Methods of Economic Investigation II (EC403)
Problem Set #4

Solutions

1. Mean square convergence is stronger than convergence in probability, so it is possible that a random sequence can converge in probabilty to a constant or a random variable but, at the same time it may not converge in mean square as it is the case of:

$$\{X_n\} = \left\{ \begin{matrix} 0 & \text{with probability} & 1-\frac{1}{n} \\ n^2 & \text{with probability} & \frac{1}{n} \end{matrix} \right\}$$

We have

$$\lim_{n\longrightarrow\infty} \Pr\left[|X_n| > \delta\right] = 0$$

because for any $\delta > 0$

$$\Pr\left[|X_n| > \delta\right] = \Pr\left[X_n = n^2\right] = \frac{1}{n}.$$

Now, if we want to have convergence in mean square we have to prove that $EX_n \longrightarrow 0$ as well as $\text{var}(X_n) \longrightarrow^P 0$ as $n \longrightarrow \infty$. But

$$EX_n = 0 \times (1 - \frac{1}{n}) + n^2 \times (\frac{1}{n}) = n \longrightarrow \infty.$$

Hence $X_n \longrightarrow^P 0$, but X_n do not converge in mean square to 0.

Now suppose that $\sqrt{T}(\hat{\theta} - 2\pi) \longrightarrow^d N(0,1)$. From Taylor's expansion evaluated at 2π is:

$$\cos(\hat{\theta}) = \cos 2\pi - \sin(2\pi)(\hat{\theta} - 2\pi) - \frac{1}{2}\cos(2\pi)(\hat{\theta} - 2\pi)^2 + \frac{1}{3!}\sin(\tilde{\theta})(\hat{\theta} - 2\pi)^3,$$

where $\tilde{\theta}$ is an intermediate point between $\hat{\theta}$ and 2π (see mean value theorem).

Hence $\cos(\hat{\theta}) = 1 - \frac{1}{2}(\hat{\theta} - 2\pi)^2 + \frac{1}{3!}\sin(\tilde{\theta})(\hat{\theta} - 2\pi)^3$ after some straightforward algebra and so

$$2T\left[1 - \cos(\hat{\theta})\right] = T(\hat{\theta} - 2\pi)^2 - \frac{2}{3!}T\sin(\tilde{\theta})(\hat{\theta} - 2\pi)^3$$

By Continuous mapping theorem if $\sqrt{T}(\hat{\theta} - 2\pi) \longrightarrow^d N(0,1)$ then $T(\hat{\theta} - 2\pi)^2 \longrightarrow^d \chi_{(1)}$, so if we could prove that $\frac{2}{3!}T\sin(\tilde{\theta})(\hat{\theta} - 2\pi)^3 \longrightarrow^p 0$ then using Slutsky's theorem it follows that

$$2T\left[1-\cos(\hat{\theta})\right] \longrightarrow^{d} \chi_{(1)}$$

Fortunately it is true that $\frac{2}{3!}T\sin(\tilde{\theta})(\hat{\theta}-2\pi)^3 \longrightarrow^{p} 0$, because: $\sin(\tilde{\theta})$ is bounded between -1 and +1, while $(\hat{\theta}-2\pi)^3 = O_p(T^{-\frac{3}{2}})$. Therefore, $T\sin(\tilde{\theta})(\hat{\theta}-2\pi)^3 = O_p(T^{-\frac{1}{2}})$.

2. (a) Basically we want to test $\mathrm{H}_0 : \beta_2 = 0$ vs. $\mathrm{H}_A : \beta_2 \neq 0$. The logic underlying the Hausman test is: construct the estimator that is consistent under both H_0 and H_A but asymptotically inefficient, construct the estimator that is consistent only under H_0 but asymptotically efficient in this case, and then compare the two estimators. In our case we let $\widehat{\beta}_1$ be the OLS estimator of β_1 in the full regression of y on X_1 and X_2, and $\widetilde{\beta}_1$ be the OLS estimator of β_1 in the regression of y on X_1.

$$\begin{aligned}\widehat{\beta}_1 &= (X_1'M_2X_1)^{-1}X_1'M_2y\\ \widetilde{\beta}_1 &= (X_1'X_1)^{-1}X_1'y,\end{aligned}$$

where $M_2 = I - X_2(X_2'X_2)^{-1}X_2'$. The Hausman test is

$$H = (\widehat{\beta}_1-\widetilde{\beta}_1)'\widehat{V}^{-1}(\widehat{\beta}_1-\widetilde{\beta}_1),$$

where $\widehat{V}$ is an estimator of the asymptotic variance of $\widehat{\beta}_1-\widetilde{\beta}_1$. Note that under the classical assumptions

$$V = \mathrm{var}(\widehat{\beta}_1-\widetilde{\beta}_1) = \mathrm{var}(\widetilde{\beta}_1) - \mathrm{var}(\widehat{\beta}_1) = \sigma^2\left((X_1'M_2X_1)^{-1} - (X_1'X_1)^{-1}\right),$$

which can be estimated by

$$\widehat{V} = s^2\left((X_1'M_2X_1)^{-1} - (X_1'X_1)^{-1}\right),$$

where s^2 is the sample residual variance from either regression. The statistic is compared with the critical value of the χ^2 distribution with k_2 degrees of freedom.

If we had computed the F-test (or Wald test) we should write the null hypothesis as $R\beta = r$, where $R_{K_2\times(K_1+K_2)} = \left[0\cdots\vdots I_{K_2}\right]$, $\beta_{(K_1+K_2)\times 1} = \left[\beta_1\cdots\vdots\cdots\right]'$, and $r_{K_2\times 1} = [0\cdots]'$. Then the statistic is:

$$F = (R\hat{\beta}-r)'\left[s^2R(X'X)^{-1}R'\right]^{-1}(R\hat{\beta}-r),$$

which is in our case:

$$\tilde{\beta}_2'\left\{s^2\left[0\cdots\vdots I_{K_2}\right](X'X)^{-1}\left[0\cdots\vdots I_{K_2}\right]'\right\}\tilde{\beta}_2 \longrightarrow^{d} \chi_{(K_2)}.$$

When $X_1'X_2 = 0$, the Hausman test will be identically zero because $\widehat{\beta}_1 = \widetilde{\beta}_1$ [the variance V is also zero in this case], while the Wald test can still be computed and becomes:

$$\frac{\tilde{\beta}_2'(X_2'X_2)\tilde{\beta}_2}{s^2}.$$

b) Now suppose $K_j = 1$, $j = 1, 2$

$$t^* = \frac{\hat{\beta}_2}{s^* \sqrt{(X_2' M_1 X_2)^{-1}}}$$

to test $\mathrm{H}_0 : \beta_2 = 0$ vs. $\mathrm{H}_A : \beta_2 \neq 0$, where $\hat{\beta}_2$ is computed from the full regression, i.e.,

$$\hat{\beta}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 y,$$

but

$$s^{*2} = \frac{\varepsilon^{*\prime} \varepsilon^*}{n-1},$$

with $\varepsilon^* = M_1 y$. This means that ε^* is computed as residual in the short regression $\varepsilon_i^* = y_i - X_{1i}\tilde{\beta}_1$. We know that under standard assumptions (normality of the distribution of y_i) the distribution of $\hat{\beta}_2$ is

$$N\left[\beta_2, \sigma^2 (X_2' M_1 X_2)^{-1}\right]$$

so that

$$Z_2 = \frac{\hat{\beta}_2 - \beta_2}{\sigma \sqrt{(X_2 M_1 X_2)^{-1}}} \sim N(0,1).$$

If we replace the true standard deviation with the estimated s.d. ($s^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-2}$, $\hat{\varepsilon} = y - X_1\hat{\beta}_1 - X_2\hat{\beta}_2$ and M the matrix that projects in the space orthogonal to $Colspan(X)$), we get the well known result that:

$$t = \frac{\hat{\beta}_2 - \beta_2}{s \sqrt{(X_2 M_1 X_2)^{-1}}} \sim t_{(n-2)}$$

because of the independence between $\hat{\beta}_2$ and s^2 (and normality), i.e.,

$$\frac{\frac{\hat{\beta}_2 - \beta_2}{\sigma \sqrt{(X_2 M_1 X_2)^{-1}}} \sim N(0,1)}{\sqrt{\frac{(n-2)s^2}{\sigma}} \sim \sqrt{\chi^2_{(n-2)}}}.$$

In the case of t^*, $\hat{\beta}_2$ and s^{*2} are no more independent and in finite sample t^* is no more a student's t distribution. We can show however that both t^* and t converge asymptotically in distribution to a $N(0,1)$ under H_0. In fact under the null $s^* \rightarrow^P \sigma$ and using the Cramer's theorem the two ratios converge in distribution to a standard normal.

In small sample the two ratios behave differently, as in general

$$s^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-2} \leq \frac{\varepsilon^{*\prime}\varepsilon^*}{n-1} = s^{*2}$$

and t will be bigger than t^* (if the number of observation is reasonably big), so we will end up rejecting H_0 more frequently.

3. Define the moment function

$$G_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[z_i\left(y_i - \sin(\theta x_i)\right]\right.$$

and let

$$Q_n(\theta) = G_n(\theta)' W_n G_n(\theta)$$

where W_n is a symmetric positive definite matrix of weights. The GMM estimator of θ is then

$$\hat{\theta}_{GMM} = \arg\min_{\theta\in\Theta}\left[Q_n(\theta)\right].$$

It can be shown that under certain conditions minimizing $Q_n(\theta)$ produce a consistent and asymptotically normal estimator of θ_0. (Note: If $W_n = I_n$ we are minimizing the sum of squares). Specifically, we can prove that the asymptotic distribution of $\hat{\theta}_{GMM}$ is such that:

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \longrightarrow^D N(0, (\Gamma' W \Gamma)^{-1}\Gamma' W \Omega W \Gamma (\Gamma' W \Gamma)^{-1})$$

where

$$\begin{aligned}
\Gamma &= p\lim_{n\to\infty} \frac{\partial G_n(\theta_0)}{\partial\theta} \\
\Omega &= \lim_{n\to\infty} \mathrm{var}(\sqrt{n} G_n(\theta_0)) \\
W &= p\lim_{n\to\infty} W_n.
\end{aligned}$$

Let's now work all these things out:

$$\lim_{n\longrightarrow\infty} \mathrm{var}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[z_i\left(y_i - \sin(\theta_0 x_i)\right]\right.) = \lim_{n\longrightarrow\infty}\frac{1}{n}\mathrm{var}\left[\sum_{i=1}^{n}\mathrm{var}(z_i u_i)\right],$$

which is equal, under standard contitions (z_i and u_i i.i.d.), to

$$\mathrm{var}(z_i u_i) = E(u_i^2 z_i z_i') = \Omega$$

as $E(z_i u_i) = 0$ by assumption. Then

$$\frac{\partial G_n(\theta)}{\partial\theta} = -\frac{1}{n}\sum_{i=1}^{n} z_i \cos(\theta_0 x_i) x_i$$

and so

$$\Gamma = p\lim_{n\longrightarrow\infty}\left(\frac{\partial G_n(\theta)}{\partial\theta}\right) = p\lim_{n\longrightarrow\infty}\left(-\frac{1}{n}\sum_{i=1}^{n} z_i\cos(\theta_0 x_i)x_i\right) = -E\left[z_i\cos(\theta_0 x_i)x_i\right]$$

under regularity assumption.

How to choose the weighting matrix optimally:

$$W_n^{-1} = \tilde{\Omega} = \frac{1}{n}\sum_{i}^{n}\left(y_i - \sin(\hat{\theta}x_i)\right)^2 z_i z_i'$$

in this case we hope to get, under certain conditions:

$$\tilde{\Omega} \longrightarrow^P E(u_i^2 z_i z_i')$$

and choosing $W_n = \tilde{\Omega}^{-1}$ we actually get

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \longrightarrow^D N(0, (\Gamma'\Omega^{-1}\Gamma)^{-1}).$$

The problem is that we need a preliminary estimation of $\hat{\theta}$ because we need to know $\tilde{\Omega}$. One solution is to run a two stage estimation:

1. Calculate $\hat{\theta}$ with $W_n = I_n$
2. Use $\hat{\theta}$ to get the optimal $W_n = \tilde{\Omega}^{-1}$
3. Then $\hat{\theta}_{GMM} = \arg\min\left[G_n(\theta)' W_n G_n(\theta)\right]$