Midterm I

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

1. Consider the following model

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i, \qquad i = 1, \ldots, N$$

where $(y_i, x_{2i}, x_{3i})_{i=1}^N$ are observed and have finite moments, and ε_i is an unobserved error term. Suppose this model is estimated by ordinary least squares. Denote the OLS estimator by $\mathbf{b}$.

   (a) What are the essential conditions required for unbiasedness of $\mathbf{b}$? What are the essential conditions required for consistency of $\mathbf{b}$? Explain the difference between unbiasedness and consistency.

   (b) Show how the conditions for consistency can be written as moment conditions (if you have not done so already). Explain how a method of moments estimator can be derived from these moment conditions. Is the resulting estimator any different from the OLS one?

   For the rest of the problem suppose that $\text{cov}(x_{3i}, \varepsilon_i) \neq 0$.

   (c) Give two examples of cases where one can expect a nonzero correlation between a regressor, x_{3i}, and the error ε_i.

   (d) In this case, is it possible still to make appropriate inferences based on the OLS estimator while adjusting the standard errors appropriately?

   (e) Explain how an instrumental variable, z_i, say, leads to a new moment condition and, consequently, an alternative estimator for the coefficient vector $\boldsymbol{\beta}$. In total, how many moment conditions we have? Write down these moment conditions explicitly.

   (f) Why does this alternative estimator lead to a smaller R^2 than the OLS one? What does this say of the R^2 as a measure for the adequacy of the model?

   (g) Why can't we choose $z_i = x_{2i}$ as an instrument for x_{3i}, even if we know that $E[x_{2i}\varepsilon_i] = 0$? Would it be possible to use x_{2i}^2 as an instrument for x_{3i}?

2. Consider the following model

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t^* + u_t \qquad (\star) \\ x_t &= x_t^* + e_t \end{aligned}$$

where u_t has zero mean and is uncorrelated with x_t^* and e_t. We observe y_t and x_t only. Assume that e_t has zero mean and is uncorrelated with x_t^* and that x_t^* is also has zero mean.

(a) Show that in the model that econometrician actually estimates the error term, say v_t, is negatively correlated with x_t if $\beta_1 > 0$. What does this imply about the OLS estimator of β_1 from the regression of y_t on x_t?

(b) In addition to the previous assumptions, assume that u_t and e_t are uncorrelated with all past values of x_t^* and e_t; in particular, with x_{t-1}^* and e_{t-1}. Show that $E[x_{t-1}v_t] = 0$ where v_t is the error term in the model from part (a).

(c) Suppose that this is a time series model, are x_t and x_{t-1} likely to be correlated? Explain.

(d) What do parts (b) and (c) suggest as a useful strategy for *consistently* estimating β_0 and β_1?

3. Consider the structural equation and reduced form

$$\begin{aligned} y_i &= \beta x_i^2 + \varepsilon_i \\ x_i &= \gamma z_i + u_i \\ E[z_i \varepsilon_i] &= 0 \\ E[z_i u_i] &= 0 \end{aligned}$$

with x_i^2 treated as endogenous so that $E[x_i^2 \varepsilon_i] \neq 0$. For simplicity assume there are no intercepts in the regressions. All variables are scalars and assume $\gamma \neq 0$.

Consider the following estimator. First estimate γ by OLS of x_i on z_i and obtain the fitted values $\hat{x}_i = \hat{\gamma} z_i$. Second, estimate β by OLS of y_i on $\hat{x}_i^2$.

(a) Write out this estimator $\hat{\beta}$ explicitly as a function of the sample data.

(b) Find its probability limit as $n \to \infty$

(c) In general, is $\hat{\beta}$ consistent estimator for β? Is there a reasonable condition under which $\hat{\beta}$ is consistent?

4. In the lecture we proved the Delta-Method under the assumption that g is continuous and $g'(\theta) \neq 0$. Let us now try to obtain a similar result when $g'(\theta) = 0$ but $g''(\theta) \neq 0$:. In this exercise, we are going to show the following:

If

$$\sqrt{n}\,(X_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

and g is continuous, $g'(\theta) = 0$, $g''(\theta) \neq 0$. Then

$$n\,[g(X_n) - g(\theta)] \xrightarrow{d} \frac{1}{2}\sigma^2 g''(\theta) \chi_1^2$$

(a) Carry out a second-order Taylor expansion of g about θ and using the assumptions above show that it can be written as

$$g(X_n) - g(\theta) = \frac{1}{2} g''(\theta)\,(X_n - \theta)^2 + \text{Remainder Term}$$

(b) At this step we multiplied both sides by $\sqrt{n}$ to prove the delta method in the lecture, but note that it would not work here. Argue that we should instead multiply both sides by n in order to be able to use the main hypothesis above. Finally, using the fact that $\chi_1^2 = Z^2$, where $Z \sim N(0,1)$, complete the proof.

(c) For this part, suppose that $X_n \sim \text{Bin}(n,p)$ where the positive integer n is large and $0 < p < 1$. Let $g(x) = x^3 - x$. Find the limiting distribution of

$$n\left[g\left(\frac{X_n}{n}\right) - c\right]$$

for an appropriate constant c when $p = 1/\sqrt{3}$.

5. Consider Hansen and Singleton (1982) model that we discussed in class. But now assume that the momentary utility function of the consumer is given as

$$U(C_t) = C_t - \theta\frac{C_t^2}{2}.$$

(a) Set up the consumer's optimization problem, derive the first order conditions and obtain Euler equation.

(b) Write the population moment condition(s) and the corresponding sample analog(s)?

(c) Describe how you would proceed to compute the GMM estimator for the population parameters δ and θ?

(d) What would be a good instrument? Explain.

(e) Suppose that another researcher suggested to use the following utility function instead $U(C_t) = \theta_1 C_t - \theta_2\frac{C_t^2}{2}$. Is this a good specification? Can we identify θ_1 and θ_2 separately in the data?

Reminder. In Hansen and Singleton (1982) model an agent maximizes her expected utility of current and future consumption by solving the following problem

$$\max_{\{C_{t+s}\}_{s=0}^S} E_t\left\{\sum_{s=0}^{S}\delta^s U(C_{t+s})\right\}$$

s.t.

$$C_{t+s} + q_{t+s} = w_{t+s} + (1 + r_{t+s})q_{t+s-1}$$

E_t : the expectation operator conditional upon time t information

C_t : consumption in period t

q_{t+s} : financial wealth at the end of period $t+s$

r_{t+s} : return on financial wealth

w_{t+s} : labour income