Notes on Robust Hypothesis Testing

By Paulo F. Naibert

October 18, 2019

1 General Problem

Based on Ledoit & Wolf (2008, 2011, 2018).

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

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

We do *not* assume the distribution to be normal, nor do we assume that returns are independent over time.

The parameter of interest is:

$$\Delta = \theta_i - \theta_n, \tag{1}$$

where θ is a given performance measure. Hence, θ_i is the performance measure for the strategy i, and θ_n is the performance measure for the strategy n.

We are interested in testing:

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

We consider the class of performance measure θ *that can be expressed as a smooth function of a finite number of population moments*. In particular, for $j = i, n$ let

$$v_j^{(k)} := E(r_j^k)$$

denote the (uncentered) kth population moment of the returns of strategy j. Then, for $j = i, n$ we assume that θ_j can be expressed as

$$\theta_j = h(v_j^{(1)}, \ldots, v_j^{(K)}),$$

where $K \geq 1$ is an integer and $h : \mathbb{R}^K \to \mathbb{R}$ is a smooth function (in the sense of being one time continuously differentiable).

For $j = i, n$, let $v_j' = (v_j^{(1)}, \ldots, v_j^{(K)})$. Furthermore, let $v' = (v_i', v_n')$. Then the parameter of interest Δ in (1) can be written as a function of the population moments:

$$\Delta := f(v) = f(v_i, v_n) = h(v_i) - h(v_n) = \theta_i - \theta_n,$$

so that $f : \mathbb{R}^{2K} \to \mathbb{R}$ is also a smooth function, defined as:

$$f(v_i^{(1)}, \ldots, v_i^{(K)}, v_n^{(1)}, \ldots, v_n^{(K)}) = h(v_i^{(1)}, \ldots, v_i^{(K)}) - h(v_n^{(1)}, \ldots, v_n^{(K)}) = \theta_i - \theta_n. \tag{2}$$

1.1 Sample Estimates

For $j = i, n$, denote the (uncentered) kth sample moment of the observed returns by:

$$\hat{v}_j^{(k)} := T^{-1} \sum_{t=1}^{T} r_{tj}^k.$$

Then the estimator of the parameter of interest, Δ, is given by:

$$\hat{\Delta} := \hat{\theta}_i - \hat{\theta}_n \tag{3}$$

where

$$\hat{\theta}_j := h(\hat{v}_j^{(1)}, \dots, \hat{v}_j^{(K)}). \tag{4}$$

For $j = i, n$, let $\hat{v}_j' := (\hat{v}_j^{(1)}, \dots, \hat{v}_j^{(K)})$. Furthermore, let $\hat{v}' := (\hat{v}_i', \hat{v}_n')$. Then the estimator of Δ can also be expressed as:

$$\hat{\Delta} := f(\hat{v})$$

2 Solutions

We assume that:

$$T^{1/2}(\hat{v} - v) \to^d N(0, \Psi), \tag{5}$$

where Ψ is an unknown symmetric PD matrix of dimension $2K \times 2K$. This relation holds under mild regularity conditions. For various sets of sufficient conditions in the TS case, see White (2001), for example.

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

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

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

$$T^{1/2}(\hat{\Delta} - \Delta) \to^d N\left(0; \nabla' f(v) \Psi \nabla f(v)\right). \tag{6}$$

where the $2K \times 1$ vector-valued function $\nabla f(\cdot)$ is the gradient of $f(\cdot)$.

Therefore, if a consistent estimator $\hat{\Psi}$ of Ψ is available, then an asymptotic standard error of for $\hat{\Delta}$, $s(\hat{\Delta})$, is given by:

$$s(\hat{\Delta}) := \sqrt{T^{-1} \nabla' f(\hat{v}) \hat{\Psi} \nabla f(\hat{v})}. \tag{7}$$

Given the formula (2) for $f(\cdot)$, it hold that:

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

$$\begin{aligned} \nabla' f(v) &= (\nabla' h(v_i), -\nabla' h(v_n)) \\ &= \left(\frac{\partial f(v_i)}{\partial v_i^{(1)}}, \dots, \frac{\partial f(v_i)}{\partial v_i^{(K)}}, -\frac{\partial f(v_n)}{\partial v_n^{(1)}}, \dots, -\frac{\partial f(v_n)}{\partial v_n^{(K)}} \right). \end{aligned}$$

Note 1: In the codes we use:

$$\nabla' f(v) = \left(\frac{\partial f(v_i)}{\partial v_i^{(1)}}, -\frac{\partial f(v_n)}{\partial v_n^{(1)}}, \dots, \frac{\partial f(v_i)}{\partial v_i^{(K)}}, -\frac{\partial f(v_n)}{\partial v_n^{(K)}} \right). \blacksquare$$

2.1 Estimating $\hat{\Psi}$

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

$$\Psi = \lim_{T\to\infty} T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} E[y_s y_t'],$$

where

$$y_t' = \left(r_{ti} - v_i^{(1)}, \dots, r_{ti}^K - v_i^{(K)}, r_{tn} - v_n^{(1)}, \dots, r_{tn}^K - v_n^{(K)} \right)$$

Note 2: In the codes we use:

$$y_t' = \left(r_{ti} - v_i^{(1)}, r_{tn} - v_n^{(1)}, \dots, r_{ti}^K - v_i^{(K)}, r_{tn}^K - v_n^{(K)} \right). \blacksquare$$

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

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

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

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

$$\hat{\Psi} := \hat{\Psi}_T = \frac{T}{T-2K} \sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T}\right) \hat{\Gamma}_T(j), \qquad \text{where}$$

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

where

$$\hat{y}_t' = \left(r_{ti} - \hat{v}_i^{(1)}, \dots, r_{ti}^K - \hat{v}_i^{(K)}, r_{tn} - \hat{v}_n^{(1)}, \dots, r_{tn}^K - \hat{v}_n^{(K)} \right)$$

The factor $T/(T-2K)$ 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 given by:

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

where Ψ denotes the cdf of the standard normal distribution.

2.2 Bootstrap Inference

The two-sided distribution function of the studentized statistic is approximated via the bootstrap as follows:

$$L\left(\frac{|\hat{\Delta}-\Delta|}{s(\hat{\Delta})}\right) \approx L\left(\frac{|\hat{\Delta}^*-\hat{\Delta}|}{s(\hat{\Delta}^*)}\right) \tag{8}$$

where Δ is the true difference in the performance measure, $\hat{\Delta}$ is the estimated difference from the original data, $\hat{\Delta}^*$ is the estimated difference from the bootstrap data. $L(X)$ is the distribution function of random variable X.

We use the Circular Boostrap of Politis & Romano (1992), resampling now *blocks of pairs* from the observed pairs $(r_{ti}, r_{tn})'$, $t = 1, \dots, T$ *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:

$$y_t^* = \left(r_{ti}^* - \hat{\mu}_i^*, r_{tn}^* - \hat{\mu}_n^*, r_{ti}^{2\,*} - \hat{\gamma}_i^*, r_{tn}^{2\,*} - \hat{\gamma}_n^*\right), \quad t = 1, \dots T \tag{9}$$

$$\zeta_j = b^{-1/2} \sum_{t=1}^{b} y^*_{t+(j-1)b}, \quad t = 1, \dots, l \tag{10}$$

and

$$\hat{\Psi}^* = l^{-1} \sum_{j=1}^{l} \zeta_j \zeta_j'. \tag{11}$$

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

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

2.2.1 Direct Computation of the p-value

Remark 3.2 of Ledoit & Wolf (2008) or *Remark 3.1* of Ledoit & Wolf (2018).

Denote the original studentized test statistic by:

$$d = \frac{|\hat{\Delta}|}{s(\hat{\Delta})} \tag{12}$$

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

$$\tilde{d}^{*,m} = \frac{|\hat{\Delta}^{*,m} - \hat{\Delta}|}{s(\hat{\Delta}^{*,m})} \tag{13}$$

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

$$\hat{p}^* = \frac{\#\{\tilde{d}^{*,m} \geq d\} + 1}{M+1}. \quad \blacksquare \tag{14}$$

3 Sharpe Ratio Example

Ledoit & Wolf (2008)

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

$$\Delta = f(u) = SR_i - SR_n = \frac{\mu_i}{\sqrt{\sigma_i^2}} - \frac{\mu_n}{\sqrt{\sigma_n^2}}. \tag{15}$$

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

$$s(\hat{\Delta}) = \sqrt{T^{-1}\nabla' f(\hat{u})\hat{\Psi}\nabla f(\hat{u})}. \tag{16}$$

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).

3.1 Ledoit and Wolf (2008)

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

Define γ_i as the uncentered second moment of r_{ti}. We can define Δ as a function of $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$, the vector of the two first moments.

$$\Delta = f(v) = SR_i - SR_n = \frac{\mu_i}{\sqrt{\gamma_i - \mu_i^2}} - \frac{\mu_n}{\sqrt{\gamma_n - \mu_n^2}}. \tag{17}$$

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.

4 Variance Testing

Ledoit & Wolf (2011)

The classical F-test, tests the reason of variances. So, we define the reason of variances as:

$$\Theta = \sigma_i^2/\sigma_n^2. \tag{18}$$

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:

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

4.1 Ledoit and Wolf (2011)

They reformulate the problem with a log transformation:

$$\Delta = \log(\Theta) = \log(\sigma_i^2) - \log(\sigma_n^2) = \log(\sigma_i^2/\sigma_n^2). \tag{19}$$

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

Defining Δ as a function of the first two uncentered moments, $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$:

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

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

$$\begin{aligned}\nabla' f(v) &= \left(\frac{\partial f(v)}{\partial \mu_i}, \frac{\partial f(v)}{\partial \mu_n}, \frac{\partial f(v)}{\partial \gamma_i}, \frac{\partial f(v)}{\partial \gamma_n}\right) && (20) \\ &= \left(\frac{-2\mu_i}{\gamma_i - \mu_i^2}, \frac{2\mu_n}{\gamma_n - \mu_n^2}, \frac{1}{\gamma_i - \mu_i^2}, \frac{1}{\gamma_n - \mu_n^2}\right). && (21)\end{aligned}$$

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

$$s(\hat{\Delta}) = \sqrt{T^{-1}\nabla' f(\hat{v})\hat{\Psi}\nabla f(\hat{v})}. \tag{22}$$

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

5 Mean Testing

The classical T-test verifies if the difference in means is statistically significant different from zero. So, we define the difference of returns as $r_{td} = r_{ti} - r_{tn}$ with moments:

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

Define the test statistic as:

$$
\hat{\Delta} = T^{1/2}\frac{\hat{\mu}_d}{\hat{\sigma}_d} = T^{1/2}\frac{\hat{\mu}_i - \hat{\mu}_n}{\hat{\sigma}_i^2 + \hat{\sigma}_n^2 - 2\hat{\sigma}_{in}}. \tag{23}
$$

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:

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

5.1 Reformulation

Based on Ledoit & Wolf (2018).
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:

$$
\nabla' f(v) = \left(\frac{\partial f(v)}{\partial \mu_i}, \frac{\partial f(v)}{\partial \mu_n}\right) = (1, -1)\,. \tag{24}
$$

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.

Ledoit, Olivier, & Wolf, Michael. 2018 (October). *Robust Performance Hypothesis Testing with Smooth Functions of Population Moments.* Working Paper Series 305. Department of Economics, University of Zurich.

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

Politis, Dimitris N., & Romano, Joseph P. 1992. A circular block-resampling procedure for stationary data. *In:* LePage, Raoul, & Billard, Lynne (eds), *Exploring the limits of bootstrap*. Wiley series in probability and mathematical statistics. New York: John Wiley.