A Course in Applied Econometrics
Lecture 3: Linear Panel Data Models, I

Jeff Wooldridge
IRP Lectures, UW Madison, August 2008

1. Overview of the Basic Model

- Unless stated otherwise, the methods discussed in these slides are for the case with a large cross section and small time series, although some approximations are based on T increasing.
- For a generic i in the population,

$$y_{it} = \eta_t + \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad t = 1,\ldots,T, \tag{1}$$

where η_t is a separate time period intercept (parameters that can be estimated), $\mathbf{x}_{it}$ is a $1 \times K$ vector of explanatory variables, c_i is the time-constant unobserved effect, and the $\{u_{it} : t = 1,\ldots,T\}$ are idiosyncratic errors. We view the c_i as random draws along with the observed variables.

- An attractive assumption (but, of course, not universally applicable) is *contemporaneous exogeneity conditional on* c_i :

$$E(u_{it}|\mathbf{x}_{it}, c_i) = 0, t = 1,\ldots,T. \tag{2}$$

This equation defines $\boldsymbol{\beta}$ in the sense that under (1) and (2),

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \eta_t + \mathbf{x}_{it}\boldsymbol{\beta} + c_i, \tag{3}$$

so the β_j are partial effects holding c_i fixed.

- Unfortunately, $\boldsymbol{\beta}$ is not identified only under (2). If we add the strong assumption $Cov(\mathbf{x}_{it}, c_i) = \mathbf{0}$, then $\boldsymbol{\beta}$ is identified.

- Allow any correlation between $\mathbf{x}_{it}$ and c_i by assuming *strict exogeneity conditional on* c_i :

$$E(u_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i) = 0, t = 1,\ldots,T, \tag{4}$$

which can be expressed as

$$E(y_{it}|\mathbf{x}_i, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i) = \eta_t + \mathbf{x}_{it}\boldsymbol{\beta} + c_i. \tag{5}$$

If $\{\mathbf{x}_{it} : t = 1,\ldots,T\}$ has suitable time variation, $\boldsymbol{\beta}$ can be consistently estimated by fixed effects (FE) or first differencing (FD), or generalized least squares (GLS) or generalized method of moments (GMM) versions of them.

• Make inference fully robust to heteroksedasticity and serial dependence, even if use GLS. With large N and small T, there is little excuse not to compute "cluster" standard errors.

• Violation of strict exogeneity: always if $\mathbf{x}_{it}$ contains lagged dependent variables, but also if changes in u_{it} cause changes in $\mathbf{x}_{i,t+1}$ ("feedback effect").

• *Sequential exogeneity condition on c_i*:

$$E(u_{it}|\mathbf{x}_{i1},\mathbf{x}_{i2},\ldots,\mathbf{x}_{it},c_i) = 0, t = 1,\ldots,T \tag{6}$$

or, maintaining the linear model,

$$E(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{it},c_i) = E(y_{it}|\mathbf{x}_{it},c_i). \tag{7}$$

Allows for lagged dependent variables and other feedback. Generally, $\boldsymbol{\beta}$ is identified under sequential exogeneity. (More later.)

• The key "random effects" assumption is

$$E(c_i|\mathbf{x}_i) = E(c_i). \tag{8}$$

Pooled OLS or any GLS procedure, including the RE estimator, are consistent. Fully robust inference is available for both.

• It is useful to define two *correlated random effects* assumptions. The first just defines a linear projection:

$$L(c_i|\mathbf{x}_i) = \psi + \mathbf{x}_i\boldsymbol{\xi}, \tag{9}$$

Called the *Chamberlain device*, after Chamberlain (1982).

• Mundlak (1978) used a restricted version

$$E(c_i|\mathbf{x}_i) = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi}, \tag{10}$$

where $\bar{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{it}$. Then

$$y_{it} = \eta_t + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i + u_{it}, \tag{11}$$

and we can apple pooled OLS or RE because $E(a_i + u_{it}|\mathbf{x}_i) = 0$. Both equal the FE estimator of $\boldsymbol{\beta}$.

• (10) is in the spirit of approaches for nonlinear models, where often an entire conditional distribution is specified.

• Equation (11) makes it easy to compute a fully robust Hausman test comparing RE and FE. Separate covariates into aggregate time effects, time-constant variables, and variables that change across i and t:

$$y_{it} = \mathbf{g}_t\boldsymbol{\eta} + \mathbf{z}_i\boldsymbol{\gamma} + \mathbf{w}_{it}\boldsymbol{\delta} + c_i + u_{it}. \tag{12}$$

We cannot estimate $\boldsymbol{\gamma}$ by FE, so it is not part of the Hausman test comparing RE and FE. Less clear is that coefficients on the time dummies, $\boldsymbol{\eta}$, cannot be included, either. (RE and FE estimation only with aggregate time effects are identical.) We can only compare $\hat{\boldsymbol{\delta}}_{FE}$ and $\hat{\boldsymbol{\delta}}_{RE}$ (M parameters).

• Convenient test:

$$y_{it} \text{ on } \mathbf{g}_t, \mathbf{z}_i, \mathbf{w}_{it}, \bar{\mathbf{w}}_i,\ t = 1,\ldots,T;\ i = 1,\ldots,N, \tag{13}$$

which makes it clear there are M restrictions to test. Pooled OLS or RE, fully robust!

• Regression (13) can also be used to estimate coefficients on $\mathbf{z}_i$ while allowing correlation between c_i and $\bar{\mathbf{w}}_i$. For these estimated coefficients to be consistent for $\boldsymbol{\gamma}$, we would assume $E(c_i|\mathbf{z}_i, \bar{\mathbf{w}}_i) = E(c_i|\bar{\mathbf{w}}_i) = \psi + \bar{\mathbf{w}}_i\boldsymbol{\xi}$ (or the linear projection version).

• Be cautious using canned procedures: the df are often wrong (the aggregate time variables are included) and the tests are nonrobust.

2. New Insights Into Old Estimators

• Consider an extension of the usual model to allow for unit-specific slopes,

$$y_{it} = c_i + \mathbf{x}_{it}\mathbf{b}_i + u_{it} \tag{14}$$

$$E(u_{it}|\mathbf{x}_i, c_i, \mathbf{b}_i) = 0, t = 1,\ldots,T, \tag{15}$$

where $\mathbf{b}_i$ is $K \times 1$. We act as if $\mathbf{b}_i$ is constant for all i but think c_i might be correlated with $\mathbf{x}_{it}$; we apply usual FE estimator. When does the usual FE estimator consistently estimate the population average effect, $\boldsymbol{\beta} = \mathrm{E}(\mathbf{b}_i)$?

• A sufficient condition for consistency of the FE estimator, along with along with (15) and the usual rank condition, is

$$\mathrm{E}(\mathbf{b}_i|\ddot{\mathbf{x}}_{it}) = \mathrm{E}(\mathbf{b}_i) = \boldsymbol{\beta}, \quad t = 1,\ldots,T \tag{16}$$

where $\ddot{\mathbf{x}}_{it}$ are the time-demeaned covariates. Allows the slopes, $\mathbf{b}_i$, to be correlated with the regressors $\mathbf{x}_{it}$ through permanent components. For example, if $\mathbf{x}_{it} = \mathbf{f}_i + \mathbf{r}_{it}, t = 1,\ldots,T$. Then (16) holds if $\mathrm{E}(\mathbf{b}_i|\mathbf{r}_{i1}, \mathbf{r}_{i2},\ldots,\mathbf{r}_{iT}) = \mathrm{E}(\mathbf{b}_i)$.

• Extends to a more general class of estimators. Write

$$y_{it} = \mathbf{w}_t\mathbf{a}_i + \mathbf{x}_{it}\mathbf{b}_i + u_{it}, \quad t = 1,\ldots,T \tag{17}$$

where $\mathbf{w}_t$ is a set of deterministic functions of time. FE now sweeps away $\mathbf{a}_i$ by netting out $\mathbf{w}_t$ from $\mathbf{x}_{it}$.

• In the random trend model, $\mathbf{w}_t = (1, t)$. If $\mathbf{x}_{it} = \mathbf{f}_i + \mathbf{h}_i t + \mathbf{r}_{it}$, then $\mathbf{b}_i$ can be arbitrarily correlated with $(\mathbf{f}_i, \mathbf{h}_i)$.

• Generally, need $\dim(\mathbf{w}_t) < T$.

- Can apply to models with time-varying factor loads, η_t :

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \eta_t c_i + u_{it}, t = 1,\ldots,T. \tag{18}$$

Sufficient for consistency of FE estimator that ignores the η_t is

$$\mathrm{Cov}(\ddot{\mathbf{x}}_{it}, c_i) = \mathbf{0}, t = 1,\ldots,T. \tag{19}$$

- Now let some elements of $\mathbf{x}_{it}$ be correlated with $\{u_{ir} : r = 1,\ldots,T\}$, but with strictly exogenous instruments (conditional on c_i). Assume

$$\mathrm{E}(u_{it}|\mathbf{z}_i, \mathbf{a}_i, \mathbf{b}_i) = 0. \tag{20}$$

Also, replace (16) with

$$\mathrm{E}(\mathbf{b}_i|\ddot{\mathbf{z}}_{it}) = \mathrm{E}(\mathbf{b}_i) = \boldsymbol{\beta}, \quad t = 1,\ldots,T. \tag{21}$$

Still not enough. A sufficient condition is

$$\mathrm{Cov}(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i|\ddot{\mathbf{z}}_{it}) = \mathrm{Cov}(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i), t = 1,\ldots,T. \tag{22}$$

$\mathrm{Cov}(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i)$, a $K \times K$ matrix, need not be zero, or even constant across time. The *conditional* covariance cannot depend on the time-demeaned instruments. Then, FEIV is consistent for $\boldsymbol{\beta} = E(\mathbf{b}_i)$ provided a full set of time dummies is included.

- Assumption (22) cannot be expected to hold when endogenous elements of $\mathbf{x}_{it}$ are discrete.

3. **Behavior of Estimators without Strict Exogeneity**

- Both the FE and FD estimators are inconsistent (with fixed T, $N \to \infty$) without the strict exogeneity assumption. But inconsistencies (as function of T) can differ.
- If we maintain $E(u_{it}|\mathbf{x}_{it}, c_i) = 0$ and assume $\{(\mathbf{x}_{it}, u_{it}) : t = 1,\ldots,T\}$ is "weakly dependent", can show

$$\mathrm{plim}_{N\to\infty}\, \hat{\boldsymbol{\beta}}_{FE} = \boldsymbol{\beta} + O(T^{-1}) \tag{23}$$

$$\mathrm{plim}_{N\to\infty}\, \hat{\boldsymbol{\beta}}_{FD} = \boldsymbol{\beta} + O(1). \tag{24}$$

- Interestingly, still holds if $\{\mathbf{x}_{it} : t = 1,\ldots,T\}$ has unit roots as long as $\{u_{it}\}$ is I(0) and contemporaneous exogeneity holds.

- Catch: if $\{u_{it}\}$ is I(1) – so that the time series "model" is a spurious regression (y_{it} and $\mathbf{x}_{it}$ are not *cointegrated*), then (23) is no longer true. FD eliminates any unit roots.
- Same conclusions hold for IV versions: FE has bias of order T^{-1} if $\{u_{it}\}$ is weakly dependent.
- Simple test for lack of strict exogeneity in covariates:

$$y_{it} = \eta_t + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{w}_{i,t+1}\boldsymbol{\delta} + c_i + e_{it} \tag{25}$$

Estimate the equation by fixed effects and test $H_0 : \boldsymbol{\delta} = \mathbf{0}$.

• Test contemporaneous endogeneity of a subset of certain regressors.

$$y_{it1} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + \mathbf{y}_{it3}\boldsymbol{\gamma}_1 + c_{i1} + u_{it1}, \tag{26}$$

where, in an FE environment, we want to test $H_0 : E(\mathbf{y}'_{it3}u_{it1}) = \mathbf{0}$.

• Reduced form for $\mathbf{y}_{it3}$:

$$\mathbf{y}_{it3} = \mathbf{z}_{it}\boldsymbol{\Pi}_3 + \mathbf{c}_{i3} + \mathbf{v}_{it3}. \tag{27}$$

• Obtain FE residuals, $\hat{\mathbf{v}}_{it3} = \mathbf{y}_{it3} - \mathbf{z}_{it}\hat{\boldsymbol{\Pi}}_3$ ($\hat{\boldsymbol{\Pi}}_3$ FE estimates). Estimate

$$y_{it1} = \mathbf{z}_{it1}\boldsymbol{\delta}_1 + \mathbf{y}_{it2}\boldsymbol{\alpha}_1 + \mathbf{y}_{it3}\boldsymbol{\gamma}_1 + \hat{\mathbf{v}}_{it3}\boldsymbol{\rho}_1 + error_{it1} \tag{28}$$

by FEIV, using instruments $(\mathbf{z}_{it}, \mathbf{y}_{it3}, \hat{\mathbf{v}}_{it3})$. $\mathbf{y}_{it3}$ exogenous: use (robust) test that $\boldsymbol{\rho}_1 = \mathbf{0}$; need not adjust for the first-step estimation.

4. IV Estimation under Sequential Exogeneity

We now consider IV estimation of the model

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it},\ t = 1,\ldots,T, \tag{29}$$

under sequential exogeneity assumptions; the weakest form is

$Cov(\mathbf{x}_{is}, u_{it}) = 0$, all $s \leq t$.

This leads to simple moment conditions after first differencing:

$$E(\mathbf{x}'_{is}\Delta u_{it}) = \mathbf{0},\ s = 1,\ldots,t-1;\ t = 2,\ldots,T. \tag{30}$$

Therefore, at time t, the available instruments in the FD equation are in the vector $\mathbf{x}_{it}^o \equiv (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{it})$. The matrix of instruments is

$$\mathbf{W}_i = \mathrm{diag}(\mathbf{x}_{i1}^o, \mathbf{x}_{i2}^o, \ldots, \mathbf{x}_{i,T-1}^o), \tag{31}$$

which has $T-1$ rows. Routine to apply GMM estimation.

• Simple strategy: estimate a reduced form for $\Delta\mathbf{x}_{it}$ separately for each t. So, at time t, run the regression $\Delta\mathbf{x}_{it}$ on $\mathbf{x}_{i,t-1}^o$, $i = 1,\ldots,N$, and obtain the fitted values, $\widehat{\Delta\mathbf{x}}_{it}$. Then, estimate the FD equation

$$\Delta y_{it} = \Delta\mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it},\ t = 2,\ldots,T \tag{32}$$

by pooled IV using instruments (not regressors) $\widehat{\Delta\mathbf{x}}_{it}$.

• Can suffer from a weak instrument problem when $\Delta\mathbf{x}_{it}$ has little correlation with $\mathbf{x}_{i,t-1}^o$.

• If we assume

$$E(u_{it}|\mathbf{x}_{it}, y_{i,t-1}\mathbf{x}_{i,t-1}, \ldots, y_{i1}, \mathbf{x}_{i1}, c_i) = 0, \tag{33}$$

many more moment conditions are available. Using linear functions only, for $t = 3,\ldots,T$,

$$E[(\Delta y_{i,t-1} - \Delta\mathbf{x}_{i,t-1}\boldsymbol{\beta})'(y_{it} - \mathbf{x}_{it}\boldsymbol{\beta})] = \mathbf{0}. \tag{34}$$

• Drawback: we often do not want to assume (30). Plus, the conditions in (31) are nonlinear in parameters.

• Arellano and Bover (1995) suggested instead the restrictions

$$Cov(\Delta\mathbf{x}'_{it}, c_i) = 0, \ t = 2, \ldots, T, \tag{35}$$

which imply linear moment conditions in the levels equation,

$$E[\Delta\mathbf{x}'_{it}(y_{it} - \alpha - \mathbf{x}_{it}\boldsymbol{\beta})] = \mathbf{0}, t = 2, \ldots, T. \tag{36}$$

• Simple AR(1) model:

$$y_{it} = \rho y_{i,t-1} + c_i + u_{it}, t = 1, \ldots, T. \tag{37}$$

• Typically, the minimal assumptions imposed are

$$E(y_{is}u_{it}) = 0, \ s = 0, \ldots, t-1, \ t = 1, \ldots, T, \tag{38}$$

so for $t = 2, \ldots, T$,

$$E[y_{is}(\Delta y_{it} - \rho\Delta y_{i,t-1}) = 0, s \leq t-2. \tag{39}$$

Again, can suffer from weak instruments when ρ is close to unity.

Blundell and Bond (1998) showed that if the condition

$$Cov(\Delta y_{i1}, c_i) = Cov(y_{i1} - y_{i0}, c_i) = 0 \tag{40}$$

is added to $E(u_{it}|y_{i,t-1}, \ldots, y_{i0}, c_i) = 0$ then

$$E[\Delta y_{i,t-1}(y_{it} - \alpha - \rho y_{i,t-1})] = 0 \tag{41}$$

which can be added to the usual moment conditions (38). We have two sets of moments linear in the parameters.

• Condition (40) can be intepreted as a restriction on the initial condition, y_{i0}. Write y_{i0} as a deviation from its steady state, $c_i/(1-\rho)$ (obtained for $|\rho| < 1$ by recursive subsitution and then taking the limit), as $y_{i0} = c_i/(1-\rho) + r_{i0}$. Then $(1-\rho)y_{i0} + c_i = (1-\rho)r_{i0}$, and so (40) reduces to

$$Cov(r_{i0}, c_i) = 0. \tag{42}$$

The deviation of y_{i0} from its SS is uncorrelated with the SS.

• Extensions of the AR(1) model, such as

$$y_{it} = \rho y_{i,t-1} + \mathbf{z}_{it}\boldsymbol{\gamma} + c_i + u_{it}, \ \ t = 1, \ldots, T \tag{43}$$

and use FD:

$$\Delta y_{it} = \rho\Delta y_{i,t-1} + \Delta\mathbf{z}_{it}\boldsymbol{\gamma} + \Delta u_{it}, \ \ t = 1, \ldots, T. \tag{44}$$

• Airfare example in notes. The data set can be obtained from the web site for Wooldridge (2002), and is called AIRFARE.RAW:

Dependent Variable:	*lfare*		
	(1)	(2)	(3)
Explanatory Variable	Pooled OLS	Pooled IV	Arellano-Bond
$lfare_{-1}$	$-.126$	$.219$	$.333$
	$(.027)$	$(.062)$	$(.055)$
concen	$.076$	$.126$	$.152$
	$(.053)$	$(.056)$	$(.040)$
N	$1,149$	$1,149$	$1,149$

- Arellano and Alvarez (2003) show that the GMM estimator that accounts for the MA(1) serial correlation in the FD errors has better properties when T and N are both large.