Topics in Panel Data Econometrics

Lecture 2

July 2, 2015

Return to the dynamic panel data example Last time, we considered a set of moment restrictions for the autoregressive coefficient of the dynamic panel data model. It turns out that we can collect them in a more systematic manner and this leads to the GMM estimator proposed by Arellano and Bond (1989, ReStud).

It is possible to add more moment restrictions aside from the ones mentioned. A very special moment restriction is one proposed by Ahn and Schmidt (1995, JoE):

$$\mathbb{E}\left[(\alpha_i+\varepsilon_{it})\left(\varepsilon_{it}-\varepsilon_{i,t-1}\right)\right]=0.$$

Try rewriting this moment restriction in terms of the y_{it}'s. Comment on how this moment restriction differ from what we have seen so far.

Assume that $\mathbb{E}\left[y_{i0}|\eta_i\right]=\eta_i/(1-\gamma)$. What additional moment restrictions are implied by this? The restrictions that you derive are related to those considered in Blundell and Bond (1998, JoE).

Another possible solution in the linear panel data setting Assume that $\mathbb{E}\left[\varepsilon|X\right]=0$ and $Var\left[\varepsilon|X\right]=\Omega$, where $\Omega=\sigma_\alpha^2\iota_T\iota_T'+\sigma_\varepsilon^2 I_{nT}$. We can now use GLS to derive an estimator for this random effects model. The GLS estimator has a nice property where it is a matrix-weighted average of the within and between estimator. Will this work for the dynamic panel data setting?

Yet another possible solution in the linear panel data setting Let $\alpha_i=\bar{x}_i'\delta+\nu_i$ where ν_i is uncorrelated with $\bar{x}_i$. Further assume that x is strictly exogenous. Is the OLS estimator consistent in this situation? How does this estimator related to the within estimator? Compare with the situation where we ignore α_i. What does this type of argument remind you of? The argument here is the spirit of Mundlak (1978, Ecta). Chamberlain (1984, HoE) proposes $\alpha_i=x_{i1}\delta_1+\cdots x_{iT}\delta_T+\nu_i$ instead. How will the argument in Mundlak change? Do we have overidentifying restrictions?

Mundlak-Chamberlain device for the dynamic panel data model For $i = 1, \ldots, n$, the structural equations are given by

$$y_{it} = \alpha y_{i,t-1} + \beta x_{it} + \eta_i + \epsilon_{it}. \tag{1}$$

Stacking the above equation across all t, we have

$$\underbrace{\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}}_{y_i} = \alpha \underbrace{\begin{bmatrix} y_{i0} \\ \vdots \\ y_{i,T-1} \end{bmatrix}}_{y_{i,-1}} + \beta \underbrace{\begin{bmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{bmatrix}}_{x_i} + \underbrace{\begin{bmatrix} \eta_i + \epsilon_{i1} \\ \vdots \\ \eta_i + \epsilon_{iT} \end{bmatrix}}_{\xi_i}$$

The device relies on a linear projection of y_{it} on all lagged, present and future x's, i.e.,

$$y_{it} = \pi_{t1} x_{i1} + \cdots + \pi_{tT} x_{iT} + w_{it},$$

where $E(w_{it} x_{is}) = 0$ for all s, t. Stacking the above equation across all t, we have

$$\underbrace{\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}}_{y_i} = \underbrace{\begin{bmatrix} \pi_{11} & \cdots & \pi_{1T} \\ \vdots & \ddots & \vdots \\ \pi_{T1} & \cdots & \pi_{TT} \end{bmatrix}}_{\Pi} \underbrace{\begin{bmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{bmatrix}}_{x_i} + \underbrace{\begin{bmatrix} w_{i1} \\ \vdots \\ w_{iT} \end{bmatrix}}_{w_i}.$$

As a result, $\Pi = E\left(y_i x_i'\right)\left[E\left(x_i x_i'\right)\right]^{-1}$. Multiplying both sides of (1) by x_i' and taking expectations, we have the following moment restrictions:

$$\begin{aligned} E\left(y_i x_i'\right) &= \alpha E\left(y_{i,-1} x_i'\right) + \beta E\left(x_i x_i'\right) + E\left(\xi_i x_i'\right) \\ E\left(y_i x_i'\right)\left[E\left(x_i x_i'\right)\right]^{-1} &= \alpha E\left(y_{i,-1} x_i'\right)\left[E\left(x_i x_i'\right)\right]^{-1} + \beta I + E\left(\xi_i x_i'\right)\left[E\left(x_i x_i'\right)\right]^{-1} \end{aligned} \tag{2}$$

By making assumptions about the elements of the matrix $E\left(\xi_i x_i'\right)$ through the projection of η_i on x_i, we may be able to identify α and β. For instance, let us assume $T = 2$, strict exogeneity of x, and $\eta_i = x_{i1}\delta_1 + x_{i2}\delta_2 + u_i$. As a result, we have

$$\begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} = \beta I + \alpha \begin{bmatrix} E(y_{i0} x_{i1}) & E(y_{i0} x_{i2}) \\ E(y_{i1} x_{i1}) & E(y_{i1} x_{i2}) \end{bmatrix} \left[E\left(x_i x_i'\right)\right]^{-1} + \begin{bmatrix} \delta_1 & \delta_2 \\ \delta_1 & \delta_2 \end{bmatrix} \left[E\left(x_i x_i'\right)\right]^{-1}.$$

The above matrix equation is a system of 4 linear equations in 4 unknowns $(\alpha, \beta, \delta_1, \delta_2)$. In this case, we needed data on the initial values y_{i0}. If we had a linear projection of y_{i0} on x_i, the matrix equation will become a system of 4 nonlinear equations in 6 unknowns (four plus two additional projection coefficients from a linear projection of y_{i0} on x_i). We need additional restrictions to augment the 4 nonlinear equations.

Kotlarski's lemma Suppose that $(Y_1, Y_2) = (A + U_1, A + U_2)$, where (A, U_1, U_2) are mutually independent and $\mathbb{E}(U_1) = 0$. Then the distribution of A is identified.

To prove this, consider the moment generating function of (Y_1, Y_2), i.e.

$$\phi_{Y_1,Y_2}(s_1, s_2) = \mathbb{E}[\exp((s_1 + s_2)A + s_1 U_1 + s_2 U_2)] = \phi_A(s_1 + s_2)\,\phi_{U_1}(s_1)\,\phi_{U_2}(s_2).$$

Take the derivative of this expression with respect to s_1. You will have

$$\frac{\partial \phi_{Y_1,Y_2}(s_1, s_2)}{\partial s_1} = \phi_A'(s_1 + s_2)\,\phi_{U_1}(s_1)\,\phi_{U_2}(s_2) + \phi_A(s_1 + s_2)\,\phi_{U_1}'(s_1)\,\phi_{U_2}(s_2).$$

As a result,

$$\frac{\dfrac{\partial \phi_{Y_1,Y_2}(s, -s)}{\partial s_1}}{\phi_{Y_1,Y_2}(s, -s)} = \frac{\phi_A'(0)\,\phi_{U_1}(s)\,\phi_{U_2}(-s) + \phi_A(0)\,\phi_{U_1}'(s)\,\phi_{U_2}(-s)}{\phi_A(0)\,\phi_{U_1}(s)\,\phi_{U_2}(-s)} = \frac{\phi_A'(0)}{\phi_A(0)} + \frac{\phi_{U_1}'(s)}{\phi_{U_1}(s)}.$$

Note that $\phi_A(0) = 1$, $\phi_A'(0) = \mathbb{E}(A) = \mathbb{E}(Y_1)$. Thus, we can identify $\phi_{U_1}(s)$ and the distribution of U_1. To identify the distribution of A, note that the mgf calculated above can be evaluated as $(s, 0)$.

Some nonparametric stuff How does the lemma relate to the identification of a nonparametric random effects model? Consider $Y_t = m(X_t, \alpha) + \varepsilon_t$. Take $T = 2$. Assume that $(\varepsilon_1, \varepsilon_2) \perp (X_1, X_2, \alpha)$. Further assume that these errors have mean zero. Try reframing the setting in terms of the conditions of Kotlarski's lemma.

A simple example to illustrate approximate solutions to the incidental parameter problem Let y_{it} be iid draws from a $N\left(\alpha_{i0}, \sigma_0^2\right)$ distribution for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. The log-likelihood for one observation is given by

$$\log f\left(y_{it}; \alpha_i, \sigma^2\right) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma^2 - \frac{(y_{it} - \alpha_i)^2}{2\sigma^2}.$$

The MLE satisfies the following first order conditions:

$$\begin{aligned}\sum_i\sum_t\left[-\frac{1}{2\sigma^2}+\frac{(y_{it}-\alpha_i)^2}{2\sigma^4}\right] &= 0,\\ \sum_t\left(\frac{y_{it}-\alpha_i}{\sigma^2}\right) &= 0.\end{aligned}$$

Profiling out the α_i's using the second equation above gives

$$\widehat{\alpha_i}\left(\sigma^2\right)=\frac{1}{T}\sum_t y_{it}. \tag{3}$$

Substituting this to the first equation and solving for θ gives

$$\widehat{\sigma^2}=\frac{1}{nT}\sum_i\sum_t\left(y_{it}-\bar{y}_i\right)^2. \tag{4}$$

Two things to note right away: (3) does not depend on σ^2 and both 3 and (4) are available in closed form. The normality and independence assumptions allows us to conclude that $\bar{y}_i \sim N\left(\alpha_{i0},\sigma_0^2/T\right)$. Thus, $y_{it}-\bar{y}_i\sim N\left(0,\sigma_0^2-\sigma_0^2/T\right)$. As a result,

$$\sum_i\sum_t\left[\frac{\left(y_{it}-\bar{y}_i\right)^2}{\sigma_0^2\left(1-\frac{1}{T}\right)}\right]\sim\chi^2_{nT}.$$

Furthermore,

$$\mathbb{E}\widehat{\sigma^2}=\mathbb{E}\left[\sigma_0^2\left(1-\frac{1}{T}\right)\frac{1}{nT}\sum_i\sum_t\frac{\left(y_{it}-\bar{y}_i\right)^2}{\sigma_0^2\left(1-\frac{1}{T}\right)}\right]=\sigma_0^2\left(1-\frac{1}{T}\right)\frac{1}{nT}\mathbb{E}\left[\chi^2_{nT}\right]=\sigma_0^2\left(1-\frac{1}{T}\right).$$

As a consequence, (4) is not an unbiased estimator of σ^2 in finite samples. If we want to determine if this finite sample bias disappears in large samples, we have to think of the dimensions in which sample sizes grow. If T is fixed and if we let $n\to\infty$, then the finite sample bias does not disappear at all. If we choose to let $T\to\infty$ and let n be fixed, then (4) is a consistent estimator of σ^2. But letting $T\to\infty$ may not be the most appropriate asymptotic setting when we have panels that have short T. In microeconometric panels, it is usually the case that n is large. Therefore, we have to find ways to remove the bias. Note that the bias is of order $\mathcal{O}\left(T^{-1}\right)$.

When both $n,T\to\infty$ at some unspecified rate, $\widehat{\sigma^2}$ will not be consistent for σ_0^2. Although we have consistency when $n,T\to\infty$ at some rate, the limiting distribution of $\widehat{\sigma^2}$ may be incorrectly centered. Consider the limiting distribution of $\sqrt{nT}\left(\widehat{\sigma^2}-\sigma_0^2\right)$. We have

$$\begin{aligned}
\sqrt{nT}\left(\widehat{\sigma^2}-\sigma_0^2\right) &= \sqrt{nT}\left(\frac{1}{nT}\sum_i\sum_t(y_{it}-\bar{y}_i)^2-\sigma_0^2\right)\\
&= \sqrt{nT}\left(\frac{1}{nT}\sum_i\sum_t(y_{it}-\alpha_{i0}+\alpha_{i0}-\bar{y}_i)^2-\sigma_0^2\right)\\
&= \underbrace{\sqrt{nT}\left(\frac{1}{nT}\sum_i\sum_t(y_{it}-\alpha_{i0})^2-\sigma_0^2\right)}_{Z_1}-\underbrace{\sqrt{nT}\left(\frac{1}{n}\sum_i(\bar{y}_i-\alpha_{i0})^2\right)}_{Z_2}
\end{aligned}$$

where $Z_1 \xrightarrow{d} N\left(0,2\sigma_0^4\right)$ as $n,T\to\infty$ and

$$Z_2=\sqrt{\frac{n}{T}}\sigma_0^2\left(\frac{1}{n}\sum_i\left(\frac{\bar{y}_i-\alpha_{i0}}{\sigma_0/\sqrt{T}}\right)^2\right)=\sqrt{\frac{n}{T}}\sigma_0^2\left(\frac{1}{n}\sum_i\chi_1^2\right)\xrightarrow{p}\kappa\sigma_0^2$$

as $n,T\to\infty$ while $n/T\to\kappa^2$.[1] As a result, we have

$$\sqrt{nT}\left(\widehat{\sigma^2}-\sigma_0^2+\frac{\sigma_0^2}{T}\right)\xrightarrow{d}N\left(0,2\sigma_0^4\right)$$

When will the noncentrality parameter disappear? Propose a procedure to remove the noncentrality parameter.

Changing the asymptotic setting We have seen an indication that the bias in the estimator for the parameter of interest in a model with incidental parameters is of order $\mathcal{O}\left(T^{-1}\right)$. We can think of this bias as time series finite sample bias and consider an asymptotic setting where both $n,T\to\infty$ and $n/T\to c$. Allowing this new asymptotic setting will allow us to approximate the asymptotic bias in the estimator and then reduce its impact.

Assume that

$$\begin{aligned}
\operatorname*{plim}_{n\to\infty}\widehat{\theta} &= \theta_T,\\
\lim_{n\to\infty}\theta_T &= \theta_0,\\
\theta_T &= \theta_0+\frac{B}{T}+\mathcal{O}\left(T^{-2}\right),\\
\sqrt{nT}\left(\widehat{\theta}-\theta_T\right) &\to N\left(0,\Omega\right).
\end{aligned}$$

[1]The result depends on sequential asymptotics. Here, we have $T\to\infty$ first then $n\to\infty$.

Then

$$
\begin{aligned}
\sqrt{nT}\left(\widehat{\theta}-\theta_0\right) &= \sqrt{nT}\left(\widehat{\theta}-\theta_T\right)+\sqrt{nT}\left(\theta_T-\theta_0\right) \\
&= \sqrt{nT}\left(\widehat{\theta}-\theta_T\right)+\sqrt{nT}\frac{B}{T}+\sqrt{nT}\,\mathcal{O}\left(T^{-2}\right) \\
&= \sqrt{nT}\left(\widehat{\theta}-\theta_T\right)+\sqrt{\frac{n}{T}}B+\mathcal{O}\left(\sqrt{\frac{n}{T^3}}\right) \\
&\to N\left(B\sqrt{c},\Omega\right). \qquad (5)
\end{aligned}
$$

Note that (5) is not centered at 0. Take note that under this new asymptotic setting, $\widehat{\theta}$ is consistent for θ! But the asymptotic distribution has a noncentrality term $B\sqrt{c}$. How can we get rid of this noncentrality term? Answering this and characterizing this term is essential for the practical purpose of bias reduction and for the theoretical purpose of understanding the source of incidental parameter bias.

What to read The asymptotic bias of the structural parameters in the linear AR(1) dynamic panel data model has been derived by Nickell (1981, Ecta). Bias corrections (roughly) based on these formulas have been considered by Kiviet (1995, JoE) and Bun and Carree (2005, JBES). Monte Carlo simulations for GMM estimators can be found in the articles already cited above. There are more of these simulations scattered in the literature; see Bond and Windmeijer (2005, EctRev), Bun and Kiviet (2006, JoE), Kiviet, Pleus, and Poldermans (2015), and Bun and Sarafidis (2015).

The last decade of panel data econometrics has been about characterizing the incidental parameter bias formally. This research has not really been widely used in practice and has not been featured in modern textbooks. Therefore, the papers are still the best source materials. For static models, see Hahn and Newey (2004, Ecta). For dynamic models, see Hahn and Kuersteiner (2011, ET). For other related papers, check their references for more.

The discussion of the nonparametric random effects model can be found in Evdokimov's dissertation (I think this will eventually appear in Econometrica). The nonparametric random effects model introduced here is very different to the panel data models found in the book by Li and Racine (2008).