Daeyoung Lim

Economic Time Series

Fall Semester, 2016

1. Replicate Figures 1,5,6 and 7 in the Matlab exercise. In doing so, make the following modifications to the figures.

   (1) In figure 1, use "plot" instead of "stairs", and use a dotted line instead of a solid line.

   (2) In figure 5, plot the data from Jan 2004 only and remove all grid lines.

   (3) Put figures 6 and 7 on a common figure frame in 2×2 format.

Solution:

2. Suppose that the data are generated by the following model

$$\begin{aligned} y &= X\beta + u && (1) \\ &= X_1\beta_1 + X_2\beta_2 + u && (2) \end{aligned}$$

where $X = [X_1, X_2]$. Assume that X is of full column rank, $T^{-1}X'X \xrightarrow{p} Q$, and $\beta_2 \neq 0$. Denote the OLS estimate for (1) by $\widehat{\beta}_1$ and $\widehat{\beta}_2$. Suppose you estimate the regression

$$y = X_1\beta_1 + e \quad (3)$$

by OLS and denote the resulting estimate by $\widehat{b}_1$.

   a) Show that $\widehat{b}_1$ is inconsistent for β_1, with assuming $\mathrm{E}\left(X_1'X_2\right) \neq 0$.

   b) The inconsistency of $\widehat{b}_1$ is an example of the omitted variable bias. A natural estimate would then be based on an instrumental variable procedure. Show that the OLS estimate $\widehat{\beta}_1$ can indeed be given an IV interpretation.

Solution:

a) The omitted variable bias lingers even when the sample size grows large causing the estimators to be inconsistent. We can show this by simply plugging in the true model of y to the estimator $\widehat{b}_1$.

$$\begin{aligned} \widehat{b}_1 &= \left(X_1'X_1\right)^{-1} X_1'y && (4) \\ &= \left(X_1'X_1\right)^{-1} X_1'\left(X_1\beta_1 + X_2\beta_2 + u\right) && (5) \\ &= \beta_1 + \left(X_1'X_1\right)^{-1} X_1'X_2\beta_2 + \left(X_1'X_1\right)^{-1} X_1'u && (6) \end{aligned}$$

The problem states that $\mathrm{E}\left(X_1' X_2\right) \neq 0, \beta_2 \neq 0$, preserving the second term. Note that the only *random variable* in the above equation is u, converging in probability to zero as $T \rightarrow \infty$ where T is the dimension of y $(T \times 1)$. Thus, there is no reason for the second term $\left(X_1' X_1\right)^{-1} X_1' X_2 \beta_2$ to disappear.

$$\hat{b}_1 \xrightarrow{p} \beta_1 + \beta_2 \left(X_1' X_1\right)^{-1} X_1' X_2 \tag{7}$$

b) X_2 is by assumption independent of the error term u. Therefore, using the method of moment, if we have $\mathrm{E}\left(x_{i2} u_i\right) = 0$,

$$\frac{1}{T} \sum_{t=1}^{T} x_{i2} u_i = \frac{1}{T} X_2' u = \frac{1}{T} X_2' \left(y - X_1 \beta_1\right). \tag{8}$$

Solving with respect to β_1 yields

$$\hat{\beta}_1 = \left(X_2' X_1\right)^{-1} X_2' y \tag{9}$$

which is the IV estimate.

3. Consider the linear regression model:

$$y = X\beta + u \tag{10}$$

$$y, u : T \times 1 \tag{11}$$

$$X : T \times k \tag{12}$$

$$\beta : k \times 1 \tag{13}$$

with the q moment conditions $\mathrm{E}\left(z_t u_t\right) = 0$. Let

$$J_T\left(\beta, W_T\right) = g_T\left(\beta\right)' W_T g_T\left(\beta\right) \tag{14}$$

where $g_T\left(\beta\right) = T^{-1} \sum z_t \left(y_t - x_t' \beta\right)$ and W_T is some weighting matrix. The GMM estimator for β is obtained as the minimizer of $J_T\left(\beta, W_T\right)$.

a) How does your choice of W_T affect the GMM estimator? Discuss the implications on the consistency, the asymptotic normality and efficiency.

b) If the model is exactly identified $(k = q)$, explain why the choice of W_T becomes irrelevant.

Solution:

a)

4. Let $y_t \overset{\text{iid}}{\sim} \text{Exp}(\theta)$ for $t = 1, \ldots, T$.

   a) Derive the score function, Hessian function and information matrix, using the exponential density.

   b) Derive the MLE for θ. Sketch the proof that the MLE is asymptotically normal. Be specific with the asymptotic variance.

Solution:

a) The likelihood function of $y_1, \ldots, y_T$,

$$L(\theta \mid y_1 \ldots, y_T) = \theta^T \exp\left(-\theta \sum_{t=1}^{T} y_t\right). \tag{15}$$

Taking the logarithm yields

$$\ell(\theta \mid y_1, \ldots, y_T) = T \log\theta - \theta \sum_{t=1}^{T} y_t. \tag{16}$$

By definition of the score function is the first derivative of the log-likelihood.

$$\frac{d\ell}{d\theta} = \frac{T}{\theta} - \sum_{t=1}^{T} y_t. \tag{17}$$

The Hessian gets reduced to the second derivative for a univariate function.

$$\frac{d^2\ell}{d\theta^2} = -\frac{T}{\theta^2} - \sum_{t=1}^{T} y_t. \tag{18}$$

To get the Fisher information,

$$\begin{aligned} \mathcal{I}_T(\theta) &= \frac{1}{T}\text{E}\left(\frac{T}{\theta^2} + \sum_{t=1}^{T} y_t\right) &(19)\\ &= \frac{1}{\theta^2} + \frac{1}{T}\sum_{t=1}^{T} \text{E}(y_t) &(20)\\ &= \frac{1}{\theta^2} + \frac{1}{\theta} &(21) \end{aligned}$$

b) We have obtained the first and second derivatives of the log-likelihood already. Recall:

$$\frac{d\ell}{d\theta} = \frac{T}{\theta} - \sum_{t=1}^{T} y_t \tag{22}$$

$$\frac{d^2\ell}{d\theta^2} = -\frac{T}{\theta^2} - \sum_{t=1}^{T} y_t \tag{23}$$

Using the first-order condition, the MLE comes with a closed-form expression.

$$\widehat{\theta}^{\text{MLE}} = T \Bigg/ \left(\sum_{t=1}^{T} y_t\right) \tag{24}$$

The second-order condition validates that the estimator is actually a maximum since

$$\frac{d^2\ell}{d\theta^2} < 0. \tag{25}$$

Now, even as rough a proof as what we will shortly give here takes at least 2 steps: the consistency of MLE and the asymptotic normality of MLE.

- *(Consistency)* Recall that we take the product of every single PDF of y_t through $t = 1, \ldots, T$ to compute the likelihood function, which also means taking the logarithm will convert the product into summation.

$$\ell\left(\theta \mid y_1, \ldots, y_T\right) = \sum_{t=1}^{T} \ell\left(\theta \mid y_t\right) \tag{26}$$

By the strong law of large numbers, we get the following relation:

$$\frac{1}{T}\sum_{t=1}^{T} \ell\left(\theta \mid y_t\right) \xrightarrow{a.s.} \mathrm{E}_{\theta_\circ} \ell\left(\theta \mid y_1\right) \tag{27}$$

for some unknown true parameter value $\theta_\circ$. We can then show that the expected log-likelihood function w.r.t. the true parameter is always greater than that of an arbitrary parameter θ by *Kullback-Leibler divergence*. The KL divergence is defined as follows.

$$\mathrm{KL}\left(f(y_1 \mid \theta_\circ) \,\|\, f(y_1 \mid \theta)\right) = \mathrm{E}_{\theta_\circ}\left[\log \frac{f(y_1 \mid \theta_\circ)}{f(y_1 \mid \theta)}\right] \tag{28}$$

$$= -\int \log \frac{f(y_1 \mid \theta)}{f(y_1 \mid \theta_\circ)}\; f(y_1 \mid \theta_\circ)\, dy_1 \tag{29}$$

By *Jensen's inequality*,

$$\underbrace{-\log \int \frac{f(y_1 \mid \theta)}{\cancel{f(y_1 \mid \theta_\circ)}} \cancel{f(y_1 \mid \theta_\circ)}\, dy_1}_{=0} \leq \underbrace{-\int \log \frac{f(y_1 \mid \theta)}{f(y_1 \mid \theta_\circ)}\, f(y_1 \mid \theta_\circ)\, dy_1}_{=\mathrm{KL}(f(y_1 \mid \theta_\circ) \,\|\, f(y_1 \mid \theta))}. \tag{30}$$

Therefore, it always follows that the KL divergence is nonnegative. In fact, it is strictly positive if $f(y_1 \mid \theta) \neq f(y_1 \mid \theta_\circ)$. This indicates that

$$\theta_\circ = \sup_{\theta \in \Omega} \mathrm{E}_{\theta_\circ} \ell(\theta \mid y_1). \tag{31}$$

Recall the following:

$$\widehat{\theta}^{\mathrm{MLE}} = \sup_{\theta \in \Omega} \frac{1}{T} \sum_{t=1}^{T} \ell(\theta \mid y_i). \tag{32}$$

Therefore by 27, $\widehat{\theta}^{\mathrm{MLE}} \xrightarrow{p} \theta_\circ$ for a finite parameter space Ω. We can also prove this for a compact parameter space by starting from the lemma that

$$\frac{1}{T} \sum_{t=1}^{T} \ell(\theta \mid y_t) \xrightarrow{\text{uniformly convergent}} \int \ell(\theta \mid y_1)\, f(y_1 \mid \theta_0)\, dy_1 \quad \left(= \mathrm{E}_{\theta_0} \ell(\theta \mid y_1)\right) \tag{33}$$

which is equivalent to

$$\Pr\left(\sup_{\theta \in \Omega} \left| \frac{1}{T} \sum_{t=1}^{T} \ell(\theta \mid y_t) - \mathrm{E}_{\theta_0} \ell(\theta \mid y_1) \right| > \epsilon \right) \xrightarrow{p} 0, \quad \forall \epsilon > 0. \tag{34}$$

34 is the expression which we refer to as the *almost-sure convergence* corresponding to the *almost-everywhere convergence* in measure theory literatures. Almost-everywhere convergence is not always interchangeable with uniform convergence in an arbitrary measure space. Nonetheless, *Egorov's theorem* states that in a finite measure space like a probability measure space, they are equivalent. Anyway, the proof for the consistency of MLE ends here.

- *(Asymptotic Normality)* We approximate the score function with its first-order Taylor expansion around the true parameter $\theta_\circ$ and apply the mean value theorem.

$$\frac{d\ell}{d\theta} \approx \left.\frac{d\ell}{d\theta}\right|_{\theta=\theta_\circ} + \left.\frac{d^2\ell}{d\theta^2}\right|_{\theta=\overline{\theta}} (\theta - \theta_\circ) \tag{35}$$

where $\overline{\theta}$ lies somewhere between θ and $\theta_\circ$. Since MLE is the value which sets the first derivative to zero, we can think of the following identity.

$$\left.\frac{d\ell}{d\theta}\right|_{\theta=\theta_\circ} + \left.\frac{d^2\ell}{d\theta^2}\right|_{\theta=\overline{\theta}} \left(\widehat{\theta}^{\mathrm{MLE}} - \theta_\circ\right) = 0 \tag{36}$$

This, in turn, translates to the following relationship.

$$\widehat{\theta}^{\mathrm{MLE}} - \theta_\circ = -\left(\left.\frac{d\ell}{d\theta}\right|_{\theta=\theta_\circ}\right) \Bigg/ \left(\left.\frac{d^2\ell}{d\theta^2}\right|_{\theta=\overline{\theta}}\right) \tag{37}$$

with $\overline{\theta} = s\widehat{\theta}^{\mathrm{MLE}} + (1-s)\,\theta_\circ,\ s \in [0,1]$. Let's slowly examine the RHS of 37. First, the numerator can be expressed as a summation of the log-likelihood of a single observation.

$$\left.\frac{d\ell\left(\theta \mid y_{1:T}\right)}{d\theta}\right|_{\theta=\theta_\circ} = \sum_{t=1}^{T} \left.\frac{d\ell\left(\theta \mid y_t\right)}{d\theta}\right|_{\theta=\theta_\circ} \tag{38}$$

By the *Central Limit Theorem*,

$$\mathrm{E}\left(\left.\frac{d\ell\left(\theta \mid y_1\right)}{d\theta}\right|_{\theta=\theta_\circ}\right) = 0 \tag{39}$$

$$\mathrm{Var}\left(\left.\frac{d\ell\left(\theta \mid y_1\right)}{d\theta}\right|_{\theta=\theta_\circ}\right) = T\mathcal{I}_1\left(\theta\right) \tag{40}$$

$$\left.\frac{d\ell\left(\theta \mid y_{1:T}\right)}{d\theta}\right|_{\theta=\theta_\circ} \xrightarrow{d} \mathcal{N}\left(0, T\mathcal{I}_1\left(\theta_\circ\right)\right) \tag{41}$$

where $\mathcal{I}_1$ is the Fisher information for a single observation y_1. The calculations for the expectation and the variance are given in the end. Now the denominator behaves like the following which makes use of the weak law of large numbers.

$$\left.\frac{d^2\ell\left(\theta \mid y_{1:T}\right)}{d\theta^2}\right|_{\theta=\overline{\theta}} = \sum_{t=1}^{T} \left.\frac{d^2\ell\left(\theta \mid y_t\right)}{d\theta^2}\right|_{\theta=\overline{\theta}} \xrightarrow{p} T\mathrm{E}\left(\left.\frac{d^2\ell\left(\theta \mid y_1\right)}{d\theta^2}\right|_{\theta=\overline{\theta}}\right) = -T\mathcal{I}_1\left(\overline{\theta}\right) \tag{42}$$

With 39, 42, and the *Slutsky's theorem*, we can conclude that 37 converges in distribution to a normal distribution.

$$-\left(\left.\frac{d\ell}{d\theta}\right|_{\theta=\theta_\circ}\right) \Bigg/ \left(\left.\frac{d^2\ell}{d\theta^2}\right|_{\theta=\overline{\theta}}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{T\mathcal{I}_1\left(\theta_\circ\right)}{\left(T\mathcal{I}_1\left(\overline{\theta}\right)\right)^2} = \frac{\mathcal{I}_T\left(\theta_\circ\right)}{\left(\mathcal{I}_T\left(\overline{\theta}\right)\right)^2}\right) \tag{43}$$

Finally, we know that $\overline{\theta} \in \left[\widehat{\theta}^{\mathrm{MLE}}, \theta_\circ\right]$ if $\widehat{\theta}^{\mathrm{MLE}} < \theta_\circ$ or $\overline{\theta} \in \left[\theta_\circ, \widehat{\theta}^{\mathrm{MLE}}\right]$ if $\widehat{\theta}^{\mathrm{MLE}} \geq \theta_\circ$. As

$T \to \infty$, $\widehat{\theta}^{\text{MLE}} \xrightarrow{p} \theta_\circ$ which also means $\overline{\theta} \xrightarrow{p} \theta_\circ$. Thus,

$$\widehat{\theta}^{\text{MLE}} \xrightarrow{d} \mathcal{N}\left(\theta_\circ, \left(\mathcal{I}_T\left(\theta_\circ\right)\right)^{-1}\right) \tag{44}$$

There are 3 different ways to compute the Fisher information and all three are equivalent under regularity conditions. The three are

$$\mathcal{I}\left(\theta\right) = \mathrm{E}\left(\left(\frac{d\ell}{d\theta}\right)^2\right) \tag{45}$$

$$= -\mathrm{E}\left(\frac{d^2\ell}{d\theta^2}\right) \tag{46}$$

$$= \mathrm{Var}\left(\frac{d\ell}{d\theta}\right) \tag{47}$$