THE SHARPE RATIO EFFICIENT FRONTIER

David H. Bailey
Complex Systems Group Leader - Lawrence Berkeley National Laboratory
dhbailey@lbl.gov

Marcos M. López de Prado
Head of Global Quantitative Research - Tudor Investment Corp.
and
Research Affiliate - Lawrence Berkeley National Laboratory
lopezdeprado@lbl.gov

First version: May 2008
This version: July 2012

Forthcoming, Journal of Risk
http://www.risk.net/type/journal/source/journal-of-risk

We thank the editor of the Journal of Risk and two anonymous referees for helpful comments. We are grateful to Tudor Investment Corporation, José Blanco (UBS), Sid Browne (Guggenheim Partners), Peter Carr (Morgan Stanley, NYU), David Easley (Cornell University), Steve Evans (Tudor Investment Corp.), Laurent Favre (Alternative Soft), Matthew Foreman (University of California, Irvine), Ross Garon (S.A.C. Capital Advisors), Robert Jarrow (Cornell University), David Leinweber (Lawrence Berkeley National Laboratory), Elmar Mertens (Federal Reserve Board), Attilio Meucci (Kepos Capital, NYU), Maureen O'Hara (Cornell University), Eva del Pozo (Complutense University), Riccardo Rebonato (PIMCO, University of Oxford) and Luis Viceira (HBS).

Supported in part by the Director, Office of Computational and Technology Research, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy, under contract number DE-AC02-05CH11231.

THE SHARPE RATIO EFFICIENT FRONTIER

ABSTRACT

We evaluate the probability that an estimated Sharpe ratio exceeds a given threshold in presence of non-Normal returns. We show that this new uncertainty-adjusted investment skill metric (called *Probabilistic Sharpe ratio*, or PSR) has a number of important applications: First, it allows us to establish the *track record length needed* for rejecting the hypothesis that a measured Sharpe ratio is below a certain threshold with a given confidence level. Second, it models the trade-off between track record length and undesirable statistical features (e.g., negative skewness with positive excess kurtosis). Third, it explains why track records with those undesirable traits would benefit from reporting performance with the highest sampling frequency such that the IID assumption is not violated. Fourth, it permits the computation of what we call the *Sharpe ratio Efficient Frontier* (SEF), which lets us optimize a portfolio under non-Normal, leveraged returns while incorporating the uncertainty derived from track record length. Results can be validated using the Python code in the Appendix.

Keywords: Sharpe ratio, Efficient Frontier, IID, Normal distribution, Skewness, Excess Kurtosis, track record.

JEL Classifications: C02, G11, G14, D53.

1. INTRODUCTION

Roy (1952) was the first to suggest a risk-reward ratio to evaluate a strategy's performance. Sharpe (1966) applied Roy's ideas to Markowitz's mean-variance framework, in what has become one of the best known performance evaluation metrics. López de Prado and Peijan (2004) showed that the implied assumptions (namely, that returns are independent and identically distributed (IID) Normal) may hide substantial drawdown risks, especially in the case of hedge fund strategies.

Renowned academics (Sharpe among them[1]) have attempted to persuade the investment community against using the Sharpe ratio in breach of its underlying assumptions. Notwithstanding its many deficiencies, Sharpe ratio has become the 'gold standard' of performance evaluation. Sharpe ratios are greatly affected by some of the statistical traits inherent to hedge fund strategies in general (and high frequency strategies in particular), like non-normality and reduced granularity (due to returns aggregation). As a result, Sharpe ratios from these strategies tend to be "inflated". Ingersoll, Spiegel, Goetzmann and Welch (2007) explain that sampling returns more frequently reduces the inflationary effect that some manipulation tactics have on the Sharpe ratio.

We accept the futility of restating Sharpe ratio's deficiencies to investors. Instead, a first goal of this paper is to introduce a new measure called *Probabilistic Sharpe Ratio* (PSR), which corrects those inflationary effects. This uncertainty-adjusted Sharpe ratio demands a longer track record length and/or sampling frequency when the statistical characteristics of the returns distribution would otherwise inflate the Sharpe ratio. That leads us to our second goal, which is to show that *Sharpe ratio can still evidence skill if we learn to require the proper length for a track record*. We formally define the concept of *Minimum Track Record Length* (MinTRL) needed for rejecting the null hypothesis of 'skill beyond a given threshold' with a given degree of confidence. The question of how long should a track record be in order to evidence skill is particularly relevant in the context of alternative investments, due to their characteristic non-Normal returns. Nevertheless, we will discuss the topic of "track record length" from a general perspective, making our results applicable to any kind of strategy or investment.

A third goal of this paper is to introduce the concept of *Sharpe ratio Efficient Frontier* (SEF), which permits the selection of optimal portfolios under non-Normal, leveraged returns, while taking into account the sample uncertainty associated with track record length. The portfolio optimization approach hereby presented differs from other higher-moment methods in that skewness and kurtosis are incorporated through the standard deviation of the Sharpe ratio estimator. This avoids having to make arbitrary assumptions regarding the relative weightings that higher moments have in the utility function. We feel that practitioners will find this approach useful, because the Sharpe ratio has become –to a certain extent– the default utility function used by investors. SEF can be intuitively explained to investors as the set of portfolios that maximize the expected Sharpe ratio for different degrees of confidence. The maximum Sharpe ratio portfolio is a member of the SEF, but it may differ from the portfolio that maximizes the PSR. While the former portfolio is oblivious to the resulting confidence bands around that maximized

[1] See Sharpe (1975) and Sharpe (1994). Sharpe suggested the name *reward-to-variability ratio*, another matter on which that author's plead has been dismissed.

Sharpe, the latter is the portfolio that maximizes the probability of skill, taking into account the impact that non-Normality and track record length have on the Sharpe ratio's confidence band.

We do not explicitly address the case of serially-conditioned processes. Instead, we rely on Mertens (2002), who 'originally' assumed IID non-Normal returns. That framework is consistent with the scenario that the skill and style of the portfolio manager does not change during the observation period. Fortunately, Opdyke (2007) has shown that Mertens' equation has a limiting distribution that is valid under the more general assumption of stationary and ergodic returns, and not only IID. Thus, our results are valid under such conditions, beyond the narrower IID assumption.

The rest of the paper is organized as follows: Section 2 presents the theoretical framework that will allow us to achieve the three stated goals. Section 3 introduces the concept of Probabilistic Sharpe Ratio (PSR). Section 4 relates applies this concept to answer the question of what is an acceptable track record length for a given confidence level. Section 5 presents numerical examples that illuminate how these concepts are interrelated and can be used in practice. Section 6 applies our methodology to Hedge Fund Research data. Section 7 takes this argument further by introducing the concept of *Sharpe Ratio Efficient Frontier* (SEF). Section 8 outlines the conclusions. Mathematical appendices proof statements made in the body of the paper. Results can be validated using the Python code in the Appendices 3 and 4.

2. THE FRAMEWORK

We have argued that the Sharpe ratio is a deficient measure of investment skill. In order to understand why, we need to review its theoretical foundations, and the implications of its assumption of Normal returns. In particular, we will see that non-Normality may increase the variance of the Sharpe ratio estimator, therefore reducing our confidence in its point estimate. When unaddressed, this means that investors may be comparing Sharpe ratio estimates with widely different confidence bands.

2.1. SHARPE RATIO'S POINT ESTIMATE

Suppose that a strategy's excess returns (or risk premiums), r_t, are *IID*[2]

$$r \sim N(\mu, \sigma^2) \tag{1}$$

where N represents a Normal distribution with mean μ and variance σ^2. The purpose of the Sharpe ratio (SR) is to evaluate the skills of a particular strategy or investor.

$$SR = \frac{\mu}{\sigma} \tag{2}$$

[2] Even if returns are serially correlated, there may be a sampling frequency for which their autocorrelation becomes insignificant. We leave for a future paper the analysis of returns' serial conditionality under different sampling frequencies, and their joint impact on Sharpe ratio estimates.

Since μ, σ are usually unknown, the true value *SR* cannot be known for certain. The inevitable consequence is that Sharpe ratio calculations may be the subject of substantial estimation errors. We will discuss next how to determine them under different sets of assumptions.

2.2. ASSUMING IID NORMAL RETURNS

Like any estimator, SR has a probability distribution. Following Lo (2002), in this section we will derive what this distribution is in the case of IID Normal returns. The Central Limit Theorem states that $\sqrt{n}(\hat{\mu}-\mu) \xrightarrow{a} N(0,\sigma^2)$ and $\sqrt{n}(\hat{\sigma}^2-\sigma^2) \xrightarrow{a} N(0,2\sigma^4)$, where $\xrightarrow{a}$ denotes asymptotic convergence. Let $\theta = \begin{pmatrix}\mu \\ \sigma\end{pmatrix}$ be the column-vector of the Normal distribution's parameters, with an estimate in $\hat{\theta} = \begin{pmatrix}\hat{\mu} \\ \hat{\sigma}\end{pmatrix}$. For IID returns, $\sqrt{n}\left(\hat{\theta}-\theta\right) \xrightarrow{a} N(0,V_\theta)$, where $V_\theta = \begin{pmatrix}\sigma^2 & 0 \\ 0 & 2\sigma^4\end{pmatrix}$ is the variance of the estimation error on θ.

Let's denote $\widehat{SR} = g\left(\hat{\theta}\right)$, where $g(.)$ is the function that estimates *SR*, and apply the delta method (see White (1984)),

$$\sqrt{n}\left(g\left(\hat{\theta}\right)-g(\theta)\right) \xrightarrow{a} N\left(0,V_g\right)$$
$$V_g = \frac{\partial g}{\partial \theta} V_\theta \frac{\partial g}{\partial \theta'} \tag{3}$$

V_g is the variance of the $g(.)$ function. Because $\frac{\partial g}{\partial \theta'} = \begin{pmatrix}\frac{1}{\sigma} \\ -\frac{\hat{\mu}}{2\sigma^3}\end{pmatrix}$, we obtain that $V_g = \left(\frac{\partial g}{\partial \mu}\right)^2 \sigma^2 + \left(\frac{\partial g}{\partial \sigma^2}\right)^2 2\sigma^4$. This means that the asymptotic distribution of $\widehat{SR}$ reduces to

$$\left(\widehat{SR}-SR\right) \xrightarrow{a} N\left(0, \frac{1+\frac{1}{2}SR^2}{n}\right) \tag{4}$$

If q is the number of observations per year, the point estimate of the annualized Sharpe ratio is

$$\widehat{SR}_q \xrightarrow{a} N\left(\sqrt{q}SR, qV_g\right) \tag{5}$$

Under the assumption of Normal IID returns, the *SR* estimator follows a Normal distribution with mean *SR* and a standard deviation that depends on the very value of *SR* and the number of observations. This is an interesting result, because it tells us that, *ceteris paribus*, in general we would prefer investments with a longer track record. That is hardly surprising, and is common practice in the hedge fund industry to ask for track records greater than 3 or more years of monthly returns. Furthermore, Eq. (4) tells us how a greater *n* exactly impacts the variance of the *SR* estimate, which is an idea we will expand in later sections.

2.3. SHARPE RATIO AND NON-NORMALITY

The *SR* does not characterize a distribution of returns, in the sense that there are infinite Normal distributions that deliver any given *SR*. This is easy to see in Eq. (2), as merely re-scaling the returns series will yield the same *SR*, even though the returns come from Normal distributions with different parameters. This argument can be generalized to the case of non-Normal distributions, with the aggravation that, in the non-Normal case, the number of degrees of freedom is even greater (distributions with entirely different first four moments may still yield the same *SR*).

Appendix 2 demonstrates that a simple mixture of two Normal distributions produces infinite combinations of skewness and kurtosis with equal *SR*. More precisely, the proof states that, in the most general cases, there exists a *p* value able to mix any two given Normal distributions and deliver a targeted *SR*.[3] The conclusion is that, however high a *SR* might be, it does not preclude the risk of severe losses. To understand this fact, consider the following combinations of parameters:

$$\mu_1 = \mu_1^*\left(1 - \frac{j-1}{k-1}\right) + SR^*\frac{j-1}{k};\ \mu_2 = SR^*\left(1+\frac{1}{k}\right)\left(1-\frac{j-1}{k-1}\right) + \mu_2^*\frac{j-1}{k-1}$$
$$\sigma_1 = \frac{\sigma_1^*}{k}j;\ \sigma_2 = \frac{\sigma_2^*}{k}j \qquad (6)$$
$$j = 1, \ldots, k$$

For $\mu_1^* \leq \mu_1 \leq SR^*\left(1-\frac{1}{k}\right)$ and $SR^*\left(1+\frac{1}{k}\right) \leq \mu_2 \leq \mu_2^*$, each combination implies a non-Normal mixture. For *k*=20 and $(\mu_1^*, \mu_2^*, \sigma_1^*, \sigma_2^*) = (-5,5,5,5)$, there are 160,000 combinations of $(\mu_1, \mu_2, \sigma_1, \sigma_2)$, but as determined in Appendix 2, only for 96,551 of them there exists a p^* such that $SR^* = 1$. Figure 1 plots the resulting combinations of skewness and kurtosis for mixtures of Normal distributions with the same Sharpe ratio ($SR^* = 1$). An interesting feature of modeling non-Normality through a mixture of Normal distributions is the trade-off that exists between skewness and kurtosis. In this analytical framework, the greater the absolute value of skewness is, the greater the kurtosis tends to be. López de Prado and Peijan (2004) find empirical evidence of this trade-off in their study of returns distributions of hedge fund styles. A mixture of Normal distributions seems to accurately capture this feature in the data.

[FIGURE 1 HERE]

The above set includes combinations as different as $(\mu_1, \mu_2, \sigma_1, \sigma_2, p) = (-5, 1.05, 5, 0.05, 0.015)$ and $(\mu_1, \mu_2, \sigma_1, \sigma_2, p) = (0.3237, 1.8816, 0.05, 0.05, 0.8706)$. Figure 2 displays the probability density functions of these two distributions, which have the same Sharpe ratio ($\mathrm{SR}^* = 1$). The continuous line represents the mixture of two Normal distributions, and the dashed line the Normal distribution with the same mean and standard deviation as the mixture. The mixture on the right side incorporates a 1.5% probability that a return is drawn from a distribution with mean -5 and a standard deviation of 5 (a catastrophic outcome).

[3] Readers interested in the estimation of the parameters that characterize a mixture of 2 Gaussians will find an efficient algorithm in López de Prado and Foreman (2011).

[FIGURE 2 HERE]

Consequently, for a risk averse investor, *SR* does not provide a complete ranking of preferences, unless non-Normality is taken into account. But, how accurately can skewness and kurtosis be estimated from this set of mixtures? In order to answer that question, for each of the 96,551 mixtures included in the above set we have generated a random sample of 1,000 observations (roughly 4 years of daily observations), estimated the first 4 moments on each random sample and compared those estimates with the true mixture's moments (see Eqs. (26)-(35)). Figures 3 (a-d) show that the estimation error is relatively small when moments adopt values within reasonable ranges, particularly for the first 3 moments.

[FIGURE 3 HERE]

Figure 4 reports the results of fitting the two specifications in Eq. (7) on the estimation errors (*er*) and their squares (er^2) for moments *m*=1,…,4.

$$
\begin{gathered}
er_l = \delta_{0,m} + \delta_{1,m}\gamma_m + \delta_{2,m}{\gamma_m}^2 + \varepsilon_l \\
{er_l}^2 = \theta_{0,m} + \theta_{1,m}\gamma_m + \theta_{2,m}{\gamma_m}^2 + \xi_l
\end{gathered} \tag{7}
$$

where $\gamma_1 = \mu$, $\gamma_2 = \sigma$, $\gamma_3 = \frac{E[r-\mu]^3}{\sigma^3}$ is skewness, and $\gamma_4 = \frac{E[r-\mu]^4}{\sigma^4}$ is kurtosis.

[FIGURE 4 HERE]

Consistent with the visual evidence in Figure 3, Figure 4 shows that the estimation error of the mean is not a function of the mean's value (see er_Prob column with prob values at levels usually rejected). The standard deviation's estimator is biased towards underestimating risks (the intercept's er_Prob is at levels at which we would typically reject the null hypothesis of unbiasness), but at least the estimation error does not seem affected by the scale of the true standard deviation. In the case of the third and fourth moments' estimation errors, we find bias and scale effects of first and second degree. This is evidence that estimating moments beyond the third, and particularly the fourth moment, requires longer sample lengths than estimating only the first two moments. We will retake this point in Section 4.

2.4. INCORPORATING NON-NORMALITY

The previous section argued that non-Normal distributions with very diverse risk profiles can all have the same *SR*. In this section we will discuss the key fact that, although skewness and kurtosis does not affect the point estimate of *SR*, it greatly impacts its confidence bands, and consequently its statistical significance. This fact of course has dreadful implications when, as it is customary, point estimates of *SR* are used to rank investments.

Mertens (2002) concludes that the Normality assumption on returns could be dropped, and still the estimated Sharpe ratio would follow a Normal distribution with parameters

$$\left(\widehat{SR} - SR\right) \xrightarrow{a} N\left(0, \frac{1 + \frac{1}{2}SR^2 - \gamma_3 SR + \frac{\gamma_4 - 3}{4}SR^2}{n}\right) \tag{8}$$

The good news is, *SR* follows a Normal distribution even if the returns do not. The bad news is, although most investors prefer to work in a mean-variance framework, they need to take non-Normality into account (in addition, of course, to sample length). Figure 5 illustrates how combinations of skewness and kurtosis impact the standard deviation of the *SR* estimator. This has the serious implication that non-Normal distributions may severely inflate the *SR* estimate, to the point that having a high SR may not be sufficient warranty of its statistical significance.

[FIGURE 5 HERE]

Christie (2005) uses a GMM approach to derive a limiting distribution that only assumes stationary and ergodic returns, thus allowing for time-varying conditional volatilities, serial correlation and even non-IID returns. Surprisingly, Opdyke (2007) proved that the expressions in Mertens (2002) and Christie (2005) are in fact identical. To Dr. Mertens' credit, his result appears to be valid under the more general assumption of stationary and ergodic returns, and not only IID.

2.5. CONFIDENCE BAND

We have mentioned that skewness and kurtosis will affect the confidence band around our estimate of *SR*, but we did not explicitly derive its expression. After some algebra, Eq.(8) gives the estimated standard deviation of $\widehat{SR}$ as $\hat{\sigma}_{\widehat{SR}} = \sqrt{\frac{1 - \hat{\gamma}_3 \widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}{n-1}}$, where $n - 1$ is due to Bessel's correction. The true value *SR* is bounded by our $\widehat{SR}$ estimate with a significance level α

$$Prob\left[SR \in \left(\widehat{SR} - Z_{\alpha/2}\hat{\sigma}_{\widehat{SR}}, \widehat{SR} + Z_{\alpha/2}\hat{\sigma}_{\widehat{SR}}\right)\right] = 1 - \alpha \tag{9}$$

In general it is misleading to judge strategies' performance by merely comparing their respective point estimates of $\widehat{SR}$, without considering the estimation errors involved in each calculation. Instead, we could compare $\widehat{SR}$'s translation in probabilistic terms, which we will define next.

3. PROBABILISTIC SHARPE RATIO (PSR)

Now that we have derived an expression for the confidence bands of *SR*, we are ready to aim for the first goal stated in the Introduction: Provide a de-inflated estimate of *SR*. Given a predefined benchmark[4] Sharpe ratio (SR^*), the observed Sharpe ratio $\widehat{SR}$ can be expressed in probabilistic terms as

[4] This could be set to a default value of zero (i.e., comparing against no investment skill).

$$\widehat{PSR}(SR^*) = Prob\left[\widehat{SR} > SR^*\right] = 1 - \int_{-\infty}^{SR^*} Prob\left(\widehat{SR}\right) \cdot d\widehat{SR} \tag{10}$$

We ask the question, what is the probability that $\widehat{SR}$ is greater than a hypothetical SR^*? Applying what we have learnt in the previous sections, we propose

$$\widehat{PSR}(SR^*) = Z\left[\frac{\left(\widehat{SR} - SR^*\right)\sqrt{n-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right] \tag{11}$$

where Z is the *cdf* of the Standard Normal distribution. For a given SR^*, $\widehat{PSR}$ increases with greater $\widehat{SR}$ (in the original sampling frequency, i.e. non-annualized), or longer track records (*n*), or positively skewed returns ($\hat{\gamma}_3$), but it decreases with fatter tails ($\hat{\gamma}_4$). Because hedge fund strategies are usually characterized by negative skewness and fat tails (Brooks and Kat (2002), López de Prado and Rodrigo (2004)), Sharpe ratios tend to be "inflated". $\widehat{PSR}(SR^*)$ takes those characteristics into account and delivers a corrected, atemporal[5] measure of performance expressed in terms of *probability of skill*.[6] It is not unusual to find strategies with irregular trading frequencies, such as weekly strategies that may not trade for a month. This poses a problem when computing an annualized Sharpe ratio, and there is no consensus as how skill should be measured in the context of irregular bets. Because *PSR* measures skill in probabilistic terms, it is invariant to calendar conventions. All calculations are done in the original frequency of the data, and there is no annualization. This is another argument for preferring *PSR* to traditional annualized *SR* readings in the context of strategies with irregular frequencies.

Section 2.3 made the point that estimates of skewness and kurtosis may incorporate significant errors. If the researcher believes that this is the case with their estimated $\hat{\gamma}_3$ and $\hat{\gamma}_4$, we recommend that a lower bound is inputted in place of $\hat{\gamma}_3$ and an upper bound in place of $\hat{\gamma}_4$ in Eq. (8), for a certain confidence level. However, if these estimates are deemed to be reasonably accurate, this 'worst case scenario analysis' is not needed.

An example will clarify how *PSR* reveals information otherwise dismissed by *SR*. Suppose that a hedge fund offers you the statistics displayed in Figure 6, based on a monthly track record over the last two years.

[FIGURE 6 HERE]

[FIGURE 7 HERE]

At first sight, an annualized Sharpe ratio of 1.59 over the last two years seems high enough to reject the hypothesis that it has been achieved by sheer luck. The question is, *"how inflated is this annualized Sharpe ratio due to the track record's non-normality, length and sampling*

---

[5] $\widehat{SR}$ and SR^* are expressed in the same frequency as the returns time series.

[6] After applying *PSR* on his track record, a hedge fund manager suggested this measure to be named *"The Sharpe razor"* [sic].

frequency?" Let's start by comparing this performance with the skill-less benchmark ($SR^* = 0$) while assuming Normality ($\hat{\gamma}_3 = 0, \hat{\gamma}_4 = 3$). The original sampling frequency is monthly, and so the estimate that goes into Eq. (11) is $\widehat{SR} = 0.458$. This yields a reassuring $\widehat{PSR}(0) = 0.982$. However, when we incorporate the skewness ($\hat{\gamma}_3 = -2.448$) and kurtosis information ($\hat{\gamma}_4 = 10.164$), then $\widehat{PSR}(0) = 0.913$! At a 95% confidence level, we would accept this track record in the first instance, but could not reject the hypothesis that this Sharpe ratio is skill-less in the second instance.

Figure 7 illustrates what is going on. The dashed black line is the Normal *pdf* that matches the Mean and StDev values in Figure 6. The black line represents the mixture of two Normal distributions that matches all four moments in Table 1 ($\mu_1 = -0.1\%$, $\mu_2 = 0.06\%$, $\sigma_1 = 0.12\%$, $\sigma_2 = 0.03\%$, *p*=0.15). Clearly, it is a mistake to assume normality, as that would ignore critical information regarding the hedge fund's loss potential.

What the annualized Sharpe ratio of 1.59 was hiding was a relatively small probability (15%) of a return drawn from an adverse distribution (a negative multiple of the mixed distribution's mode). This is generally the case in track records with negative skewness and positive excess kurtosis, and it is consistent with the signs of $\hat{\gamma}_3$ and $\hat{\gamma}_4$ in Eq. (11).

This is not to say that a track record of 1.59 Sharpe ratio is worthless. As a matter of fact, should we have 3 years instead of 2, $\widehat{PSR}(0) = 0.953$, enough to reject the hypothesis of skill-less performance even after considering the first four moments. In other words, *a longer track record may be able to compensate for the uncertainty introduced by non-Normal returns*. The next Section quantifies that "compensation effect" between non-Normality and the track record's length.

PSR takes into account the statistical accuracy of the point estimate of *SR* for different levels of skewness and kurtosis (and length of track record). In this sense, it incorporates information regarding the non-Normality of the returns. However, we caution the reader that *PSR* does not, and does not attempt to, incorporate the effect of higher moments on preferences. The investor still only cares about mean and variance, but she is rightly worried that in the presence of skewness and kurtosis –about which she does not care *per se*– her estimates may be inaccurate and 'flattering'.

4. TRACK RECORD LENGTH

Understanding that Sharpe ratio estimations are subject to significant errors begs the question: *"How long should a track record be in order to have statistical confidence that its Sharpe ratio is above a given threshold?"* In mathematical terms, for $\widehat{SR} > SR^*$, this is equivalent to asking

$$\left\{n \middle| \widehat{PSR}(SR^*) > 1 - \alpha\right\} \tag{12}$$

with minimum track record length (*MinTRL*) in

$$MinTRL = n^* = 1 + \left[1 - \hat{\gamma}_3 \widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4} \widehat{SR}^2\right] \left(\frac{Z_\alpha}{\widehat{SR} - SR^*}\right)^2 \tag{13}$$

And again we observe that a longer track record will be required the smaller $\widehat{SR}$ is, or the more negatively skewed returns are, or the greater the fat tails, or the greater our required level of confidence. A first practical implication is that, if a track record is shorter than *MinTRL*, we do not have enough confidence that the observed $\widehat{SR}$ is above the designated threshold SR^*. A second practical implication is that a portfolio manager will be penalized because of her non-Normal returns, however she can regain the investor's confidence over time (by extending the length of her track record).

It is important to note that *MinTRL* is expressed in terms of number of observations, not annual or calendar terms. A note of caution is appropriate at this point: Eqs. (11) and (13) are built upon Eq. (8), which applies to an asymptotic distribution. CLT is typically assumed to hold for samples in excess of 30 observations (Hogg and Tanis (1996)). So even though a *MinTRL* may demand less than 2.5 years of monthly data, or 0.5769 years of weekly data, or 0.119 years of daily data, etc. the moments inputted in Eq. (13) must be computed on longer series for CLT to hold. This is consistent with practitioners' standard practice of requiring similar lengths during the due diligence process.

5. NUMERICAL EXAMPLES

Everything we have learnt in the previous sections can be illustrated in a few practical examples. Figure 8 displays the minimum track record lengths (*MinTRL*) in years required for various combinations of measured $\widehat{SR}$ (rows) and benchmarked SR^* (columns) at a 95% confidence level, based upon *daily* IID Normal returns. For example, a *2.73* years track record is required for an annualized Sharpe of 2 to be considered greater than 1 at a 95% confidence level.

[FIGURE 8 HERE]

We ask, what would the *MinTRL* be for a weekly strategy with also an observed annualized Sharpe of 2? Figure 9 shows that, if we move to *weekly* IID Normal returns, the requirement is *2.83* years of track record length, a 3.7% increase.

[FIGURE 9 HERE]

Figure 10 indicates that the track record length needed increases to *3.24* years if instead we work with *monthly* IID Normal returns, an 18.7% increase compared to daily IID Normal returns. This increase in *MinTRL* occurs despite the fact that both strategies have the same observed annualized Sharpe ratio of 2, and it is purely caused by a decrease in frequency.

[FIGURE 10 HERE]

Let's stay with monthly returns. Brooks and Kat (2002) report that the HFR Aggregate Hedge Fund returns index exhibits $\hat{\gamma}_3 = -0.72$ and $\hat{\gamma}_4 = 5.78$. In these circumstances, Figure 11 tells us that the track record should now be *4.99* years long. This is 54% longer than what we required with Normal monthly returns, and 82.8% longer than what was needed with Normal daily returns.

[FIGURE 11 HERE]

6. SKILLFUL HEDGE FUND STYLES

We are now ready to run our model on real data. Figure 12 applies our methodology on HFR Monthly indices from January 1st 2000 to May 1st 2011 (134 monthly observations, or 11.167 years). *MinTRL* is expressed in years, subject to a confidence level of 95%.

A $PSR(0) > 0.95$ indicates that a *SR* is greater than 0 with a confidence level of 0.95. Similarly, a $PSR(0.5) > 0.95$ means that a *SR* is greater that 0.5 (annualized) with a confidence level of 0.95. The *Probabilistic Sharpe ratio* has taken into account multiple statistical features present in the track record, such as its length, frequency and deviations from Normality (skewness, kurtosis).

Because our sample consists of 11.167 years of monthly observations, a $PSR(0) > 0.95$ is consistent with a $MinTRL(0) < 11.167$ at 95% confidence, and a $PSR(0.5) > 0.95$ is consistent with a $MinTRL(0.5) < 11.167$ at 95% confidence. Our calculations show that most hedge fund styles evidence some level of skill, i.e. their *SR* are above the zero benchmark. However, looking at *PSR(0.5)*, we observe that only 9 style indices substantiate investment skill over an annualized Sharpe ratio of 0.5 at a 95% confidence level:

- Distressed Securities
- Equity Market Neutral
- Event Driven
- Fixed Asset-Backed
- Macro
- Market Defensive
- Mortgage Arbitrage
- Relative Value
- Systematic Diversified

[FIGURE 12 HERE]

This is not to say that only hedge funds practicing the 9 styles listed above should be considered. Our analysis has been performed on indices, not specific track records. However, it could be argued that special care should be taken when analyzing performance from styles other than the 9 mentioned. We would have liked to complete this analysis with a test of structural breaks, however the amount and quality of data does not allow for meaningful estimates.

7. THE SHARPE RATIO EFFICIENT FRONTIER

PSR evaluates the performance of an individual investment in terms of an uncertainty-adjusted *SR*. It seems natural to extend this argument to a portfolio optimization or capital allocation context. Rather than a mean-variance frontier of portfolio returns on capital, we will build a mean-variance frontier of portfolio returns on risk.

Following Markowitz (1952), a portfolio w belongs to the Efficient Frontier if it delivers maximum expected excess *return on capital* ($E[rw]$) subject to the level of uncertainty surrounding those portfolios' excess returns ($\hat{\sigma}_{(rw)}$).

$$\max_{w} \quad E[rw] \big| \hat{\sigma}_{(rw)} = \sigma^* \tag{14}$$

Similarly, we define what we denote the *Sharpe ratio Efficient Frontier* (SEF) as the set of portfolios $\{w\}$ that deliver the highest expected excess *return on risk* (as expressed by their Sharpe ratios) subject to the level of uncertainty surrounding those portfolios' excess returns on risk (the standard deviation of the Sharpe ratio).

$$\max_{w} \quad \widehat{SR}(rw) \big| \hat{\sigma}_{\widehat{SR}(rw)} = \sigma^* \tag{15}$$

But why would we compute an efficient frontier of Sharpe ratios while accepting that returns (r) are non-Normal? Because a great majority of investors use the *SR* as a proxy for their utility function. Even though they do not care about higher moments *per se*, they must de-inflate their estimates of *SR* (a mean-variance metric) using the third and fourth moments. A number of additional reasons make this analysis interesting:

1. *SEF* deals with efficiency within the *return on risk* (or Sharpe ratio) space rather than *return on capital*. Unlike returns on capital, Sharpe ratios are invariant to leverage.
2. Even if returns are non-Normally distributed,
    a. the distribution of Sharpe ratios follows a Normal, therefore an efficient frontier–style of analysis still makes sense.
    b. as long as the process is IID, the cumulative returns distribution asymptotically converges to Normal, due to the Central Limit Theorem.
3. Performance manipulation methods like those discussed by Ingersoll, Spiegel, Goetzmann and Welch (2007) generally attempt to inflate the Sharpe ratio by distorting the returns distribution. As *SEF* considers higher moments, it adjusts for such manipulation.
4. It is a *second degree of uncertainty* analysis. The standard (Markowitz) portfolio selection framework measures uncertainty in terms of standard deviation on returns. In the case of *SEF*, uncertainty is measured on a function ($\widehat{SR}(rw)$) that already incorporates an uncertainty estimate ($\hat{\sigma}_{(rw)}$). Like in Black-Litterman (1992), this approach does not assume perfect knowledge of the mean-variance estimates, and deals with uncertainty in the model's input variables. This in turn increases the robustness of the solution, which contrasts with the instability of mean-variance optimization (see Best and Grauger (1991)).

5. Computing the *SEF* will allow us to identify the portfolio that delivers the highest *PSR* for any given SR^* threshold, thus dealing with non-Normality and sample uncertainty due to track record length in the context of portfolio selection. From Eq. (11), the highest *PSR* portfolio is the one such that

$$\arg\max_{w} \frac{\widehat{SR}(rw)}{\hat{\sigma}_{\widehat{SR}(rw)}} = \arg\max_{w} \frac{\widehat{SR}(rw)\sqrt{n-1}}{\sqrt{1-\hat{\gamma}_3(rw)\widehat{SR}(rw)+\frac{\hat{\gamma}_4(rw)-1}{4}\widehat{SR}(rw)^2}} \tag{16}$$

A numerical example will clarify this new analytical framework. There exist 43,758 fully invested long portfolios that are linear combinations of the 9 HFR indices identified in the previous section, with weightings

$$w_i = \frac{j}{10}, \qquad j = 0, \dots, 10, \qquad i = 1, \dots, 9$$
$$\sum_{i=1}^{9} w_i = 1 \tag{17}$$

Because non-Normality and sample length impact our confidence on each portfolio's risk-adjusted return, selecting the highest Sharpe ratio portfolio is suboptimal. This is illustrated in Figure 13, where the highest *SR* portfolio (right end of the *SEF*) comes at the expense of substantial uncertainty with regards that estimate, since $\left(\hat{\sigma}_{\widehat{SR}(rw)}, \widehat{SR}(rw)\right) = (0.155, 0.818)$. The portfolio that delivers the highest *PSR* is indeed quite different, as marked by the encircled cross ($\left(\hat{\sigma}_{\widehat{SR}(rw)}, \widehat{SR}(rw)\right) = (0.103, 0.708)$). Recall that the x-axis in this figure does not represent the risk associated with an investment, but the statistical uncertainty surrounding our estimation of *SR*.

[FIGURE 13 HERE]

Figure 14 illustrates how the composition of the *SEF* evolves as $\hat{\sigma}_{\widehat{SR}(rw)}$ increases. The vertical line at $\hat{\sigma}_{\widehat{SR}(rw)} = 0.103$ indicates the composition of the highest *PSR* portfolio, while the vertical line at $\hat{\sigma}_{\widehat{SR}(rw)} = 0.155$ gives the composition of the highest *SR* portfolio. The transition across different regions of the *SEF* is very gradual, as a consequence of the robustness of this approach.

[FIGURE 14 HERE]

Figure 15 shows why the *Max PSR* solution is preferable: Although it delivers a lower Sharpe ratio than the *Max SR* portfolio (0.708 vs. 0.818 in monthly terms), its better diversified allocations allow for a much greater confidence (0.103 vs. 0.155 standard deviations). *Max PSR* invests in 5 styles, and the largest holding is 30%, compared to the 4 styles and 50% maximum holding of the *Max SR* portfolio.

[FIGURE 15 HERE]

The *Max PSR* portfolio displays better statistical properties than the *Max SR* portfolio, as presented in Figure 16: *Max PSR* is very close to Normal (almost null skewness and kurtosis close to 3, $JB\ prob \approx 0.6$), while the *Max SR* portfolio features a left fat-tail ($JB\ prob \approx 0$). A risk averse investor should not accept a 17.4% probability of returns being drawn from an adverse distribution in exchange for aiming at a slightly higher Sharpe ratio (Figures 17-18).

[FIGURE 16 HERE]

[FIGURE 17 HERE]

[FIGURE 18 HERE]

In other words, taking into account higher moments has allowed us to *naturally* find a better balanced portfolio that is optimal in terms of uncertainty-adjusted Sharpe ratio. We say "naturally" because this result is achieved without requiring constraints on the maximum allocation permitted per holding. The reason is, *PSR* recognizes that concentrating risk increases the probability of catastrophic outcomes, thus it penalizes such concentration.

8. CONCLUSIONS

A probabilistic translation of Sharpe ratio, called *PSR*, is proposed to account for estimation errors in an IID non-Normal framework. When assessing Sharpe ratio's ability to evaluate skill, we find that a longer track record may be able to compensate for certain statistical shortcomings of the returns probability distribution. Stated differently, *despite Sharpe ratio's well-documented deficiencies, it can still provide evidence of investment skill, as long as the user learns to require the proper track record length*.

Even under the assumption of IID returns, the track record length required to exhibit skill is greatly affected by the asymmetry and kurtosis of the returns distribution. A typical hedge fund's track record exhibits negative skewness and positive excess kurtosis, which has the effect of "inflating" its Sharpe ratio. One solution is to compensate for such deficiencies with a longer track record. When that is not possible, a viable option may be to provide returns with the highest sampling frequency such that the IID assumption is not violated. The reason is, for negatively skewed and fat-tailed returns distributions, the number of years required may in fact be lowered as the sampling frequency increases. This has led us to affirm that "badly behaved" returns distributions have the most to gain from offering the greatest transparency possible, in the form of higher data granularity.

We present empirical evidence that, despite the high Sharpe ratios publicized for several hedge fund styles, in many cases they may not be high enough to indicate statistically significant investment skill beyond a moderate annual Sharpe ratio of 0.5 for the analyzed period, confidence level and track record length.

Finally, we discuss the implications that this analysis has in the context of capital allocation. Because non-Normality, leverage and track record length impact our confidence on each portfolio's risk-adjusted return, selecting the highest Sharpe ratio portfolio is suboptimal. We develop a new analytical framework, called the *Sharpe ratio Efficient Frontier* (SEF), and find that the portfolio of hedge fund indices that maximizes Sharpe ratio can be very different from the portfolio that delivers the highest *PSR*. Maximizing for *PSR* leads to better diversified and more balanced hedge fund allocations compared to the concentrated outcomes of Sharpe ratio maximization.

APPENDICES

A.1. HIGHER MOMENTS OF A MIXTURE OF *m* NORMAL DISTRIBUTIONS

Let z be a random variable distributed as a standard normal, $z \sim N(0,1)$. Then, $\eta = \mu + \sigma z \sim N(\mu, \sigma^2)$, with *characteristic function*:

$$\vartheta_\eta(s) = E\left[e^{is\eta}\right] = E\left[e^{is\mu}\right]E\left[e^{is\sigma z}\right] = e^{is\eta}\,\underbrace{\vartheta_z(s\sigma)}_{e^{-\frac{1}{2}s^2\sigma^2}} = e^{is\mu-\frac{1}{2}s^2\sigma^2} \tag{18}$$

Let r be a random variable distributed as a mixture of m normal distributions, $r \sim M(\mu_1, \dots, \mu_m, \sigma_1, \dots, \sigma_m, p_1, \dots, p_m)$, with $\sum_{j=1}^{m} p_j = 1$. Then:

$$\vartheta_r(s) = E\left[e^{isr}\right] = e^{\sum_{j=1}^{m} p_j\left(is\mu_j - \frac{1}{2}s^2\sigma_j^2\right)} \tag{19}$$

The k^{th} moment centered about zero of any random variable x can be computed as:

$$E[x^k] = \frac{\left|\frac{\partial^k \vartheta_x(s)}{\partial s^k}\right|_{s=0}}{i^k} \tag{20}$$

In the case of r, the first 5 moments centered about zero can be computed as indicated above, leading to the following results:

$$E[r] = \sum_{j=1}^{m} p_j \mu_j \tag{21}$$

$$E[r^2] = \sum_{j=1}^{m} p_j\left(\sigma_j^2 + \mu_j^2\right) \tag{22}$$

$$E[r^3] = \sum_{j=1}^{m} p_j\left(3\sigma_j^2\mu_j + \mu_j^3\right) \tag{23}$$

$$E[r^4] = \sum_{j=1}^{m} p_j\left(3\sigma_j^4 + 6\sigma_j^2\mu_j^2 + \mu_j^4\right) \tag{24}$$

$$E[r^5] = \sum_{j=1}^{m} p_j\left(15\sigma_j^4\mu_j + 10\sigma_j^2\mu_j^3 + \mu_j^5\right) \tag{25}$$

The first 5 central moments about the mean are computed by applying Newton's binomium:

$$E\left[r - E[r]\right]^k = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (E[r])^j E\left[r^{k-j}\right] \tag{26}$$

$$E\left[r-E[r]\right] \quad = 0 \tag{27}$$

$$E\left[r-E[r]\right]^2 = E[r^2]-(E[r])^2 \tag{28}$$

$$E\left[r-E[r]\right]^3 = E[r^3]-3E[r^2]E[r]+2(E[r])^3 \tag{29}$$

$$E\left[r-E[r]\right]^4 = E[r^4]-4E[r^3]E[r]+6E[r^2](E[r])^2-3(E[r])^4 \tag{30}$$

$$\begin{gathered} E\left[r-E[r]\right]^5 = E[r^5]-5E[r^4]E[r]+10E[r^3](E[r])^2 \\ -10E[r^2](E[r])^3+4(E[r])^5 \end{gathered} \tag{31}$$

A.2. TARGETING SHARPE RATIO THROUGH A MIXTURE OF TWO NORMAL DISTRIBUTIONS

Suppose that $r \sim M\left(\mu_1, \mu_2, \sigma_1\ , \sigma_2\ , p, 1-p\right)$. We ask for what value p the Mixture of two Normal distributions is such that

$$\frac{E[r]}{\sqrt{E\left[r-E[r]\right]^2}} = SR^* \tag{32}$$

where SR^* is a targeted Sharpe ratio. Setting SR^* implies that p will now be a function of the other parameters, $p = f\left(\mu_1, \sigma_1\ , \mu_2, \sigma_2\ , SR^*\right)$. In this section we will derive that function *f*.

From Eq. (32), $(E[r])^2 = SR^{*2}E\left[r-E[r]\right]^2$. Applying Eq. (28), this expression simplifies into

$$(E[r])^2\left(1+SR^{*2}\right) = SR^{*2}E[r^2] \tag{33}$$

From Eq. (21) and Eq. (22),

- $(E[r])^2 = \left(\mu_1 p + \mu_2(1-p)\right)^2 = p^2(\mu_1^2+\mu_2^2-2\mu_1\mu_2)+2p(\mu_1\mu_2-\mu_2^2)+\mu_2^2$
- $E[r^2] = (\sigma_1^2+\mu_1^2)p+(\sigma_2^2+\mu_2^2)(1-p) = p(\sigma_1^2+\mu_1^2-\sigma_2^2-\mu_2^2)+\sigma_2^2+\mu_2^2$

Let $\alpha = \mu_1^2+\mu_2^2-2\mu_1\mu_2$ and $\beta = 1+\frac{1}{SR^{*2}}$. Then, Eq. (33) can be rewritten as

$$(p^2\alpha+2p(\mu_1\mu_2-\mu_2^2)+\mu_2^2)\beta = p(\sigma_1^2+\mu_1^2-\sigma_2^2-\mu_2^2)+\sigma_2^2+\mu_2^2 \tag{34}$$

which can be reduced into

$$p^2\alpha\beta+p(2\beta(\mu_1\mu_2-\mu_2^2)-\mu_1^2-\sigma_1^2+\mu_2^2+\sigma_2^2)+\mu_2^2(\beta-1)-\sigma_2^2 = 0 \tag{35}$$

For $a = \alpha\beta$, $b = 2\beta(\mu_1\mu_2 - \mu_2^2) - \mu_1^2 - \sigma_1^2 + \mu_2^2 + \sigma_2^2$, $c = \mu_2^2(\beta - 1) - \sigma_2^2$, Eq. (35) leads to the monic quadratic equation

$$p^2 + p\frac{b}{a} + \frac{c}{a} = 0 \tag{36}$$

with solution in

$$p^* = \frac{1}{2}\left(-\frac{b}{a} \pm \sqrt{\left(\frac{b}{a}\right)^2 - 4\frac{c}{a}}\right) \tag{37}$$

where

- $a = \left(1 + \frac{1}{SR^{*2}}\right)(\mu_1^2 + \mu_2^2 - 2\mu_1\mu_2)$
- $b = 2\left(1 + \frac{1}{SR^{*2}}\right)(\mu_1\mu_2 - \mu_2^2) - \mu_1^2 - \sigma_1^2 + \mu_2^2 + \sigma_2^2$
- $c = \frac{\mu_2^2}{SR^{*2}} - \sigma_2^2$

Let's discuss now the condition of existence of the solution: In order to be a probability, solutions with an imaginary part must be discarded, which leads to the condition that

$$b^2 \geq 4ac \tag{38}$$

Furthermore, because in Eq. (33) we squared both sides of the equality, p^* could deliver $\frac{E[r]}{\sqrt{E\left[r - E[r]\right]^2}} = -SR^*$. So a second condition comes with selecting the root p^* such that

$$Sgn(SR^*) = Sgn(p^*\mu_1 + (1 - p^*)\mu_2) \tag{39}$$

Finally, in order to have $0 \leq p^* \leq 1$, it is necessary that either

$$\begin{gathered}\frac{\mu_2}{\sigma_2} \geq SR^* \geq \frac{\mu_1}{\sigma_1} \\ \text{or} \\ \frac{\mu_1}{\sigma_1} \geq SR^* \geq \frac{\mu_2}{\sigma_2}\end{gathered} \tag{40}$$

This result allows us to simulate a wide variety of non-Normal distributions delivering the same targeted Sharpe ratio (SR^*).

A.3. IMPLEMENTATION OF PSR STATISTICS

PSR and *MinTRL* calculations are implemented in the following code. The input parameters are set to replicate the result obtained in Figure 11 ($\widehat{SR} = \frac{2}{\sqrt{12}}$, $\hat{\gamma}_3 = -0.72$, $\hat{\gamma}_4 = 5.78$, $SR^* = \frac{1}{\sqrt{12}}$, where $\sqrt{12}$ factor recovers the monthly *SR* estimates). Then, $MinTRL(0.95) = 59.895$ months, or approx. 4.99 years. This result is corroborated by computing the PSR with a sample length of 59.895, which gives value of $PSR\left(\frac{1}{\sqrt{12}}\right) = 0.95$.

```
#!/usr/bin/env python
# PSR class for computing the Probabilistic Sharpe Ratio
# On 20120502 by MLdP <lopezdeprado@lbl.gov>

from scipy.stats import norm
#------------------------------------------
# PSR class
class PSR:
    def __init__(self,stats,sr_ref,obs,prob):
        self.PSR=0
        self.minTRL=0
        self.stats=stats
        self.sr_ref=sr_ref
        self.obs=obs
        self.prob=prob
#------------------------------------------
    def set_PSR(self,moments):
        stats=self.stats[:moments]+[0 for i in \
            range(len(self.stats)-moments)]
        sr=self.stats[0]/self.stats[1]
        self.PSR=norm.cdf((sr-self.sr_ref)*(self.obs-1)**0.5/ \
            (1-sr*stats[2]+sr**2*(stats[3]-1)/4.)**0.5)
#------------------------------------------
    def set_TRL(self,moments):
        stats=self.stats[:moments]+[0 for i in \
            range(len(self.stats)-moments)]
        sr=self.stats[0]/self.stats[1]
        self.minTRL=1+(1-stats[2]*sr+(stats[3]-1)/4.*sr**2)* \
            (norm.ppf(self.prob)/(sr-self.sr_ref))**2
#------------------------------------------
    def get_PSR(self,moments):
        self.set_PSR(moments)
        return self.PSR
#------------------------------------------
    def get_TRL(self,moments):
        self.set_TRL(moments)
        return self.minTRL
#------------------------------------------
#------------------------------------------
# Main function
def main():
    #1) Inputs (stats on excess returns)
    stats=[2,12**0.5,-0.72,5.78]        #non-annualized stats
    sr_ref=1/12**0.5            #reference Sharpe ratio (non-annualized)
    obs=59.895
    prob=0.95
```

```
    #2) Create class
    psr=PSR(stats,sr_ref,obs,prob)

    #3) Compute and report values
    print 'PSR(2m,3m,4m):',[psr.get_PSR(i) for i in \
        range(2,5,1)]
    print 'minTRL(2m,3m,4m):',[psr.get_TRL(i) for i in \
        range(2,5,1)]
#-------------------------------------------
# Boilerplate
if __name__=='__main__': main()
```

A.4. COMPUTING THE PSR OPTIMAL PORTFOLIO

A.4.1. TAYLOR'S EXPANSION

We would like to find the vector of weights ω that maximize the expression

$$\widehat{PSR}(SR^*) = Z\left[\frac{\left(\widehat{SR} - SR^*\right)\sqrt{n-1}}{\sqrt{1 - \hat{\gamma}_3\widehat{SR} + \frac{\hat{\gamma}_4 - 1}{4}\widehat{SR}^2}}\right] \tag{41}$$

where $r = \sum_{i=1}^{I} \omega_i \tilde{r}_i$ is the return of the portfolio with weightings ω (of dimension I), $\mu = \sum_{i=1}^{I} \omega_i \tilde{\mu}_i = \sum_{i=1}^{I} \omega_i E[\tilde{r}_i]$ is the mean portfolio return, $\sigma = \sqrt{E[(r-\mu)^2]}$ its standard deviation, $\gamma_3 = \frac{E\left[(r-\mu)^3\right]}{\sigma^3}$ its skewness, $\gamma_4 = \frac{E\left[(r-\mu)^4\right]}{\sigma^4}$ its kurtosis and $SR = \frac{\mu}{\sigma}$ its Sharpe ratio. Because $\widehat{PSR}(SR^*) = Z\left[\widehat{Z^*}\right]$ is a monotonic increasing function of $\widehat{Z^*}$, it suffices to compute the vector ω that maximizes $\widehat{Z^*}$. This optimal vector is invariant to the value adopted by the parameter SR^*.

A second degree Taylor expansion of the $\widehat{PSR}$ function takes the form:

$$\begin{aligned}\widehat{Z^*}(\omega_i + \Delta\omega_i) = \widehat{Z^*}(\omega_i) + \frac{\partial \widehat{Z^*}(\omega_i)}{\partial \omega_i}\Delta\omega_i + \frac{1}{2}\frac{\partial^2 \widehat{Z^*}(\omega_i)}{\partial {\omega_i}^2}(\Delta\omega_i)^2 \\ + \sum_{k=3}^{\infty}\frac{1}{k!}\frac{\partial^k \widehat{Z^*}(\omega_i)}{\partial {\omega_i}^k}(\Delta\omega_i)^k\end{aligned} \tag{42}$$

So we need to compute an analytical expression for the first and second partial derivatives.

A.4.2. FIRST DERIVATIVE

We would like to compute the derivative of the expression $\widehat{Z^*} = \frac{\widehat{SR}-SR^*}{\hat{\sigma}_{SR}}$,

$$\frac{\partial \widehat{Z^*}}{\partial \omega_j} = \frac{1}{\hat{\sigma}_{SR}^2}\left(\frac{\partial \widehat{SR}}{\partial \omega_j}\hat{\sigma}_{SR} - \frac{\partial \hat{\sigma}_{SR}}{\partial \omega_j}\left(\widehat{SR} - SR^*\right)\right) \tag{43}$$

This requires us to compute $\frac{\partial \widehat{SR}}{\partial \omega_j}$ and $\frac{\partial \hat{\sigma}_{SR}}{\partial \omega_j}$, where $\hat{\sigma}_{SR} = \sqrt{\frac{1-\hat{\gamma}_3\widehat{SR}+\frac{\hat{\gamma}_4-1}{4}\widehat{SR}^2}{n-1}}$:

$$\frac{\partial SR}{\partial \omega_j} = \frac{1}{\sigma^2}\left(\frac{\partial \mu}{\partial \omega_j}\sigma - \frac{\partial \sigma}{\partial \omega_j}\mu\right) \tag{44}$$

$$\begin{aligned}\frac{\partial \hat{\sigma}_{SR}}{\partial \omega_j} = \frac{1}{2\hat{\sigma}_{SR}(n-1)}\Bigg(&-\frac{\partial \hat{\gamma}_3}{\partial \omega_j}\widehat{SR} - \frac{\partial \widehat{SR}}{\partial \omega_j}\hat{\gamma}_3 \\ &+ \frac{1}{4}\left(\frac{\partial \hat{\gamma}_4}{\partial \omega_j}\widehat{SR}^2 + 2\widehat{SR}\frac{\partial \widehat{SR}}{\partial \omega_j}(\hat{\gamma}_4 - 1)\right)\Bigg)\end{aligned} \tag{45}$$

We are still missing $\frac{\partial \mu}{\partial \omega_j}$, $\frac{\partial \sigma}{\partial \omega_j}$, $\frac{\partial \hat{\gamma}_3}{\partial \omega_j}$ and $\frac{\partial \hat{\gamma}_4}{\partial \omega_j}$:

$$\frac{\partial \mu}{\partial \omega_j} = \tilde{\mu}_j \tag{46}$$

$$\frac{\partial \sigma}{\partial \omega_j} = \frac{1}{2\sigma}\frac{\partial E[(r-\mu)^2]}{\partial \omega_j} = \frac{1}{\sigma}E\left[\left(\tilde{r}_j - \tilde{\mu}_j\right)\sum_{i=1}^{I}\omega_i(\tilde{r}_i - \tilde{\mu}_i)\right] \tag{47}$$

$$\frac{\partial \gamma_3}{\partial \omega_j} = \sigma^{-6}\left(\frac{\partial E[(r-\mu)^3]}{\partial \omega_j}\sigma^3 - 3\sigma^2\frac{\partial \sigma}{\partial \omega_j}E[(r-\mu)^3]\right) \tag{48}$$

$$\frac{\partial \gamma_4}{\partial \omega_j} = \sigma^{-8}\left(\frac{\partial E[(r-\mu)^4]}{\partial \omega_j}\sigma^4 - 4\sigma^3\frac{\partial \sigma}{\partial \omega_j}E[(r-\mu)^4]\right) \tag{49}$$

Since we are working with a finite sample, for $k \geq d > 0$,

$$\frac{\partial^d E[(r-\mu)^k]}{\partial {\omega_j}^d} = E\left[\left(\tilde{r}_j - \tilde{\mu}_j\right)^d\left(\sum_{i=1}^{I}\omega_i(\tilde{r}_i - \tilde{\mu}_i)\right)^{k-d}\right]\prod_{\delta=0}^{d-1}(k-\delta) \tag{50}$$

and $\frac{\partial^d E[(r-\mu)^k]}{\partial \omega_j{}^d} = 0$ for $d > k$.

A.4.3. SECOND DERIVATIVE

Following up with the previous results, we would like to compute $\frac{\partial^2 \widehat{Z^*}}{\partial \omega_j{}^2}$:

$$\frac{\partial^2 \widehat{Z^*}}{\partial \omega_j{}^2} = \frac{-2}{\hat{\sigma}_{SR}^3}\left(\frac{\partial \widehat{SR}}{\partial \omega_j}\hat{\sigma}_{SR} - \frac{\partial \hat{\sigma}_{SR}}{\partial \omega_j}\left(\widehat{SR} - SR^*\right)\right) + \frac{1}{\hat{\sigma}_{SR}^2}\left(\frac{\partial^2 \widehat{SR}}{\partial \omega_j{}^2}\hat{\sigma}_{SR} + \frac{\partial \hat{\sigma}_{SR}}{\partial \omega_j}\frac{\partial \widehat{SR}}{\partial \omega_j} - \frac{\partial^2 \hat{\sigma}_{SR}}{\partial \omega_j{}^2}\left(\widehat{SR} - SR^*\right) - \frac{\partial \widehat{SR}}{\partial \omega_j}\frac{\partial \hat{\sigma}_{SR}}{\partial \omega_j}\right) \tag{51}$$

So we still need to calculate the expressions $\frac{\partial^2 \widehat{SR}}{\partial \omega_j{}^2}$ and $\frac{\partial^2 \hat{\sigma}_{SR}}{\partial \omega_j{}^2}$.

$$\frac{\partial^2 \widehat{SR}}{\partial \omega_j{}^2} = -2\sigma^{-3}\left(\frac{\partial \mu}{\partial \omega_j}\sigma - \frac{\partial \sigma}{\partial \omega_j}\mu\right) + \left(\frac{\partial^2 \mu}{\partial \omega_j{}^2}\sigma - \frac{\partial^2 \sigma}{\partial \omega_j{}^2}\mu\right)\sigma^{-2} \tag{52}$$

$$\begin{aligned}\frac{\partial^2 \hat{\sigma}_{SR}}{\partial \omega_j{}^2} = &-\frac{1}{2\hat{\sigma}_{SR}^2(n-1)}\frac{\partial \hat{\sigma}_{SR}}{\partial \omega_j}\left(-\frac{\partial \hat{\gamma}_3}{\partial \omega_j}\widehat{SR} - \frac{\partial \widehat{SR}}{\partial \omega_j}\hat{\gamma}_3 + \frac{1}{4}\left(\frac{\partial \hat{\gamma}_4}{\partial \omega_j}\widehat{SR}^2 + 2\widehat{SR}\frac{\partial \widehat{SR}}{\partial \omega_j}(\hat{\gamma}_4 - 1)\right)\right) \\ &+ \frac{1}{2\hat{\sigma}_{SR}(n-1)}\left(-\frac{\partial^2 \hat{\gamma}_3}{\partial \omega_j{}^2}\widehat{SR} - 2\frac{\partial \hat{\gamma}_3}{\partial \omega_j}\frac{\partial \widehat{SR}}{\partial \omega_j} - \frac{\partial^2 \widehat{SR}}{\partial \omega_j{}^2}\hat{\gamma}_3 \right. \\ &\left. + \frac{1}{4}\left(\frac{\partial^2 \hat{\gamma}_4}{\partial \omega_j{}^2}\widehat{SR}^2 + 4\widehat{SR}\frac{\partial \widehat{SR}}{\partial \omega_j}\frac{\partial \hat{\gamma}_4}{\partial \omega_j} + 2(\hat{\gamma}_4 - 1)\left(\left(\frac{\partial \widehat{SR}}{\partial \omega_j}\right)^2 + \frac{\partial^2 \widehat{SR}}{\partial \omega_j{}^2}\widehat{SR}\right)\right)\right)\end{aligned} \tag{53}$$

which requires us to derive $\frac{\partial^2 \hat{\gamma}_3}{\partial \omega_j{}^2}$, $\frac{\partial^2 \hat{\gamma}_4}{\partial \omega_j{}^2}$, $\frac{\partial^2 \mu}{\partial \omega_j{}^2}$, $\frac{\partial^2 \sigma}{\partial \omega_j{}^2}$:

$$\frac{\partial^2\hat{\gamma}_3}{\partial\omega_j{}^2} = -6\sigma^{-7}\frac{\partial\sigma}{\partial\omega_j}\left(\frac{\partial E[(r-\mu)^3]}{\partial\omega_j}\sigma^3 - 3\sigma^2\frac{\partial\sigma}{\partial\omega_j}E[(r-\mu)^3]\right) + \sigma^{-6}\left(\frac{\partial^2 E[(r-\mu)^3]}{\partial\omega_j{}^2}\sigma^3 + \frac{\partial E[(r-\mu)^3]}{\partial\omega_j}3\sigma^2\frac{\partial\sigma}{\partial\omega_j} - 3\left(2\sigma\left(\frac{\partial\sigma}{\partial\omega_j}\right)^2 + \frac{\partial^2\sigma}{\partial\omega_j{}^2}\sigma^2\right)E[(r-\mu)^3] - 3\frac{\partial E[(r-\mu)^3]}{\partial\omega_j}\sigma^2\frac{\partial\sigma}{\partial\omega_j}\right) \tag{54}$$

$$\frac{\partial^2\hat{\gamma}_4}{\partial\omega_j{}^2} = -8\sigma^{-9}\frac{\partial\sigma}{\partial\omega_j}\left(\frac{\partial E[(r-\mu)^4]}{\partial\omega_j}\sigma^4 - 4\sigma^3\frac{\partial\sigma}{\partial\omega_j}E[(r-\mu)^4]\right) + \sigma^{-8}\left(\frac{\partial^2 E[(r-\mu)^4]}{\partial\omega_j{}^2}\sigma^4 + \frac{\partial E[(r-\mu)^4]}{\partial\omega_j}4\sigma^3\frac{\partial\sigma}{\partial\omega_j} - 4\left(3\sigma^2\left(\frac{\partial\sigma}{\partial\omega_j}\right)^2 + \frac{\partial^2\sigma}{\partial\omega_j{}^2}\sigma^3\right)E[(r-\mu)^4] - 4\frac{\partial E[(r-\mu)^4]}{\partial\omega_j}\sigma^3\frac{\partial\sigma}{\partial\omega_j}\right) \tag{55}$$

$$\frac{\partial^2\mu}{\partial\omega_j{}^2} = 0 \tag{56}$$

$$\frac{\partial^2\sigma}{\partial\omega_j{}^2} = \frac{-1}{2\sigma^2}\frac{\partial E[(r-\mu)^2]}{\partial\omega_j} + \frac{1}{2\sigma}\frac{\partial^2 E[(r-\mu)^2]}{\partial\omega_j{}^2} \tag{57}$$

A.4.4. STEP SIZE

Finally, assuming $\sum_{k=3}^{\infty}\frac{1}{k!}\frac{\partial^k\widehat{Z^*}(\omega_i)}{\partial\omega_i{}^k}(\Delta\omega_i)^k \approx 0$, we can replace these derivatives into Taylor's expansion:

$$\Delta\widehat{Z^*} = \widehat{Z^*}(\omega_i + \Delta\omega_i) - \widehat{Z^*}(\omega_i) = \frac{\partial\widehat{Z^*}(\omega_i)}{\partial\omega_i}\Delta\omega_i + \frac{1}{2}\frac{\partial^2\widehat{Z^*}(\omega_i)}{\partial\omega_i{}^2}(\Delta\omega_i)^2 \tag{58}$$

Let's define

$$a = \frac{1}{2}\frac{\partial^2\widehat{Z^*}(\omega_i)}{\partial\omega_i{}^2}$$
$$b = \frac{\partial\widehat{Z^*}(\omega_i)}{\partial\omega_i} \tag{59}$$
$$c = -\Delta\widehat{Z^*}$$

Then, for $a \neq 0$ we will choose the smallest step size (to reduce the error due to Taylor's approximation, which grows with $|\Delta\omega_i|$):

$$\Delta\omega_i = \begin{cases} \dfrac{-\dfrac{\partial\widehat{Z^*}(\omega_i)}{\partial\omega_i} + \sqrt{\left(\dfrac{\partial\widehat{Z^*}(\omega_i)}{\partial\omega_i}\right)^2 + 2\dfrac{\partial^2\widehat{Z^*}(\omega_i)}{\partial{\omega_i}^2}\Delta\widehat{Z^*}}}{\dfrac{\partial^2\widehat{Z^*}(\omega_i)}{\partial{\omega_i}^2}} & if\ b \geq 0 \\ -\dfrac{\dfrac{\partial\widehat{Z^*}(\omega_i)}{\partial\omega_i} + \sqrt{\left(\dfrac{\partial\widehat{Z^*}(\omega_i)}{\partial\omega_i}\right)^2 + 2\dfrac{\partial^2\widehat{Z^*}(\omega_i)}{\partial{\omega_i}^2}\Delta\widehat{Z^*}}}{\dfrac{\partial^2\widehat{Z^*}(\omega_i)}{\partial{\omega_i}^2}} & if\ b < 0 \end{cases} \tag{60}$$

For $a = 0$, the solution coincides with a first degree Taylor approximation:

$$\Delta\omega_i = \frac{\Delta\widehat{Z^*}}{\dfrac{\partial\widehat{Z^*}(\omega_i)}{\partial\omega_i}} \tag{61}$$

A.4.5. IMPLEMENTATION OF A PSR PORTFOLIO OPTIMIZATION

We can use the equations derived earlier to develop a PSR portfolio optimization algorithm. The example that follows is coded in Python. It uses a gradient-ascent logic to determine the ω that maximizes $\widehat{Z^*} = \frac{\widehat{SR}-SR^*}{\hat{\sigma}_{SR}}$, subject to the condition that $\sum_{i=1}^{I}\omega_i = 1$, as enunciated in Section 7. Gradient-ascent only requires the first derivative, so in this particular implementation we are not making use of our calculated $\frac{\partial^2\widehat{Z^*}}{\partial{\omega_j}^2}$. The solution reported in Section 7 is reached after only 118 iterations.

The user can specify boundary conditions using the variable *bounds*, in the *main()* function. By default, weights are set to be bounded between 0 and 1.

```
#!/usr/bin/env python
# PSR class for Portfolio Optimization
# On 20120502 by MLdP <lopezdeprado@lbl.gov>

import numpy as np
#-------------------------------------------
#-------------------------------------------
class PSR_Opt:
    def __init__(self,series,seed,delta,maxIter,bounds=None):
        # Construct the object
        self.series,self.w,self.delta=series,seed,delta
        self.z,self.d1Z=None,[None for i in range(series.shape[1])]
        self.maxIter,self.iter,self.obs=maxIter,0,series.shape[0]
        if len(bounds)==None or seed.shape[0]!=len(bounds):
            self.bounds=[(0,1) for i in seed]
```

```
        else:
            self.bounds=bounds
#------------------------------------------
    def optimize(self):
        # Optimize weights
        mean=[self.get_Moments(self.series[:,i],1) for i in range(self.series.shape[1])]
        w=np.array(self.w)
        # Compute derivatives
        while True:
            if self.iter==self.maxIter:break
            # Compute gradient
            d1Z,z=self.get_d1Zs(mean,w)
            # Evaluate result
            if z>self.z and self.checkBounds(w)==True:
                # Store new local optimum
                self.z,self.d1Z=z,d1Z
                self.w=np.array(w)
            # Find direction and normalize
            self.iter+=1
            w=self.stepSize(w,d1Z)
            if w==None:return
        return
#------------------------------------------
    def checkBounds(self,w):
        # Check that boundary conditions are satisfied
        flag=True
        for i in range(w.shape[0]):
            if w[i,0]<self.bounds[i][0]:flag=False
            if w[i,0]>self.bounds[i][1]:flag=False
        return flag
#------------------------------------------
    def stepSize(self,w,d1Z):
        # Determine step size for next iteration
        x={}
        for i in range(len(d1Z)):
            if d1Z[i]!=0:x[abs(d1Z[i])]=i
        if len(x)==0:return
        index=x[max(x)]
        w[index,0]+=self.delta/d1Z[index]
        w/=sum(w)
        return w
#------------------------------------------
    def get_d1Zs(self,mean,w):
        # First order derivatives of Z
        d1Z=[0 for i in range(self.series.shape[1])]
        m=[0 for i in range(4)]
        series=np.dot(self.series,w)[:,0]
        m[0]=self.get_Moments(series,1)
        for i in range(1,4):m[i]=self.get_Moments(series,i+1,m[0])
        stats=self.get_Stats(m)
        meanSR,sigmaSR=self.get_SR(stats,self.obs)
        for i in range(self.series.shape[1]):
            d1Z[i]=self.get_d1Z(stats,m,meanSR,sigmaSR,mean,w,i)
        return d1Z,meanSR/sigmaSR
#------------------------------------------
    def get_d1Z(self,stats,m,meanSR,sigmaSR,mean,w,index):
```

```
        # First order derivatives of Z with respect to index
        d1Mu=self.get_d1Mu(mean,index)
        d1Sigma=self.get_d1Sigma(stats[1],mean,w,index)
        d1Skew=self.get_d1Skew(d1Sigma,stats[1],mean,w,index,m[2])
        d1Kurt=self.get_d1Kurt(d1Sigma,stats[1],mean,w,index,m[3])
        d1meanSR=(d1Mu*stats[1]-d1Sigma*stats[0])/stats[1]**2
        d1sigmaSR=(d1Kurt*meanSR**2+2*meanSR*d1meanSR*(stats[3]-1))/4
        d1sigmaSR-=d1Skew*meanSR+d1meanSR*stats[2]
        d1sigmaSR/=2*sigmaSR*(self.obs-1)
        d1Z=(d1meanSR*sigmaSR-d1sigmaSR*meanSR)/sigmaSR**2
        return d1Z
#-------------------------------------------
    def get_d1Mu(self,mean,index):
        # First order derivative of Mu
        return mean[index]
#-------------------------------------------
    def get_d1Sigma(self,sigma,mean,w,index):
        # First order derivative of Sigma
        return self.get_dnMoments(mean,w,2,1,index)/(2*sigma)
#-------------------------------------------
    def get_d1Skew(self,d1Sigma,sigma,mean,w,index,m3):
        # First order derivative of Skewness
        d1Skew=self.get_dnMoments(mean,w,3,1,index)*sigma**3
        d1Skew-=3*sigma**2*d1Sigma*m3
        d1Skew/=sigma**6
        return d1Skew
#-------------------------------------------
    def get_d1Kurt(self,d1Sigma,sigma,mean,w,index,m4):
        # First order derivative of Kurtosis
        d1Kurt=self.get_dnMoments(mean,w,4,1,index)*sigma**4
        d1Kurt-=4*sigma**3*d1Sigma*m4
        d1Kurt/=sigma**8
        return d1Kurt
#-------------------------------------------
    def get_dnMoments(self,mean,w,mOrder,dOrder,index):
        # Get dOrder derivative on mOrder mean-centered moment with respect to w index
        x0,sum=1.,0
        for i in range(dOrder):x0*=(mOrder-i)
        for i in self.series:
            x1,x2=0,(i[index]-mean[index])**dOrder
            for j in range(len(i)):x1+=w[j,0]*(i[j]-mean[j])
            sum+=x2*x1**(mOrder-dOrder)
        return x0*sum/self.obs
#-------------------------------------------
    def get_SR(self,stats,n):
        # Set Z*
        meanSR=stats[0]/stats[1]
        sigmaSR=((1-meanSR*stats[2]+meanSR**2*(stats[3]-1)/4.)/(n-1))**.5
        return meanSR,sigmaSR
#-------------------------------------------
    def get_Stats(self,m):
        # Compute stats
        return [m[0],m[1]**.5,m[2]/m[1]**(3/2.),m[3]/m[1]**2]
#-------------------------------------------
    def get_Moments(self,series,order,mean=0):
        # Compute a moment
```

```
        sum=0
        for i in series:sum+=(i-mean)**order
        return sum/float(self.obs)
#-------------------------------------------
#-------------------------------------------
def main():
    #1) Inputs (path to csv file with returns series)
    path='H:/ TimeSeries.csv'
    maxIter=1000 # Maximum number of iterations
    delta=.005 # Delta Z (attempted gain per interation)

    #2) Load data, set seed
    series=np.genfromtxt(path,delimiter=',') # load as numpy array
    seed=np.ones((series.shape[1],1))/series.shape[1] # initialize seed
    bounds=[(0,1) for i in seed] # min and max boundary per weight

    #3) Create class and solve
    psrOpt=PSR_Opt(series,seed,delta,maxIter,bounds)
    psrOpt.optimize()

    #4) Optimize and report optimal portfolio
    print 'Maximized Z-value: '+str(psrOpt.z)
    print '# of iterations: '+str(psrOpt.iter)
    print 'PSR optimal portfolio:'
    print str(psrOpt.w)
#-------------------------------------------
# Boilerplate
if __name__=='__main__': main()
```

FIGURES

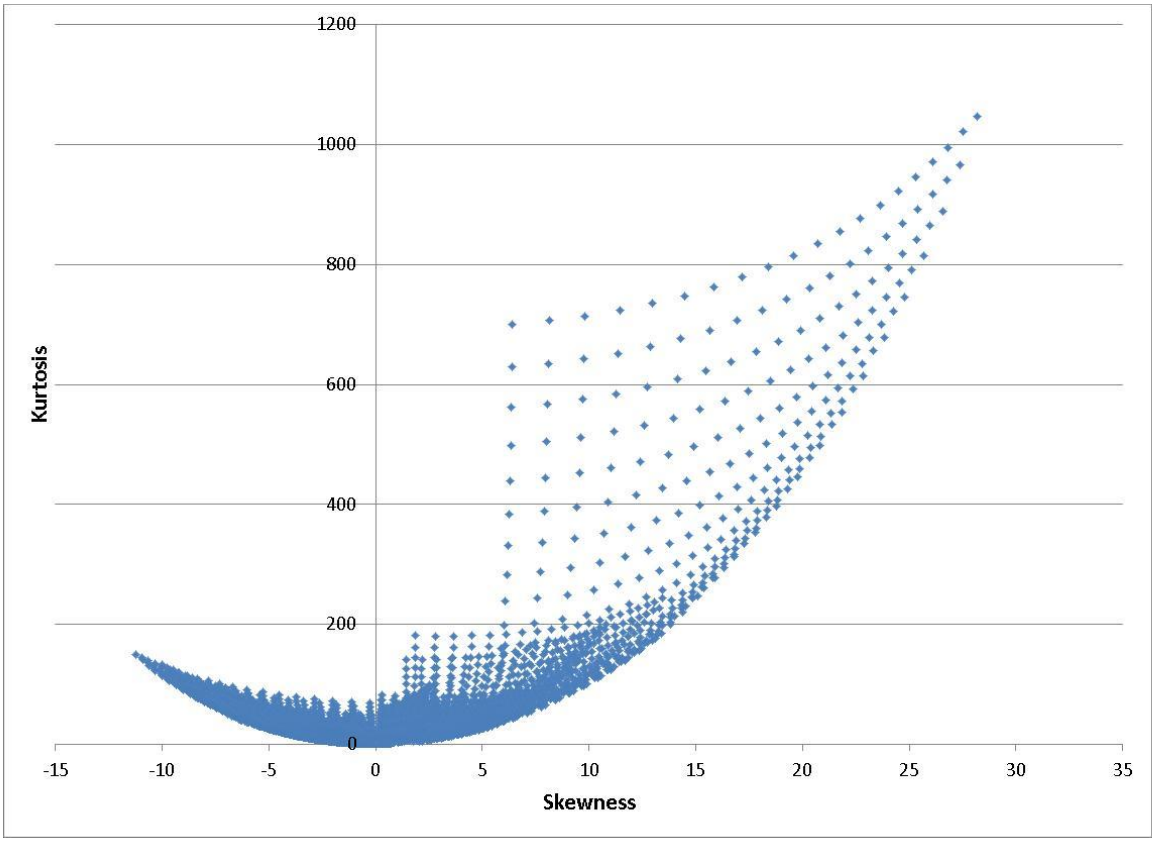

Figure 1 – Combinations of skewness and kurtosis from Mixtures of two Gaussians with the same Sharpe ratio ($SR^* = 1$)

An infinite number of mixtures of two Gaussians can deliver any given *SR*, despite of having widely different levels of skewness and kurtosis. This is problematic, because high readings of *SR* may come from extremely risky distributions, like combinations on the left side of this figure (negative skewness and positive kurtosis).

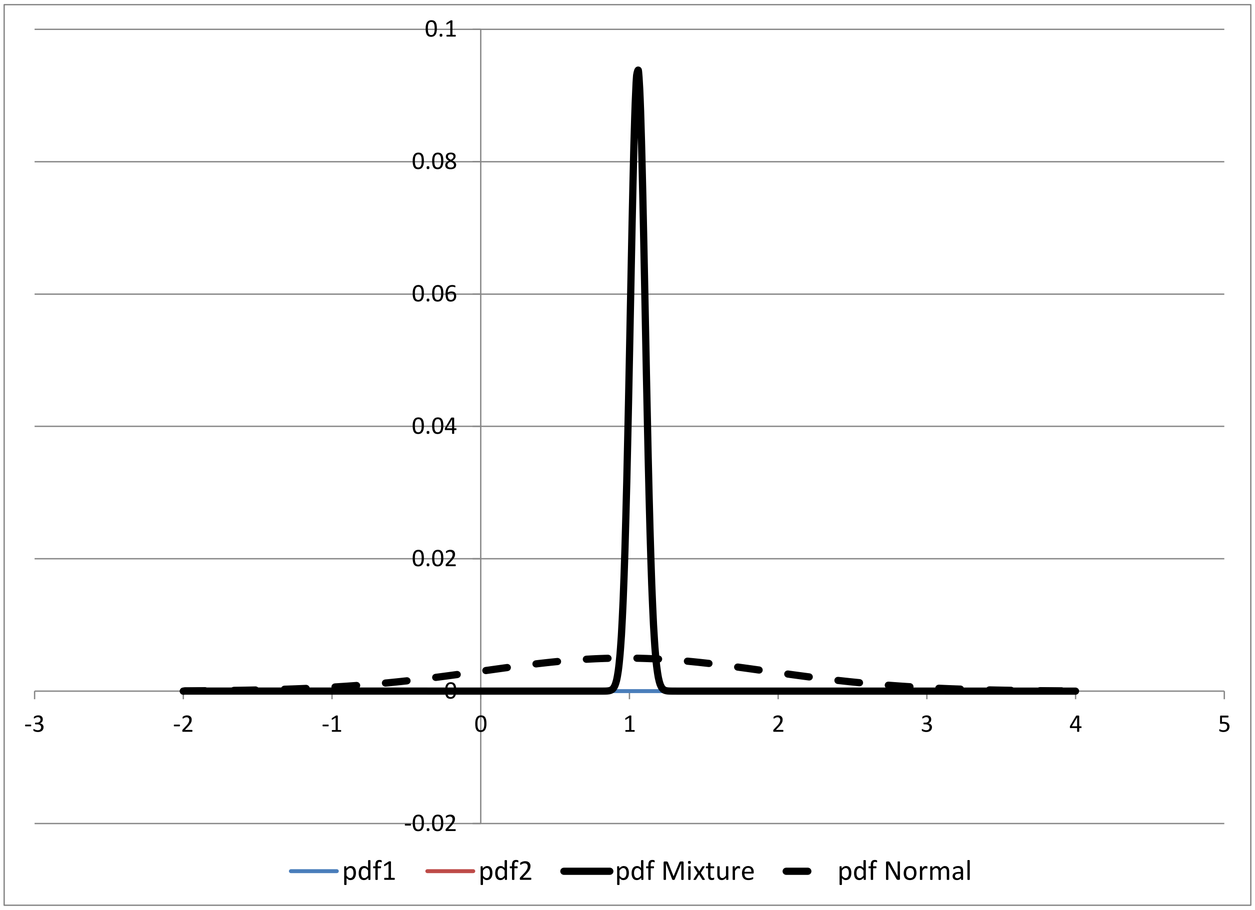

Figure 2(a) – Probability density function for a Mixture of two Gaussians with parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, p) = (-5, 1.05, 5, 0.05, 0.015)$

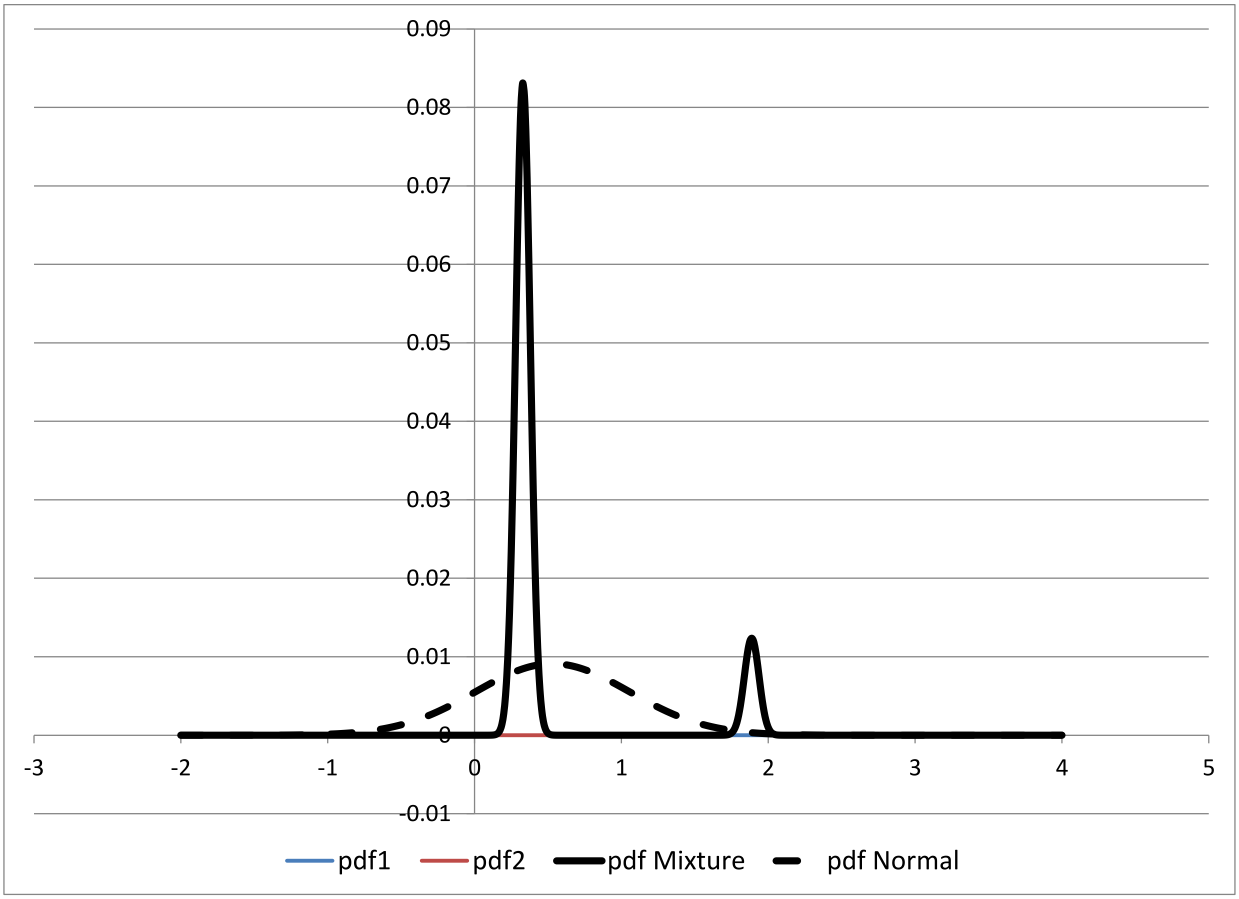

Figure 2(b) – Probability density function for a Mixture of two Gaussians with parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, p) = (0.3237, 1.8816, 0.05, 0.05, 0.8706)$

These two distributions were drawn from the combinations plotted in Figure 1. Both have a Sharpe ratio of 1, despite of their evidently different risk profile. The dashed black line represents the probability distribution function of a Normal distribution fitted of each of these mixtures. The variance not only underestimates non-Normal risks, but its own estimator is greatly affected by non-Normality. A minimal change in the mixture's parameters could have a great impact on the estimated value of the mixture's variance.

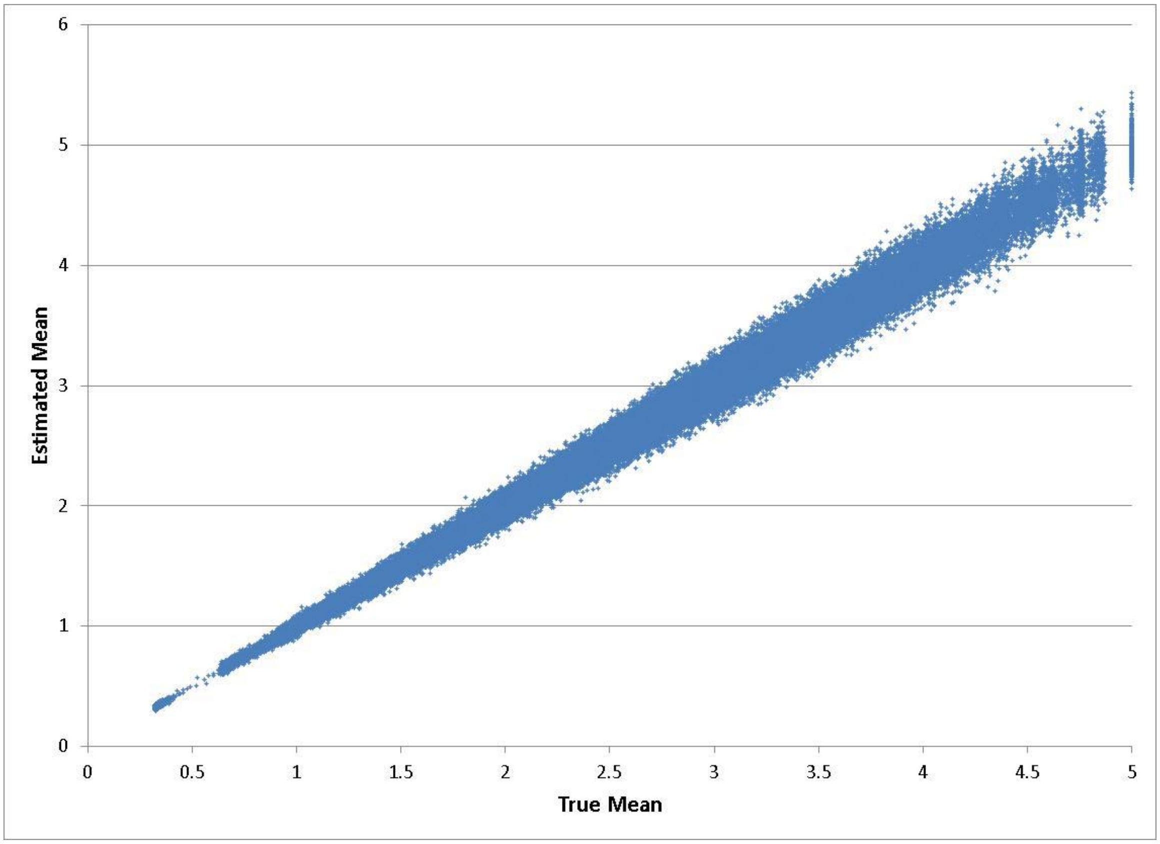

Figure 3(a) – True vs. estimated mean

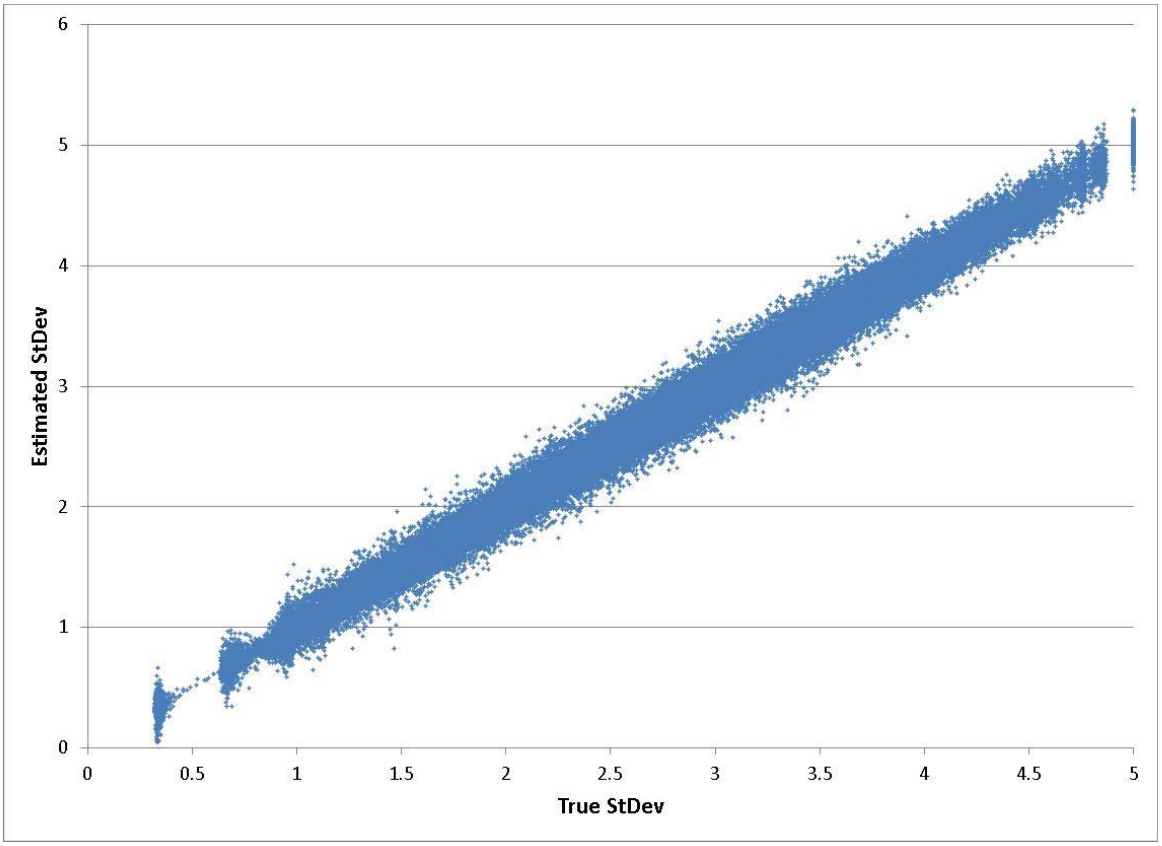

Figure 3(b) – True vs. estimated standard deviation

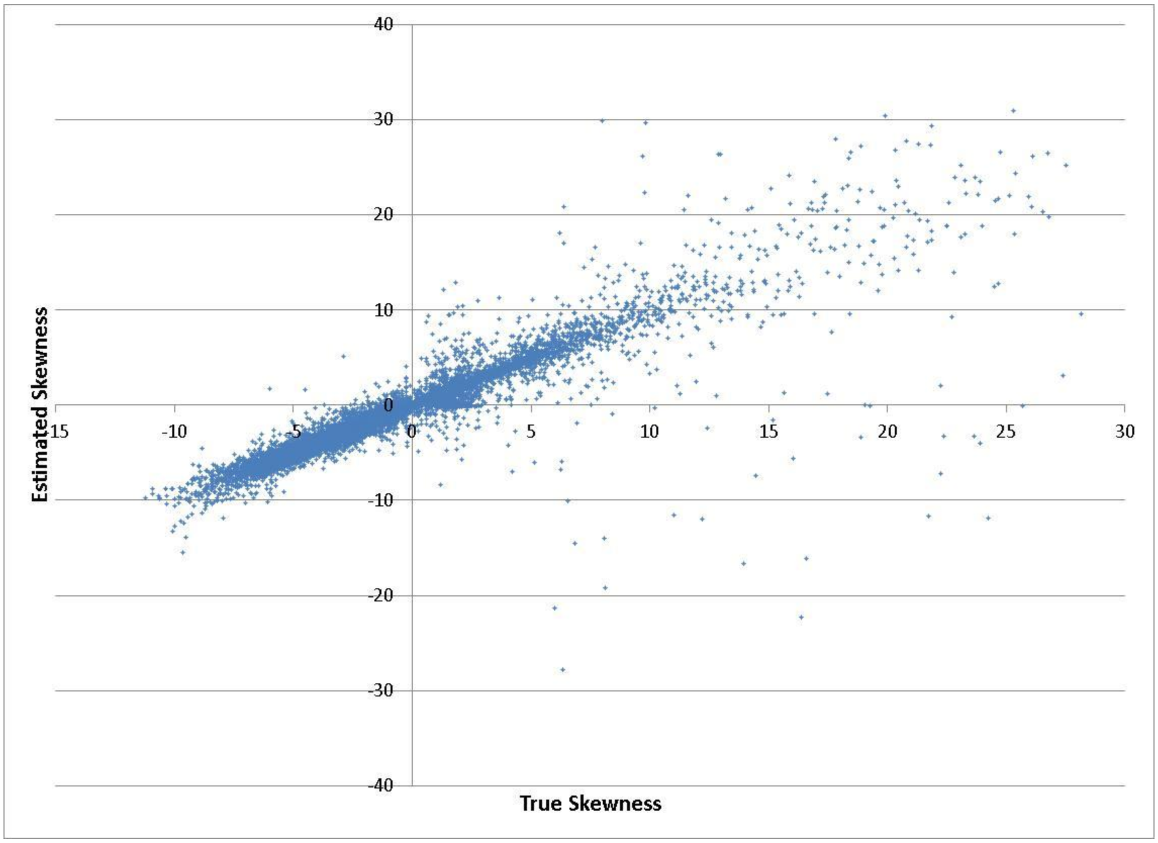

Figure 3(c) – True vs. estimated skewness

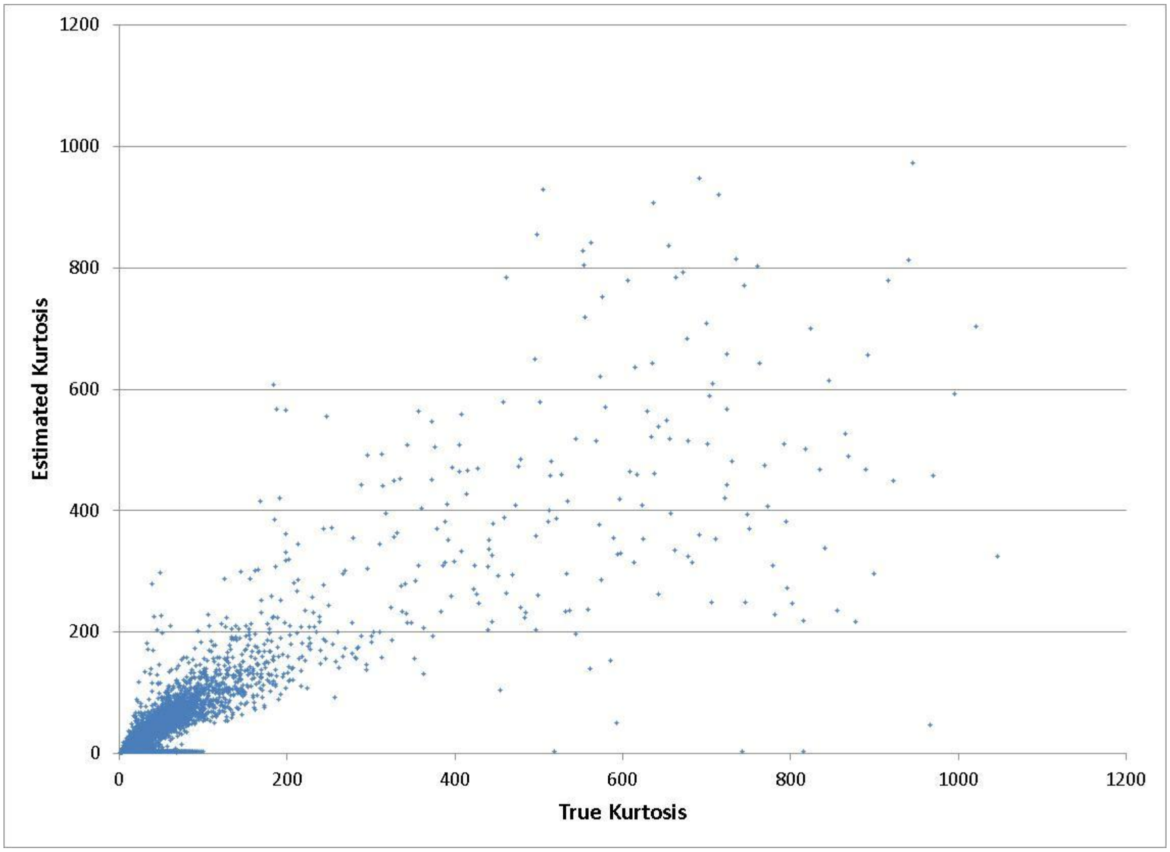

Figure 3(d) – True vs. estimated kurtosis

Estimations errors increase with higher moments, requiring longer sample sizes.

MEAN					STDEV				
Degree	**er_δ**	**er_Prob**	**er2_θ**	**er2_Prob**	**Degree**	**er_δ**	**er_Prob**	**er2_θ**	**er2_Prob**
0	-0.0016	0.3756	0.0001	0.7863	0	0.0048	0.0134	0.0049	0.0000
1	0.0013	0.3738	-0.0001	0.7877	1	-0.0014	0.3687	0.0017	0.0000
2	-0.0002	0.4007	0.0010	0.0000	2	0.0002	0.4276	0.0000	0.3292
SKEW					**KURT**				
Degree	**er_δ**	**er_Prob**	**er2_θ**	**er2_Prob**	**Degree**	**er_δ**	**er_Prob**	**er2_θ**	**er2_Prob**
0	0.0063	0.0115	0.3421	0.0000	0	0.3532	0.0000	269.3399	0.0000
1	0.0300	0.0000	0.7558	0.0000	1	-0.0342	0.0000	-37.2576	0.0000
2	0.0071	0.0000	0.1979	0.0000	2	0.0005	0.0000	0.2841	0.0000

adj-R2	er	er2
Mean	0.0000	0.1233
StDev	0.0000	0.0063
Skew	0.0685	0.1196
Kurt	0.3937	0.4904

Figure 4 –Estimation error models for various moments and levels

If we draw samples from random mixtures of two Gaussians, we can study how the estimation errors on their moments are affected by the moment's values.

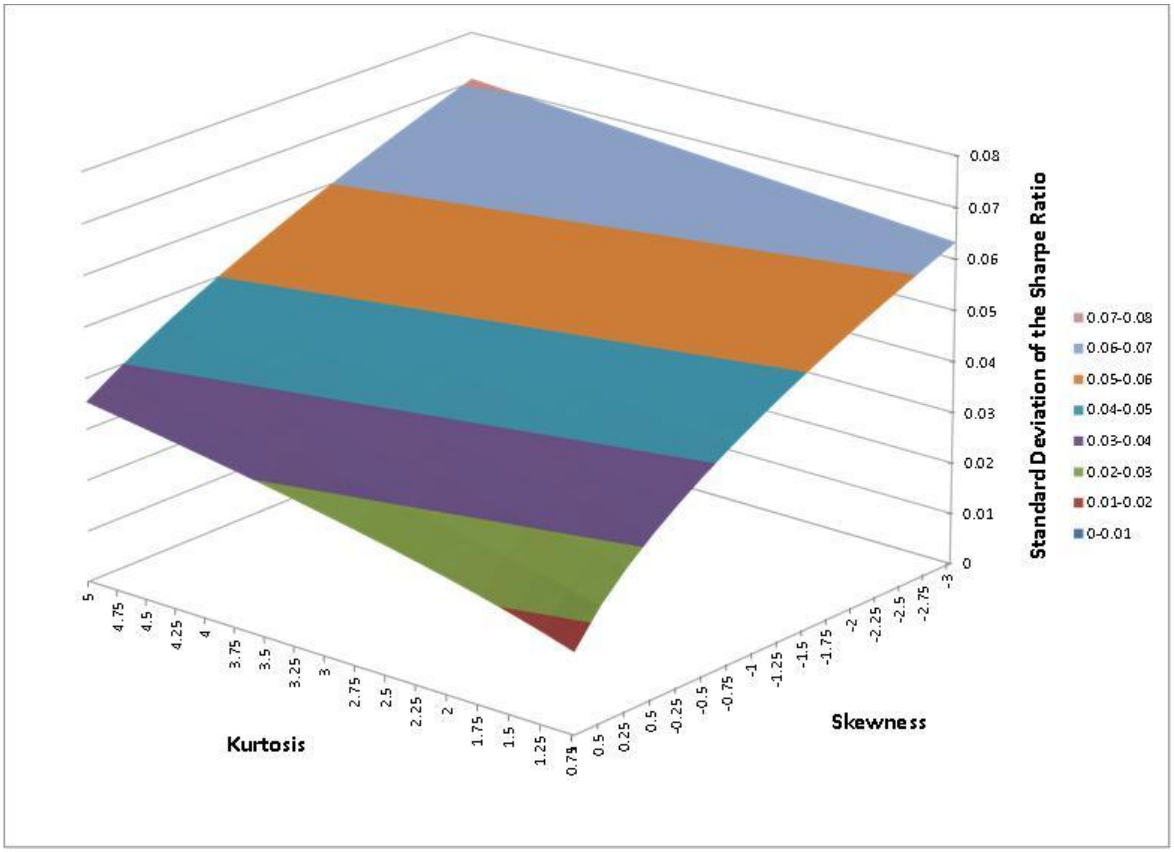

Figure 5 - $\sigma_{\widehat{SR}}$ as a function of (γ_3, γ_4), with n=1000, SR=1

The standard deviation of the SR estimator is sensitive to skewness and kurtosis. For *SR=1*, we see that $\sigma_{\widehat{SR}}$ is particularly sensitive to skewness, as we could expect from inspecting Eq. (8).

Stats	Values
Mean	0.036
StDev	0.079
Skew	-2.448
Kurt	10.164
SR	0.458
Ann. SR	1.585

Figure 6 – Hedge fund track record statistics

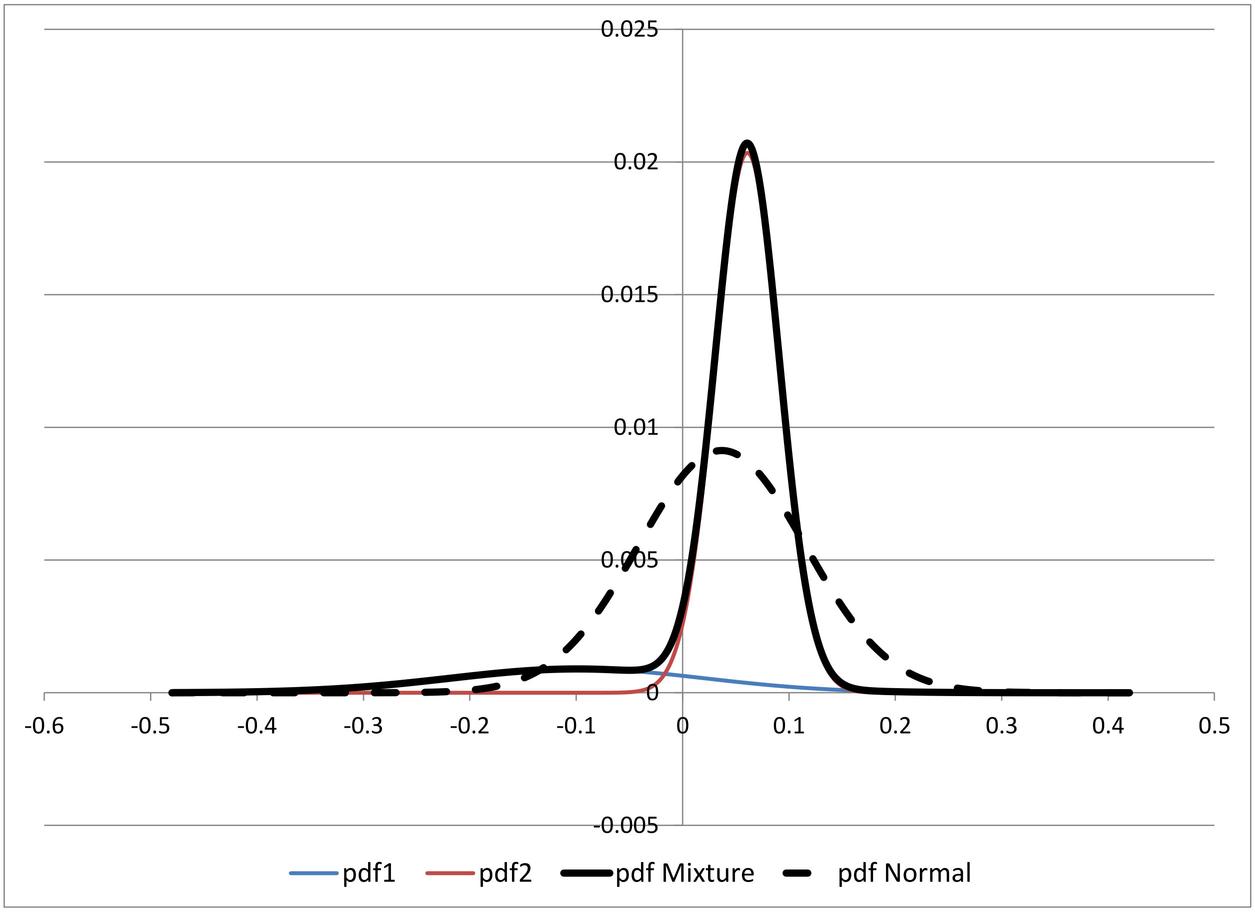

Figure 7 – Probability distributions assuming Normality (dashed black line) and considering non-Normality (black line)

This mixture of two Gaussians exactly matches the moments reported in Figure 6. The dash line shows that a Normal fit severely underestimates the downside risks for this portfolio manager. Moreover, there is a significant probability that this portfolio manager may have no investment skill, despite of having produced an annualized Sharpe ratio close to 1.6.

		True Sharpe Ratio									
		0	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**
Observed Sharpe	**0**										
	0.5	10.83									
	1	2.71	10.85								
	1.5	1.21	2.72	10.87							
	2	0.69	1.22	**2.73**	10.91						
	2.5	0.44	0.69	1.22	2.74	10.96					
	3	0.31	0.44	0.69	1.23	2.76	11.02				
	3.5	0.23	0.31	0.45	0.70	1.24	2.78	11.09			
	4	0.18	0.23	0.31	0.45	0.70	1.24	2.80	11.17		
	4.5	0.14	0.18	0.23	0.32	0.45	0.71	1.25	2.82	11.26	
	5	0.12	0.14	0.18	0.24	0.32	0.46	0.71	1.27	2.84	11.36

Figure 8 – Minimum track record in years, under daily IID Normal returns

	True Sharpe Ratio									
Observed Sharpe	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**
0										
0.5	10.87									
1	2.75	10.95								
1.5	1.25	2.78	11.08							
2	0.72	1.27	**2.83**	11.26						
2.5	0.48	0.74	1.29	2.89	11.49					
3	0.35	0.49	0.75	1.33	2.96	11.78				
3.5	0.27	0.36	0.50	0.78	1.36	3.04	12.12			
4	0.21	0.27	0.37	0.52	0.80	1.41	3.14	12.51		
4.5	0.18	0.22	0.28	0.38	0.54	0.83	1.46	3.25	12.95	
5	0.15	0.18	0.23	0.29	0.39	0.56	0.86	1.51	3.38	13.44

Figure 9 – Minimum track record in years, under weekly IID Normal returns

	True Sharpe Ratio									
Observed Sharpe	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**
0										
0.5	11.02									
1	2.90	11.36								
1.5	1.40	3.04	11.92							
2	0.87	1.49	**3.24**	12.71						
2.5	0.63	0.94	1.60	3.49	13.72					
3	0.50	0.68	1.01	1.74	3.80	14.96				
3.5	0.42	0.54	0.74	1.10	1.90	4.17	16.43			
4	0.37	0.45	0.58	0.80	1.21	2.09	4.59	18.12		
4.5	0.33	0.40	0.49	0.64	0.88	1.33	2.30	5.07	20.04	
5	0.30	0.36	0.43	0.53	0.70	0.97	1.46	2.54	5.61	22.18

Figure 10 – Minimum track record in years, under monthly IID Normal returns

	True Sharpe Ratio									
Observed Sharpe	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**
0										
0.5	12.30									
1	3.62	14.23								
1.5	1.93	4.24	16.70							
2	1.31	2.26	**4.99**	19.72						
2.5	1.01	1.53	2.66	5.88	23.26					
3	0.84	1.17	1.79	3.11	6.90	27.35				
3.5	0.73	0.97	1.36	2.08	3.63	8.06	31.98			
4	0.66	0.84	1.11	1.57	2.40	4.20	9.35	37.15		
4.5	0.61	0.75	0.96	1.27	1.79	2.76	4.84	10.78	42.85	
5	0.57	0.69	0.85	1.08	1.44	2.04	3.15	5.53	12.34	49.09

Figure 11 – Minimum track record in years, under monthly IID returns with $\hat{\gamma}_3 = -0.72$ and $\hat{\gamma}_4 = 5.78$

HFR Index	Code	SR	StDev(SR)	An. SR	Low An. SR	PSR(0)	PSR(0.5)	MinTRL (0)	MinTRL (0.5)
Conserv	HFRIFOFC Index	0.251	0.116	0.871	0.210	0.985	0.822	6.456	35.243
Conv Arbit	HFRICAI Index	0.253	0.124	0.875	0.170	0.979	0.809	7.282	39.246
Dist Secur	HFRIDSI Index	0.414	0.116	1.433	0.771	1.000	0.990	2.448	5.661
Divers	HFRIFOFD Index	0.208	0.099	0.719	0.158	0.982	0.740	6.841	72.870
EM Asia	HFRIEMA Index	0.200	0.092	0.691	0.168	0.985	0.726	6.423	82.857
EM Global	HFRIEMG Index	0.258	0.100	0.892	0.325	0.995	0.872	4.559	23.242
EM Latin Amer	HFRIEMLA Index	0.173	0.093	0.598	0.068	0.968	0.620	8.782	323.473
Emerg Mkt	HFRIEM Index	0.259	0.100	0.896	0.324	0.995	0.873	4.602	23.214
Equity Hedge	HFRIEHI Index	0.196	0.092	0.681	0.158	0.984	0.715	6.608	92.752
Equity Neutral	HFRIEMNI Index	0.413	0.099	1.432	0.866	1.000	0.997	1.817	4.176
Event Driven	HFRIEDI Index	0.348	0.108	1.205	0.589	0.999	0.970	2.982	8.548
Fixed Asset-Back	HFRIFIMB Index	0.657	0.153	2.276	1.405	1.000	1.000	1.706	2.749
Fixed Hig	HFRIFIHY Index	0.283	0.120	0.980	0.294	0.991	0.875	5.513	22.716
Fund of Funds	HFRIFOF Index	0.213	0.099	0.739	0.174	0.984	0.757	6.560	61.984
Macro	HFRIMI Index	0.381	0.087	1.320	0.824	1.000	0.997	1.649	4.138
Mkt Defens	HFRIFOFM Index	0.388	0.087	1.343	0.847	1.000	0.997	1.596	3.922
Mrg Arbit	HFRIMAI Index	0.496	0.112	1.717	1.080	1.000	0.999	1.611	3.124
Multi-Strategy	HFRIFI Index	0.361	0.138	1.252	0.468	0.996	0.943	4.426	12.118
Priv/Regulation	HFRIREGD Index	0.225	0.082	0.780	0.312	0.997	0.837	4.083	31.061
Quant Direct	HFRIENHI Index	0.146	0.090	0.506	-0.005	0.948	0.508	11.400	77398.739
Relative Value	HFRIRVA Index	0.470	0.163	1.630	0.702	0.998	0.977	3.676	7.561
Russia-East Euro	HFRICIS Index	0.278	0.104	0.964	0.369	0.996	0.900	4.303	18.285
Sec Energy	HFRISEN Index	0.278	0.094	0.963	0.427	0.998	0.922	3.522	14.951
Sec Techno	HFRISTI Index	0.067	0.086	0.231	-0.261	0.780	0.184	50.420	n/a
Short Bias	HFRISHSE Index	0.043	0.086	0.148	-0.344	0.690	0.120	122.495	n/a
Strategic	HFRIFOFS Index	0.149	0.091	0.517	-0.004	0.949	0.521	11.348	10935.740
Sys Diversified	HFRIMTI Index	0.316	0.085	1.094	0.610	1.000	0.978	2.252	7.434
Wgt Comp	HFRIFWI Index	0.287	0.097	0.994	0.441	0.998	0.929	3.515	13.974
Wgt Comp CHF	HFRIFWIC Index	0.229	0.088	0.792	0.291	0.995	0.831	4.513	32.660
Wgt Comp GBP	HFRIFWIG Index	0.181	0.093	0.626	0.097	0.974	0.653	7.986	194.050
Wgt Comp GBP	HFRIFWIG Index	0.181	0.093	0.626	0.097	0.974	0.653	7.986	194.050
Wgt Comp JPY	HFRIFWIJ Index	0.167	0.090	0.580	0.065	0.968	0.601	8.805	459.523
Yld Alternative	HFRISRE Index	0.310	0.108	1.073	0.456	0.998	0.937	3.748	12.926

Figure 12 – Performance analysis on HFR Monthly indices

Only a few hedge fund investment styles evidence skill beyond a Sharpe ratio of 0.5 with a confidence level of 95%.

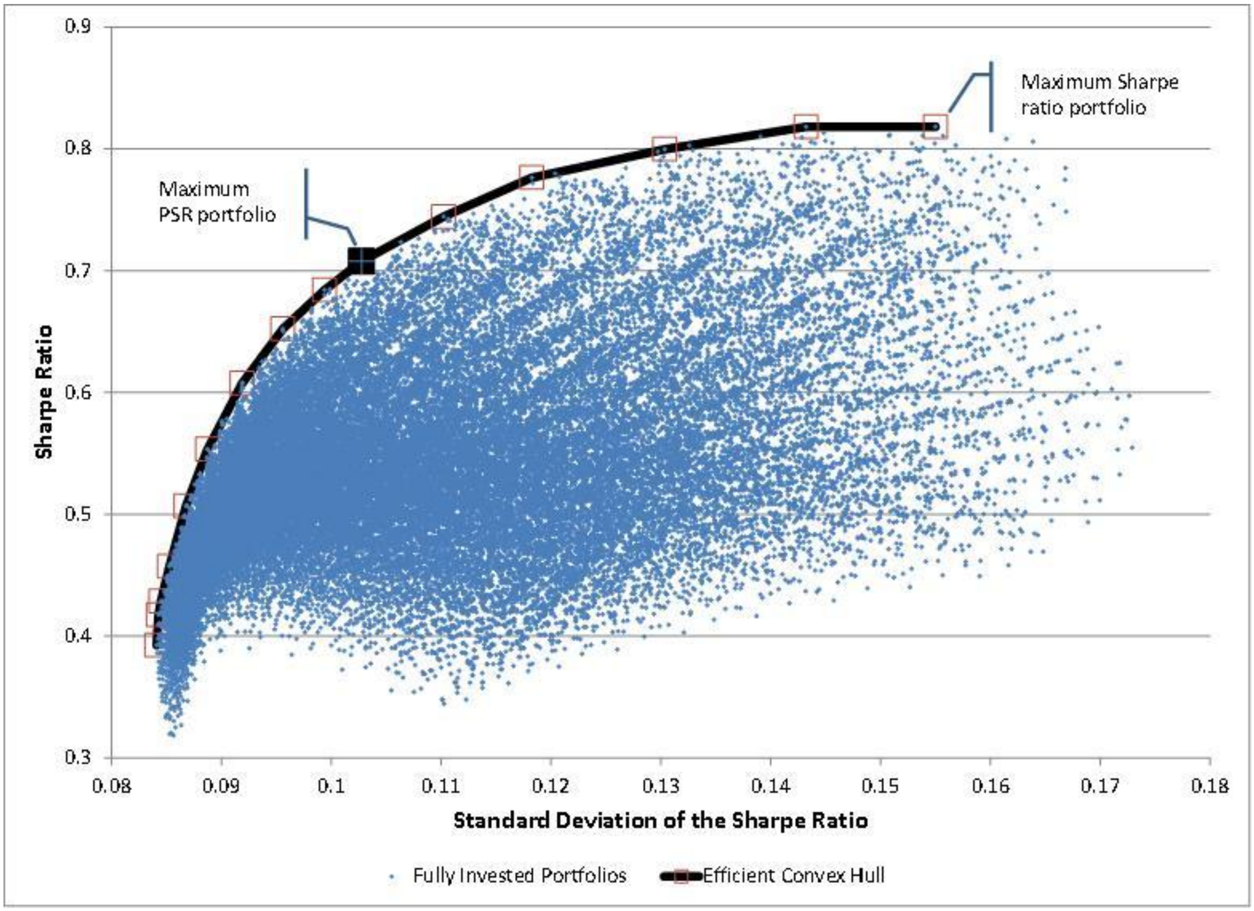

Figure 13 – The Sharpe ratio Efficient Frontier (SEF)

A Sharpe ratio Efficient Frontier can be derived in terms of optimal mean-variance combinations of risk-adjusted returns.

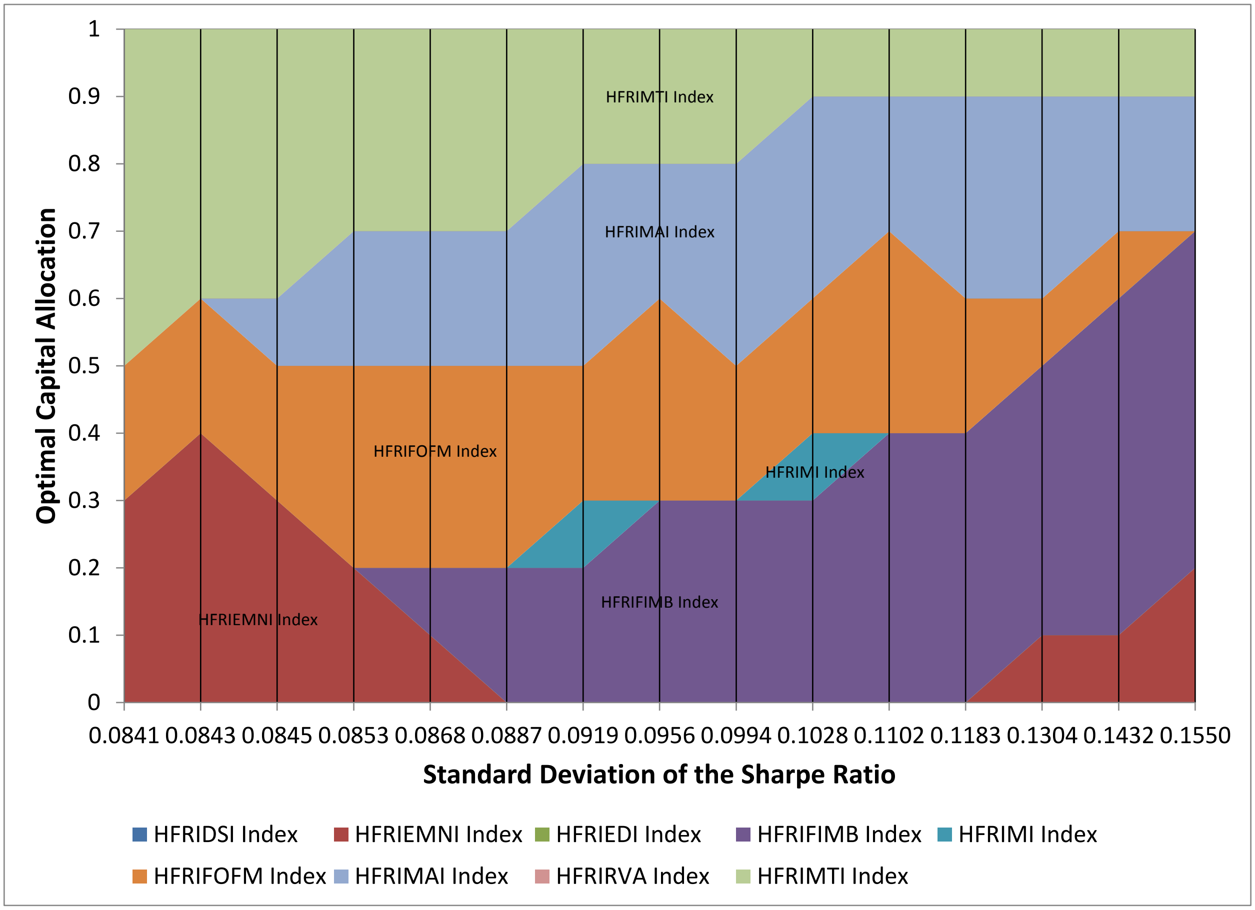

Figure 14 – Composition of the SEF for different $\hat{\sigma}_{\widehat{SR}(rw)}$ values

We can compute the capital allocations that deliver maximum Sharpe ratios for each confidence level. The difference with Markowitz's Efficient Frontier is that *SEF* is computed on risk-adjusted returns, rather than returns on capital

HFR Index	Code	Max PSR	Max SR
Dist Secur	HFRIDSI Index	0	0
Equity Neutral	HFRIEMNI Index	0	0.2
Event Driven	HFRIEDI Index	0	0
Fixed Asset-Back	HFRIFIMB Index	0.3	0.5
Macro	HFRIMI Index	0.1	0
Mkt Defens	HFRIFOFM Index	0.2	0
Mrg Arbit	HFRIMAI Index	0.3	0.2
Relative Value	HFRIRVA Index	0	0
Sys Diversified	HFRIMTI Index	0.1	0.1

Figure 15 – Composition of the Max PSR and Max SR portfolios

Stat	Max PSR	Max SR
Average	0.0061	0.0060
StDev	0.0086	0.0073
Skew	-0.2250	-1.4455
Kurt	2.9570	7.0497
Num	134	134
SR	0.7079	0.8183
StDev(SR)	0.1028	0.1550
An. SR	2.4523	2.8347
Low An. SR	1.8667	1.9515
PSR(0)	1.00000	1.00000
PSR(0.5)	1.00000	0.99999
MinTRL (0)	0.7152	1.1593
MinTRL (0.5)	1.0804	1.6695

Figure 16 – Stats of Max PSR and Max SR portfolios

Maximum *PSR* portfolios are risk-adjusted optimal, while maximum *SR* portfolios are risk-adjusted suboptimal. The reason is, although a maximum *SR* portfolio may be associated with a high expected Sharpe ratio (point estimate), the confidence bands around that expectation may be rather wide. Consequently, maximum *PSR* portfolios are distributed closer to a Normal, and demand a lower *MinTRL* than maximum *SR* portfolios.

Max PSR			Max SR		
Param.	**Dist.1**	**Dist.2**	**Param.**	**Dist.1**	**Dist.2**
Avg	-0.0118	0.0069	Avg	-0.0021	0.0077
StDev	0.0027	0.0078	StDev	0.0111	0.0047
Prob	0.0451	0.9549	Prob	0.1740	0.8260

Figure 17 – Mixture of Normal distributions that recover first four moments for the Max PSR and Max SR portfolios (parameters)

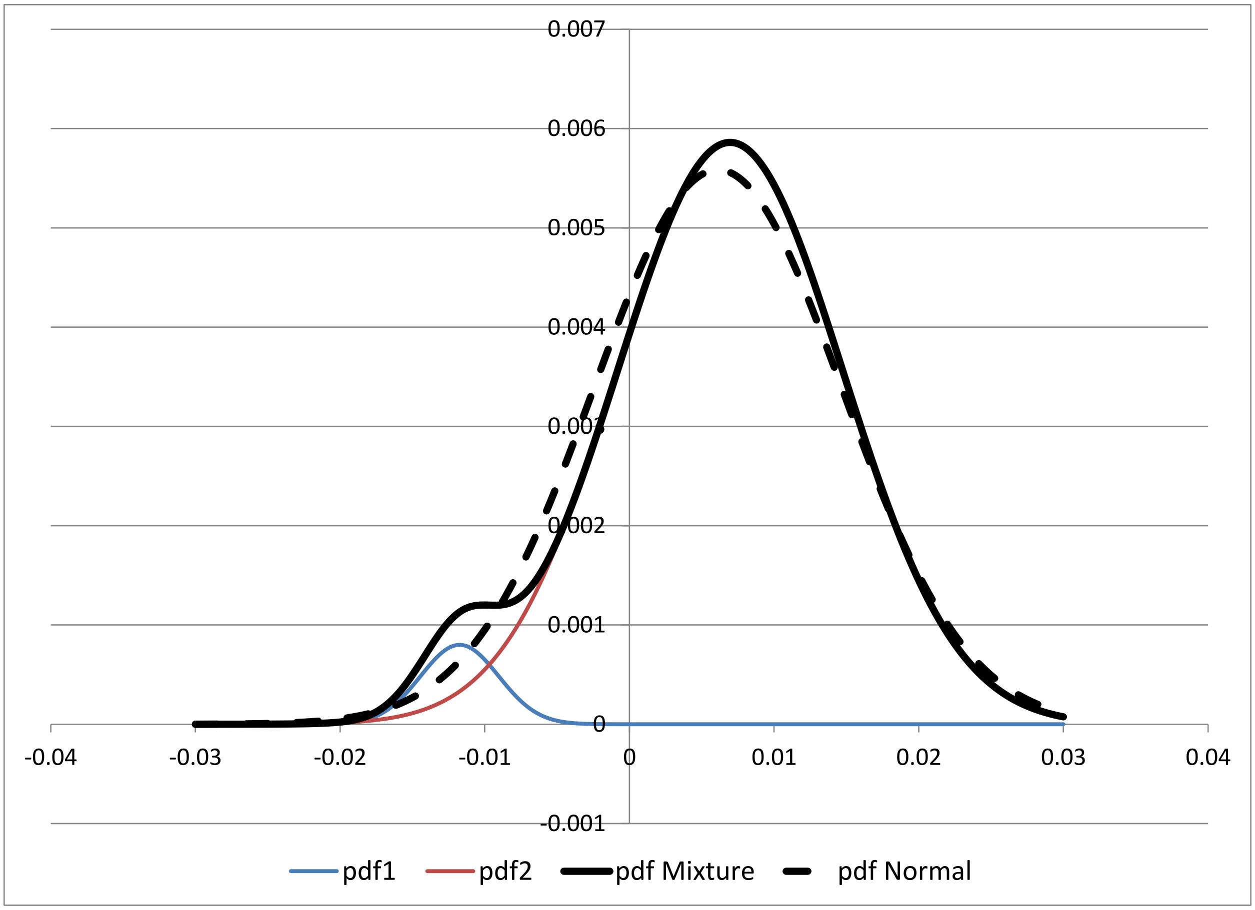

Figure 18(a) – Mixture of Normal distributions that recover the first four moments for the Max PSR

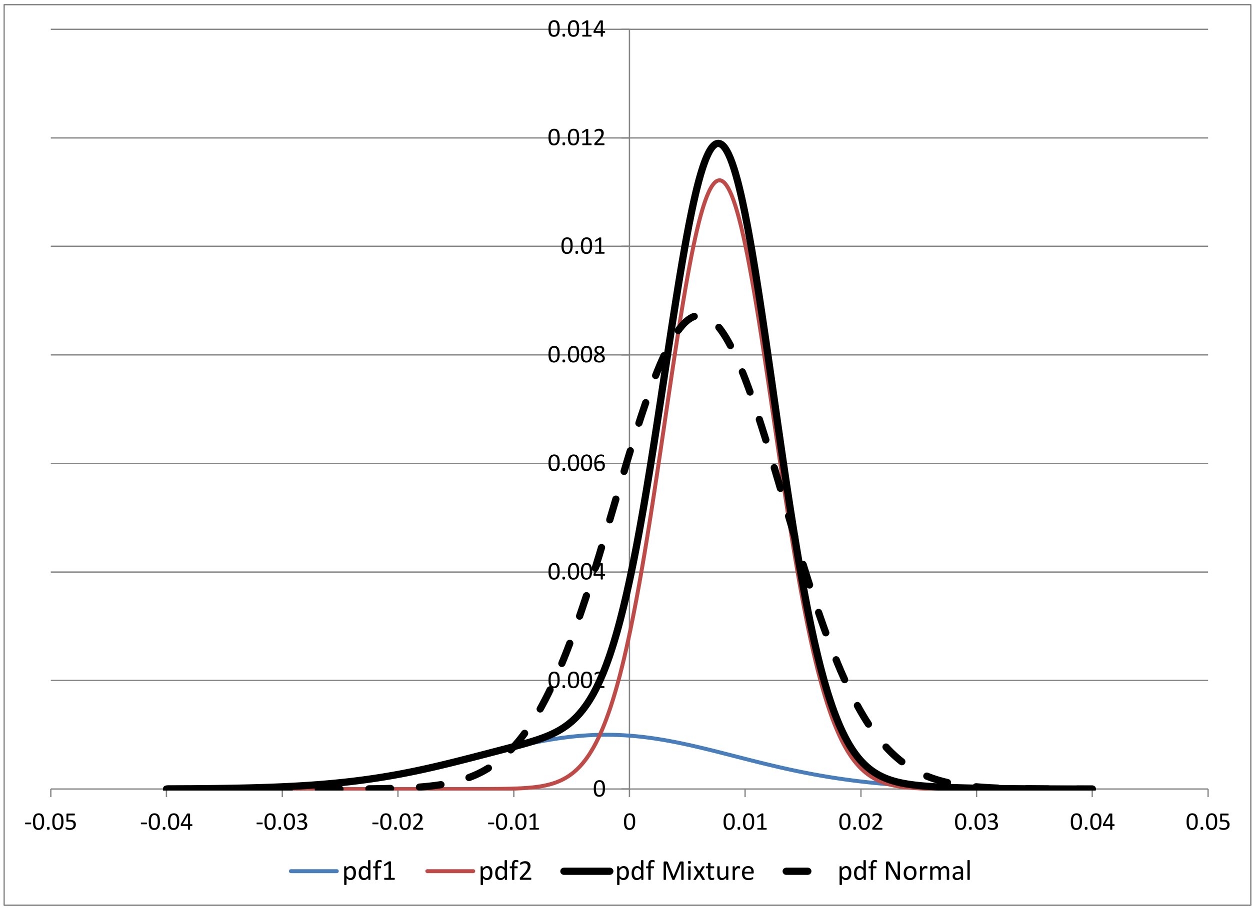

Figure 18(b) – Mixture of Normal distributions that recover the first four moments for the Max SR

REFERENCES

- Best, M., R. Grauger (1991) "On the sensitivity of Mean-Variance-Efficient portfolios to changes in asset means: Some analytical and computational results". Review of Financial Studies, January, pp. 315-342.
- Black, F., R. Litterman (1992) "Global portfolio optimization". Financial Analysts Journal, September-October, pp. 28-43.
- Brooks, C., H. Kat (2002) "The Statistical Properties of Hedge Fund Index Returns and Their Implications for Investors". Journal of Alternative Investments, Vol. 5, No. 2, Fall, pp. 26-44.
- Christie, S. (2005): "Is the Sharpe Ratio Useful in Asset Allocation?", MAFC Research Papers No.31, Applied Finance Centre, Macquarie University.
- Hogg, R. and Tanis, E. (1996) "Probability and Statistical Inference", Prentice Hall, 5th edition.
- Ingersoll, J., M. Spiegel, W. Goetzmann, I. Welch (2007) "Portfolio performance manipulation and manipulation-proof performance measures". The Review of Financial Studies, Vol. 20, No. 5, pp. 1504-1546.
- Lo, A. (2002) "The Statistics of Sharpe Ratios". Financial Analysts Journal (July), pp. 36-52.
- López de Prado, M., A. Peijan (2004) "Measuring Loss Potential of Hedge Fund Strategies". Journal of Alternative Investments, Vol. 7, No. 1, Summer, pp. 7-31. http://ssrn.com/abstract=641702
- López de Prado, M., C. Rodrigo (2004) "Invertir en Hedge Funds". 1st ed. Madrid: Díaz de Santos.
- López de Prado, M. and M. Foreman (2011) "Markowitz meets Darwin: Portfolio Oversight and Evolutionary Divergence". RCC at Harvard University, Working paper. http://ssrn.com/abstract=1931734
- Markowitz, H.M. (1952) "Portfolio Selection". The Journal of Finance 7(1), pp. 77–91.
- Mertens, E. (2002) "Variance of the IID estimator in Lo (2002)". Working paper, University of Basel.
- Opdyke, J. (2007): "Comparing Sharpe ratios: so where are the p-values?", Journal of Asset Management 8 (5), 308–336
- Roy, Arthur D. (1952) "Safety First and the Holding of Assets". Econometrica (July), pp. 431–450.
- Sharpe, W. (1966) "Mutual Fund Performance", Journal of Business, Vol. 39, No. 1, pp. 119–138.
- Sharpe, W. (1975) "Adjusting for Risk in Portfolio Performance Measurement", Journal of Portfolio Management, Vol. 1, No. 2, Winter, pp. 29-34.
- Sharpe, W. (1994) "The Sharpe ratio", Journal of Portfolio Management, Vol. 21, No. 1, Fall, pp. 49-58.
- White, H. (1984) "Asymptotic Theory for Econometricians". Academic Press, New York.

DISCLAIMER

The views expressed in this paper are those of the authors and not necessarily reflect those of Tudor Investment Corporation. No investment decision or particular course of action is recommended by this paper.