Dynamic Panel Data Models

Laura Magazzini

Sant'Anna School of Advanced Studies

laura.magazzini@santannapisa.it

Dynamic panel data models

Notation & Assumptions

- One of the advantage of panel data is that allows the study of *dynamics*
- For a randomly drawn cross section observation, the basic linear multivariate dynamic panel data model can be written as $(t = 2, 3, ..., T)$

$$y_{it} = \rho y_{it-1} + \mathrm{x}_{it}'\beta + c_i + u_{it}$$

  - As usual u_{it} denotes the idiosyncratic error term
  - c_i captures individual heterogeneity
  - More complicated dynamic structures can be accomodated in this framework (e.g., additional lags of the dependent variables and/or a distributed lag structure for the variables in x)

Why a dynamic model?

$$y_{it} = \rho y_{it-1} + \mathrm{x}_{\mathrm{it}}'\beta + \mathrm{c_i} + \tau_{\mathrm{t}} + \mathrm{u_{it}}$$

- In panel data: correlation of the dependent variable over time
- From a policy perspective, it is of interest to distinguish among the two explanations (Hsiao, 2003; Cameron Trivedi, 2005)
  1. *True state dependence* when y_{it-1} has a causal effect on y_{it}
     - $\rho \neq 0$: y_{it} is determined by y_{it-1}; changes in $\mathrm{x_{it}}$ also have a long-lasting effect
  2. *Unobserved heterogeneity*: correlation can arise even absent a causal relation ($\rho = 0$), driven by unobserved characteristics at the unit level, u_i

- Dynamic models are of interest in a wide range of economic applications including

Partial adjustment framework

- y^* as (unobservable) desired value of unit i at time t
- The adjustment process is defined as

$$y_{it} - y_{it-1} = \theta(y^*_{it} - y_{it-1}) + u_i + e_{it}$$

  with θ the coefficient of adjustment, that is the proportion of the gap between the observed and desired outcome that is closed over the period
- The desired outcome is then specified as a function of $\mathrm{x_{it}}$

$$y^*_{it} = \mathrm{x}'_{\mathrm{it}}\beta + \tau_{\mathrm{t}}$$

- Substitute to obtain the dynamic specification:

$$\begin{aligned} y_{it} - y_{it-1} &= \theta(y^*_{it} - y_{it-1}) + u_i + e_{it} \\ y_{it} - y_{it-1} &= \theta(\mathrm{x}'_{\mathrm{it}}\beta + \tau_{\mathrm{t}} - \mathrm{y_{it-1}}) + \mathrm{u_i} + \mathrm{e_{it}} \\ y_{it} &= (1-\theta)y_{it-1} + \mathrm{x}'_{\mathrm{it}}(\theta\,\beta) + \theta\,\tau_{\mathrm{t}} + \mathrm{u_i} + \mathrm{e_{it}} \\ y_{it} &= \rho y_{it-1} + \mathrm{x}'_{\mathrm{it}}\delta + \mathrm{d_t} + \mathrm{u_i} + \mathrm{e_{it}} \end{aligned}$$

Correlation in the error term

- One of the most famous application of dynamic panel data considered a Cobb-Douglas production function with autocorrelated productivity shocks (Blundell & Bond, 2000)

$$y_{it} = \beta_n n_{it} + \beta_k k_{it} + \gamma_t + (\eta_i + v_{it} + m_{it})$$

  - ▷ $v_{it} = \rho r_{it-1} + e_{it}$ an AR (productivity) shock (with $\rho < 1$ and e_{it} homosk. and uncorrelated)
  - ▷ m_{it} reflecting measurement error (homosk. and uncorrelated)
- The model can be written as a dynamic specification by considering $y_{it} - \rho y_{it-1}$

Roadmap

- Why “standard” methods (OLS, FE) fail – endogeneity issue
- Estimation by GMM: the idea of “internal” instrument
    - ▷ Anderson & Hsiao (1981)
    - ▷ Arellano & Bond (1991)
    - ▷ Ahn & Schmidt (1995)
    - ▷ Blundell & Bond (1998)
- Models with x
- Testing GMM assumptions
- Problems with GMM estimation
- Other estimation framework