ECO 522 - Fall 2021 Midterm Exam

Haixiang Zhu

November 19, 2021

1. A just-identified method-of-moments model (m=k)

   (a) Moment condition

   $$E[g_i^j(\theta_0)] = 0, j = 1, 2, \ldots, m$$

   (b) The variance of the asymptotic distribution
   To estimate θ with $m = k$, we define a quadratic form below,

   $$S_1(\theta) = \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} g_i(\theta)\right)' W_n \left(\frac{1}{n}\sum_{i=1}^{n} g_i(\theta)\right)$$

   where W_n is symmetric and positve definite.
   FOC for $\hat{\theta}$ gives

   $$\left(\frac{1}{n}\sum_{i=1}^{n} G_i(\hat{\theta})\right)' W_n \left(\frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\theta})\right) = 0_{k\times 1}$$

   where

   $$G_i = \begin{pmatrix} \frac{\partial g_i^1}{\partial \theta_1} & \frac{\partial g_i^1}{\partial \theta_2} & \cdots & \frac{\partial g_i^1}{\partial \theta_k} \\ \frac{\partial g_i^2}{\partial \theta_1} & \frac{\partial g_i^2}{\partial \theta_2} & \cdots & \frac{\partial g_i^2}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_i^k}{\partial \theta_1} & \frac{\partial g_i^k}{\partial \theta_2} & \cdots & \frac{\partial g_i^k}{\partial \theta_k} \end{pmatrix}$$

   By a uniform law of large numbers and the consistency of $\hat{\theta}$, FOC becomes

   $$\Gamma' W \left(\frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\theta})\right) \stackrel{a}{=} 0$$

   where $\Gamma = E[G_i(\theta_0)]$ and W_n converges to a positive definite symmetric matrix W.

Together with Taylor expansion,

$$\frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta_0) + \frac{1}{n}\sum_{i=1}^{n} G_i(\theta^*)(\hat{\theta}-\theta_0)$$

where θ^* lies between θ_0 and $\hat{\theta}$.
Plugging it back, we obtain an asymptotic version of FOC,

$$\Gamma' W\left(\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_0)\right) + \Gamma' W\Gamma\cdot\sqrt{n}(\hat{\theta}-\theta_0) \overset{a}{=} 0$$

Rearranging,

$$\sqrt{n}(\hat{\theta}-\theta_0) \overset{a}{=} -(\Gamma' W\Gamma)^{-1}\Gamma' W\left(\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_0)\right)$$

Along with moment condition in (a),

$$\begin{aligned} Var[\sqrt{n}\frac{1}{n}\sum_{i=1}^{n} g_i(\theta_0)] &= \frac{1}{n}\sum_{i} Var[g_i(\theta_0)] \\ &= E[g_i^2(\theta_0)] + [Eg_i(\theta_0)]^2 \\ &= E[g_i(\theta_0)g_i(\theta_0)'] \end{aligned}$$

Applying delta method,

$$Var[\sqrt{n}(\hat{\theta}-\theta_0)] = (\Gamma' W\Gamma)^{-1}\Gamma' W\Delta W\Gamma(\Gamma' W\Gamma)^{-1}$$

where $\Delta = E[g_i(\theta_0)g_i(\theta_0)']$

(c) Hypothesis testing
Construct a Wald test statistic

$$n(R\hat{\theta}-r)'(RVR')^{-1}(R\hat{\theta}-r) \sim \chi_q^2$$

where q is the rank of matrix R, $V = (\Gamma' W\Gamma)^{-1}\Gamma' W\Delta W\Gamma(\Gamma' W\Gamma)^{-1}$,
$\Delta = E[g_i(\theta_0)g_i(\theta_0)']$, $\Gamma = E[G_i(\theta_0)]$ and $G_i = \begin{pmatrix} \frac{\partial g_i^1}{\partial\theta_1} & \frac{\partial g_i^1}{\partial\theta_2} & \cdots & \frac{\partial g_i^1}{\partial\theta_k} \\ \frac{\partial g_i^2}{\partial\theta_1} & \frac{\partial g_i^2}{\partial\theta_2} & \cdots & \frac{\partial g_i^2}{\partial\theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_i^k}{\partial\theta_1} & \frac{\partial g_i^k}{\partial\theta_2} & \cdots & \frac{\partial g_i^k}{\partial\theta_k} \end{pmatrix}$.

2. Dynamic nonlinear panel-data model

(a) θ can be estimated consistently using the GMM method. Two important assumptions, so-called "weak exoneneity", are as follows.

$$\begin{cases} E(\epsilon_{it}|X_{it}, X_{it-1}) & = 0 \\ E(\epsilon_{it-1}|X_{it}, X_{it-1}) & = 0 \end{cases}$$

Proof. First, let $\phi(X_{it}, \theta) \equiv \phi_{it}(\theta)$. Applying the first-differences approach,

$$\begin{cases} Y_{it} & = \phi_{it}(\theta)u_i + \alpha Y_{it-1} + \epsilon_{it} \\ Y_{it-1} & = \phi_{it-1}(\theta)u_i + \alpha Y_{it-2} + \epsilon_{it-1} \end{cases}$$
$$\Rightarrow Y_{it} - \frac{\phi_{it}(\theta)}{\phi_{it-1}(\theta)}Y_{it-1} = \alpha\left(Y_{it-1} - \frac{\phi_{it}(\theta)}{\phi_{it-1}(\theta)}Y_{it-2}\right) + \epsilon_{it} - \frac{\phi_{it}(\theta)}{\phi_{it-1}(\theta)}\epsilon_{it-1}$$

Note that the fixed effect u_i has been eliminated, we have the possibility of using X_{it}, X_{it-1} as instruments in a moment specification

$$\begin{aligned} g_{it}(\theta) &= \begin{bmatrix} X_{it} \\ \\ X_{it-1} \end{bmatrix} \left[\left(Y_{it} - \frac{\phi_{it}(\theta)}{\phi_{it-1}(\theta)}Y_{it-1}\right) - \alpha\left(Y_{it-1} - \frac{\phi_{it}(\theta)}{\phi_{it-1}(\theta)}Y_{it-2}\right)\right] \\ &= \begin{bmatrix} X_{it} \\ \\ X_{it-1} \end{bmatrix} \left(\epsilon_{it} - \frac{\phi_{it}(\theta)}{\phi_{it-1}(\theta)}\epsilon_{it-1}\right) \end{aligned}$$

Since two weak exoneneity conditions hold, the moment condition $Eg_{it}(\theta_0) = 0$ with θ_0 being the true value of θ, is met. □

3. Evaluating programs with panel data

(a) Consider two groups of workers, participants A and non-participants B.

For participants,

$$\begin{cases} Y_{A0} & = \alpha + u_A + \epsilon_{A0} \\ Y_{A1} & = \alpha + \beta + \delta + u_A + \epsilon_{A1} \end{cases}$$
$$\Rightarrow (Y_{A1} - Y_{A0}) = \beta + \delta + (\epsilon_{A1} - \epsilon_{A0})$$

For non-participants,

$$\begin{cases} Y_{B0} & = \alpha + u_B + \epsilon_{B0} \\ Y_{B1} & = \alpha + \beta + u_B + \epsilon_{B1} \end{cases}$$
$$\Rightarrow (Y_{B1} - Y_{B0}) = \beta + (\epsilon_{B1} - \epsilon_{B0})$$

Two important assumptions for DID method:

- Parallel time trend
  It means the pre-program and post-program averages are same ($\beta_1 = \beta_2 = \beta$)
- No "Ashenfelter dip"

$$\begin{cases} E(\epsilon_{A1} - \epsilon_{A0}|D_{it} = 1) & = 0 \\ E(\epsilon_{B1} - \epsilon_{B0}|D_{it} = 0) & = 0 \end{cases}$$

By the assumptions above and assume that large sample n_1 for participants and n_2 for non-participants.

$$\hat{\delta} = \frac{1}{n_1}\sum_{A=1}^{n_1}(Y_{A1} - Y_{A0}) - \frac{1}{n_2}\sum_{B=1}^{n_2}(Y_{B1} - Y_{B0})$$

(b) Applying first difference approach for participants group,

$$\begin{cases} Y_{it} = \alpha + \beta + D_{it}\delta + u_i + \epsilon_{it} \\ Y_{it-1} = \alpha + \beta + D_{it-1}\delta + u_i + \epsilon_{it-1} \end{cases}$$
$$\Rightarrow Y_{it} - Y_{it-1} = (D_{it} - D_{it-1})\delta + (\epsilon_{it} - \epsilon_{it-1})$$

Rewrite the model as

$$\tilde{Y}_{it} = \tilde{D}_{it}\delta + \tilde{\epsilon}_{it}$$

where $\tilde{Y}_{it} = Y_{it} - Y_{it-1}, \tilde{D}_{it} = D_{it} - D_{it-1}, \tilde{\epsilon}_{it} = \epsilon_{it} - \epsilon_{it-1}$
Running OLS on the panel data,

$$\hat{\delta} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{D}_{it}\tilde{D}'_{it}\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{D}_{it}\tilde{Y}_{it}\right)$$

and

$$\sqrt{n}(\hat{\delta} - \delta) \overset{a}{=} \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{D}_{it}\tilde{D}'_{it}\right)^{-1} \cdot \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{D}_{it}\tilde{\epsilon}_{it}$$

Since

$$E(\tilde{D}_{it}\tilde{\epsilon}_{it}) = 0$$
$$\Rightarrow E(\sum_{t=1}^{T}\tilde{D}_{it}\tilde{\epsilon}_{it}) = 0$$

and

$$Var(\sum_{t=1}^{T}\tilde{D}_{it}\tilde{\epsilon}_{it}) = \sum_{t=1}^{T}\sum_{s=1}^{T}E(\tilde{D}_{it}\tilde{\epsilon}_{it}\tilde{\epsilon}'_{is}\tilde{D}'_{is})$$
$$\equiv V_i$$

If we assume that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{N} V_i = V$$

Then

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{D}_{it}\tilde{\epsilon}_{it} \xrightarrow{d} N(0,V)$$

Substituting back and further assume that $\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{D}_{it}\tilde{D}'_{it} \xrightarrow{p} W$,

$$\begin{aligned}\sqrt{n}(\hat{\delta}-\delta) &\overset{a}{=} W^{-1}\cdot N(0,V)\\ &\xrightarrow{d} N(0, W^{-1}VW^{-1})\end{aligned}$$

Therefore, the standard errors of the limiting distribution of $\sqrt{n}(\hat{\delta}-\delta)$ is $\sqrt{W^{-1}VW^{-1}}$, where $\hat{W} = plim\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{D}_{it}\tilde{D}'_{it}, \hat{V} = plim\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} E(\tilde{D}_{it}\tilde{\epsilon}_{it}\tilde{\epsilon}'_{is}\tilde{D}'_{is})$.

(c) Following the same method in (b),

$$Y_{it} - Y_{it-1} = (X'_{i,t} - X'_{i,t-1})\gamma + (D_{it} - D_{it-1})\delta + (\epsilon_{it} - \epsilon_{it-1})$$

Rewrite the model as

$$Y = B\xi + \varepsilon$$

where $Y = Y_{it} - Y_{it-1}, B = \begin{pmatrix} X'_{i,t} - X'_{i,t-1} & D_{it} - D_{it-1}\end{pmatrix}, \xi = \begin{pmatrix}\gamma \\ \delta\end{pmatrix}, \varepsilon = \epsilon_{it} - \epsilon_{it-1}$

Then run the OLS on the panel data.

$$\hat{\xi} = (B'B)^{-1}BY$$

Hence, we can get $\hat{\delta}$ from $\hat{\xi}$.

4. Sample selection model

(a) The program participation rule is

$$D_i = 1 \text{ if } Y_{i,1} \geq Y_{i,0}$$

The expected value of earnings among all women who decide to participate in the program is

$$\begin{aligned}E(Y_{i,1}|D_i = 1) &= \mu_1 + E(\epsilon_{i,1}|D_i = 1)\\ &= \mu_1 + E(\epsilon_{i,1}|\epsilon_{i,1} - \epsilon_{i,0} \geq (\mu_1 - \mu_0))\end{aligned}$$

(b) The expected value of their earnings had they not participated is

$$\begin{aligned} E(Y_{i,0}|D_i = 1) &= \mu_0 + E(\epsilon_{i,0}|D_i = 1) \\ &= \mu_0 + E(\epsilon_{i,0}|\epsilon_{i,1} - \epsilon_{i,0} \geq (\mu_1 - \mu_0)) \end{aligned}$$

(c) The first problem is that NLS is used for nolinear model such that $y_i = f(x_i, \beta) + \epsilon_i$, which is not suitable here.

The second problem is that NLS requires user to provide initial values for the unknown parameters before the software can begin the optimization. Bad starting values can lead to local minimum instead of global minimum.

(d) The average program effect is

$$\begin{aligned} E(Y_{1,i} - Y_{0,i}) &= D \cdot E(Y_1|D) - (1 - D) \cdot E(Y_0|D) \\ &= D \cdot E(Y_1) - (1 - D) \cdot E(Y_0) \\ &= D\mu_1 - (1 - D)\mu_0 \end{aligned}$$