Chapter 2
Outline: Panel Data

Badi H. Baltagi, László Mátyás and Alain Pirotte

1. Introduction

- The interest in panel data is partly related to developments in economic theory, in computer technology and software programs, the progress in the elaboration and implementation of appropriate statistical and econometric methods, and the availability of panel data sets.
- In dealing with cross-sections and time series, panel econometrics must obviously solve the problems of each of these data sets, but it encounters a significant variety of particular statistical phenomena related to multi-dimensionality.
- 1990s: the predominance of micro-panels (large N, small T) is declining. Macro-panels introduce new asymptotic theory to complement that of micro-panels. Emergence of the heterogeneous and "time series" (large T) panels literatures.
- Although at its birth, the panel data econometrics drew its research topics from other areas of econometrics, it has always integrated them by providing solutions to specific problems. Thus, it has its own history and it became one of the most productive areas of quantitative economics since the 1960s. The 50th anniversary of the first 1977 International Panel Data Conference,[1] is a "living proof" of the success of this branch of econometrics.

2. Early Statistical Foundations of Panel Data

Topics:

Badi H. Baltagi ✉
Syracuse University, New York, USA, e-mail: bbaltagi@syr.edu

László Mátyás
Central European University, Budapest, Hungary and Vienna, Austria, e-mail: matyas@ceu.edu

Alain Pirotte
Paris-Panthéon-Assas University, Paris, France, e-mail: alain.pirotte@assas-universite.fr

[1] In 2027 it will be the 32nd edition, as it was only in 2004 that this conference adopted an annual schedule.

- Airy (1861), Chauvenet (1863): random effects (RE) models in the analysis of astronomical data.
- Fisher (1918, 1925): fixed effects (FE) ANOVA (Fisher's work mainly on agronomic research)[2].
- Daniels (1939), Eisenhart (1947): clarification of the distinction between fixed-effects and random-effects.
- Eisenhart (1947), Hildreth (1950), Scheffé (1956, 1959): fundamental contributions for the emergence of panel data econometrics.

Lessons:

- New ground by distinguishing between different sources of error (Airy).
- Foundations of RE and FE models.
- Emergence of a clear distinction between FE and RE and what will become the foundation for the development of panel data econometrics.

3. Fixed vs Random Effects and Related Tests

Topics:

- FE: Kuh (1959), Hoch (1955, 1957, 1958, 1962), Mundlak (1961, 1963, 1964).
- RE: Balestra and Nerlove (1966), Wallace and Hussain (1969), Amemiya (1971), Nerlove (1971), Swamy and Arora (1972).
- Tests: Mundlak (1978), Hausman (1978), Chamberlain (1982).
- Correlated individual effects and time-invariant regressors: Hausman and Taylor (1981), Amemiya and MaCurdy (1986), Breusch, Mizon and Schmidt (1989).

Lessons:

- Econometricians appropriated these two models for their own needs, i.e., to represent individual unobservable characteristics, time-persistent, that influence individual behaviors.
- The main reasons why the random effects models have not aged well.
- Then, the concepts of fixed and random effects models have long been key elements in the development of linear and non-linear models for multi-dimensional panel data.

4. Dynamic Panels

Topics:

- VI: Balestra and Nerlove (1966), Anderson and Hsiao (1981, 1982).

[2] Nerlove (2014, p. 6) said: *"Fisher's work established variance components, or random-effects models, as a method of allowing individual heterogeneity to play a role in the analysis of biometric data. But he also invented ANOVA tables, that is, fixed-effects models, for allowing for individual heterogeneity in agronomic data."*, and on page 7, *"Although Fisher may have been perfectly clear in his own mind what the distinction between fixed effects and random effects was, by eschewing the use of the expectation operator and working from a standard ANOVA table but giving it a population interpretation appropriate to a random-effects model, he greatly muddied the waters for those who followed."*

- GMM: Arellano and Bond (1991), Arellano and Bover (1995), Ahn and Schmidt (1995).
- SYS-GMM: Blundell and Bond (1998).

Lessons:

- For $N \longrightarrow \infty$ and T finite, inconsistency of the usual least squares methods (OLS, FGLS, LSDV (Within), etc.).
- Poor performance of the standard VI and GMM estimators – Weak instruments problem.
- To what extent has big data changed the way dynamics is dealt with in this context.

5. Non-Stationary Panels

Topics:

- Asymptotic theory: Phillips and Moon (1999a, 1999b, 2000).
- Unit roots: Levin and Lin (1992) (Levin, Lin and Chu (2002)), Im, Pesaran and Shin (2003).
- Cointegration tests: Pedroni (1997, 1999), Kao (1999), McCoskey and Kao (1998), Larsson, Lyhagen and Löthgren (1998).
- FM estimators: Phillips and Moon (1999), Choi (1999), Kao and Chiang (2000), Pedroni (2000).

Lessons:

- Macro-panels (relatively large N, large T), emergence of new and more complicated asymptotic theory that was for micro-panels.
- The assumption of cross-sectional independence is inappropriate in many empirical works.
- Initially, the panel test outcomes are often difficult to interpret if the null of unit root or cointegrated is rejected.
- Presence of different forms of cross-sectional dependence impacts the methods.
- As in the case of time series this approach aged badly, very few empirical applications, if any, rely on it lately.

6. Large Heterogeneous Panels

Topics:

- Testing homogeneity slopes: Zellner (1962), Swamy (1970), Pesaran, Smith and Im (1996), Phillips and Sul (2003), Pesaran and Yamagata (2008), Blomquist and Westerlund (2013).
- Estimation methods: Swamy (1970), Pesaran and Smith (1995), Pesaran, Shin and Smith (1999), Pesaran and Zhao (1999), Hsiao, Pesaran and Tahmiscioglu (1999), Pesaran (2006).

Lessons:

- If N is small and T large, the standard approach is to treat the equations from the different cross-section units as a system of seemingly unrelated regression equations (SURE).
- If T is large, homogenous estimators no longer hold. Heterogenous estimators should be used.
- Random coefficients models and factor structures as a framework benchmark.

7. Cross-Sectional Dependence in Panels

Topics:

- Spatial and/or factor cross-dependence: Pesaran (2004), Ng (2006), Pesaran, Ullah and Yamagata (2008), Baltagi, Feng and Kao (2012), Pesaran (2015).

Lessons:

- Distinction between weak and strong cross-dependence.
- Cross-sectional dependence may arise from unobserved global shocks, local interactions, or can be seen as pure idiosyncratic correlation.
- Serious consequences of ignoring factors behind cross-correlations which are themselves correlated with the regressors. This is a challenge for empirical applications.

8. Limited Dependent, Discrete and Qualitative Panels

Topics:

- Measurements are often qualitative, discrete, or limited by truncation or censoring. In this context, the models are highly non-linear, which implies that specific methods/approaches should be used.
- Prolific literature: Heckman and Willis (1975), Manski (1975, 1977, 1987), Heckman (1978, 1981a, 1981b, 1991), Heckman and Singer (1985), Chamberlain (1980), Honoré (1992, 1993), Kyriazidou (1997)...

Lessons:

- The incidental parameters problem (Neyman and Scott 1948, Lancaster, 2000) is not an easy problem in the context of panel data, especially in the context of non-linear fixed effects models.
- More complex numerical methods are needed, often with properties not fully understood.

9. Multi-dimensional Panels

Topics:

- Early work in trade and other flows.
- Fixed effects models and many orthogonal projections.
- Complex dynamic specifications
- The fixed effects approach takes over and becomes the main empirical tool.

- Exponential use of multi-dimensional data set.

Lessons and questions:

- The formulation of individual, time and cross heterogeneity, and dynamics became much more complex, but simplicity and interpretability matters.
- Sometimes these data sets cannot be considered anymore in the standard statistical way as samples from a population; rather, they should be viewed as complete snapshots of a given process, for a period of time.
- This opens the question about how and to what extent the usual statistical tools (e.g. hypothesis testing, etc.) can be applied and how the results should be interpreted. New ways of looking at 'old' tools are needed.
- Panel data econometrics is steadily merging with other newer areas in the field like, for example, networks, etc. What may be the implications of this radical change?

10. General Lessons from the History of Panel Data Econometrics

- Panel data econometrics is at the crossroads of time series and cross-section econometrics.
- An important strand of the literature found its original motivation by controlling unobserved individual heterogeneity.
- Better understanding of economic behaviors (taking into account the observable and unobservable heterogeneity of economic agents in the analysis of their behavior).
- Panel data econometrics has enriched the set of possible identification arrangements.
- Still going strong as it has been able to interact with and integrate into the new technical and data developments.

11. Looking Ahead

- Challenge: Big data and the many different forms of panel data sets that are emerging.
- Integration and interaction with artificial intelligence, machine learning techniques and networks
- Multidimensional asymptotic theory and inference with high-dimensional data.
- Disaggregate to aggregate predictions (policy relevance).

12. Conclusions

Miscellaneous other topics that may be given some scope in the chapter

Topics:

- Incomplete panels: Biorn (1981), Deaton (1985), Verbeek and Nijman (1996), Nijman, Verbeek and van Soest (1991). Consequences of incomplete panels: Gronau (1974), Heckman (1976, 1979), Hausman and Wise (1979).
- Pseudo-Panels: Heckman and Robb (1985), Deaton (1985), Verbeek (1986), Moffitt (1993), with spatial dimension Anselin (1998, 2000).

- Systems of equations: Baltagi (1981), Balestra and Krishnakumar (1987), Krishnakumar (1988), Pagan (1979).