Advanced Dynamic Panel Data Methods

Major Characteristics of Micro Panel Relationships

Sebastian Kripfganz

University of Exeter Business School, Department of Economics, Exeter, UK
www.kripfganz.de @Kripfganz

Universidad de Salamanca

July 17–19, 2023

Strict Exogeneity

- As seen before, the linear-in-coefficients "error components model"

$$y_{it} = \mathbf{x}_{it}'\beta + \alpha_i + \varepsilon_{it}$$

  can accommodate various forms of heterogeneity, including time-specific effects and (restricted) heterogeneous slope coefficients, by including time dummies and interaction terms in $\mathbf{x}_{it}$. The same applies to nonlinear functions (polynomials) of $\mathbf{x}_{it}$.
- The FE and RE estimators rely on the strong assumption that all regressors are strictly exogenous, such that $E[\mathbf{x}_{it}\varepsilon_{is}] = 0$ for all s, t.
  - This assumption can be reasonable – e.g., in a controlled laboratory experiment – but often is difficult to justify.

Weak Exogeneity

- Strict exogeneity precludes dynamic feedback from the outcome variable to the explanatory variables – i.e., reactions of the regressors to past outcomes or expectations about future outcomes. If the outcome variable determines future realizations of the regressors, the latter are predetermined (or weakly exogenous):

$$E[\mathbf{x}_{it}\varepsilon_{is}] \begin{cases} = 0 & , s \geq t \\ \neq 0 & , s < t \end{cases}$$

  - Importantly, a lagged dependent variable is predetermined by construction, assuming that the idiosyncratic error component ε_{it} is serially uncorrelated.

Endogeneity

- If the explanatory variables are simultaneously determined with the outcome variable, they are endogenous. This includes instantaneous feedback. In general, this also may or may not encompass dynamic feedback. For the purpose of this course, an endogenous regressor satisfies

$$E[\mathbf{x}_{it}\varepsilon_{is}] \begin{cases} = 0 & , s > t \\ \neq 0 & , s \leq t \end{cases}$$

  - A lagged dependent variable becomes endogenous if the idiosyncratic error component ε_{it} is serially correlated.
- Strictly speaking, any regressor that is correlated with the idiosyncratic error component α_i would be endogenous. Here, we use the term endogeneity in the narrower sense with respect to ε_{it} only.

Unit-Specific Error Component

- Often, a debate arises whether to treat the unit-specific intercept α_i as deterministic or random variable.
  - From the perspective of a single unit i, α_i is a constant (fixed) population parameter.
  - From the panel data perspective, there is a collection of unobserved parameters α_i, $i = 1, 2, \dots, N$. As we increase the cross-sectional sample size N, more parameters are added. We can think about this as drawing units (and therefore unit-specific intercepts α_i) randomly from some distribution, $\alpha_i \sim (\mu_\alpha, \sigma_\alpha^2)$, which justifies treating α_i as a random variable. With repeated random sampling, the true (but unknown) intercepts α_i will be different in each sample.
- Similarly, the regressors $\mathbf{x}_{it}$ are usually treated as random variables. In many cases, it is reasonable to assume that $\mathbf{x}_{it}$ and α_i are correlated, $E[\mathbf{x}_{it}\alpha_i] \neq \mathbf{0}$, as the latter often represent relevant omitted variables.

Idiosyncratic Error Component

- Estimator properties crucially depend on the assumptions about the idiosyncratic error component ε_{it}.
- Baseline assumptions:
  - $E[\varepsilon_{it}] = 0$ (without loss of generality because any nonzero mean will be captured by α_i or an overall regression intercept)
  - $Var(\epsilon_{it}) = E[\varepsilon_{it}^2] = \sigma_\varepsilon^2$ (homoskedasticity)
  - $Cov(\varepsilon_{it}, \varepsilon_{is}) = E[\varepsilon_{it}\varepsilon_{is}] = 0$ for all $s \neq t$ (no serial correlation)
  - $E[\varepsilon_{it}\varepsilon_{js}] = 0$ for all $j \neq i$ and all s, t (no cross-sectional dependence)
  - $Cov(\alpha_i, \varepsilon_{it}) = E[\alpha_i\varepsilon_{it}] = 0$ (rather inconsequential)
- $\varepsilon_{it} \overset{iid}{\sim} (0, \sigma_\varepsilon^2)$ would imply those assumptions, but independence is stronger than needed.
- The homoskedasticity assumption is occasionally useful, but can often easily be relaxed.

Idiosyncratic Error Component

- Absence of serial correlation in ε_{it} is a crucial assumption for dynamic panel data models (or, more generally, models with dynamic feedback).
  - The error term is not intrinsically serially correlated. It is implicitly defined by the model specification.
  - Ideally, the choice of $\mathbf{x}_{it}$ should yield a dynamically complete model – i.e., any dynamic feedback is accounted for by the variables in the model, which often requires including one or more lags of the dependent and/or independent variables.
  - In other words, evidence of serial correlation in ε_{it} indicates that important variables capturing the model dynamics are omitted.
  - Testing this assumption will be an important part of any model specification check.
- Cross-sectional dependence should probably receive more attention (here and in the microeconometric literature), but we will ignore it in this course.

Interim Conclusion

- The classic least-squares estimators require assumptions which are typically too strong for most microeconometric applications (where T is small). In particular, they all rule out the presence of a lagged dependent variable.
  - The pooled OLS estimator makes the mild assumption that the regressors are contemporaneously exogenous with respect to ε_{it} – i.e., there are no instantaneous feedbacks – but it requires that they are not co-related to the unit-specific intercepts.
  - The RE estimator requires the stronger assumption of strict exogeneity – i.e., no instantaneous or dynamic feedback.
  - The FE estimator is agnostic about α_i, but still requires strict exogeneity with respect to ε_{it}.
  - The first-difference estimator is also agnostic about α_i, but only requires contemporaneous exogeneity of the first differences, which is weaker than strict exogeneity.