Notes on Robust Hypothesis Testing

By Paulo F. Naibert

October 17, 2019

1 Generalized Testing

Based on Ledoit & Wolf (2008, 2011).

We observe T pairs of returns, $(r_{1i}, r_{1n})', \dots, (r_{Ti}, r_{Tn})'$, with a bivariate return distribution:

$$\mu = \begin{bmatrix} \mu_i \\ \mu_n \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_i^2 & \sigma_{in} \\ \sigma_{in} & \sigma_n^2 \end{bmatrix}.$$

NOTE: Any quantity x when denoted as $\hat{x}$ simply means its sample counterpart, e.g. $\hat{\mu}$ denotes the sample counterpart of μ.

Define a primitive statistic τ as a function of data, define a "derived" statistic δ as a function of the primitive statistics, and finally, define Δ as the difference of derived statistics, then we have:

$$\Delta = \delta_i - \delta_n = \delta(\tau(r_{ti})) - \delta(\tau(r_{tn}))$$

We can also define Δ as a function of the $K \times 1$ vector of primitive statistics v:

$$\Delta = f(v)$$

Then we can define the $\nabla f(v)$ as the gradient of $f(v)$:

$$\nabla' f(v) = \left(\frac{\partial f(u)}{\partial v_1}, \dots \frac{\partial f(u)}{\partial v_n}, \right)$$

Assume that:

$$T^{1/2}(v - \hat{v}) \to^d N(0, \Psi),$$

where Ψ is an unknown symmetric PSD matrix.

If we apply a function on the vector v of parameters, the Taylor expansion (Delta method) implies:

$$T^{1/2}[f(v) - f(\hat{v})] \to^d N\left(0; \nabla' f(v) \Psi \nabla f(v)\right).$$

Well, we will use $f(\cdot)$ as defined earlier, and we denote $f(v) = \Delta$, so we have:

$$T^{1/2}(\Delta - \hat{\Delta}) \to^d N\left(0; \nabla' f(v) \Psi \nabla f(v)\right).$$

Now, if a consistent estimator $\hat{\Psi}$ of Ψ is available, then the standard error of the estimate $\hat{\Delta}$, $s(\hat{\Delta})$, is given by:

$$s(\hat{\Delta}) = \sqrt{T^{-1} \nabla' f(\hat{v}) \hat{\Psi} \nabla f(\hat{v})}. \tag{1}$$

1.1 Estimating $\hat{\Psi}$

Let's estimate $\hat{\Psi}$:

$$\Psi = \lim_{T\to\infty} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} E[y_s y_t'], \quad \text{with} \quad y_t' = \left(r_{ti} - \mu_i, r_{tn} - \mu_n, r_{ti}^2 - \gamma_i, r_{tn}^2 - \gamma_n\right)$$

$$\Psi = \lim_{T\to\infty} \Psi_t, \quad \text{with} \quad \Psi = \sum_{j=-T+1}^{T-1} \Gamma_T(j), \qquad \text{where}$$

$$\Gamma_T(j) = \begin{cases} T^{-1} \sum_{t=j+1}^{T} E[y_t y_{t-j}'] & \text{for } j \geq 0 \\ T^{-1} \sum_{t=-j+1}^{T} E[y_{t+j} y_t'] & \text{for } j < 0 \end{cases}$$

Are those $E(\cdot)$ necessary, since we are already taking means?

$$\hat{\Psi} = \hat{\Psi}_T = \frac{T}{T-K} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T}\right) \hat{\Gamma}_T(j), \qquad \text{where}$$

$k(\cdot)$ kernel. S_T bandwidth.

$$\hat{\Gamma}_T(j) = \begin{cases} T^{-1} \sum_{t=j+1}^{T} \hat{y}_t \hat{y}_{t-j}' & \text{for } j \geq 0 \\ T^{-1} \sum_{t=-j+1}^{T} \hat{y}_{t+j} \hat{y}_t' & \text{for } j < 0 \end{cases}$$

> The factor $T/(T-K)$ is a small sample degrees of freedom adjustment that is introduced to offset the effect of the estimation of the $K \times 1$ vector v in the computation $\hat{\Gamma}_T(j)$, that is, the use of $\hat{y}_t$ rather than y_t.

A two-sided p-value for the Null hypothesis $H_0 : \Delta = 0$ is givenby:

$$\hat{p} = 2\Phi\left(\frac{|\hat{\Delta}|}{s(\hat{\Delta})}\right)$$

where Ψ denotes the cdf of the standard normal distribution.

1.2 Bootstrap Inference

Circular Boostrap of Politis and Romano (1992), resampling now *blocks of pairs* from the observed pairs $(r_{ti}, r_{tn})'$, $t = 1, \ldots, T$ *with replacement*. These block have a *fixed* size $b \geq 1$. Standard error is computed as:

$$se(\hat{\Delta}) = \sqrt{T^{-1} \nabla' f(\hat{v}) \hat{\Psi} \nabla f(\hat{v})}.$$

The estimator (estimate?) $\hat{\Psi}$ is obtained via kernel estimation.

Standard error $se(\hat{\Delta}^*)$ is the "natural" standard error computed from the bootstrap data. More specifically, letting $l = \lfloor T/b \rfloor$, define:

$$y_t^* = \left(r_{ti}^* - \hat{\mu}_i^*, r_{tn}^* - \hat{\mu}_n^*, r_{ti}^{2\,*} - \hat{\gamma}_i^*, r_{tn}^{2\,*} - \hat{\gamma}_n^*\right), \quad t = 1, \dots T \tag{2}$$

$$\zeta_j = b^{-1/2} \sum_{t=1}^{b} y_{t+(j-1)b}^*, \quad t = 1, \dots, l \tag{3}$$

and

$$\hat{\Psi}^* = l^{-1} \sum_{j=1}^{l} \zeta_j \zeta_j'. \tag{4}$$

With this more genral definition, of $\hat{\Psi}^*$, the bootstrap error is given by:

$$se(\hat{\Delta}^*) = \sqrt{T^{-1} \nabla' f(\hat{v}^*) \hat{\Psi}^* \nabla f(\hat{v}^*)}.$$

1.2.1 Direct Computation of the p-value

Remark 3.2 on the original paper

Denote the original studentized test statistic by:

$$d = \frac{|\hat{\Delta}|}{s(\hat{\Delta})} \tag{5}$$

and denote the *centered* studentized test statistic computed from the mth bootstrap sample by:

$$\tilde{d}^{*,m} = \frac{|\hat{\Delta}^{*,m} - \hat{\Delta}|}{s(\hat{\Delta}^{*,m})} \tag{6}$$

where M is the number of bootstrap resamples. Then the p-value is computed as:

$$PV = \frac{\#\{\tilde{d}^{*,m} \geq d\} + 1}{M + 1}. \tag{7}$$

2 Sharpe Ratio Example

Define the difference in Sharpe Ratios as function of the vector of primitive statistics $u = (\mu_i, \mu_n, \sigma_i^2, \sigma_n^2)'$:

$$\Delta = f(u) = SR_i - SR_n = \frac{\mu_i}{\sqrt{\sigma_i^2}} - \frac{\mu_n}{\sqrt{\sigma_n^2}}. \tag{8}$$

Now, we need an estimator $\hat{\Psi}$ for Ψ, to find the standar error:

$$s(\hat{\Delta}) = \sqrt{T^{-1}\nabla' f(\hat{u})\hat{\Psi}\nabla f(\hat{u})}. \tag{9}$$

The problem in Jobson & Korkie (1981) is that they uses a formula for $\hat{\Psi}$ that crucially relies on *iid* return data for a bivariate normal distribution:

$$\Psi = \begin{bmatrix} \sigma_i^2 & \sigma_{in} & 0 & 0 \\ \sigma_{in} & \sigma_n^2 & 0 & 0 \\ 0 & 0 & 2\sigma_i^4 & 2\sigma_{in}^2 \\ 0 & 0 & 2\sigma_{in}^2 & 2\sigma_n^4 \end{bmatrix}$$

Further information can be found in: Lo (2002); Jobson & Korkie (1981).

2.1 Ledoit and Wolf (2008)

They correct Jobson & Korkie (1981) using robust estimates for Ψ. They also use the bootstrap to find better p-values.

They work with the uncentered second moments, so we define

$$\gamma_i = E(r_{it}^2),\ \gamma_n = E(r_{nt}^2).$$

Now, define Δ as a function of $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$, the vector of primitive statistics:

$$\Delta = f(v) = SR_i - SR_n = \frac{\mu_i}{\sqrt{\gamma_i - \mu_i^2}} - \frac{\mu_n}{\sqrt{\gamma_n - \mu_n^2}}. \tag{10}$$

The gradient of $f(v)$ is:

$$\begin{aligned} \nabla' f(v) &= \left(\frac{\partial f(v)}{\partial \mu_i}, \frac{\partial f(v)}{\partial \mu_n}, \frac{\partial f(v)}{\partial \gamma_i}, \frac{\partial f(v)}{\partial \gamma_n}\right) \\ &= \left(\frac{\gamma_i}{(\gamma_i - \mu_i^2)^{3/2}}, \frac{-\gamma_n}{(\gamma_n - \mu_n^2)^{3/2}}, \frac{-\mu_i}{2(\gamma_i - \mu_i^2)^{3/2}}, \frac{\mu_n}{2(\gamma_n - \mu_n^2)^{3/2}}\right). \end{aligned}$$

And the standard error, $s(\hat{\Delta})$, is given by:

$$s(\hat{\Delta}) = \sqrt{T^{-1}\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v})}.$$

where $\hat{\Psi}$ is a consistent estimator for Ψ.

3 Variance Testing

The classical F-test:
Define the reason of variances as:

$$\Theta = \sigma_i^2/\sigma_n^2. \tag{11}$$

What we want to test is if the reason of variances is different from one statistically significant.

$$H_0 : \Theta = 1 \text{ vs. } H_1 : \Theta \neq 1.$$

Defining the test statistic as $F = \hat{\sigma}_i^2/\hat{\sigma}_n^2$. F_{λ,k_1,k_2} is the λ-quantile of $F_{k1,k2}$, or the F distribution with k_1 and k_2 degrees of freedom. The F-test rejects H_0 at a significance level α iff:

$$F < F_{\alpha/2,T-1,T-1} \text{ or, } F > F_{1-\alpha/2,T-1,T-1}.$$

3.1 Ledoit and Wolf (2011)

First, we reformulate the problem with a log transformation:

$$\Delta = \log(\Theta) = \log(\sigma_i^2) - \log(\sigma_n^2) = \log(\sigma_i^2/\sigma_n^2). \tag{12}$$

OBS: Log transformation, see Efron and Tibshirani 1993, sec 12.6.

What we are going to test is whether the difference in log variances, Δ, is zero or not. In other terms:

$$H_0 : \Delta = 0 \text{ vs. } H_1 : \Delta \neq 0.$$

Define the uncentered second moments as:

$$\gamma_i = E(r_{it}^2), \; \gamma_n = E(r_{nt}^2),$$

Then, we define Δ as a function of primitive statistics, $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$:

$$\begin{aligned}\Delta = f(v) &= \log\left(\gamma_i - \mu_i^2\right) - \log\left(\gamma_n - \mu_n^2\right) \\ &= \log\left(\frac{\gamma_i - \mu_i^2}{\gamma_n - \mu_n^2}\right).\end{aligned}$$

The gradient of the function $f(v)$ is:

$$\nabla' f(v) = \left(\frac{\partial f(v)}{\partial \mu_i}, \frac{\partial f(v)}{\partial \mu_n}, \frac{\partial f(v)}{\partial \gamma_i}, \frac{\partial f(v)}{\partial \gamma_n}\right) \tag{13}$$

$$= \left(\frac{-2\mu_i}{\gamma_i - \mu_i^2}, \frac{2\mu_n}{\gamma_n - \mu_n^2}, \frac{1}{\gamma_i - \mu_i^2}, \frac{1}{\gamma_n - \mu_n^2}\right). \tag{14}$$

And the standard error of $\hat{\Delta}$, $s(\hat{\Delta})$, is given by:

$$s(\hat{\Delta}) = \sqrt{T^{-1}\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v})}. \tag{15}$$

where $\hat{\Psi}$ is a consistent estimator for Ψ.

4 Mean Testing

The classical T-test:
Define the difference of returns as $r_{td} = r_{ti} - r_{tn}$ with moments:

$$\begin{aligned}
E(r_{td}) &= \mu_d = \mu_i - \mu_n \\
V(r_{dt}) &= E[(r_{ti} - r_{tn})^2] - [E(r_{ti} - r_{tn})]^2 \\
V(r_{dt}) &= E(r_{ti}^2) + E(r_{tn}^2) - 2E(r_{ti}r_{tn}) - \mu_i^2 + \mu_n^2 - 2\mu_i\mu_n \\
V(r_{dt}) &= V(r_{ti}) + V(r_{tn}) - 2Cov(r_{ti}, r_{tn}) \\
\sigma_d^2 &= \sigma_i^2 + \sigma_n^2 - 2\sigma_{in}
\end{aligned}$$

What we want to test is if the difference in means is different from zero statistically significant.

$$H_0 : \hat{\mu}_d = 0 \text{ vs. } H_1 : \hat{\mu}_d \neq 0.$$

Define the test statistic as:

$$\Theta = T^{1/2}\frac{\hat{\mu}_d}{\hat{\sigma}_d} = T^{1/2}\frac{\hat{\mu}_i - \hat{\mu}_n}{\hat{\sigma}_i^2 + \hat{\sigma}_n^2 - 2\hat{\sigma}_{in}}. \tag{16}$$

Further, let's define $t_\lambda(k)$ as the λ-quantile of $t(k)$, or the t distribution with k degrees of freedom. The test rejects H_0 at a significance level α iff:

$$|\Theta| > t_{1-\alpha/2}(T-1).$$

5 Reformulation

Based on Ledoit & Wolf (2008, 2011).

What we are going to test is whether the difference in mean, Δ, is zero or not. In other terms:

$$H_0 : \Delta = 0 \text{ vs. } H_1 : \Delta \neq 0.$$

Define Δ as a function of $v = (\mu_i, \mu_n)'$:

$$\Delta = f(v) = \mu_i - \mu_n.$$

The gradient of the function $f(v)$ is:

$$\nabla' f(v) = \left(\frac{\partial f(v)}{\partial \mu_i}, \frac{\partial f(v)}{\partial \mu_n}\right) = (1, -1)\,. \tag{17}$$

And the standard error of $\hat{\Delta}$, $s(\hat{\Delta})$, is given by:

$$\begin{aligned}
s(\hat{\Delta}) &= \sqrt{T^{-1}\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v})} \\
\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v}) &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \hat{\psi}_{11} & \hat{\psi}_{12} \\ \hat{\psi}_{21} & \hat{\psi}_{22} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
&= \hat{\psi}_{11} + \hat{\psi}_{22} - 2\hat{\psi}_{12}
\end{aligned}$$

If we use $\hat{\Psi}$ as the sample covariance matrix:

$$\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v}) = \hat{\sigma}_1^2 + \hat{\sigma}_{22}^2 - 2\hat{\sigma}_{12}.$$

And we have the same case as the classical t-test. So we use HAC consistend methods to estimate $\hat{\Psi}$.

6 CEQ Testing

Based on Ledoit & Wolf (2008, 2011).
Define Δ as a function of $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$:

$$\begin{aligned}\Delta = f(v) &= CEQ_i - CEQ_n \\ &= \mu_i - \frac{\theta}{2}(\gamma_i - \mu_i^2) - \mu_n + \frac{\theta}{2}(\gamma_n - \mu_n^2).\end{aligned}$$

The gradient of the function $f(v)$ is:

$$\begin{aligned}\nabla' f(v) &= \left(\frac{\partial f(v)}{\partial \mu_i}, \frac{\partial f(v)}{\partial \mu_n} \frac{\partial f(v)}{\partial \gamma_i}, \frac{\partial f(v)}{\partial \gamma_n} \right) \\ &= \left(1 + \theta \mu_i, -(1 + \theta \mu_n), -\frac{\theta}{2}, \frac{\theta}{2} \right).\end{aligned}$$

And the standard error of $\hat{\Delta}$, $s(\hat{\Delta})$, is given by:

$$s(\hat{\Delta}) = \sqrt{T^{-1} \nabla' f(\hat{v}) \hat{\Psi} \nabla f(\hat{v})}.$$

where $\hat{\Psi}$ is a consistent estimator for Ψ.

References

JOBSON, J. D., & KORKIE, BOB M. 1981. Performance Hypothesis Testing with the Sharpe and Treynor Measures. *The Journal of Finance*, **36**(4), 889–908.

LEDOIT, OLIVER, & WOLF, MICHAEL. 2008. Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, **15**(5), 850–859.

LEDOIT, OLIVER, & WOLF, MICHAEL. 2011. Robust performance hypothesis testing with the Variance. *Wilmott Magazine*, September, 86–89.

LO, ANDREW W. 2002. The Statistics of Sharpe Ratios. *Financial Analysts Journal*, **58**(4), 36–52.