Section 6: Asymptotic theory and identification (solutions)

ARE 210

October 10, 2017

1) $\mathbf{O_p}$ **and Taylor approximations**: Let $\{X_i\}_{i=1}^{\infty}$ iid, and $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, and let $\mu = \mathbf{E}[X_i]$.

a) Show that $(\overline{X}_n - \mu)^k$ is $O\left(n^{-k/2}\right)$.

First, $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\text{d}} N(0, V[X_i])$. $g(x) = x^k$ is continuous, by the continuous mapping theorem $n^{k/2}(\overline{X}_n - \mu)^k$ converges in distribution. By definition of O_p, $(\overline{X}_n - \mu)^k = O_p(n^{-k/2})$.

b) Show that if $T_n = O(n^k)$, then $T_n = o(n^{k+\epsilon})$ $\forall$ $\epsilon > 0$.

By definition, $\frac{T_n}{n^k} \xrightarrow{\text{d}} T$ for some T. $\frac{T_n}{n^{k+\epsilon}} = n^{-\epsilon}\frac{T_n}{n^k}$. $n^{-\epsilon} \xrightarrow{\text{p}} 0$ and $\frac{T_n}{n^k} \xrightarrow{\text{d}} T$, so $\frac{T_n}{n^{k+\epsilon}} \xrightarrow{\text{d}} 0T \Rightarrow \frac{T_n}{n^{k+\epsilon}} \xrightarrow{\text{p}} 0$.

2) **OLS**: Consider the OLS model - $Y_i = X_i'\beta + \epsilon_i$, with $(Y_i, X_i', \epsilon_i) \sim P$ iid, and assume $\mathbf{E}_P[X'\epsilon] = 0$ and $\mathbf{E}_P[XX']$. Assume (Y_i, X_i) is observed

a) Show P is point identified.

First, note that $X_iY_i = X_iX_i'\beta + X_i\epsilon$. Taking expectations, $\mathbf{E}_P[X_iY_i] = \mathbf{E}_P[X_iX_i']\beta$, using $\mathbf{E}_P[X'\epsilon] = 0$. Using the full rank condition, this means $\mathbf{E}_P[X_iX_i']^{-1}\mathbf{E}_P[X_iY_i] = \beta$.

Second, P_X is identified from the distribution of (Y, X). Since β is identified, $P_{Y-X'\beta|X}$ is also identified, or $P_{\epsilon|X}$. This gives us $P_{X,\epsilon}$, which with β is sufficient for $P_{Y,X,\epsilon} = P$.

b) Suggest an estimator of β.

By the analogy principle, we can use $\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_iY_i\right)$.

c) Additionally, assume $\epsilon_i \perp X_i$. Derive the asymptotic distribution of $\hat{\beta}$.

First, note that $\sqrt{n}(\beta - \hat{\beta}) = \left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_iY_i\right) - \mathbf{E}_P[X_iX_i']^{-1}\sqrt{n}\mathbf{E}_P[X_iY_i]$. $\frac{1}{n}\sum_{i=1}^{n} X_iX_i' \xrightarrow{\text{p}} \mathbf{E}[X_iX_i']$. Since the inverse matrix operator is continuous, this implies

$\left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1} \xrightarrow{\text{p}} \mathbf{E}[X_i X_i']^{-1}$. Second,

$$\begin{aligned}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i Y_i - \mathbf{E}_P[X_i Y_i]\right) &= \sqrt{n}\frac{1}{n}\sum_{i=1}^n X_i \epsilon_i \\ &\xrightarrow{\text{d}} N(0, V_P[X_i\epsilon_i])\end{aligned}$$

By LLCLT, making use of the definition of Y_i and that $\mathbf{E}_P[X'\epsilon] = 0$. Next, note that $V_P[X_i\epsilon_i] = \mathbf{E}_P[X_i\epsilon_i^2 X_i'] = \mathbf{E}_P[X_i X_i']V_P[\epsilon_i]$ using $\epsilon_i \perp X_i$. Putting this together, this yields $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{\text{d}} N(0, \mathbf{E}_P[X_i X_i']^{-1} V_P[\epsilon_i])$.

3) **Selection models and potential outcomes**: Consider the potential outcomes model (Y_{i1}, Y_{i0}, D_i), where Y_{i1} and Y_{i0} are binary, and let $Y_i = Y_{i1}D_i + Y_{i0}(1 - D_i)$. Additionally, consider the selection model $Y_i = \mathbf{1}\{(1, D_i)\beta - U_i \geq 0\}$, where $U_i \sim N(0,1)$, and assume $U_i \perp D_i$. Suppose (Y_i, D_i) are observed.

a) Show that β is identified, and that the two models are observationally equivalent. Did the normality assumption matter?

Note that $P_\beta[Y_i = 1|D_i = 0] = \Phi(\beta_0)$ and $P_\beta[Y_i = 1|D_i = 1] = \Phi(\beta_0 + \beta_1)$. Therefore $\beta_0 = \Phi^{-1}(P[Y_i = 1|D_i = 0])$ and $\beta_1 = \Phi^{-1}(P[Y_i = 1|D_i = 1]) - \beta_0$, so β is identified. Moreover, since these two probabilities fully characterize the joint distribution of (Y_i, D_i), the two models are observationally equivalent. The normality assumption did not matter.

b) Show that $\mathbf{E}[Y_{i1}|D_i = 0]$ is not identified. Calculate $\mathbf{E}[Y_{i1}|D_i = 0]$ using the selection model.

In a simple case, suppose $\mathbf{P}[Y_{i1}|D_i = 1] = \mathbf{P}[Y_{i0}|D_i = 0] = 1$, and take $\mathbf{E}[Y_{i1}|D_i = 0] = 1$ and $\mathbf{E}[Y_{i1}|D_i = 0] = 0$. Both of these result in the same observed data, so $\mathbf{E}[Y_{i1}|D_i = 0] = 0$ is not identified. Under the selection model, $\mathbf{E}[Y_{i1}|D_i = 0] = \mathbf{E}[Y_i|D_i = 0]$.

c) How can we make the selection model more flexible to accomodate $\mathbf{E}[Y_{i1}|D_i = 0] \neq \mathbf{E}[Y_i|D_i = 1]$?

We can relax the assumption that $U_i \perp D_i$. One way to do this is to allow $D_i = \mathbf{1}[(1, U_i)\gamma - V_i > 0]$, where $V_i \sim N(0,1)$, and assuming $V_i \perp U_i$.

What value of γ would correspond to the assumption that $Y_i \perp D_i$? Is this model equivalent to a fully flexible potential outcomes framework? What are we implicitly assuming?