The Shrinkage Adjusted Sharpe Ratio: 
An Improved Method for Mutual Fund Selection

Richard Roll
California Institute of Technology
With
Moshe Levy, Hebrew University
The Shrinkage Adjusted Sharpe Ratio: 
An Improved Method for Mutual Fund Selection

Moshe Levy* and Richard Roll**

Abstract

Mutual fund selection is a notoriously difficult task, because past performance is a poor predictor of future performance. We propose a fund performance measure that incorporates a simple idea: shrinkage, in the sense of Bayes-James-Stein, should be applied to gross return parameters, but not to fees, which are known. The proposed Shrinkage Adjusted Sharpe ratio (SAS) substantially improves the prediction of out-of-sample performance relative to existing methods. The best prediction is obtained when fees are weighed 5 times heavier than sample returns.

Keywords: Mutual funds, Sharpe ratio, shrinkage, management fees, investment performance, alpha.

JEL Code: G11.

*The Hebrew University, Jerusalem 91905, Israel. 972-2-588-1329. mslm@huji.ac.il.
**The California Institute of Technology, 1200 East California Blvd, Pasadena, CA 91125. 626-395-3890. rroll@caltech.edu.
1. Introduction

Delegated asset management plays a central role in financial markets. While delegated management accounted for less than 5% of U.S. equity ownership in the early 1950’s, this figure has risen to about 50% in recent years (Gârleanu and Pedersen 2021). In 2020, U.S. mutual funds managed $23.9 trillion, and 47.4% of households held funds.\(^1\) As about 50% of U.S. households do not participate in the stock market at all (Bogan 2008), this implies that the vast majority of those holding stocks do so, at least in part, via mutual funds. Thus, fund selection is a central task facing most investors.

This is a notoriously difficult task, because a fund’s past return parameters are poor predictors of its future return parameters. This is true both for alphas, as well as Sharpe ratios. While alpha may be a good measure of a fund manager’s ability to generate excess return relative to the fund’s systematic risk exposure,\(^2\) from the perspective of an investor who holds a single fund, it is the fund’s total risk that matters, and therefore the Sharpe ratio is the relevant performance measure.\(^3\) Panel A of Figure 1 depicts the relationship between in-sample and out-of-sample Sharpe ratios for all U.S. domestic equity funds (the data and empirical methodology are described in Section 3). The figure shows that the relationship between in-sample Sharpe ratios and out-of-sample Sharpe ratios is very weak. As Morningstar ratings are closely related

---

\(^1\) Investment Company Institute 2021 Factbook https://www.icifactbook.org/.


\(^3\) To illustrate this point, consider, for example, a fund that has the same standard deviation as the market portfolio, an expected return that is 1% lower than the market’s expected return, and returns that are uncorrelated with the market. This fund has a large positive CAPM alpha, because its CAPM beta is zero. Obviously, all investors should prefer the market portfolio over this fund, even though the market’s alpha is by definition zero. In the mean-variance framework, investors who hold a single fund should rank funds by their Sharpe ratio (Sharpe 1966, 1994). This is true even if the return distributions are not precisely normal - see Levy and Markowitz (1979) and Kroll, Levy and Markowitz (1984). The inadequacy of alpha for an investor holding a single fund (or a small number of funds) is discussed by Levy and Roll (2016). The appendix expands on this point. The median number of funds held by U.S. investors is 4, and 35% of investors who hold mutual funds hold either a single fund or two funds (Investment Company Institute 2021 Factbook). Thus, the Sharpe ratio seems as the most appropriate measure of performance from investors’ perspective.
to the sample Sharpe ratios (Sharpe 1998), it is not surprising that they too are not very good predictors of future performance. This is a very troubling picture: millions of households invest trillions of dollars based on the sample performance, which is not very indicative of future performance. Can this situation be improved? The present study is an attempt towards this goal.

Figure 1: The relationship between in-sample performance and the out-of-sample Sharpe ratio. All U.S. domestic equity funds are included in the analysis, and the sample period is December 1991 – September 2021 (see Section 3 for a detailed description of the data and empirical methodology). Funds are ranked according to their in-sample performance, and sorted into deciles. For each decile, the figure shows the average in-sample performance and the average out-of-sample Sharpe ratio (net of fees). In Panel A the standard Sharpe ratio (net of fees) is employed as the in-sample performance measure. In Panel B the SAS (eq. 6-7) is employed as the in-sample performance measure.
The following example conveys the main idea. Consider two funds: fund A has a sample average annual return of 11% and charges annual fees of 1%. Fund B has a sample average return of 12%, and charges fees of 2%. Assume for simplicity that the two funds have the same volatility. Which fund should the investor prefer? At first glance, it may seem that the two funds are equivalent, because they both yield a net average return of 10%. Note, however, that the sample average gross returns are noisy estimates, while the fees are known.\footnote{Fees may also change, but they are typically very stable over time. For example, the standard deviation in monthly fees is lower than 0.01\% for 88\% of U.S. domestic equity funds. It is lower than 0.02\% for 97\% of funds (for the 120-month sample period ending in September 2021). These standard deviations are approximately one tenth and one fifth of the average level of monthly fees, which is about 0.1\%. This variation in fees is very low compared to the variation of monthly returns, with a typical average return of 1\% and a standard deviation of about 4\%.} Intuitively, this suggests that fees should be weighted more heavily than the sample returns. In the extreme case in which the sample returns are completely uninformative, one should ignore them and rank funds only according to fees. In the more realistic and interesting case, where the sample return parameters do convey a noisy signal about the future returns, the question is how much weight one should assign to this signal relative to the fees. This is the main question addressed in this paper.

We suggest a generalization of the Sharpe ratio that manifests a simple idea: while “shrinkage”, in the sense of Bayes-James-Stein, should be applied to the sample gross return parameters, the fees are known, and should be taken as is. The suggested Shrinkage adjusted Sharpe ratio, or SAS, captures this idea, and formalizes the intuition of over-weighing fees relative to the sample returns. The standard sample Sharpe ratio and the shrinkage employed by Levy and Roll (2018) are obtained as special cases of the more general SAS.

We find that the sample SAS predicts future performance much better than the sample Sharpe ratio. This can be seen in Panel B of Figure 1: the relationship between the in-sample SAS and the out-of-sample Sharpe ratio is almost monotonic, with $R^2 = 0.934$ (compared with an $R^2$ of only 0.298 for the in-sample Sharpe ratio, panel A). Figure 1 depicts the relationship
between the sample performance measures and out-of-sample performance for the set of all funds. Investors naturally focus their attention on the funds with the highest in-sample performance, so the out-of-sample performance of these top funds is of central importance. If one invests in the top 10 funds with the highest in-sample Sharpe ratios (equally weighted) one obtains an out-of-sample monthly Sharpe ratio of 0.107 for the 1991-2021 sample period. This is much lower than the monthly Sharpe ratio of the S&P500 index over the same period, which is 0.148. In contrast, investing in the 10 funds with the highest in-sample SAS yields an out-of-sample Sharpe ratio of 0.169 (see Section 3 for more detail). Thus, the top sample SAS funds are not only much better than the funds with the highest sample Sharpe ratios, they even beat the market by about 1.1% on an annual risk-adjusted basis.5

2. The Shrinkage Adjusted Sharpe Ratio (SAS)

From the perspective of an investor selecting a mutual fund, performance should be measured based on returns net of fees. The key idea that the SAS aims to capture is that while one should imply shrinkage, in the sense of Bayes-James-Stein, to the funds’ sample gross mean returns, the fund fees are typically known, and therefore fees should be excluded from shrinkage. Thus, when estimating a fund’s expected net return, one should first apply shrinkage to the average gross return, and only then subtract the fees (rather than applying shrinkage to the net returns).

Consider an investor who observes \( T \) returns of fund \( i \), with a sample gross mean \( \bar{R}_i \). All returns are in excess of the risk-free rate. The investor forms his posterior belief regarding the expected gross return based on the sample mean and the investor’s prior. For normally

5 The S&P’s monthly standard deviation is 4.46% in our sample period. Thus, the difference of 0.169 – 0.148 = 0.021 in Sharpe ratios translates to a monthly risk-adjusted return difference of 0.021 · 4.46% = 0.094%, and to an annual risk-adjusted excess return of (approximately) 0.094 · 12 = 1.13%.
distributed returns, and a normal prior with mean \( \mu \) and standard deviation \( \sigma_\mu \), the investor’s posterior expected gross return for fund \( i \) is given by:

\[
E(\mu_i | R_i^{gross}) = \frac{T \sigma_i^2 R_i^{gross}}{\sigma_i^2 (T + 1) \sigma_i^2 + 1} + \frac{1}{\sigma_i^2} \mu \tag{1}
\]

where \( \sigma_i \) is the standard deviation of fund \( i \)'s returns (see, for example, Lee 2012, p. 46). This can be written as:

\[
E(\mu_i | R_i^{gross}) = \gamma_i R_i^{gross} + (1 - \gamma_i)\mu, \tag{2}
\]

where \( \gamma_i \), which can be viewed as a fund-specific shrinkage factor, is given by:

\[
\gamma_i = \frac{T \sigma_i^2}{\sigma_i^2 + 1} = \frac{1}{1 + \frac{1}{T \sigma_i^2}} \tag{3}
\]

\( \gamma_i \) determines how much the sample average return is shrunk towards the prior. \( \gamma_i = 1 \) implies no shrinkage at all, while on the other extreme \( \gamma_i = 0 \) implies that the sample returns are completely ignored. The derivation of eq.(1) assumes that the return distributions are normal, and, perhaps more importantly, that fund returns are drawn from a distribution that is stable over time. This is the setup employed by Levy and Roll (2018). In practice, these assumptions do not hold, and a better estimation of the fund’s future expected return may be obtained with a higher (or lower) shrinkage intensity (i.e. with a lower or higher value of \( \gamma_i \)). This is an issue investigated empirically in the next section. In order to allow for “extra shrinkage” relative to equations (2-3), we introduce a market-wide shrinkage parameter \( \gamma_\mu \) defined on the interval \([0,1]\), and we generalize eq.(3) to:

\[
\gamma_i = \frac{1}{1 + \frac{\sigma_i^2}{T \sigma_\mu^2} \left( \frac{1}{\gamma_\mu} - 1 \right)} \tag{4}
\]
For $\gamma_\mu = \frac{1}{2}$ the standard eq.(3) is obtained as a special case. Values of $\gamma_\mu$ lower than $\frac{1}{2}$ imply more shrinkage relative to eq.(3) (i.e. a lower $\gamma_i$), and values higher than $\frac{1}{2}$ imply less shrinkage. In the extremes, $\gamma_\mu = 0$ implies $\gamma_i = 0$, i.e. this is the case of maximal shrinkage where the sample average return is completely ignored; $\gamma_\mu = 1$ implies $\gamma_i = 1$, i.e. the case of no shrinkage at all. In our empirical analysis we will investigate the value of $\gamma_\mu$ that yields the best out-of-sample performance prediction.

Our goal is an estimate of the fund’s Sharpe ratio, i.e. the expected net return (net of fees and in excess of the risk-free rate), divided by the standard deviation of returns. In their derivations, Levy and Roll (2018) assume that the standard deviation is known. In practice, of course, the standard deviation is also measured with error, and shrinkage in the sense of James-Stein (1961) may provide a better estimate of the out-of-sample standard deviation than the sample standard deviation. Thus, we will employ the following estimate of the fund’s standard deviation:

$$\sigma_i = \gamma_\sigma \hat{\sigma}_i + (1 - \gamma_\sigma) \bar{\sigma}, \quad (5)$$

where $\hat{\sigma}_i$ is the sample standard deviation of fund $i$, $\bar{\sigma}$ is the average standard deviation across all funds, and $\gamma_\sigma$ determines the shrinkage intensity for the standard deviations.

The Shrinkage Adjusted Sharpe ratio (SAS) of fund $i$ is thus given by:

$$SAS_i = \frac{\gamma_i \bar{R}^{\text{gross}}_i + (1 - \gamma_i) \mu - fee_i}{\gamma_\sigma \hat{\sigma}_i + (1 - \gamma_\sigma) \bar{\sigma}} \quad (6)$$

where $\bar{R}^{\text{gross}}_i$ is the fund’s average gross return (in excess of the risk-free rate), and $\gamma_i$ is given by eq.(4), i.e. it is determined by the parameter $\gamma_\mu$. Thus, the SAS formula depends on the two market-wide (i.e. not fund-specific) shrinkage intensities, $\gamma_\mu$ and $\gamma_\sigma$. The standard sample Sharpe ratio is obtained as a special case with $\gamma_\mu = 1$ and $\gamma_\sigma = 1$. The analysis of Levy and
Roll (2018) can be viewed as a special case with $\gamma_\mu = \frac{1}{2}$ (i.e. eq.(3)) and $\gamma_\sigma = 1$. As the return distributions may change over time, by a process that is unknown, and as the distributions are not normal, the optimal values of $\gamma_\mu$ and $\gamma_\sigma$ cannot be derived analytically. In the next section we empirically examine which combination of $\gamma_\mu$ and $\gamma_\sigma$ provides the best fund ranking, in the sense of providing the highest prediction of future performance.

3. Data and Results

We employ data from the CRSP survivorship-bias-free Mutual Fund Database. Monthly fee data are mostly unavailable before December 1991, thus, we take the sample period of December 1991 – September 2021 (358 months). Monthly risk-free rates are taken from the CRSP Monthly Risk-free Series.

Our main analysis is conducted on all US domestic equity funds (all funds with CRSP style codes starting with “ED”). In the next section we also separately examine all Foreign Equity funds (all funds with CRSP style codes starting with “EF”), and corporate and municipal fixed-income funds (all funds with CRSP style codes starting with “IC” and “IU”).

We evaluate performance measures by employing two approaches. In the first approach we examine the out-of-sample performance of the top funds that are ranked highest by the measure. This is the most relevant criterion from the perspective of an investor employing the performance measure, because these are the funds that she will presumably invest in. In the second approach we evaluate performance by looking at the relationship between the ranking and out-of-sample performance for all funds, as in the decile analysis of Figure 1. This provides a wider perspective, and has the statistical advantage of employing data for all funds, not just the top funds.

At month $t$ we employ the preceding $T$ months, $t - T$ to $t - 1$, to calculate the in-sample SAS (eq. 6) for each fund. In our main analysis we employ an estimation window of $T=60$. 
months (in the next section we also consider a different value of $T$). $\bar{R}_i$ and $\bar{\tilde{R}}_i$ are the sample average return in excess of the risk-free rate, gross of fees, and the sample standard deviation of fund $i$, respectively. $\sigma_{\mu}$, the cross-sectional standard deviation of mean returns across all funds (eq. 3-4), is evaluated by its sample value. Similarly, $\sigma$ (eq. 5) is taken as the average of sample standard deviations across all funds. $\mu$, the mean prior expected return (eq.1), is taken as the average return of the market portfolio in excess of the risk-free rate over the entire 1991-2021 sample period, which is 0.77%.

The SAS depends on the intensity of shrinkage employed, i.e. on the parameters $\gamma_{\mu}$ and $\gamma_{\sigma}$. As discussed above, the optimal values of these parameters depend on the stability of the return generating process, which is unknown. Our goal is to find the combination of $\gamma_{\mu}$ and $\gamma_{\sigma}$ that provides the most informative fund ranking. In the first approach (focus on the top funds) this means finding the values that yield the best out-of-sample performance of the top ranked funds. For each combination of $\gamma_{\mu}$ and $\gamma_{\sigma}$ we evaluate the out-of-sample Sharpe ratio (net of fees) of the strategy of holding the top 10 funds with the highest in-sample SAS (with these values of $\gamma_{\mu}$ and $\gamma_{\sigma}$). These 10 funds are held equally-weighted for the next year. At the end of the year the process is repeated: we again calculate the SAS for all funds, and choose the top 10 funds according to the updated SAS values. This is repeated over the entire 1991-2021 sample period. Figure 2 shows the out-of-sample monthly Sharpe ratio, calculated over the entire sample period, for each $(\gamma_{\mu}, \gamma_{\sigma})$ combination ($\gamma_{\mu}$ and $\gamma_{\sigma}$ are varied at increments of 0.01, thus a total of $101^2 = 10,201$ combinations are evaluated). The colors (shades) in the map indicate the out-of-sample Sharpe ratio for each $(\gamma_{\mu}, \gamma_{\sigma})$ combination. Brighter colors (lighter shades) correspond to better performance, as indicated by the legend (vertical color bar) to the right of the figure.

Note that $(\gamma_{\mu} = 1, \gamma_{\sigma} = 1)$, the top right corner, implies no shrinkage at all. This corresponds to the widespread practice of employing the sample Sharpe ratios. This strategy
yields an out-of-sample Sharpe ratio of 0.109. Levy and Roll (2018) employ shrinkage of the average return, assuming that the standard deviation is known and that the return distribution is stable over time. This corresponds to the case \((\gamma_\mu = \frac{1}{2}, \gamma_\sigma = 1)\), top center in the figure, and yields a slightly higher out-of-sample Sharpe ratio of 0.115. The combination that yields the highest Sharpe ratio is:

\[
\gamma_\mu = 0.15, \quad \gamma_\sigma = 0.04,
\]

(7)
denoted by the star in the figure. This combination yields an out-of-sample Sharpe ratio of 0.169. This is not only much higher compared to the cases of no-shrinkage or Levy-Roll shrinkage, but it is even higher than the Sharpe ratio of the market portfolio over the corresponding period, which is 0.148. This corresponds to an annual risk-adjusted excess return of about 1.1\% (see footnote 5), and implies not only that some funds are able to beat the market, but also that these funds can be identified \textit{ex-ante} and exploited by employing SAS.

\[\text{Figure 2: The out-of-sample Sharpe ratio of the top 10 funds with the highest SAS, as a function of the shrinkage intensities, } \gamma_\mu \text{ and } \gamma_\sigma. \text{ Lighter colors (shades) indicate better performance, as shown by the vertical color-bar legend to the right of the figure. The best out-of-sample performance is obtained with the combination } \gamma_\mu = 0.15 \text{ and } \gamma_\sigma = 0.04, \text{ denoted by the star. This shrinkage yields an out-of-sample Sharpe ratio of 0.169, which is higher not only of that obtained by ranking according to the standard sample Sharpe (0.109, top right corner), and Levy-Roll shrinkage (0.115, top center), but it is also higher than the Sharpe ratio of the market portfolio over the same period, which is 0.148.}\]
Of course, a key question is the robustness of the values in eq.(7) to different asset classes, sample periods, and lengths of the estimation window ($T$). This issue is examined in the next section, where we find that the parameter values in (7) do indeed provide a good prediction of out-of-sample performance in a wide range of settings. They work well not only when considering the top funds, but also in the second approach, where the performance of all funds is considered: panel B of Figure 1, which describes the relationship between in-sample SAS and out-of-sample Sharpe ratios for all funds, employs the parameter values in eq.(7).

How much overweighing of fees relative to the sample returns does $\gamma_\mu = 0.15$ imply? For a typical fund with $\sigma_i = 4\%$, $\gamma_\mu = 0.15$ implies $\gamma_i = 0.2$.\(^6\) This means that the weight of fees in the SAS is 5 times larger than that of the sample return (see eq.(6)). In other words, for two funds with the same volatility, a 5 basis-point difference in the sample average returns is required to compensate for a 1 basis-point difference in fees. Going back to the example in the introduction, $\gamma_i = 0.2$ implies that Fund B should be preferred to Fund A only if its sample gross return is 16% or larger.\(^7\) This implies a very substantial overweighing of fees, and reflects the large errors involved in estimating mean returns.

Given that the estimation problem is typically more severe for mean returns than it is for standard deviations, it may seem surprising at first that the optimal out-of-sample performance is obtained with such a small value of $\gamma_\sigma$. However, note that $\gamma_\sigma$ determines not only the shrinkage of the sample $\tilde{\sigma}$ towards the cross-sectional average, $\bar{\sigma}$, but also the importance of the sample standard deviation relative to the fees. To illustrate this point, consider the simplified case of $\gamma_\mu = 0$, where the SAS becomes: $R_i^\text{gross-fee} = \frac{\gamma_i \sigma_i + (1-\gamma_i)\sigma}{\bar{\sigma}}$. The higher the value of $\gamma_\sigma$, the larger the variation in the denominator, and the more dominant the role of $\sigma_i$ relative to

\(^6\) See eq.(4). $\sigma_\mu$, the standard deviation of average returns across funds is empirically about 0.6% (this value changes somewhat from one month to another). Thus, for $T=60$ we have: $\gamma_i = \frac{1}{1 + \frac{4.4^2}{60 \cdot 0.6^2} \left(1 - \frac{1}{0.15^2} \right)} \cong 0.2$.

\(^7\) $0.2 \cdot 11 - 1 = 0.2 \cdot 16 - 2$. 

10
fee. The fact that the optimal performance is obtained with very low values of \( \gamma_\sigma \) indicates that the role of the fees in predicting out-of-sample performance is much more pronounced than the role of the sample standard deviations.  

To get a better idea about the difference in rankings between SAS and the standard Sharpe ratio, let us look at the top 10 funds identified by these two measures. Table I provides these top funds, as of September 2016, with sample parameters based on the preceding 60 months. The table reports the in-sample average gross return and standard deviation for each fund, its fees, as well as its out-of-sample Sharpe ratio (net of fees) in the period October 2016 – September 2021. The most obvious difference between these two groups of funds is the fees: the average (monthly) fees are 0.068\% for the top Sharpe funds, compared to only 0.028\% for the top SAS funds. This follows directly from the over-weighting of fees implied by SAS. Because of the very small weight attached to the sample variance, SAS funds have higher sample volatilities than the Sharpe funds. As higher volatility funds tend to also have higher average returns, this relaxed constraint on the volatility also allows SAS funds to have higher sample returns. The average out-of-sample Sharpe ratio is 39.6\% higher for the SAS funds: 0.261 compared to 0.187. Moreover, if funds are ranked by their out-of-sample Sharpe ratios, each SAS fund has a higher Sharpe ratio than its counterpart Sharpe fund (this is analogous to the First-order Stochastic Dominance of one distribution over another, see Hadar and Russell 1969, and Hanoch and Levy 1969).

---

8 This is consistent with previous research suggesting that fees may be the best predictors of future performance (Carhart 1997, Kinnel 2010). This does not imply, however, that the sample return parameters should be ignored - as evident from the fact that the maximal out-of-sample performance is obtained with non-zero values of \( \gamma_\mu \) and \( \gamma_\sigma \).

9 Note that these values are different than those reported in Figure 2, which employs the entire sample period, compared to the “snapshot” analysis reported here.
Table I
The top 10 funds by in-sample Sharpe ratio (panel A) and by in-sample SAS (panel B), as of September 2016. Return parameters are monthly, in %, and are estimated from the previous \( T=60 \) months. The out-of-sample Sharpe ratio is net of fees, calculated over the period October 2017 – September 2021. Four funds appear in both groups.

### A. Top Funds by Sample Sharpe Ratio

<table>
<thead>
<tr>
<th>rank</th>
<th>Fund i.d.</th>
<th>Fund Name</th>
<th>( \bar{R}_i^{\text{gross}} )</th>
<th>( \hat{\sigma}_i )</th>
<th>( fee_i )</th>
<th>Out-of-sample Sharpe ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49277</td>
<td>Fidelity Advisor Real Estate Class I</td>
<td>0.98</td>
<td>1.74</td>
<td>0.064</td>
<td>0.154</td>
</tr>
<tr>
<td>2</td>
<td>12006</td>
<td>Fidelity Real Estate Income Fund</td>
<td>0.97</td>
<td>1.74</td>
<td>0.068</td>
<td>0.154</td>
</tr>
<tr>
<td>3</td>
<td>49275</td>
<td>Fidelity Advisor Real Estate Class A</td>
<td>0.97</td>
<td>1.73</td>
<td>0.086</td>
<td>0.149</td>
</tr>
<tr>
<td>4</td>
<td>49278</td>
<td>Fidelity Advisor Real Estate Class T</td>
<td>0.97</td>
<td>1.73</td>
<td>0.089</td>
<td>0.148</td>
</tr>
<tr>
<td>5</td>
<td>24710</td>
<td>PowerShares High Yield Equity</td>
<td>1.44</td>
<td>2.93</td>
<td>0.045</td>
<td>0.162</td>
</tr>
<tr>
<td>6</td>
<td>49276</td>
<td>Fidelity Advisor Real Estate Class C</td>
<td>0.96</td>
<td>1.72</td>
<td>0.148</td>
<td>0.133</td>
</tr>
<tr>
<td>7</td>
<td>43925</td>
<td>US Managed Volatility</td>
<td>1.22</td>
<td>2.60</td>
<td>0.020</td>
<td>0.218</td>
</tr>
<tr>
<td>8</td>
<td>36082</td>
<td>PIMCO StocksPLUS Long Duration</td>
<td>1.83</td>
<td>3.91</td>
<td>0.049</td>
<td>0.340</td>
</tr>
<tr>
<td>9</td>
<td>27507</td>
<td>SEI Institutional Managed Trust</td>
<td>1.24</td>
<td>2.58</td>
<td>0.085</td>
<td>0.205</td>
</tr>
<tr>
<td>10</td>
<td>51184</td>
<td>PowerShares KBW</td>
<td>1.58</td>
<td>3.48</td>
<td>0.029</td>
<td>0.210</td>
</tr>
</tbody>
</table>

**Average:** 1.22 2.42 0.068 0.187

### B. Top Funds by Sample SAS

<table>
<thead>
<tr>
<th>rank</th>
<th>Fund i.d.</th>
<th>Fund Name</th>
<th>( \bar{R}_i^{\text{gross}} )</th>
<th>( \hat{\sigma}_i )</th>
<th>( fee_i )</th>
<th>Out-of-sample Sharpe ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31366</td>
<td>Vanguard Health Care Admiral Class</td>
<td>1.54</td>
<td>3.61</td>
<td>0.008</td>
<td>0.304</td>
</tr>
<tr>
<td>2</td>
<td>31357</td>
<td>Vanguard Health Care ETF Class Shares</td>
<td>1.54</td>
<td>3.61</td>
<td>0.008</td>
<td>0.304</td>
</tr>
<tr>
<td>3</td>
<td>36082</td>
<td>PIMCO StocksPLUS Long Duration</td>
<td>1.83</td>
<td>3.91</td>
<td>0.049</td>
<td>0.340</td>
</tr>
<tr>
<td>4</td>
<td>51184</td>
<td>PowerShares KBW</td>
<td>1.58</td>
<td>3.48</td>
<td>0.029</td>
<td>0.210</td>
</tr>
<tr>
<td>5</td>
<td>24710</td>
<td>PowerShares High Yield Equity</td>
<td>1.44</td>
<td>2.93</td>
<td>0.045</td>
<td>0.162</td>
</tr>
<tr>
<td>6</td>
<td>27428</td>
<td>PowerShares S&amp;P 500 Quality</td>
<td>1.58</td>
<td>3.67</td>
<td>0.033</td>
<td>0.279</td>
</tr>
<tr>
<td>7</td>
<td>16463</td>
<td>iShares US Healthcare</td>
<td>1.54</td>
<td>3.53</td>
<td>0.037</td>
<td>0.299</td>
</tr>
<tr>
<td>8</td>
<td>49219</td>
<td>PowerShares S&amp;P SmallCap</td>
<td>1.47</td>
<td>3.64</td>
<td>0.024</td>
<td>0.172</td>
</tr>
<tr>
<td>9</td>
<td>43925</td>
<td>US Managed Volatility Fund; Class A</td>
<td>1.22</td>
<td>2.60</td>
<td>0.020</td>
<td>0.218</td>
</tr>
<tr>
<td>10</td>
<td>24748</td>
<td>PowerShares S&amp;P 500 Quality</td>
<td>1.30</td>
<td>2.95</td>
<td>0.024</td>
<td>0.318</td>
</tr>
</tbody>
</table>

**Average:** 1.51 3.39 0.028 0.261
4. **Robustness**

We examine the robustness of the out-of-sample performance of SAS with the shrinkage parameters in eq.(7) in several ways. First, we look at different fund classes. We separately examine foreign equity funds, and fixed-income funds. Second, we split the sample period and look at the performance of domestic equity funds in the two sub-periods: 1991-2006, and 2006-2021. Finally, we examine robustness relative to the length of the estimation window, $T$.

**Different Asset Classes**

Our main analysis is conducted on U.S. domestic equity funds, and the shrinkage parameters in eq.(7) are optimal for this class of mutual funds. How well do these parameters perform with other mutual fund classes? Figure 3 shows the out-of-sample Sharpe ratio of the top 10 SAS funds for different $(\gamma_\mu, \gamma_\sigma)$ combinations, for foreign equity funds (“EF” CRSP style codes, panel A), and fixed income funds (corporate and municipal bonds, CRSP style codes “IC” and “IU”, panel B). The star depicts the parameter values in eq.(7), i.e. those that were found to be optimal for domestic equity funds. For each asset class, the highest out-of-sample performance is obtained at different $(\gamma_\mu, \gamma_\sigma)$ combinations, but the parameters in eq.(7) yield good out-of-sample performance for both foreign equity funds and fixed-income funds. In both cases, SAS with these parameters yields superior performance relative to both the standard Sharpe ratio and the Levy-Roll Shrinkage. In the case of foreign equity funds, SAS yields an out-of-sample Sharpe ratio of 0.154, compared with 0.137 of the standard Sharpe ranking (top right corner), and 0.136 of Levy-Roll shrinkage (top center). For fixed income funds SAS yields an out-of-sample Sharpe ratio of 0.151, compared with 0.048 of the standard Sharpe ranking, and 0.066 of Levy-Roll shrinkage.
Figure 3: The out-of-sample Sharpe ratio of the top 10 funds with highest SAS, as a function of the shrinkage intensities, $\gamma_\mu$ and $\gamma_\sigma$. Panel A: foreign equity funds. Panel B: corporate and municipal bond funds. The star represents the shrinkage which was found to be optimal for domestic equity funds: $\gamma_\mu = 0.15, \gamma_\sigma = 0.04$. In both cases, this shrinkage is close to optimal, and yields out-of-sample performance superior to both the standard sample Sharpe ratio (top right corner: $\gamma_\mu = 1, \gamma_\sigma = 1$) and to Levy-Roll shrinkage (top center: $\gamma_\mu = \frac{1}{2}, \gamma_\sigma = 1$).
Figure 4 reports the relationship between in-sample performance and out-of-sample performance for all funds in each asset class (rather than just the top 10 funds, as in Figure 3). In both cases, in-sample SAS provides higher explanatory power than the standard Sharpe ratio. The R²’s are lower relative to the case of domestic equity funds (panel B of Figure 1); this may be due to the larger number of domestic equity funds.

![Figure 4](image-url)

**Figure 4**: The relationship between in-sample performance and the out-of-sample Sharpe ratio for all funds in each asset class. All funds are sorted into deciles according to the in-sample performance. Panel A: foreign equity funds, the standard sample Sharpe ratio is employed as the in-sample performance measure. Panel B: foreign equity funds, SAS (with parameters in eq.(7)) is employed as the in-sample performance measure. Panel C: corporate and municipal bond funds, the standard Sharpe ratio is employed as the in-sample performance measure. Panel D: corporate and municipal bond funds, SAS is employed as the in-sample performance measure.
Different Sample Periods

To examine the consistency of the results over time, we divide the sample period into two subperiods, and analyze each sub-period separately. The first subperiod is December 1991- October 2006, and the second November 2006 – September 2021. Figures 5 and 6 report the results – Figure 5 for the top 10 domestic equity funds, and Figure 6 for all domestic equity funds. In both subperiods the shrinkage parameters in eq.(7) (denoted by the star) perform well, and are better than both the standard Sharpe ratio (no shrinkage), and the Levy-Roll shrinkage. In the first subperiod the top 10 SAS funds yield an out-of-sample Sharpe ratio of 0.198, compared to 0.137 for the top 10 in-sample Sharpe funds, and 0.132 for Levy-Roll shrinkage. In the second subperiod the top SAS funds yield an out-of-sample Sharpe ratio of 0.142, compared to 0.127 for the top in-sample Sharpe funds, and 0.115 for Levy-Roll shrinkage.
Figure 5: The out-of-sample Sharpe ratio of the top 10 funds with highest sample SAS, as a function of the shrinkage intensities, $\gamma_\mu$ and $\gamma_\sigma$, for two subperiods. Panel A: December 1991- October 2006. Panel B: November 2006 – September 2021. The star represents the shrinkage parameters in eq.(7): $\gamma_\mu = 0.15$, $\gamma_\sigma = 0.04$. In both sub-periods, this shrinkage is close to optimal, and yields out-of-sample performance superior to both the standard sample Sharpe ratio and to Levy-Roll shrinkage.
Figure 6: The relationship between in-sample performance and the out-of-sample Sharpe ratio for all domestic equity funds, in two sub-periods. Funds are sorted into deciles according to in-sample SAS with the parameters in eq.(7).
Shorter Estimation Window

The formula for $\gamma_i$ takes the number of observations, $T$, into account, see eq. (4). Thus, we do not expect the optimal $\gamma_{\mu}$ to be very sensitive to the choice of the length of the estimation window, $T$. The optimal $\gamma_{\sigma}$ could potentially depend on $T$, but its value is very small anyway ($\gamma_{\sigma} = 0.04$), implying that the sample standard deviations are almost ignored. In the preceding analysis we employ $T=60$ months (5 years), which is standard practice in many applications. Another popular alternative is the choice of $T=36$ months, i.e. 3 years. Figure 7 shows the results for this case. While the optimal out-of-sample performance is obtained at a slightly lower value of $\gamma_{\mu}$ ($\gamma_{\mu} = 0.1, \gamma_{\sigma} = 0.05$), these parameter values are very close to those in eq.(7). The parameters in eq.(7), denoted in the figure by the star, yield an out-of-sample Sharpe ratio of 0.197, which is almost maximal, and it is much higher than the one obtained with the standard Sharpe ratio (0.109), or the Levy-Roll (2018) shrinkage (0.115).

![Figure 7: The out-of-sample Sharpe ratio of the top 10 domestic equity funds with highest sample SAS, as a function of the shrinkage intensities, $\gamma_{\mu}$ and $\gamma_{\sigma}$, when the estimation window length is $T=36$ months. The star represents the optimal parameters obtained with $T=60$ months (eq.(7): $\gamma_{\mu} = 0.15, \gamma_{\sigma} = 0.04$).]
5. Conclusions

There are many applications of statistical shrinkage in finance. These include, among others, shrinkage of the return parameters in portfolio optimization (Jorion 1985, 1986, DeMiguel, Martin-Utrera and Nogales 2013, Ledoit and Wolf 2004, 2017, 2020, and Kircher and Rösch 2021), shrinkage of betas for cost of capital estimation (Jorion 1991, Genton and Ronchetti 2008, Levi and Welch 2017), and shrinkage in the context of model uncertainty (Wang 2005, Garlappi, Uppal and Wang 2007). In this study we apply the concept of shrinkage in the context of mutual fund selection. The key point in this context is that while the sample gross return parameters should be “shrunk”, the fees are known, and should therefore be taken as is. We propose the Shrinkage Adjusted Sharpe ratio (SAS) to capture this idea.

If returns are drawn from a stationary distribution, the optimal shrinkage intensity can be derived analytically. However, in practice, the return distributions may change over time, in an unknown fashion, and thus, the optimal shrinkage intensity cannot be derived analytically. We analyze this issue empirically, and find that the optimal shrinkage intensity is quite robust to the mutual fund asset class, the sample period, and the length of the window employed for estimating the return parameters. This is true both when the out-of-sample performance is evaluated for the top funds with the highest in-sample rankings, as well as when performance is evaluated for all funds. The optimal shrinkage we find implies that the weight of fees is about 5 times larger than the weight of the sample average return. This means that a difference of 1 basis point in fees is justified only if it is accompanied by a 5 basis point difference in the sample returns.

The out-of-sample performance of the top SAS funds is not only better than that of the top funds ranked by the standard sample Sharpe ratio or by the Levy and Roll (2018) measure, it is also substantially better than that of the market portfolio, by 1.1% on an annual risk-adjusted
basis. Thus, some funds are able to consistently beat the market, and SAS allows us to identify these funds *ex-ante*. 
References


Appendix: Sharpe versus Alpha in Mutual Fund Selection

Alpha is popular as a measure of the performance of fund managers. Its appeal is that it provides a measure of excess return after controlling for various systematic risk factors. While this may be a good measure of a fund manager’s ability to generate excess return relative to systematic risk-exposure, it is not a very relevant measure for the investor selecting a single mutual fund, because the investor cares about the fund’s variance, including its non-systematic risk component, as illustrated by the example in footnote 3. In other words, fund alphas are not well aligned with the expected utility of an investor holding the fund. Panel B of Figure A1 illustrates this point.

Under the assumption of normal return distributions and the existence of a risk-free asset, expected utility maximization of any risk-averse investor is perfectly aligned with the maximization of the Sharpe ratio (Sharpe 1966, 1994, Hanoch and Levy 1969). This result has subsequently been extended to all investors with non-decreasing preferences (Levy and Levy 2004, Levy, De Giorgi and Hens 2012), including, for example, investors with Prospect Theory preferences and investors with various aspiration levels. Levy and Markowitz (1979) and Kroll, Levy and Markowitz (1984) show that MV optimization serves as an excellent approximation for direct expected utility maximization even if return distributions are not precisely normal.

Figure A1 shows the relationship between alpha, Sharpe, and the expected utility of an investor with log utility. All U.S. domestic equity mutual funds are employed, and return distributions are taken as the 60 monthly returns in the 60-month period of October 2016 – September 2021. For each mutual fund we calculate its 4-factor alpha, its Sharpe ratio, and the expected utility it yields under the optimal asset allocation (which is found numerically). The top panel shows that there is an almost monotonic relationship between the Sharpe ratio and the expected utility, with a Spearman rank correlation of 0.987. Thus, even though the return distributions are not normal, the Sharpe ratio is very closely aligned with the investor’s
expected utility. The bottom panel shows the relationship between the fund’s Fama-French-Carhart 4-factor alpha and expected utility. In this case the Spearman correlation is only 0.391. Thus, while the Sharpe ratio is almost perfectly aligned with expected utility maximization, choosing the fund with the highest alpha implies a below-average expected utility. Similar results are obtained for different utility functions, and for Fama-French (2015) 5-factor alphas.

**Figure A1.** The top panel shows the relationship between a fund’s Sharpe ratio and the expected utility of an investor with log-utility who holds the fund (and chooses the optimal asset allocation between the fund and the risk-free asset). All U.S. domestic equity funds are shown, and the return distributions are taken as the sample distributions in the 60-month period October 2016 – September 2021. The relationship between the Sharpe ratio and expected utility is almost monotonic, with a Spearman correlation of 0.987. This implies that although the return distributions are not normal, the Sharpe ratio is closely aligned with expected utility. The bottom panel shows the relationship between a fund’s 4-factor alpha and the expected utility (again, the investor is assumed to hold the optimal asset allocation). Alphas are not aligned well with expected utility, with a Spearman correlation of only 0.391. Similar results hold for other utility functions and for Fama-French (2015) 5-factor alphas. Levy (2021) shows similar results for the subset of active mutual funds, and a different sample period.