Midterm I - Solutions

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.

The essential conditions for $\mathbf{b}$ *to be unbiased for* $\boldsymbol{\beta}$ *are that* $E[\varepsilon_i] = 0$ *and that* $\{\varepsilon_1, \ldots, \varepsilon_N\}$ *is independent of* $\{x_1, \ldots, x_N\}$. *The essential condition for* $\mathbf{b}$ *to be consistent is* $E[\varepsilon_i x_i] = 0$ *plus some regularity conditions conditions on the data matrix, i.e.* $(1/N)(\mathbf{Z}'\mathbf{X}) \xrightarrow{p} \mathbf{Q}_{zx}$ *where the last matrix is a positive definite matrix.*

An estimator that is unbiased provides, in repeated sampling, an estimate that is - on average - equal to the true value. That is, we do not expect the estimator to be systematically too high or too low, no matter how big or small the sample is. An estimator is consistent if its probability limit is equal to the true value. That is, the probability that the estimator differs from the true value becomes infinitesimally small when the sample size becomes infinitely large. While we hope that the estimator does not systematically under- or overestimate the true value when the sample size is reasonably large, this result is only true in the limit.

(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?

We can write $E[\varepsilon_i x_{ki}] = 0$ *as*

$$E[(y_i - \mathbf{x}_i'\boldsymbol{\beta})x_{ki}] = 0, \qquad (k = 2, 3).$$

This is a set of moment conditions as the expression in brackets is a function of observable data and unknown parameters. A method of moments estimator is obtained by taking the sample equivalent of the above expectations, setting it *to zero and then solving for the unknown parameters. This reproduces the OLS estimator* **b**.

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.

Some examples of such situations

1) Endogeneity and Omitted Variable Bias (Omitted variable bias)

2) Simultaneity and Reverse Causality

3) Measurement error in the regressors

4) Autocorrelation with a Lagged Dependent Variable

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

No. If $E[x_{3i}\varepsilon_i] \neq 0$ the OLS estimator is biased and inconsistent, no matter what other assumptions we are making. Correcting standard errors does not solve the problem.

(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.

We already have two moment conditions since $E[\varepsilon_i] = 0$ and $E[x_{2i}\varepsilon_i] = 0$:

$$E[(y_i - \mathbf{x}_i'\boldsymbol{\beta})] = 0$$
$$E[(y_i - \mathbf{x}_i'\boldsymbol{\beta})x_{2i}] = 0$$

Then an instrumental variable z_i gives rise to a new moment condition that can replace the invalid one. In this case, we have the following three moment conditions:

$$E[(y_i - \mathbf{x}_i'\boldsymbol{\beta})] = 0$$
$$E[(y_i - \mathbf{x}_i'\boldsymbol{\beta})x_{2i}] = 0$$
$$E[(y_i - \mathbf{x}_i'\boldsymbol{\beta})z_i] = 0$$

These moment conditions typically provide sufficient information to identify the three parameters in $\boldsymbol{\beta}$.

(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?

OLS minimizes the residual sum of squares and therefore maximizes the R^2. Any other estimator, including instrumental variables, results in a lower R^2. Note *that we are typically not interested in obtaining an R^2 that is as high a possible, but in obtaining consistent (or unbiased) estimates for the coefficients of interest that are as accurate as possible. The R^2 does not tell us which estimator is the preferred one. The R^2 tells us how well the model fits the data (in a given sample) and typically is only interpreted in this way when the model is estimated by ordinary least squares.*

(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}?

We cannot use x_{2i} as instrument for x_{3i} because x_{2i} is already included in the model. The corresponding moment condition is based on $E[x_{2i}\varepsilon_i] = 0$ and this is already exploited. In theory, it is possible to use x_{2i}^2 as an instrument for x_{3i}. This produces the additional moment condition $E[(y_i - \mathbf{x}_i'\boldsymbol{\beta})x_{2i}^2] = 0$. However, identification is then based upon the assumption that the functional form of the model is correct and that x_{2i}^2 is appropriately excluded from the model. Economic theory rarely allows us to be that confident in specifying our functional form. In practice one should view a linear model as only an approximation, and employing such artificial instruments can be easily criticized.

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?

Plug $x_t^* = x_t - e_t$ into $(*)$ to get

$$y_t = \beta_0 + \beta_1 x_t + v_t$$

where $v_t = u_t - \beta_1 e_t$. By assumption, u_t is uncorrelated with x_t^* and e_t; therefore, u_t is uncorrelated with x_t. Since e_t is uncorrelated with x_t^*,

$$E[x_t e_t] = E[(x_t^* + e_t)e_t] = E[x_t^* e_t] + E[e_t^2] = \sigma_e^2$$

Therefore, with v_t defined as above

$$Cov(x_t, v_t) = Cov(x_t, u_t) - \beta_1 Cov(x_t, e_t) = -\beta_1\sigma_e^2 < 0\,.$$

Because the explanatory variable and the error have negative covariance, the OLS estimator of β_1 has a downward bias.

(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).

By assumption

$$E[x_{t-1}^* u_t] = E[e_{t-1}u_t] = E[x_{t-1}^* e_t] = E[e_{t-1}e_t] = 0$$

and also since $x_t = x_t^* + e_t$

$$E[x_{t-1}u_t] = E[x_{t-1}e_t] = 0$$

Therefore,

$$E[x_{t-1}v_t] = E[x_{t-1}u_t] - \beta_1 E[x_{t-1}e_t] = 0$$

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

Most economic time series, unless they represent the first difference of a series or the percentage change, are positively correlated over time. If the initial equation is in levels or logs, x_t *and* x_{t-1} *are likely to be positively correlated. If the model is for first differences or percentage changes, there still may be positive or negative correlation between* x_t *and* x_{t-1}.

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

Under the assumptions made, x_{t-1} *is exogenous in*

$$y_t = \beta_0 + \beta_1 x_t + v_t$$

as we showed in part (b): $Cov(x_{t-1}, v_t) = E(x_{t-1}v_t) = 0$. *Second,* x_{t-1} *will often be correlated with* x_t*, and we can check this easily enough by running a regression of* x_t *of* x_{t-1}. *This suggests estimating the equation by instrumental variables, where* x_{t-1} *is the IV for* x_t. *The IV estimator will be consistent for* β_0 *and* β_1 *and asymptotically normally distributed.*

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.

In the second stage we are estimating the model

$$y_i = \theta \hat{x}_i^2 + v_i$$

where $\hat{x}_i = \hat{\gamma} z_i$ *and*

$$\hat{\gamma} = \frac{\sum_{i=1}^N z_i y_i}{\sum_{i=1}^N z_i^2}$$

is estimated in the first stage. Therefore, the OLS estimator for θ *is the estimator described in the question:*

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^N \hat{x}_i^2 y_i}{\sum_{i=1}^N \hat{x}_i^4} \\ &= \frac{\sum_{i=1}^N z_i^2 y_i}{\hat{\gamma}^2 \sum_{i=1}^N z_i^4}\end{aligned}$$

Here everything is sample data as desired.

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

First note that $\hat{\gamma} \xrightarrow{p} \gamma$ *as the OLS assumptions are satisfied there, in particular it is assumed that* $E[z_i u_i] = 0$.

$$\begin{aligned}\hat{\beta} &= \frac{\sum_{i=1}^N z_i^2 y_i}{\hat{\gamma}^2 \sum_{i=1}^N z_i^4} \\ &\xrightarrow{p} \frac{E[z_i^2 y_i]}{\gamma^2 E[z_i^4]} \qquad \left(WLLN \text{ and } \hat{\gamma} \xrightarrow{p} \gamma\right) \\ &= \frac{E[z_i^2 \left(\beta x_i^2 + \varepsilon_i\right)]}{\gamma^2 E[z_i^4]} \\ &= \frac{E\left[z_i^2 \left(\beta \left(\gamma z_i + u_i\right)^2 + \varepsilon_i\right)\right]}{\gamma^2 E[z_i^4]} \\ &= \beta + \frac{1}{\gamma^2 E[z_i^4]} \left(2\beta\gamma E[z_i^3 u_i] + \beta E[z_i^2 u_i^2] + E[z_i^2 \varepsilon_i]\right)\end{aligned}$$

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

In general, it's not a consistent estimator for β. *However, under certain conditions consistency can be established. A set of such conditions:* z_i *and* u_i *are independent,* $E[u_i] = 0$, *and* $E[z_i^2 \varepsilon_i] = 0$.

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

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

A second-order Taylor expansion of g about θ:

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

using $g'(\theta) = 0$ and then rearranging the terms gives

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

(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.

For convenience let's drop the remainder term and then multiply both sides by n to get

$$n\,[g(X_n) - g(\theta)] = \frac{1}{2}g''(\theta)\left[\sqrt{n}\,(X_n - \theta)\right]^2$$

Note that the main hypothesis can be expressed as

$$\frac{1}{\sigma}\sqrt{n}\,(X_n - \theta) \xrightarrow{d} N(0, 1) \equiv Z$$

Then by Slutsky theorem

$$\left[\frac{1}{\sigma}\sqrt{n}\,(X_n - \theta)\right]^2 \xrightarrow{d} Z^2 = \chi_1^2$$

$$\implies \left[\sqrt{n}\,(X_n - \theta)\right]^2 \xrightarrow{d} \sigma^2\chi_1^2$$

Therefore we have

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

(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 $p = 1/\sqrt{3}$.

First note that we have this result by the CLT

$$\sqrt{n}\left[\frac{X_n}{n} - p\right] \xrightarrow{d} N(0, p(1-p))$$

Now, in order to be able to apply the delta method we must choose

$$c = g(p) = \left(\frac{1}{\sqrt{3}}\right)^3 - \frac{1}{\sqrt{3}} = -\frac{2\sqrt{3}}{9}$$

Also compute that $g'(\frac{1}{\sqrt{3}}) = 0$ *and* $g''(\frac{1}{\sqrt{3}}) = 2\sqrt{3} \neq 0$. *Now note that the conditions of the previous exercise are satisfied so we can apply that result to write*

$$n\left[g\left(\frac{X_n}{n}\right) - c\right] \xrightarrow{d} \frac{1}{2}\sigma^2 g''(p)\chi_1^2.$$

Finally, substituting $c = -\frac{2\sqrt{3}}{9}$, $g''(\frac{1}{\sqrt{3}}) = 2\sqrt{3}$, *and* $\sigma^2 = p(1-p) = \frac{1}{\sqrt{3}}\left(1 - \frac{1}{\sqrt{3}}\right)$ *we obtain*

$$n\left[g\left(\frac{X_n}{n}\right) + \frac{2\sqrt{3}}{9}\right] \xrightarrow{d} \left(1 - \frac{1}{\sqrt{3}}\right)\chi_1^2.$$

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.

$$E_t\left\{\delta\frac{U'(C_{t+1})}{U'(C_t)}(1 + r_{t+1}) - 1\right\} = 0$$

$$\Longrightarrow E_t\left\{\delta\left(\frac{1 - \theta C_{t+1}}{1 - \theta C_t}\right)(1 + r_{t+1}) - 1\right\} = 0$$

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

$$E_t\left\{\left[\delta\left(\frac{1 - \theta C_{t+1}}{1 - \theta C_t}\right)(1 + r_{t+1}) - 1\right] z_{it}\right\} = 0$$

$$\frac{1}{T}\sum_{t=1}^{T}\left[\delta\left(\frac{1 - \theta C_{t+1}}{1 - \theta C_t}\right)(1 + r_{t+1}) - 1\right] z_{it} = 0, \qquad i = 1, 2$$

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

We need at least two moments conditions... Discussed in PS 5

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

For example, past consumption would be a good instrument... Discussed in PS 5

(e) Suppose that a 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?

No, it is not a good specification and we cannot identify these two parameters separately because this utility function is just a multiple of the previous one: $U(C_t) = \theta_1 \left(C_t - (\theta_2/\theta_1) \frac{C_t^2}{2} \right)$. *So, defining* $\theta = \theta_2/\theta_1$ *would reduce this to the previous specification.*