.194	0.143	0.118	0.096
(t-stat)	(2.500)**	(2.589)***	(2.449)**	(1.915)*	(1.586)	(1.268)
EFW	0.425	0.439	0.418	0.388	0.355	0.317
(t-stat)	(3.838)***	(4.290)***	(4.316)***	(4.326)***	(4.182)***	(3.897)***

Panel C: Carhart alpha of the (Q5-Q1)-portfolio for every fund group assuming annual rebalancing						
Group	**αCar**	**βERM**	**βSMB**	**βHML**	**βWML**	**Adj. R2**
EFG	0.036	0.062	0.167	0.006	0.030	0.252
(t-stat)	(0.357)	(2.754)***	(4.412)***	(0.157)	(1.514)	
EFEMU	0.223	0.081	0.216	0.073	-0.076	0.353
(t-stat)	(2.148)**	(3.620)***	(4.239)***	(0.810)	(-1.912)*	
EFE	0.151	0.059	0.178	0.090	0.067	0.478
(t-stat)	(2.095)**	(3.584)***	(5.711)***	(1.984)**	(3.546)***	
EFP	0.196	-0.013	0.093	-0.150	0.086	0.272
(t-stat)	(2.164)**	(-0.680)	(2.889)***	(-3.087)***	(2.788)***	
EFA	0.141	-0.020	0.110	0.046	0.025	0.176
(t-stat)	(1.616)	(-0.873)	(2.945)***	(1.016)	(1.096)	
EFW	0.254	0.003	0.229	0.092	0.065	0.305
(t-stat)	(2.350)**	(0.122)	(4.217)***	(1.158)	(1.996)**	

Table 4: Alternative Predictors: Dummy Variable Regression Results

This table presents results from monthly regressions of mutual fund performance in the subsequent year (from month t+1 to month t+12) on alternative predictor quintile dummies. Results are reported for the following predictors as alternatives to the Stiftung Warentest rating: the benchmark-adjusted fund return (BaR) in Panel A, the Jensen (1968) one factor alpha (α^{Jen}) in Panel B, and the Carhart (1997) four factor alpha (α^{Car}) in Panel C. All predictors are measured using 60 months of return data prior to the sorting month. Mutual fund performance is assessed using the same measures. We restrict ourselves to the total fund sample in this table. Fama-MacBeth regression coefficients are the times-series average of the monthly cross-sectional regression coefficients. t-statistics (in parentheses) are calculated using the time-series standard error of the mean and adjusted for heteroskedasticity using the Newey-West procedure with a lag of eleven months. See section 2.3 for a description of the methodology. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level. The last column of this table reports studentised p-values from the monotonic relation (MR) test of Patton and Timmermann (2010) applied to the quintile portfolios based on all possible pair-wise comparisons (MR^{all}). To compute the p-values we use the Matlab code provided by Andrew Patton on his web-site.

Performance Measure (in %)	**γ1**	**γ2**	**γ3**	**γ4**	**γ5**	**Adj. R2**	**F-Test**	**MR^{all} p-value**
Panel A: BaR as alternative predictor								
BaR	-0.118	-0.018	0.014	0.065	0.130	0.029	6.278	0.000
(t-stat)	(-4.374)***	(-0.679)	(0.416)	(2.287)**	(2.246)**			
Jensen alpha	-0.211	0.025	0.060	0.111	0.180	0.029	5.893	0.000
(t-stat)	(-4.168)***	(0.999)	(2.002)**	(5.060)***	(4.710)***			
Carhart alpha	-0.248	0.045	0.051	0.097	0.159	0.015	5.278	0.001
(t-stat)	(-4.942)***	(1.683)*	(1.756)*	(3.056)***	(3.575)***			
Panel B: Jensen alpha as alternative predictor								
BaR	-0.139	-0.001	0.026	0.089	0.177	0.031	6.303	0.000
(t-stat)	(-5.429)***	(-0.037)	(0.785)	(3.428)***	(3.834)***			
Jensen alpha	-0.215	0.002	0.023	0.065	0.149	0.039	7.429	0.000
(t-stat)	(-3.629)***	(0.086)	(0.827)	(2.184)**	(3.154)***			
Carhart Alpha	-0.245	0.045	0.043	0.098	0.148	0.015	5.120	0.047
(t-stat)	(-4.788)***	(1.794)*	(1.169)	(3.052)***	(3.224)***			
Panel C: Carhart alpha as alternative predictor								
BaR	-0.141	0.007	0.034	0.090	0.171	0.024	5.011	0.000
(t-stat)	(-7.930)***	(0.377)	(1.449)	(6.390)***	(4.646)***			
Jensen alpha	-0.217	0.042	0.069	0.113	0.180	0.023	4.931	0.000
(t-stat)	(-4.816)***	(1.775)*	(2.555)***	(5.630)***	(5.531)***			
Carhart alpha	-0.264	0.061	0.084	0.114	0.167	0.011	3.558	0.000
(t-stat)	(-6.934)***	(4.274)***	(5.629)***	(4.622)***	(4.588)***			

Table 5: Alternative Predictors: Trading Strategy Results

This table presents results of a fund trading strategy using the Jegadeesh and Titman (1993) methodology. Fund portfolios are of equal size and sorted on the basis of the following alternative predictors to the Stiftung Warentest rating: the benchmark-adjusted fund return (BaR) in Panel A, the Jensen (1968) one factor alpha (α^{Jen}) in Panel B, and the Carhart (1997) four factor alpha (α^{Car}) in Panel C. All predictors are measured using 60 months of return data prior to the sorting month. We report benchmark-adjusted returns for portfolios Q1 and Q5 which are computed as equal-weighted average of individual benchmark-adjusted returns of all funds within quintile 1 and quintile 5. (Q5 - Q1) reports the return to a hypothetical zero-cost strategy investing long (short) in the Q5 (Q1) portfolio. We restrict ourselves to the total fund sample in this table and investigate holding periods of one, three, six, twelve, 24 and 36 months in all Panels. See section 2.3 for a description of the methodology. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level. In addition, the last row of each Panel reports studentised p-values from the monotonic relation (MR) test of Patton and Timmermann (2010) applied to the quintile portfolios based on all possible pair-wise comparisons (MR^{all}). To compute the p-values we use the Matlab code provided by Andrew Patton on his web-site.

Panel A: BaR over previous 60 months as alternative predictor						
Portfolio	**Rebalancing Frequency**					
	T=1	**T=3**	**T=6**	**T=12**	**T=24**	**T=36**
Q1	-0.164	-0.146	-0.135	-0.118	-0.089	-0.072
(t-stat)	(-1.804)*	(-1.592)	(-1.475)	(-1.273)	(-0.944)	(-0.756)
Q5	0.094	0.089	0.085	0.074	0.058	0.042
(t-stat)	(0.747)	(0.719)	(0.705)	(0.643)	(0.524)	(0.395)
Q5-Q1	0.258	0.235	0.220	0.192	0.147	0.114
(t-stat)	(3.089)***	(2.901)***	(2.946)***	(2.866)***	(2.551)**	(2.323)**
MR^{all} p-value	0.001	0.004	0.006	0.027	0.140	0.288

Panel B: Jensen alpha over previous 60 months as alternative predictor						
Portfolio	**Rebalancing Frequency**					
	T=1	**T=3**	**T=6**	**T=12**	**T=24**	**T=36**
Q1	-0.170	-0.174	-0.159	-0.139	-0.111	-0.094
(t-stat)	(-1.911)*	(-1.927)*	(-1.754)*	(-1.528)	(-1.194)	(-1.002)
Q5	0.119	0.116	0.118	0.107	0.090	0.072
(t-stat)	(0.954)	(0.939)	(0.976)	(0.912)	(0.798)	(0.663)
Q5-Q1	0.290	0.290	0.277	0.246	0.201	0.166
(t-stat)	(3.497)***	(3.567)***	(3.564)***	(3.461)***	(3.134)***	(2.877)***
MR^{all} p-value	0.008	0.000	0.004	0.036	0.208	0.392

Panel C: Carhart alpha over previous 60 months as alternative predictor						
Portfolio	**Rebalancing Frequency**					
	T=1	**T=3**	**T=6**	**T=12**	**T=24**	**T=36**
Q1	-0.125	-0.140	-0.139	-0.122	-0.110	-0.100
(t-stat)	(-1.384)	(-1.530)	(-1.533)	(-1.351)	(-1.211)	(-1.095)
Q5	0.076	0.076	0.077	0.072	0.063	0.054
(t-stat)	(0.645)	(0.651)	(0.666)	(0.630)	(0.559)	(0.482)
Q5-Q1	0.201	0.216	0.216	0.194	0.174	0.154
(t-stat)	(2.930)***	(3.248)***	(3.365)***	(3.277)***	(3.000)***	(2.751)***
MR^{all} p-value	0.118	0.001	0.005	0.003	0.006	0.013

Table 6: Performance Predictors and Fund Expenses

This table reports average total expense ratios (TER) for fund portfolios sorted into quintiles according to the following performance predictors: Stiftung Warentest rating, benchmark-adjusted fund return (BaR), Jensen (1968) one factor alpha (α^{Jen}), and Carhart (1997) four factor alpha (α^{Car}). Total expense data is from Morningstar Direct and covers the time period since 2005. It is measured in % per year and refers to the latest financial year of the fund preceding the end of the calendar years 2005, 2006 or 2007. All predictors are measured using 60 months of return data prior to December of every calendar year. The second last row (Q1-Q5) gives the difference in the total expense ratio between low ranked and high ranked funds. The last row gives the t-statistic associated with this value. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level.

	Total Expense Ratio (in %)			
Predictor Quintile	**Rating Stiftung Warentest**	**BaR**	**Jensen alpha**	**Carhart alpha**
Q1	1.877	1.884	1.885	1.951
Q2	1.796	1.743	1.765	1.779
Q3	1.762	1.656	1.656	1.712
Q4	1.701	1.723	1.692	1.640
Q5	1.695	1.838	1.843	1.762
Q5-Q1	-0.182	-0.047	-0.042	-0.189
t-stat	(-5.238)***	(-1.263)	(-1.116)	(-5.042)***

Table 7: Fund Activity: Dummy Variable Regression Results

This table presents results from monthly regressions of mutual fund performance in the subsequent year (from month t+1 to month t+12) on three measures of fund activity which are used as performance predictors. In Panel A, funds are sorted into quintiles based on their R^2 obtained from a four factor regression. Funds with the lowest R^2 are assigned to quintile 5; funds with the highest R^2 are assigned to quintile 1. In Panel B, the standard deviation of the residual ($Tracking\ Error$) from a four factor regression is used as a predictor. Funds in quintile 1 (5) have the lowest (highest) $Tracking\ Error$. R^2 and $Tracking\ Error$ are estimated using 60 months of return data prior to the sorting month. In Panel C, we report the results for $Active\ Share$, which is calculated as in Cremers and Petajisto (2009) with the adjustments described in section 4. Mutual fund performance is assessed using the benchmark-adjusted fund return (BaR), the Jensen (1968) one factor alpha (α^{Jen}), and the Carhart (1997) four factor alpha (α^{Car}). We restrict ourselves to the total fund sample in this table. Fama-MacBeth regression coefficients are the times-series average of the monthly cross-sectional regression coefficients. t-statistics (in parentheses) are calculated using the time-series standard error of the mean and adjusted for heteroskedasticity using the Newey-West procedure with a lag of eleven months. See section 2.3 for a description of the methodology. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level. The last column of this table reports studentised p-values from the monotonic relation (MR) test of Patton and Timmermann (2010) applied to the quintile portfolios based on all possible pair-wise comparisons (MR^{all}). To compute the p-values we use the Matlab code provided by Andrew Patton on his web-site.

Performance Measure (in %)	**γ1**	**γ2**	**γ3**	**γ4**	**γ5**	**Adj. R2**	**F-Test**	**MR^{all} p-value**
Panel A: R^2 from four factor model as alternative predictor								
BaR	-0.181	0.060	0.075	0.159	0.204	0.031	7.260	0.005
(t-stat)	(-7.301)***	(3.336)***	(4.236)***	(4.601)***	(3.870)***			
Jensen alpha	-0.217	0.067	0.080	0.132	0.126	0.038	8.412	0.141
(t-stat)	(-8.305)***	(2.832)***	(3.028)***	(2.932)***	(1.694)*			
Carhart alpha	-0.214	0.036	0.033	0.072	0.040	0.010	2.682	0.826
(t-stat)	(-5.946)***	(1.481)	(1.484)	(2.578)***	(0.987)			
Panel B: Tracking Error from four factor model as alternative predictor								
BaR	-0.180	0.054	0.076	0.142	0.225	0.040	8.345	0.005
(t-stat)	(-7.791)***	(3.021)***	(3.900)***	(4.838)***	(3.470)***			
Jensen alpha	-0.197	0.044	0.071	0.102	0.138	0.044	10.086	0.001
(t-stat)	(-8.505)***	(2.029)**	(2.869)***	(2.632)***	(1.858)*			
Carhart alpha	-0.219	0.016	0.062	0.049	0.077	0.011	2.855	0.590
(t-stat)	(-6.678)***	(0.640)	(2.643)***	(1.814)*	(1.646)			
Panel C: Active Share as alternative predictor								
BaR	-0.177	0.033	0.116	0.221	0.330	0.166	9.745	0.000
(t-stat)	(-6.243)***	(1.428)	(3.378)***	(2.406)**	(2.322)**			
Jensen alpha	-0.226	0.029	0.083	0.186	0.193	0.121	6.741	0.165
(t-stat)	(-5.591)***	(1.252)	(1.798)*	(1.651)*	(1.356)			
Carhart alpha	-0.214	0.027	0.040	0.073	0.094	0.062	4.931	0.609
(t-stat)	(-3.299)***	(0.968)	(0.886)	(1.519)	(0.817)			

Table 8: Fund Activity: Trading Strategy Results

This table presents results of a fund trading strategy using the Jegadeesh and Titman (1993) methodology. Fund portfolios are of equal size and sorted on the basis of three measures of fund activity which are used as performance predictors. In Panel A, funds are sorted into quintiles based on their R^2 obtained from a four factor regression. Funds with the lowest R^2 are assigned to quintile 5; funds with the highest R^2 are assigned to quintile 1. In Panel B, the standard deviation of the residual ($Tracking\ Error$) from a four factor regression is used as a predictor. Funds in quintile 1 (5) have the lowest (highest) $Tracking\ Error$. R^2 and $Tracking\ Error$ are estimated using 60 months of return data prior to the sorting month. In Panel C, we report the results for $Active\ Share$, which is calculated as in Cremers and Petajisto (2009) with the adjustments described in section 4. We report benchmark-adjusted returns for portfolios Q1 and Q5 which are computed as equal-weighted average of individual benchmark-adjusted returns of all funds within quintile 1 and quintile 5. (Q5-Q1) reports the return to a hypothetical zero-cost strategy investing long (short) in the Q5 (Q1) portfolio. We restrict ourselves to the total fund sample in this table and investigate holding periods of one, three, six, twelve, 24 and 36 months in all Panels. See section 2.3 for a description of the methodology. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level. In addition, the last row of each Panel reports studentised p-values from the monotonic relation (MR) test of Patton and Timmermann (2010) applied to the quintile portfolios based on all possible pair-wise comparisons (MR^{all}). To compute the p-values we use the Matlab code provided by Andrew Patton on his web-site.

Panel A: R^2 from four factor model as alternative predictor

Portfolio	**Rebalancing Frequency**					
	T=1	**T=3**	**T=6**	**T=12**	**T=24**	**T=36**
Q1	-0.176	-0.174	-0.172	-0.169	-0.159	-0.153
(t-stat)	(-2.756)***	(-2.657)***	(-2.559)***	(-2.469)**	(-2.255)**	(-2.129)**
Q5	0.054	0.063	0.064	0.058	0.053	0.050
(t-stat)	(0.405)	(0.477)	(0.491)	(0.450)	(0.407)	(0.387)
Q5-Q1	0.230	0.237	0.236	0.228	0.211	0.203
(t-stat)	(2.127)**	(2.207)**	(2.259)**	(2.189)**	(2.057)**	(2.018)**
MR^{all} p-value	0.000	0.000	0.002	0.006	0.019	0.012

Panel B: Tracking Error from four factor model as alternative predictor

Portfolio	**Rebalancing Frequency**					
	T=1	**T=3**	**T=6**	**T=12**	**T=24**	**T=36**
Q1	-0.178	-0.176	-0.173	-0.170	-0.159	-0.153
(t-stat)	(-2.791)***	(-2.716)***	(-2.564)**	(-2.485)**	(-2.286)**	(-2.158)**
Q5	0.069	0.074	0.076	0.078	0.079	0.077
(t-stat)	(0.474)	(0.510)	(0.530)	(0.542)	(0.556)	(0.544)
Q5-Q1	0.246	0.250	0.249	0.247	0.238	0.230
(t-stat)	(2.037)**	(2.092)**	(2.110)**	(2.097)**	(2.035)**	(2.007)**
MR^{all} p-value	0.002	0.004	0.005	0.006	0.006	0.007

Panel C: Active Share as alternative predictor						
Portfolio	**Rebalancing Frequency**					
	T=1	**T=3**	**T=6**	**T=12**	**T=24**	**T=36**
Q1	-0.151	-0.163	-0.172	-0.174	-0.174	-0.175
(t-stat)	(-2.360)**	(-2.861)***	(-3.098)***	(-3.195)***	(-3.250)***	(-3.340)***
Q5	0.248	0.252	0.234	0.233	0.247	0.255
(t-stat)	(1.293)	(1.311)	(1.211)	(1.194)	(1.265)	(1.304)
Q5-Q1	0.399	0.415	0.407	0.407	0.421	0.430
(t-stat)	(2.171)**	(2.260)**	(2.226)**	(2.225)**	(2.310)**	(2.352)**
MR^{all} p-value	0.193	0.109	0.059	0.031	0.025	0.035

Table 9: Double Sorts Based on Stiftung Warentest Rating and Fund Activity Measure

This table presents results of a fund trading strategy using the Jegadeesh and Titman (1993) methodology with an annual rebalancing frequency. Funds are first sorted by their Stiftung Warentest rating into five different groups of equal size. For each group we further subdivide funds into five portfolios based on their measure of fund activity ($Tracking\ Error$ in Panel A and $Active\ Share$ in Panel B). $Tracking\ Error$ is obtained from a four factor regression. For every sorting, we consider only funds for which we have non-missing information about the Stiftung Warentest rating and the specific fund activity measure. We report average benchmark-adjusted returns for the different fund portfolios which are computed as equal-weighted average of individual benchmark-adjusted returns of all funds within the particular portfolio. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level. In addition, for every row and every column of each Panel, the table reports studentised p-values from the monotonic relation (MR) test of Patton and Timmermann (2010). The final row (final column) in each panel presents the p-value from a joint test for a monotonic relation between the portfolio sorting variable and benchmark-adjusted fund returns, computed across all column (row) portfolios. The bottom-right number in each panel is the p-value for the joint test for a monotonic relation in both variables.

Panel A: BaR of 25 different fund portfolios sorted by their Stiftung Warentest rating and Tracking Error assuming annual rebalancing

Rating	Tracking Error quintile						MR	Joint MR
quintile	Low	2	3	4	High	High-Low	p-value	p-value
Low	-0.239	-0.230	-0.201	-0.132	0.077	0.316	0.010	
(t-stat)	(-3.821)***	(-2.615)***	(-2.067)**	(-1.132)	(0.508)	(2.502)**		
2	-0.235	-0.213	-0.220	-0.132	0.050	0.285	0.171	
(t-stat)	(-3.115)***	(-2.500)**	(-2.242)**	(-1.164)	(0.347)	(2.227)**		
3	-0.151	-0.161	-0.148	-0.010	0.075	0.225	0.306	0.038
(t-stat)	(-2.284)**	(-1.992)**	(-1.464)	(-0.092)	(0.511)	(1.836)*		
4	-0.125	-0.095	-0.046	0.086	0.172	0.296	0.019	
(t-stat)	(-2.101)**	(-1.101)	(-0.468)	(0.679)	(1.035)	(2.153)**		
High	-0.086	-0.082	0.034	0.191	0.281	0.367	0.060	
(t-stat)	(-1.208)	(-0.972)	(0.341)	(1.403)	(1.660)*	(2.677)***		
High-Low	0.153	0.149	0.235	0.322	0.204			
(t-stat)	(3.946)***	(2.753)***	(3.747)***	(3.446)***	(2.600)***			
MR p-value	0.012	0.017	0.608	0.055	0.209			
Joint MR p-value			0.114					0.040

Panel B: BaR of 25 different fund portfolios sorted by their Stiftung Warentest rating and Active Share assuming annual rebalancing								
Rating	Active Share quintile						MR	Joint MR
quintile	Low	2	3	4	High	High-Low	p-value	p-value
Low	-0.244	-0.231	-0.258	-0.023	0.242	0.485	0.156	
(t-stat)	(-3.758)***	(-3.117)***	(-2.853)***	(-0.219)	(1.073)	(2.303)**		
2	-0.209	-0.137	-0.159	-0.055	0.139	0.348	0.171	
(t-stat)	(-3.260)***	(-1.438)	(-1.658)	(-0.429)	(0.854)	(2.227)**		
3	-0.178	-0.142	-0.069	-0.013	0.208	0.387	0.101	0.147
(t-stat)	(-2.520)**	(-2.154)**	(-0.676)	(-0.096)	(1.066)	(2.157)**		
4	-0.123	-0.065	-0.036	0.289	0.312	0.436	0.168	
(t-stat)	(-1.649)	(-0.706)	(-0.338)	(1.701)*	(1.336)	(1.990)**		
High	-0.023	-0.074	0.092	0.166	0.293	0.316	0.479	
(t-stat)	(-0.287)	(-0.884)	(0.751)	(0.951)	(1.450)	(1.953)**		
High-Low	0.221	0.176	0.312	0.189	0.052			
(t-stat)	(3.684)***	(2.331)**	(3.459)***	(1.604)	(0.396)			
MR p-value	0.001	0.002	0.003	0.769	0.430			
Joint MR p-value			0.318					0.221