Final Examination

Introduction to Econometrics, 2016
Written by
In Choi

Instructions: 1. Consulting an A4-sized sheet of paper on which you wrote down necessary information is allowed.
2. Answering questions simply by writing "yes" or "no" carries no points.

1. (20 points) We are concerned with the regression model

$$y_t = \beta_1 + \beta_2 D_t + u_t,\ (t = 1, ..., T) \tag{1}$$

where

$$D_t = \begin{cases} 0, & 1 \le t \le T_0 \\ 1, & T_0 + 1 \le t \le T \end{cases},$$

T_0 is known and $u_t \sim iid(0, \sigma^2)$. This model shows that the mean of y_t changes from β_1 to $\beta_1 + \beta_2$.

(a) If T_0 is quite small relative to T, what problems do you expect for the regression using model (1)?
Ans. Because D_t does not have much variation, the OLS estimator of β has high variance. Or, because $\{D_t\}$ is almost equivalent to $\{1\}$, there is a problem of near multicollinearity.

(b) Show that the OLS estimator of β_2 using Model (1) is consistent if $T_0 = [\frac{T}{2}]$ where $[x]$ denotes the integer part of x. (e.g., $[2.89] = 2$.)
Ans. The OLS estimator is unbiased. Its variance is $\frac{\sigma^2}{\sum_{t=1}^{T}(D_t - \bar{D})^2} = \frac{\sigma^2}{T_0 - \frac{T_0^2}{T}} = \frac{\sigma^2}{\frac{T+c}{2} - \frac{\left(\frac{T}{2}+c\right)^2}{T}}$, where c is a constant. The variance diverges to infinity, so the estimator is consistent.

2. (10 points) Consider the AR(1) process

$$y_t = \alpha y_{t-1} + u_t,\ (t = 2, ..., T)$$

where $y_1 = 0$, $u_t \sim iid\ (0, \sigma^2)$ and $\alpha = -1$. Show that this stochastic process is nonstationary.
Ans. Write $y_t = u_t + (-1)u_{t-1} + ... + (-1)^{t-2}u_2$. Then, we find $Var(y_t) = (t-1)\sigma^2$, which depends on t. Thus, $\{y_t\}$ is not stationary.

3. (20 points) Let the true regression model be

$$y_i = \alpha + \beta x_i + u_i,\ (i = 1, ..., n),$$

where $u_i \sim iid\,(0, \sigma_u^2)$ and $x_i \sim iid\,(0, \sigma_x^2)$. But we observe

$$x_i^* = x_i + w_i,\ w_i \sim iid\left(0, \sigma_w^2\right)$$

instead of x_i due to measurement error. Assume that $\{x_i\}$, $\{u_i\}$ and $\{w_i\}$ are independent (this implies that $\sum_{i=1}^n x_i u_i/n$, $\sum_{i=1}^n x_i w_i/n \xrightarrow{p} 0$ as $n \to \infty$). The regression model we use is

$$y_i = \alpha + \beta x_i^* + v_i.$$

(a) Is the OLS estimator of β consistent?
Ans. Because $\sum_{i=1}^n x_i u_i/n$, $\sum_{i=1}^n x_i w_i/n \xrightarrow{p} 0$ and $\sum_{i=1}^n x_i/n$, $\sum_{i=1}^n w_i/n \xrightarrow{p} 0$, we obtain

$$\hat{\beta} - \beta = \frac{\sum_{i=1}^n (x_i^* - \bar{x}^*) v_i}{\sum_{i=1}^n (x_i^* - \bar{x}^*)^2} = \frac{\sum_{i=1}^n (x_i + w_i - \bar{x} - \bar{w})(u_i - \beta w_i)/n}{\sum_{i=1}^n (x_i + w_i - \bar{x} - \bar{w})^2/n} \xrightarrow{p} \frac{-\beta \sigma_w^2}{\sigma_x^2 + \sigma_w^2}.$$

The OLS estimator is not consistent.

(b) Suppose that there is another measurement of x_i that is also subject to error and represented as

$$z_i = x_i + a_i, \;\; a_i \sim iid\left(0, \sigma_a^2\right).$$

Assume $\{a_i\}$, $\{u_i\}$ and $\{w_i\}$ are independent. Is the IV estimator using $\{z_i\}$ consistent?
Ans. The IV estimator is written as

$$\hat{\beta}_{IV} - \beta = \frac{\sum_{i=1}^n (z_i - \bar{z})(u_i - \beta w_i)}{\sum_{i=1}^n (x_i^* - \bar{x}^*)(z_i - \bar{z})}.$$

Since $\sum_{i=1}^n (z_i - \bar{z})(u_i - \beta w_i)/n \xrightarrow{p} 0$ and $\sum_{i=1}^n (x_i^* - \bar{x}^*)(z_i - \bar{z})/n \xrightarrow{p} \sigma_x^2$, the IV estimator is consistent.

4. (20 points) Consider the panel data model

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \gamma z_i + a_i + u_{it}, \;\; (i = 1, ..., N; \; t = 1, ..., T),$$

where a_i is unobserved for every i.

(a) When is it appropriate to use the random effects estimator?
Ans. If $a_i + u_{it}$ are uncorrelated with all the regressors, it is appropriate to sue the random effects estimator because it is an efficient estimator.

(b) Can we estimate the coefficient γ if we use the random effects estimator?
Ans. Yes, we can because differencing or subtraction of time series means is not performed.

5. (45 points) GINI coefficient of Korea was regressed on annual real GDP of Korea, the results of which are shown the following page.

(a) Is the coefficient of LOG(RGDP) statistically significant at the 5% level?

(b) If real GDP increases by 10%, how much does GINI coefficient change?

(c) If we want to use a large sample approach for the significance of entire coefficients, how should we modify value of the F-statistic in the following table? What distribution should we use?

Dependent Variable: GINI
Method: Least Squares
Date: 10/26/15 Time: 18:14
Sample: 1990 2014
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.213349	0.050568	-4.219088	0.0003
LOG(RGDP)	0.055437	0.005579	9.937452	0.0000

R-squared	0.811093	Mean dependent var	0.288760
Adjusted R-squared	0.802880	S.D. dependent var	0.022842
S.E. of regression	0.010142	Akaike info criterion	-6.267722
Sum squared resid	0.002366	Schwarz criterion	-6.170212
Log likelihood	80.34653	F-statistic	98.75296
Durbin-Watson stat	1.004660	Prob(F-statistic)	0.000000

Figure 1: