Comparing Sharpe Ratios: So Where are the p-values?

J.D. OPDYKE*

Until recently, since Jobson & Korkie (1981) derivations of the asymptotic distribution of the Sharpe ratio that are practically useable for generating confidence intervals or for conducting one- and two-sample hypothesis tests have relied on the restrictive, and now widely refuted, assumption of normally distributed returns. This paper presents an easily implemented formula for the asymptotic distribution that is valid under very general conditions – stationary and ergodic returns – thus permitting time-varying conditional volatilities, serial correlation, and other non-iid returns behavior. It is consistent with that of Christie (2005), but it is more mathematically tractable and intuitive, and simple enough to be used in a spreadsheet. Also generalized beyond the normality assumption is the small sample bias adjustment presented in Christie (2005). A thorough simulation study examines the finite sample behavior of the derived one- and two-sample estimators under the realistic returns conditions of concurrent leptokurtosis, asymmetry, and importantly (for the two-sample estimator), strong positive correlation between funds, the effects of which have been overlooked in previous studies. The two-sample statistic exhibits reasonable level control and good power under these real world conditions. This makes its application to the ubiquitous Sharpe ratio rankings of mutual funds and hedge funds very useful, since the implicit pairwise comparisons in these orderings have little inferential value on their own. Using actual returns data from twenty mutual funds, the statistic yields statistically significant results for many such pairwise comparisons of the ranked funds. It should be useful for other purposes as well, wherever Sharpe ratios are used in performance assessment.

JEL:C10,C12,C13,G10,G11   Keywords: performance, risk, portfolio, mutual fund, hedge fund, asymmetry, heavy tails

One of the most widely used statistics in financial analysis is the reward-to-variability ratio, or Sharpe ratio (see Sharpe, 1966, 1975, and 1994).[1] This simple statistic is a measure of risk-adjusted performance: it measures the average excess returns of a stock or fund (beyond some risk-free rate) relative to its volatility as measured by its standard deviation: $SR = (\mu - R_f) / \sigma$. Thus will SR provide a measure of returns per unit of volatility: SR not only will score higher for higher returns, but also will score higher (lower), all else equal, under less (more) volatility.[2]

Debate over the years regarding the utility of SR as a metric for evaluating market performance has been extensive,[3] but in comparison, surprisingly little has been written about its statistical properties,[4] especially given its ubiquitous usage. The latter is the focus of this paper, which must begin with the recognition that the components of SR – both the mean and the standard deviation – are statistics subject to random variation. This makes $\widehat{SR}$, too, a statistic subject to random variation ($\widehat{SR}$ is the sample-based estimate of SR). Yet as such, all three follow specific statistical distributions, and knowing the statistical distribution that SR follows will allow researchers and financial analysts to make probabilistic inferences about its values with specified levels of certainty. For example, is $SR = 0.5$, for a certain fund over a certain period of time, statistically significantly different from zero? In other words, does that 0.5 value indicate a *risk-adjusted* positive excess return with some level of confidence (say, 95%), or is it just an artifact of random variation about a true value of zero (or less)? Similarly, is the SR of one fund statistically significantly larger than that of another, or is any observed difference just a reflection of market volatility? This is a very important question, since many thousands of times every week, around the globe, the performance of funds and fund managers are ranked according to their Sharpe ratios. Yet no information ever accompanies these rankings to indicate whether the observed differences are actually statistically significant! Without such information, in the form of p-values and/or confidence intervals, the entire ranking exercise is of limited inferential value. Yet it is exactly questions like these that can be answered with knowledge of the statistical distribution that $\widehat{SR}$ follows as it is subject to random variation.

* J.D. Opdyke is President, Senior Statistical Data Miner, DataMineIt, 40 Tioga Way, Commerce Center – Suite 240, Marblehead, MA 01945 (E-mail: JDOpdyke@DataMineIt.com). I express thanks to Keith Ord, Hrishikesh Vinod, and Andrew Clark for useful comments, and sincere appreciation to Steve Christie for spotting an error in an earlier draft.

[1] Christie (2005) refers to the Sharpe ratio as "ubiquitous in the finance industry … arguably the most widely used general measure of fund manager performance." (p.5). McLeod and van Vuuren describe how it "quickly gained widespread popular acceptance and today enjoys almost ubiquitous implementation in the financial world" (p.1). And Lo (2002) calls it "one of the most commonly cited statistics in financial analysis" (p.1). In light of such assessments, it would appear difficult to overstate the importance of correctly understanding the statistical properties of the Sharpe ratio.

[2] Although this is not true when excess returns are negative, many argue that the interpretation of the Sharpe ratio under these conditions does not change: a larger Sharpe ratio still indicates better risk-adjusted performance (see Akeda, 2003, Sharpe, 1998, and Vinod & Morey, 2000). Others disagree (see Scholz, 2007).

[3] Eling & Schuhmacher (2007) present strong new evidence, even under the highly non-normal data conditions of hedge fund returns, in support of the Sharpe ratio compared to other more complex risk-adjusted performance metrics, the statistical properties of most of which are far less well understood.

[4] Scherer (2004) believes this is due to "the extreme difficulty of working out the required statistics for most risk-return ratios," yet the derivations presented herein are fairly straightforward.

This paper derives the asymptotic (large sample) distribution of $\widehat{SR}$ under very general conditions – stationary and ergodic returns – thus permitting time-varying conditional volatilities, serial correlation, and otherwise non-iid returns behavior (i.e. returns that are not independent and identically distributed). The derivation shows that those of both Mertens (2002), which is valid under iid returns, and Christie (2005), which is more broadly valid under stationary and ergodic returns, are, in fact, identical. It thus generalizes the far more restrictive iid requirement of the former, while greatly simplifying the more complex formula of the latter, making it more mathematically tractable and intuitive, and far easier to calculate and implement when conducting hypothesis tests or generating corresponding confidence intervals (it is simple enough to be used in a spreadsheet).

In the non-asymptotic realm, the small sample bias adjustment proposed by Christie (2005) is generalized beyond the restrictive and unrealistic assumption of iid normality, and used in an extensive simulation study that a) demonstrates the empirical level and power of the one-sample estimator of *SR* under leptokurtosis ("heavy-tails") and asymmetry, which have been widely cited as characterizing stock market returns (see Be, 2000, Brännäs & Nordman, 2003, Harris & Coskun, 2001, Dillen & Stoltz, 1999, Patterson & Heravi, 2003, and Richardson & Smith, 1993); and b) demonstrates the empirical level and power of an analogous two-sample statistic that tests whether the *SR* of one stock or fund is larger than that of another, especially under returns with strong, positive correlation between funds. No other easily implemented statistic exists, as a simple distribution-based formula rather than a complex computer program, to perform such a comparison under real-world financial data conditions.

The paper concludes by applying both the one- and two-sample estimators to the actual returns data of twenty mutual funds. The fund lists are ranked by $\widehat{SR}$, as is widely done in practice, and pairwise comparisons of the funds' Sharpe ratios, for all possible pairs, are made using the two-sample statistic. Unlike some preliminary research conducted under more restrictive and unrealistic assumptions (i.e. iid normality of returns), these hypothesis tests yield many statistically significant results, indicating that, given the good level control shown in the simulation study, the two-sample statistic should be very useful for this purpose in practice. It should be useful for other purposes as well, wherever Sharpe ratios are used for performance assessment.

I. The Commonly Used Sharpe Ratio

Many 'modified' versions of the Sharpe ratio have been presented in the finance literature, but the basic Sharpe ratio in common usage takes one of two forms:

$SR_e = \dfrac{\mu_e}{\sigma_e}$ (and for the actual samples, $\widehat{SR}_e = \dfrac{\hat{\mu}_e}{\hat{\sigma}_e}$), where $\hat{\mu}_e$ is the sample mean of the *excess* returns of a stock or fund beyond some risk-free rate (R_f), $\hat{\mu}_e = \dfrac{\sum_{t=1}^{T} R_{et}}{T}$ and $R_{et} = \left(R_t - R_{ft}\right)$; and $\hat{\sigma}_e = \sqrt{\dfrac{\sum_{t=1}^{T}\left(R_{et} - \hat{\mu}_e\right)^2}{T-1}}$ is the sample standard deviation of the *excess* returns. This is the definition of the Sharpe ratio as presented by Jobson and Korkie (1981), Sharpe (1994), Memmel (2003), and others.

An arguably more widely used version is presented in Lo (2002) and Christie (2005) as: $SR_S = \dfrac{\mu_S - R_f}{\sigma_S}$ (and for the actual samples, $\widehat{SR}_S = \dfrac{\hat{\mu}_S - R_f}{\hat{\sigma}_S}$), where $\hat{\mu}_S = \dfrac{\sum_{t=1}^{T} R_t}{T}$ is the sample mean of the stock returns, $\hat{\sigma}_S = \sqrt{\dfrac{\sum_{t=1}^{T}\left(R_t - \hat{\mu}_S\right)^2}{T-1}}$ is the sample standard deviation of the stock returns, and R_f is the risk-free rate. If R_f is constant over the periods used to calculate $\widehat{SR}$, as is often the case and assumed by this second definition of $\widehat{SR}$, then $R_f = \hat{\mu}_f$ and, of course, $\widehat{SR}_e = \widehat{SR}_S$, because the mean of the difference is equal to the difference of the means. Even if R_f is not literally constant over the specified time period, its variance is so small relative to that of a typical stock or fund that its arithmetic mean often is treated as its constant value, a convention that is assumed throughout the remainder of this paper.[5]

[5] Treating the risk-free rate as a constant is further justified by the fact that, even when its variance is not literally zero over the time period being examined, its covariance with stocks or funds will be.

II. The Statistical Distribution of The Sharpe Ratio

A. IID Normality

A quarter century ago, Jobson and Korkie (1981) presented a derivation of the asymptotic distribution of *SR* under the assumption of normal iid returns:

$$SR \overset{a}{\sim} N\left(\frac{\mu_e}{\sigma},\ \frac{1}{n}\left(1+\frac{\mu_e^2}{2\sigma^2}\right)\right) \tag{1}$$

where μ_e is the average *excess* return. Lo (2002) presented this again more recently, but because of his iid notation (2) and the fact that the presumption of normality is implicitly stated in a footnote, the requirement of normality is easy to miss. Consequently, this result has been widely (and incorrectly) cited as being valid under iid generally (see Lee (2003), Getmansky et al. (2004), Pinto and Curto (2005), Hennard and Aparicio (2003), and McLeod and van Vurren (2004) for a few examples).[6]

$$\sqrt{T}\left(\widehat{SR}-SR\right)\overset{a}{\sim}N\left(0,\ V_{IID}\right),\ V_{IID}=1+\frac{\left(\mu-R_f\right)^2}{2\sigma^2}=1+\frac{1}{2}SR^2$$

(2)

B. IID Generally

Mertens (2002) correctly notes that Lo's (2002) derivation is valid only under iid *normality*, and presents a derivation that is valid under iid *generally*:

$$\sqrt{T}\left(\widehat{SR}-SR\right)\overset{a}{\sim}N\left(0,\ 1+\frac{1}{2}SR^2-SR\cdot\gamma_3+SR^2\left[\frac{\gamma_4-3}{4}\right]\right),\text{ where }\gamma_3=\frac{\mu_3}{\sigma^3}\text{ and }\gamma_4=\frac{\mu_4}{\sigma^4} \tag{3}$$

Note that this is very straightforward: the distribution is normal, and its variance is determined by just three easily calculated values: the value of $\widehat{SR}$ itself, the kurtosis of the returns,[7] and the skewness of the returns. No higher moments affect the asymptotic distribution of $\widehat{SR}$.[8] This result also can be derived using the similar delta method approach for the ratio of two random variables presented in Stuart & Ord (1994), pp. 351-352, and in Appendix A.

C. Stationarity and Ergodicity (no IID requirement)

More recently, Christie (2005) went beyond the iid restriction.[9] He used a generalized method of moments (GMM) approach – based on a single system of moment restrictions that *jointly* tests the restrictions – to provide a careful

[6] In their letters to the editor of the *Financial Analysts Journal*, Morillo & Pohlman (2002) point out the earlier, identical result of Jobson & Korkie (1981), and Wolf (2003) points out that Lo's (2002) "iid" derivation is valid only under normality. Lo (2003) acknowledges both points in his response, but emphasizes the illustrative nature of his (normal) "iid" derivation while urging readers to instead use his more robust GMM estimator when analyzing actual financial data. However, Lo's (2002) GMM estimator is not a simple formulaic solution and requires a modestly complex computer program to implement (i.e. Newey & West's (1987) procedure). Both of these shortcomings to quick, simple, and practical implementation are overcome by the one- and two-sample estimators derived in this paper.

[7] Also note that the term $1/4\left(\gamma_4-1\right)$, obtained after combining the SR^2 terms in (3), may be recognized as the relative variance of the estimate of the standard deviation, $\hat{\sigma}$ (see Hansen et al. (1953), p.99, 102).

[8] It is very important to note, however, that there appears to be growing empirical evidence for many financial instruments that the fourth moment of returns sometimes simply does not exist – kurtosis diverges rather than converges as sample sizes (the number of periods) increases (see Gençay et al., 2001). While this may be related to the (high) frequency of the returns data examined, it is an important potential limitation of using (3) and the related derivations in this paper, as well as other statistical approaches, when making inferences about $\widehat{SR}$.

[9] Bao & Ullah (2006) also recently presented a derivation of the distribution of the Sharpe ratio that does not require a presumption of independence, but it does require normality, so it is much less general than Christie's (2005) derivation, and less useful in practice given the vast empirical evidence in the finance literature that returns are non-normal. And as mentioned above, Lo's (2002) GMM estimator, while not requiring iid returns, is not a simple formulaic solution, but rather requires a modestly complex computer program to implement (i.e. Newey & West's (1987) procedure), making it less preferable to Christie's (2005) formulaically straightforward estimator.

derivation of the asymptotic distribution of $\widehat{SR}$ under only the restrictions of stationarity and ergodicity. Consequently, his result is valid under the more realistic conditions of time-varying conditional volatilities, serial correlation, and otherwise non-iid returns. He obtains the somewhat unwieldy formula for the variance of $\widehat{SR}$ in (4)

$$Var\left(\sqrt{T}\,\widehat{SR}\right) = E\left[\frac{SR^2\mu_4}{4\sigma^4} - \frac{SR\left[\left(R_t - R_{ft}\right)\left(R_t-\mu\right)^2 - \left(R_t - R_{ft}\right)\sigma^2\right]}{\sigma^3} + \frac{\left(R_t-\mu\right)^2}{\sigma^2} - \frac{2\left(R_t-\mu\right)}{\sigma} + \frac{3SR^2}{4}\right] \tag{4}$$

The price paid for a more general result, however, appears to be a less intuitive and more difficult implementation compared to that of Mertens (2002). This is especially true if one needs to construct from this a two-sample statistic to compare two Sharpe ratios to see if one fund's *SR* is larger than that of another. Such comparisons are made implicitly, many thousands of times a week, whenever mutual funds or hedge funds are ranked based on their Sharpe ratios. However, tests of statistical significance never accompany such rankings because of their heretofore unrealistic assumptions (e.g. iid normality) and lack of quick and easy implementation. This paper provides solutions to both of these shortcomings – i) an easily implemented single-sample estimator valid under general conditions, and ii) a corresponding (easily implemented) two-sample estimator for comparing two Sharpe ratios.

i) In Appendix B, it is shown that Christie's (2005) derivation (4) is, in fact, identical to that of Mertens (2002) (3), making the far more general conditions of the former valid for the latter (5).[10]

$$E\left[\frac{SR^2\mu_4}{4\sigma^4} - \frac{SR\left[\left(R_t - R_{ft}\right)\left(R_t-\mu\right)^2 - \left(R_t - R_{ft}\right)\sigma^2\right]}{\sigma^3} + \frac{\left(R_t-\mu\right)^2}{\sigma^2} - \frac{2\left(R_t-\mu\right)}{\sigma} + \frac{3SR^2}{4}\right] = 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4} - 1\right] - SR\frac{\mu_3}{\sigma^3} \tag{5}$$

So the distribution of $\widehat{SR}$, under very general, "real world" financial data conditions, is simply (6) below:[11]

$$\sqrt{T}\left(\widehat{SR} - SR\right) \overset{a}{\sim} N\left(0,\ 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4} - 1\right] - SR\frac{\mu_3}{\sigma^3}\right) \tag{6}$$

Of course the standard error of $\widehat{SR}$, based on (6), is

$$SE\left(\widehat{SR}\right) = \sqrt{\left[1 + \frac{SR^2}{4}\left(\frac{\mu_4}{\sigma^4} - 1\right) - SR\frac{\mu_3}{\sigma^3}\right]\Big/T} \tag{7}$$

and the estimated standard error is:[12]

$$\widehat{SE\left(\widehat{SR}\right)} = \sqrt{\left[1 + \frac{\widehat{SR}^2}{4}\left(\frac{\hat{\mu}_4}{\hat{\sigma}^4} - 1\right) - \widehat{SR}\frac{\hat{\mu}_3}{\hat{\sigma}^3}\right]\Big/(T-1)} \tag{8}$$

Likewise, confidence bounds for $\widehat{SR}$ are determined with:

$$\widehat{SR} \pm z_{crit} \times \widehat{SE\left(\widehat{SR}\right)} \tag{9}$$

where z_{crit} is the critical value of the standard normal distribution corresponding to the desired level of confidence the financial analyst wants associated with the confidence interval (e.g. $z_{crit} \approx 1.645$ and ± 1.96 correspond to one-tailed (upper tail) and two-tailed 95% confidence intervals, respectively). The easily calculated and intuitive nature of (6) allows us to examine how the distribution of $\widehat{SR}$ will vary based on the distribution of the returns, as shown in Graph 1 below (although (6) remains valid even if returns are not iid). The Asymmetric Power Distribution (APD) of Komunjer (2006) is described in detail further below, and is used in Figure 1 with skewness and kurtosis parameters that reflect the (negatively) skewed and leptokurtotic ("heavy tailed") characteristics of "real world" returns data. Figure 1 makes very

---

[10] As noted in Appendix B, the presumption of a constant risk-free rate, or an *essentially* constant risk-free rate, is required for this simplification. As an empirical matter, this assumption is justified.

[11] In addition to being more easily calculated and understandable, (6), unlike (4), makes readily apparent the requirement of the existence of third and fourth moments (in addition to stationarity and ergodicity). As mentioned above, the fourth moment does not appear to exist for some financial instruments (see Gençay et al., 2001, and Jondeau & Rockinger, 2003), in which case transformations to normality sometimes may be a viable alternative (for example, see Malevergne & Sornette, 2005).

[12] Common practice in the financial services sector notwithstanding, dividing by *T*-1, rather than *T*, in the standard error will provide a less biased, albeit still slightly biased, estimate of the population standard error (see Zar, 1999, p.39).

clear that even if one was to ignore the extensive empirical evidence against normality, simply assuming normality of returns could provide a misleading basis for making statistical inferences about $\widehat{SR}$.[13]

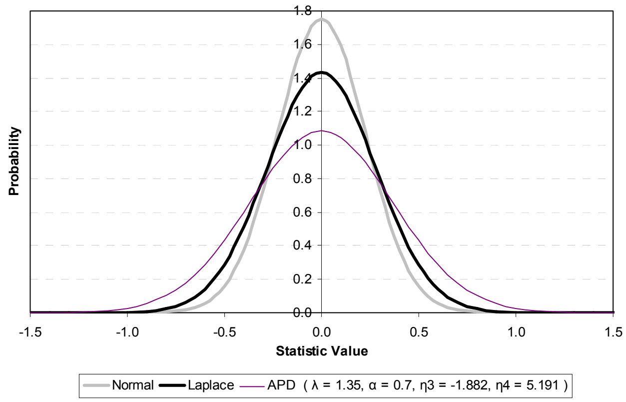


Figure 1. Distribution of $\widehat{SR}$ by Distribution of Returns, SR = 1.0, T=30

ii) For comparing two Sharpe ratios, an easily implemented two-sample statistic, which is the two-sample analog to (6), is derived and presented below. But first, the issue of small sample bias is addressed.

D. Small Sample Bias

Unbiased estimates of *SR* have been derived under iid normality (see Miller & Gehr, 1978) and iid lognormality (see Knight and Satchell, 2005), but both are relatively unwieldy and, of course, restricted by their parametric assumptions. Christie (2005) takes a more general, and arguably more practical approach to small sample bias. He begins by correctly pointing out that, due to division by σ, *SR* is convex, so its estimator will be biased due to Jensen's inequality:

$$E\left[\widehat{SR}(\hat{\mu},\hat{\sigma})\right] \geq SR\left(E[\hat{\mu}],E[\hat{\sigma}]\right) = SR(\mu,\sigma) \tag{10}$$

To obtain an estimate of the bias of $\widehat{SR}$, Christie (2005) first obtains a second-order Taylor-series expansion of $\widehat{SR}$ about σ, the cause of $\widehat{SR}$'s convexity, and then uses a first-order Taylor-series expansion of $\hat{\sigma}^2$ about σ^2 to obtain the distribution of $\hat{\sigma}$. Ironically, however, after cautioning against using bias adjustments that rely on parametric assumptions, he uses an estimator of the variance of the sample variance, $2\hat{\sigma}^4$, that only is valid under normality (see Kmenta (1986), p.139). Therefore, his result below is valid only under normality (see Christie, 2005).

$$E\left[\widehat{SR}(\hat{\mu},\hat{\sigma})\right] = SR(\mu,\sigma)\left(1+\frac{1}{2}\frac{1}{T}\right) \tag{11a}$$

It is not surprising that this resembles Lo's (2002) asymptotic distribution of *SR* because he also uses the same estimate of the variance of $\hat{\sigma}^2$, namely $2\hat{\sigma}^4$, which only is valid under normality. Substituting for $2\hat{\sigma}^4$ the term $\left(\hat{\mu}_4 - \hat{\sigma}^4\right)$, which is valid asymptotically for any distribution (see Randles & Wolf (1979), pp. 73-74), yields the bias adjustment of (11b) below.

$$E\left[\widehat{SR}(\hat{\mu},\hat{\sigma})\right] = SR(\mu,\sigma)\left(1+\frac{1}{4}\frac{\left[\mu_4/\sigma^4 - 1\right]}{T}\right) \tag{11b}$$

Not surprisingly, this resembles the asymptotic distribution in (6), which correctly and more generally takes into account the kurtosis of the returns. This bias adjustment – dividing $\widehat{SR}$ by the coefficient on $SR(\mu,\sigma)$ in (12) – is used in the simulation study below to scrutinize the actual small sample behavior of $\widehat{SR}$, especially under leptokurtotic and/or skewed data. Also examined in the simulation study is the two-sample analog to (6), which can be used to test the hypothesis that the *SR* of one fund is larger than that of another: $H0: SR_a \leq SR_b$ vs. $Ha: SR_a > SR_b$ (of course, it also can be used in two-tailed tests of whether two *SR*'s are equal). The two-sample estimator is presented below.

[13] The fact that the distribution of the Sharpe ratio (6) takes into account higher moments of the returns distribution (i.e. skewness and kurtosis) at least partially mitigates criticism of the Sharpe ratio for not *explicitly* incorporating such moments into its actual formula (which, of course, is based only on the mean and the standard deviation). And as previously noted, Eling & Schuhmacher (2007) present strong new evidence, even under the "difficult" data conditions of highly non-normal hedge fund returns, that the Sharpe ratio performs virtually identically to far more complex metrics that attempt, with mixed success, to explicitly incorporate higher moments.

III. A Two-Sample Statistic for Comparing Sharpe Ratios

One approach to testing Ho: $SR_a \le SR_b$ v. Ha: $SR_a > SR_b$ is to derive the distribution of $\widehat{SR}_{diff} = \left(\widehat{SR}_a - \widehat{SR}_b\right) - \left(SR_a - SR_b\right)$.[14] Because $\widehat{SR}_a$ and $\widehat{SR}_b$ are asymptotically unbiased normally distributed random variables, based on the Central Limit Theorem (for dependent variables – see, for example, White, 2001, Ch.5) their linear combination will be asymptotically unbiased and normally distributed. The expected value is zero, and the variance, of course, is:

$$Var\left[\left(\widehat{SR}_a - \widehat{SR}_b\right) - \left(SR_a - SR_b\right)\right] = Var\left(\widehat{SR}_a - \widehat{SR}_b\right) = Var\left(\widehat{SR}_{diff}\right) = Var\left(\widehat{SR}_a\right) + Var\left(\widehat{SR}_b\right) - 2Cov\left(\widehat{SR}_a, \widehat{SR}_b\right) \quad (12)$$

The first two terms are from (6), and the covariance term is derived in Appendix C, so that, letting $a = R_{at}$ and $b = R_{bt}$,

$$\sqrt{T}\left(\widehat{SR}_{diff}\right) \overset{a}{\sim} N\left(0,\ Var_{diff}\right), \text{ where}$$

$$\begin{aligned}
Var_{diff} = {} & 1 + \frac{SR_a^2}{4}\left[\frac{\mu_{4a}}{\sigma_a^4} - 1\right] - SR_a \frac{\mu_{3a}}{\sigma_a^3} + \\
& 1 + \frac{SR_b^2}{4}\left[\frac{\mu_{4b}}{\sigma_b^4} - 1\right] - SR_b \frac{\mu_{3b}}{\sigma_b^3} \\
& -2\left[\rho_{a,b} + \frac{SR_a SR_b}{4}\left[\frac{\mu_{2a,2b}}{\sigma_a^2\sigma_b^2} - 1\right] - \frac{1}{2}SR_a \frac{\mu_{1b,2a}}{\sigma_b\sigma_a^2} - \frac{1}{2}SR_b \frac{\mu_{1a,2b}}{\sigma_a\sigma_b^2}\right]
\end{aligned} \quad (13)$$

where $\mu_{2a,2b} = E\left[\left(a - E(a)\right)^2\left(b - E(b)\right)^2\right]$ is the joint second central moment of the joint distribution of *a* and *b*, and $\mu_{1a,2b} = E\left[\left(a - E(a)\right)\left(b - E(b)\right)^2\right]$ and $\mu_{1b,2a} = E\left[\left(b - E(b)\right)\left(a - E(a)\right)^2\right]$ (unbiased estimators for these three terms are provided in Appendix C). Note that when $\rho_{a,b} = 0$, $\mu_{2a,2b} = \sigma_a^2\sigma_b^2$, $\mu_{1a,2b} = 0$, and $\mu_{1b,2a} = 0$, so the entire covariance term disappears, as it should.

Following (12), we can see that (13) is the two-sample analog to (6), and since it also was derived using the delta method, which for the one-sample estimator (6) was shown to be identical to the more generally valid GMM method, we suspect the more general conditions of stationarity and ergodicity are the only requirements for (13) as well. Proving this is the topic of continuing research.

(13) is very easily implemented as a test of $H0: SR_a \le SR_b$ vs. $Ha: SR_a > SR_b$ – so easy, in fact, that it can be implemented in a spreadsheet (one that implements the hypothesis tests and confidence intervals corresponding to both (6) and (13) is available for download at the author's website at www.DataMineIt.com). And under iid normality, $\mu_3/\sigma^3 = 0$, $\mu_{1,2} = 0$, $\mu_4/\sigma^4 = 3$, and $\mu_{2a,2b} = \left(1 + 2\rho_{a,b}^2\right)\sigma_a^2\sigma_b^2$ (see Stuart & Ord, 1994, p.105): inserting these values into (13), as shown in Appendix D, yields Memmel's (2003) correction of Jobson and Korkie's (1981) two-sample statistic, thus providing further independent validation of these derivations. The distribution of $\widehat{SR}_{diff}$ is graphed in Figure 2 below for illustrative purposes. Assuming zero correlation between the two returns, we can see in Figure 2a that the variance of (13) is twice as large, all else equal, as that of (6) in Figure 1.

[14] Of course, this is not the only approach. To test this two-sample hypothesis, Christie (2005) jointly tests, consistent with his asymptotic derivation, moment restrictions within a single system of moment restrictions. However, the benefits of this paper's approach over Christie's (2005) GMM approach are two-fold: i) implementation of the former does not require custom coding a moderately complex statistical software program (rather, it can be implemented in a spreadsheet), and ii) it provides confidence intervals as well as p-values, while Christie's (2005) approach provides only p-values. While Vinod & Morey's (2000) bootstrap approach does not require derivation of the distribution of the difference between Sharpe ratios, it does require a computationally intensive computer program (and a *very* computationally intensive program for their double bootstrap method), and may be less powerful than the asymptotic approach taken in this paper. In addition, it should be noted that the variance estimates produced by many bootstrap procedures have been shown in the literature to be notoriously poor under asymmetric heavy tails, and even under symmetric heavy tails (see Rocke & Downs, 1981, Gosh et al., 1984, and Salibián-Barrera, 1998), and these are the defining characteristics of financial market returns. Consequently, in the absence of a rigorous, validating bootstrap simulation study providing results based on simulated returns of *known* distributions rather than actual returns data, such bootstrap variance estimators of Sharpe ratios should be interpreted with caution.

Figure 2a: $\widehat{SR}_{diff}$ by Distribution of Returns, $\rho = 0$ Figure 2b: $\widehat{SR}_{diff}$ Under "Real World" APD, by ρ

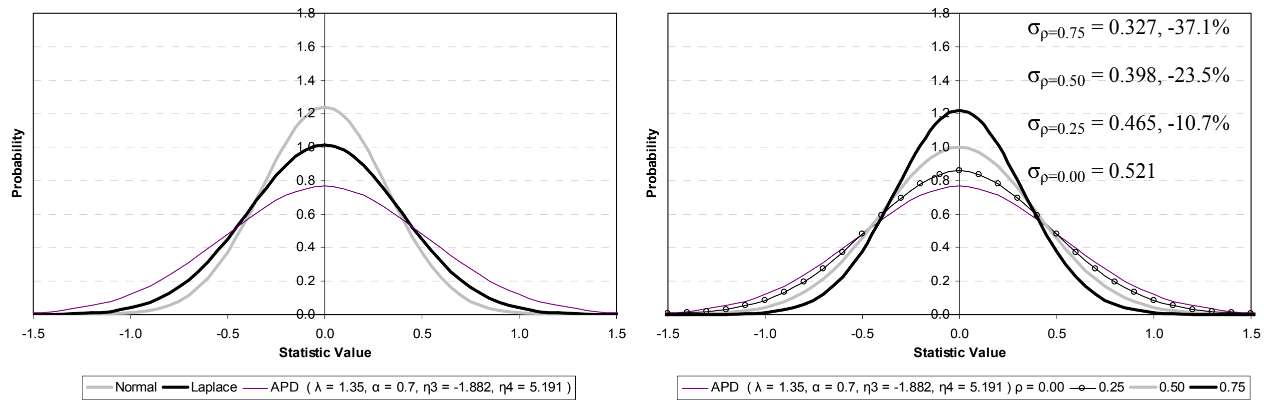


Figure 2. Distribution of $\widehat{SR}_{diff}$ *,* $SR_a = SR_b = 1.0$*,* $T = 30$

However, the variance of (13) decreases markedly as correlation between the funds increases, as shown in Figure 2b for APD with "real world" values for its skewness and kurtosis parameters.[15] This causes a dramatic increase in the power of (13), which is a major finding of both the simulation results presented below, and the empirical results from the analysis of actual mutual fund returns in Section V, so its presentation analytically in Figure 2b is important and useful. The standard deviation corresponding to each between-fund correlation, and their decreases relative to that under no correlation – $\sigma_{\rho=0.00}$ – are shown as line labels, in order, in Figure 2b.[16]

IV. Simulation Study

A. Simulating Returns with Komunjer's (2006) Asymmetric Power Distribution

The non-asymptotic properties of (6) and (13) are examined below under a wide range of sample sizes, mean-variance configurations, between-fund correlations, and distributions. Regarding distributions, notwithstanding the extensive empirical evidence in the literature of the non-normality, (negative) skewness, and leptokurtosis of financial returns, examining (6) and (13) under normality provides an important baseline. And while extensive study has been made of returns under a Laplacian (double exponential) framework (see Cajigas & Urga, 2005, Kotz, Kozubowski & Podgorski, 2001, Kozubowski & Podgorski, 2001, and Linden, 2001), including, more recently, the asymmetric Laplacian distribution, there exists empirical evidence that the kurtosis of returns lies in the "in-between" range between that of the normal and Laplacian distributions (see Haas et al., 2005, and Komunjer, 2006). So to robustly test (6) and (13) under the widest range of possible conditions, including as a subset of distributions that reflect the characteristics of "real world" financial returns, we turn to a very flexible and relevant distribution – Komunjer's (2006) Asymmetric Power Distribution (APD). APD nests both the Laplace and normal distributions, as well as asymmetric versions of each (the asymmetric Laplace of Kozubowski & Podgorski, 1999, and the two-piece normal (see Johnson, Kotz & Balakrishnan, 1994, vol. 1 p.173 and vol. 2 p.190)[17]), and every combination of skewness and kurtosis "in-between."[18] One parameter of APD, α, controls skewness: $0 < \alpha < 1$, with symmetry at $\alpha = 0.5$ (when it is equivalent to the Generalized Power Distribution (GPD), which nests the normal and Laplace distributions). The other parameter, $\lambda > 0$, controls kurtosis, such that when $\alpha = 0.5$, $\lambda = \infty \rightarrow$ the uniform distribution, $\lambda = 1.0 \rightarrow$ the Laplace distribution (with variance = 2.0), and $\lambda = 2.0 \rightarrow$ the normal distribution (with variance = 0.5). Thus does APD allow simultaneous control over skewness and kurtosis.

[15] The joint moment terms of (13), for $\rho \neq 0$, are very accurately estimated in simulations of N=100,000.

[16] Although one might be tempted to say that the distribution of $\widehat{SR}_{diff}$ under a naïve assumption of normality with no between-fund correlation (Figure 2a) is virtually identical to that under more realistic distributional conditions and strong positive correlation (Figure 2b), this is only true asymptotically. In practice, using actual finite data samples, these two distributions are very different, and the simplifying but naïve (and incorrect) assumption of normality can cause very misleading inferences.

[17] This is not to be confused with the skew-normal distribution of Azzalini (1985), which is very similar.

[18] Similar densities recently have been developed, such as the asymmetric exponential power (AEP) distribution of Ayebo & Kozubowski (2003), and the Gauss-Laplace Mixture (GLaM) and Gauss-Laplace Sum (GLaS) distributions used by Haas et al (2005).

The APD for many of the combinations of α and λ used in the simulations is shown in Figure 3E in Appendix E. Although positive skewness (APD with $\alpha < 0.5$) is not shown to maintain graphical clarity, it is simulated in the study to test the robustness of the estimators, even though returns typically are negatively skewed (i.e., have a longer left tail – see Komunjer, 2006, Cajigas & Urga, 2005, and Cappiello et al., 2003 for just a few examples). Sometimes, however, they are positively skewed (see Komunjer, 2006, and for the case of bonds, Cappiello et al., 2003).

In addition to using "evenly spaced" and sometimes extreme parameter values for APD to test the full range of behavior of (6) and (13), a set of simulations is conducted using APD parameter values that reflect the (negative) skewness and leptokurtosis of actual "real world" financial returns. These are $\alpha = 0.7$ and $\lambda = 1.35$, which yield skewness and kurtosis coefficients of $\eta_3 = -1.882$ and $\eta_4 = 5.191$, respectively (see Table E1 in Appendix E for APD skewness and kurtosis corresponding to the values of α and λ used in the simulations). This distribution (Figure 3) is at least as "extreme," i.e. at least as skewed and leptokurtotic, as those of typical financial returns.[19] For example, this leptokurtosis lies in the middle of the ranges of those reported by Haas et al. (2005), Cajigas & Urga (2005), and Cappiello et al. (2003), and the skewness is far more extreme than those reported in the latter two papers. And Komunjer's (2006) maximum likelihood estimates for the values of α and λ range from 0.462 to 0.586, and 1.21 to 1.55, respectively. So the "real world" simulations reflected by $\alpha = 0.7$ and $\lambda = 1.35$ provide a reasonable test of the robustness of (6) and (13) as they would be used in practice on actual returns data.

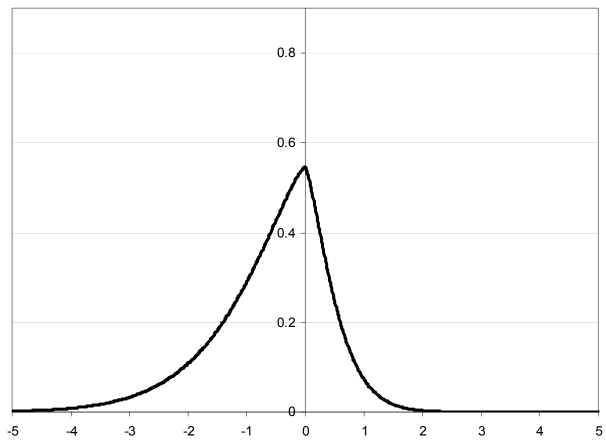


Figure 3: Standardized Asymmetric Power Distribution with Parameter Values reflecting "Real World" Returns (APD with $\alpha = 0.7$, $\lambda = 1.35$, so $\eta_3 = -1.882$ and $\eta_4 = 5.191$)

In addition to this full range of distributions, the simulation study examines two-sided, as well as one-sided (Ho: $SR_a \leq$ c vs. Ha: $SR_a >$ c, and Ho: $SR_{diff} \leq 0$ vs. Ha: $SR_{diff} > 0$) coverage. It uses sample sizes of # periods = T = 15, 30, 50, 100, and 300; mean-variance configurations (with unit variance) yielding SR values of $SR_a = 0$ & $SR_b = 0.0, 0.1, 0.2, \& 0.5$; $SR_a = 0.2$ & $SR_b = 0.2, 0.4$; $SR_a = 1.0$ & $SR_b = 1.0, 1.5$; and $SR_a = 3.0$ & $SR_b = 3.0, 3.5$; (only SR_a is used for the one-sample (6) results[20]); and correlations between the two series of returns of $\rho_{a,b} = 0.0$, $\rho_{a,b} \approx 0.25$, $\rho_{a,b} \approx 0.50$, and $\rho_{a,b} \approx 0.75$, making the total number of scenarios 25 x 5 x 10 x 4 = 5,000; with the "real world" (Figure 3) returns simulations, the total is 5,200. Dependence was induced either directly in the distributional simulations, or via Gaussian copulae, so for any simulations but the normal, Pearson's linear correlation coefficient is approximate (but almost always within ±0.01) because the non-linear transformations required to simulate these distributions cannot preserve the linear correlation function[21] (even though rank correlations, like Spearman's rho and Kendall's tau, are exactly preserved). The point estimates of SR_{diff} used the bias corrected versions of each SR, but the non-corrected point estimates were used when calculating the variances of both (6) and (13). The estimator used for skewness $\eta_3 = \mu_3/\sigma^3$ is $\sqrt{b_1}$ (14) (see Zar,

[19] Graphically, Figure 4 is very similar to the empirically estimated skew-normal density used by Vinod (2005) (p.854) (Vinod used Azzalini, 1985) and has similar coefficients of skewness and kurtosis. APD, however, is not only more flexible, since it nests a version of the skew normal, but also appears more appropriate from an empirical perspective, since Komunjer's (2006) hypothesis tests reject both the symmetric and asymmetric versions of the normal distribution using actual financial returns data. So APD would appear to be the better choice.

[20] For the one-sample power results, i.e. when $SR_a \neq$ c, $SR_a = 0.3, 0.4$ also are included.

[21] While a Cholesky decomposition will exactly preserve the linear correlations, it typically will not preserve the distributions of the returns.

1999, p.71, and Stuart & Ord, 1994, p.440). Only for the simulations, to increase numerical stability and precision, the biased estimators of (g_2 + 3) (15) (see Zar, 1999, p.68, and Stuart & Ord, 1994, pp.108-109) and $m_{2,2}$ (16) (see Rose & Smith, 2002, p.261) were used to estimate kurtosis $\eta_4 = \mu_4/\sigma^4$ and the second joint central moment, $\mu_{2,2}$, respectively.[22] However, for analyzing actual mutual fund data in Section V, the less biased b_2 (17) (see Zar, 1999, p. 71, and Stuart & Ord, 1994, p.452) and unbiased $h_{2,2}$ (see Appendix C, Halmös, 1946, and Rose & Smith, 2002, pp.253-260), respectively, are used.

$$\sqrt{b_1} = \left[(n-2)\Big/\sqrt{n(n-1)}\right] \times \left[\left(n\sum_{i=1}^{n} x_i^3 - 3\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^2 + 2\left(\sum_{i=1}^{n} x_i\right)^3 \Big/ n\right) \Big/ \left(s^3 \times [n-1][n-2]\right)\right] \tag{14}$$

$$g_2 + 3 = 3 + \left[\left((n^3+n^2)\sum_{i=1}^{n} x_i^4 - 4(n^2+n)\sum_{i=1}^{n} x_i^3 \sum_{i=1}^{n} x_i - 3(n^2-n)\left(\sum_{i=1}^{n} x_i^3\right)^2 + 12n\sum_{i=1}^{n} x_i^2 \left(\sum_{i=1}^{n} x_i\right)^2 - 6\left(\sum_{i=1}^{n} x_i\right)^4\right) \Big/ \left(s^4 \times n[n-1][n-2][n-3]\right)\right] \tag{15}$$

$$m_{2a,2b} = -\frac{3\left(\sum_{i=1}^{n} a_i\right)^2 \left(\sum_{i=1}^{n} b_i\right)^2}{n^4} + \frac{\sum_{i=1}^{n} a_i^2 \left(\sum_{i=1}^{n} a_i\right)^2}{n^3} + \frac{4\sum_{i=1}^{n} a_i \sum_{i=1}^{n} b_i \sum_{i=1}^{n} a_i b_i}{n^3} - \frac{2\sum_{i=1}^{n} a_i \sum_{i=1}^{n} a_i b_i^2}{n^2} + \frac{\sum_{i=1}^{n} a_i^2 \left(\sum_{i=1}^{n} b_i\right)^2}{n^3} - \frac{2\sum_{i=1}^{n} b_i \sum_{i=1}^{n} a_i^2 b_i}{n^2} + \frac{\sum_{i=1}^{n} a_i^2 b_i^2}{n} \tag{16}$$

$$b_2 = 3\frac{(n-1)}{(n+1)} + \frac{(n-2)(n-3)}{(n+1)(n-1)} \times \left[(n^3+n^2)\sum_{i=1}^{n} x_i^4 - 4(n^2+n)\sum_{i=1}^{n} x_i^3 \sum_{i=1}^{n} x - 3(n^2-n)\left(\sum_{i=1}^{n} x_i^2\right)^2 + 12n\sum_{i=1}^{n} x_i^2 \left(\sum_{i=1}^{n} x\right)^2 - 6\left(\sum_{i=1}^{n} x\right)^4\right] \Big/ \left[s^4 \times n(n-1)(n-2)(n-3)\right] \tag{17}$$

B. Results

B.1 One-Sample Statistic: Level Control

For the one-sample statistic, it should be emphasized that in this setting, the null hypothesis of interest is whether the risk-adjusted performance reflects positive excess returns with statistical significance – i.e., Ho: $SR \leq 0$ vs. Ha: $SR > 0$. In all the results from this simulation study, under all distributional conditions, convergence of (6) to the nominal level of $\alpha = 0.05$ (not to be confused with APD-α) under these hypotheses is virtually immediate – i.e. it occurs even for small samples.[23] However, to fully put (6) through its paces and obtain a thorough understanding of its nonasymptotic behavior under all conditions, we also test Ho $SR \leq c$ vs. Ha: $SR > c$ where $c > 0$.

Under the "real-world" returns conditions simulated by APD-α = 0.7 and APD-λ = 1.35 (yielding skewness and kurtosis coefficients of η_3 = -1.882 and η_4 = 5.191, respectively), generally quick convergence of (6) to α = 0.05 is shown in Table I and Figure 4 below for different values of SR = c (two-tailed convergence is very similar). While still fast for larger values of SR, convergence is even faster as SR = c approaches zero (for SR = 0, μ = 0 and σ = 1.0).

[22] Using biased but more efficient estimators for simulations is common statistical practice.

[23] Complete simulation results of the 5,200 scenarios examined, for both one- and two-sided tests, are available from the author upon request.

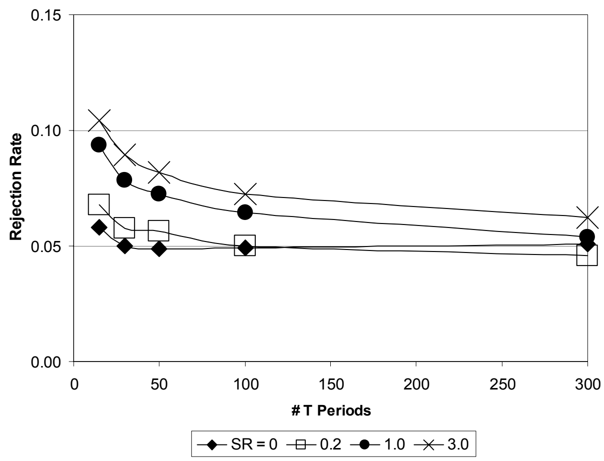


Figure 4. Rejection Rate (N=10,000) of One-Tailed Test (Ho: SR ≤ c), Bias Corrected, by SR = c by T, under "Real World" Simulated Returns (APD with $\alpha = 0.7$, $\lambda = 1.35$, so $\eta_3 = -1.882$ and $\eta_4 = 5.191$)

Table I: # of Simulations of $\widehat{SR}$ Beyond the Upper 95% Confidence Interval of (6),
APD ($\alpha = 0.7$ & $\lambda = 1.35$, η_3=-1.882, η_4=5.191) by Sample Size by *SR* = c (#Simulations = N = 10,000)

	Ho: $SR \leq 0$ ($\mu = 0$, $\sigma = 1.0$)		Ho: $SR \leq 1$ ($\mu = 1$, $\sigma = 1.0$)	
T	No adj.	Bias adj.	No adj.	Bias adj.
15	651	579	1,107	935
30	535	501	892	786
50	500	486	821	724
100	509	491	712	644
300	512	508	580	540

Under all other skewness and kurtosis combinations, we can see generally good (quick) convergence, except under the most extreme conditions, where convergence is slowed by the following factors, in the approximate order of the magnitude of the slowing effect:

1) size of *SR = c*: the larger the value of *SR = c*, the slower the convergence to α
2) skewness: usually, the more skewed the returns, the slower the convergence to α
3) kurtosis: usually, the more leptokurtotic the returns, the slower the convergence to α
4) bias correction & 1-sided vs. 2-sided coverage: 1-sided (upper-tail) coverage typically converges faster with bias correction, but 2-sided coverage typically converges faster without bias correction

As expected from (6), the slowest convergence occurs under, concurrently, large values of *SR = c*, extreme positive skewness (as high as $\eta_3 = 2.23$), and extreme leptokurtosis (as high as $\eta_4 = 6.65$), as shown in Figures 5a & 5b below (convergence patterns for two-tailed coverage are similar). Also as expected from (6), all else equal, convergence improves noticeably if extreme positive skewness is replaced with extreme negative skewness (see Figures 5c & 5d), even for two-tailed coverage. Mathematically, this is due to the *SR**skewness term in (6), and is important to note since returns typically are negatively skewed, albeit at the less extreme values shown in Figure 4.

Figure 5a: Bias-Corrected, Extreme Positive Skewness

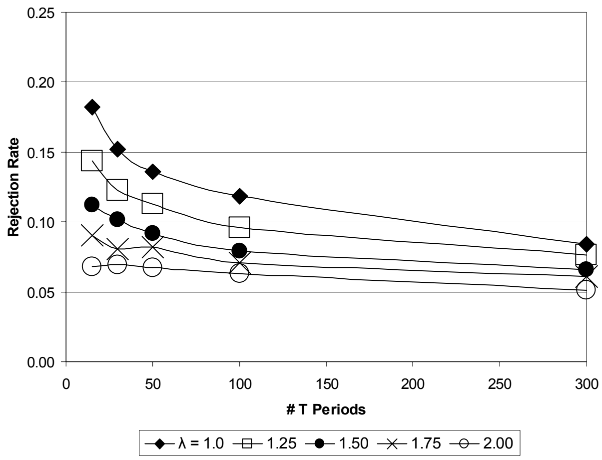


Figure 5b: No Correction, Extreme Positive Skewness

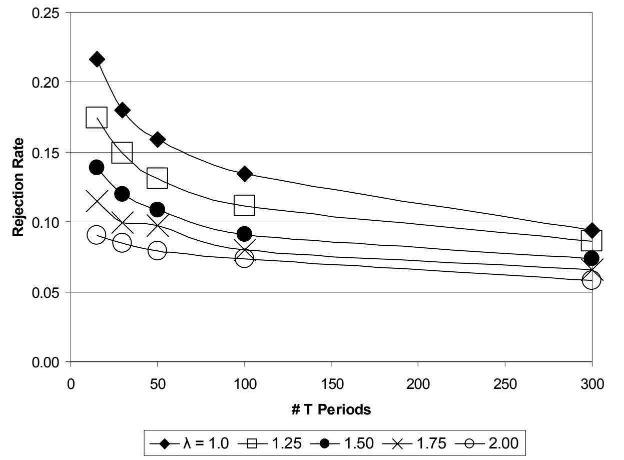


Figure 5c: Bias-Corrected, Extreme Negative Skewness

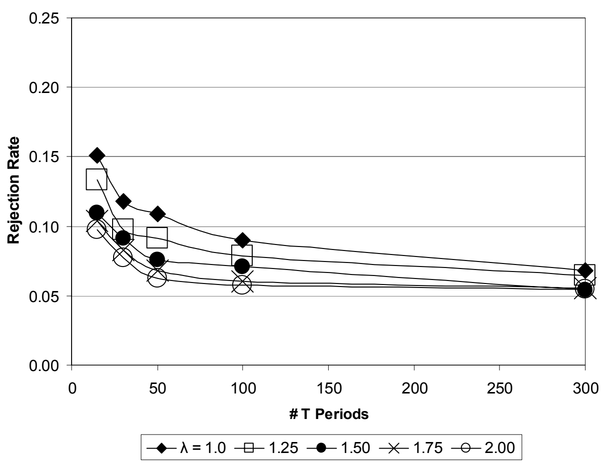


Figure 5d: No Correction, Extreme Negative Skewness

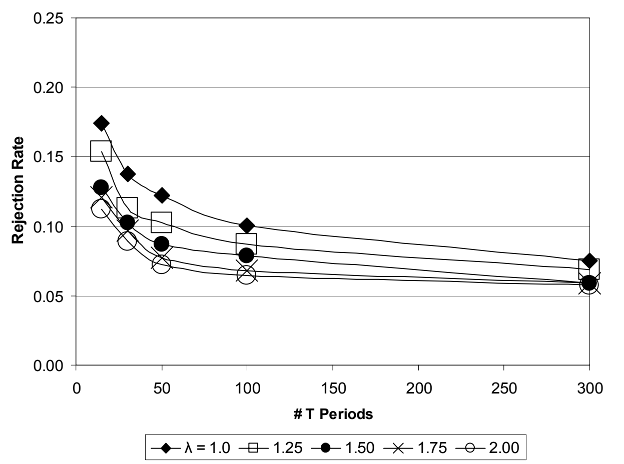


Figure 5. Rejection Rate (N=10,000) of One-Tailed Test (Ho: SR ≤ 3.0), by Bias-Correction v. No Correction by Extreme Positive v. Extreme Negative Skewness (APD-α = 0.1/0.9) by λ by T, under large SR (= c = 3.0)

B.2 One-Sample Statistic: Bias Correction

As expected, correcting for bias in the estimation of *SR* has noticeable, but not dramatic, effects on convergence to the nominal level when sample sizes are small. The largest effects can be seen under the "extreme" returns conditions shown in Figures 5a & 5b. Under the "real-world" returns conditions simulated by APD-α = 0.7 and APD-λ = 1.35, we see more modest improvement in convergence when using bias-corrected estimates (see Figure 6 below). For two-tailed coverage, the effects typically appear to be smaller than shown in Figure 7. Note again that convergence under Ho: *SR* ≤ 0.0 is virtually perfect – i.e. virtually equal to α for even very small samples.

Figure 6a: Bias-Corrected Figure 6b: No Bias Correction

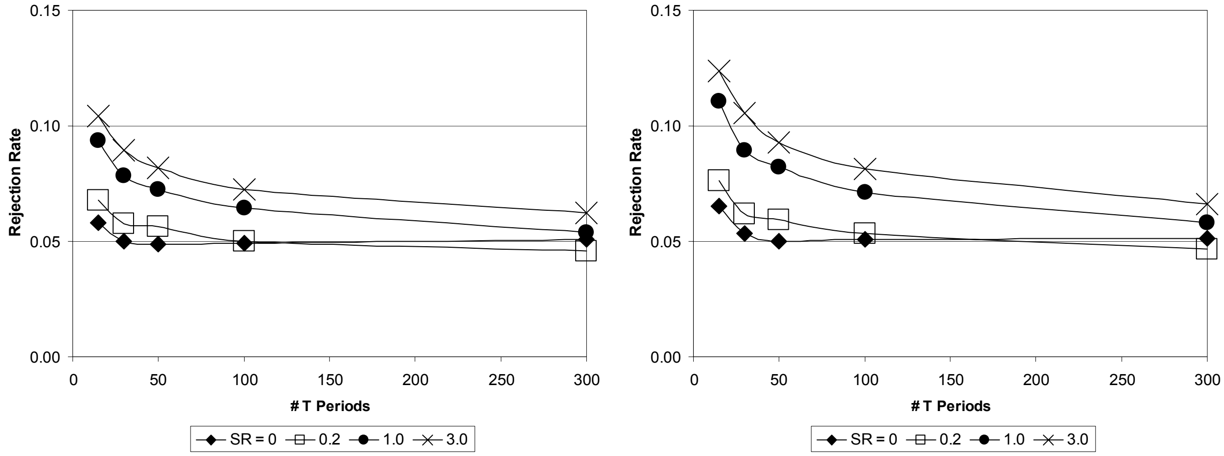


Figure 6. Rejection Rate (N=10,000) of One-Tailed Test (Ho: SR ≤ c), by Bias-Correction v. No Correction by SR = c by T, under "Real World" Simulated Returns (APD with α = 0.7, λ = 1.35, so η_3 = -1.882 and η_4 = 5.191)

B.3 One-Sample Statistic: Power

To reemphasize, for the one-sample statistic, the null hypothesis of interest is whether the risk-adjusted performance reflects positive excess returns with statistical significance – i.e., Ho *SR* ≤ c vs. Ha: *SR* > c when c = 0. For c = 0, Figure 7 below shows generally modest power for this test under the "real world" conditions of APD-α = 0.7 and APD-λ = 1.35. Under other combinations of skewness and kurtosis, for both one- and two-sided tests of c = 0, positive skewness yields more power than symmetry, which yields more power than negative skewness. And leptokurtosis yields slightly more power than mesokurtosis under positive skewness, but slightly less power under negative skewness. The one-sided test always is noticeably, but not dramatically, more powerful than the two-sided test, with the greatest differences occurring under negative skewness. Bias correction, for both one- and two-sided tests, have very little affect on power.

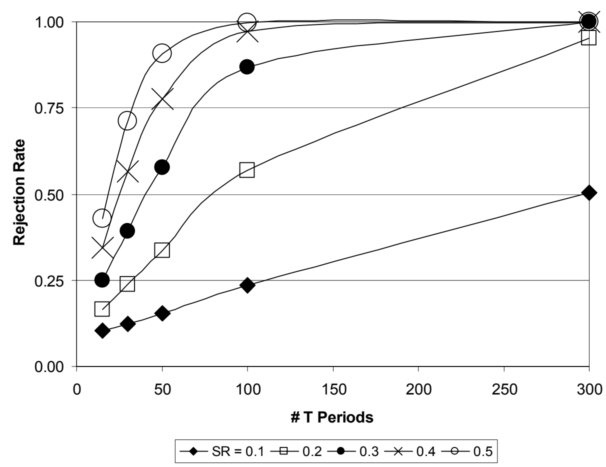


Figure 7. Rejection Rate (N=10,000) of One-Tailed Test (Ho: SR ≤ 0), Bias Corrected, by SR by T, under "Real World" Simulated Returns (APD with α = 0.7, λ = 1.35, so η_3 = -1.882 and η_4 = 5.191)

For c > 0, Figure 8a below shows generally modest power for this test under the "real world" conditions of APD-α = 0.7 and APD-λ = 1.35. Figure 8b shows that the relative size of the difference between c and *SR* matters: power to detect a difference is greater for c = 0.0 & *SR* = 0.5 than it is for c = 1.0 & *SR* = 1.5 than it is for c = 3.0 & *SR* = 3.5, all else equal, even though the absolute difference is the same. This would appear to be due to the increased variance of (6), all else equal, when the values of *SR* are larger, and with a larger variance, power decreases.

Figure 8a: Power for c > 0

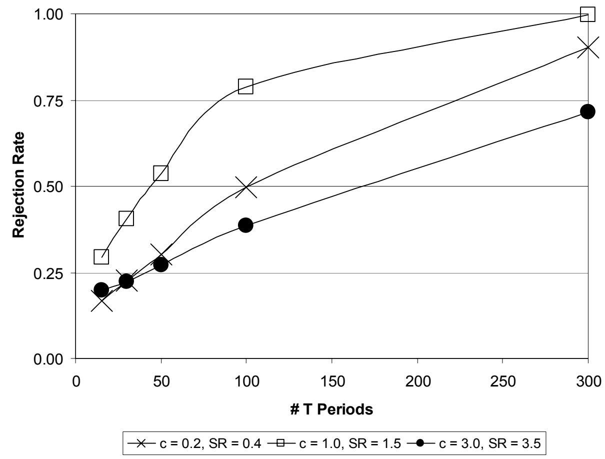


Figure 8b: Relative vs. Absolute Difference

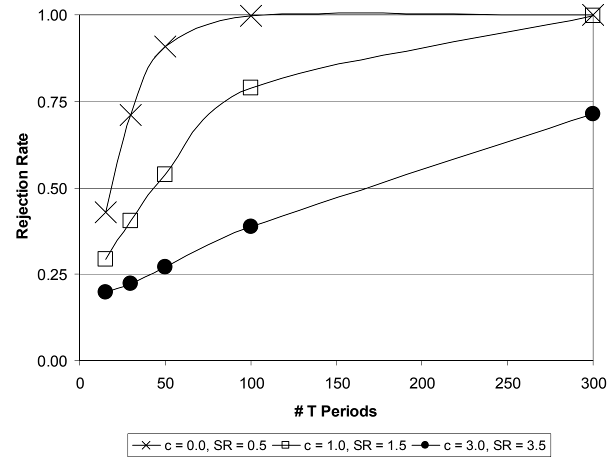


Figure 8. Rejection Rate (N=10,000) of One-Tailed Test (Ho: SR ≤ c), Bias Corrected, by SR by T, under "Real World" Simulated Returns (APD with $\alpha = 0.7$, $\lambda = 1.35$, so $\eta_3 = -1.882$ and $\eta_4 = 5.191$)

Under other combinations of skewness and kurtosis, for both one- and two-sided tests of c > 0, positive skewness generally yields more power than symmetry, which generally yields more power than negative skewness. For smaller values of c and *SR* (c = 0.2, *SR* = 0.4), leptokurtosis yields slightly more power than mesokurtosis under positive skewness, but slightly less power under negative skewness; for larger values (c = 1.0, *SR* = 1.5; and c = 3.0, *SR* = 3.5), leptokurtosis always yields less power. The one-sided test always was noticeably more powerful than the two-sided test. Bias correction, for both one- and two-sided tests, had little affect on power, except under concurrent small samples, positive skewness, and moderately sized c and *SR* (c = 1.0, *SR* = 1.5), where it decreased power noticeably.

B.4 Two-Sample Statistic: Level Control

Under the "real-world" returns conditions simulated by APD-α = 0.7 and APD-λ = 1.35, (13) exhibits excellent convergence to the nominal level of $\alpha = 0.05$ (see Figure 9). In fact, under strong positive correlation between the two returns (see Figure 9a), which is the rule rather than the exception when making apples-to-apples Sharpe ratio comparisons of similarly categorized funds, (13) is actually conservative, never notably violating the nominal level. Only under small samples, larger values of $SR_a = SR_b$, and no or low correlation between the two returns does (13) exhibit slightly inflated levels (see Figures 9c & 9d). Similar results hold for two-sided tests, except under large values of $SR_a = SR_b$ (e.g. $SR_a = SR_b = 3.0$) when convergence is noticeably slower. However, this is only true for low or no correlation between the two funds: this level inflation disappears almost entirely for two-sided tests when the two series of returns are strongly positively correlated, as is typically the case in practice when comparing Sharpe ratios.

Figure 9a: Correlation $\rho = 0.75$

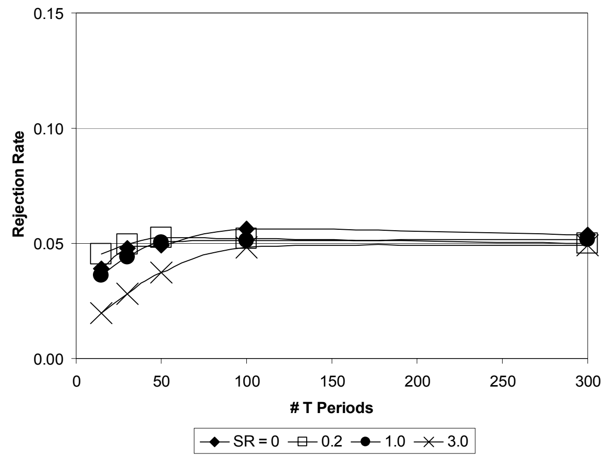


Figure 9b: Correlation $\rho = 0.50$

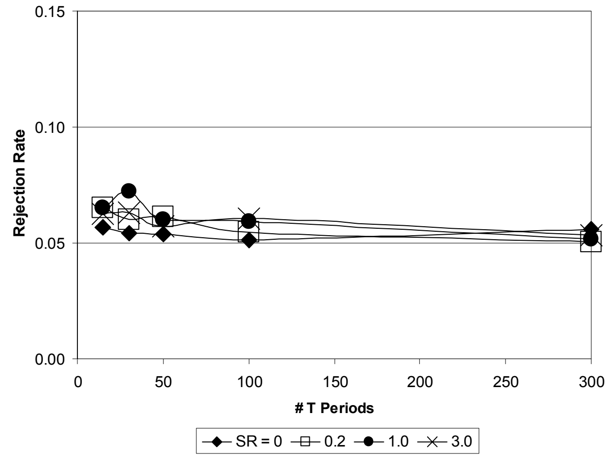


Figure 9c: Correlation $\rho = 0.25$

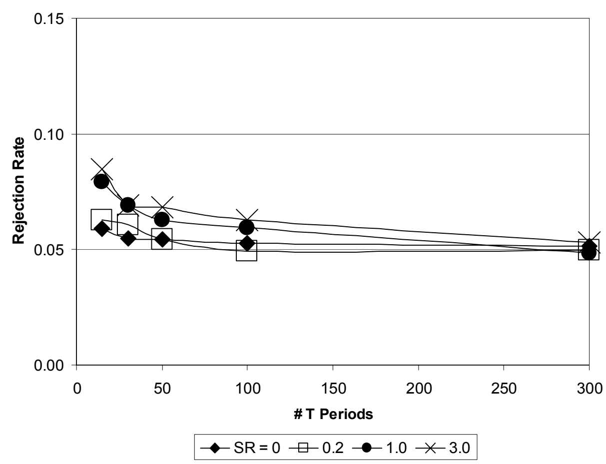


Figure 9d: Correlation $\rho = 0.00$

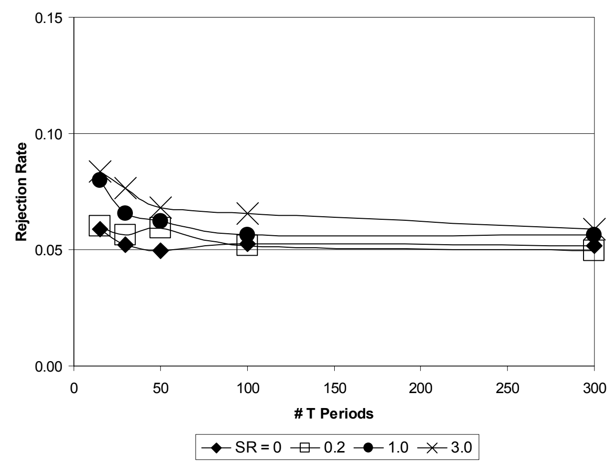


Figure 9. Rejection Rate (N=10,000) of One-Tailed Test (Ho: $SR_{diff} \leq 0$), Bias Corrected, by SR_a=SR_b by ρ by T, under "Real World" Simulated Returns (APD with $\alpha = 0.7$, $\lambda = 1.35$, so $\eta_3 = -1.882$ and $\eta_4 = 5.191$)

Under other combinations of skewness and kurtosis in simulated returns, (13) exhibits generally quick convergence to the nominal level of $\alpha = 0.05$, except under the most extreme conditions: concurrently large values of $SR_a = SR_b$ (= 3.0), extreme leptokurtosis, extreme (positive) skewness, and zero correlation between the two series of returns (see Figure 10a). All else equal, (13) achieves slightly quicker convergence under extreme negative skewness. But even under extreme positive skewness, this level inflation, all else equal, largely evaporates under the more realistic presumption of strong positive correlation between the two returns (see Figure 10b). For two-sided tests, similar patterns of convergence hold for all but large values of $SR_a = SR_b$ (= 3.0), when convergence is noticeably slower under no or low correlation. However, under strong positive correlation between funds, convergence is similar to one-sided tests.

Figure 10a: $SR_a = SR_b = 3.0$, Correlation $\rho = 0.00$

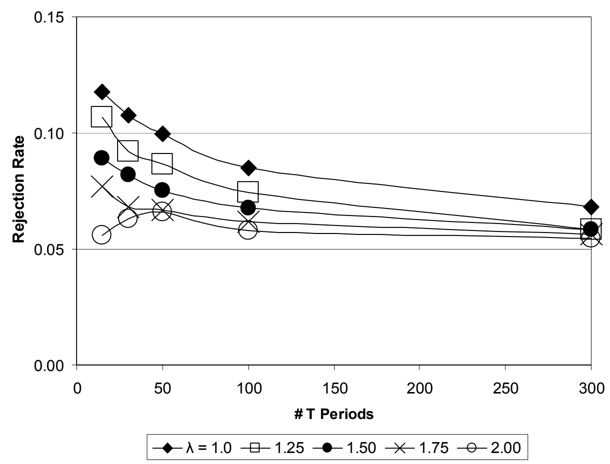


Figure 10b: $SR_a = SR_b = 3.0$, Correlation $\rho = 0.75$

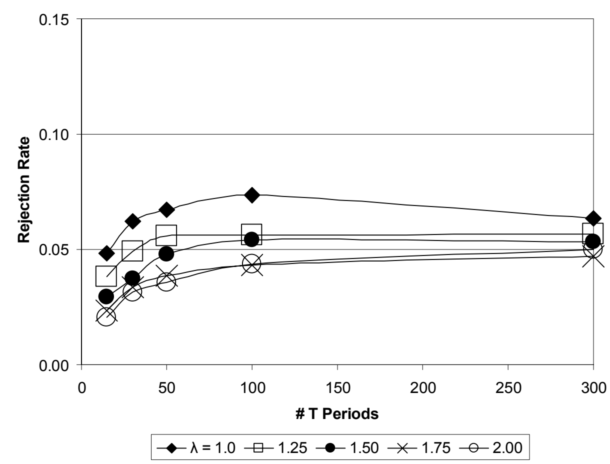


Figure 10. Rejection Rate (N=10,000) of One-Tailed Test (Ho: $SR_{diff} \leq 0$), Bias-Corrected, by ρ by λ by T under Large $SR_a = SR_b$ (= 3.0) and "Extreme" Positive Skewness (APD with α = 0.1, so largest $\eta_3 = 2.23$)

B.5 Two-Sample Statistic: Bias Correction

As expected, bias correction affects results minimally for the two-sample statistic. All two-sample results use bias-corrected point estimates for SR_a and SR_b. As with simulation results for the one-sample statistic, uncorrected estimates are used when calculating variances.

B.6 Two-Sample Statistic: Power

The major finding related to the power of the two-sample statistic (13) is that strong positive correlation between the two series of returns increases power dramatically, under virtually all conditions. An example is shown in Figure 11 under the "real-world" returns conditions simulated by APD-α = 0.7 and APD-λ = 1.35. This finding is important for two reasons: first, it has not been documented adequately in previous research. The only previous study of a two-sample estimator that explicitly examined the effects of correlation between the two returns is Jobson & Korkie (1981). This study only examined between-fund correlations as high as 0.50 (under iid normality), for which it reported only modest power. Yet we can see from the results below that power increases appear to be nonlinear in increases in (positive) correlation: increasing correlation from 0.50 to 0.75 typically increases power far more than increasing it from 0.00 to 0.25, or even from 0.25 to 0.50 (this also can be seen in Figure 2b). This relates to the second and more important point, which is that, as an empirical matter, correlations of 0.75 and above are the rule rather than the exception for most Sharpe ratio comparisons in practice. Day in and day out, most financial analysts are making apples-to-apples comparisons of similarly categorized funds, such as comparisons of competing large growth mutual funds. Not surprisingly, similar *types* of funds are almost always very strongly positively correlated with each other. Pairwise correlations above 0.9 for such funds are not uncommon (as seen in Section V below). Therefore, as it would be used in practice, the two-sample estimator derived in this paper (13) not only is easily calculated and implemented, but also has good power (see Figure 11), contrary to the preliminary results of some earlier research.

Figure 11a: Power – SR_a=0.0, SR_b=0.1

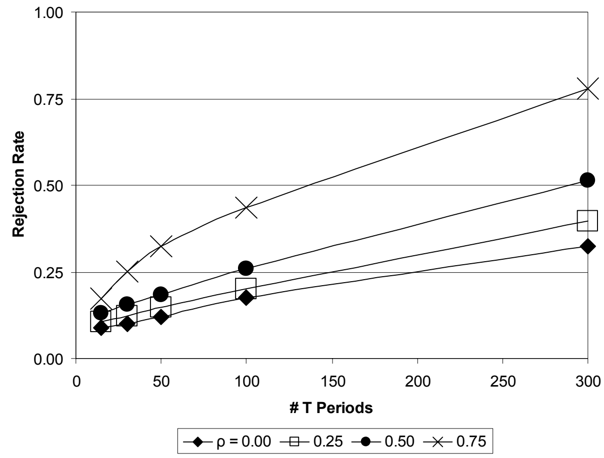


Figure 11b: Power – SR_a=0.0, SR_b=0.2

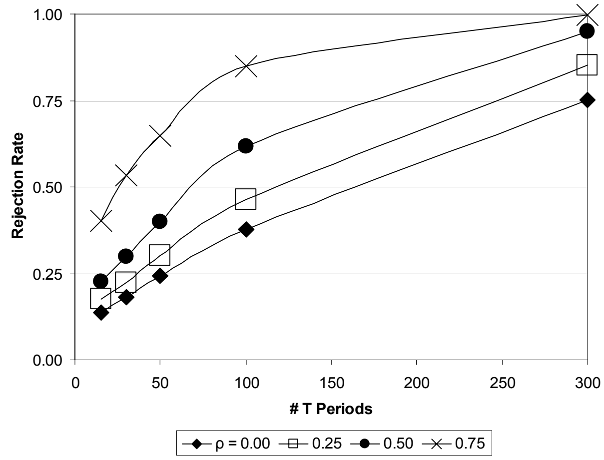


Figure 11c: Power – SR_a=0.0, SR_b=0.5

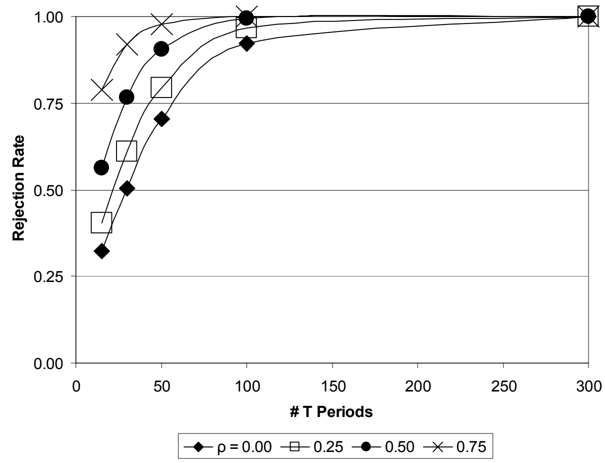


Figure 11d: Power – SR_a=0.2, SR_b=0.4

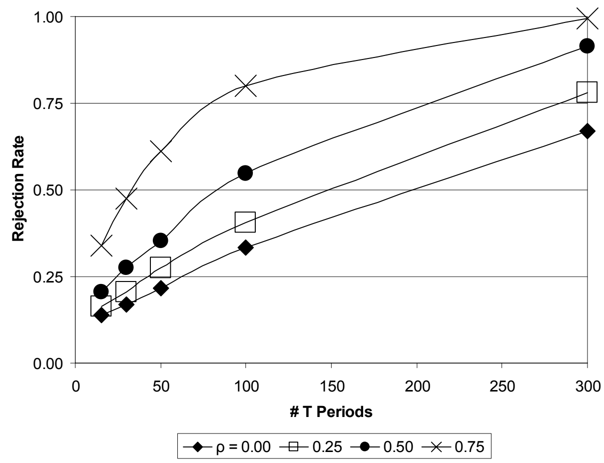


Figure 11e: Power – SR_a=1.0, SR_b=1.5

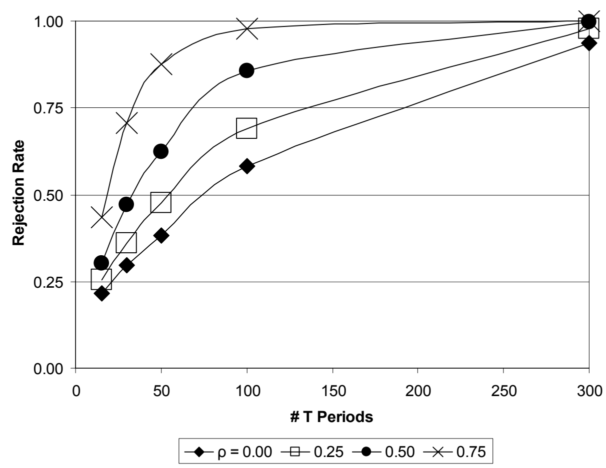


Figure 11f: Power – SR_a=3.0, SR_b=3.5

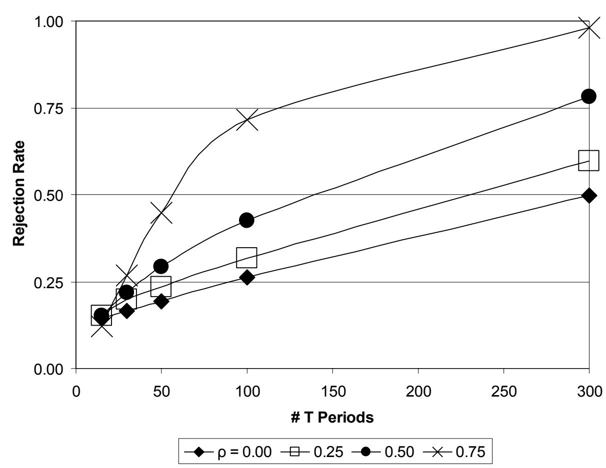


Figure 11. Rejection Rate (N=10,000) of One-Tailed Test (Ho: $SR_{diff} \leq 0$), Bias Corrected, by SR_a & SR_b by ρ by T, under "Real World" Simulated Returns (APD with $\alpha = 0.7$, $\lambda = 1.35$, so $\eta_3 = -1.882$ and $\eta_4 = 5.191$)

Some additional findings for the empirical power of (13), which also are valid for two-tailed coverage, include:

1) all else equal, typically greater power when both returns are positively skewed, as opposed to negatively skewed to the same degree (see Figures 12a & 12b vs. Figures 12c & 12d). This is consistent with the formula of (13), since negative skewness, all else equal, will increase the variance, and consequently decrease power. Evidently, the skewness terms in (13) typically will dominate the opposite-signed "bivariate" skewness terms in the covariance term of (13).

2) all else equal, typically less power under larger kurtosis when the *SR*s of both returns are large, especially under strong positive correlation between the two returns (see Figures 12a & 12c vs. Figures 12b & 12d). This is consistent with the formula of (13), since larger kurtosis will increase the variance, and consequently decrease power. However, since skewness is not independent of kurtosis (even with distinct skewness and kurtosis parameters in APD), positive skewness sometimes can cause slightly greater power under lepto- vs. mesokurtosis, which occurs in this study at times when values of *SR* are smaller.

3) The *relative* size, not just the absolute size, of the difference between the two *SR*s affects power. For example, power to detect a difference is greater for $SR_a = 0.0$ & $SR_b = 0.5$ than it is for $SR_a = 1.0$ & $SR_b = 1.5$ than it is for $SR_a = 3.0$ & $SR_b = 3.5$, all else equal, even though the absolute difference is the same (see Figures 11c, 11e, and 11f). This would appear to be due to the increased variance of (13), all else equal, when the values of *SR* are larger. And with a larger variance, power decreases.

Figure 12a: Extreme Pos. Skew, Leptokurtosis (Laplacian)
(α = 0.1 and λ = 1.0)

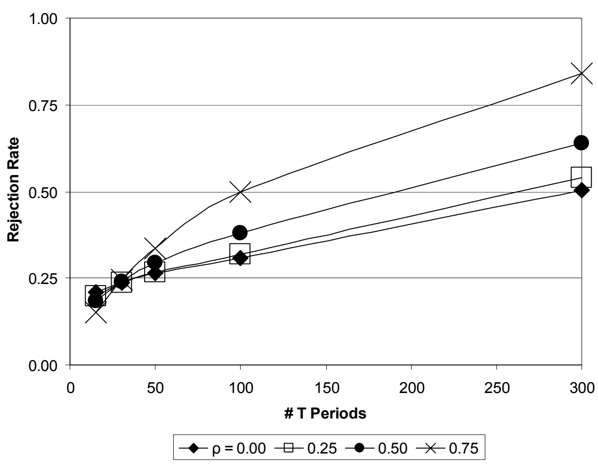


Figure 12b: Extreme Pos. Skew, Mesokurtosis (Normal)
(α = 0.1 and λ = 2.0)

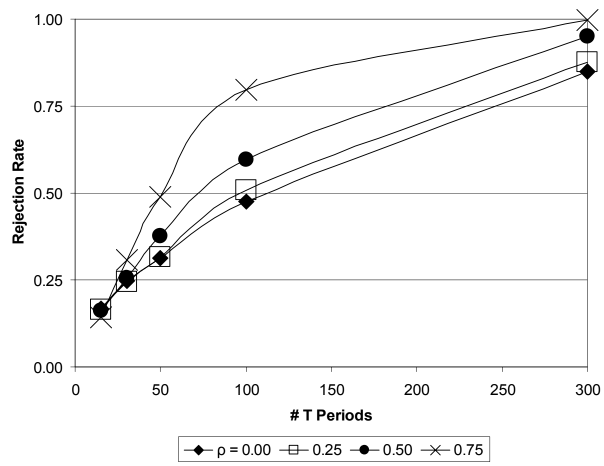


Figure 12c: Extreme Neg. Skew, Leptokurtosis (Laplacian)
(α = 0.9 and λ = 1.0)

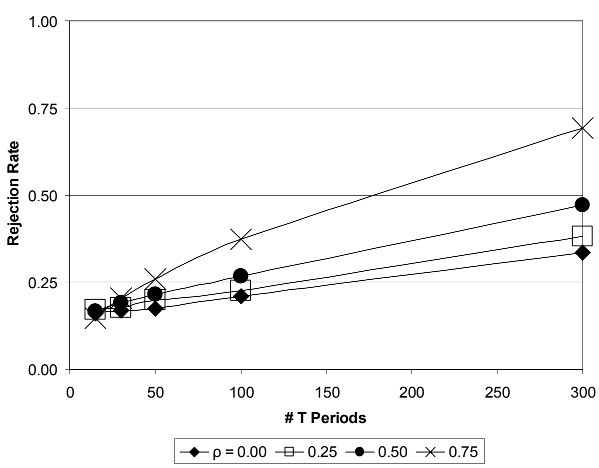


Figure 12d: Extreme Neg. Skew, Mesokurtosis (Normal)
(α = 0.9 and λ = 2.0)

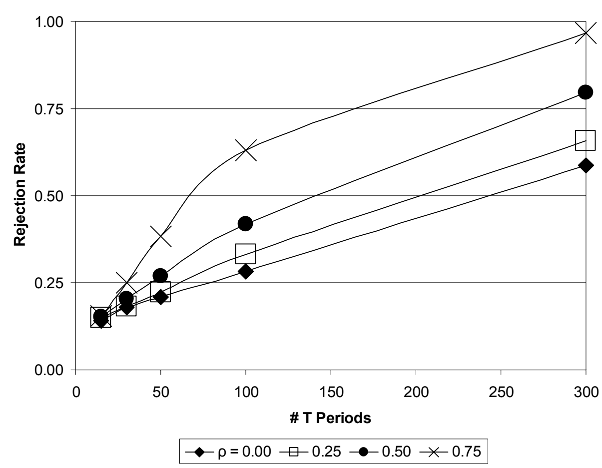


Figure 12: Rejection Rates (N=10,000) of One-Tailed Test (Ho: $SR_{diff} \leq 0$), SR_a=3.0 & SR_b=3.5, by α by λ by ρ by T

B.7 Results Summary

Table II: Summary of Simulation Study Results

Level / Power	One-sample Estimator (6) (Ho: $SR \le$ c, Ha: $SR >$ c)	Two-sample Estimator (13) (Ho: $SR_{diff} \le 0$, Ha: $SR_{diff} > 0$)
Level Control (Type I error)	For c = 0, excellent For c > 0, generally acceptable	For $\rho \ge 0.75$, excellent otherwise, generally acceptable
Power (1 – Type II error)	For c = 0, modest For c > 0, fairly low	For $\rho \ge 0.75$, good, and excellent as $\rho \to 1.0$ otherwise, fairly low

V. Sharpe Ratio Comparisons of Actual Mutual Fund Returns

The practical purpose for deriving the estimators (6) and (13) is to test, under very general, real-world conditions, the hypotheses a) that a Sharpe ratio is larger than zero – i.e. that the market behavior of the security, when adjusted for risk, reflects positive excess returns – and b) that the Sharpe ratio of one fund is larger than that of another. The latter hypothesis is posed implicitly, thousands of times daily, whenever mutual funds are ranked according to their Sharpe ratios. However, the implicit pairwise comparisons on these lists[24] never are accompanied by confidence intervals or p-values indicating whether the larger *SR* is actually larger *with statistical significance*, rather than simply as an artifact of random chance (i.e. due solely to the volatility of returns data). To be able to say, "Fund Y's Sharpe ratio is larger than that of Fund X, with over 95% confidence." would be valuable for a wide variety of purposes: indeed, wherever Sharpe ratios are used to assess and compare risk-adjusted performance. Thus, (6) and (13) are applied below to the actual returns data of twenty mutual funds: the top 20 large growth mutual funds by net assets as of 09/06/06 as obtained from http://finance.yahoo.com. Weekly returns, calculated from opening price to closing price,[25] were obtained for the three year period from the week of 12/24/03 to that of 12/20/06. The risk free rate used is the 90-day U.S. treasury bill (nominal) (series TCMNOMM3) downloaded from the Federal Reserve Board website at http://www.federalreserve.gov/releases/h15/data.htm. The arithmetic mean was used to enforce the constant risk-free rate assumption.[26] All the returns data and a flexible, fully parameterized SAS® program implementing these results can be downloaded from the author's website at www.DataMineIt.com.

The results (see Tables III, IV and V) show the major finding is that which we draw from the simulation study: strong positive correlation between funds appears to dramatically increase what might otherwise be lackluster power. This can be inferred from two results: first, between-funds correlations shown in the funds' correlation matrix (Table V) roughly match the corresponding matrix of two-sample p-values a la (13) (Table IV), with higher correlations generally matching more significant (smaller) p-values. Secondly, the Sharpe ratios of individual funds whose performance is in the top of their class do not achieve statistical significance a la (6) for the one-sided test of positive excess returns (i.e. Ho: $SR \le 0$ vs. Ha: $SR > 0$), at least at the $\alpha = 0.05$ level (two are significant at $\alpha = 0.10$ – see Table III). However, when the Sharpe ratios of the top four or five funds are compared a la (13) to those of their competitors, most of whom have Sharpe ratios well above zero, the top performers are better, with statistical significance, than about half of their competitors at α = 0.05. This is due, of course, to the strong positive correlation between these funds and their competitors (almost four fifths – 149 of 190 between-fund correlations – exceed $\rho \ge 0.9$). Mathematically, this strong positive correlation increases the covariance term in (13) which, when subtracted from the overall variance, decreases it notably, thus increasing power. It is the apparent magnitude of this effect, the fact that it has been missed in earlier research, and the fact that the vast majority of Sharpe ratio comparisons in practice will involve funds that are strongly, positively correlated with each other, that makes it a very noteworthy finding.[27]

[24] In situations where many related hypothesis tests are being conducted and the cost of type I error (false positives) is high (e.g. genome research), multiple comparisons procedures often are used to control the family-wise error rate (FWE) or the false discovery rate (FDR) rather than the pairwise error rate (i.e., α). Although the objective here as shown in Table IV is different – only to examine specific columns of interest individually – such procedures could be very useful in this setting if the hypotheses being examined *do* involve sizeable numbers of multiple comparisons. See J.D. Storey (2002, 2003, 2004, 2007) and J. Hsu (1996) for details.

[25] Kelly (2007) relies on the Sharpe ratio, and the estimators derived herein ((6) and (13)), to test whether open-to-close ETF returns are different from close-to-open (after hours trading) ETF returns. He finds large differences with strong statistical significance.

[26] Identical results were obtained when the variable risk-free rate was incorporated into the returns themselves, confirming that, as an empirical matter, the simplifying assumption of a constant risk-free rate is acceptable for practical usage.

[27] This finding is similar to and consistent with that of Pastor & Stambaugh (2003) who showed that the use of returns of seemingly unrelated assets, which often are correlated with the particular fund being examined, can dramatically increase the precision with which one can estimate the *SR* of that particular fund, and that the estimate of *SR* can differ dramatically as a result.

Table III – Sharpe Ratio Rankings of Top 20 Large Growth Mutual Funds by Net Assets (as of 09/06/06): Weekly Returns, 8/20/03-8/30/06

Fund Name	Symbol	Rank	Sharpe Ratio	SR–Bias Corrected	Prob(SR>0)	Mean	Risk-Free Rate	Standard Deviation	Median	Skewness $\sqrt{b_1}$	Kurtosis b_2
American Funds Grth Fund of Amer F	GFAFX	1	0.1070	0.1067	0.905	0.00202	0.00052	0.01401	0.00333	-0.40932	2.82945
American Funds Grth Fund of Amer A	AGTHX	2	0.1059	0.1056	0.902	0.00200	0.00052	0.01401	0.00316	-0.41639	2.84999
American Funds Grth Fund of Amer C	GFACX	3	0.1020	0.1017	0.894	0.00194	0.00052	0.01401	0.00321	-0.42284	2.86003
American Funds Grth Fund of Amer B	AGRBX	4	0.1018	0.1015	0.894	0.00194	0.00052	0.01398	0.00326	-0.41823	2.85798
Fidelity Contrafund	FCNTX	5	0.0939	0.0937	0.874	0.00195	0.00052	0.01526	0.00292	-0.52592	2.79629
Janus Twenty	JAVLX	6	0.0906	0.0903	0.867	0.00203	0.00052	0.01672	0.00262	-0.40782	2.88478
Janus Growth & Income	JAGIX	7	0.0831	0.0828	0.847	0.00175	0.00052	0.01490	0.00336	-0.32041	3.10304
American Funds New Economy A	ANEFX	8	0.0820	0.0817	0.845	0.00177	0.00052	0.01528	0.00390	-0.29591	2.63672
Fidelity Growth Company	FDGRX	9	0.0661	0.0659	0.792	0.00182	0.00052	0.01970	0.00345	-0.54738	3.32944
T. Rowe Price Growth Stock	PRGFX	10	0.0631	0.0629	0.784	0.00140	0.00052	0.01395	0.00263	-0.13681	3.05440
Janus	JANSX	11	0.0454	0.0452	0.714	0.00121	0.00052	0.01539	0.00209	-0.17905	3.09362
T. Rowe Price Blue Chip Growth	TRBCX	12	0.0434	0.0432	0.706	0.00114	0.00052	0.01446	0.00169	-0.18664	3.06391
American Funds Amcap A	AMCPX	13	0.0400	0.0399	0.691	0.00100	0.00052	0.01215	0.00154	-0.22222	3.10797
AIM Constellation A	CSTGX	14	0.0368	0.0367	0.676	0.00116	0.00052	0.01738	0.00242	-0.42497	3.08331
Fidelity OTC	FOCPX	15	0.0336	0.0335	0.662	0.00123	0.00052	0.02131	0.00360	-0.33902	3.37859
Fidelity Capital Appreciation	FDCAX	16	0.0309	0.0307	0.649	0.00113	0.00052	0.01997	0.00198	-0.73058	4.27309
Harbor Capital Appreciation Instl	HACAX	17	0.0296	0.0295	0.644	0.00098	0.00052	0.01576	0.00184	-0.36896	2.97183
Vanguard Growth Index	VIGRX	18	0.0109	0.0109	0.554	0.00067	0.00052	0.01415	0.00115	-0.21855	3.08313
Fidelity Blue Chip Growth	FBGRX	19	0.0088	0.0088	0.544	0.00064	0.00052	0.01354	0.00137	-0.18337	2.79949
American Century Ultra Inv	TWCUX	20	-0.0226	-0.0225	0.388	0.00018	0.00052	0.01467	0.00137	-0.27888	3.07084

Table IV – All Pairwise P-Values of Ho: $SR_a \leq SR_b$ using (13), Top 20 Large Growth Mutual Funds by Net Assets (as of 09/06/06): Weekly Returns, 8/20/03-8/30/06

Fund	GFAFX	AGTHX	GFACX	AGRBX	FCNTX	JAVLX	JAGIX	ANEFX	FDGRX	PRGFX	JANSX	TRBCX	AMCPX	CSTGX	FOCPX	FDCAX	HACAX	VIGRX	FBGRX
AGTHX	-1.000																		
GFACX	-1.000	-1.000																	
AGRBX	-1.000	-1.000	-1.000																
FCNTX	0.334	0.348	0.395	0.398															
JAVLX	0.348	0.359	0.393	0.394	0.468														
JAGIX	0.207	0.218	0.256	0.258	0.375	0.430													
ANEFX	0.185	0.196	0.236	0.237	0.370	0.429	0.488												
FDGRX	0.117	0.123	0.146	0.146	0.235	0.297	0.329	0.312											
PRGFX	0.087	0.092	0.111	0.111	0.205	0.260	0.271	0.260	0.465										
JANSX	**0.023**	**0.025**	**0.031**	**0.031**	0.098	0.172	0.120	0.102	0.260	0.279									
TRBCX	**0.022**	**0.024**	**0.030**	**0.030**	0.085	0.129	0.107	0.098	0.236	**-1.000**	0.472								
AMCPX	**0.018**	**0.020**	**0.025**	**0.025**	0.077	0.129	0.112	0.095	0.219	0.209	0.428	0.448							
CSTGX	**0.004**	**0.005**	**0.006**	**0.006**	**0.035**	0.101	0.054	0.057	0.132	0.179	0.362	0.401	0.453						
FOCPX	**0.022**	**0.024**	**0.030**	**0.029**	0.072	0.128	0.102	0.055	0.088	0.194	0.357	0.380	0.427	0.455					
FDCAX	**0.042**	**0.044**	0.053	0.053	0.084	0.138	0.133	0.121	0.217	0.238	0.367	0.389	0.421	0.442	0.475				
HACAX	**0.007**	**0.008**	**0.011**	**0.010**	**0.039**	0.076	0.068	**0.034**	0.080	0.103	0.292	0.278	0.351	0.383	0.437	0.489			
VIGRX	**0.001**	**0.001**	**0.002**	**0.002**	**0.014**	**0.036**	**0.015**	**0.008**	**0.027**	**0.006**	0.084	**0.034**	0.086	0.132	0.225	0.328	0.191		
FBGRX	**0.001**	**0.001**	**0.002**	**0.002**	**0.013**	**0.031**	**0.013**	**0.008**	**0.029**	**0.002**	0.087	**0.015**	0.100	0.136	0.218	0.317	0.180	0.431	
TWCUX	**0.000**	**0.000**	**0.000**	**0.000**	**0.002**	**0.006**	**0.003**	**0.001**	**0.003**	**0.000**	**0.014**	**0.001**	**0.006**	**0.020**	**0.043**	0.126	**0.019**	**0.035**	0.070

Table V – Correlation Matrix, Top 20 Large Growth Mutual Funds by Net Assets (as of 09/06/06): Weekly Returns, 8/20/03-8/30/06

Fund	GFAFX	AGTHX	GFACX	AGRBX	FCNTX	JAVLX	JAGIX	ANEFX	FDGRX	PRGFX	JANSX	TRBCX	AMCPX	CSTGX	FOCPX	FDCAX	HACAX	VIGRX	FBGRX
AGTHX	1.000																		
GFACX	1.000	1.000																	
AGRBX	1.000	1.000	1.000																
FCNTX	0.955	0.955	0.955	0.955															
JAVLX	0.891	0.891	0.892	0.892	0.894														
JAGIX	0.956	0.956	0.956	0.957	0.928	0.877													
ANEFX	0.964	0.964	0.963	0.964	0.921	0.844	0.925												
FDGRX	0.929	0.929	0.929	0.930	0.904	0.853	0.900	0.931											
PRGFX	0.942	0.943	0.943	0.944	0.913	0.876	0.936	0.946	0.921										
JANSX	**0.947**	**0.947**	**0.948**	**0.948**	0.909	0.838	0.933	0.949	0.931	0.938									
TRBCX	**0.943**	**0.943**	**0.944**	**0.945**	0.914	0.880	0.936	0.942	0.932	0.993	0.944								
AMCPX	**0.941**	**0.941**	**0.942**	**0.942**	0.907	0.859	0.916	0.932	0.921	0.945	0.938	0.953							
CSTGX	**0.963**	**0.964**	**0.964**	**0.964**	**0.940**	0.877	0.946	0.946	0.956	0.943	0.960	0.952	0.946						
FOCPX	**0.914**	**0.915**	**0.914**	**0.915**	0.885	0.818	0.893	0.938	0.963	0.915	0.924	0.923	0.907	0.940					
FDCAX	**0.876**	**0.875**	0.875	0.875	0.862	0.787	0.845	0.869	0.859	0.853	0.869	0.856	0.845	0.879	0.861				
HACAX	**0.941**	**0.941**	**0.942**	**0.943**	**0.915**	0.872	0.914	**0.945**	0.957	0.951	0.941	0.962	0.947	0.959	0.954	0.857			
VIGRX	**0.944**	**0.944**	**0.945**	**0.945**	**0.908**	**0.862**	**0.927**	**0.942**	**0.945**	**0.972**	0.955	**0.979**	0.968	0.962	0.932	0.852	0.967		
FBGRX	**0.940**	**0.940**	**0.940**	**0.941**	**0.905**	**0.867**	**0.928**	**0.940**	**0.939**	**0.978**	0.948	**0.984**	0.957	0.954	0.924	0.842	0.963	0.990	
TWCUX	**0.916**	**0.916**	**0.916**	**0.917**	**0.898**	**0.856**	**0.899**	**0.927**	**0.927**	**0.954**	**0.929**	**0.966**	**0.956**	**0.938**	**0.919**	0.838	**0.953**	**0.973**	0.964

VI. Conclusions

The presumption of normal stock returns has been widely refuted in the empirical finance literature. Yet until recently, since Jobson & Korkie (1981) the only formulaic derivation of the asymptotic distribution of the Sharpe ratio that was easily used for generating confidence intervals or for conducting hypothesis tests relied on this restrictive and incorrect distributional assumption. This study simplifies a very recent and somewhat complex derivation to a much simpler, more easily implemented, and more readily understood result, but one that still is valid under the very general conditions of Christie's (2005) more complex estimator – namely, stationary and ergodic returns. Thus is the one-sample estimator presented here valid under realistic financial returns conditions that include serial correlation, time-varying conditional volatilities, and other non-iid behavior, along with negative skewness and leptokurtosis. Also derived is an analogous two-sample statistic for comparing two Sharpe ratios, and both estimators are tested in an extensive simulation study. Under real world conditions, both estimators, when used with a simple and generally valid bias-correction presented here, demonstrate good level control. Power of the one-sample estimator appears to be less impressive than that of the two-sample estimator because strong positive correlation between funds dramatically increases the power of the latter. This finding is notable for several reasons: a) it was missed in previous research on two-sample estimators for comparing Sharpe ratios, in part because the other estimators (Christie, 2005, and Vinod & Morey, 2000) are not based on a straightforward formula from a distributional derivation; b) the size of the effect is dramatic as between-fund correlation increases above 0.5; and c) the vast majority of funds whose Sharpe ratios are being compared are, in fact, very strongly, positively correlated, with between-fund correlations typically above 0.8. This is due to the fact that almost all comparisons, in practice, are between competing funds that are similarly categorized – e.g., large growth funds are being compared against other competing large growth funds (not against small value funds). Consequently, as it would be used in practice, the two-sample estimator has good power, a finding that has not been presented in previous research on this topic.

This study concludes with an application of the one- and two-sample estimators to returns data from twenty mutual funds, with the former yielding few statistically significant results and the latter yielding many. The two-sample statistic, therefore, would appear to have great utility when used in conjunction with the ubiquitous mutual fund and hedge fund rankings based on Sharpe ratios. Although very important in the industry, these rankings never are accompanied by p-values indicating whether differences between Sharpe ratios are, in fact, statistically significant, rather than produced simply as an artifact of the volatility of financial returns. Without such statistical discipline, the widespread ranking exercise is of little *inferential* value.

To conclude, although recent empirical evidence (see Eling & Schuhmacher, 2007) suggests that the Sharpe ratio is, after all, just as good at ranking risk-adjusted performance as numerous other far more complex metrics, even under "difficult," highly non-normal hedge fund returns data, debate about its utility per se in this regard is beyond the scope of this paper. What this paper does provide is a scientific basis for *easily and effectively* using the Sharpe ratio *inferentially* under very general, real-world conditions. This will help to inform such debates in the arena of financial analysis, and bring statistical discipline to the widespread usage of the Sharpe ratio in the industry. In other settings, the mean-divided-by-standard-deviation statistic, and its generally valid large-sample distribution presented in this paper, may have even greater practical utility, if not ubiquity.

APPENDIX A – Alternate Derivation of the Distribution of $\widehat{SR}$ Under IID via the Delta Method

According to Stuart & Ord (1994), pp. 350-352, the "delta method" can be used to derive the variance of a function, $g(x_1, x_2, x_3, \ldots)$, of a number of random variables, and if the function is the ratio of two random variables, then:

$$\operatorname{var}(x_1/x_2) = \left[\frac{E(x_1)}{E(x_2)}\right]^2 \left[\frac{\operatorname{var}(x_1)}{E(x_1)^2} + \frac{\operatorname{var}(x_2)}{E(x_2)^2} - 2\frac{\operatorname{cov}(x_1, x_2)}{E(x_1)E(x_2)}\right] \tag{A1}$$

If $x_1 = \mu$ and $x_2 = \sigma$ then

$$\operatorname{var}(\mu/\sigma) = \frac{\mu^2}{\sigma^2}\left[\frac{\sigma^2/n}{\mu^2} + \frac{(\mu_4 - \sigma^4)/4n\sigma^2}{\sigma^2} - 2\frac{\mu_3/(2n\sigma)}{\mu\sigma}\right] \tag{A2}$$

$$\operatorname{var}(\mu/\sigma) = \frac{1}{n}\frac{\mu^2}{\sigma^2}\left[\frac{\sigma^2}{\mu^2} + \frac{(\mu_4 - \sigma^4)}{4\sigma^4} - \frac{\mu_3}{\mu\sigma^2}\right] \tag{A3}$$

since $x_1 = \mu - R_{ft}$ yields the same results above, we can treat $\mu/\sigma = SR$, so

$$\operatorname{var}(\mu/\sigma) = \frac{1}{n}\left[1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4} - 1\right] - \frac{\mu\mu_3}{\sigma\sigma^3}\right] \tag{A4}$$

$$\operatorname{var}(\mu/\sigma) = \frac{1}{n}\left[1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4} - 1\right] - SR\frac{\mu_3}{\sigma^3}\right] \tag{A5}$$

which is Merten's (2002) result. The more technical conditions required for the valid use of the delta method are discussed in Appendix C.

APPENDIX B – Equivalence of Merten's (2002) iid, and Christie's (2005) GMM, Derivations of the Distribution of $\widehat{SR}$

Under only the requirements of stationarity and ergodicity, Christie (2005) derives (C21),

$$Var\left(\sqrt{T}\,\widehat{SR}\right) = E\left[\frac{SR^2\mu_4}{4\sigma^4} - \frac{SR\left[\left(R_t - R_{ft}\right)\left(R_t-\mu\right)^2 - \left(R_t - R_{ft}\right)\sigma^2\right]}{\sigma^3} + \frac{\left(R_t - R_{ft}\right)^2}{\sigma^2} - \frac{2\left(R_t - R_{ft}\right)SR}{\sigma} + \frac{3SR^2}{4}\right] \qquad \text{(C21)}$$

which can be simplified as below[28]:

$$= \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}+3\right] - SR\cdot E\left[\frac{\left(R_t-R_f\right)\left(R_t-\mu\right)^2}{\sigma^3}\right] - SR\cdot E\left[\frac{\left(R_t-R_f\right)}{\sigma}\right] + E\left[\frac{R_t^2 - 2R_tR_f + R_f^2}{\sigma^2}\right]$$

since $E\left[R_t^2\right] = \sigma^2 + \mu^2$,

$$= \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}+3\right] - SR\cdot E\left[\frac{\left(R_t-R_f\right)\left(R_t^2 - 2\mu R_t + \mu^2\right)}{\sigma^3}\right] - SR^2 + \frac{\sigma^2+\mu^2-2\mu R_f + R_f^2}{\sigma^2}$$

$$= \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}+3\right] - SR\cdot E\left[\frac{R_t^3 - 2\mu R_t^2 + \mu^2 R_t - R_t^2 R_f + 2\mu R_t R_f - \mu^2 R_f}{\sigma^3}\right] - SR^2 + 1 + SR^2$$

since $E\left[R_t^3\right] = \mu_3 + 3\sigma^2\mu + \mu^3$,

$$= 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}+3\right] - SR\cdot\left[\frac{\mu_3 + 3\mu\sigma^2 + \mu^3 - 2\mu\left(\sigma^2+\mu^2\right) + \mu^3 - \left(\sigma^2+\mu^2\right)R_f + 2\mu^2 R_f - \mu^2 R_f}{\sigma^3}\right]$$

$$= 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}+3\right] - SR\cdot\left[\frac{\mu_3 + 3\mu\sigma^2 + \mu^3 - 2\mu\sigma^2 - 2\mu^3 + \mu^3 - \sigma^2 R_f - \mu^2 R_f + \mu^2 R_f}{\sigma^3}\right]$$

$$= 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}+3\right] - SR\cdot\left[\frac{\mu_3 + 3\mu\sigma^2 - 2\mu\sigma^2 - \sigma^2 R_f}{\sigma^3}\right]$$

$$= 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}+3\right] - SR\cdot\left[\frac{\mu_3}{\sigma^3} + \frac{\mu - R_f}{\sigma}\right]$$

$$= 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}+3\right] - SR\frac{\mu_3}{\sigma^3} - SR^2$$

$$= 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}-1\right] - SR\frac{\mu_3}{\sigma^3}$$

So $Var\left(\sqrt{T}\,\widehat{SR}\right) = 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}-1\right] - SR\frac{\mu_3}{\sigma^3}$ which is Merten's (2002) result, and that derived in Appendix A.

[28] As previously mentioned, even if the variance of the risk-free rate is not literally zero, as is often the case, as a practical empirical matter it can be treated as zero, and its arithmetic mean used as the presumed constant rate (so above, let $R_{ft} = \hat{\mu}_f = R_f$). Covariances of the risk-free rate with fund returns, too, can be treated as zero as an empirical matter. Mathematically, these assumptions are necessary for the above simplification.

APPENDIX C – Variance of the Difference Between Two Sharpe Ratios

If SR_a and SR_b are the respective Sharpe ratios for the returns (R_{at} and R_{bt}) of funds "a" and "b," then use the "delta method"[29] (see Greene, 1993, and Stuart & Ord, 1994) to obtain the asymptotic variance of $\left(\widehat{SR}_a - \widehat{SR}_b\right) - \left(SR_a - SR_b\right)$: Assuming $\sigma_f^2 = 0$, which is always essentially, if not literally true, $SR = \dfrac{\mu - R_f}{\sigma} = f\left(\mu, \sigma^2\right)$, so let $u = \left(\mu_a, \mu_b, \sigma_a^2, \sigma_b^2\right)$ and $\hat{u} = \left(\hat{\mu}_a, \hat{\mu}_b, \hat{\sigma}_a^2, \hat{\sigma}_b^2\right)$, then $\sqrt{T}\left(\hat{u} - u\right) \sim N\left(0, \Omega\right)$ where Ω is the variance-covariance matrix of u:

$$\Omega = \begin{pmatrix} \sigma_a^2 & \sigma_{a,b} & \mu_{3a} & \mu_{1a,2b} \\ \sigma_{a,b} & \sigma_b^2 & \mu_{1b,2a} & \mu_{3b} \\ \mu_{3a} & \mu_{1b,2a} & \left(\mu_{4a} - \sigma_a^4\right) & Cov\left(\sigma_a^2, \sigma_b^2\right) \\ \mu_{1a,2b} & \mu_{3b} & Cov\left(\sigma_a^2, \sigma_b^2\right) & \left(\mu_{4b} - \sigma_b^4\right) \end{pmatrix}$$

where $\sigma_{a,b} = Cov\left(a,b\right)$, $\mu_{3a} = E\left[\left(a - \mu_a\right)^3\right] = Cov\left(\mu_a, \sigma_a^2\right)$,

$\mu_{3b} = E\left[\left(b - \mu_b\right)^3\right] = Cov\left(\mu_b, \sigma_b^2\right)$ (see Mertens, 2002), $\mu_{1a,2b} = E\left[\left(a - \mu_a\right)\left(b - \mu_b\right)^2\right] = Cov\left(\mu_a, \sigma_b^2\right)$,

and $\mu_{1b,2a} = E\left[\left(b - \mu_b\right)\left(a - \mu_a\right)^2\right] = Cov\left(\mu_b, \sigma_a^2\right)$ (see Espejo & Singh, 1999). Now,

$$\sqrt{T}\left(\left(\widehat{SR}_a - \widehat{SR}_b\right) - \left(SR_a - SR_b\right)\right) \sim N\left(0, Var_{diff}\right),\quad Var_{diff} = \left(\frac{\partial f}{\partial u}\right)\Omega\left(\frac{\partial f}{\partial u}\right)',\quad \frac{\partial f}{\partial u} = \left(\frac{1}{\sigma_a}, -\frac{1}{\sigma_b}, -\frac{\left(\mu_a - R_f\right)}{2\sigma_a^3}, \frac{\left(\mu_b - R_f\right)}{2\sigma_b^3}\right),$$

$$Var_{diff} = 1 - \frac{\sigma_{a,b}}{\sigma_a\sigma_b} - \frac{\mu_{3a}\left(\mu_a - R_f\right)}{2\sigma_a^4} + \frac{\mu_{1a,2b}\left(\mu_b - R_f\right)}{2\sigma_a\sigma_b^3} - \frac{\sigma_{a,b}}{\sigma_a\sigma_b} + 1 + \frac{\mu_{1b,2a}\left(\mu_a - R_f\right)}{2\sigma_a^3\sigma_b} - \frac{\mu_{3b}\left(\mu_b - R_f\right)}{2\sigma_b^4} - \frac{\mu_{3a}\left(\mu_a - R_f\right)}{2\sigma_a^4} + \frac{\mu_{1b,2a}\left(\mu_a - R_f\right)}{2\sigma_a^3\sigma_b}$$

$$+ \frac{\left(\mu_{4a} - \sigma_a^4\right)\left(\mu_a - R_f\right)^2}{4\sigma_a^6} - \frac{\left(\mu_a - R_f\right)\left(\mu_b - R_f\right)Cov\left(\sigma_a^2, \sigma_b^2\right)}{4\sigma_a^3\sigma_b^3} + \frac{\mu_{1a,2b}\left(\mu_b - R_f\right)}{2\sigma_a\sigma_b^3} - \frac{\mu_{3b}\left(\mu_b - R_f\right)}{2\sigma_b^4} - \frac{\left(\mu_a - R_f\right)\left(\mu_b - R_f\right)Cov\left(\sigma_a^2, \sigma_b^2\right)}{4\sigma_a^3\sigma_b^3} + \frac{\left(\mu_{4b} - \sigma_b^4\right)\left(\mu_b - R_f\right)^2}{4\sigma_b^6}$$

$$= 2 - 2\rho_{a,b} - \frac{\mu_{3a}\left(\mu_a - R_f\right)}{\sigma_a^4} - \frac{\mu_{3b}\left(\mu_b - R_f\right)}{\sigma_b^4} + \frac{\mu_{1b,2a}\left(\mu_a - R_f\right)}{\sigma_a^3\sigma_b} + \frac{\mu_{1a,2b}\left(\mu_b - R_f\right)}{\sigma_a\sigma_b^3} + \frac{\left(\mu_{4a} - \sigma_a^4\right)\left(\mu_a - R_f\right)^2}{4\sigma_a^6} + \frac{\left(\mu_{4b} - \sigma_b^4\right)\left(\mu_b - R_f\right)^2}{4\sigma_b^6} - \frac{\left(\mu_a - R_f\right)\left(\mu_b - R_f\right)Cov\left(\sigma_a^2, \sigma_b^2\right)}{2\sigma_a^3\sigma_b^3}$$

$$= 2 - 2\rho_{a,b} - SR_a\frac{\mu_{3a}}{\sigma_a^3} - SR_b\frac{\mu_{3b}}{\sigma_b^3} + SR_a\frac{\mu_{1b,2a}}{\sigma_b\sigma_a^2} + SR_b\frac{\mu_{1a,2b}}{\sigma_a\sigma_b^2} + \frac{SR_a^2}{4}\left[\frac{\left(\mu_{4a} - \sigma_a^4\right)}{\sigma_a^4}\right] + \frac{SR_b^2}{4}\left[\frac{\left(\mu_{4b} - \sigma_b^4\right)}{\sigma_b^4}\right] - \frac{SR_aSR_b}{2}\frac{Cov\left(\sigma_a^2, \sigma_b^2\right)}{\sigma_a^2\sigma_b^2}$$

Since $= Var\left(\sigma_a^2\right) = Cov\left(\sigma_a^2, \sigma_a^2\right) = \mu_{4a} - \sigma_a^4 = E\left[\left(a - E\left[a\right]\right)^4\right] - \sigma_a^4 = E\left[\left(a - E\left[a\right]\right)^2\left(a - E\left[a\right]\right)^2\right] - \sigma_a^2\sigma_a^2$,

then $Cov\left(\sigma_a^2, \sigma_b^2\right) = E\left[\left(a - E\left[a\right]\right)^2\left(b - E\left[b\right]\right)^2\right] - \sigma_a^2\sigma_b^2 = \mu_{2a,2b} - \sigma_a^2\sigma_b^2$, where $\mu_{2a,2b}$ is the joint second central

---

[29] The delta method is a widely used technique that provides an asymptotic approximation of the variance of a particular function (see Greene, 1993, pp.297-298, and Stuart & Ord, 1994, p.350). It is valid as long as the random variables used in the function are asymptotically normal, and the function is (loosely speaking) continuous and continuously differentiable. The former assumption is true in this case, since the sample mean and the sample variance are asymptotically normal. The latter assumption clearly is violated if the variance of returns is zero. This will never actually occur in practice using real data samples, but if the variance *approaches* zero, making the Sharpe ratio highly nonlinear, delta method estimates will become unstable, as correctly noted by Vinod & Morey (2000). However, this scenario, too, arguably will affect few, if any cases in practice, as the variances of the returns of most, if not all funds or stocks that would be of enough interest to be subjected to Sharpe ratio comparisons are quite far from zero; if they were not, there would be nothing to compare! Still, it is important to note the limitations of analytical methods relied upon in any study, in case their domain of application changes. Jobson & Korkie (1981), Lo (2002), Memmel (2003), and Mertens (2002) all use the delta method in their studies of Sharpe ratios, thus supporting its practical use here.

moment of the joint distribution of *a* and *b*. The same result can be obtained using Stuart & Ord's (1994) (pp.457-458) result of $Cov\left(\hat{\sigma}_a^2, \hat{\sigma}_b^2\right) = \kappa_{2a,2b}/n + 2\kappa_{1a,1b}^2 / (n-1)$, where $\kappa_{2a,2b}$ is the second joint cumulant of the joint distribution of *a* and *b*, and $\kappa_{1a,1b}$ is the first joint cumulant, equal to the first joint central moment, $\mu_{1a,1b}$, which is the covariance. Dropping the *n* coefficients due to the use of the estimators $\hat{\sigma}_a^2, \hat{\sigma}_b^2$ for σ_a^2, σ_b^2 yields

$Cov\left(\sigma_a^2, \sigma_b^2\right) = \kappa_{2a,2b} + 2\kappa_{1a,1b}^2 = \kappa_{2a,2b} + 2\mu_{1a,1b}^2 = \kappa_{2a,2b} + 2\sigma_{a,b}^2$. Recognizing that the joint cumulant also can be expressed in terms of central moments, $\kappa_{2a,2b} = \mu_{2a,2b} - \mu_{2a,0} \times \mu_{0,2b} - 2\mu_{1a,1b}^2 = \mu_{2a,2b} - \sigma_a^2\sigma_b^2 - 2\sigma_{a,b}^2$ (see Stuart & Ord, 1994, p.107, and Smith, 1995), we have:

$Cov\left(\sigma_a^2, \sigma_b^2\right) = \kappa_{2a,2b} + 2\kappa_{1a,1b}^2 = \mu_{2a,2b} - \sigma_a^2\sigma_b^2 - 2\sigma_{a,b}^2 + 2\sigma_{a,b}^2 = \mu_{2a,2b} - \sigma_a^2\sigma_b^2$. Thus,

$$Var_{diff} = 2 - SR_a \frac{\mu_{3a}}{\sigma_a^3} - SR_b \frac{\mu_{3b}}{\sigma_b^3} + SR_a \frac{\mu_{1b,2a}}{\sigma_b \sigma_a^2} + SR_b \frac{\mu_{1a,2b}}{\sigma_a \sigma_b^2} + \frac{SR_a^2}{4}\left[\frac{\mu_{4a}}{\sigma_a^4} - 1\right] + \frac{SR_b^2}{4}\left[\frac{\mu_{4b}}{\sigma_b^4} - 1\right] - 2\rho_{a,b} - SR_a SR_b \frac{1}{2}\left[\frac{\mu_{2a,2b} - \sigma_a^2\sigma_b^2}{\sigma_a^2\sigma_b^2}\right]$$

$$= 2 - SR_a \frac{\mu_{3a}}{\sigma_a^3} - SR_b \frac{\mu_{3b}}{\sigma_b^3} + SR_a \frac{\mu_{1b,2a}}{\sigma_b \sigma_a^2} + SR_b \frac{\mu_{1a,2b}}{\sigma_a \sigma_b^2} + \frac{SR_a^2}{4}\left[\frac{\mu_{4a}}{\sigma_a^4} - 1\right] + \frac{SR_b^2}{4}\left[\frac{\mu_{4b}}{\sigma_b^4} - 1\right] - 2\left[\rho_{a,b} + \frac{SR_a SR_b}{4}\left[\frac{\mu_{2a,2b}}{\sigma_a^2\sigma_b^2} - 1\right]\right]$$

So analogous to the variance of the distribution of a single $\widehat{SR}$, (6), the variance of the difference between two $\widehat{SRs}$ is

$$\begin{aligned} Var_{diff} = {} & 1 + \frac{SR_a^2}{4}\left[\frac{\mu_{4a}}{\sigma_a^4} - 1\right] - SR_a \frac{\mu_{3a}}{\sigma_a^3} + \\ & 1 + \frac{SR_b^2}{4}\left[\frac{\mu_{4b}}{\sigma_b^4} - 1\right] - SR_b \frac{\mu_{3b}}{\sigma_b^3} \\ & -2\left[\rho_{a,b} + \frac{SR_a SR_b}{4}\left[\frac{\mu_{2a,2b}}{\sigma_a^2\sigma_b^2} - 1\right] - \frac{1}{2} SR_a \frac{\mu_{1b,2a}}{\sigma_b \sigma_a^2} - \frac{1}{2} SR_b \frac{\mu_{1a,2b}}{\sigma_a \sigma_b^2}\right] \end{aligned}$$

Note that when $\rho_{a,b} = 0$, $\mu_{2a,2b} = \sigma_a^2\sigma_b^2$, $\mu_{1a,2b} = 0$, and $\mu_{1b,2a} = 0$, so the entire covariance term of Var_{diff} disappears, as it should.

Minimum variance unbiased estimators of $\mu_{1a,2b}$, $\mu_{1b,2a}$, & $\mu_{2a,2b}$ are the respective *h*-statistics $h_{1a,2b}$, $h_{1b,2a}$, & $h_{2a,2b}$, where $h_{1,2} = \left[2s_{0,1}^2 s_{1,0} - ns_{0,2}s_{1,0} - 2s_{0,1}s_{1,1} + n^2 s_{1,2}\right] / \left[n(n-1)(n-2)\right]$, and $h_{2,2} =$

$$= \left[-3s_{0,1}^2 s_{1,0}^2 + ns_{0,2}s_{1,0}^2 + 4ns_{0,1}s_{1,0}s_{1,1} - 2(2n-3)s_{1,1}^2 - 2\left(n^2 - 2n + 3\right)s_{1,0}s_{1,2} + s_{0,1}^2 s_{2,0} - (2n-3)s_{0,2}s_{2,0} - 2\left(n^2 - 2n + 3\right)s_{0,1}s_{2,1} + n\left(n^2 - 2n + 3\right)s_{2,2}\right] /$$

$/ \left[(n-3)(n-2)(n-1)n\right]$, where $s_{x,y}$ are the simple power sums of $s_{x,y} = \sum_{i=1}^{n} a_i^x b_i^y$ (see Rose & Smith, 2002, pp.259-260).

This derivation is valid under iid returns, but because the one-sample estimator (6), derived using the same (delta) method (a la Mertens, 2002), was shown in Appendix B to be valid under the more general conditions afforded by its (identical) GMM derivation (a la Christie, 2005), we suspect those more general conditions of stationarity and ergodicity are the only requirements for the two-sample estimator of (13) as well. Proving this is the topic of continuing research.

APPENDIX D – Equivalence of Var_{diff} with Memmel (2003) and Jobson & Korkie (1981)

Under iid normality, Memmel's (2003) correction of Jobson and Korkie's (1981) variance of the two-sample statistic for the difference between two Sharpe ratios is:

$$Var = TV = 2 - 2\rho_{a,b} + \frac{1}{2}\left(SR_a^2 + SR_b^2 - 2SR_a SR_b \rho_{a,b}^2\right)$$

Under iid normality, TV is identical to Var_{diff}, as shown below:

$$Var_{diff} = 1 + \frac{SR_a^2}{4}\left[\frac{\mu_{4a}}{\sigma_a^4} - 1\right] - SR_a \frac{\mu_{3a}}{\sigma_a^3} + SR_a \frac{\mu_{1b,2a}}{\sigma_b \sigma_a^2} + 1 + \frac{SR_b^2}{4}\left[\frac{\mu_{4b}}{\sigma_b^4} - 1\right] - SR_b \frac{\mu_{3b}}{\sigma_b^3} + SR_b \frac{\mu_{1a,2b}}{\sigma_a \sigma_b^2} - 2\left[\rho_{a,b} + \frac{SR_a SR_b}{4}\left[\frac{\mu_{2a,2b}}{\sigma_a^2 \sigma_b^2} - 1\right]\right]$$

Under iid normality, $\mu_3/\sigma^3 = 0$, $\mu_{1,2} = 0$, $\mu_4/\sigma^4 = 3$, & $\mu_{2a,2b} = \left(1 + 2\rho_{a,b}^2\right)\sigma_a^2\sigma_b^2$ (see Stuart & Ord, 1994, p.105), so

$$Var_{diff} = 1 + \frac{SR_a^2}{4}[3-1] - 0 + 0 + 1 + \frac{SR_b^2}{4}[3-1] - 0 + 0 - 2\left[\rho_{a,b} + \frac{SR_a SR_b}{4}\left[\frac{\left(1+2\rho_{a,b}^2\right)\sigma_a^2\sigma_b^2 - \sigma_a^2\sigma_b^2}{\sigma_a^2\sigma_b^2}\right]\right] =$$

$$= 2 + \frac{SR_a^2}{2} + \frac{SR_b^2}{2} - 2\left[\rho_{a,b} + \frac{SR_a SR_b}{4}\left[2\rho_{a,b}^2\right]\right]$$

$$= 2 - 2\rho_{a,b} + \frac{1}{2}\left[SR_a^2 + SR_b^2 - 2SR_a SR_b \rho_{a,b}^2\right] = TV$$

, which is Memmel's (2003) result.

APPENDIX E – Simulation Distributions: APD of Komunjer (2006)

Komunjer (2006) gives the density of the asymmetric power distribution (APD) below:

$$f(u) = \begin{cases} \dfrac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \exp\left[-\dfrac{\delta_{\alpha,\lambda}}{\alpha^{\lambda}}|u|^{\lambda}\right] & \text{if } u \le 0, \\ \dfrac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \exp\left[-\dfrac{\delta_{\alpha,\lambda}}{(1-\alpha)^{\lambda}}|u|^{\lambda}\right] & \text{if } u > 0, \end{cases} \quad \text{where } 0 < \alpha < 1,\ \lambda > 0, \text{ and } \delta_{\alpha,\lambda} \equiv \frac{2\alpha^{\lambda}(1-\alpha)^{\lambda}}{\alpha^{\lambda} + (1-\alpha)^{\lambda}}$$

The α parameter ($0 < \alpha < 1$) controls skewness, with symmetry at $\alpha = 0.5$, and $\lambda > 0$ controls kurtosis, such that when $\alpha = 0.5$, $\lambda = \infty \rightarrow$ the uniform distribution, $\lambda = 1.0 \rightarrow$ the Laplace distribution (with variance = 2.0), $\lambda = 2.0 \rightarrow$ the normal distribution (with variance = 0.5) and any $\lambda \rightarrow$ the Generalized Power Distribution. When $\alpha \neq 0.5$, $\lambda = 1.0 \rightarrow$ the Asymmetric Laplace distribution of Kozubowski & Podgorski (1999), and $\lambda = 2.0 \rightarrow$ the two-piece normal distribution (see Johnson, Kotz & Balakrishnan, 1994, vol. 1 p.173 and vol. 2 p.190). Thus does APD allow simultaneous control over skewness and kurtosis, nesting the normal and Laplace densities, and asymmetric versions of each, as well as any "in between" combination of asymmetry and kurtosis.

Location and scale are accommodated via the simple transformation: $X \equiv \theta + \phi U$

APD moments are given by:

$$E\left(U^{r}\right) = \frac{\Gamma\left((1+r)/\lambda\right)}{\Gamma(1/\lambda)} \frac{(1-\alpha)^{1+r} + (-1)^{r}\alpha^{1+r}}{\delta_{\alpha,\lambda}^{r/\lambda}}$$

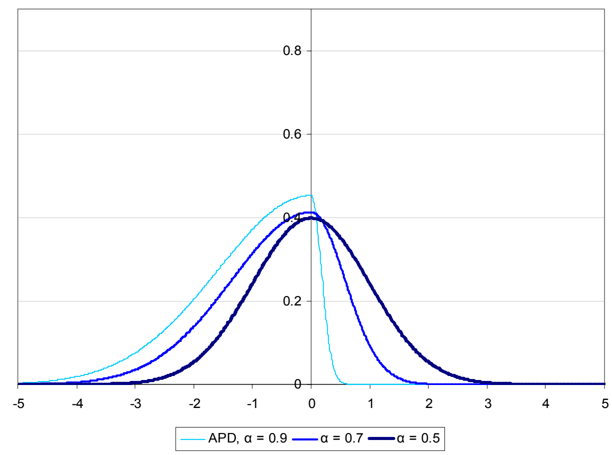
(see Table F1 below). So for example,

$$E(U) = \frac{\Gamma(2/\lambda)}{\Gamma(1/\lambda)}(1-2\alpha)\delta_{\alpha,\lambda}^{-1/\lambda} \quad \text{and} \quad Var(U) = \frac{\Gamma(3/\lambda)\Gamma(1/\lambda)\left[1-3\alpha+3\alpha^{2}\right] - \left[\Gamma(2/\lambda)\right]^{2}\left[1-2\alpha\right]^{2}}{\left[\Gamma(1/\lambda)\right]^{2}}\delta_{\alpha,\lambda}^{-2/\lambda}$$

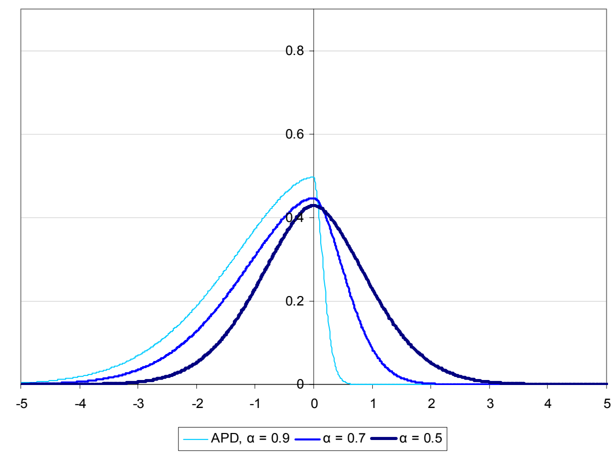
To standardize the APD for the simulations presented in this study, U is modified by $u' = u$ / sqrt[$Var(u)$] (because, for example, when $\alpha = 0.5$ and $\lambda = 1.0$, $Var(U) = 2.0$, and when $\alpha = 0.5$ and $\lambda = 2.0$, $Var(U) = 0.5$).

Table E1: Skewness η_3 and Kurtosis η_4 of APD by Values of α and λ

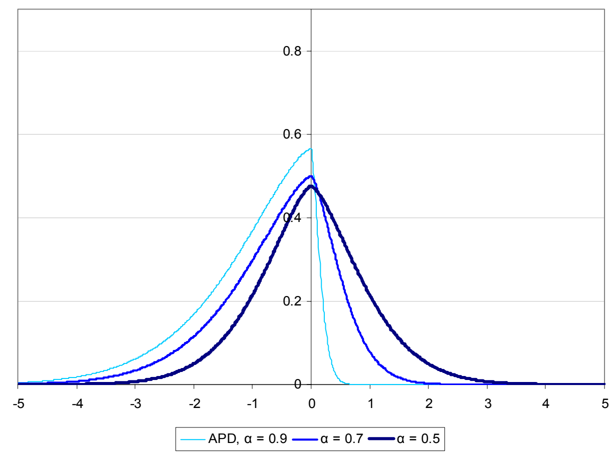
Special-case Nested Distribution	α	λ	Skewness η_3	Kurtosis η_4
Asymmetric Laplace	0.1 / 0.9	1.00	± 2.2311	6.6485
	0.1 / 0.9	1.25	± 1.9870	5.0165
	0.1 / 0.9	1.50	± 1.8415	4.1686
	0.1 / 0.9	1.75	± 1.7457	3.6595
Two-piece normal	0.1 / 0.9	2.00	± 1.6784	3.3243
Asymmetric Laplace	0.3 / 0.7	1.00	± 2.1867	7.4726
	0.3 / 0.7	1.25	± 1.9474	5.6383
	0.3 / 0.7	1.50	± 1.8048	4.6853
	0.3 / 0.7	1.75	± 1.7109	4.1131
Two-piece normal	0.3 / 0.7	2.00	± 1.6450	3.7363
Laplace (variance = 2.0)	0.5	1.00	0.0000	6.0000
GPD	0.5	1.25	0.0000	4.5272
GPD	0.5	1.50	0.0000	3.7620
GPD	0.5	1.75	0.0000	3.3026
Normal (variance = 0.5)	0.5	2.00	0.0000	3.0000



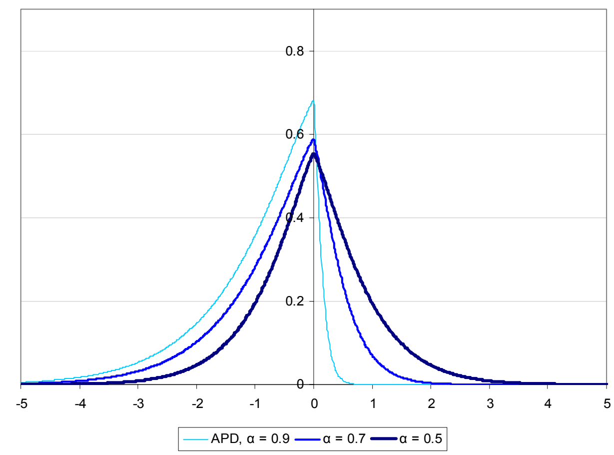
λ = 2.00 (Normal at α = 0.5)



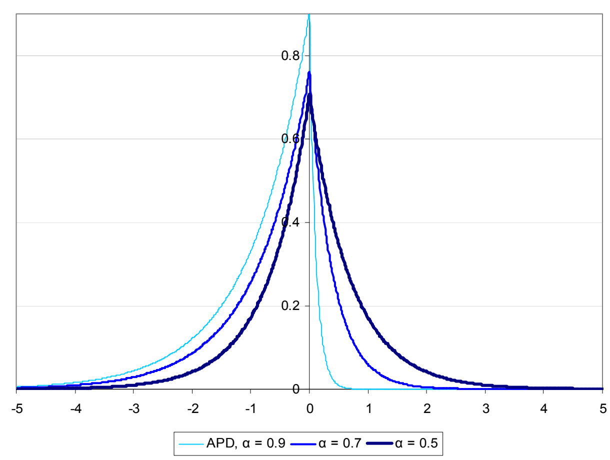
λ = 1.75



λ = 1.50



λ = 1.25



λ = 1.00 (Laplace at α = 0.5)

Figure 3E: Asymmetric Power Distribution by α by λ (all densities standardized so that Variance = 1.0)

REFERENCES

Akeda, Y. (2003), "Another Interpretation of Negative Sharpe Ratio," *The Journal of Performance Measurement*, 7(3), 19-23.

Ayebo, A. & Kozubowski, T. (2003), "An Asymmetric Generalization of Gaussian and Laplace Laws," *Journal of Probability and Statistical Science*, 1(2), 187-210.

Azzalini, A. (1985), "A Class of Distributions Which Includes the Normal Ones," *Scandinavian Journal of Statistics*, 12, 171-178.

Bao, Y., & Ullah, A. (2006), "Moments of the Estimated Sharpe Ratio when the Observations are not IID," *Finance Research Letters*, 3, 49-56.

Be, T.W. (2000), "Excess Kurtosis of Conditional Distribution for Daily Stock Returns: the Case of Japan," *Applied Economics Letters*, 7(6), 353-355.

Brännäs, K., & Nordman, N. (2003), "Conditional Skewness Modeling for Stock Returns," *Applied Economics Letters*, 10, 725-728.

Cappiello, L., Engle, R., & Sheppard, K. (2003), "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns," European Central Bank Working Paper Series, Working Paper No. 204.

Cajigas, J, & Urga, G. (2005), "Dynamic Conditional Correlation Models with Asymmetric Multivariate Laplace Innovations," presentation, Simulation Based and Finite Sample Inference in Finance II Conference in Quebec City, April 29-30.

Christie, S. (2005), "Is the Sharpe Ratio Useful in Asset Allocation?" *MAFC Research Papers No.31*, Applied Finance Centre, Macquarie University.

Dillen, H. & Stoltz, B. (1999), "The Distribution of Stock Market Returns and the Market Model," *Finnish Economic Papers*, 12(1) 41-56.

Eling, M. & Schuhmacher, F., "Does the Choice of Performance Measure Influence the Evaluation of Hedge Funds?," *Journal of Banking & Finance*, forthcoming.

Espejo, R., & Singh, H. (1999), "Covariance of Sample Moments for Simple Random Sampling with Replacement of a Finite Population," *Mathematical Proceedings of the Royal Irish Academy*, 99A (2), 171-177.

Federal Reserve Board, data series TCMNOMM3, from http://www.federalreserve.gov/releases/h15/data.htm

Gençay, R., Dacorogna, M., Muller, U., Pictet, O., & Olsen, R. (2001), *An Introduction to High-Frequency Finance*, Academic Press, San Diego.

Getmansky, M., Lo, A., & Makarov, I. (2004), "Model of Serial Correlation and Illiquidity in Hedge-Fund Returns," *Journal of Financial Economics*, 74(3), 529-609.

Gosh, M., Parr, W., Singh, K., & Babu, G. (1984), "A Note on Bootstrapping the Sample Median," *The Annals of Statistics*, Vol. 12, 1130-1135.

Greene, W. (1993), *Econometric Analysis*, 2nd Edition, Macmillan Publishing Company, New York.

Haas, M., Mittnik, S., & Paolella, M. (2005), "Modeling and Predicting Market Risk with Laplace-Gaussian Mixture Distributions," Center for Financial Studies, Ifo Institute for Economic Research, CFS Working Paper No. 2005/11.

Halmös, P. (1946), "The Theory of Unbiased Estimation," *Annals of Mathematical Statistics*, Vol. 17, 34-43.

Hansen, M.H., Hurwitz, W.N., & Madow, W.G. (1953), *Sample Survey Methods and Theory*, Walter Shewhart, Ed., John Wiley & Sons, Inc., New York.

Harris, R.D.F., & Coskun, K. (2001), "The Empirical Distribution of Stock Returns: Evidence from an Emerging European Market," *Applied Economics Letters*, 8, 367-371.

Hennard, Y. & Aparicio, M. (2003), "Multi-class Asset Allocation for Swiss Institutional Investors," Masters Thesis, Universite de Lausanne, Ecole des Hautes Etudes Commerciales (HEC), Master of Science in Finance.

Hsu, J. (1996), *Multiple Comparisons*, Chapman & Hall/CRC, New York.

Jobson, J. & Korkie, B. (1981) "Performance Hypothesis Testing with the Sharpe and Treynor Measures," *The Journal of Finance*, Vol. 36(4), 889-908.

Johnson, N., Kotz, S., & Balakrishnan, N. (1994), *Continuous Univariate Distributions*, New York: Wiley, 2nd ed.

Jondeau, E., Rockinger, M. (2003), "Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements," *Journal of Economic Dynamics and Control*, 27(10), 1699–1737.

Kelly, M. (2007), "Returns in Trading versus Non-Trading Hours: The Difference is Day and Night," Working Paper.

Kmenta, J. (1986), *Elements of Econometrics*, 2nd Edition, Macmillan Publishing Company, New York.

Knight, J. & Satchell, S. (2005), "A Re-Examination of Sharpe's Ratio for Log-Normal Prices," *Applied Mathematical Finance*, 12(1), 87-100.

Komunjer, I. (2006), "Asymmetric Power Distribution: Theory and Applications to Risk Measurement," *Journal of Applied Econometrics*, forthcoming.

Kotz, S., Kozubowski, T., & Podgorski, K. (2001), *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*, Birkhäuser, Boston.

Kozubowski, T., & Podgorski, K. (1999), "A Class of Asymmetric Distributions," *Actuarial Research Clearing House*, Vol. 1, 113-134.

Kozubowski, T., & Podgorski, K. (2001), "Asymmetric Laplace Laws and Modeling Financial Data," *Mathematical and Computer Modeling*, Vol. 34, 1003-1021.

Lee, Youngju (2003), *Society of Actuary Annual Meeting 2003*, Keynote: "Diversification with Fund of Hedge Funds."
Linden, M. (2001), "A Model for Stock Return Distribution," *International Journal of Finance & Economics*, Vol. 6(2), 159-169.
Lo, Andrew W. (2002), "The Statistics of Sharpe Ratios," *Financial Analysts Journal*, 58(4), 36-52.
Lo, Andrew W. (2003), "The Statistics of Sharpe Ratios: Author's Response," *Financial Analysts Journal*, 59(5), 17-17.
Malevergne, Y. & Sornette, D. (2005) "High-Order Moments and Cumulants of Multivariate Weibull Asset Returns Distributions: Analytical Theory and Empirical Tests: II," *Finance Letters*, 3(1), 54-63.
McLeod, W. & van Vurren, G. (2004), "Interpreting the Sharpe Ratio when Excess Returns are Negative," *Investment Analysts Journal*, 59, 15-20.
Memmel, C. (2003), "Performance Hypothesis Testing with the Sharpe Ratio," *Finance Letters*, 1, 21-23.
Mertens, E. (2002), "Comments on the correct variance of estimated Sharpe Ratios in Lo (2002, FAJ) when returns are IID." Research Note (www.elmarmertens.org).
Miller, R. and Gehr, A. (1978) "Sample Bias and Sharpe's Performance Measure: A Note," *Journal of Financial and Quantitative Analysis*, 13, 943–946.
Morillo, D., & Pohlman, L. (2002), "The Statistics of Sharpe Ratios: A Comment," *Financial Analysts Journal*, 58(6), 18-18.
Newey, W., & West, D. (1987), "A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.
Pastor, L., & Stambaugh, R. (2003) "Mutual Fund Performance and Seemingly Unrelated Assets," *Journal of Financial Economics*, Vol. 63(3), 315-349.
Patterson, K.D. & Heravi, S. (2003), "The Impact of Fat-tailed Distributions on Some Leading Unit Root Tests," *Journal of Applied Statistics*, 30, 635-667.
Pinto, J. & Curto, J. (2005), "The Generalized Method of Moments: An Application to the Sharpe Ratio," Temas em Métodos Quantitativos (Subjects in Quantitative Methods), 5, forthcoming.
Randles, R. & Wolfe, D. (1979), *Introduction to the Theory of Nonparametric Statistics*, John Wiley & Sons, New York.
Richardson, M. & Smith, T. (1993), "A Test for Multivariate Normality in Stock Returns," *Journal of Business*, 1993, 66(2), 295-321.
Rocke, D., & Downs, G. (1981), "Estimating the Variances of Robust Estimators of Location: Influence Curve, Jackknife and Bootstrap," *Communications in Statistics – Simulation and Computation*, Vol. 10(3), 221-248
Rose, C. & Smith, M. (2002), *Mathematical Statistics with Mathematica*, Springer-Verlag, New York.
Salibián-Barrera, M. (1998), "On Globally Robust Confidence Intervals for Regression Coefficients," The University of British Columbia Department of Statistics, Technical Report #176.
Scherer, B. (2004), "An Alternative Route to Performance Hypothesis Testing," *Journal of Asset Management*, 5, 5-12.
Scholz, H. (2007), "Refinements to the Sharpe ratio: Comparing Alternatives for Bear Markets," *Journal of Asset Management*, 5, 5-12.
Sharpe, W. (1966), "Mutual Fund Performance," *Journal of Business*, January, 119-138.
Sharpe, W. (1975), "Adjusting for Risk in Portfolio Performance Measurement," *Journal of Portfolio Management*, Winter, 29-34.
Sharpe, W. (1994), "The Sharpe Ratio," *Journal of Portfolio Management*, Fall, 49-58.
Sharpe, W. (1998), "Morningstar's Risk Adjusted Ratings," *Financial Analyst Journal*, July/August, 1998, 21-33.
Smith, P. (1995), "A Recursive Formulation of the Old Problem of Obtaining Moments From Cumulants and Vice Versa," *The American Statistician*, Vol.49 (2), 217-218.
Storey, J.D. (2002), "The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value," The Annals of Statisticis, Vol. 31(6), 2013-2035.
Storey, J.D. (2003), "A Direct Approach to False Discovery Rates," *Journal of the Royal Statistical Society, Series B*, 64, Part 3, 479-498.
Storey, J.D. (2004), "Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach," *Journal of the Royal Statistical Society, Series B*, 66, Part 1, 187-205.
Storey, J.D. (2007), "The Optimal Discovery Procedure: A New Approach to Simultaneous Significance Testing," *Journal of the Royal Statistical Society, Series B*, 69, Part 3, 347-368.
Stuart, A. & Ord, K. (eds.) (1994), *Kendall's Advanced Theory of Statistics, Vol.1 Distribution Theory*, 6th ed., Oxford University Press, New York.
Vinod, H. (2005), "Skew Densities and Ensemble Inference for Financial Economics," *The Mathematica Journal*, 9(4), 852-862.
Vinod, H., & Morey, M. (2000), "Confidence Intervals and Hypothesis Testing for the Sharpe and Treynor Performance Measures: A Bootstrap Approach," in *Computational Finance 1999*, eds. Yaser S. abu-Mostafa, Blake LeBaron, Andrew W. Lo, and Andreas S. Weigend, The MIT Press, Cambridge, Massachusetts.
White, H. (2001), *Asymptotic Theory for Econometricians*, Revised Edition, Academic Press, San Diego.
Wolf, M. (2003), "The Statistics of Sharpe Ratios: A Comment," *Financial Analysts Journal*, 59(5), 17-17.
Yahoo – Finance, http://finance.yahoo.com, data on returns from 40 mutual funds.
Zar, J. (1999), *Biostatistical Analysis*, 4th ed., Upper Saddle River, NJ: Prentice-Hall.