

# Fooled by Randomness: Investor Perception of Fund Manager Skill\*

Justus Heuer, Christoph Merkle, Martin Weber<sup>†</sup>

July 2014

## Abstract

Return chasing investors almost exclusively consider top-performing funds for their investment decision. When drawing conclusions about managerial skill of these top performers, they neglect cross-sectional information and volatility. We show that they fail to understand that, in large populations of mutual funds, a few will outperform by pure chance. In a series of surveys, we demonstrate that investors entirely ignore cross-sectional information and regard fund information in isolation only. In addition, investors do not sufficiently account for volatility and are thus likely to confuse risk taking with skill. In large mutual fund populations, this can lead to an over-allocation of capital to lucky past winners and to excessive risk taking by fund managers in order to attract inflows.

*JEL-Classification Codes:* D14, D83, G02, G11.

*Keywords:* Mutual Fund Performance, Skill, Luck, Return Chasing, False Discovery.

---

\*We are grateful to FAZ and Handelsblatt for allowing us to present the studies and invite participants in their newspapers. We thank Sebastian Müller, Philipp Wiederhold, and seminar participants in Mannheim.

<sup>†</sup>Department of Banking and Finance, University of Mannheim, L5, 2, 68131 Mannheim, Germany (corresponding author: [chmerkle@mail.uni-mannheim.de](mailto:chmerkle@mail.uni-mannheim.de), +49 621 1811531).

# 1 Introduction

Mutual fund investors chase returns. The Investment Company Institute (2006) reports that historical returns are the piece of information considered most important when purchasing a fund. Sirri and Tufano (1998) show that investors chase returns of top performers and fail to flee poor performers. Among others, Lynch and Musto (2003) confirm this convex relation between past returns and fund flows. Barber et al. (2014) suggest that investors actually chase the risk-adjusted performance (alpha in a one-factor model) and that this alpha chasing persists even in a four-factor model when the momentum effect is controlled for. Return chasing behavior is at odds with Carhart (1997), who shows that there is no persistence in fund returns (after controlling for the momentum effect).<sup>1</sup>

So why do mutual fund investors chase returns or alphas if this strategy is controversial at best and even if almost every mutual fund information document includes the phrase that “*Past performance is no guarantee for future results*”? We show that two misperceptions of fund manager performance data induce investors to chase alpha: First, they fail to sufficiently consider the influence of volatility on the reliability of past performance as an indicator of fund manager skill. Everything else constant, higher volatility of fund returns will lead to less significant—and thus less reliable—alpha estimates. No matter whether investors chase returns or alphas, when they disregard volatility, they are more likely to end up with a risky than with a safe fund. Second, we show that investors are subject to a selection bias when chasing top-performers. They would have to look at the entire population of fund returns to obtain an unbiased estimate of the likelihood a past outperformer is managed by a skilled manager. The severity of this selection neglect increases with the size of the fund population.

Throughout the paper, we will refer to a manager with a true positive after fees alpha as a *skilled* manager. Skilled managers stand in competition to *zero-skilled* managers, who have true after fees alphas indistinguishable from zero and *unskilled* managers, who have true after fees negative alphas.<sup>2</sup> In order to test for the reliability of alpha estimates, significance levels are determined, which are a function of alpha and its standard error. Therefore, when investors chase alpha, they should

---

<sup>1</sup>A vast strand of literature examines the profitability of return chasing: Gruber (1996), Zheng (1999) and Keswani and Stolin (2008) show that money is “smart” and funds outperform following inflows. This smart money effect is very short term, however, and investors hold funds longer than the effect persists. Frazzini and Lamont (2008), to the contrary, suggest investors are “dumb money” and the smart money effect is limited to about one quarter, while in the long run the dumb money effect dominates. Friesen and Sapp (2007) show that the smart money effect can be explained by momentum. They find that investors—compared to a buy-and-hold strategy—underperform by 1.56% annually. Choi et al. (2010) show that investors are willing to pay higher fees for index funds with higher past returns than for otherwise identical index funds with lower past returns, where return differences were purely a result of time horizon since inception.

<sup>2</sup>The definitions are adapted from Barras et al. (2010).

not only consider mere alpha values but have to account for the standard error—depending on the fund’s volatility—as well.

The literature provides extensive evidence that people have difficulties dealing with volatility. Ehm et al. (2013) find that investors are unable to distinguish between different levels of volatility. Huang et al. (2012) show that naïve investors disregard performance volatility when building expectations about managerial skill from past performance. Zion et al. (2010) show that adaptive behavior in the form of return chasing leads to underdiversification. Extending their argument, alpha chasing without accounting for volatility would lead investors to concentrate on few funds with high idiosyncratic risk. Ert and Erev (2007) demonstrate that, as the number of investment alternatives increases, individuals are more likely to select a risky alternative even though it has a lower expected payoff. This is because the best performer is more likely to be drawn from a risky distribution than from a less risky one.

But even if investors did account for volatility, even if they were chasing significant alpha instead of large alpha, they would be a long way from having identified a skilled manager: even a significant alpha does not safely indicate skill. A positive and significant alpha at 5% level (two-sided) might be with 2.5% probability just a lucky realization from a distribution of alphas with a true mean of zero. Consequently, in a population of 100 true zero-alpha funds, on average 2.5 funds will appear to have significantly positive alpha by pure chance. If there are 1000 funds with a true alpha of zero this number will increase to 25 and so on. Put differently, in a sample entirely made up of zero-alpha funds, if an investor chases alphas significant at the 5% level, there still is a 100% probability to end up with a zero-alpha fund, or, as Barras et al. (2010) call it, to have a *false discovery*.

The larger the fund population becomes, the more likely it is that there are several funds with significant alpha by pure chance. When solely considering funds with significant alpha, investors suffer from a selection neglect. In order to evaluate the reliability of their alpha estimate, investors have to take into account information from the cross-section of all fund returns. Specifically, they would have to determine whether the set of significant alpha funds was larger than predicted by chance. Only if this is the case, they can expect there to be skilled managers within that set. Following this reasoning, Barras et al. (2010) show that it is possible to estimate the proportion of skilled, unskilled, and zero-skilled fund managers within a set of funds. Rational investors should select a fund from the set of funds with the highest proportion of skilled managers.

These population effects are not intuitive. Consider the probability that a fund outperforms the market over ten consecutive years. Under the assumption that the probability of outperformance in

each year is 50% and independent across years (implying little or no skill), this probability amounts to 0.1%. However, in a large fund population of 10,000 funds it is almost certain ( $p=0.9999$ ) that at least one fund outperforms the market in all ten years. The difficulty in understanding lies in a presumably unlikely event becoming very likely in large populations. While the ex ante probability of a specific fund to consistently outperform is very low, it is very high for one of the top performers. But exactly these very good funds are the ones which investors look at when they chase returns. Without realizing, they are evaluating a biased sample. To debias the sample, investors have to consider the entire cross section of mutual fund returns.

Koehler and Mercer (2009) are—to our knowledge—the first to experimentally analyze if subjects are able to improve conclusions about fund manager skill by looking at cross-sectional information. They ask subjects to predict future performance of certain advertised funds and find that investors deem advertised funds representative for the entire population of funds by the same manager. However, if informed previously that, in addition to the advertised fund, the manager manages several other funds, subjects estimate future performance to be significantly worse than without the prompt. Investors realize that the sample of advertised funds they see is likely biased because fund managers will prefer to advertise successful funds instead of unsuccessful funds.

In a social context, Han and Hirshleifer (2012) describe a similar problem caused by the tendency of investors to talk more frequently about their winning investments than their failures. Again, this favors more volatile investments and produces a selection bias. Heimer and Simon (2012) confirm that success of investing in particular assets induces increased investment in these assets by the social peer group. However, as in the advertisement context, the discussed investments are more likely to include top performers. As in the unprompted treatment of Koehler and Mercer (2009), investors ignore the fact that they only see a preselected part of the total population of investment results and misinterpret this sample as representative, which then creates biased return expectations.

With the help of methodology introduced by Barras et al. (2010), we show formally that the volatility of fund returns and the number of funds in the population both influence the reliability of alpha as an indicator of true fund manager skill. Further, in a series of surveys, we investigate whether investors take into account volatility and are able to infer information from the cross section of fund returns that is needed to obtain an unbiased estimate of fund manager skill.

We conduct three online surveys, with different groups of participants. In our first survey, we show participants the price charts of fund populations with top-performers that all had different alphas and volatilities but equally significant alphas. Of every fund population, we generate versions

with different population sizes and therefore different degrees of selection neglect. Participants are asked to estimate the probability that the top-performing fund is managed by a skilled manager. Despite equal significance, participants have far larger confidence in fund manager skill of volatile top performers with high absolute alpha than in the skill of less volatile low-alpha top-performers. Further, participants entirely disregard information from the cross section and ignore that in larger populations, it is more likely that at least one zero-alpha fund has a significant alpha estimate. They were equally confident in the skill of fund managers managing top-performers in large populations as in small populations when both top-performers had equally significant alphas.

To validate our results on volatility, we conduct a second survey with readers of the online version of a large German financial newspaper. In a number of different scenarios, we ask subjects to compare the probability of skill of two funds with equally significant alphas and equal Sharpe ratios, but different alphas, annual returns and volatilities. Participants, on average, have significantly larger confidence in the skill of managers of volatile funds with higher alphas. A control group of participants, who compare the risk-return relation of the two funds, does not systematically assign higher Sharpe ratios to the riskier funds, confirming the validity of our survey set up. In a final experiment, we repeat this set up with more scenarios in order to confirm results. Subjects evaluate price charts showing two funds with equally significant alphas and equal Sharpe ratios. In every price chart, one fund has higher alpha and volatility and the other fund has lower alpha and volatility. Again, a significant majority of subjects had higher confidence in the abilities of the manager of the more volatile fund.

We are aware that it is virtually impossible to precisely judge the probability that a fund is managed by a skilled manager by simply looking at a price chart—and this is not the aim of the paper. For the price charts we present in our surveys, we ensured that it is evident the funds are part of fund populations of different size and that participants can easily recognize that funds have different volatilities. We test whether subjects understand the general idea of these concepts, not whether they provide an exact probability estimate. Our results indicate that investors chase large alphas and large returns without sufficiently accounting for volatility, which impairs the reliability of the past performance as an indicator of manager skill. We further show that investors look at top performers in isolation, without realizing that ignoring the cross-section of other funds introduces a selection bias. When chasing alphas or returns, investors are thus subject to selection neglect.

Our paper makes a threefold contribution to the literature: We are the first to conduct an in detail analysis of investor behavior that leads to return chasing and to show that there are two

hurdles investors fail to overcome to arrive at an unbiased estimation of fund manager skill. This complements recent empirical literature on the distinction of skill and luck in real fund populations (Barras et al. (2010); Cuthbertson et al. (2008)). Second, we confirm the hypothesis by Huang et al. (2012) that naïve investors fail to sufficiently account for volatility when drawing conclusions about managerial skill from past performance. Thirdly, we show that investors are subject to a selection neglect when estimating the skill of top-performers. Similar to findings by Heimer and Simon (2012), investors fail to take into account that they limit their analysis to a biased sample when chasing alphas.

Our findings have major policy implications: Traditionally, regulators have tried to fight return chasing by requiring a mandatory performance warning on any type of promotional material that uses past performance data.<sup>3</sup> Apparently, this warning does not have much impact, investors still chase alphas in empirical data as new as 2012 (Barber et al., 2014). It might be more successful to attack the problem by its roots and find appropriate phrases that directly target the two mistakes investors make in interpreting past return data.

The remainder of the paper is structured as follows: Section 2 formally shows how volatility and population size influence the reliability of alpha estimates and develops our hypotheses on investor behavior. Section 3 describes the setup of our first survey and presents the main results, section 4 describes the setup and results of the second and third survey. Section 5 sums up our findings and concludes.

## 2 Theoretical Foundation

The variance of a fund  $i$ 's returns can be decomposed into a systematic part, which is explained by the market return and an idiosyncratic part, which is orthogonal to the market.

$$\sigma_i^2 = \beta_i^2 \sigma_m^2 + \sigma_{\varepsilon,i}^2, \quad (1)$$

where  $\sigma_i$  is the volatility of the fund returns,  $\sigma_m$  is the volatility of the market returns,  $\beta$  is the systematic risk exposure of fund  $i$  to the market and  $\sigma_\varepsilon$  is the idiosyncratic volatility of fund  $i$ . Any fund manager who within the same strategy doubles the amount of risk, e.g., by doubling the amount invested in the risky portfolio, therefore doubles both, systematic or market risk exposure

---

<sup>3</sup>Typical wording is “*Past performance is not a reliable indicator of future performance*”, or, as mentioned before, “*Past performance is no guarantee for future results*”. Regulators requiring such a warning include the ASIC in Australia, the SEC in the US, the FCA in the UK and the BaFin in Germany.

and idiosyncratic risk exposure. Barber et al. (2014) show that investors are able to account for the market risk exposure of a fund and chase alpha in the one-factor market model proposed by Jensen (1969):

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \varepsilon_{i,t}. \quad (2)$$

The p-value of alpha estimated from a regression of fund  $i$ 's returns gives the probability that this alpha value has been generated by pure chance by a fund with a true alpha of zero:

$$p(\hat{\alpha}_i) = F_n(t). \quad (3)$$

Here  $F_n(t)$  is the cumulative distribution function of a standardized Student's t-distribution with  $n$  degrees of freedom, mean zero, variance one and

$$t = \frac{\hat{\alpha}_i}{SE(\hat{\alpha}_i)}, \quad (4)$$

where  $SE(\hat{\alpha}_i)$  is the standard error of alpha.  $SE$  is an increasing function in the idiosyncratic risk  $\sigma_\varepsilon$  of fund  $i$ . Therefore,  $t$  is decreasing in the idiosyncratic risk of fund  $i$ . Since  $F_n(t)$  is an increasing function of  $t$ , it holds that

$$\frac{\partial p(\hat{\alpha}_i)}{\partial \sigma_{\varepsilon,i}} < 0. \quad (5)$$

The significance of alpha is decreasing in the idiosyncratic volatility of the fund. Therefore, even if idiosyncratic risk can be diversified and is not rewarded, it decreases the reliability of the alpha estimate and therefore has to be considered by the investor in order to evaluate the probability that a fund manager is skilled.

The significance of alpha is equivalent to the probability that a *randomly picked* single time series of returns with true zero-alpha has the alpha estimate  $p(\hat{\alpha}_i)$ . If investors chase alpha and exclusively consider the top-performer of an entire fund population, this top-performer will, by virtue of the return ranking, have a relatively low p-value even if all funds in the population have true alphas of zero. The probability that the manager of fund  $i$  is really skilled is thus equal to the probability that there is no true zero-alpha fund in a population of size  $N$  with significance of at least  $p(\hat{\alpha}_i)$ :

$$p_i(\textit{skilled}) = \left(1 - \frac{p(\hat{\alpha}_i)}{2}\right)^{\pi_0 N} \quad (6)$$

$\pi_0$  is the share of true zero-alpha funds in the population, which needs to be estimated. We use a framework developed by Barras et al. (2010) (BSW, henceforth) to determine this share. BSW separate the entire mutual fund population into three groups: Funds with true negative alphas are managed by *unskilled* managers, funds with true alphas of zero are managed by *zero-skilled* managers and funds with true positive alphas are managed by *skilled* managers. Figure 1 shows the distribution of t-values for the three groups. For demonstration purposes, we assume t-values of unskilled funds are distributed around -3 and t-values of skilled funds are distributed around 3. In the right tail of the distribution of zero-alpha funds, there are some funds, which have positive and significant alpha estimates but true alphas of zero. Since the area under the curve represents a certain proportion of the population of zero-alpha funds, the total number of lucky funds increases with the population size.

We determine p-values for the hypothesis  $H_0 : \alpha_i = 0$  for every fund in the population. In a population of 100% zero-alpha funds, p-values of the hypothesis  $H_0 : \alpha_i = 0$  should be uniformly distributed between zero and 1. If the true population of funds also includes an unknown number of unskilled and skilled funds, we should find clustering of p-values near zero, indicating a high number of positively or negatively significant alpha estimates. Figure 2 contrasts the distribution of p-values in a hypothetically perfect distribution of zero-alpha funds (upper panel) and in a distribution that includes skilled and unskilled fund managers (lower panel). To estimate the share of zero-skilled fund managers ( $\pi_0$ ) in the lower panel of figure 2, BSW select a p-value to the right of the point where the empirical distribution of p-values begins to be approximately uniform. The share of funds with p-values above that selected threshold, multiplied by  $\frac{1}{1-p}$ , will yield the total share of zero-skilled fund managers.

In the example, the p-value distribution is uniform to the right of 0.3. The density in every p-value block of width 0.1 is 0.075, indicating that, e.g., 15% of all funds in the population have p-values larger than 0.8. According to the estimation, about 75% of all funds have true zero-alpha. Turning to the left part of the distribution in figure 2, we see that 20% of all funds in the population have p-values lower than 0.1. Therefore in this subset of funds, there are  $\frac{7.5\%}{20\%} = 37.5\%$  funds with true zero-alpha, which were simply lucky, and  $\frac{12.5\%}{20\%} = 62.5\%$  skilled funds. This means, even with a p-value lower than 0.1 suggesting a fund is lucky with less than 10% probability, the actual

probability to purchase a fund with true zero-alpha when purchasing a fund with p-value lower 0.1 is 37.5%.<sup>4</sup>

Now, we exclusively look at the top-performing fund in terms of significance of alpha in the fund population. The probability that not even one of the zero-alpha funds has at least the p-value of the alpha estimate of this top-performer is calculated according to equation 6. Differentiating equation 6 with respect to the number of funds  $N$  in the population, we find that

$$\frac{\partial p_i(\textit{skilled})}{\partial N} < 0 \quad \forall p(\alpha_i) < 1, \pi_0 > 0, \quad (7)$$

because  $p(\hat{\alpha}_i)$  does not depend on  $N$ .

Figure 3 summarizes the relationship between the probability that the top-performing fund is managed by a skilled fund manager, the fund's volatility and the number of zero-alpha funds in the population. To generate the figure, we use an exemplary time series of returns along with the population sizes and volatility range from our first survey. We keep annual returns constant for the fund and add linearly scaled disturbances to increase and decrease volatility. We calculate the probability that the fund manager is skilled according to equation 6. In the simple case with only one fund in the population, it is the converse probability of the p-value of the fund. Figure 3 shows that, all else equal, the probability that the fund is managed by a skilled manager decreases in the volatility of the fund returns and in the number of zero-alpha funds in the population. The probability eventually approaches zero for large  $N$  and high volatility.

### 3 Survey 1: Cross Section and Skill

#### 3.1 Survey design

In the first survey, we concentrate on the implications of cross-sectional information. To test how investors judge the probability of a fund manager to be skilled, we ask 656 readers of a large German newspaper (or its online version) to complete our survey. This group of subjects had previously participated in an unrelated study and had communicated their willingness to be considered for future research projects. 234 recipients responded and completed the questionnaire.

In the survey, we confront participants with price charts showing the development of a mutual

---

<sup>4</sup>For a complete description of the methodology refer to Barras et al. (2010), in particular to their internet appendix. An empirical decomposition of the US fund market into the three types of managers can be found in appendix A.

fund population and the stock market in 2012. They then provide an estimate of the probability that the top-performing fund in the population was managed by a skilled fund manager. In total they see eight charts with different fund populations to which they are randomly assigned. Figure 4 gives an overview of the survey design. Our base setup consists of nine different scenarios, out of which each participant sees six. Scenarios differed in two dimensions: the price series of the top-performer and the number of funds in the fund population.

We use three different sets of funds such that the intra-year return pattern was different for every top performer. The first set of funds (see panel A of table 1) includes the top-performer (fund 1) with the largest annual return of all fund sets. In this case, most of the annual return was generated in the first two quarters of 2012. The second top-performer (fund 1 in panel B) has the highest return in the last two quarters of 2012 and generated the majority of its annual return during the third quarter. In the third scenario, the top performer (fund 1 in panel C) has lower annual returns than in the first two scenarios along with lower volatility. It was characterized by a particularly strong fourth quarter. While we do not test any hypotheses regarding the intra-year return structure, the different price charts were created to counterbalance for effects potentially produced by particular patterns or trends in the price development.

The second and more important variation is in terms of population size. The price charts present either two, five, or nine funds, and the rightmost columns of 1 show which funds are included in each scenario. All participants see six of these scenarios, which are randomly selected in such a way that each population size and each set of funds is used twice. The particular population sizes were chosen, because ten price paths (nine funds and the stock market) seem to be the maximum that can be easily identified in one price chart. Since investors have to consider the cross section of p-values, we make sure the fund populations have comparable p-values in all three sets of funds. In the scenarios with nine funds, the range of t-values is between -1 and 1 in steps of 0.25. In the medium population size scenarios, funds with t-values of around -1, -0.5, 0, 0.5 and 1 are included. The two-fund scenario contains only the funds with t-values of -1 and 1.

Market beta of all funds is between 0.9 and 1.1 as we aim to focus the analysis on alpha.<sup>5</sup> Across all three sets of funds (panels A-C), top performers had three performance measures in common: Market beta was slightly larger than one, the t-value of the alpha estimate was around one and the significance (p-value) of the alpha estimate was around 30%. Therefore, the three top

---

<sup>5</sup>Betas could not be fixed to exactly one as the adjustment on other dimensions (e.g., t-value) sometimes required small deviations.

performers had almost identical systematic risk exposure and identically reliable alpha estimates, but different values of alpha and different idiosyncratic risk exposures. Absolute alpha values for the top performers range between 7.0% in fund set C and 17.4% in set A, volatility is between 15.7% and 21.0%. This allows us to obtain first insights on the role of absolute level of alpha and volatility for judgments of skill.

When compared to the other funds in their respective set of funds, the top performers always have the highest absolute return, but the margin by which they outperform the other funds is highest in set B and lowest in set C. The latter set represents also the only instance that the top performer does not have the highest alpha. The top-performing fund has the highest volatility of all funds in set B, but not in the other sets. But despite these differences the sets of funds are very similar in all three dimensions driving the probability that the top-performer is skilled: (1) the p-value of the top performer, (2) the distribution of p-values of all other funds and consequently the share of zero-skilled funds in the population and (3) the number of funds in the population. Therefore, given equal population size, the mathematical probability that the top performer is managed by a skilled manager was about equal in all scenarios.

As figure 4 shows, besides the six randomly drawn baseline scenarios, participants saw two more price charts. The first was one out of three *down* scenarios, which were generated by multiplying all daily returns of the three fund sets A-C with  $-1$ . Consequently, all previous top performers turned into bottom-performers, and participants were asked to estimate the probability that the bottom-performer was managed by an unskilled manager. This was to check whether the evaluation of underperformance was any different from that of outperformance. For this scenario, to which we refer to as the *down* scenario, the population size was fixed at five.

In a final scenario, participants were presented the fund population in panel D of table 1. Panel D differs from panels A-C in the significance level of all funds. Instead of t-values between  $-1$  and  $1$  in  $0.25$  steps in the population of nine funds, funds now have t-values between  $-2$  and  $2$  in  $0.5$  steps. The alpha estimate of the top performer was therefore close to significant at the 5% level. Further, the top-performer in terms of significance of alpha had only the third largest annual alpha and the third largest annual return. In exchange, it was the safest positive return fund in the entire population. Contrary to the seven other scenarios, in this case we asked participants to estimate two probabilities: the probability that fund 1 (the top performer in terms of significance of alpha) and the probability that fund 2 (the top performer in terms of alpha) was managed by a skilled manager. The comparison reveals whether investors chase alpha or significant alpha. We refer to

this scenario as the *wide* scenario.

At the beginning of each page of the survey it is made clear that all funds of the population are presented and subjects do not see a mere sample. Subjects are told that “*The market for equity funds on stocks from country 1 [2,...,8] consists of 2 [5,9] funds. Below, you can see the price development of both [all five, all nine] funds that were traded on stocks from country 1 [2,...,8] in 2012.*” The top performers in each scenario are displayed in pink, all other funds were grey. All fund prices were set to 100 at the beginning of the year. In addition to the prices of the funds, each chart included a price chart of the market, also normalized to 100 at the beginning of the year and presented in red.

Below each of the charts, we ask subjects to estimate the probability (in %) that the pink (top-performing) fund was managed by a skilled fund manager by asking: “*How high would you judge the probability that the fund presented in pink is managed by a skilled fund manager?*” We define the term “skilled fund manager” at the beginning of the survey and below each question as “*managers who will—in the long term—perform better than the market after costs and fees.*” In the *wide* scenario, we additionally ask to estimate the probability that the fund with the highest annual return but not the most significant alpha was managed by a skilled manager. In the *down* scenario, we ask for the probability that the worst performing fund was managed by an unskilled manager. We define unskilled managers as managers, whose fees “*exceed the return they can earn over the market.*” It was made clear that “*in the long term, these funds will underperform the market after fees and costs.*” The order in which all eight scenarios were presented was randomized. Examples for the presented price charts can be found in appendix B.

At the end of the survey, Participants are asked to answer six of the Van Rooij et al. (2011) advanced literacy questions. All selected questions are different from questions previously used with the same subject pool to avoid learning effects. Statistical knowledge is self assessed according to German school grading system. We also ask for socio-demographic details. Finally, subjects have to provide an estimate of the share of skilled, unskilled and zero-skilled managers in the market and self-assess their ability to identify a skilled manager.

### **3.2 Summary statistics and survey responses**

Table 2 shows the average characteristics of all participants in our surveys. In our first survey, between 226 and 234 participants answer questions on their socio-demographic status, their financial literacy and their views of the mutual fund market. Subjects are predominantly male (86%) and around 50 years old. On average, subjects are very well educated with a Master’s degree being the

median education. Median net income is between EUR 2000 and EUR 6000, and 70% are currently invested in mutual funds. 88% have invested in funds before and 86% have invested in stocks. 16% of our subjects categorized themselves as “financial professionals”. Participants grade their statistics knowledge, on average, as satisfactory. Average financial literacy was very high and more than 50% answered all six questions correctly.

Clearly, our sample is not representative for the German general population but closely resembles the relevant population of mutual fund investors. According to Deutsches Aktieninstitut e.V. (2013), only 10% of the German population is invested in funds, compared to over 70% in our sample. Among these investors more than 80% had a monthly net income above EUR 2000, which is in line with answers provided by our participants. Additionally, about half of fund investors are over fifty years old and almost 40% of investors in funds or stocks hold the highest German school degree Abitur. In our sample, around 50% of the subjects are 50 years and older and median education was even a little higher than in the investor population.

Participants did not have much confidence in their ability to identify skilled managers with more than 50% answering they “cannot” or “can rather not” identify skilled managers. We also ask subjects to provide an estimate of the share of skilled, zero-skilled and unskilled managers in the market. They estimate a comparably high 18% share of all fund managers to be skilled and an extremely skeptical 48% share of all managers to be unskilled. Compared to the empirical evidence (see appendix A), a 34% share of zero-skilled managers is low.

In table 3, we evaluate whether personal characteristics influence how people judge the share of skilled and unskilled managers in the real fund market and in our scenarios. The first column presents results of an OLS regression analysis of the answer to the question “*How high would you estimate the share of fund managers who can outperform their benchmark after fees in the long run?*”. Explanatory variables are personal characteristics of the participant. In the second column, the explained variable is the answer to the analogous question for the share of unskilled managers. Interestingly, older participants assume a higher share of skilled managers in the market. Barras et al. (2010) show that the proportion of skilled managers decreases from 14.4% in early 1990 to 0.6% in late 2006. If older people formed their market views earlier, this could explain why they believe the share of skilled managers is higher. Men, on average, estimate the share of skilled managers to be 5.6% lower than women and the share of unskilled managers to be 11.1% higher than women. Higher educated participants estimate a lower share of skilled and a higher share of unskilled managers. In column 3, the dependent variable is the self-assessed ability to identify a

skilled manager. Subjects in a higher income bracket judged their ability to be slightly worse, other personal characteristics do not seem to explain this self-assessment.

Columns 4 to 6 show results from a regression of participants' estimates of the probability that the top-performing fund manager in our eight scenarios is skilled. Consistent with their more optimistic views of the share of skilled fund managers in the real fund market, older participants also assign a higher probability of skill to the fund managers in our survey. There is slight evidence that a sound statistical knowledge leads participants to be more sceptical of fund manager skill. And those who estimate a higher share of managers in the real fund market to be skilled also believe in a higher probability of skill in the survey. A higher estimate of the share of unskilled managers in the real fund market indicates a lower estimate of the skill probability in the survey. We find strong evidence that participants, who are very confident in their ability to identify a skilled manager, estimate a higher probability that the top performer is skilled.

### 3.3 Results

We now turn to the main analysis, whether population size has any impact on submitted skill probabilities. Table 4 provides objective probabilities of fund manager skill, which clearly depend on the number of funds in the population. With two funds the probability of skill for the top performer in all fund sets is above 75%, while it is around 25% for the large populations. These probabilities are calculated by equation 6 using the share of zero-skilled managers, the p-value of the respective top performer, and the population size. The share of zero-skilled managers  $\pi_0$  is estimated according to the BSW methodology. In the base scenarios it is between 70% and 94% given the modest t-values within the fund sets. It drops to 51% for fund set D with its more spread out t-values.

We propose two alternatives to estimate the probability for fund manager skill. One is based on the actual empirical distribution of fund manager skill in the US market 2012, for which we estimated a 100% share of zero-skilled fund managers (see appendix A). Consequently, the probabilities for fund manager skill are even lower than in the BSW case. The other alternative uses the fund market views of participants themselves to generate skill predictions. As they believe in a relatively low share of zero-skilled managers, these probability estimates are considerably higher. But what all methods have in common is that the effect of population size on skill probabilities is strongly visible.

In contrast, participants' actual estimates in the fund price scenarios make no such difference. Not only are the probability estimates for different population sizes statistically indistinguishable,

there also is no trend in either direction. The consequence of the neglect of cross-sectional information is that participants overestimate skill in large populations and underestimate skill in small populations when compared to the objective probabilities. We interpret this as first indication in favor of our hypothesis. The estimates vary between sets of funds but not within each set. Given that the distribution of t-values and p-values was (almost) identical across sets, this presumed skill differential is completely spurious. Participants seem to favor the top performer in fund set A for reasons other than its statistical properties. In the scenario with more pronounced t-values (fund set D), we uncover another bias. The top performer in terms of significance of alpha (fund 1) is assigned a lower probability of skill than the top-performer in terms of absolute alpha and annual return (fund 2), although the theoretical probabilities suggest the opposite.

For the *down* scenarios, in which participants estimate the probability of the fund manager to be unskilled, they submit larger values ( $p < 0.01$ ) than for the corresponding up scenario. This means that they react more harshly if a fund underperforms than when it outperforms. A possible explanation could be Morewedge’s (2009) negativity bias, according to which people are more likely to attribute an outcome to an external agent when the event is negative than when it is positive. This means that subjects estimate negative returns to be more likely generated on purpose—through negative skill—than positive returns. The result also fits well with findings by Chang et al. (2013), who show that delegated assets like mutual funds are subject to a reverse-disposition effect, because investors like to blame poor performance on the manager and punish her for the losses. The fact that investors assume a high probability that funds with poor performance were managed by unskilled instead of unlucky managers could contribute to this reverse disposition effect.

To more formally test whether investors are able to de-bias their conclusions about the top-performer by incorporating cross-sectional information, table 5, panel A, provides t-tests whether probability estimates significantly differ by population size. We find that the average estimates in populations of two, five or nine funds are almost identical. All mean differences in estimates between population sizes are lower than 1% while the mathematical differences, as reported in table 4, are around 30% when moving from a population of two funds to five funds and around 20% when moving from a population of five funds to nine funds. This also holds in the individual scenarios, as for none of the fund sets A-C a population effect is visible. As hypothesized, participants in our survey do not adjust their probability estimates with population size, ignoring the properties from equation 7.

Since skill estimates vary however by the presented fund sets, in panel B, we compare the

skill estimates by the set presented. All differences, aggregated as well as split up by scenarios, are economically and statistically strongly significant, indicating that participants see patterns in the price charts they interpret as evidence for skill. Objectively, as reported in table 4, there are no differences in probability of skill between the fund sets. To examine where these erroneous impressions come from, we turn to a multivariate setting, which allows controlling for different properties of the top performer in each fund set.

In table 6, we explain probability estimates of skill by return, alpha, and volatility of the evaluated fund, and the population size in the respective scenario. We use OLS regressions and Tobit regressions, accounting for the fact that probabilities are bounded between 0 and 100%. We include participant fixed effects to control for all personal characteristics that potentially drive probability estimates, and cluster standard errors by scenario. Consistent with the observation of return and alpha chasing, we find in columns 1 and 2 that annual return and alpha positively influence the estimated probability of skill. For every additional percentage point of annual return, subjects assign about 2% higher probability of skill. The result for alpha is slightly weaker, probably suggesting that investors focus on overall return.

In columns 3-7, we add volatility and population size. As already the insignificant univariate differences suggested, we find no significant impact of the dummy variables for five and nine fund populations. The coefficients are very small given the theoretical values and often point in the wrong direction as larger fund populations should have a negative effect on skill estimates. We conclude that participants behave as hypothesized and fail to account for cross-sectional information. Since, in reality, not only nine but hundreds of funds are competing for investors' money, the importance of cross-sectional information is even larger than in our survey. The failure to take this information into account will lead to an overreliance on performance in determining fund manager skill and offers an explanation for return or alpha chasing.

If subjects sufficiently considered volatility, coefficients for the standard deviation of returns would be negative. This is only the case in two specifications and has to be interpreted with caution. Standard deviation and alpha are, by construction, highly correlated ( $\rho = 0.79$ ) because all funds, except for fund set D, have equal t-values of alpha. Given that a particular year (2012) is used to derive the fund sets, also annual return and alpha are highly correlated ( $\rho = 0.93$ ). Therefore, the design primarily chosen for studying selection neglect is not optimally able to test for the role of volatility and to distinguish between return and alpha, which motivated the follow-up surveys 2 and 3, which we discuss below.

The large differences in estimated probability of skill between the fund sets might be explained by intra-year returns. As explained before price charts are characterized by different return patterns. While the top performer in fund set A generates much of its return in the first half of the year, the top performer in sets B and C have a strong performance in the second half or last quarter of the year, respectively. We examine intra-year performance in column 4 and 5 of table 6. Interestingly, coefficients for both, the second half year return and the fourth quarter return, are smaller than for the rest of the year. A Wald test for the difference in coefficients is highly significant ( $p < 0.01$ ). Presumably, in the presentation of price charts, a fund with strong early return looks much more dominating as it trades above the other funds for the entire year. In contrast, a fund with high last quarter return may end up with the same annual return but runs with the average funds for most of the year (see the fund charts in appendix B).

## 4 Survey 2 and 3: Volatility and Skill

### 4.1 Survey design

The survey design of our second and third survey has as main objective to clarify the role of volatility in the estimates of fund manager skill. As survey one demonstrates that population size is ignored by participants, we do not vary the fund scenarios along this dimension any longer. Instead, the second survey is characterized by the following features: First, we use different years and market return realizations to provide more variety in market environments and to disentangle total return, alpha, and volatility. However, *within* scenarios we only compare funds with identical intra-year return patterns to exclude them as determinant of the probability estimate. Finally, we ask a control group explicitly for volatilities to investigate whether participants are able to infer volatility information from the price charts.

In the third survey we fill some remaining gaps in survey design to rule out two alternative explanations for our findings on the underestimated influence of volatility. First, to exclude the possibility that investors are attributing higher skill to the riskier fund because they believe its higher market beta is a sign of market timing, we add scenarios with negative realizations of the market risk premium. In these years, a higher market beta of the riskier fund would even be evidence of negative market timing skill. Secondly, in case investors can only connect volatility and the probability of skill when the more volatile fund has a lower value than the safe fund for a certain time during the observation period, we ensure that in each scenario the riskier fund trades for a

lower price for a varying number of days.

In study 2, we consider three different markets, the US stock market of either 2010, 2011 or 2012, and for each market two sets of funds, which results in a total of six scenarios. All scenarios contain only two funds as larger populations make (as shown before) no difference in skill estimates but complicate the price charts. One of the two funds always is a risky fund (high return, high volatility) and the other a safer fund (lower return, lower volatility). However, the volatility differential differs between the scenarios within each market year, it is either large or small. Table 7 shows the properties of the used funds. Importantly, p-values of alpha and Sharpe ratios are again very similar in each market, there are no objectively superior funds.

The price charts are generated in an automated fashion by selecting a (real world) mutual fund with a slightly positive alpha for each year and multiplying its daily returns by either 0.7, 0.9, 1.5, and 1.7. We choose this procedure to ensure that the intra-year return patterns for all funds within a given year are identical. The two more extreme funds (0.7, 1.7) are included the scenario with large volatility differential and the more moderate fund in the scenario with small volatility differential. Examples for the price charts can be found in appendix C. Unlike in study one, the procedure does not allow to keep beta constant. As the table shows, there is large variety in return, alpha, and volatility across funds. And while there is marginally significant skill in 2011, there is no skill in the other years, in particular in 2012 where p-values for alpha are close to one.

The test group question set up is comparable to the set up in survey one. A definition of fund manager skill is provided, and the price charts for the funds and the stock market in each scenario are shown in different colors. In contrast to survey one, subjects are not asked to provide a numerical probability estimate that the top-performer was managed by a skilled manager but simply indicate which of the two funds they believe is more likely to be managed by a skilled manager.<sup>6</sup> Subjects can choose either one of the two funds or state that the funds have equal probability of skill. Objectively, since both the riskier and the safer fund in each scenario have almost identical p-values, we should not see systematic preferences for one of the funds.

The control group question was designed to test whether subjects are in principle able to appropriately infer volatility from price charts. We are aware that charts alone do not allow a precise estimation of volatility but aim to rule out the presence of a systematic misestimation potentially driving our results. We ask participants to compare the risk-return relationship of both funds, which

---

<sup>6</sup>Given that the total fund population is potentially larger than the two funds and the entire cross-section is not shown, the information provided is not sufficient for an unbiased numerical probability of skill estimate.

means that, since the annual return is easily visible by the final price, they have to gauge whether it sufficiently compensates for volatility. Again, participants can choose one of the funds as better in terms of risk-return relationship or state that they are equivalent. The (almost) identical Sharpe ratios within the scenarios favor neither of the funds.

Within the survey and prior to the first task, each participant is randomly assigned to either the test or control group. The assignment to either probability of skill or risk-return relationship is permanent; every participant is only asked one type of question. For each year, participants see one of the two scenarios (with small or large volatility difference). The order of years presented is random. After completing all questions on performance, subjects answer the same socio-demographic and market views questions as in survey one.

Similar to survey 2, participants in survey 3 are presented price charts of the market and two different funds. Again, both fund price paths are based on one return sequence but scaled by different factors. Due to this scaling and the mostly very positive returns, in the previous survey the more volatile fund often outperforms in price the less volatile for almost the entire year. In survey 3, we have the two price paths intersect, and the high volatility fund trades at a lower price for between 4 and 195 days (out of 250 trading days). This is to counter the argument that volatility is ignored because the more volatile fund dominates the other fund, or that volatility on the upside does not matter (for estimating the probability of skill it is inconsequential anyway, whether investors associate volatility with risk).

The sixteen new scenarios based on eight different market environments can be found in appendix D. In two markets, the realization of the market risk premium is negative. In all scenarios, the riskier funds have higher annual returns, higher market betas, higher alphas, and higher volatility than the safer funds. All funds in the same market have identical p-values and almost equal Sharpe ratios; p-values range from 12.5% to 65%. In contrast to survey 2, we do not mention the year of the returns and do not make any reference to the US fund market (instead we use the neutral instruction “*the mutual fund market from country A [to H]*”). This is to ensure participants do not base their judgments on memories or associations with a particular year or market.

As in survey 2, participants are assigned to one question mode (for probability of skill or risk-return relationship) at the beginning of the survey and stay within this treatment throughout the survey. The wording of the questions and the response alternatives are the same as in survey 2. Participants are randomly shown one scenario out of two for each of the eight markets. Again, the scenarios are characterized by either high or low difference in volatilities between the two presented

funds. At the end of the survey, we ask the same questions on socio-demographics, market views and financial knowledge as in the first two surveys.

## 4.2 Summary statistics and survey responses

Participants for survey 2 are recruited through the online version of a large German financial newspaper, which is not identical to the outlet used for the first survey. An advertisement on their website invites participation to support research at the University of Mannheim. A total of 1,417 readers respond to the link provided on the news website, 905 complete the survey. We drop every participant who spends less than 150 seconds on our survey, most of them quit on the introduction page or at the first question (nevertheless, available answers of these participants are comparable to those of the remaining participants).

For survey 3, we contact 639 individuals who have previously communicated their willingness to participate in surveys of the University of Mannheim. 200 of them respond, out of which 186 finish the questionnaire. All respondents have participated in a previous (unpublished) survey on presentation formats of asset returns. As the latter survey was conducted five years prior to ours, we do not expect any spillover effects from the older survey. The subject pool is different for all three surveys; however, we cannot rule out with certainty that some individuals participate in more than one survey since they all were (at some point) recruited through the media and remain anonymous.

Table 2 shows average characteristics of the participants in our surveys. As in the first survey, they were predominately male (87% and 89%), but slightly younger in survey 2 and older in survey 3, possibly because the latter were initially recruited five years prior to our survey. Participants are very well educated with a Bachelor's degree as median education level (survey 2), or even a Master's degree (survey 3). Investment experience is also high; most participants have invested in mutual funds and stocks before. A considerable fraction classify themselves as financial professionals (24% and 21%). Self-reported statistics knowledge and tested financial literacy is high and comparable to survey 1. The general beliefs about the distribution of skilled, zero-skilled, and unskilled managers are also very similar across surveys. We conclude that our subject pool is very homogeneous over all three surveys and results should not be driven by idiosyncratic properties of the samples.

## 4.3 Results

If it is the case that investors, on average, underestimate the influence of volatility on the reliability of past performance, then they would select the more volatile fund with higher return as the one

with a superior probability of fund manager skill. This would be consistent with Huang et al. (2012), who suggest that some investors learn about fund manager skill from past performance without sufficiently considering volatility. Secondly, the survey design allows to test whether this bias increases with the difference in volatilities between the presented funds, i.e., whether a larger fraction of participants assigns a higher probability of skill to the more volatile fund when there is a high difference in volatilities (and returns) between the two funds.

In the question mode asking participants to compare the risk-return relation of the two funds, participants do not have to take the additional step to convert return and volatility into skill probabilities. They only have to judge performance based on return and risk, which shows where in the thought process the bias occurs. If participants already favor the riskier fund due to its risk-return characteristics, they would select this fund as the “better” fund also in this treatment. If however the price charts are to communicate returns and volatilities without systematic bias and participants realize that Sharpe ratios are identical, they would be indifferent between the funds or at least would not systematically prefer one of the funds.

In table 8, panel A, we present the responses to the question asking to select the fund with the higher probability of fund manager skill. The table shows the fraction of participants favoring each of the funds and of those who think both funds have equal probability of skill. On aggregate in both surveys, the risky fund is believed to be managed by a skillful manager by a much larger fraction of the sample. The difference is 16%-points in survey 2 and even 24%-points in survey 3, both highly significant ( $p < 0.01$  in a proportion test). The risky fund is even more preferred in the scenarios with large volatility difference between the funds (20 and 26%-points). But also in the scenarios with small volatility difference, participants significantly more often choose the risky fund as the one with higher fund manager skill. The higher return and alpha of this fund seems to dominate its higher volatility from the viewpoint of investors.

The direction of the effect holds in all individual markets in both surveys and is significant for all but one market (2010 of survey 2). It is also significant in 4 out of six scenarios in survey 2 (individual scenarios in survey 3 are not reported due to low number of observations). This confirms the general presence of an underestimation of the role of volatility in fund manager skill and portrays it as a highly robust phenomenon. There is a considerable fraction of participants who evaluate both funds as equally likely to be managed by a skillful manager. They give the theoretically correct response as the funds have equal p-values of alpha. However, it is likely that this fraction is overstated as any participant, who is uncertain about the correct solution will probably opt for

this category as well.

Within the individual markets, there are several interesting observations. Some markets have a low or even negative market risk premium (e.g., 2011 of survey 2 and markets A, D, and F of survey 3). In these cases, most of the difference in return is explained by higher (and less reliable) alpha. We can thus exclude the possibility that investors reward market exposure or market timing skills as it might be possible when the realized risk premium is high. We still find that participants to a greater proportion select the riskier fund as the one more likely to have a skillful manager, even though the difference is somewhat less pronounced. In other markets (e.g., 2012 of survey 2 and market B of survey 3), alphas are zero or relatively low. A complete absence of skill is also visible in p-values of 99% and 65%. Differences in return and volatility are mostly due to exposure to systematic risk. Nevertheless, the risky fund is hugely favored in these cases, which is perhaps the clearest indication for volatility neglect.

In panel B of table 8, we report responses to the question for superior risk-return relationship instead for the probability of skill. We find that on aggregate participants do not as clearly prefer one of the funds, proportions differ by only 5%-points (survey 2) and 8%-points (survey 3). In contrast to the question on probability of skill, the slight but significant majority ( $p = 0.09$  and  $p = 0.02$ ) believes that the safe fund offers the better risk-return profile. The differences are even a bit larger in the scenarios with high volatility differential between the two funds. The direction of the effect is in favor of the safe fund also in the majority of the individual markets and scenarios although often not significant due to the lower number of observations. But we make no claim in the regard that the safe fund is necessarily viewed as superior. For our hypothesis it is sufficient that in this control group participants are apparently taking volatility into account to reach their judgments.

This confirms that our survey design and the presentation of fund performance in price charts is in principle able to communicate fund returns and fund volatility. Participants in our surveys do not systematically misestimate the volatility or the returns due to features of the design. However, there are few participants reaching the conclusion that both funds are equal in terms of risk-return relationship. As one might expect, it is not possible to calculate Sharpe ratios exactly from the price charts alone and most likely other factors than pure Sharpe ratios will contribute to impressions of risk and return. This could explain why participants prefer one of the funds. In addition, the question might seem easier as the one for fund manager skill, which increases the confidence to make a choice between the funds.

We now take a closer look at the preference for either the risky or the safe fund in the two surveys. Table 9 reports differences between the proportions of participants choosing the risky and the safe fund for different treatments and scenarios (difference in differences). Our previous results revealed that participants tend to favor the risky fund in terms of probability of skill but are divided in their risk-return assessments. A direct comparison of the two groups confirms that the question mode has a significant influence on the fund preference. The risky vs. safe difference is in both surveys 21%-points larger for the probability of skill group than for the risk-return relationship group. It is significant not only for the overall sample but in every single market, which demonstrates the robustness of this result. We conclude that participants make a distinction between estimating probability of skill and evaluating the trade-off between risk and return. In particular, they seem to overweight return or alpha (or underweight volatility) in judgments of skill. They do not realize that high returns are a less reliable indicator of skill when these returns are highly volatile.

The other dimension by which we split the sample is large and small difference in volatility between the two funds. We report results separately for the two treatments. Results are in general less strong, if anything large differences in volatility slightly increase the preference for the risky fund in the probability of skill questions (significant in survey 2 but not in survey 3). In contrast, when asked for the risk-return relationship high volatility differences seem to favor the safe fund (not significant). This is in line with a focus on return in the probability of skill judgments as the difference in return between the funds is also larger in the scenarios with high difference in volatility.

## 5 Conclusion

Empirical evidence shows that investors over-invest into actively managed funds even if, on average, these funds underperform the market. Moreover, there is ample evidence that investors, once they have decided to invest into actively managed funds, buy the wrong funds. They chase returns and, by doing so, lose money on average.

We suggest that two factors contribute to these investment mistakes: First, investors fail to derive information from the cross section. They are unable to understand that in a large cross section of fund returns, there must be a few strong performers by pure chance and therefore, in a large fund population, a rare outperformer is less likely to be skilled than in a small fund population. Second, investors fail to incorporate the fund volatility when determining the probability of fund manager skill. Because of this, gambling fund managers will find it easier to attract investor money

than safer managers. While investors are able to notice different degrees of riskiness in time-series of returns, they are unable to draw conclusions about the uncertainty of fund manager skill. As ex-post past performance is certain, the role of volatility in the genesis of this performance is an elusive concept.

In a set of surveys with financially competent private investors, we show that participants entirely fail to take into account cross-sectional information but solely rely on the price development of the fund in question to judge the skill of its manager. They seem to believe that the number of competitors should not have an impact on the skill of an individual fund manager (except perhaps for motivational effects). What they miss is that they do not evaluate a random fund drawn from the population but the top performer. In a task that involves a combination of skill and luck they should ask themselves about the likelihood of a fund reaching this level of return by pure luck. By calculating theoretical values for the used fund samples, we demonstrate that this cross-sectional effect is huge even in small populations of between two to nine funds—and thus of first-order importance in real fund markets with hundreds of funds.

Further, we show that investors fail to sufficiently incorporate risk when inferring fund manager skill from past returns. They focus on return or alpha and underestimate the influence of volatility on the reliability of alpha. Participants apparently do not realize that it is more difficult to generate a high alpha value with a safe fund than with a risky fund. This result is in line with previous evidence by Huang et al. (2012), who theoretically show that investors have to take into account volatility when drawing conclusions about managerial ability and empirically find that naïve investors fail to do so. In our controlled experiment we can rule out that investors ignore volatility in general or that they are unable to infer it from price charts as they take it into account when asked for risk-return trade-off.

Our findings are also consistent with prior results from the psychological literature discussing the intentionality bias. According to this bias, individuals perceive an ambiguous outcome to be intended until proven otherwise (Rosset, 2008). Only with time and by realizing that outcomes are undesired, individuals learn to overwrite this bias. Rabin and Vayanos (2010) discuss the relationship between gambler’s fallacy and hot-hand fallacy and conclude that due to the expectation that luck will reverse overly quickly, investors will over-interpret streaks of above-average performance. People who interpret every result as intentional or excessively believe in continuation of streaks (hot-hand) are likely to be fooled by randomness.

All this feeds into Kahneman’s 2011 anecdotal observation of an *illusion of skill*, illustrated by

the tendency of asset managers to be convinced that strong returns were the result of their personal skill even when there is clear evidence that they were lucky. We show that this illusion of skill is also maintained on the side of investors and is unaffected by an increase in the likelihood that fund managers were lucky, e.g., because funds were more volatile or because the fund population was larger. As a consequence, investors underestimate the probability that a track record was generated by pure chance, especially in large fund populations and when fund managers take excessive risks.

These biases can lead to a misallocation of capital to unskilled managers and excessive risk taking of fund managers in order to attract new capital. Jordan and Riley (2014) show that volatile funds significantly underperform safer funds. Our results indicate that investors channel flows towards these underperforming funds. This could—in theory—lead to a race for riskiness in the fund industry as the riskiest of all funds will, luck permitting, be attributed the highest likelihood of skill by investors. Huddart (1999) shows that, with low barriers to entry, it is attractive for unskilled managers to take risk, hoping they will by chance generate a track record which falsely indicates skill. The fact that, on aggregate, significantly more subjects selected the risky fund in our surveys is evidence that this race to riskiness could indeed pay off.

Future research might show which interventions are successful to help investors understand randomness. A possibility would be to introduce a new performance measure: the probability of skill. While classic performance measures such as alpha or the Sharpe ratio are not always easy to interpret for retail investors, an indicator assigning funds to probability classes similar to the false discovery rates in BSW could give investors a clearer representation of luck and skill. However, a more general agreement on the definition of fund manager skill is likely needed for this suggestion to become reality. In addition, future research could attempt to gain a deeper understanding of the difficulties investors face when drawing conclusions about uncertain future returns from realized (and thus certain) past performance.

## References

- Barber, B. M., X. Huang, and T. Odean (2014). Which risk factors matter to investors? Evidence from mutual fund flows. Unpublished working paper.
- Barras, L., O. Scaillet, and R. Wermers (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance* 65(1), 179–216.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance* 52(1), pp. 57–82.
- Chang, T., D. H. Solomon, and M. M. Westerfield (2013). Looking for someone to blame: delegation, cognitive dissonance, and the disposition effect. Working paper.
- Choi, J. J., D. Laibson, and B. C. Madrian (2010). Why does the law of one price fail? An experiment on index mutual funds. *Review of Financial Studies* 23(4), 1405–1432.
- Cuthbertson, K., D. Nitzsche, and N. O’Sullivan (2008). Uk mutual fund performance: Skill or luck? *Journal of Empirical Finance* 15(4), 613–634.
- Deutsches Aktieninstitut e.V. (2013). *DAI-Factbook 2012* (Blatt 08.3 ed.). Frankfurt a. M.: Deutsches Aktieninstitut e.V.
- Ehm, C., C. Kaufmann, and M. Weber (2013). Volatility inadaptability: Investors care about risk, but cannot cope with volatility. *Review of Finance forthcoming*.
- Ert, E. and I. Erev (2007). Replicated alternatives and the role of confusion, chasing, and regret in decisions from experience. *Journal of Behavioral Decision Making* 20(3), 305–322.
- Frazzini, A. and O. A. Lamont (2008). Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics* 88(2), 299–322.
- Friesen, G. C. and T. R. Sapp (2007). Mutual fund flows and investor returns: An empirical examination of fund investor timing ability. *Journal of Banking & Finance* 31(9), 2796–2816.
- Gruber, M. J. (1996). Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance* 51(3), 783–810.
- Han, B. and D. Hirshleifer (2012). Self-enhancing transmission bias and active investing. Available at SSRN 2032697.
- Heimer, R. Z. and D. Simon (2012). Facebook finance: How social interaction propagates active investing. Working Paper, Brandeis University.
- Huang, J., K. Wei, and H. Yan (2012). Investor learning and mutual fund flows. AFA 2012 Chicago Meetings Paper.
- Huddart, S. (1999). Reputation and performance fee effects on portfolio choice by investment advisers. *Journal of Financial Markets* 2(3), 227–271.
- Investment Company Institute (2006). *Understanding Investor Preferences for Mutual Fund Information*. Investment Company Institute.

- Jensen, M. (1969). Risk, the pricing of capital assets, and the evaluation of investment portfolios. *The Journal of Business* 42(2), 167–247.
- Jordan, B. and T. Riley (2014). Volatility and mutual fund manager skill. Working Paper.
- Kahneman, D. (2011). Don't blink! The hazards of confidence. *The New York Times*, 10/19/2011.
- Keswani, A. and D. Stolin (2008). Which money is smart? Mutual fund buys and sells of individual and institutional investors. *The Journal of Finance* 63(1), 85–118.
- Koehler, J. J. and M. Mercer (2009). Selection neglect in mutual fund advertisements. *Management Science* 55(7), 1107–1121.
- Lynch, A. W. and D. K. Musto (2003). How investors interpret past fund returns. *The Journal of Finance* 58(5), 2033–2058.
- Morewedge, C. K. (2009). Negativity bias in attribution of external agency. *Journal of Experimental Psychology. General* 138(4), 535.
- Rabin, M. and D. Vayanos (2010). The gambler's and hot-hand fallacies: Theory and applications. *Review of Economic Studies* 77(2), 730–778.
- Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition* 108(3), 771–780.
- Sirri, E. R. and P. Tufano (1998). Costly search and mutual fund flows. *The Journal of Finance* 53(5), 1589–1622.
- Van Rooij, M., A. Lusardi, and R. Alessie (2011). Financial literacy and stock market participation. *Journal of Financial Economics* 101(2), 449–472.
- Zheng, L. (1999). Is money smart? A study of mutual fund investors' fund selection ability. *The Journal of Finance* 54(3), 901–933.
- Zion, U. B., I. Erev, E. Haruvy, and T. Shavit (2010). Adaptive behavior leads to under-diversification. *Journal of Economic Psychology* 31(6), 985–995.

## A Empirical separation of the US fund market into unskilled, skilled and zero-alpha funds according to BSW

We obtain daily fund return data from CRSP for the 1999-2012 period for all US equity funds (classified by CRSP objective code as domestic equity). Information from CRSP are on a share class level, while managers make investment decisions on a portfolio level. Therefore, we weight share class level returns by TNA to obtain portfolio level returns. Wherever TNA figures are not available, we use equal weighting of the share class level returns instead. Risk free rates and market returns are obtained from Kenneth French's website<sup>7</sup>. Based on these data, we calculate one factor Jensen alphas year by year and for the entire lifetime of the fund. We use alphas to determine the shares of unskilled, zero-skilled and skilled managers according to the method described in section 2. Results are reported in table A.1.

Table A.1: Share of skilled, zero-skilled and unskilled fund managers 1999 - 2012

Performance is measured by the Jensen one-factor and by the Carhart four-factor alphas. Years 1999 to 2012 are based on annual alphas. Lifetime is based on alphas calculated on all data available between 1999 and 2012. Shares of skilled, zero-skilled and unskilled managers are estimated according to the BSW methodology.

	Jensen Alphas (%)			Market	No.
	Skilled	Zero-Skld.	Unskld.	Ret (%)	
1999	19	47	34	26	2,216
2000	24	76	0	-12	2,551
2001	6	68	25	-11	2,873
2002	0	82	18	-21	3,064
2003	17	63	20	32	3,220
2004	7	72	21	12	3,323
2005	4	86	10	6	3,469
2006	1	63	36	15	3,689
2007	17	56	27	6	3,918
2008	0	88	12	-37	4,497
2009	11	87	2	28	5,272
2010	0	83	17	17	5,787
2011	0	89	11	0	4,825
2012	0	100	0	16	4,838
Mean	8	76	17		
Lifetime	1	92	7		10,530

In the one-factor Jensen model, the share of skilled managers is 0 in 2002, 2008 and 2010-2012. Around the peak of the dot-com bubble (1999 and 2000) an unusually large share of managers has

<sup>7</sup>[http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

been able to beat the market. Apparently, market inefficiencies were particularly extreme during that time. Another peak in the share of skilled managers can be observed in 2007, just prior to the outbreak of the financial crisis. In total, about three quarters of all managers have zero skill, 17% have negative skill and destroy value, while only 8% of all managers create value. In 2012, the year our first survey refers to, the BSW model concludes that all the cross sectional variation of fund returns could be explained by chance and there were no truly skilled managers.

If the entire lifetime of each fund between 1999 and 2012 is considered (the smaller of the 14 years in our sample and the lifetime of the fund), the share of skilled managers becomes even smaller with 1% and the vast majority (92%) of cross sectional differences in alpha are the result of chance.

## B Examples for price charts used in survey 1

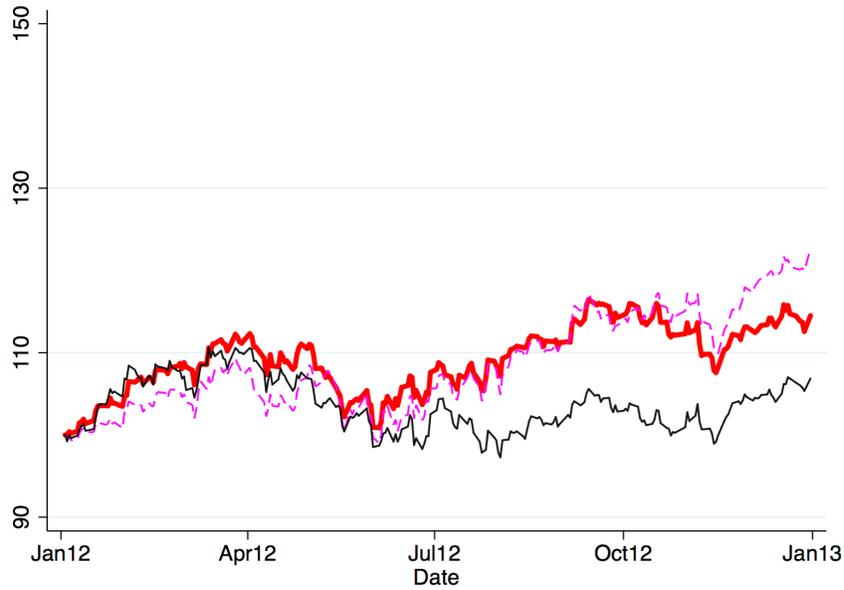


Figure B.1: Fund set C with 2 funds in the population. The top performer is shown in pink, the stock market in red, and the other available fund in grey.

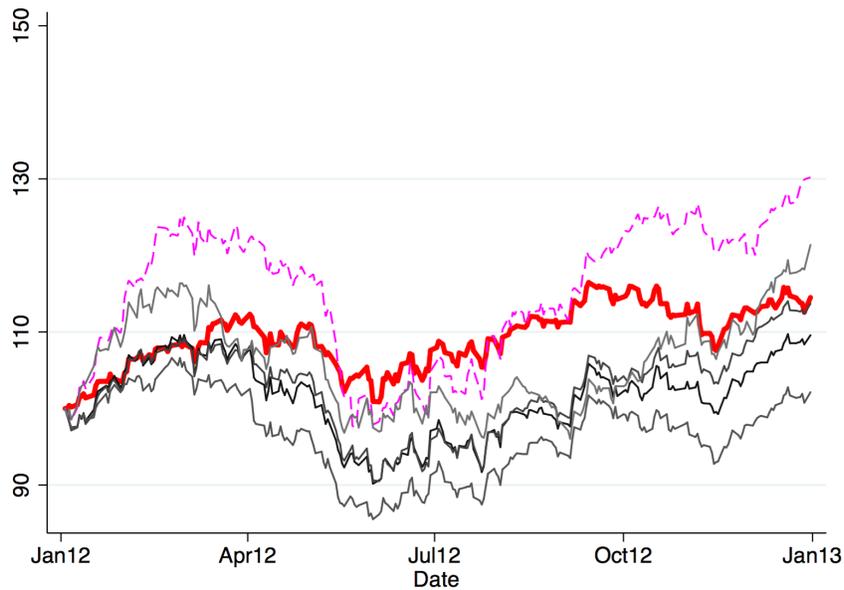


Figure B.2: Fund set B with 5 funds in the population. The top performer is shown in pink, the stock market in red, and the other available funds in grey.

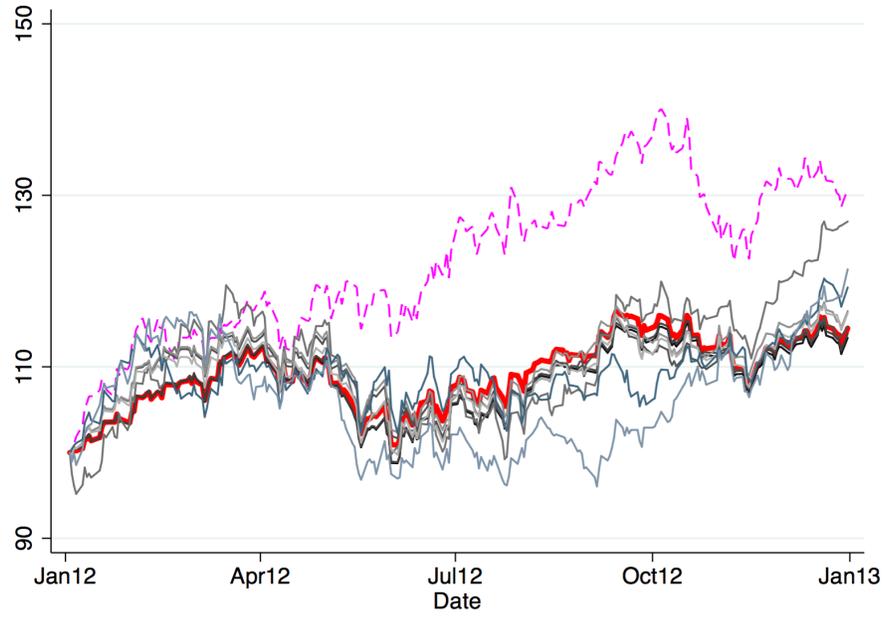


Figure B.3: Fund set A with 9 funds in the population. The top performer is shown in pink, the stock market in red, and the other available funds in grey.

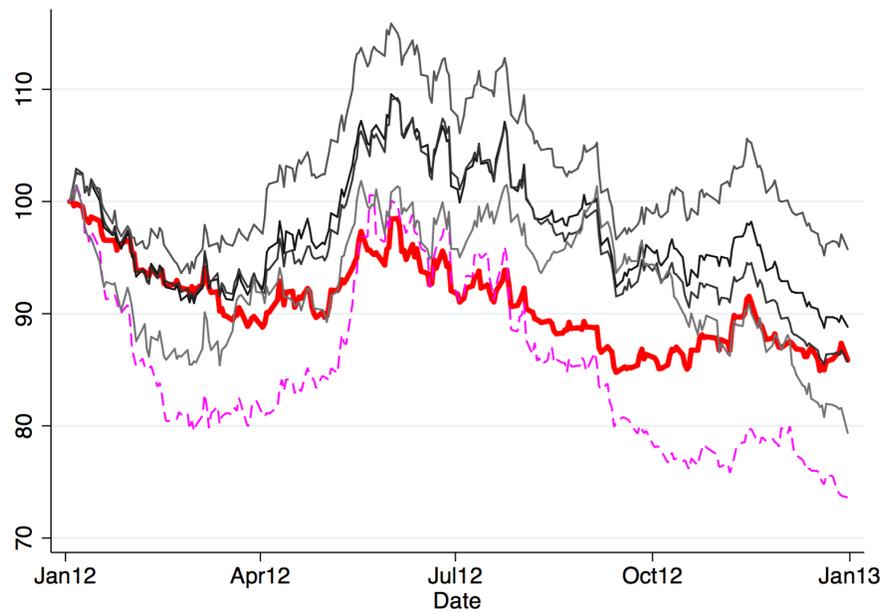


Figure B.4: Fund set B with 5 funds in the population in the *down* scenario. The bottom performer is shown in pink, the stock market in red, and the other available funds in grey.

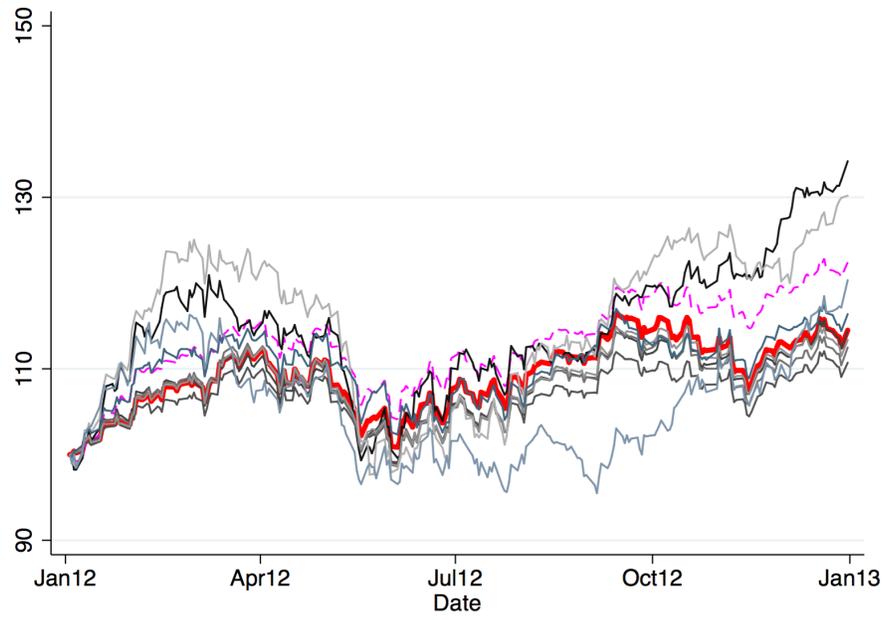


Figure B.5: Fund set D with 9 funds in the population in the *wide* scenario. The top performer in terms of significant alpha is shown in pink, the top performer in terms of absolute alpha in black, the stock market in red, and the other available funds in grey.

## C Examples for price charts used in survey 2

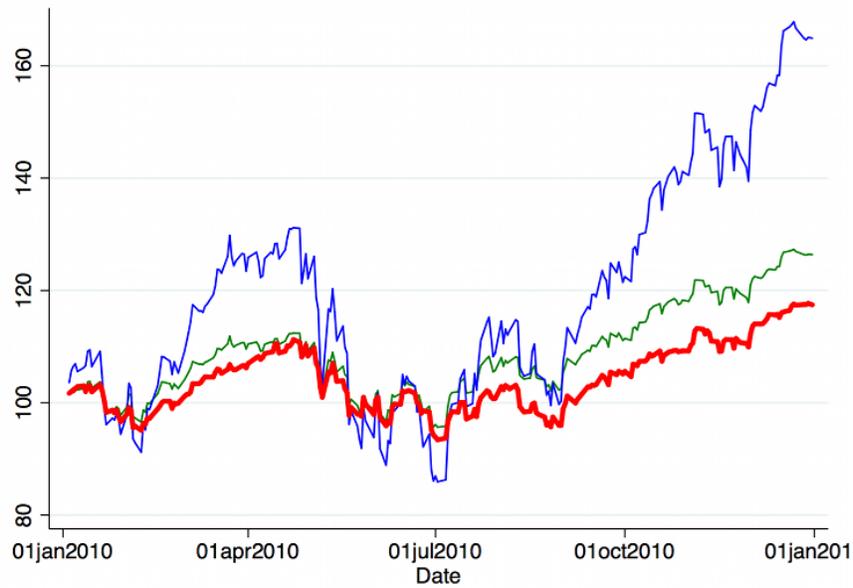


Figure C.1: Scenario 1 in year 2010 with large difference in volatility between presented funds.

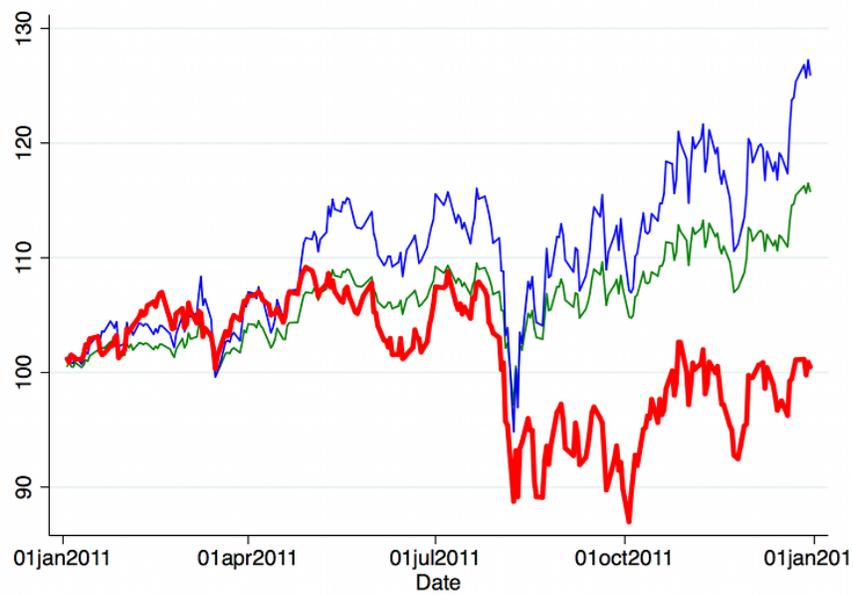


Figure C.2: Scenario 4 in year 2011 with small difference in volatility between presented funds.

## D Funds used in survey 3

Table D.1: Overview of funds used in survey 3

The table shows returns, volatility, alpha, significance of alpha, beta, and Sharpe ratio for all funds used in survey 3 and the market. *Days worse* is the number of days the price path of the risky fund is below the safe fund.

Mkt.	Scenario	Fund	Days worse	Ret.(%)	Std.dev.(%)	Alpha(%)	p-Alpha(%)	Beta	Sharpe Ratio
A	1+2	market		-3.1	11.2	0.0	—	1.00	-0.28
	1	risky	50	8.4	17.7	13.9	15.4	1.36	0.47
	1	safe		5.1	10.1	7.7	15.4	0.77	0.50
	2	risky	50	7.9	16.5	12.8	15.4	1.26	0.48
	2	safe		6.3	12.7	9.7	15.4	0.97	0.49
B	3+4	market		8.9	10.3	0.0	—	1.00	0.87
	3	risky	195	17.4	17.7	4.3	64.6	1.48	0.98
	3	safe		10.0	10.1	2.5	64.6	0.84	0.99
	4	risky	194	16.2	16.5	4.0	64.6	1.37	0.98
	4	safe		12.5	12.7	3.1	64.6	1.05	0.99
C	5+6	market		9.5	10.4	0.0	—	1.00	0.91
	5	risky	139	23.2	17.3	9.2	33.7	1.41	1.34
	5	safe		13.1	9.9	5.1	33.7	0.81	1.32
	6	risky	139	21.5	16.1	8.5	33.7	1.31	1.34
	6	safe		16.4	12.4	6.5	33.7	1.01	1.33
D	7+8	market		4.0	11.6	0.0	—	1.00	0.35
	7	risky	4	13.4	18.3	8.2	39.4	1.36	0.73
	7	safe		7.9	10.5	4.6	39.4	0.78	0.75
	8	risky	4	12.5	17.0	7.6	39.4	1.27	0.73
	8	safe		9.7	13.1	5.8	39.4	0.97	0.75
E	9+10	market		9.4	10.9	0.0	—	1.00	0.86
	9	risky	41	23.1	17.5	9.7	31.3	1.37	1.32
	9	safe		13.0	10.0	5.4	31.3	0.78	1.30
	10	risky	41	21.4	16.2	8.9	31.3	1.27	1.32
	10	safe		16.3	12.5	6.8	31.3	0.98	1.31
F	11+12	market		-2.6	11.2	0.0	—	1.00	-0.24
	11	risky	102	9.6	18.8	15.0	12.5	1.47	0.51
	11	safe		5.9	10.7	8.3	12.5	0.84	0.55
	12	risky	102	9.0	17.4	13.8	12.5	1.36	0.52
	12	safe		7.2	13.4	10.5	12.5	1.05	0.54
G	13+14	market		8.0	11.2	0.0	—	1.00	0.71
	13	risky	64	20.9	17.5	9.9	30.0	1.34	1.19
	13	safe		11.9	10.0	5.5	30.0	0.76	1.19
	14	risky	64	19.4	16.3	9.2	30.0	1.24	1.19
	14	safe		14.9	12.5	7.0	30.0	0.95	1.19
H	15+16	market		9.9	11.3	0.0	—	1.00	0.88
	15	risky	114	20.6	17.5	7.0	45.6	1.33	1.17
	15	safe		11.7	10.0	4.0	45.6	0.76	1.17
	16	risky	112	19.1	16.3	6.5	45.6	1.24	1.17
	16	safe		14.7	12.5	5.0	45.6	0.95	1.17

Table 1: Overview of funds used in survey 1

List of all funds used in survey 1. Panels A to D show the four different sets of funds used. Displayed fund attributes include market beta, absolute annual alpha in %, the t-value and p-value of alpha, the annual return and the annual standard deviation of returns. Funds presented in the scenarios of varying population size are marked with an x.

<b>Panel A: High Annual Return Configuration</b>									
Fund	Beta	Alpha(%)	t-Alpha	p-Alpha(%)	Ret.(%)	Std.dev.(%)	9 Funds	5 Funds	2 Funds
1	1.05	13.45	1.03	30.16	30.76	18.43	x	x	x
2	1.02	12.81	0.78	43.59	26.93	20.38	x		
3	1.07	7.14	0.51	61.04	21.37	19.46	x	x	
4	1.08	3.17	0.25	80.04	19.25	18.78	x		
5	1.07	-0.07	-0.01	98.88	16.5	14.97	x	x	
6	1.09	-0.82	-0.2	83.76	16.27	14.92	x		
7	1.07	-2.68	-0.53	59.58	14.69	14.91	x	x	
8	1.07	-1.18	-0.74	45.52	13.86	14.17	x		
9	1.07	-1.6	-1.01	31.15	13.37	14.16	x	x	x
<b>Panel B: High Second Half Year Return Configuration</b>									
Fund	Beta	Alpha(%)	t-Alpha	p-Alpha(%)	Ret.(%)	Std.dev.(%)	9 Funds	5 Funds	2 Funds
1	1.06	17.32	1.01	31.19	30.19	20.96	x	x	x
2	0.96	7.7	0.74	45.95	20.47	16.05	x		
3	1.07	7.14	0.51	61.04	21.37	19.46	x	x	
4	1.04	4.32	0.29	76.71	17.38	19.76	x		
5	1.04	-0.16	-0.01	98.47	13.68	16.22	x	x	
6	1.02	-1.79	-0.21	82.9	11.25	15.84	x		
7	1.09	-4.01	-0.47	63.3	9.57	16.71	x	x	
8	1.01	-6.86	-0.78	43.51	5.43	16.09	x		
9	0.94	-8.46	-0.97	32.82	2.12	15.3	x	x	x
<b>Panel C: High Last Quarter Return Configuration</b>									
Fund	Beta	Alpha(%)	t-Alpha	p-Alpha(%)	Ret.(%)	Std.dev.(%)	9 Funds	5 Funds	2 Funds
1	1.07	7	0.97	32.99	22.61	15.7	x	x	x
2	0.91	8.87	0.76	44.38	21.36	16.26	x		
3	1.07	7.14	0.51	61.04	21.37	19.46	x	x	
4	1.04	4.32	0.29	76.71	17.38	19.76	x		
5	0.94	0.09	0	99.44	11.52	18.05	x	x	
6	1.08	-2.28	-0.25	79.98	11.12	16.81	x		
7	1.09	-4.01	-0.47	63.3	9.57	16.71	x	x	
8	1.01	-6.86	-0.78	43.51	5.43	16.09	x		
9	0.96	-6.35	-0.99	31.93	6.85	14.3	x	x	x
<b>Panel D: Wide t-Statistic Intervals Configuration</b>									
Fund	Beta	Alpha(%)	t-Alpha	p-Alpha(%)	Ret.(%)	Std.dev.(%)	9 Funds	5 Funds	2 Funds
1	0.92	8.3	1.96	5.11	22.42	12.75	x		
2	0.99	21.22	1.48	13.84	34.21	18.37	x		
3	1.06	17.32	1.01	31.19	30.19	20.96	x		
4	1.06	6.32	0.45	65.03	20.33	19.39	x		
5	1.07	-0.18	-0.03	97.06	16.37	14.96	x		
6	1.06	-0.96	-0.47	63.33	14.39	14.06	x		
7	1.07	-1.63	-1.02	30.43	13.36	14.14	x		
8	1.04	-2.13	-1.5	13.26	12.49	13.81	x		
9	0.96	-2.71	-2	4.59	10.71	12.73	x		

Table 2: Summary statistics of all survey participants

The table shows mean, median, standard deviation (SD), minimum, maximum, and number of observations for the answers to our questions on socio-demographics, market views and financial literacy. *Male* is a dummy variable that is 0 for female and 1 for male participants. *Education* indicates the level of the highest education with 1 indicating “no school graduation”, 2 German “Hauptschule”, 3 German “Realschule”, 4 non-academic apprenticeship, 5 German “Abitur”, 6 a Bachelor’s degree, 7 a Master’s degree or a Diploma, and 8 a PhD. *Income* is 1 for less than EUR 2000 monthly net income, 2 for income between EUR 2000 and EUR 6000 and 3 for income higher than EUR 6000. *Financial professional*, *Invested in fund*, *Ever invested in stock* are dummy variables with 0 for no and 1 for yes. Statistics knowledge is self attributed with German school grades form 1 (very good) to 5 (poor). *Fin. literacy score* is the number of six financial literacy questions answered correctly. For *Identifying skill* subjects were asked whether they can identify skilled fund managers with answer choices “no”, “rather no”, “neutral”, “rather yes”, “yes”, represented in the table by numbers from 1 to 5. Skilled, zero-skilled, and unskilled managers are the estimated shares of each respective manager type in the fund market. Definitions were provided as stated in section 3.1.

	Survey 1					Survey 2					Survey 3							
	Mean	Med.	SD	Min	Max	N	Mean	Med.	SD	Min	Max	N	Mean	Med.	SD	Min	Max	N
Age	48.70	49	13.39	18	78	228	45.65	45	16.08	10	89	893	55.45	57	14.45	22	82	183
Gender	0.86	1	0.35	0	1	228	0.87	1	0.33	0	1	893	0.89	1	0.32	0	1	183
Education	6.47	7	1.39	2	8	228	5.81	6	1.70	1	8	893	6.64	7	1.37	2	8	183
Income	2.06	2	0.59	1	3	228	1.87	2	0.64	1	3	893	2.11	2	0.52	1	3	183
Financial professional	0.16	0	0.37	0	1	228	0.24	0	0.42	0	1	893	0.21	0	0.41	0	1	183
Invested in fund	0.70	1	0.46	0	1	228	0.71	1	0.45	0	1	893	0.75	1	0.43	0	1	183
Ever invested in fund	0.88	1	0.32	0	1	228	0.82	1	0.38	0	1	893	0.91	1	0.29	0	1	183
Ever invested in stock	0.86	1	0.34	0	1	228	0.79	1	0.41	0	1	893	0.90	1	0.31	0	1	183
Statistics knowledge (1-5)	2.86	3	0.92	1	5	228	2.89	3	0.87	1	5	893	2.77	3	0.89	1	5	183
Fin. literacy score (1-6)	5.63	6	0.98	0	6	234	5.57	6	1.01	0	6	923	5.73	6	0.69	0	6	185
Identifying skill (1-5)	2.46	2	1.14	1	5	228	2.67	3	1.19	1	5	905	2.27	2	1.00	1	4	183
Skilled managers (%)	18.15	15	14.27	0	90	227	18.31	10	16.64	0	100	868	17.93	15	13.66	0	70	179
Zero-skilled managers (%)	34.00	32	16.82	0	85	226	31.87	30	18.14	0	100	852	30.70	30	19.00	0	80	178
Unskilled managers (%)	47.56	45	22.95	2	100	226	51.42	50	23.10	0	100	858	52.03	50	23.54	10	100	181

Table 3: Fund market views and skill estimates explained by personal characteristics

The table shows the result of OLS regression analyses of participant's market views explained by personal characteristics. In column (1) the dependent variable is the estimated percentage of skilled managers in the general fund market. In column (2) the dependent variable is the estimated percentage of unskilled managers. In column (3) the dependent variable is the self-assessed ability to identify a skilled manager on a scale from 1 ("no") to 5 ("yes"). In columns (4) to (6) dependent variable is the probability estimate for skill in the eight scenarios. Independent variables are personal characteristics of participants as explained in table 2 (the variable statistical knowledge is reversed). Displayed are coefficients and their t-values in parentheses (using robust standard errors in regressions (1)-(3), standard errors clustered by scenario in regressions (4)-(6)). Coefficients are significant at  $*p < 0.10$ ,  $**p < 0.05$ ,  $***p < 0.01$ .

	% Skilled	% Unskilled	Identif. skill	Prob. est. of fund manager skill		
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.21** (2.43)	-0.16 (-1.23)	0.01 (1.14)	0.32*** (2.63)	0.24** (2.07)	0.25** (2.27)
Male	-5.58* (-1.81)	11.10*** (2.76)	0.04 (0.26)	-1.78 (-0.49)	-0.38 (-0.12)	1.72 (0.53)
Education	-2.03** (-2.43)	2.43* (1.75)	0.07 (1.25)	-0.45 (-0.41)	-0.25 (-0.23)	0.05 (0.05)
Income	-1.52 (-0.85)	2.49 (0.69)	-0.28* (-1.78)	-2.54 (-0.87)	-0.96 (-0.36)	-0.74 (-0.31)
Financial professional	-1.89 (-0.69)	6.98 (1.35)	0.37 (1.42)	-2.66 (-0.57)	-4.12 (-0.93)	-2.13 (-0.50)
Invested in fund	-1.17 (-0.42)	5.82 (1.45)	0.22 (1.09)	0.46 (0.13)	-0.37 (-0.11)	1.29 (0.39)
Ever invested (fund)	-0.07 (-0.02)	2.08 (0.33)	0.18 (0.66)	-1.04 (-0.19)	-2.62 (-0.48)	-1.86 (-0.34)
Ever invested (stock)	-1.00 (-0.38)	-1.54 (-0.37)	-0.21 (-0.95)	-5.43 (-1.40)	-3.96 (-1.11)	-4.82 (-1.40)
Statistics knowledge	1.16 (1.13)	-0.59 (-0.31)	0.11 (1.25)	-1.86 (-1.13)	-2.85* (-1.76)	-2.66* (-1.71)
Fin. literacy score	-1.62 (-0.95)	0.37 (0.13)	0.03 (0.23)	-2.90 (-1.21)	-2.83 (-1.23)	-3.10 (-1.34)
Skilled (%)					0.25*** (2.78)	
Unskilled (%)						-0.31*** (-5.09)
Identifying skill					5.03*** (3.82)	4.37*** (3.35)
Constant	37.06*** (3.54)	18.46 (1.13)	1.47* (1.70)	76.29*** (5.10)	61.11*** (3.76)	77.02*** (5.10)
$R^2$	0.11	0.11	0.06	0.04	0.09	0.13
N	206	206	206	1855	1847	1847

Table 4: Estimated probabilities of fund manager skill by scenario

*Fund set* and *Population size* identify the scenario from table 1. The BSW estimates for the share of zero-skilled managers ( $\pi_0$ ) and the probability that the manager of the top performing fund is skill ( $p(skilled)$ ) are calculated according to the BSW methodology. Alternative probability estimates are based on the assumption that the market share of zero-skilled funds is (1) as in the real US fund market in 2012 (*Market*) or (2) as expressed by the participants (*Beliefs*). *Survey estimates* show the average probability estimates for each scenario that the top-performing fund is managed by a skilled manager, and that the worst performer in the *down* scenario is managed by an unskilled manager. All probabilities are expressed in %.

Fund set	Population size	BSW estimates		Market	Beliefs	Survey estimates	
		$\pi_0$	$p(skilled)$	$p(skilled)$	$p(skilled)$	$p(skilled)$	$p(unskilled)$
A	2	72.5	78.9	72.1	89.9	66.5	
A	5	92.3	47.0	44.2	76.3	67.6	70.3
A	9	93.6	25.2	23.0	62.1	65.3	
B	2	75.0	77.5	71.2	88.9	55.4	
B	5	92.3	45.7	42.8	75.7	53.0	62.4
B	9	93.6	24.0	21.7	62.4	54.3	
C	2	70.0	77.7	69.7	88.8	36.8	
C	5	92.3	43.5	40.6	74.7	39.7	51.1
C	9	93.6	21.9	19.7	59.2	37.7	
D (fund 1)	9	51.3	88.7	79.2	92.5	50.9	
D (fund 2)	9	51.3	71.8	52.5	80.3	57.4	

Table 5: Differences between the skill estimates for different scenarios

In panel A, skill estimates in percent are reported for different population sizes, both aggregated and by fund set. Panel B contains values for each fund set, aggregated and by population size. Values between the panels differ as only those investors are compared, who see both scenarios.  $\Delta$  is the difference between two scenarios in each column. T-values of a standard t-test are reported, \* indicates significance at 10% level, \*\* at 5% level and \*\*\* at 1% level.

Panel A: Difference in estimates between populations sizes												
	Overall			A			By fund set			C		
							B					
2 Funds	52.9	52.9		67.6	66.1		53.9	57.5		39.2	34.4	
5 Funds	53.2		53.2	66.9		69.1	53.7		52.2	39.1		40.3
9 Funds		52.5	52.5		62.7	68.4		57.7	51.1		32.8	42.6
$\Delta$	-0.3	0.4	0.7	0.6	3.4*	0.7	0.2	-0.2	1.1	0.1	1.6	-2.3
t-statistic	-0.22	0.42	0.69	0.29	1.69	0.53	0.08	-0.11	0.67	0.03	0.66	-1.17
N	230	230	230	73	80	75	77	74	78	79	75	76
Panel B: Difference in estimates between fund sets												
	Overall			2 Funds			By population size			9 Funds		
							5 Funds					
Fund set A	66.4	66.4		68.1	64.9		66.5	69.6		64.3	66.3	
Fund set B	54.1		54.1	53.9		57.0	54.0		51.9	55.0		53.5
Fund set C		38.1	38.1		34.9	38.8		43.1	37.2		36.0	40.0
$\Delta$	12.3***	28.3***	16.1***	14.3***	30.0***	18.2***	12.4***	26.5***	14.8***	9.3***	30.3***	13.5***
t-statistic	11.23	14.51	10.76	6.26	7.70	5.65	5.33	7.38	5.55	5.47	8.16	4.26
N	230	230	230	76	78	76	75	73	80	79	77	73

Table 6: Estimates of probability of skill explained by chart type

The table shows OLS and Tobit regressions to explain the probability that a fund is managed by a skilled manager as estimated by survey participants ( $p(skilled)$ ). Independent variables are fund annual return, annual alpha, and annual standard deviation of returns (volatility). Additionally, intra-year return patterns and population size are included (2 fund scenarios are the baseline, dummy variables for 5 and 9 fund scenarios are included). Regressions include participant fixed effects and standard errors are clustered by scenarios. T-statistics are reported in parentheses, \* indicates significance at 10% level, \*\* at 5% level and \*\*\* at 1% level.

Survey estimates of $p(skilled)$	OLS					Tobit	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ann. Return	1.85*** (3.35)		2.75** (2.71)			3.08*** (2.94)	
Alpha		1.26** (2.95)					1.60* (1.75)
Return H1				1.87*** (5.04)			
Return H2				1.10** (2.49)			
Return Q1-Q3					2.20*** (13.23)		
Return Q4					1.18*** (5.83)		
Volatility			-1.88 (-1.49)	0.28 (0.35)	-1.08** (-2.69)	-2.17* (-1.66)	-0.46 (-0.28)
5 Funds dummy			0.43 (0.10)	0.53 (0.25)	0.54 (0.40)	0.54 (0.12)	0.42 (0.05)
9 Funds dummy			-2.85 (-0.63)	1.01 (0.42)	0.70 (0.43)	-3.44 (-0.73)	-2.05 (-0.24)
Constant	-37.67** (-2.54)	-2.56 (-0.44)	-27.58* (-2.06)	-32.35*** (-3.29)	-23.69*** (-3.70)	3.74 (0.20)	37.50 (1.47)
Participant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$R^2$	0.65	0.62	0.66	0.69	0.69	—	—
N	1828	1828	1828	1828	1828	1828	1828

Table 7: Overview of funds used in survey 2

The table shows returns, volatility, alpha, significance of alpha, beta, and Sharpe ratio for all funds used in survey 2 and for the US stock market in each year.

Year	Scenario	Fund	Ret.(%)	Std.dev.(%)	Alpha(%)	p-Alpha(%)	Beta	Sharpe Ratio
2010	1+2	market	17.4	18.5	0.0	—	1.00	0.93
	1	risky	64.9	48.1	20.2	28.9	2.45	1.34
	1	safe	26.4	19.8	7.8	29.6	1.01	1.32
	2	risky	57.4	42.5	17.6	29.0	2.16	1.35
	2	safe	34.2	25.5	10.1	29.3	1.29	1.33
2011	3+4	market	0.5	24.2	0.0	—	1.00	0.02
	3	risky	29.2	29.5	30.1	9.5	1.06	0.99
	3	safe	12.3	12.2	11.5	9.5	0.44	1.01
	4	risky	25.9	26.0	26.2	9.5	0.94	1.00
	4	safe	15.8	15.6	15.0	9.5	0.56	1.01
2012	5+6	market	16.2	13.1	0.0	—	1.00	1.24
	5	risky	39.8	36.1	0.1	99.3	2.51	1.10
	5	safe	16.6	14.9	0.1	99.4	1.03	1.12
	6	risky	35.3	31.8	0.1	99.3	2.21	1.11
	6	safe	21.3	19.1	0.1	99.4	1.33	1.12

Table 8: Fund choice by year, volatility scale and question type in all questions with equally skilled managers

Panel A presents for survey 2 and 3 the responses to the probability of skill question (treatment group), disaggregated by market and volatility differential. It provides the fractions of participants who believe the risky fund has a higher probability of a skilled fund manager, the safe fund has a higher probability, and of those who think both funds have equal probability. Significance of differences is tested by a two-tailed proportion test. Panel B analogously shows results for the risk-return relationship question (control group).

<b>Panel A: Probability of skill group</b>												
Survey 2	By market			Low $\Delta$ Vola			High $\Delta$ Vola			Aggregated		
	2010	2011	2012	2010	2011	2012	2010	2011	2012	All	Low $\Delta$	High $\Delta$
Risky fund	0.40	0.39	0.59	0.34	0.40	0.48	0.46	0.39	0.68	0.46	0.40	0.51
Safe fund	0.36	0.32	0.23	0.35	0.27	0.27	0.36	0.38	0.19	0.30	0.30	0.31
Equal	0.24	0.28	0.18	0.31	0.34	0.24	0.18	0.23	0.13	0.24	0.30	0.18
Risky – Safe	0.04	0.07	0.36	-0.01	0.13	0.21	0.10	0.01	0.49	0.16	0.10	0.20
P-value	0.28	0.09	0.00	0.80	0.02	0.00	0.09	0.82	0.00	0.00	0.00	0.00
<i>N</i>	414	418	414	201	200	190	213	218	224	1,246	591	655
<b>Panel B: Risk-return relationship group</b>												
Survey 3	By market			By market			By market			Aggregated		
	A	B	C	D	E	F	G	H	All	Low $\Delta$	High $\Delta$	
Risky fund	0.40	0.42	0.45	0.42	0.48	0.37	0.43	0.41	0.42	0.36	0.49	
Safe fund	0.24	0.18	0.20	0.20	0.15	0.18	0.15	0.18	0.18	0.14	0.23	
Equal	0.36	0.40	0.35	0.38	0.37	0.45	0.41	0.41	0.39	0.50	0.29	
Risky – Safe	0.16	0.24	0.25	0.22	0.33	0.19	0.28	0.23	0.24	0.22	0.26	
P-value	0.05	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	
<i>N</i>	98	97	98	97	98	97	97	97	779	390	389	
Survey 2	By market			Low $\Delta$ Vola			High $\Delta$ Vola			Aggregated		
	2010	2011	2012	2010	2011	2012	2010	2011	2012	All	Low $\Delta$	High $\Delta$
Risky fund	0.42	0.38	0.46	0.41	0.40	0.40	0.43	0.36	0.54	0.42	0.40	0.44
Safe fund	0.48	0.51	0.42	0.42	0.45	0.45	0.54	0.58	0.37	0.47	0.44	0.50
Equal	0.10	0.11	0.12	0.17	0.15	0.14	0.03	0.06	0.09	0.11	0.16	0.06
Risky – Safe	-0.06	-0.13	0.04	-0.01	-0.05	-0.05	-0.11	-0.22	0.17	-0.05	-0.04	-0.06
P-value	0.21	0.01	0.33	0.81	0.49	0.41	0.14	0.00	0.02	0.09	0.31	0.16
<i>N</i>	380	386	386	192	199	209	188	187	177	1,152	600	552
Survey 3	By market			By market			By market			Aggregated		
	A	B	C	D	E	F	G	H	All	Low $\Delta$	High $\Delta$	
Risky fund	0.27	0.38	0.41	0.38	0.40	0.30	0.43	0.39	0.37	0.36	0.38	
Safe fund	0.55	0.48	0.41	0.45	0.43	0.48	0.39	0.39	0.45	0.39	0.50	
Equal	0.18	0.14	0.18	0.16	0.17	0.22	0.18	0.22	0.18	0.24	0.12	
Risky – Safe	-0.28	-0.10	0.00	-0.07	-0.03	-0.18	0.04	0.00	-0.08	-0.03	-0.12	
P-value	0.00	0.30	1.00	0.49	0.82	0.05	0.73	1.00	0.02	0.51	0.01	
<i>N</i>	89	90	90	91	89	92	89	90	720	361	359	

Table 9: Difference in differences in fund selection

The table shows the difference in the difference between selection of the risky and the safe fund for various subsamples (diff-in-diff). The first row compares results for the probability of fund manager skill question (treatment group) to results for the risk-return relationship question (control group). The other comparisons are between the large and small volatility differential scenarios, separately for treatment and control group. Significance of the difference in differences is tested by a Wilcoxon ranksum test.

Risky – Safe	Survey 2				Survey 3								
	2010	2011	2012	All	A	B	C	D	E	F	G	H	All
<i>Probability of skill (treatment group) – return-risk relationship (control group)</i>													
Diff in diff	0.11	0.20	0.32	0.21	0.43	0.35	0.24	0.29	0.35	0.37	0.24	0.24	0.21
P-value	0.10	0.00	0.00	0.00	0.00	0.01	0.07	0.03	0.01	0.00	0.08	0.07	0.00
<i>Large volatility difference scenarios – small volatility difference scenarios (treatment group only)</i>													
Diff in diff	0.11	-0.12	0.28	0.10	0.11	0.08	0.10	0.00	-0.07	-0.15	0.03	0.26	0.05
P-value	0.15	0.18	0.00	0.02	0.41	0.47	0.32	0.87	0.91	0.38	0.56	0.06	0.13
<i>Large volatility difference scenarios – small volatility difference scenarios (control group only)</i>													
Diff in diff	-0.10	-0.17	0.22	-0.02	0.07	-0.12	-0.09	0.05	-0.27	-0.25	-0.12	0.00	-0.09
P-value	0.31	0.05	0.02	0.67	0.78	0.51	0.64	0.79	0.16	0.14	0.57	1.00	0.14

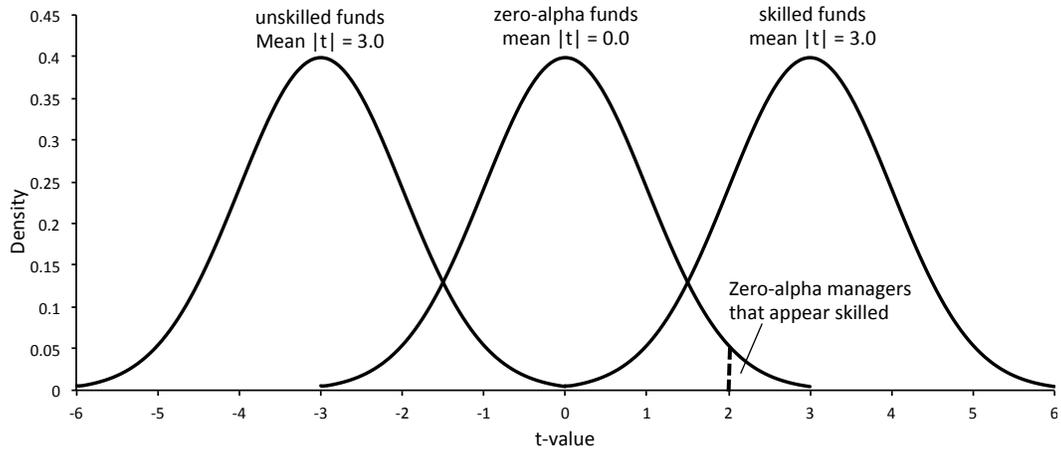


Figure 1: Unskilled, zero-skilled and skilled fund managers and their distribution of t-values according to BSW

Exemplary distribution of t-values of the three skill groups according to BSW. t-values of zero-alpha funds are distributed around zero. For demonstration purposes, we assume t-values of unskilled funds are distributed around -3 and t-values of skilled funds are distributed around 3. Zero-alpha funds that appear skilled are the proportion of zero-alpha funds that have positive and significant alpha by chance. The figure is adapted from Barras et al. (2010).

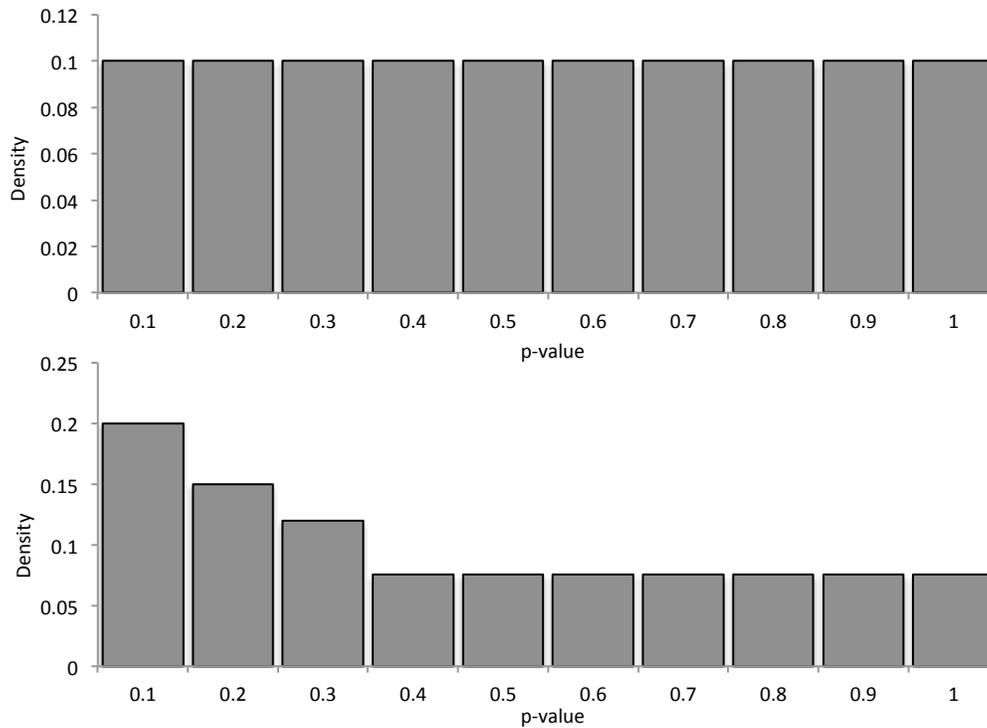


Figure 2: Distribution of p-values in a population with 100% zero-alpha funds and in a population with zero-alpha, skilled and unskilled funds

The figure displays a hypothetical distribution of p-values of  $H_0 : \alpha_i = 0$  of a population of zero-alpha funds (above) and of a population that includes funds with true negative and true positive alphas. The figure is adapted from Barras et al. (2010).

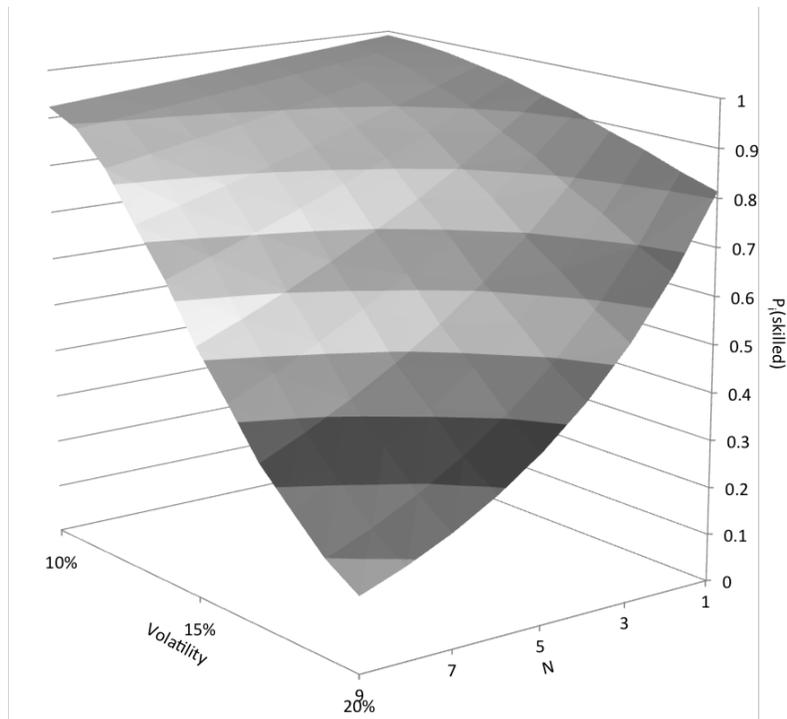


Figure 3: Probability of skill mapped against the number of zero-skilled funds and the volatility of the top performer

The figure shows the probability to be skilled for the manager of a hypothetical top-performing fund depending on population size and volatility. Fund populations vary between 1-9 funds, and volatility of fund returns of is scaled such that the actual annual returns stay constant. One factor alphas are calculated for all funds.  $p_i(skilled)$  is calculated according to equation 6.  $p_i(skilled)$  is on the y-axis, the number of zero-skilled funds on the x-axis and the annualized volatility of the daily returns on the z-axis.

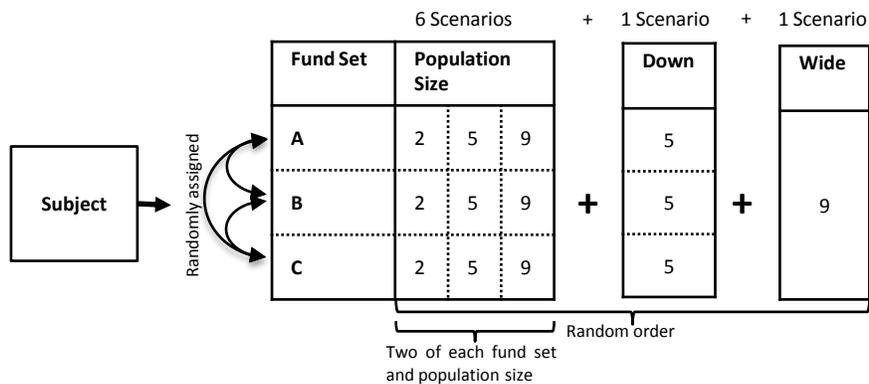


Figure 4: Survey design and fund scenarios in survey 1

Each participant sees price charts of eight scenarios in random order. From the base scenarios six are randomly selected including two of each fund set (A, B, or C) and two of each population size (2, 5, or 9 funds). Additionally, a down scenario is shown (negative mirror image of fund set A, B, or C) and the wide scenario (fund set D).