DataMineIt SM
©J.D. Opdyke

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

J.D. Opdyke, DataMineIt

JSM Activity #468: Confidence Intervals and Hypothesis Testing – Contributed - Papers

DataMineIt SM
©J.D. Opdyke

Contents

DataMineIt SM
©J.D. Opdyke

1. Why Consider $\widehat{SR}$?

- **Ubiquitous in Financial Analysis**
- **Funds and Fund Managers worldwide are continuously ranked according to their Sharpe Ratios. This ordering means little without statistical inference: how certain are we that one sample's $\widehat{SR}$ really is larger than that of another? Comparisons via ranking are implicit pairwise hypothesis tests:**

$$H_0: SR_b \leq SR_a \;\text{ v. }\; H_A: SR_b > SR_a$$

- **Important theoretical foundations: CAPM & MPT**

2. Two Common Definitions

i. Using StdDev of EXCESS returns

$$\widehat{SR}_e = \frac{\hat{\mu}_e}{\hat{\sigma}_e}$$

where $\hat{\mu}_e = \frac{\sum_{t=1}^{T} R_{et}}{T}$,

$R_{et} = \left(R_t - R_{ft} \right)$,

$$\hat{\sigma}_e = \sqrt{\frac{\sum_{t=1}^{T} \left(R_{et} - \hat{\mu}_e \right)^2}{T-1}}$$

$T = \#$ time periods,

$R_{ft} =$ period's risk free rate,

$R_t =$ period's return

See Jobson & Korkie (1981), Memmel (2003), Sharpe (1994)

©J.D. Opdyke

2. Two Common Definitions

ii. Using StdDev of returns

$$\widehat{SR}_s = \frac{\hat{\mu}_s - \hat{\mu}_f}{\hat{\sigma}_s}$$

$$\hat{\mu}_s = \frac{\sum_{t=1}^{T} R_t}{T},$$

$$\hat{\mu}_f = \frac{\sum_{t=1}^{T} R_{ft}}{T},$$

$$\hat{\sigma}_s = \sqrt{\frac{\sum_{t=1}^{T} \left(R_t - \hat{\mu}_s\right)^2}{T-1}}$$

$T = \#$ time periods,

$R_{ft} =$ period's risk free rate,

$R_t =$ period's return

See Christie (2005), Lo (2002)

2. Two Common Definitions

- **As an empirical matter, if the risk-free rate is not actually constant, it will be nearly so, making $\hat{\sigma}_s \gg \hat{\sigma}_{ft}$, so for all practical purposes, definition ii) is appropriate.**

DataMineIt SM
©J.D. Opdyke

3. Asymptotic Distribution of $\widehat{SR}$ Under iid Normality

- **Jobson & Korkie (1981) under iid normality:**

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

- **Lo (2002) presented same result, but misread by many as iid generally, not *NORMAL* iid (normality *implied* in a footnote). See Getmansky et al. (2004), Hennard & Aparicio (2003), Lee (2003), McLeod & van Vurren (2004), and Pinto & Curto (2005).**

DataMineIt[SM]
©J.D. Opdyke

3. Asymptotic Distribution of $\widehat{SR}$ Under iid Normality

- Lo (2002) uses variance of estimated variance *for a normal distribution* when using delta method to get variance of $\widehat{SR}$. $Var\left(\widehat{\sigma^2}\right) = 2\sigma^4$

- More generally, $Var\left(\widehat{\sigma^2}\right) = \mu_4 - \sigma^4$

DataMineIt SM
©J.D. Opdyke

4. Asymptotic Distribution of $\widehat{SR}$ Generally

- Using the more general result leads to

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

- Mertens (2002) presents this, but he does not generalize beyond iid returns, as is done by Christie (2005).

4. Asymptotic Distribution of $\widehat{SR}$ Generally

- Christie (2005) uses a GMM approach to obtain:

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

- This is quite unwieldy, but valid under very general conditions, requiring only stationary and ergodic returns. It thus allows for time-varying conditional volatilities, serial correlation, and even non-iid returns.

- However, this is identical to Mertens (2002)!

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

$$Var\left(\sqrt{T}\widehat{SR}\right)=E\left[\frac{SR^2 m_4}{4s^4}-\frac{SR\left[\left(R_t-R_{ft}\right)\left(R_t-m\right)^2-\left(R_t-R_{ft}\right)s^2\right]}{s^3}+\frac{\left(R_t-R_{ft}\right)^2}{s^2}-\frac{2\left(R_t-R_{ft}\right)SR}{s}+\frac{3SR^2}{4}\right] \qquad \text{(C21)}$$

which can be simplified as below[13]:

$$=\frac{SR^2}{4}\left[\frac{m_4}{s^4}+3\right]-SR\cdot E\left[\frac{\left(R_t-R_f\right)\left(R_t-m\right)^2}{s^3}\right]-SR\cdot E\left[\frac{\left(R_t-R_f\right)}{s}\right]+E\left[\frac{R_t^2-2R_tR_f+R_f^2}{s^2}\right]$$

since $E\left[R_t^2\right]=s^2+m^2$,

$$=\frac{SR^2}{4}\left[\frac{m_4}{s^4}+3\right]-SR\cdot E\left[\frac{\left(R_t-R_f\right)\left(R_t^2-2mR_t+m^2\right)}{s^3}\right]-SR^2+\frac{s^2+m^2-2mR_f+R_f^2}{s^2}$$

$$=\frac{SR^2}{4}\left[\frac{m_4}{s^4}+3\right]-SR\cdot E\left[\frac{R_t^3-2mR_t^2+m^2R_t-R_t^2R_f+2mR_tR_f-m^2R_f}{s^3}\right]-SR^2+1+SR^2$$

since $E\left[R_t^3\right]=m_3+3s^2m+m^3$,

$$=1+\frac{SR^2}{4}\left[\frac{m_4}{s^4}+3\right]-SR\cdot\left[\frac{m_3+3ms^2+m^3-2m\left(s^2+m^2\right)+m^3-\left(s^2+m^2\right)R_f+2m^2R_f-m^2R_f}{s^3}\right]$$

$$=1+\frac{SR^2}{4}\left[\frac{m_4}{s^4}+3\right]-SR\cdot\left[\frac{m_3+3ms^2+m^3-2ms^2-2m^3+m^3-s^2R_f-m^2R_f+m^2R_f}{s^3}\right]$$

$$=1+\frac{SR^2}{4}\left[\frac{m_4}{s^4}+3\right]-SR\cdot\left[\frac{m_3+3ms^2-2ms^2-s^2R_f}{s^3}\right]$$

$$=1+\frac{SR^2}{4}\left[\frac{m_4}{s^4}+3\right]-SR\cdot\left[\frac{m_3}{s^3}+\frac{m-R_f}{s}\right]$$

$$=1+\frac{SR^2}{4}\left[\frac{m_4}{s^4}+3\right]-SR\frac{m_3}{s^3}-SR^2$$

$$=1+\frac{SR^2}{4}\left[\frac{m_4}{s^4}-1\right]-SR\frac{m_3}{s^3}$$

So $Var\left(\sqrt{T}\widehat{SR}\right)=1+\frac{SR^2}{4}\left[\frac{m_4}{s^4}-1\right]-SR\frac{m_3}{s^3}$ which is Merten's (2002) result, and that derived in Appendix A.

[13] 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 $R_{ft}=\hat{m}_f=R_f$). 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.

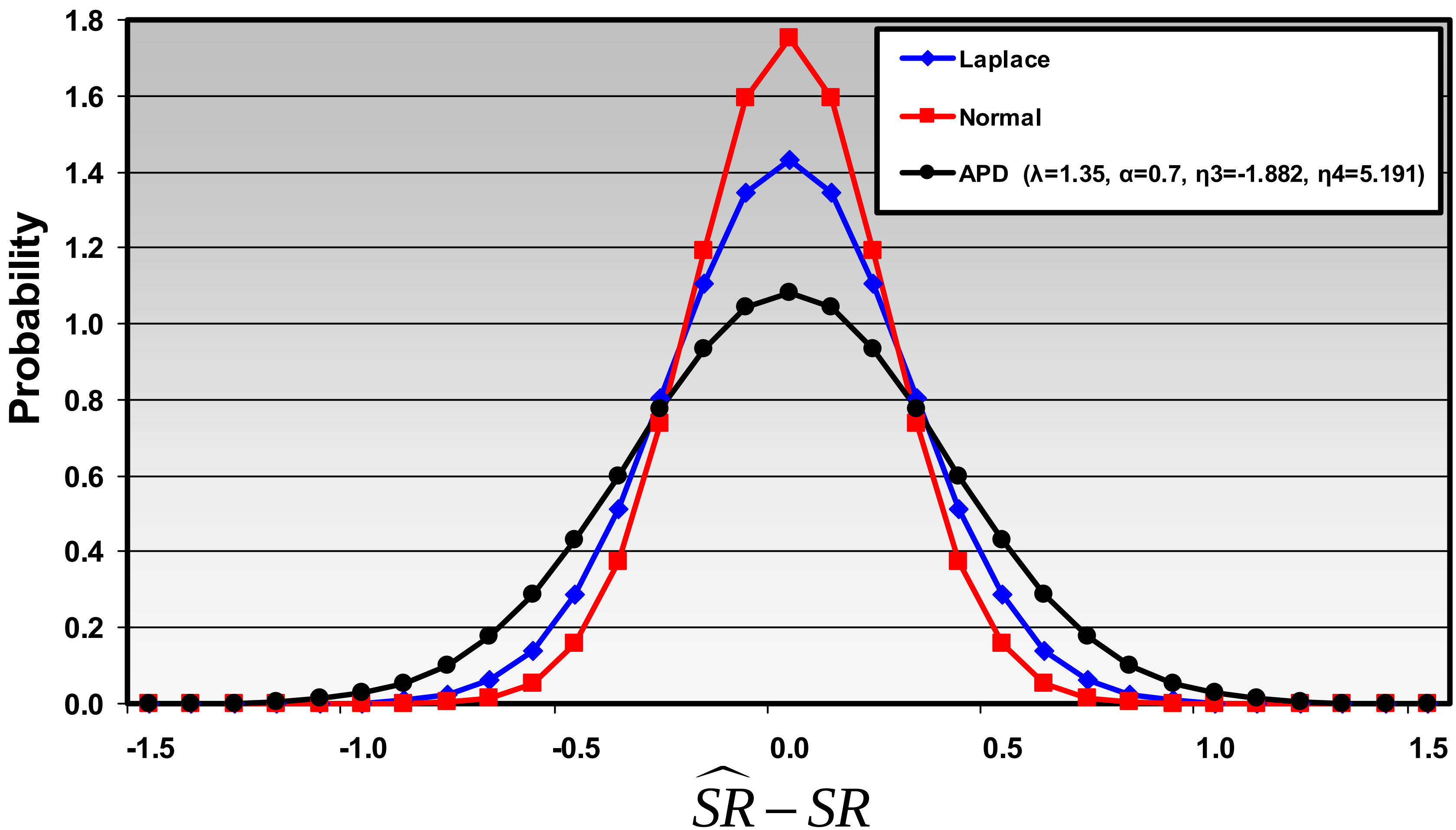
4. Asymptotic Distribution of $\widehat{SR}$ Generally

Consider

- $$\sqrt{T}\left(\widehat{SR}-SR\right)\overset{a}{\sim}N\left(0,\ 1+\frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}-1\right]-SR\frac{\mu_3}{\sigma^3}\right)$$
- **Note that only SR, skewness, and kurtosis of the returns determine the distribution of $\widehat{SR}$.**
- **So asymptotically, no moments beyond the fourth affect the distribution of $\widehat{SR}$. This is shown by distribution in Graph 1. Consistent with the empirical evidence, it shows that assuming normality could be a poor basis for inference.**

DataMineIt SM
©J.D. Opdyke

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



4. Asymptotic Distribution of $\widehat{SR}$ Generally

- **However, with this more mathematically tractible derivation we can see now that, in addition to stationarity and ergodicity, converging third and fourth moments are required for the convergence of the asymptotic distribution of $\widehat{SR}$.**
- **An apparently non-trivial number of financial instruments have diverging fourth moments, so this is important to note.**

5. Small Sample Bias of $\widehat{SR}$

- Because SR is convex, its estimator will be biased due to Jensen's inequality:

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

- Christie (2005) obtains a 2nd order Taylor series expansion of $\widehat{SR}$ about σ, and then a 1st order expansion of $\hat{\sigma}^2$ about σ^2 to obtain the distribution of $\hat{\sigma}$. However, like Lo (2002), he uses Var($\hat{\sigma}^2$) = $2\sigma^4$, **valid only under normality!**

DataMineIt SM
©J.D. Opdyke

5. Small Sample Bias of $\widehat{SR}$

- Christie (2005) obtains:

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

which not surprisingly, resembles the asymptotic distribution under normal returns.

- Using the more appropriate Var($\hat{\sigma}^2$) = $\mu_4-\sigma^4$,

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

which not surprisingly, resembles the asymptotic distribution generally.

DataMineIt SM
©J.D. Opdyke

5. Small Sample Bias of $\widehat{SR}$

- **So for small sample estimates of $\widehat{SR}$, divide $\widehat{SR}$ by**

$$\left(1+\frac{1}{4}\frac{\left[\hat{\mu}_4/\hat{\sigma}^4\right]}{T}\right)$$

to obtain an approximately unbiased estimate.

- **Simulations show this works very well**

DataMineIt[SM]
©J.D. Opdyke

6. Asymptotic Distribution of $\widehat{SR}_{diff}$

- $\widehat{SR}_a$ & $\widehat{SR}_b$ **are asymptotically unbiased normally distributed variables, so by CLT, their linear combination is asymptotically unbiased & normal.**

- **Statistic =** $\left(\widehat{SR}_b - \widehat{SR}_a\right) - \left(SR_b - SR_a\right)$, **Ho:** $SR_b = SR_a$,

$$Var\left[\left(\widehat{SR}_b - \widehat{SR}_a\right) - \left(SR_b - SR_a\right)\right] = Var\left(\widehat{SR}_b - \widehat{SR}_a\right) =$$

$$Var\left(\widehat{SR}_{diff}\right) = Var\left(\widehat{SR}_b\right) + Var\left(\widehat{SR}_a\right) - 2Cov\left(\widehat{SR}_b, \widehat{SR}_a\right)$$

6. Asymptotic Distribution of $\widehat{SR}_{diff}$

- Jobson & Korkie (1981) solved for normal iid returns; improved upon by Memmel (2003).
- But we now know returns are *not* normal.
- Vinod & Morey (2000) used the bootstrap and the double bootstrap, but this is computationally intensive, and bootstrap-based variance estimates are notoriously poor under asymmetric heavy tails, and even symmetric heavy tails (see Rocke & Downs, 1981, Gosh et al., 1984, and Salibian-Barrera, 1998), so caution is warranted with this approach.

DataMineIt SM
©J.D. Opdyke

6. Asymptotic Distribution of $\widehat{SR}_{diff}$

- **Previous work on distribution of a single $\widehat{SR}$ used the delta method (Jobson & Korkie, 1981; Lo, 2002; and Memmel, 2003), so why not use it to derive the two-sample statistic?**

- **If a function is continuous and continuously differentiable (loosely speaking), then the delta method obtains its variance if the random variables it uses are asymptotically normal.**

6. Asymptotic Distribution of $\widehat{SR}_{diff}$

So by delta, $Var\left(\widehat{SR}_{diff}\right)=\left(\frac{\partial f}{\partial u}\right)\Omega\left(\frac{\partial f}{\partial u}\right)'$**, where**

$SR=\frac{\mu}{\sigma}=f\left(\mu,\sigma\right)$, $\quad u=\left(\mu_a,\mu_b,\sigma_a^2,\sigma_b^2\right)$, and

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

Variance/covariance matrix of *u*

DataMineIt SM
©J.D. Opdyke

6. Asymptotic Distribution of $\widehat{SR}_{diff}$

delta method yields:

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

$$1 + \frac{SR_b^2}{4}\left[\frac{\mu_{4b}}{\sigma_b^4} - 1\right] - SR_b \frac{\mu_{3b}}{\sigma_b^3}$$

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

where $\mu_{2a,2b} = E\left[\left(a - E(a)\right)^2 \left(b - E(b)\right)^2\right]$ **= 2nd central moment of joint distribution**

and $\mu_{1a,2b} = E\left[\left(a - E(a)\right)\left(b - E(b)\right)^2\right]$

DataMineIt SM
©J.D. Opdyke

6. Asymptotic Distribution of $\widehat{SR}_{diff}$

- **When returns are normal iid,** $\mu_4/\sigma^4 = 3,\ \mu_3/\sigma^3 = 0,\ \mu_{1,2} = 0$ **and** $\mu_{2a,2b} = \left(1+2\rho_{a,b}^2\right)\sigma_a^2\sigma_b^2$, so Var_{diff} **reduces to Jobson & Korkie (1981), as shown below.**

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

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"[14] (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 $s_f^2 = 0$, which is always essentially, if not literally true, $SR = \dfrac{m - R_f}{s} = f\left(m, s^2\right)$, so let $u = \left(m_a, m_b, s_a^2, s_b^2\right)$ and $\hat{u} = \left(\hat{m}_a, \hat{m}_b, \hat{s}_a^2, \hat{s}_b^2\right)$, then $\sqrt{T}\left(\hat{u} - u\right) \sim N\left(0, \Omega\right)$ where Ω is the variance-covariance matrix of u:

$$\Omega = \begin{pmatrix} s_a^2 & s_{a,b} & m_{3a} & m_{1a,2b} \\ s_{a,b} & s_b^2 & m_{1b,2a} & m_{3b} \\ m_{3a} & m_{1b,2a} & \left(m_{4a} - s_a^4\right) & Cov\left(s_a^2, s_b^2\right) \\ m_{1a,2b} & m_{3b} & Cov\left(s_a^2, s_b^2\right) & \left(m_{4b} - s_b^4\right) \end{pmatrix}$$
where $s_{a,b} = Cov(a,b)$, $m_{3a} = E\left[\left(a - m_a\right)^3\right] = Cov\left(m_a, s_a^2\right)$,

$m_{3b} = E\left[\left(b - m_b\right)^3\right] = Cov\left(m_b, s_b^2\right)$ (see Mertens, 2002), $m_{1a,2b} = E\left[\left(a - m_a\right)\left(b - m_b\right)^2\right] = Cov\left(m_a, s_b^2\right)$,

and $m_{1b,2a} = E\left[\left(b - m_b\right)\left(a - m_a\right)^2\right] = Cov\left(m_b, s_a^2\right)$ (see Espejo & Singh, 1999). Now,

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

$$Var_{diff} = 1 - \frac{s_{a,b}}{s_a s_b} - \frac{m_{3a}\left(m_a - R_f\right)}{2s_a^4} + \frac{m_{1a,2b}\left(m_b - R_f\right)}{2s_a s_b^3} - \frac{s_{a,b}}{s_a s_b} + 1 + \frac{m_{1b,2a}\left(m_a - R_f\right)}{2s_a^3 s_b} - \frac{m_{3b}\left(m_b - R_f\right)}{2s_b^4} - \frac{m_{3a}\left(m_a - R_f\right)}{2s_a^4} + \frac{m_{1b,2a}\left(m_a - R_f\right)}{2s_a^3 s_b}$$

$$+ \frac{\left(m_{4a} - s_a^4\right)\left(m_a - R_f\right)^2}{4s_a^6} - \frac{\left(m_a - R_f\right)\left(m_b - R_f\right)Cov\left(s_a^2, s_b^2\right)}{4s_a^3 s_b^3} + \frac{m_{1a,2b}\left(m_b - R_f\right)}{2s_a s_b^3} - \frac{m_{3b}\left(m_b - R_f\right)}{2s_b^4} - \frac{\left(m_a - R_f\right)\left(m_b - R_f\right)Cov\left(s_a^2, s_b^2\right)}{4s_a^3 s_b^3} + \frac{\left(m_{4b} - s_b^4\right)\left(m_b - R_f\right)^2}{4s_b^6}$$

$$= 2 - 2r_{a,b} - \frac{m_{3a}\left(m_a - R_f\right)}{s_a^4} - \frac{m_{3b}\left(m_b - R_f\right)}{s_b^4} + \frac{m_{1b,2a}\left(m_a - R_f\right)}{s_a^3 s_b} + \frac{m_{1a,2b}\left(m_b - R_f\right)}{s_a s_b^3} + \frac{\left(m_{4a} - s_a^4\right)\left(m_a - R_f\right)^2}{4s_a^6} + \frac{\left(m_{4b} - s_b^4\right)\left(m_b - R_f\right)^2}{4s_b^6} - \frac{\left(m_a - R_f\right)\left(m_b - R_f\right)Cov\left(s_a^2, s_b^2\right)}{2s_a^3 s_b^3}$$

$$= 2 - 2r_{a,b} - SR_a\frac{m_{3a}}{s_a^3} - SR_b\frac{m_{3b}}{s_b^3} + SR_a\frac{m_{1b,2a}}{s_b s_a^2} + SR_b\frac{m_{1a,2b}}{s_a s_b^2} + \frac{SR_a^2}{4}\left[\frac{\left(m_{4a} - s_a^4\right)}{s_a^4}\right] + \frac{SR_b^2}{4}\left[\frac{\left(m_{4b} - s_b^4\right)}{s_b^4}\right] - \frac{SR_a SR_b}{2}\frac{Cov\left(s_a^2, s_b^2\right)}{s_a^2 s_b^2}$$

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

---

[14] 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.

then $Cov\left(s_a^2, s_b^2\right) = E\left[\left(a - E[a]\right)^2\left(b - E[b]\right)^2\right] - s_a^2 s_b^2 = m_{2a,2b} - s_a^2 s_b^2$, where $m_{2a,2b}$ is the joint second central 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 $Cov\left(\hat{s}_a^2, \hat{s}_b^2\right) = k_{2a,2b}/n + 2k_{1a,1b}^2/(n-1)$, where $k_{2a,2b}$ is the second joint cumulant of the joint distribution of a and b, and $k_{1a,1b}$ is the first joint cumulant, equal to the first joint central moment, $m_{1a,1b}$, which is the covariance. Dropping the n coefficients due to the use of the estimators $\hat{s}_a^2, \hat{s}_b^2$ for s_a^2, s_b^2 yields

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

$Cov\left(s_a^2, s_b^2\right) = k_{2a,2b} + 2k_{1a,1b}^2 = m_{2a,2b} - s_a^2 s_b^2 - 2s_{a,b}^2 + 2s_{a,b}^2 = m_{2a,2b} - s_a^2 s_b^2$. Thus,

$$Var_{diff} = 2 - SR_a\frac{m_{3a}}{s_a^3} - SR_b\frac{m_{3b}}{s_b^3} + SR_a\frac{m_{1b,2a}}{s_b s_a^2} + SR_b\frac{m_{1a,2b}}{s_a s_b^2} + \frac{SR_a^2}{4}\left[\frac{m_{4a}}{s_a^4} - 1\right] + \frac{SR_b^2}{4}\left[\frac{m_{4b}}{s_b^4} - 1\right] - 2r_{a,b} - SR_a SR_b\frac{1}{2}\left[\frac{m_{2a,2b} - s_a^2 s_b^2}{s_a^2 s_b^2}\right]$$

$$= 2 - SR_a\frac{m_{3a}}{s_a^3} - SR_b\frac{m_{3b}}{s_b^3} + SR_a\frac{m_{1b,2a}}{s_b s_a^2} + SR_b\frac{m_{1a,2b}}{s_a s_b^2} + \frac{SR_a^2}{4}\left[\frac{m_{4a}}{s_a^4} - 1\right] + \frac{SR_b^2}{4}\left[\frac{m_{4b}}{s_b^4} - 1\right] - 2\left[r_{a,b} + \frac{SR_a SR_b}{4}\left[\frac{m_{2a,2b}}{s_a^2 s_b^2} - 1\right]\right]$$

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

$$Var_{diff} = 1 + \frac{SR_a^2}{4}\left[\frac{m_{4a}}{s_a^4} - 1\right] - SR_a\frac{m_{3a}}{s_a^3} +$$

$$1 + \frac{SR_b^2}{4}\left[\frac{m_{4b}}{s_b^4} - 1\right] - SR_b\frac{m_{3b}}{s_b^3}$$

$$-2\left[r_{a,b} + \frac{SR_a SR_b}{4}\left[\frac{m_{2a,2b}}{s_a^2 s_b^2} - 1\right] - \frac{1}{2}SR_a\frac{m_{1b,2a}}{s_b s_a^2} - \frac{1}{2}SR_b\frac{m_{1a,2b}}{s_a s_b^2}\right]$$

Note that when $r_{a,b} = 0$, $m_{2a,2b} = s_a^2 s_b^2$, $m_{1a,2b} = 0$, and $m_{1b,2a} = 0$, so the entire covariance term of Var_{diff} disappears, as it should.

Minimum variance unbiased estimators of $m_{1a,2b}$, $m_{1b,2a}$, & $m_{2a,2b}$ 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.

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:

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

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

$$Var_{diff} = 1 + \frac{SR_a^2}{4}\left[\frac{m_{4a}}{s_a^4} - 1\right] - SR_a\frac{m_{3a}}{s_a^3} + SR_a\frac{m_{1b,2a}}{s_b s_a^2} + 1 + \frac{SR_b^2}{4}\left[\frac{m_{4b}}{s_b^4} - 1\right] - SR_b\frac{m_{3b}}{s_b^3} + SR_b\frac{m_{1a,2b}}{s_a s_b^2} - 2\left[r_{a,b} + \frac{SR_a SR_b}{4}\left[\frac{m_{2a,2b}}{s_a^2 s_b^2} - 1\right]\right]$$

Under iid normality, $m_3/s^3 = 0$, $m_{1,2} = 0$, $m_4/s^4 = 3$, & $m_{2a,2b} = \left(1 + 2r_{a,b}^2\right)s_a^2 s_b^2$ (see Stuart & Ord, 1994, p.105), so

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

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

$$= 2 - 2r_{a,b} + \frac{1}{2}\left[SR_a^2 + SR_b^2 - 2SR_a SR_b r_{a,b}^2\right] = TV$$, which is Memmel's (2003) result.

6. Asymptotic Distribution of $\widehat{SR}_{diff}$

- Limitations of delta: Estimates become unstable when function is highly nonlinear. This would occur for $\widehat{SR}$ if σ^2 approaches zero. Also, $\widehat{SR}$ obviously is not continuous or continuously differentiable at $\sigma^2 = 0$.

- However, this is not a problem for practical usage, since there would be nothing to compare if σ^2 was close to zero! The problem in practice is not too little variance, but rather, too much (see Christie, 2005).

DataMineIt SM
©J.D. Opdyke

7. Simulation Study

- **Asymptotics are fine, but how does Var_{diff} perform under realistic data conditions – returns that are:**

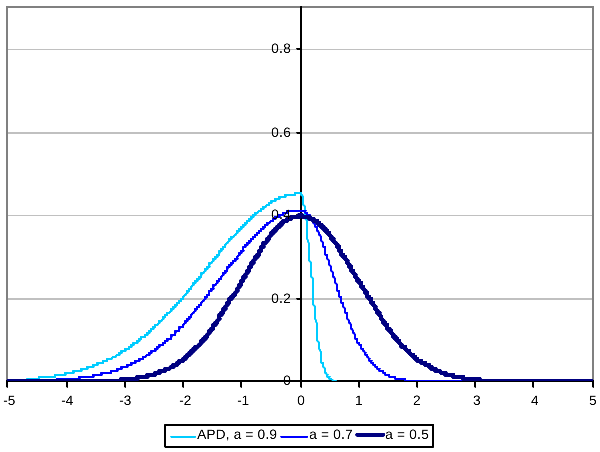
  i. **Leptokurtotic (i.e. "heavy tailed")**
  ii. **Asymmetric**
  iii. **Strongly (positively) correlated with each other**
  iv. **Based on finite sample sizes**

DataMineIt SM
©J.D. Opdyke

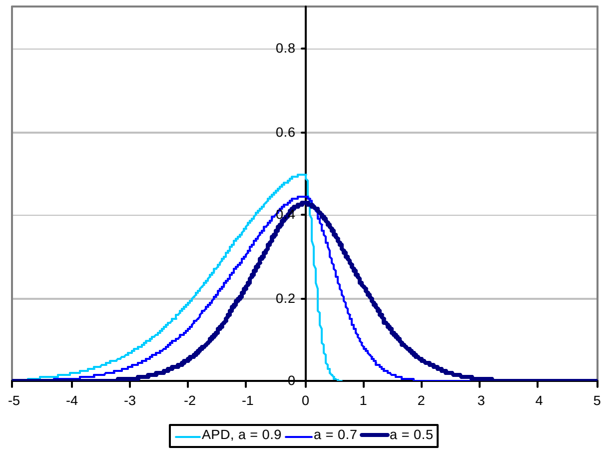
7. Simulation Study

- Use Komunjer's (2006) Asymmetric Power Distribution (APD) for simulation study testing empirical level and power.

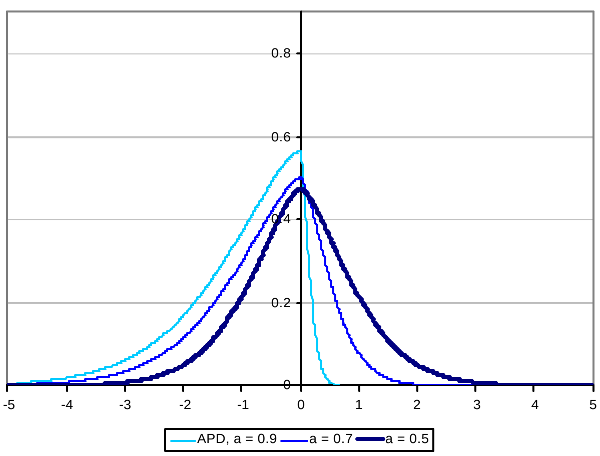
- APD has skewness (α) and kurtosis (λ) parameters and nests the normal, Laplace, asymmetric (2-piece) normal, and asymmetric Laplace. These all have been used extensively in the empirical finance literature.



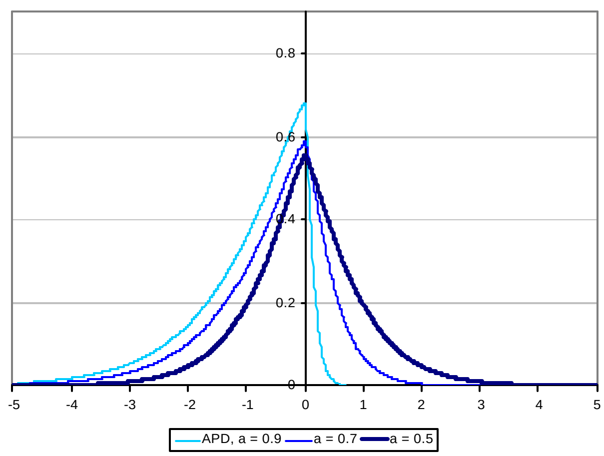
? = 2.00 (Normal at a = 0.5)



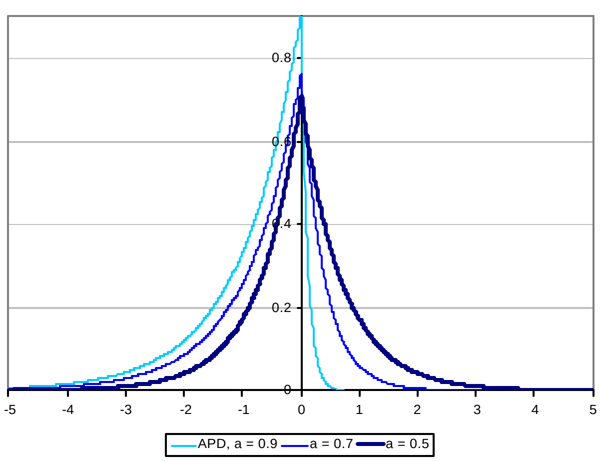
? = 1.75



? = 1.50



? = 1.25



? = 1.00 (Laplace at a = 0.5)

Figure 2: Asymmetric Power Distribution by a by ? (all densities standardized so that Variance = 1.0)

Simulation Distributions: APD of Komunjer (2006)

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

$$f(u)=\begin{cases}\dfrac{d_{a,l}^{1/l}}{\Gamma(1+1/l)}\exp\left[-\dfrac{d_{a,l}}{a^{l}}|u|^{l}\right] & \text{if } u\le 0,\\[2ex] \dfrac{d_{a,l}^{1/l}}{\Gamma(1+1/l)}\exp\left[-\dfrac{d_{a,l}}{(1-a)^{l}}|u|^{l}\right] & \text{if } u>0,\end{cases}\qquad \text{where } 0 < a < 1,\ ? > 0, \text{ and } d_{a,l}\equiv\frac{2a^{l}(1-a)^{l}}{a^{l}+(1-a)^{l}}$$

The a parameter controls skewness, with symmetry at a = 0.5, and ? controls kurtosis, such that when a = 0.5, ? = 8 ? the uniform distribution, ? = 1.0 ? the Laplace distribution (with variance = 2.0), and ? = 2.0 ? the normal distribution (with variance = 0.5). When a ? 0.5, ? = 1.0 ? the Asymmetric Laplace distribution of Kozubowski & Podgorski (1999), and ? = 2.0 ? 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: $X \equiv q + fU$

APD moments are given by:

$$E\left(U^{r}\right)=\frac{\Gamma\left((1+r)/l\right)}{\Gamma(1/l)}\frac{(1-a)^{1+r}+(-1)^{r}a^{1+r}}{d_{a,l}^{r/l}}$$ (see Table III below). So for example,

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

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

Table III: Skewness $?_3$ and Kurtosis $?_4$ of APD by Values of a and ?

Special-case Nested Distribution	a	?	Skewness $?_3$	Kurtosis $?_4$
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

7. Simulation Study

N = 10,000 simulations run for over 5,200 combinations of:

- **Sample size (*T* = 15, 30, 50, 100, 300)**
- μ / σ^2 **configurations; sizes of** $\widehat{SR}$
- **Skewness (range used = η_3 = ±2.23)**
- **Kurtosis (range used = η_4 = ±7.47)**
- **Correlation between the two series of returns ($\rho_{a,b} =$ 0.00, 0.25, 0.50, 0.75)**

7. Simulation Study - Results

Many results, but key results include:

- **Under <u>concurrent skewness and leptokurtosis</u> at least as extreme as typical returns (used α = 0.7 & λ = 1.35 for skewness & kurtosis of η_3 = -1.88 & η_4 = 5.19, respectively; see Haas et al., 2005; Cajigas & Urga, 2005; Cappiello et al., 2003; and Vinod, 2005)**
- **Under positive <u>correlation</u> at least as extreme as typical returns (IMPORTANT! NOT carefully examined in literature on a 2-sample $\widehat{SR}$ tests)**

JSM 2006

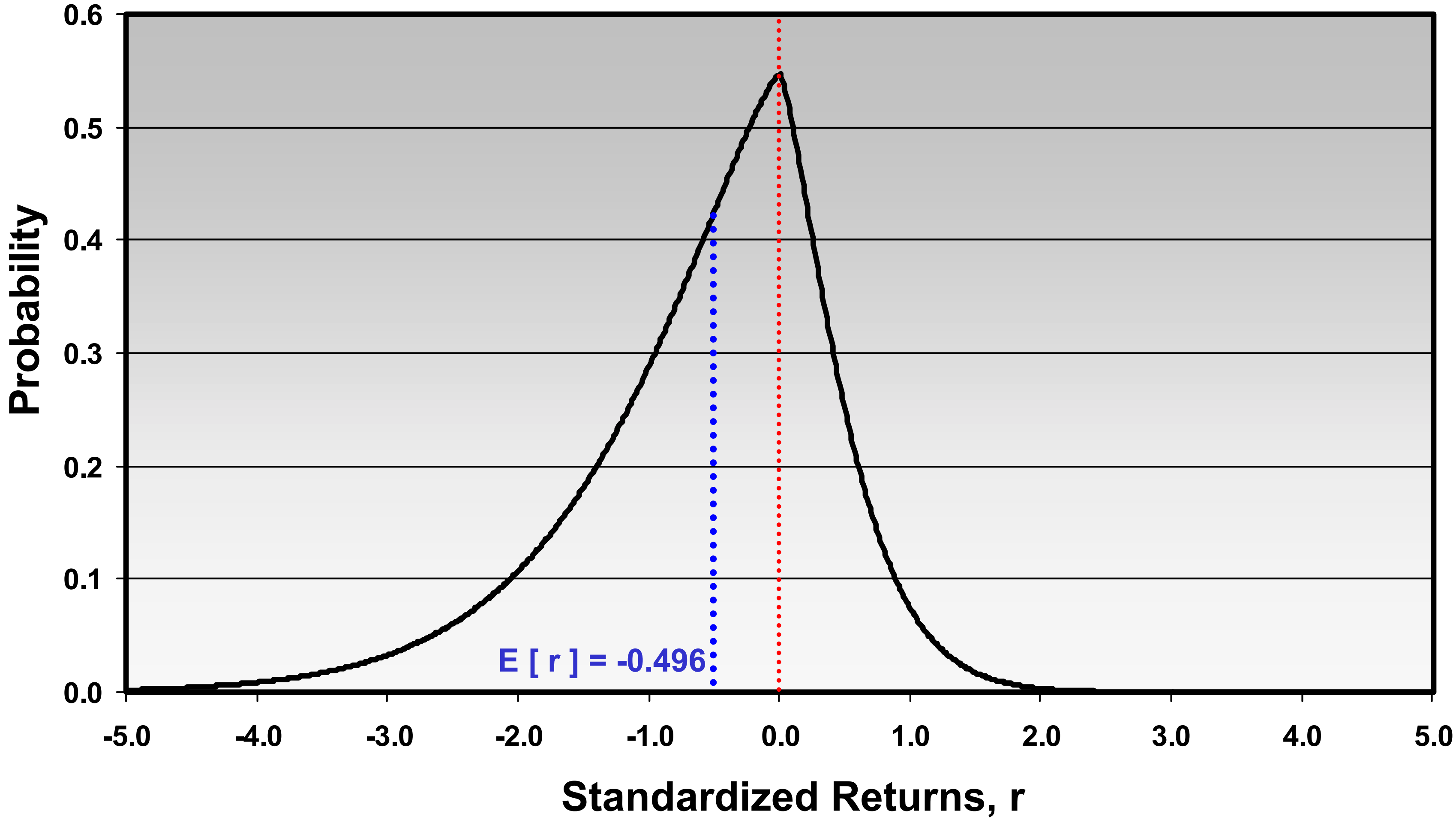
DataMineIt[SM]

©J.D. Opdyke

GRAPH 3

APD "Real World" Simulated Returns

(α = 0.7, λ = 1.35, η_3 = -1.88, η_4 = 5.19)



7. Simulation Study - Results

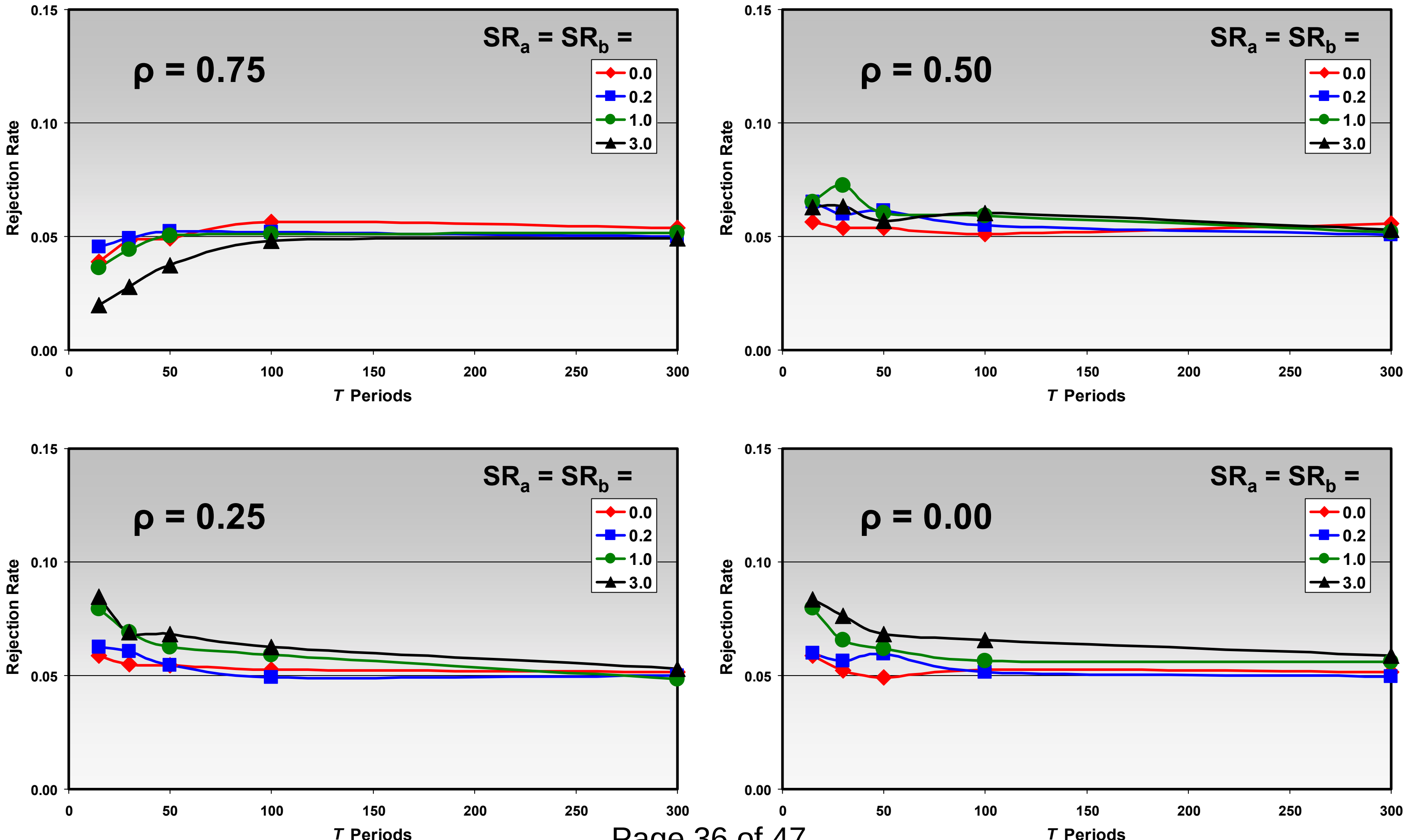
- **As shown in Graph 4, excellent convergence to nominal level α under “real world” skewness and leptokurtosis of APD simulated returns, even for large values of $SR_a = SR_b$.**

JSM 2006

DataMineIt[SM]
©J.D. Opdyke

GRAPH 4

Level of 2-sample Estimator Under "Real World" APD Returns by ρ by SR by T



7. Simulation Study - Results

- Under “realistic” skewed and leptokurtotic APD-simulated returns, note the dramatic increases in power under strong, positive correlation, which is typical for Sharpe ratio comparisons in practice.

Figure 5a: Power – SR_a=0.0, SR_b=0.1

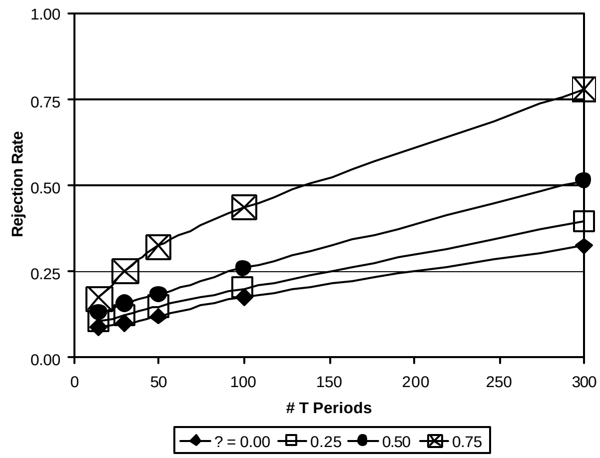


Figure 5b: Power – SR_a=0.0, SR_b=0.2

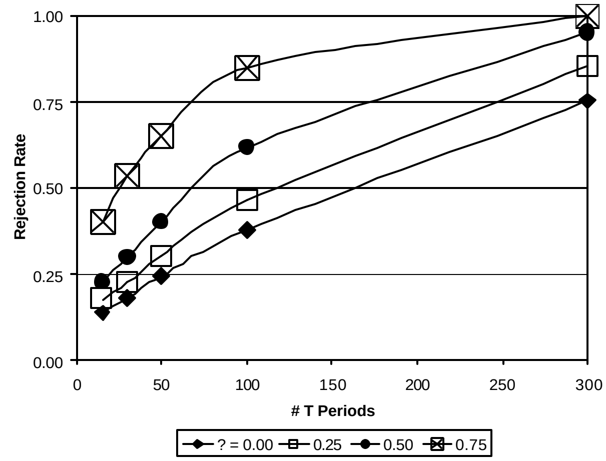


Figure 5c: Power – SR_a=0.0, SR_b=0.5

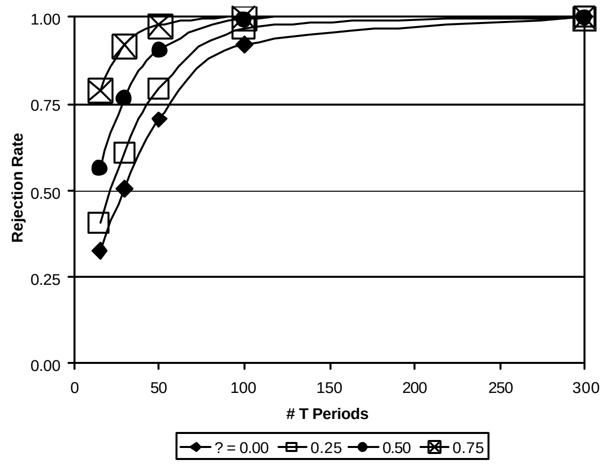


Figure 5d: Power – SR_a=0.2, SR_b=0.4

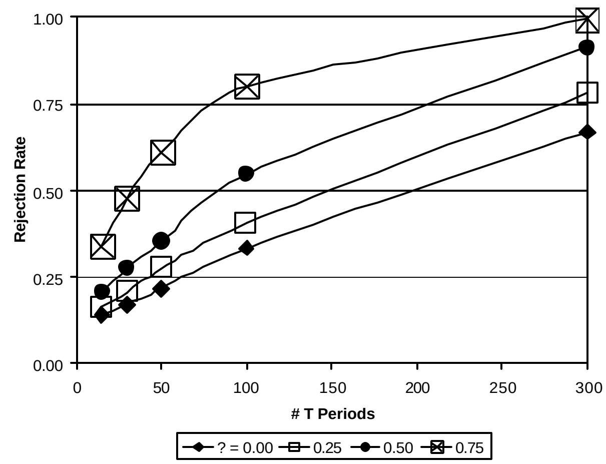


Figure 5e: Power – SR_a=1.0, SR_b=1.5

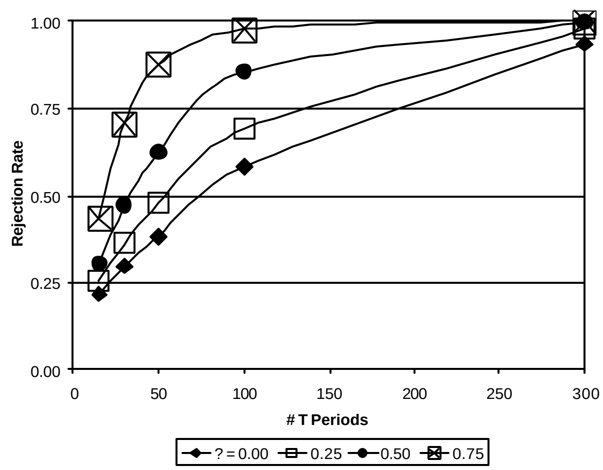
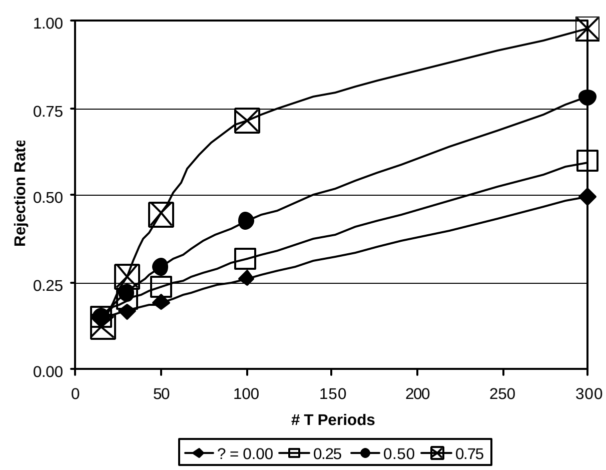


Figure 5f: Power – SR_a=3.0, SR_b=3.5



DataMineIt SM
©J.D. Opdyke

8. Using Actual Returns - Mutual Funds

- Christie (2005) finds that, when based on actual returns data, statistically significant differences between Sharpe ratios remain elusive due to large variances.

- However, Var_{diff} finds statistically significant differences between the Sharpe ratios of several randomly selected large growth mutual funds right off the bat!

8. Using Actual Returns - Mutual Funds

- For the period 09/01 through 08/06, take the monthly returns of:
  - Fidelity's Contrafund (FCNTX)
  - Janus Growth & Income (JAGIX)
  - Vanguard Growth Index (VIGRX)
  - 90-day Treasury Bill rate (divided by 12).

- Only Fidelity approaches statistically significant positive excess returns, as indicated by $SR > 0$ (one-sided p-values of 0.074, 0.320, and 0.477, respectively).

DataMineIt SM
©J.D. Opdyke

8. Using Actual Returns - Mutual Funds

- However, the strong, positive correlations between them give Var_{diff} greater precision, and thus, greater power to detect differences between their Sharpe Ratios.

- Ho: SR-Fidelity ≤ SR-Janus, p = 0.047
  Ho: SR-Fidelity ≤ SR- Vanguard, p = 0.030
  Ho: SR-Janus ≤ SR-Vanguard, p = 0.195

DataMineIt℠
©J.D. Opdyke

8. Using Actual Returns - Mutual Funds

Null Hypothesis, Ho:	$\widehat{SR}$, $\widehat{SR}_{diff}$	Pearson's Sample Correlation Coefficient r	Opdyke (2006) one-sided p-value
SR-Fidelity ≤ 0	0.20		0.074
SR-Janus ≤ 0	0.06		0.320
SR-Vanguard ≤ 0	0.01		0.477
SR-Fidelity ≤ SR-Janus	0.14	0.86	0.047
SR-Fidelity ≤ SR-Vanguard	0.19	0.77	0.030
SR-Janus ≤ SR-Vanguard	0.05	0.90	0.195

9. Conclusions

Major contributions of this study:

- It generalized the only useable version of the asymptotic distribution of $\widehat{SR}$ to very realistic conditions, requiring only stationary and ergodic returns with converging 3rd & 4th moments.

- It derived an easily used 2-sample statistic for ($\widehat{SR}_b - \widehat{SR}_a$) that nests the normal iid derivation of Jobson & Korkie (1981) and has excellent level control under real-world data conditions (i.e. asymmetric, leptokurtotic, and highly correlated returns based on finite samples).

DataMineIt SM
©J.D. Opdyke

9. Conclusions

- Power of the 2-sample statistic is generally modest, but it increases dramatically under strong, positively correlated returns: since most Sharpe Ratio comparisons are apples-to-apples, this is the rule rather than the exception!

- Therefore, as it would be used in practice, the statistic appears to have GOOD power, as demonstrated by a comparison of actual mutual fund returns.

DataMineIt SM
©J.D. Opdyke

Selected References:

- Cajigas & Urga, 2005
- Cappiello et al., 2003
- Christie, 2005
- Getmansky et al., 2003
- Gosh et al., 1984
- Haas et al., 2005
- Hennard & Aparicio, 2003
- Jobson & Korkie, 1981
- Johnson, Kotz, & Balakrishnan, 1994
- Komunjer, 2006
- Kozubowski & Podgorski, 1999
- Lee, 2003
- Lo, 2002
- McLeod & van Vurren, 2004
- Memmel, 2003
- Mertens, 2002
- Pinto & Curto, 2005
- Rocke & Downs, 1981
- Salibian-Barrerra, 1998
- Sharpe, 1966, 1975, 1994
- Vinod, 2005
- Vinod & Morey, 2000

Acknowledgments:

- I express thanks to Keith Ord, Andrew Clark, and Hrishikesh Vinod for useful comments, and sincere appreciation to Steve Christie for spotting an error in an earlier draft, and to Geri S. Costanza for numerous and valuable insightful discussions.

Contact:

- J.D. Opdyke, President

DataMineIt SM
When You Need to Know SM

JDOpdyke@DataMineIt.com
www.DataMineIt.com
203-249-4837 (ph) • 781-639-6463 (ph/fax)