Topics in Panel Data Econometrics

Lecture 1

July 2, 2015

The overarching assumption I will assume that we have random sampling over i. Sometimes I might consider random sampling as well over t. Take note that i and t are just indices for the two dimensions by which the data was generated or gathered.

Omitted Variable Bias Assume that the data are generated by the following model:

$$y_i = \beta x_i + \eta z_i + \varepsilon_i,$$

where ε_i is an error term, x_i and z_i are scalar zero-mean regressors, and β and η are parameters. If, for some reason, we compute the OLS estimator from the regression of y_i on x_i, then we have the following:

$$\widehat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \beta + \frac{\sum x_i z_i}{\sum x_i^2}\eta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2} = \beta + \frac{\frac{1}{n}\sum x_i z_i}{\frac{1}{n}\sum x_i^2}\eta + \frac{\frac{1}{n}\sum x_i \varepsilon_i}{\frac{1}{n}\sum x_i^2} \xrightarrow{p} \beta + \frac{\mathbb{E}\left(x_i z_i\right)}{\mathbb{E}\left(x_i^2\right)}\eta + \frac{\mathbb{E}\left(x_i \varepsilon_i\right)}{\mathbb{E}\left(x_i^2\right)}.$$

Think about the conditions under which a consistent estimator for β can be obtained. Can you also give some reasons why we can only perform an OLS regression of y_i on x_i? Suppose I tell you that (a) you get to observe the same i for two periods and (b) z_i does not change over time. How does this small bit of information help in consistently estimating β?

Panel Data Blues, Part 1 Are there any other issues we may have neglected? For instance, what are the properties of the new error term as a result of the new bit of information? How will it affect the estimator for β?

Deriving the LSDV estimator Consider the panel data linear regression model:

$$y_{it} = \alpha_i + x_{it}'\beta + \varepsilon_{it}, \tag{1}$$

where $i = 1, \ldots, n$ and $t = 1, \ldots, T$. The parameters to be estimated are $(\alpha_1, \ldots, \alpha_n, \beta)$ but β is of interest. It would be useful to use matrix algebra here. Let

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, X_i = \begin{bmatrix} x'_{i1} \\ x'_{i2} \\ \vdots \\ x'_{iT} \end{bmatrix}, \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}, \iota_T = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

As a result we can stack (1) across time series observations to get

$$y_i = \iota_T \alpha_i + X_i \beta + \varepsilon_i. \tag{2}$$

Further let

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

We can stack this equation across cross sectional observations to get

$$y = D\alpha + X\beta + \varepsilon, \tag{3}$$

where $D = I_n \otimes \iota_T$. Since the parameter of interest is β, we have to somehow "get rid" of the other regressors in D. We can use ideas from the projection interpretation of least squares. Note that least squares can be thought of as trying to construct an orthogonal decomposition of regressand into a component predicted from the regressors and a residual component. Suppose we wish to construct a linear predictor of y using regressors Z_1 and Z_2. We can estimate regression coefficients $\widehat{\gamma_1}$ and $\widehat{\gamma_2}$ using least squares and we have $y = \widehat{\gamma_1} Z_1 + \widehat{\gamma_2} Z_2 + e$. Here e is not an error term but a residual. Let

$$P_2 = Z_2 \left(Z_2' Z_2\right)^{-1} Z_2', \; M_2 = I - P_2.$$

The matrix M_2 satisfies $M_2 M_2 = M_2$, $M_2 P_2 = 0$, $M_2 Z_2 = 0$, and $M_2 e = e$. As a result, we have $M_2 y = \widehat{\gamma_1} M_2 Z_1 + e$. We now have a closed form for the OLS estimator for $\widehat{\gamma_1}$ which is given by

$$\widehat{\gamma_1} = \left(Z_1' M_2 Z_1\right)^{-1} Z_1' M_2 y. \tag{4}$$

Applying (4) to get the OLS estimator for β alone, we have

$$\widehat{\beta} = \left(X' M_D X\right)^{-1} X' M_D y, \tag{5}$$

where $M_D = I_n \otimes M_{\iota_T}$. Try to explore what M_D does to a vector or a matrix. (5) can be rewritten as

$$\widehat{\beta} = \left(\sum_i \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'\right)^{-1} \sum_i \sum_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i).$$

Think again about the conditions required for consistency of $\widehat{\beta}$.

Panel Data Blues, Part 2 Something always overlooked when conducting panel data analysis is the possibility of low longitudinal variation. This means that although you have additional data on the time-varying regressors, there may not be much variation across time.

Is there an alternative transformation that removes the fixed effects? From the previous discussion, we have $M_D D = 0$ or $M_{\iota_T} \iota_T = 0$. You know what M_D looks like. What does it do to the error vector? Can you thinnk of an alternative matrix that could eliminate the fixed effects? Or what properties should this alternative matrix possess? What are the consequences of selecting these alternatives in terms of their effects on the error vector? If you try working this out, you will be able to recover what Arellano and Bover (1995, JoE) try to show.

Yet another way to derive the LSDV estimator The projection approach earlier is restricted to linear models. To link our simple example to more complicated ones, it helps to consider what we call profiled (or concentrated) objective functions. Consider once again the model in (1). The least squares objective function is given by

$$\min_{(\alpha_1,\ldots,\alpha_n,\beta)} \sum_i \sum_t \left(y_{it} - \alpha_i - x_{it}'\beta\right)^2. \tag{6}$$

Because we envision the possibility of having a large value for n, it is possible to solve a simpler problem if we could find a way to "get rid" of the α_i's. We can optimize step-by-step. The first order conditions for the problem (6) are given by

$$\sum_t \left(y_{it} - \alpha_i - x_{it}'\beta\right) = 0, \tag{7}$$

$$\sum_i \sum_t \left(y_{it} - \alpha_i - x_{it}'\beta\right) x_{it} = 0. \tag{8}$$

Note that we can solve for α_i in terms of β in (7), i.e.

$$\widehat{\alpha_i}(\beta) = \frac{1}{T}\sum_t \left(y_{it} - x_{it}'\beta\right) = \bar{y}_i - \bar{x}_i'\beta. \tag{9}$$

It can be shown that (9) is a solution to the following minimization problem with β fixed:

$$\min_{\alpha_i} \sum_t \left(y_{it} - \alpha_i - x_{it}'\beta\right)^2.$$

Furthermore, once we substitute (9) into (8), we can solve for β. This solution can be obtained from the following optimization problem:

$$\min_{\beta} \sum_i \sum_t \left(y_{it} - \widehat{\alpha_i}(\beta) - x_{it}'\beta\right)^2. \tag{10}$$

We usually call (10) a profiled (or concentrated) objective function. Try to solve for β and show that it coincides with the LSDV estimator.

Nonlinear models Is it a good idea to extend this insight from linear models to the nonlinear case? Consider the following logistic regression model where $y_{i1}, \ldots, y_{iT}$ are independently distributed according to

$$y_{it} \sim Bernoulli(p_{it}),\; p_{it} = \frac{\exp\left(\alpha_i + x_{it}'\beta\right)}{1 + \exp\left(\alpha_i + x_{it}'\beta\right)},$$

where $i = 1, \ldots, n$ and $t = 1, \ldots, T$. What happens if one uses maximum likelihood to estimate β? Explore this when $T = 2$. Assume that x_{it} is fixed in repeated sampling.

Failure of strict exogeneity Return to the original panel data regression model in (1). Let $x_{it} = y_{i,t-1}$. Will the LSDV estimator still be consistent?

Pooled OLS How will the identification conditions change when $\alpha_i = \alpha$?

A unified way of looking at the previous results The previous results can be thought of as manifestations of the so-called incidental parameter problem. Usually, the parameter of interest β is called the structural parameter while the nuisance parameters $\alpha_1, \ldots, \alpha_n$ are called incidental parameters. Notice that β appears in the probability law of every observation while α_i appears only in the probability law of unit i and nowhere else. As a result, we will be unable to estimate α_i consistently when T is finite. The inability to consistently estimate α_i affects the consistency of β in general.

We can formalize this idea by looking at it from a likelihood perspective. Let $f(y_{it};\theta,\alpha_i)$ be the density function of y_{it}. Assume that y_{it} are independent across i and t. The log-likelihood is given by

$$\sum_i \sum_t \log f(y_{it};\theta,\alpha_i).$$

Profiling out the incidental parameters, we obtain the fixed effects estimator $\widehat{\theta}$ which solves the first order conditions

$$\frac{1}{nT}\sum_i \sum_t \frac{\partial}{\partial\theta}\log f(y_{it};\theta,\widehat{\alpha_i}(\theta)) = 0.$$

But note that

$$\begin{aligned}\frac{1}{nT}\sum_i \sum_t \frac{\partial}{\partial\theta}\log f(y_{it};\theta,\widehat{\alpha_i}(\theta)) \approx\ & \frac{1}{nT}\sum_i \sum_t \frac{\partial}{\partial\theta}\log f(y_{it};\theta,\alpha_i)\\ & + \frac{1}{nT}\sum_i \sum_t \frac{\partial^2}{\partial\theta\partial\alpha_i}\log f(y_{it};\theta,\alpha_i)(\widehat{\alpha_i}(\theta)-\alpha_i)\end{aligned}$$

Regardless of the asymptotic scheme, the first term has the following stochastic behavior:

$$\frac{1}{nT}\sum_i \sum_t \frac{\partial}{\partial\theta}\log f(y_{it};\theta,\alpha_i) \xrightarrow{p} \lim_{n\to\infty}\frac{1}{nT}\sum_i \sum_t \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(y_{it};\theta,\alpha_i)\right].$$

With T fixed and $n\to\infty$, the second term does not disappear. This is the incidental parameter problem. Question: Will allowing $T\to\infty$ along with $n\to\infty$ be enough? What if we insist on fixed T?

One way to resolve the incidental parameter problem in the dynamic panel data model
Recall that we have

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \varepsilon_{it}, \tag{11}$$

where $Cov\left(y_{i,t-1},\varepsilon_{it}\right)=0$ (Any sufficient condition for this?). If we take first differences, we have

$$\triangle y_{it} = \rho\triangle y_{i,t-1} + \triangle\varepsilon_{it}.$$

For one thing, note that $Cov\left(\triangle y_{i,t-1},\triangle\varepsilon_{it}\right)\neq 0$. What do you notice about $Cov\left(y_{i,t-2},\triangle\varepsilon_{it}\right)$? Given this information, what can we do in this situation? Is there any other alternative? Congratulations! You were able to recover the arguments by Anderson and Hsiao (1980 JASA; 1982 JOE). Can we generalize this idea to what some call general predetermined regressors, i.e. $\mathbb{E}(x_{is}\varepsilon_{it}) = 0$ for $s \leq t$? What if we only have contemporaneous exogeneity?

One way to resolve the incidental parameter problem in the static logit model It is possible to use a portion of the data to form the log-likelihood. In particular, try deriving $\Pr(y_{i2} = 1|x_{i1}, x_{i2}, y_{i1} + y_{i2} = 1)$ and $\Pr(y_{i1} = 1|x_{i1}, x_{i2}, y_{i1} + y_{i2} = 1)$. Write out the resulting likelihood. Will the maximizer of the resulting likelihood be consistent? Will the argument work for the static probit model? Try extending this to the AR(1) logit model.

What to read Textbook-level treatments are available. Try Hsiao's Analysis of Panel Data, Wooldridge's Econometric Analysis of Cross-Section and Panel Data, or Arellano's Panel Data Econometrics. Surveys on the state of the literature are also available. Try Chamberlain (1984, HoE Vol 2) and Arellano and Honoré (2001, HoE Vol 5). If you want books for specialists but have wide coverage, try The Econometrics of Panel Data: Fundaments and Recent Developments in Theory and Practice by Mátyás and Sevestre (2008) or the more recent 2015 Oxford Handbook of Panel Data (in stores now).

If you want to know more about the history of the incidental parameter problem, check Neyman and Scott (1948, Ecta) and a 50-year retrospective by Lancaster (2000, JoE). They are both highly recommended.

If you want a history of panel data from the perspective of the man who might have started it all, I suggest Nerlove's (2002) Essays in Panel Data Econometrics. One of the chapters include Balestra and Nerlove's (1966, Ecta) attempts to estimate a dynamic panel data model with random effects. Fixed-effects treatments of dynamic panel data models can be traced to the two papers by Anderson and Hsiao (1981, JASA; 1982, JoE).

A starting point for nonlinear models are the papers by Chamberlain (1980, ReStud; 1985, Longitudinal Analysis of Labor Market Data). The 1980 paper also has discussions of the integrated likelihood approach to the incidental parameter problem.

If you want to know more about Monte Carlo simulations, please have a look at Kiviet's paper (2012) entitled Monte Carlo Simulation for Econometricians. A copy is available at `econ.ucsb.edu/~doug/245a/Papers/Monte%20Carlo%20Simulation.pdf`. The paper contains exercises and sample scripts. He also discusses some best practices for Monte Carlo experimental design.