Just Unlucky? –
A Bootstrapping Simulation to Measure Skill in Individual Investors’ Investment Performance*  

Steffen Meyer¹, Dennis Schmoltzi, Christian Stammschulte, Simon Kaesler, Benjamin Loos, Andreas Hackethal  
Goethe University Frankfurt

ABSTRACT

Few studies have focused on the measurement of individual investors’ investment performance and more extensive research has been conducted on biases and investment mistakes – such as the disposition effect, security selection bias and lacking ability of market timing. No study so far has focused on measuring whether the realized performance was driven by skill or mere luck. This paper disentangles skill and luck in individual investors’ investment performance using a four-factor model and apply bootstrapping simulations pioneered in the mutual fund literature to distinguish skill from luck. We use a comprehensive dataset of 8,621 individual investor portfolios from a German online broker, spanning a timeframe from September 2005 to April 2010. We find that 89\% of individual investors exhibit negative skill (\(\alpha \leq 0\)) when measured on a gross basis and 91\% when considering returns net of costs and expenses. An individual investor with an average level of risk-taking depicts an investment performance of approximately -7.5\% per year for gross returns and of -8.5\% per year for net returns.

Keywords: Individual Investors, Retail Investors, Portfolio Performance, Investment Policy, Bootstrapping

JEL-Classification: D14, G11

* We would like to thank the anonymous bank for providing us with data and the employees and administrators at the Frankfurt Cloud to provide us with highly secured access and the necessary computing power to run the simulations.

¹ Corresponding author: Grüneburgplatz 1, 60323 Frankfurt am Main, Germany, Tel.: +49 (69) 798 33675; Fax: +49 (69) 798 33530. E-mail address: meyer@finance.uni-frankfurt.de
Just Unlucky? –
A Bootstrapping Simulation to Measure Skill in Individual Investors’ Investment Performance†

ABSTRACT

Few studies have focused on the measurement of individual investors’ investment performance and more extensive research has been conducted on biases and investment mistakes – such as the disposition effect, security selection bias and lacking ability of market timing. No study so far has focused on measuring whether the realized performance was driven by skill or mere luck. This paper disentangles skill and luck in individual investors’ investment performance using a four-factor model and apply bootstrapping simulations pioneered in the mutual fund literature to distinguish skill from luck. We use a comprehensive dataset of 8,621 individual investor portfolios from a German online broker, spanning a timeframe from September 2005 to April 2010. We find that 89% of individual investors exhibit negative skill (α ≤ 0) when measured on a gross basis and 91% when considering returns net of costs and expenses. An individual investor with an average level of risk-taking depicts an investment performance of approximately -7.5% per year for gross returns and of -8.5% per year for net returns.

Keywords: Individual Investors, Retail Investors, Portfolio Performance, Investment Policy, Bootstrapping

JEL-Classification: D14, G11

† We would like to thank the anonymous bank for providing us with data and the employees and administrators at the Frankfurt Cloud to provide us with highly secured access and the necessary computing power to run the simulations.
I. Introduction

As per 2010, about 52% of U.S. households’ and 24% of German households’ total financial assets were invested in securities.\(^1\) These figures suggest a worldwide investment volume of private households in the high double-digit trillions. Despite representing an important investor class in our worldwide economies, very few studies have to date comprehensively measured individual investors’ investment performance.

In the literature it is undisputed that private investors are prone to investment mistakes, like the disposition effect (Shefrin and Statman (1985); Odean (1998)), security selection bias (Barber and Odean (2000), Barber et al. (2009)) and a lack of market timing ability of (Odean (1999) and Barber et al. (2009)), literature which comprehensively analyses the performance of private investors is scarce and the results point in different directions.

The first to look at the performance of private investors were Schlarbaum et al. (1978) who do not find individual investors to perform significantly different from the market. Moreover, Barber and Odean (2000) analyze the benchmark performance of a cross-section of individual investors and find neither significant out- nor underperformance before trading costs and expenses. However, taking costs associated with trading into account alters the results. The more investors trade, the worse the performance gets. Hence, after costs an underperformance is discernible. Using data from Sweden and applying a methodology which looks at the ex-ante performance of investors in Sweden, Calvet et al. (2007) find that Swedish households on aggregate do not underperform considerably. In contrast, Barber et al. (2009) compare returns between 1995 and 1999 of private and institutional investors in Taiwan and find that private investors underperform.

Thoughts governed by the efficient market hypothesis and insights of the literature which repeatedly document a strikingly low level of financial literacy make it tempting to believe that individual investors represent a random sample of the overall market and therefore do not

\(^1\) Excluding pension funds and life insurances. Cf. Federal Reserve Board: Flow of Funds Accounts of the United States, Deutsche Bundesbank: Ergebnisse der gesamtwirtschaftlichen Finanzierungsrechnung für Deutschland.
underperform the market as a group. Therefore, any over- or underperformance found in the literature would be due to luck of private investors in the sample period and by substantial skill differences. On the other hand, the extensive literature documenting the behavioral biases and investment mistakes suggests that investors may in fact be unskilled and hence systematically underperform. Research which extracts the skill and the luck component from the time series of private investor returns is to the best of our knowledge not existing.

This paper is exactly focusing on this issue and sets out to answer two crucial questions. First, do individual investors exhibit positive or negative skills, and second what is the individual investor investment skill worth measured in percentage terms?

Insights into the value of private investors’ skills are important for research and policy making. If underperformance of private investors would be temporal and driven be luck, then for policy makers there would be limited need for action, except for ensuring that private investors are not deceived by financial intermediaries, and participation in equity market is increasing to allow individuals to participate in the equity premium, which helps avoiding old age poverty. Yet, if the underperformance were substantial and due to negative skill, policy making and researchers should follow Campbell’s (2006) call for coming up with solutions to investment mistakes even more serious. Beyond the measures already suggested, transparency on the past performance of their portfolios might help people to learn whether they are skilled or not, since learning is only possible if individuals receive feedback (cf. Glaser and Weber 2007). Moreover, in the discussion of the participation puzzle negative skills of individuals may contribute to the literature explaining the low rate of participation, since the realized equity premium is directly affected by a potential negative skill of private investors. As a result, the equity premium computed by just comparing returns on equity and money market indices would be overstated and hence less people would be expected to participate due to their negative skill.

In order to disentangle skill from luck, we will draw on methods introduced by the mutual fund literature which has studied α generation since Jensen (1968), Gruber (1996) and Carhart (1997). Particularly, this paper uses a very recent methodology to differentiate
between skill and luck (Fama and French (2010)). Building on the latest methodological findings we back out individual investors’ skill from their return series.

The methodology by Fama and French (2010) is applied on a comprehensive dataset of 8,621 individual investor portfolios from a German online broker, spanning a timeframe from September 2005 to April 2010. Using a dataset from an online broker may not be exactly representative of the German population, but since Dorn and Huberman (2008) note online investor are more sophisticated than the average investor, we expect our results to be slightly overstating the skill of investors if there is any bias. Therefore, sample selection is from our point of view unlikely to drive the results. Since we are looking at individuals’ returns before (gross) and after (net) trading costs, we are able to control for any underperformance or negative skill induced through excessive trading. (Barber/Odean (1998)

Our results show that 89% of individual investors exhibit negative skill ($\alpha \leq 0$) when measured on a gross basis and 91% when considering returns net of costs and expenses. When backing out the value of individual investor skill we find it to be in the order of magnitude of -7.5% per year for gross returns and of -8.5% per year for net returns on average.

Nevertheless, even this results requires further attention to determine who underperforms on an individual level and why and may also trigger research investigating who in particularly benefits from the negative skill of private investors. Analyzing this question requires, however, a different and more comprehensive dataset. Moreover, banks, politicians as well as individual investors themselves might want to take our substantive finding into account to reconsider investment strategies and policies.

The remainder of this paper is organized as follows: First, Section I provides an overview of the literature which comprehensively addresses the measurement of individual investors’ investment performance. Additionally, it summarizes the most important biases and investment mistakes by individual investors and presents a synopsis of the general mutual fund performance literature and methodological approaches. In Section II we describe the dataset and necessary adjustments, before coming to the three main analyses: Equal-weighted
portfolios, simulation with true $\alpha$ of zero and simulations with a distribution of $\alpha$. For all three approaches we describe the methodology first and then discuss the results afterwards. Finally, Section III concludes.

II. Literature Review

A. Individual Investors’ Performance and Behavior

Academic studies on the comprehensive measurement of individual investors’ investment performance are scarce and – to the best of our knowledge – have never tried to disentangle skill from luck.

As Ivkovic and Weisbenner (2005) note:

“The Finance Literature has yielded a large number of in-depth studies concerning the investments managed by professional money managers, yet historically, relatively little has been known about the individual investors’ money management.”

Only Schlarbaum et al. (1978), Barber and Odean (2000) and Calvet et al. (2007) investigate the performance of a full cross-section of individual investors. Schlarbaum et al. (1978) analyze the increasing trend away from direct to indirect investments managed by professionals given the belief of professionals’ superior performance. The authors find that neither gross nor net individual performances significantly differ from buy-and-hold strategies. Barber and Odean (2000) confirm that before trading costs and expenses, individual investors do not perform significantly different from the benchmark. Yet, they find that net returns on average underperform by about 3.7% annually when comparing with Fama and French (1993) three-factor model. However, their analysis has a shortcoming as it approximates intra-month trading. Calvet et al. (2007) use a comprehensive Swedish dataset on household portfolios and find that the majority of Swedish households outperform the Sharpe ratio of their domestic stock index. Using a global CAPM, they determine that Swedish households suffer a median return loss of only 1.2%. Their results also suggest that households do not substantially underperform the market.
Recently, research has emphasized on identifying behavioral biases and investment mistakes exhibited by individual investors. The impact on overall performance has only been a side-issue in these studies. The disposition effect, security selection bias and lacking ability of market timing are repeatedly documented in the literature. The disposition effect has been brought up by Shefrin and Statman (1985) and analyzed in depth by Odean (1998), who found that investors tend to hold losers too long and sell winners too early. The security selection bias has been discussed by Barber et al. (2009) using a comprehensive Taiwanese sample. They show that individual investors lose on average with their trades, which is largely attributable to aggressive and obviously overconfident orders. This is also consistent with the overconfidence of individual investors described by Barber and Odean (2000) and the observation of the particularly strong overconfidence of males as pointed out by Barber and Odean (2001). Barber and Odean also show that overconfidence leads to excessive trading and consequently to a poor net performance because of trading costs and expenses. Finally, Barber et al. (2009) find that individual investors lack the ability to time the market and, hence, lose money on it. Furthermore, there are other papers providing only partial evidence on the individual investors’ performance puzzle. Odean (1998) presents a theoretical framework which attributes the underperformance of active individual investors to overconfidence and costly information. Campbell (2006) argues that individual investors face complex problems when it comes to investment decisions and that in particular less educated investors, exhibit serious investment mistakes resulting in underperformance. Moreover, French (2008) estimates that fees, expenses and trading costs of active investing amount to 67 basis points. However, no study so far has focused on disentangling skill from luck in private investors’ portfolio performance.

B. Methodological Approach in the Mutual Funds’ Performance Literature

In order to investigate the distribution of skill of individual investors in depth, we will draw on thorough methodological foundations of the mutual fund literature. In this area, there are a number of studies which have addressed similar research questions.
Gruber (1996) points out that it is crucial to question whether some investors, in this case mutual fund managers, possess true skills in selecting the right stocks – also described as “hot hand”⁴ by Hendricks et al. (1993) – or whether observed fund performances are determined by pure chance and hence only reflect managers’ luck.

Generally, there has been some consensus in the fund literature that active investing cannot beat a passive benchmark and that funds rather exhibit underperformance when trying to do so. Hence, the general conclusion is that buy-and-hold appears as the dominant strategy. The most important studies in this respect include Elton et al. (1993), Grinblatt et al. (1995), Gruber (1996) and Carhart (1997). The aforementioned authors find that common factors in stock returns, persistent differences in mutual fund expenses and transaction costs explain nearly all of the predictability in mutual fund returns. Theoretical foundations were also laid by Berk and Green (2004) who argue that, following a long-run equilibrium theory, abnormal fund returns are bid away in competitive markets. The authors show that mutual funds face costs which can be described as an increasing convex function of assets under management. Following this thought, a fund with a positive expected $\alpha$ before costs attracts inflows until its assets under management reach the point where expected $\alpha$, net of costs, is zero while outflows drive out funds with negative expected $\alpha$ vice versa. Nevertheless, the picture remains fragmented for mutual funds as well. In contrast to previous findings, Grinblatt and Titman (1992), Goetzmann and Ibbotson (1994), Brown and Goetzmann (1995) as well as Elton et al. (1996) and Wermers (2000) provide partial evidence of positive $\alpha$ generation related to managers’ skills and investment strategies.

More recent studies have tried to shed additional light on the performance question by applying statistically more sophisticated methods. Using a CRSP/CDA sample of 2,188 open-end mutual funds for a time period spanning from 1975 to 2002, Kosowski et al. (2006) simulate return series for individual portfolios with an underlying $\alpha$ of zero using 1,000 bootstrap simulations. They compare the obtained average $t(\alpha)$ statistics to the $t$-values derived from the actual return series. This approach does not directly investigate the cross-sectional distribution of $\alpha$ and their significance levels, which would only be valid for the rare

---

⁴ Cf. Hendricks et al., p. 94.
case of residuals drawn from a multivariate normal distribution, zero correlations in the residuals and identical risk levels pursued by investors. The authors observe some stock picking ability by mutual fund managers and differences in the cross-section related to the investment objectives chosen (growth-oriented vs. income-oriented). Using a UK sample of 935 open-end mutual funds for the time period 1975 to 2002 and applying a methodology similar to Kosowski et al. (2006), Cuthbertson et al. (2008) also discover evidence for stock picking ability and, hence, positive skills by mutual fund managers. Moreover, the authors present strong evidence for negative skill and its persistence. Also applying a similar methodology, Fama and French (2010) analyze mutual fund performance for the time period 1984 to 2006 based on a sample of 3,156 funds from the CRSP database. They find proof of net underperformance on an aggregated level, independent of the benchmark model used (CAPM, three-factor model, four-factor model). Looking at individual funds, Fama and French (2010) discover that only few funds are able to cover their costs, while on a gross basis some superior und inferior performance can be observed in the extreme tails of the distribution.

Finally, Barras et al. (2010) suggest a different method to distinguish lucky and unlucky funds from skilled and unskilled funds, respectively. They form three groups of funds (unskilled funds, zero-α funds and skilled funds) and apply a false discovery rate approach as developed by Storey (2002) to examine the distribution of p-values on sampled t-statistics. They find very few “hot hands” for very short-term and in the extreme tails of the distribution as well as significant persistence in truly negative-α funds performance using a sample of 2,076 actively managed U.S. open-end, domestic equity mutual funds for the period from 1975 to 2006.

III. Empirical Analysis

In this section we present the empirical analysis of this paper. After presenting the dataset and descriptive statistics, we turn to the analysis which consists of three steps. In the first one we compute the alpha of individual investors in the way it has been done in studies like Charhart

3 Cf. Kosowoski et al. (2006) for a detailed methodology overview.
(1997), then we use a bootstrap approach introduced by Fama and French (2010) to disentangle skill from luck and finally we fit a distribution to the results in order determine how much the skill of private investors is worth in percentage terms. For presentation purposes we first describe the methodology to then present and discuss the results for each of the three steps.

A. Dataset

Our dataset consists of weekly portfolio returns of 8,805 retail investors of a German online broker between September 2005 and April 2010 (242 weeks). The returns exist both gross as well as net of trading fees. The procedure of generating the series of daily gross and net returns for each individual works as described in Bhattacharya et al. (2012).

We drop 184 portfolios, which have less than half a year of returns, i.e., less than 26 weeks. We thereby obtain a final sample of 8,621 portfolios.

The investors in our sample display a strong focus on equity and funds as Table I shows.

<table>
<thead>
<tr>
<th>Asset Class</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equity</td>
<td>58.3%</td>
<td>36.5%</td>
</tr>
<tr>
<td>thereof German equity</td>
<td>69.9%</td>
<td>30.4%</td>
</tr>
<tr>
<td>Bonds</td>
<td>2.0%</td>
<td>8.6%</td>
</tr>
<tr>
<td>Funds</td>
<td>35.5%</td>
<td>35.2%</td>
</tr>
<tr>
<td>Certificates</td>
<td>3.3%</td>
<td>9.1%</td>
</tr>
<tr>
<td>Other</td>
<td>0.9%</td>
<td>5.1%</td>
</tr>
</tbody>
</table>

On average, they invest 58.3% of their portfolio in equity and 35.5% in funds. Bonds, certificates (structured financial products) and other financial instruments (mainly warrants and participation rights) have a share of only 6.2%. Moreover, consistent with the existing
literature, the investors exhibit a strong home bias. About 69.9% of the equity is domestic equity.

The sample period contains the financial crisis of 2008 and 2009. We find that skill levels of individual investors are less negative in our sample period prior to the crisis, i.e. from September 2005 to the end of 2007. However, we will nonetheless focus on the full sample period as we try to generate a perspective on the skill through a full cycle. Even if investors in certain time periods would be skilled enough to outperform the market, we need to consider their skill level also across all other periods to assess their overall ability to make good investment decisions.

An issue that arises is, however, that we can only observe investors’ investment portfolios. If most investors withdrew their money from the market prior to the crisis and, thus, the negative skill during the crisis holds only for a small subsample of investors, our results might turn out to be too negative. However, the number of investors is rather steadily increasing as an outcome of the growth of the broker as it can be seen in Exhibit I.

\begin{center}
Exhibit I

\textbf{Number of portfolios}
\end{center}

The exhibit shows the total number of portfolios, having an account value larger than zero, over the sample period from September 2005 to April 2010.
Moreover, both the relative addition of new portfolios and the full sale of portfolios per week remain relatively constant over time with only a few extreme peaks (see Exhibit II). It does not allow the conclusion that the sample contains a relevant self-selection or survivorship bias.

Exhibit II

Change in number of portfolios

The exhibit shows the weekly relative change in the number of portfolios in our sample compared to the previous week split into new portfolios and portfolios fully sold. The new portfolios are those portfolios that are observed for the first time in the given week or previous portfolios that changed from being empty to containing at least one security. Similarly, portfolios fully sold are those portfolios in a given week that either were deleted completely or have seen a reduction in the number of securities to zero, i.e. being empty through full divestment.

B. Equal-weighted portfolios

Methodology

A standard method to determine the average under- or outperformance of a group of portfolios is to compute an equal-weighted portfolio and estimate its \( \alpha \) following Jensen (1968). We do so applying the four-factor model of Carhart (1997):

\[
R_{i,t} - R_{ft} = \alpha_i + \beta_i \cdot (R_{Mt} - R_{ft}) + s_t \cdot SMB_t + h_i \cdot HML_t + m_t \cdot MOM_t + \varepsilon_{i,t},
\]

where \( R_{i,t} \) is the return of portfolio \( i \) for week \( t \), \( R_{ft} \) is the 1 month EURIBOR, \( R_{Mt} \) is the week \( t \) benchmark return, \( SMB_t \) and \( HML_t \) are the returns for the size and value-growth
portfolios according to Fama and French (1993), and \( MOM_t \) is the one-year momentum return as defined by Carhart (1997).

As investors exhibit a strong home bias, the benchmark chosen is the CDAX. It is the value-weighted index of all German stocks listed in the Prime or General Standard on the Frankfurt Stock Exchange, which is Germany’s most important exchange. Our choice of the CDAX needs to be briefly substantiated. Obviously, a domestic benchmark cannot constitute the market portfolio as it lacks international equities. Yet, in the time span from September 2005 to April 2010, the CDAX outperformed the MSCI All Country Investable Market Index, an index which is closer to the international market portfolio (see Exhibit III).

Exhibit III

**Performance of MSCI AC IMI and CDAX**

The exhibit shows the performance (rebased to 100) of MSCI AC IMI and CDAX from September 2005 to March 2010.

Since investors in our sample hold a large share of German securities, the outperformance of domestic equity in that specific period might lead to an overestimation of \( \alpha \) for that period when applying the international benchmark. In contrast, applying the CDAX will account for
the outperformance of the German market. Another issue with using the CDAX could be that the home bias of the individual investors is no reason to apply a similarly biased, i.e., domestic benchmark. However, consider the three possible reasons for the home bias: Either it is an efficient limitation, an efficient selection skill or an inefficient phenomenon. Assuming the first, home bias would be a rational investment limitation to German investors, which we simply account for by choosing the CDAX. Assuming the second, investors would tilt their portfolio towards German equity in the sample period because of superior selection skill. However, this would require that the home bias phenomenon varies across time, which is not the case as it is a persistent phenomenon (see e.g. Lewis (1999)).

Finally, assuming the third explanation, we would observe the same inefficiency in our benchmark as the investors in their portfolio by choosing the CDAX. Applying the CDAX will, hence, filter that inefficiency and we ultimately underestimate \( \alpha \). However, the outperformance of the CDAX is unlikely to be a stable phenomenon over time, assuming semi-efficient markets. Thus, the outperformance in the sample period is rather an outcome of luck. As we aim at filtering all luck, we will generate a more accurate perspective on the skill level of individual investors when filtering the random effect of the CDAX outperformance coupled with an inefficient home bias, which just by chance improved individual investors’ performance in our sample. Consequently, we choose the CDAX as our benchmark. Nonetheless, we also tested the results with the MSCI AC IMI as shown in the Appendix, which, as predicted, leads to higher estimated mean skill levels, but not a substantially higher share of skilled investors.

Moreover, choosing the CDAX over an international index is consistent with both the broad literature stream using US-focused benchmarks for US-based analyses as well as suggestions from Daniel et al. (1997) and Koijen (2010) to determine a best fitting benchmark as the correct benchmark.

Furthermore, a concern could be related to the multiple asset classes in investors’ portfolios, which are not limited to equity, but also consist of bonds, mutual funds, certificates and little investment in other asset classes. In fact, individual bonds only account for 2.0% in investors’ portfolios compared to 58.3% invested in equity. Assuming that the remaining 39.7% of
funds, certificates and other securities do contain a similar split between equity and bonds, individual investors would have a total share of only 3.3% bonds in their portfolios. Hence, we assume that an all equity index like the CDAX is quite a fair benchmark as other asset classes like bonds, commodities or real estate seem to have a marginal share in individual investors’ brokerage portfolios. Nonetheless, we perform a robustness test in the Appendix with a German and international multi-asset index as suggested by Jacobs et al. (2010). The results do not change substantially when using the multi-asset indices and our conclusions are robust to these alternative benchmarks.

The computation of weekly $SMB_t$ and $HML_t$ for the German market applies the methodology of Fama and French (1993) to the CDAX constituents. For the momentum factor $MOM_t$ we follow Carhart (1997). The only difference to Carhart (1997) is that we lag the eleven-month returns used for constructing the high and low-momentum portfolios only by one week instead of one month, since our dataset consists of weekly returns. We use Datastream to obtain the constituent list of the CDAX, return series, market capitalization and book equity values (common equity, preferred equity and deferred taxes).

We compute the equal-weighted portfolios for both gross and net returns.

Results

The results for the equal-weighted portfolio analyses are presented in Table II. Panel A contains the results for net returns and Panel B shows the results for the returns before expenses and trading fees. We will briefly discuss the coefficients $\beta_i$, $s_i$, $h_i$ and $m_i$ before turning to the estimated $\alpha$.

In both panels, the equal-weighted portfolio has a market premium coefficient of 0.89. The largest part in the portfolio is domestic equity and the benchmark used is the value-weighted domestic equity. Hence, it is clear that the market premium has strong explanatory power and is relatively high. That it is statistically significantly different from 1, though, can be explained by the fact that foreign equity and non-equity, which the portfolio also contains, is usually not perfectly explained through the market premium for domestic equity. In its most extreme form of no correlation at all, the coefficient would be zero. Consequently, the larger
share of foreign equity and non-equity reduces the coefficient. Moreover, the coefficient might indicate that individual investors prefer less risk than implied by the market portfolio.

While the coefficients for SMB and HML are not statistically significant, the estimated coefficient of MOM is weakly significant at -0.07. It implies a negative correlation with a momentum strategy, which might result from the tendency to sell past winners and hold or even buy past losers. That would be in line with the findings of Odean (1998).

Finally, we turn to the estimated $\alpha$. For easier interpretation we have annualized the weekly $\alpha$. The net $\alpha$ is substantially negative at -5.84%. Yet, $\alpha$ is only different from zero at a 10% significance level. The magnitude itself is a bit larger than the estimate of Barber and Odean (2000), who find an $\alpha$ after transaction costs of -3.7%. The annualized $\alpha$ for gross returns is also quite large with -4.94%, but not significant at any reasonable significance level. This is consistent with the insignificant gross $\alpha$ of Schlarbaum et al. (1978) and Barber and Odean (2000). Hence, using this standard approach of equal-weighted portfolios, we can only identify a statistically weak on-average underperformance of individual investors after transaction costs and no underperformance at all before costs.

One additional observation can be made by looking at the equal-weighted portfolio estimates. The estimated net $\alpha$ is about 90 basis points lower than the gross $\alpha$. It implies that the

---

**Table II**  
Results of the 4-Factor Model Applied To Equal-Weighted Portfolio

The table presents the results of the 4-factor model for the equal-weighted portfolio. Panel A contains results for net portfolio returns and Panel B for gross portfolio returns before expenses and trading fees. The table displays the annualized intercept in percent ($52^*\alpha$), the slope of the four factors, i.e., market premium ($\beta$), size ($s$), value ($h$) and momentum ($m$) as well as the $R^2$ of the regression model. Values in brackets are the corresponding t-statistics. The null hypothesis for the coefficient estimate on the market premium is $H_0: \beta = 1$.

<table>
<thead>
<tr>
<th>$52^*\alpha$</th>
<th>$\beta$</th>
<th>$s$</th>
<th>$h$</th>
<th>$m$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Net Returns</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5.84 *</td>
<td>0.89 ***</td>
<td>0.12</td>
<td>0.05</td>
<td>-0.07 *</td>
<td>0.91</td>
</tr>
<tr>
<td>(-1.73 )</td>
<td>(-2.86 )</td>
<td>(1.39 )</td>
<td>(0.92 )</td>
<td>(-1.84 )</td>
<td></td>
</tr>
<tr>
<td>Panel B: Gross Returns</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4.94</td>
<td>0.89 ***</td>
<td>0.12</td>
<td>0.05</td>
<td>-0.07 *</td>
<td>0.91</td>
</tr>
<tr>
<td>(-1.46 )</td>
<td>(-2.88 )</td>
<td>(1.38 )</td>
<td>(0.91 )</td>
<td>(-1.86 )</td>
<td></td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1
expenses and trading costs amount to about 90 basis points on average. This is similar to French (2008), who estimates that they amount to 67 basis points for the entire U.S. market.

C. Bootstrapping simulation with a true $\alpha$ of zero

Methodology

The results for the equal-weighted portfolio are of limited value when it comes to the question of judging individual investors’ investment skill. Thus, we investigate this issue further by identifying how many individual investors exhibit a positive $\alpha$ and how many of those have true skill versus mere luck. We use the bootstrapping simulation approach of Fama and French (2010), which allows the disentangling of skill and luck. To do so, it simulates return series for individual portfolios with a true $\alpha$, i.e., an underlying $\alpha$ of zero, and compares this simulation outcome with the actual return series.

The bootstrapping simulation overcomes fundamental statistical issues of directly investigating the cross-sectional distribution of $\alpha$: For the significance levels to be appropriate, we would have to make the unrealistic assumption that residuals are drawn from a multivariate normal distribution, that correlations in the residuals are zero and investors have identical risk levels (see Kosowski et al. (2006) for an in-depth discussion of the advantages of the approach chosen). Consequently, the bootstrapping approach is clearly superior for drawing inferences from the cross-section of $\alpha$.

Additionally and irrespective of statistical issues, an individual $\alpha$ significantly different from zero might be the outcome of the investor’s mere luck during that observation period. Due to the lack of further observations, one could then not reasonably conclude whether an $\alpha$ is the outcome of luck or true skill. Since the return of luck is centered around zero and assuming no cross-sectional correlation and a normal distribution, a joint test of cross-sectional $\alpha$ could mitigate the problem. However, the unknown distribution of $\alpha$ prevents such a joint test. To illustrate this problem, assume that an investor with a true $\alpha$ of zero was lucky over the observation period and let us estimate a significant $\alpha$ of 5%. If we knew that all investors in the cross-section also had a true $\alpha$ of zero, we could take the estimated $\alpha$ of all other investors into account and jointly test whether the $\alpha$ are different from zero. Yet, we do not know
whether all investors have that true $\alpha$ of zero and, henceforth, cannot perform the joint test. This is a problem the bootstrapping simulation overcomes by generating a controlled environment which maintains most of investors’ characteristics, but allows prespecifying a known cross-sectional distribution of $\alpha$. Hence, the bootstrapping allows us to differentiate much more precisely between luck and skill and identify the underlying distribution of true $\alpha$.

The simulation consists of five steps: (i) Estimating an $\alpha$ and its t-statistic for each portfolio and calculating the value of the cross-sectional $t(\alpha)$ distribution at each percentile, (ii) subtracting the estimated actual $\alpha$ of a portfolio from its weekly returns to generate a return series with a true $\alpha$ of zero, (iii) randomly drawing weeks to generate new return series per simulation run, (iv) estimating a new simulated $\alpha$ and $t(\alpha)$ per portfolio in each simulation run, (v) computing the $t(\alpha)$ value at each percentile as an average of the percentile values from all simulation runs. These five steps will subsequently be described in more detail.

The first step is to calculate $\alpha$ and $t(\alpha)$ for each portfolio for the full time period of up to 242 weeks. This is done by applying the four-factor model as described in the previous section on the equal-weighted portfolio analysis. In the next step, the estimated $\alpha$ is used to generate an adjusted return series for each portfolio with $\alpha$ of zero. This is done by simply subtracting the estimated $\alpha$ from the actual weekly returns. Subsequently, we construct a new return series for each portfolio by randomly drawing from these adjusted returns 242 times with replacement. Since the returns for all portfolios are drawn simultaneously, the same weeks are used across portfolios. Thereby, the analysis captures any cross-sectional effects and correlations. Possible effects across time will, however, be lost.\(^4\)

The newly created return series are then used to estimate $\alpha$ for each portfolio. Due to the random sampling, however, the estimated $\alpha$ will regularly deviate from zero, although the underlying $\alpha$ is zero by design. Any estimated $\alpha$ different from zero is then obviously just one generated by luck. This procedure of random drawing and estimating new $\alpha$ and $t(\alpha)$ constitutes one simulation run. It is repeated 1,000 times.

\(^4\) For further discussion of the advantages and disadvantages of the methodology, see Fama and French (2010).
The distribution of the actual $t(\alpha)$ and the simulated $t(\alpha)$ can be then compared to infer whether the actual distribution is generated by mere luck or whether some investors exhibit skill. We follow the approach of Fama and French (2010) by comparing the values at the percentiles. For each of the 1,000 simulations we calculate the value at every percentile. For the comparison, we then compute the average value at the percentiles as well as a figure representing how many simulations in percent generated a value at the respective percentile that was below the actual percentile value.

Following Kosowski et al. (2006), Fama and French (2010) as well as Barras et al. (2010), we use the t-statistics of $\alpha$, $t(\alpha)$, instead of $\alpha$ for the analysis. This is because $t(\alpha)$ has a major advantage over $\alpha$ for our purposes: As a measure of significance it accounts for differences in the precision of $\alpha$. Accordingly, the t-statistic normalizes differences in risk-taking and number of observations in the cross-section. Consider, for example, an investor who takes high risks or who only has a short time series of portfolio returns. That investor can more easily exhibit extreme estimated $\alpha$. However, those extreme $\alpha$ will likely be spurious outliers. The t-statistic corrects for that by scaling the $\alpha$ by its standard error. Through this rescaling, the t-statistic accounts even more generally for differences in risk-taking and number of observations in the cross-section. The cross-sectional distribution of $t(\alpha)$ hence has the statistically preferable attribute of being closer to a normal distributed than the cross-sectional distribution of $\alpha$.

Throughout the paper we use Newey and West (1987) heteroskedasticity-robust standard errors to compute those t-statistics. Heteroskedasticity is found in the majority of regressions. We also tested for serial correlation using the test of Breusch-Godfrey, but in the majority of regressions we did not find serial correlation given a 5% confidence level. Hence, we did not account for serial correlation, which has the advantage that it enhances comparability between actual and simulated t-statistics through a uniform test specification. This is because the later simulations consist of random drawings of individual weeks, which means the time series drawn cannot contain any true underlying serial correlation by design. So we also do not account for serial correlation during the simulations.
Results

Table III depicts the results obtained from the bootstrapping simulations with a true $\alpha$ of zero. Panel A exhibits the t($\alpha$) values at selected percentiles of the distribution for net returns, while Panel B exhibits the t($\alpha$) values at selected percentiles of the distribution for gross return calculations. In both panels, the actual values (“Act”) are presented in the first column versus average t($\alpha$) values for the 1,000 simulations (“Sim”) in the second column. Additionally, we provide the share of the 1,000 simulation runs that have produced lower t($\alpha$) values at the given percentile than the corresponding actual t($\alpha$) value (“%<Act”).

Similar to Fama and French (2010) we face a multiple comparison issue and a correlation in likelihoods for the different percentiles. However, in order to not lose too much information we follow their approach by examining all likelihoods and placing emphasis on the extreme tails.

The average t($\alpha$) values from our simulations are similar for net and gross returns given the true $\alpha$ is set to zero by design.

At first glance we observe that actual investors’ portfolio t($\alpha$) are always below the corresponding average t($\alpha$) values from the simulations for both, net and gross returns, except for the 99th percentile.

Investigating the left tail of the distribution on a net basis, we observe a t($\alpha$) value of -2.18 for the simulations versus an actual t($\alpha$) value amounting to -3.20 for the first percentile. This represents a strong underperformance compared to the zero-$\alpha$ simulation. The same holds true independent of the position in the performance distribution which we examine. For the fifth percentile we observe a simulated t($\alpha$) value of -1.52 against an actual value of -2.46 while we see t($\alpha$) value of -0.04 and -0.84 for the 50th percentile. Also when looking at the right tail of the performance distribution private investors underperform the true-$\alpha$ equal to zero simulations on a gross and net basis. For the 95th percentile and on a net basis, the simulated t($\alpha$) value of 1.51 compares to an actual t($\alpha$) of 0.94. Even for the 99th
and thus in the very tail, are the simulated $t(\alpha)$ statistics with 2.22 larger than the actual $t(\alpha)$ values of 1.92.

Table III

Percentiles of $t(\alpha)$ for Actual and Simulated Returns with True $\alpha$ of Zero

The table shows the estimated $t$-statistics for $\alpha$ at selected percentiles (Pct) for actual net returns in Panel A and actual gross returns in Panel B (Act). The table also contains the average of all the values of $t(\alpha)$ at the respective percentile from 1,000 simulations with a true $\alpha$ of zero (Sim). Finally, the last column presents the percent of the 1,000 simulations that produced a $t(\alpha)$ at the respective percentile that was below the actual value of $t(\alpha)$. The results are for the 4-factor model and the time period from September 2005 to April 2010 for all 8,621 portfolios in the sample.

<table>
<thead>
<tr>
<th>Panel A: Net Returns</th>
<th>Panel B: Gross Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t(\alpha)$</td>
<td>$t(\alpha)$</td>
</tr>
<tr>
<td>Pct</td>
<td>Act</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>1</td>
<td>-3.20</td>
</tr>
<tr>
<td>2</td>
<td>-2.90</td>
</tr>
<tr>
<td>3</td>
<td>-2.73</td>
</tr>
<tr>
<td>4</td>
<td>-2.58</td>
</tr>
<tr>
<td>5</td>
<td>-2.46</td>
</tr>
<tr>
<td>6</td>
<td>-2.06</td>
</tr>
<tr>
<td>7</td>
<td>-1.62</td>
</tr>
<tr>
<td>8</td>
<td>-1.32</td>
</tr>
<tr>
<td>9</td>
<td>-1.07</td>
</tr>
<tr>
<td>10</td>
<td>-0.84</td>
</tr>
<tr>
<td>11</td>
<td>-0.59</td>
</tr>
<tr>
<td>12</td>
<td>-0.30</td>
</tr>
<tr>
<td>13</td>
<td>0.04</td>
</tr>
<tr>
<td>14</td>
<td>0.52</td>
</tr>
<tr>
<td>15</td>
<td>0.94</td>
</tr>
<tr>
<td>16</td>
<td>1.07</td>
</tr>
<tr>
<td>17</td>
<td>1.26</td>
</tr>
<tr>
<td>18</td>
<td>1.47</td>
</tr>
<tr>
<td>19</td>
<td>1.92</td>
</tr>
</tbody>
</table>

Previous observations also apply to our measurements on a gross return basis while we observe a small general upward shift in actual values, which is attributable to the fact that gross returns exclude trading costs and expenses. Again, actual $t(\alpha)$ values are significantly smaller than corresponding zero-$\alpha$ simulated values throughout the distribution. For the first percentile we observe a negative actual $t(\alpha)$ value of -3.08 against the simulated value of -2.19. For the right tails of the distribution, our results indicate some convergence in actual to simulated $\alpha$. Yet, for the 99th percentile the actual gross $t(\alpha)$ value of 1.99 is still smaller than the simulated average $t(\alpha)$ value of 2.22.
Examining the number of simulations that were below the actual $t(\alpha)$ ("%<Act") amends our findings. Generally, insufficient skill to cover costs can be postulated if low fractions of the simulation runs produce left tail percentiles of $t(\alpha)$ below those from actual net fund returns. In contrast, we would argue for returns that more than cover costs if large fractions of the simulation runs produce right tail percentiles of $t(\alpha)$ below those from actual net fund returns.

Starting with the left tail of the distribution we find that for the 5th percentile, actual individual investors’ returns beat those of the zero-\(\alpha\) simulation in only 2.1% of the simulated cases when measured on a net basis and in only 4.5% of the gross cases. Even at the 90th percentile, only 4.4% of actual net $t(\alpha)$ and 7.3% of actual gross $t(\alpha)$ were better than the zero-\(\alpha\) simulations. Thus, there is again evidence that the large majority of individual investors are neither able to generate sufficient returns to cover costs nor to generate positive\(\alpha\) even when disregarding costs.

Although the gap between actual and simulated zero-\(\alpha\) t-statistics closes somewhat in the upper tail of the skill distribution, we cannot say that those investors are skilled enough to outperform the market. For the 99th percentile, we observe that investors beat the simulated zero-\(\alpha\) benchmark in 25.4% of the cases on a net return basis and in 31.9% of the cases on a gross return basis. This is a higher percentage than for most other percentiles, but is still too low to indicate skill. It rather suggests that even the best performing investors are more driven by luck than by skill when outperforming the market.

This evidence reveals that the large majority of retail investors do not possess sufficient skill to generate positive abnormal returns, neither on a net nor gross basis. Even worse, private investors consistently show negative\(\alpha\) performance as investors apparently underperform our bootstrapped zero-\(\alpha\) benchmark at all percentiles based on 1,000 simulations. Hence, the subsequent question is how many individual investors are skilled and how many are unskilled and to what extent?

**D. Bootstrapping simulation with a simulated \(\alpha\) distribution**

**Methodology**
In the previous section we found that the underlying true $\alpha$ has to be negative for most investors since the actual values of $t(\alpha)$ are lower at almost any percentile than those from a simulation with true $\alpha$ of zero. Yet, not all investors will exhibit the same $\alpha$. True $\alpha$ will stem rather from an unknown distribution. This distribution is what we want to identify subsequently in order to infer how many individual investors exhibit skill and how many do not, as well as to what extent they are skilled or unskilled.

To identify the underlying $\alpha$ distribution, we use the same simulation approach as before, but before drawing from the return series with a true $\alpha$ of zero, we draw a new $\alpha$ to be added to the return series for each portfolio and for each simulation run. This $\alpha$ is drawn from a distribution specified beforehand. We start our analysis with normal distributions with negative means. Yet, as we find a tendency of the upper tail to be too low, we also apply slightly skewed normal distributions with a bit of fat tails (skewness $\gamma = 1$, kurtosis $\kappa = 4$), which fit the actual $t(\alpha)$ distribution in our sample better as will be discussed and shown below.

Before adding the drawn $\alpha$ to the return series, we rescale $\alpha$ by the ratio of the residual standard error of the individual investor to the average standard error of all investors. This rescaling captures the fact that more diversified investors are less able to generate extreme $\alpha$ than less diversified investors. It also accounts for the issue that investors with the same level of diversification have different probabilities to generate large $\alpha$, depending on their risk. Investors with a higher risk will more easily generate large $\alpha$ compared to an investor with the same level of diversification but less risky assets. This approach is also in line with Fama and French (2010).

**Results**

We performed various simulations testing for different normal distributions from which to draw $\alpha$. Specifically, for both net and gross returns we test 25 normal distributions based on any combination of 5 different means and standard deviations. Values for the mean $\mu$ are $-4\%$, $-6\%$, $-8\%$, $-10\%$ and $-12\%$ and values for the standard deviation $\sigma$ are $2\%$, $4\%$, $6\%$, $8\%$ and $10\%$. The results of those 25 different combinations can be found in Exhibit IV for net
returns and in Exhibit V for gross returns. The full set of underlying data can be obtained from the authors upon request.

We use the information of how many of the simulation runs yielded values of $t(\alpha)$ that were above or below the actual $t(\alpha)$ to evaluate how well the simulated $\alpha$ distribution fits the unknown actual $\alpha$ distribution. If most simulation runs were above the actual $t(\alpha)$, the simulated $\alpha$ distribution is likely too high. If the simulation runs yielded mostly $t(\alpha)$ below the actual value, the simulated $\alpha$ distribution is likely too low. The perfect match would mean that 50% of the simulations are above and 50% are below the actual $t(\alpha)$.

Moreover, we need a measure to decide which values would be too extreme and could reasonably be rejected. To do so, we first have to determine a confidence level. Following Fama and French (2010) we set it to 20%. Specifically, we are willing to accept a 20% chance of setting a lower bound for $\mu$ and $\sigma$ that is too high and a 20% chance of setting an upper bound that is too low. This rule implies that a minimum of 20% and a maximum of 80% of the simulation runs have to generate a $t(\alpha)$ lower than the actual $t(\alpha)$ at every percentile. Exceeding this range, we would reject the respective combination of $\mu$ and $\sigma$ as too extreme.

Taking a look at the results, most combinations of $\mu$ and $\sigma$ exceed the valid range of 20% to 80%. For net returns, five combinations lie within the limits: $\mu = -6\%$ with $\sigma = 4\%$ and 6\% and $\mu = -8\%$ with $\sigma = 4\%, 6\%$ and 8\%. For gross returns there are also five combinations, which seem feasible: $\mu = -6\%$ with $\sigma = 2\%, 4\%, 6\%$ and 8\% as well as $\mu = -8\%$ with $\sigma = 6\%$.

All these combinations imply quite a negative skill for individual investors: For net returns, 84\% to 91\% of individual investors would have a negative skill with a mean of about -7\% or -10\% per year, depending on which standard deviation to assume. In contrast, only a meager 9\% to 16\% of investors would have a positive skill with a mean of 1\% to 4\% per year. For gross returns the bandwidth is a bit broader, but still very negative: 77\% up to 99.9\% exhibit a negative skill with a mean between -6\% and -9\%. In contrast, only 0.1\% to 23\% of the individual investors exhibit positive skill levels with a mean between 1\% and 4\%.
However, it is quite remarkable that in all simulations the generated $t(\alpha)$ tend to be too low at the upper percentiles as indicated by the rightward trend at the bottom of each graph. It indicates that a normal distribution tends to underestimate the upper bound of individual investors’ positive skill. Conversely, it also indicates that the results might still overestimate
Exhibit IV

Percentage of Estimated t-Statistics Below Actual t(α) at Percentiles Given Normal Distribution and Net Returns

The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a t(α) below the actual value of t(α) at the respective percentiles (y-axis). Simulation results are presented for net returns with a normal distribution with a prespecified mean μ ranging from μ=-4% to μ=-12% and standard deviation σ ranging from σ=2% to σ=10%. All results are for the 4-factor model and the time period from September 2005 to April 2010 for all 8,621 portfolios in the sample.
Exhibit V

Percentage of Estimated t-Statistics Below Actual $t(\alpha)$ at Percentiles Given Normal Distribution and Gross Returns

The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a $t(\alpha)$ below the actual value of $t(\alpha)$ at the respective percentiles (y-axis). Simulation results are presented for gross returns with a normal distribution with a prespecified mean $\mu$ ranging from $\mu=-4\%$ to $\mu=-12\%$ and standard deviation $\sigma$ ranging from $\sigma=2\%$ to $\sigma=10\%$. All results are for the 4-factor model and the time period from September 2005 to April 2010 for all 8,621 portfolios in the sample.
the middle and lower bound of the skill distribution. This is due to the correlation in the resulting $t(\alpha)$ distribution: Adjusting the distribution from which $\alpha$ is drawn to increase the likelihood of large, positive $t(\alpha)$ outcomes will lead to higher $t(\alpha)$ across the whole distribution.

In order to account for the underestimation of the upper bound of positive skill, we perform the same kind of simulations with a slightly skewed normal distribution with fat tails for $\alpha$, specifically with a skewness of $\gamma = 1.5$ and kurtosis of $\kappa = 6$. It tilts the normal distribution in a way that the likelihood to draw $\alpha$ from the right tail of the distribution increases. Hence, it should reduce the problem of underestimating the right tail. We use -6%, -8%, -8.5%, -9% and -11% for the mean $\mu$ for net returns and -5%, -7%, -7.5%, -8% and -10%. For the standard deviation $\sigma$ we use 2%, 4%, 6%, 8% and 10% in both cases. The results can be found in Exhibit VI for net returns and Exhibit VII for gross returns.

The skewness seems to largely account for the underestimation of the upper tail as the persistent rightward drift at the bottom of the graph is missing in both net and gross results. Moreover, one can easily identify the best fitting $\alpha$-distributions which lead to values of $t(\alpha)$ close to the actual values: For net returns, a mean of -8.5% and a standard deviation of 6% is very close to the 50% mark at any percentile. Similarly, for gross returns a mean of -7.5% and a standard deviation of 6% seem to generate the best fitting $\alpha$ distribution.

While there are also a select number of other combinations of means and standard deviations scattered around those two best fitting combinations, we will focus our subsequent interpretation on the aforementioned best fitting values for reasons of comprehensibility. The other possible values do not significantly change the overall results, anyway.

The results imply that individual investors do not have much skill: Only about 11% of individual investors have a gross $\alpha$ of zero or above. Considering $\alpha$ net of expenses and trading costs, even fewer investors exhibit skill. Just about 9% of individual investors have a net $\alpha \geq 0$. Accordingly, about 89% of individual investors already have an overall negative skill and when considering returns net of expenses, a total of 91% have a negative skill. Moreover, the magnitude of the negative skill is remarkable.
Exhibit VI

Percentage of Estimated t-Statistics Below Actual t(α) at Percentiles Given Skewed Normal Distribution and Net Returns

The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a t(α) below the actual value of t(α) at the respective percentiles (y-axis). Simulation results are presented for net returns with a skewed normal distribution with a prespecified mean $\mu$ ranging from $\mu=-6\%$ to $\mu=-11\%$, a standard deviation $\sigma$ ranging from $\sigma=2\%$ to $\sigma=10\%$ and a skewness $\gamma$ of $\gamma=1.5$ and a kurtosis $\kappa$ of $\kappa=6$. All results are for the 4-factor model and the time period from September 2005 to April 2010 for all 8,621 portfolios in the sample.
The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a t(α) below the actual value of t(α) at the respective percentiles (y-axis). Simulation results are presented for gross returns with a skewed normal distribution with a prespecified mean µ ranging from µ=−5% to µ=−10%, a standard deviation σ ranging from σ=2% to σ=10% and a skewness γ of γ=1.5 and a kurtosis κ of κ=6. All results are for the 4-factor model and the time period from September 2005 to April 2010 for all 8,621 portfolios in the sample.
Based on gross returns, individual investors have an average skill of approximately -7.5% per year. Net of costs and expenses the average skill is about 1 percentage point lower with about -8.5% per year.

How does this average net skill of -8.5% translate into returns? Assuming an average diversification and risk of an individual investor, the average investor will underperform the market by 8.5% per year net of expenses due to her insufficient skill. Would she take less risk or be better diversified than the average investor, this negative skill would not fully show and, hence, lead to a proportionally lower underperformance. In contrast, if she takes more risk or is less diversified, the investor will underperform proportionally more. Only by mere luck, might she be able to beat the market in a given year – or perform even worse.

IV. Conclusion

A few studies have focused on the measurement of individual investors’ investment performance and more extensive research has been conducted on biases and investment mistakes – such as the disposition effect, security selection bias and lacking ability of market timing. No study so far has focused on measuring whether it was driven by skill or mere luck. We contribute to the literature by analyzing whether individual investors exhibit negative skill or not by using a comprehensive set of 8,621 portfolios from a German online broker from September 2005 to April 2010.

By applying bootstrapping simulations introduced to the mutual fund literature by Fama and French (2010), we found significant negative skill – both in statistical and in economical terms.Analyzing gross returns, we found that about 89% of individual investors have negative skill. This implies that 89% of investors underperform the market when pursuing an active strategy – unless luck is on their side. The magnitude of the average skill of all investors is also very large with about -7.5% per year.

Taking expenses and trading fees into account, the figures are even worse: 91% of individual investors have negative skill which does not even suffice to cover their expenses and trading fees. As expenses and trading fees amount to about 1% per year, the average skill of all
individual investors amounts to -8.5% per year. It can thus be concluded that the large majority of individual investors do not have skill to outperform the market – and if they do, it is mere luck.

With this clear evidence of individual investors underperforming the market and the main puzzle solved, the key questions for further research becomes what biases in particular cause the negative skill and who underperforms? Moreover, the results are a clear case for passive strategies. Hence, banks, politicians and individual investors might want to reconsider investment strategies and policies to help investors improving the investment skill. In the light of these findings and Campbell’s (2006) call for financial economists to come up with solutions to the investment mistakes of individual investors becomes even more urgent.

V. Appendix

For robustness, we performed the bootstrapping simulations also using other indices than the CDAX. Specifically, we use the MSCI AC IMI as an international equity index and construct a German and international multi-asset index as suggested by Jacobs et al. (2010). The German multi-asset index consists of the CDAX (60%), the German bond index Rex (25%) and the Goldman Sachs Commodity Index GSCI (15%). In contrast, the international multi-asset index consists of the MSCI AC IMI (60%), the international bond index iBoxx in EUR (25%) and again the GSCI (15%). For the two international indices we also adjust the additional factors SMB, HML and MOM and use the returns for these factors as provided by French. Although these factors are US-based, not international, the US has the largest stake in the world market portfolio making the data from French a reasonable proxy.

The results are presented in Exhibit VIII to Exhibit XIII. Exhibit VIII contains net results for the MSCI AC IMI, Exhibit IX the gross results. Exhibit X and XI depict the net and gross results for the German multi-asset index, respectively. Finally, Exhibit XII and XIII show the net and gross results for the international multi-asset index.

---

In any case, the mean skill level improves by applying any of the three alternative benchmarks. The best fitting distribution for the German multi-asset index has a mean of -5% for net returns and -4% for gross returns compared to -8.5% and -7.5% estimated using the CDAX. The increase in the mean is even larger for the two international indices. In both cases, the best fitting distribution has a mean of -2% for gross and only -1% for net returns.

Yet, the share of unskilled investors is still remarkably high even with the improved mean. For net returns, the best fitting distributions suggest 80% unskilled investors using an international index and the 83% using the German multi-asset index. Compared to the estimate of 91% using the CDAX this is a somewhat smaller share, but still remarkably large.

Moreover, as discussed earlier, we still believe the CDAX to be the most accurate index choice and the improved skill by using an international index is rather due to the outperformance of the German index during the selected time period rather than true skill.
Exhibit VIII

Percentage of Estimated t-Statistics Below Actual t(α) for Net Returns Using the MSCI AC IMI

The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a t(α) below the actual value of t(α) at the respective percentiles (y-axis). Simulation results are presented for net returns with a skewed normal distribution with a prespecified mean µ ranging from µ=2% to µ=6%, a standard deviation σ ranging from σ=0% to σ=6% and a skewness γ of γ=1 and a kurtosis κ of κ=4. All results are for the 4-factor model using the MSCI AC IMI and the US-based factors of SMB, HML and MOM. The time period is from September 2005 to April 2010 and all 8,621 portfolios in the sample are included.
Exhibit IX

Percentage of Estimated t-Statistics Below Actual t(α) for Gross Returns Using the MSCI AC IMI

The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a t(α) below the actual value of t(α) at the respective percentiles (y-axis). Simulation results are presented for gross returns with a skewed normal distribution with a prespecified mean µ ranging from µ=2% to µ=-4%, a standard deviation σ ranging from σ=0% to σ=6% and a skewness γ of γ=1 and a kurtosis κ of κ=4. All results are for the 4-factor model using the MSCI AC IMI and the US-based factors of SMB, HML and MOM. The time period is from September 2005 to April 2010 and all 8,621 portfolios in the sample are included.
Exhibit X

Percentage of Estimated t-Statistics Below Actual t(α) for Net Returns Using an International Multi-Asset Benchmark

The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a t(α) below the actual value of t(α) at the respective percentiles (y-axis). Simulation results are presented for net returns with a skewed normal distribution with a prespecified mean µ ranging from µ=2% to µ=-6%, a standard deviation σ ranging from σ=0% to σ=6% and a skewness γ of γ=1 and a kurtosis κ of κ=4. All results are for the 4-factor model using an international multi-asset benchmark consisting of 60% MSCI AC IMI, 25% iBoxx EUR and 15% GSCI. Moreover, the US-based factors of SMB, HML and MOM are used. The time period is from September 2005 to April 2010 and all 8,621 portfolios in the sample are included.
Exhibit XI

Percentage of Estimated t-Statistics Below Actual t(α) for Gross Returns Using an International Multi-Asset Benchmark

The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a t(α) below the actual value of t(α) at the respective percentiles (y-axis). Simulation results are presented for gross returns with a skewed normal distribution with a prespecified mean µ ranging from µ=2% to µ=-4%, a standard deviation σ ranging from σ=0% to σ=6% and a skewness γ of γ=1 and a kurtosis κ of κ=4. All results are for the 4-factor model using an international multi-asset benchmark consisting of 60% MSCI AC IMI, 25% iBoxx EUR and 15% GSCI. Moreover, the US-based factors of SMB, HML and MOM are used. The time period is from September 2005 to April 2010 and all 8,621 portfolios in the sample are included.
Exhibit XII

Percentage of Estimated t-Statistics Below Actual t(α) for Net Returns Using a Domestic Multi-Asset Benchmark

The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a t(α) below the actual value of t(α) at the respective percentiles (y-axis). Simulation results are presented for net returns with a skewed normal distribution with a prespecified mean μ ranging from μ=−2% to μ=−8%, a standard deviation σ ranging from σ=2% to σ=10% and a skewness γ of γ=1 and a kurtosis κ of κ=4. All results are for the 4-factor model using a domestic multi-asset benchmark consisting of 60% CDAX, 25% REX and 15% GSCI. Moreover, the domestic values for the factors of SMB, HML and MOM are used. The time period is from September 2005 to April 2010 and all 8,621 portfolios in the sample are included.
Exhibit XIII

Percentage of Estimated t-Statistics Below Actual t(α) for Gross Returns Using a Domestic Multi-Asset Benchmark

The exhibit shows the percentage (x-axis) of estimated t-statistics drawn from 1,000 simulations that produced a t(α) below the actual value of t(α) at the respective percentiles (y-axis). Simulation results are presented for gross returns with a skewed normal distribution with a prespecified mean µ ranging from µ=0% to µ=−8%, a standard deviation σ ranging from σ=2% to σ=10% and a skewness γ of γ=1 and a kurtosis κ of κ=4. All results are for the 4-factor model using a domestic multi-asset benchmark consisting of 60% CDAX, 25% REX and 15% GSCI. Moreover, the domestic values for the factors of SMB, HML and MOM are used. The time period is from September 2005 to April 2010 and all 8,621 portfolios in the sample are included.
References


Lewis, Karen K., 1999, Trying to explain home bias in equities and consumption, Journal of Economic Literature 37, 571-608.


