Panel Data Methods: Recent advances
Prerequisites & Notation

Thierry Magnac Toulouse School of Economics

Advanced EEE, Fall 2021

General References

- Arellano, M., 2003, *Panel Data Econometrics*, Cambridge UP, Cambridge.
- Hsiao, C., 2004, *Panel Data*, Cambridge UP.
- Wooldridge, J. M., 2010, *Econometric analysis of cross section and panel data*. MIT press.

Panel data

Panel data of households, countries, firms ... over several time periods and is more elaborate than the pooling of cross sections (see Pseudo-Panels).

Examples:

- PSID : Income, Food consumption
- NLSY, CEX: Labor, income, consumption
- HRS, ELSA, SHARE: Older people, retirement; health
- ECHP, SILC: Labor, income in European countries
- German SEP
- World Bank surveys
- Matched employer/employee data: many countries
- Consumer panels
- Other double dimension: networks in trade

Pros and cons

- More information, more variability
- Better ability to identify effects
  Example 1: Age and cohort (but not time, age and cohort)
  Example 2: 50% of women are found to be working at t.
  Does it mean that:
  1. Half of them work all the time.
  2. All of them work every other year
- Controlling for heterogeneity or correlated effects
- Dynamics

but some limitations:

- Costly design and data collection
- Measurement errors could be more severe
- Selectivity: non response and attrition.

Correlated effects: an example

Consider a linear treatment effect regression

$$y_{it} = x_{it}\beta + \gamma D_{it} + \varepsilon_{it}$$

- in which $D_{it} =$ "treatment" dummy variable (1 or 0)
  Unionization, training last period, regulating the market etc

  The issue is that the correlation between D_{it} and ε_{it} blurs the distinction between unobserved heterogeneity and state dependence
  What panel data brings up is the possibility of differencing an individual effect from which could come the endogeneity bias (principle of difference -in-difference methods).

Linear models

Panel data; double index y_{it} and x_{it} for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

Hypothesis: no missing value so the sample is balanced, y_{it} and x_{it} are observed for any i and t.

A linear panel data model is:

$$\begin{aligned} y_{it} &= x_{it}\beta + \varepsilon_{it} \\ &= x_{it}\beta + \alpha_i + \delta_t + u_{it} \end{aligned}$$

by the variance decomposition of ε_{it}.

Parameters: δ_t are macro shocks, α_i are individual effects, the parameter of interest is β.

Period effects and time dummies

Most of the time, $\{\delta_t\}_{t=1,.,T}$ are treated as parameters and included among the x_{it} as coefficients of period indicators or time dummies $x_{it}^{(\tau)}$ such that $x_{it}^{(\tau)} = \mathbf{1}\{t = \tau\}$ and we write:

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}$$

Rule: Always include time dummies at least in the first run of the model so that you can test for their presence. Always justify their absence.

Alternatively, in linear models, take deviations with respect to aggregate means of all variables (i.e. Frisch Waugh), $z = (y, x, u)$:

$$z_{it} - z_{.t} := z_{it} - \frac{1}{N}\sum_{i=1}^{N} z_{it}$$

Fixed or random individual effects?

Is α_i in:

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}$$

random or fixed. It gives rise to three situations:

- *"Pure" random effects*: α_i is a random variable, independent of u_{it} and its distribution function is independent of x_i
- *Correlated individual effects*: α_i is a random variable, independent of u_{it} and its distribution function conditional on x_i is left free and in particular:

$$E(x_{it}\alpha_i) \neq 0$$

- *Fixed effects:* $\{\alpha_i\}_{i=1,.,n}$ are unknown parameters and the analysis is conditionned upon them.

Remark: The two last cases are very close to each other. In one, the distribution $f(\alpha_i \mid x_i)$ is the nuisance functional, in the other one, the list $\{\alpha_i\}_{i=1,.,n}$ is. Technicalities are somewhat different.

Exogeneity

The relationship between x_{it} and u_{it} in $y_{it} = x_{it}\beta + \alpha_i + u_{it}$ commands what are called exogeneity assumptions:

- strict exogeneity: if $x_i = (x_{i1}, ., x_{iT})$

$$E(u_{it} \mid x_i) = 0.$$

  We also say that the model is *static*

- weak exogeneity: if $x_i^{(t)} = (x_{i1}, ., x_{it})$

$$E(u_{it} \mid x_i^{(t)}) = 0.$$

  We also say that the model is *dynamic.*

Remark: if x_{it} contains lagged endogenous variables i.e. y_{it-1}, x_{it} is weakly exogeneous if u_{it} is independent over time (see below the definition of dynamic completeness).
Remark 2: When the model is dynamic and have a lagged endogenous variable on the right hand side, than another nuisance parameter is the distribution of the initial condition y_{i0} i.e. $f(y_{i0} \mid \alpha_i)$.

Within group & First differences

Linear operators W of dimension $T \times T$ and Δ of dimension $(T-1) \times T$ and mapping $y_i = (y_{i1}, ., y_{iT})$ into Wy_i or Δy_i :

- Within operator: Wy_i of elements $(y_{it} - j_T' y_i / T)$ in which j_T is a vector of dimension T whose elements are equal to 1.
- First differences: Δy_i of elements $(y_{it} - y_{it-1})_{t \geq 2}$.

Forward orthogonal differences

Let $u_i = (u_{i1}, ., u_{iT})$ and denote $u_i^* = Au_i$:

$$u_{it}^* = \frac{T-t}{T-t+1}(u_{it} - \frac{1}{T-t}(u_{it+1} + . + u_{iT})), t < T$$

A is a linear operator of dimension $(T-1) \times T$ such that $A'A = W$ and $AA' = I_{T-1}$.

Properties:

- If $V(u_i) = \sigma^2 I_T$ then $V(u_i^*) = \sigma^2 AV(u_i)A' = \sigma^2 I_{T-1}$.
- Furthermore, $Aj_T = AA'Aj_T = AWj_T = A.0_T = 0$

The Within as an OLS estimator

We have:

$$y_i^* = x_i^* \beta + u_i^*$$

in which the individual effect has been differenced out.
The OLS estimate of this model is:

$$\begin{aligned} \hat{\beta}^* &= (\sum_i x_i^{*\prime} x_i^*)^{-1} \sum_i x_i^{*\prime} y_i^* \\ &= (\sum_i x_i^\prime A^\prime A x_i)^{-1} (\sum_i x_i^\prime A^\prime A y_i) = \hat{\beta}_{Within} \end{aligned}$$

Opertaor A is easier to work with since it preserves homoskedacticity and the absence of autocorrelation

Other estimators

- OLS: Usually biased and inconsistent since $E(x'_{it}\alpha_i) \neq 0$.
- MLE: A pseudo likelihood approach
  We assume $(u_{i1},.,u_{iT})$ are independent normally distributed and of variance σ^2. The likelihood function is:

$$l(\beta,\alpha) \propto \frac{NT}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - x_{it}\beta - \alpha_i)^2$$

  in which $\alpha = (\alpha_1,.,\alpha_N)$.

Proposition: $\hat{\beta}_{MLE}$ is consistent and asymptotically normal and $\hat{\beta}_{MLE} = \hat{\beta}_W$.
However, $\hat{\sigma}^2_{MLE}$ is not consistent when $N \longrightarrow \infty$.
Incidental parameter issue: The number of parameters to be estimated and specifically, α_i, grows with the sample size.

Linear Dynamic models

- Dynamics: lagged and predetermined variables
- Initial conditions and mean-stationarity
- Asymptotic properties : T fixed
- Asymptotic properties in N and T

References

- Alvarez, J., & Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4), 1121-1159.
- Alvarez, J., & Arellano, M. (2004). Robust Likelihood Estimation of Dynamic Panel Data models, working paper.
- Hahn J., & Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica*, 70(4), 1639-1657.

Linear set-up

Panel data; double index y_{it} and x_{it} for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}$$

under weak exogeneity: if $x_i^{(t)} = (x_{i1}, ., x_{it})$

$$E(u_{it} \mid x_i^{(t)}) = 0.$$

- fixed effects : α_i are treated as parameters
- correlated random effects : df of $\alpha_i \mid x_i$ is unrestricted

Dynamic completeness

Suppose that:

$$y_{it} = x_{it}\beta + v_i + u_{it}$$

where $u_{it} = \rho u_{it-1} + \eta_{it}$ is a AutoRegressive process of order 1 and $\mid \rho \mid < 1$. The individual and period shocks (as demand shocks in product demand of a firm) decay over time. We obtain the dynamic model:

$$y_{it} = \rho y_{it-1} + (x_{it} - \alpha x_{it-1})\beta + (1-\rho)v_i + \eta_{it}$$

in which η_{it} is independent over time. The last specification is *dynamically complete*.

Initial conditions

The dynamically complete AR(1) model

$$y_{it} = \rho y_{it-1} + \alpha_i + u_{it},$$

can be rewritten by repetitive replacement:

$$y_{it} = \rho^t y_{i0} + \alpha_i \frac{1-\rho^t}{1-\rho} + \sum_{\tau=0}^{t-1} \rho^\tau u_{it-\tau}$$

so that the distribution of y_{it} depends on the distribution of y_{i0}, of α_i and of u_{it}.
Nuisance parameters are individual effects α_i and initial conditions y_{i0}.

Initialized processes

Use the previous formula as if the dynamic process was in place since $y_{i(-t)}$:

$$y_{i0} = \rho^t y_{i(-t)} + \alpha_i \frac{1-\rho^t}{1-\rho} + \sum_{\tau=0}^{t-1} \rho^\tau u_{i(-\tau)}$$

and take the limit when $t \to -\infty$ to get:

$$y_{i0} = \frac{\alpha_i}{1-\rho} + \sum_{\tau=0}^{\infty} \rho^\tau u_{i(-\tau)}.$$

Mean stationarity

Initialized processes are stationary and they depend only on the distribution of α_i and of u_{it}:

$$y_{it} = \frac{\alpha_i}{1-\rho} + \sum_{\tau=0}^{\infty} \rho^{\tau} u_{i(t-\tau)}.$$

Other initial conditions

- Initial conditions y_{i0} are **exogenous** if

$$E\left(y_{i0}\alpha_i\right) = 0, E\left(y_{i0}u_{it}\right) = 0.$$

  The first assumption is deemed stronger than what is standard and is a special case (to be tested) of:

- Initial conditions y_{i0} are *independently distributed* between individuals, homoskedastic and correlated with individual effects α_i :

$$Ey_{i0}^2 = \sigma_0^2 < +\infty \quad Ey_{i0}\alpha_i = \sigma_{0\alpha}.$$

  and σ_0^2 and $\sigma_{0\alpha}$ are to be estimated.

Within estimation

By definition:

$$y_{it} - y_{i.} = \rho(y_{it-1} - y_{i.}^0) + (u_{it} - u_{i.})$$

where:

$$y_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}\,, \quad y_{i.}^0 = \frac{1}{T}\sum_{t=0}^{T-1} y_{it}\,, \quad u_{i.} = \frac{1}{T}\sum_{t=1}^{T} u_{it}\,.$$

No correlation assumptions imply:

$$E(y_{it-1}u_{it}) = 0$$

but $y_{i.}^0$ is correlated with u_{it} and $u_{i.}$ and y_{it-1} correlated with $u_{i.}$

Nickell bias

The within estimator, using the dynamic model, is biased when T is finite

$$\hat{\rho}_{W,T} \to \rho - \frac{1+\rho}{T} + o(1/T)$$

Example: The true model is not dynamic ($\rho = 0$), the within estimator of the dynamic term converges to the quantity $-\frac{1}{T}$.

WG bias, Arellano, 2003, p86

$T \downarrow$: $\rho \to$	0.05	0.5	0.95
2	-0.52	-0.75	-0.97
3	-0.35	-0.54	-0.73
10	-0.11	-0.16	-0.26
15	-0.07	-0.11	-0.17

Consistent estimators

Based on the first difference operator defined by:

$$\Delta y_{it} = y_{it} - y_{it-1} = \rho\left(y_{it-1} - y_{it-2}\right) + \left(u_{it} - u_{it-1}\right)$$

Remark: Variable y_{it-1} is correlated with u_{it-1} but not variables $y_{it-k}, k > 1$.

- Anderson & Hsiao (1982): Use instruments y_{it-2} and $\left(y_{it-2} - y_{it-3}\right)$ (which are generically valid).
- Arellano and Bond (1991), i.e. the GMM approach: Use all $\{y_{it-k}\}_{k>1}$ as instruments i.e. use all moments:

$$E((\Delta y_{it} - \rho\Delta y_{it-1})y_{it-k}) = 0; k > 1$$

Optimal GMM

Rewrite the $(T-1)(T-2)/2$ moment conditions

$$E(h(y_i;\rho)) = 0$$

and minimize:

$$\left[\sum_{i=1}^{n} h(y_i;\rho)\right]' \Sigma \left[\sum_{i=1}^{n} h(y_i;\rho)\right]$$

Two steps:

- Use a weight matrix $\Sigma = I$ and estimate $\hat{\rho}$.
- Use residuals $h(y_i;\hat{\rho})$, compute its estimated covariance matrix and use its inverse as weight.

Other estimators

Exploit the mean stationarity assumption

$$y_{it} = \frac{\alpha_i}{1-\rho} + \sum_{\tau=0}^{\infty} \rho^{\tau} u_{i(t-\tau)},$$

and formulate additional moments.
See Arellano & Bover (1995): system GMM, Blundell & Bond (1998)

Incidental parameter issue

Arises in a likelihood or pseudo-likelihood framework.
As moment estimation approaches can be framed into pseudo likelihoods, the incidental parameter issue implies bias issues for many estimation methods.

Example: static linear MLE

The MLE estimate is the solution to:

$$\max_{\sigma^2,\beta,\alpha} l(\sigma^2,\beta,\alpha) = \max_{\sigma^2,\beta} \max_{\alpha} l(\sigma^2,\beta,\alpha).$$

The quantity $\max_\alpha l(\sigma^2,\beta,\alpha) = l^p(\sigma^2,\beta)$ is called the profile likelihood or the concentrated likelihood. We denote $\hat{\alpha}_i(\beta)$ the maximizers that satisfy the first order condition:

$$-2\sum_{t=1}^{T}(y_{it} - x_{it}\beta - \hat{\alpha}_i(\beta)) = 0,$$

$$\implies \hat{\alpha}_i(\beta) = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - x_{it}\beta) = y_{i.} - x_{i.}\beta$$

Note they do not depend on σ^2.
Replacing these parameters in the likelihood function delivers the profile likelihood.

Static linear MLE: the profile likelihood

$$l^p(\sigma^2, \beta) =\propto \frac{NT}{2}\log(\sigma^2) + \frac{1}{2\sigma^2}\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - y_{i.} - (x_{it} - x_{i.})\beta)^2$$

Maximizing with respect to β yields:

$$\hat{\beta}_{MLE} = \hat{\beta}_W = \arg\min \sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - y_{i.} - (x_{it} - x_{i.})\beta)^2$$

and the MLE estimate of σ^2 is a function of residuals, $\hat{u}_{it} = y_{it} - x_{it}\hat{\beta}_{MLE}$:

$$\hat{\sigma}^2 = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}(\hat{u}_{it} - \hat{u}_{i.})^2 \underset{N\to\infty}{\overset{P}{\to}} \frac{1}{T}\sum_{t=1}^{T} E\left[(u_{it} - u_{i.})^2\right]$$

Static linear MLE: Neyman and Scott, 1948

We have:

$$E\left[(u_{it} - u_{i.})^2\right] = \frac{T-1}{T}\sigma^2$$

and :

$$\underset{N\to\infty}{plim}\,\hat{\sigma}^2 = \frac{T-1}{T}\sigma^2 \neq \sigma^2$$

Arises in likelihood estimation when the number of parameters to estimate grows with the number of observations.
Specifically, in panel data, the number of individual effects ($\simeq N$) grows with the sample size (NT).
A variety of cases:

- No bias due to incidental parameters: Within estimation with correlated fixed effects
- An exact and computable bias : Autoregressive models of order 1
- An approximate bias correction : non linear models like Probit / Logit
- Identification may fail: Set identification

(Pseudo) likelihood set-up

Panel data; double index y_{it} and x_{it} for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

Denote $z_{it} = (y_{it}, x_{it})$ and its density is $f(z_{it}; \theta_0, \alpha_{i0})$ in which θ_0 is a finite-dimensional parameter and α_{i0} are treated as fixed parameters.

The likelihood function:

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \log f(z_{it}; \theta, \alpha_i)$$

is first maximized with respect to α_i, i.e.

$$\hat{\alpha}_i(\theta) = \arg\max_{\alpha_i} \frac{1}{T} \sum_{t=1}^{T} \log f(z_{it}; \theta, \alpha_i)$$

to get the profile, or concentrated likelihood $l_N^P(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \log f(z_{it}; \theta, \hat{\alpha}_i(\theta))$.

The MLE is $\hat{\theta} = \arg\max_\theta l_N^P(\theta)$.

Parameters of interest

- "Common" parameters indexing the density function e.g. θ.
- Average marginal parameters or average partial effects:

$$\psi = E(m(z_i, \theta, \alpha_i))$$

in which $m(.)$ is a known function, for instance a derivative:

$$m(z_i, \theta, \alpha_i) = \frac{\partial}{\partial x_i} E(y_i \mid x_i; \theta, \alpha_i)$$

so that ψ is an average derivative in which x_i and α_i have been integrated out.
Another example is even more simpler if you want to estimate moments of α_i e.g. $\psi = V(\alpha_i)$.

Incidental parameter bias

Assume that $N \longrightarrow \infty$ and T fixed. Then if we denote $\theta_T = plim_{N \to \infty} \hat{\theta}$, it is often the case that $\theta_T \neq \theta_0$ and the MLE is not consistent. We call θ_T a pseudo-true value.
The issue arises because $\hat{\alpha}_i(\theta)$ cannot be consistent for α_{i0} since T is fixed. There is estimation noise $\hat{\alpha}_i(\theta) - \alpha_{i0}$ which does not disappear when $N \longrightarrow \infty$.
Had we known α_{i0}, we would have defined the (infeasible) MLE estimator as:

$$\tilde{\theta} = \arg\max_{\theta} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \log f(z_{it}; \theta, \alpha_{i0})$$

which is consistent and asymptotically normal.

Example: Neyman and Scott, 1948

Let $z_{it} \sim N(\alpha_{i0}, \sigma_0^2)$ be independent draws. As the MLE for α_{i0} is the empirical mean, $z_{i.} = \frac{1}{T}\sum_{t=1}^{T} z_{it}$, we can define the MLE as:

$$\hat{\sigma}^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(z_{it} - z_{i.})^2$$

from which we can derive:

$$plim_{N\to\infty}\hat{\sigma}^2 = \sigma_0^2 - \frac{\sigma_0^2}{T}$$

We lose degrees of freedom because of the estimation of the mean.
Remark: the bias is computable easily and a bias corrected estimate is

$$\hat{\sigma}^2_{BC} = \frac{T}{T-1}\hat{\sigma}^2.$$

Example 2: Hahn and Kuesteiner, 2002

Let $z_{it} \sim N(\alpha_{i0} + \rho_0 z_{it-1}, \sigma_0^2)$ be independent draws. The MLE for ρ_0 is such that:

$$plim_{N\to\infty}\hat{\rho} = \rho_0 - \frac{1+\rho_0}{T} + O(T^{-2}),$$

in which $\lim_{T\to\infty} \left|O(T^{-2})T^2\right| \in (0, \infty)$
Remark: the first-order bias can be corrected and a bias corrected estimate is

$$\hat{\rho}_{BC} = \frac{T}{T-1}(\hat{\rho} + \frac{1}{T}).$$

since

$$plim_{N\to\infty}\hat{\rho}_{BC} = \frac{T}{T-1}(\rho_0\frac{T-1}{T} + \frac{1}{T} - \frac{1}{T}) + O(T^{-2}) = \rho_0 + O(T^{-2})$$

Extension to "square" asymptotics

Assume that $N \longrightarrow \infty$ and $T \longrightarrow \infty$ and $\lim_{N,T\to\infty} N/T = c \in (0,\infty)$. Then the MLE recovers its consistency:

$$plim_{N\to\infty,T\to\infty}\hat{\theta} = \theta_0$$

is asymptotically normal, but the asymptotic distribution is not centered at 0 (unless T grows faster than N) i.e.

$$\sqrt{NT}(\hat{\theta} - \theta_0) => N(B, V).$$

in which B is a bias term.
This strongly affects the coverage of confidence intervals since they are incorrectly centered.

Summary: Incidental parameters

- Every time a common parameter θ and individual parameters α_i are estimated either jointly or starting by individual parameters, there should be a suspicion of an incidental parameter issue.
- It is only if θ can be estimated through a condition that does not depend on α_i that the issue is absent.
- the within estimator of β is very specific since it belongs to both classes:
  - it solves a moment condition which does not depend on α_i
  - it can be obtained as the profile MLE
- when $T \longrightarrow \infty$, the issue of inconsistency disappears But an issue of asymptotic bias likely arises.

References

Analytical correction and Jackknife:

- Hahn, J., & Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica*, 70(4), 1639-1657.
- Hahn, J., & Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27(6), 1152-1191.
- Dhaene, G., & Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3), 991-1030.
- Fernández-Val, I., & Weidner, M. (2018). Fixed effects estimation of large-t panel data models. *Annual Review of Economics*.