# Chapter 2
# Panel Data

Badi H. Baltagi, László Mátyás and Alain Pirotte

**Abstract** The chapter traces the development of panel data econometrics from its origins in the 19th century to its remarkable expansion in contemporary empirical research. It reviews the main models and methods that form the foundation of the field and examines why some approaches have endured more successfully than others. Moving beyond the traditional linear model where individual effects captured behavioural heterogeneity, the econometrics of panel data has advanced substantially. This progress reflects developments in economic theory, the growing availability of panel datasets—facilitated in part by the Internet—and remarkable improvements in computing power and statistical software. These latter two factors, in particular, have enabled applied economists to use sophisticated econometric tools with relative ease and have encouraged the wider adoption of advanced models and methods. This top-down perspective also provides a natural introduction to several related topics explored in other chapters of these volumes.

## 2.1 Introduction

Over the past few decades, panel data econometrics has become an important workhorse of empirical research. This shift has been driven by the rapidly expanding availability of panel datasets across diverse fields of economics. The ability to link multiple independent data sources has made it possible to track 'individuals' over time, giving rise to a variety of new and increasingly rich panel data structures. This surge in data availability has been reinforced by advances in modelling techniques,

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

as well as rapid progress in computing technology and statistical software. Because panel econometrics must handle both cross-sectional and time-series dimensions, it inherits the challenges of each while also confronting additional statistical issues arising from multi-dimensionality. Although the field initially drew many of its research questions from other branches of econometrics, it consistently adapted and extended these ideas to address its own distinctive problems. As a result, panel data econometrics has developed a clear identity and has become one of the most dynamic and productive areas of quantitative economics since the 1960s. The many editions of Baltagi's, Hsiao's and Mátyás and Sevestre's volumes, and the widely used textbook of Wooldridge, among others, eloquently illustrate this process. This chapter offers a bird's-eye view of this trajectory, highlighting both the approaches that have aged well and those that have not. Several other chapters in this volume provide concrete illustrations of how the use of panel data has expanded across numerous fields and directions (see, for example, Chapters 4, 6, 7, 10, 18, and 20). The chapter is structured as follows. Section 2.2 deals with the early foundations. Section 2.3 contrasts the two classic panel-data model specifications—random effects (RE) and fixed effects (FE)—and reports that, in practice, fixed effects models have become more dominant. Section 2.4 presents dynamic models, while Section 2.5 extends them to nonstationary processes. Large heterogeneous panels and cross sectional dependence are discussed in Sections 2.6 and 2.7, qualitative variables, multi-dimensional panels and some further miscellaneous topics in Sections 2.8, 2.9, and 2.10. Finally, Section 2.11 concludes.

## 2.2 Early Statistical Foundations

It is well known that the theory of least squares was developed independently by Legendre (1805) and Gauss (1809) in their astronomical problems books.[1] The study of these astronomical problems gave rise to the analysis of variance approach (fixed effects models). It is not so well known that astronomers also formulated variance components models (random effects models) particularly through the work of Airy (1861) and Chauvenet (1863). Nevertheless, the concept and methods of these two models were developed extensively in Ronald Aylmer Fisher (Fisher, 1925, especially in Chapters 7 and 8). He introduced variance components (in effect the random effects model) as a method to account for individual heterogeneity in the analysis of biometric data. He also established ANOVA tables (the fixed effects model) to account for heterogeneity in agronomic data. Nevertheless, Fisher did not clearly distinguish between them. To sum up, 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*

[1] Section 2.2 draws heavily on the paper by the Dupont-Kieffer and Pirotte (2011).

*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."* It was Daniels (1939) who clearly established the difference between the two approaches, although the relevance of that distinction for experimental versus non-experimental data was put forward for applied research by Eisenhart (1947).

Daniels (1939) and Eisenhart (1947) truly initiated the debate on the nature of the error terms and on variance analysis. More broadly, the issue was then how to deal with latent variables. Some econometricians relied on this distinction with Hildreth (1950), Hoch in the mid-1950s, and then with Mundlak (1961), and Balestra and Nerlove (1966). They were the heirs of Fisher's (1925) contributions, but they clarified the debate on the use of fixed- versus random-effects models according to the nature of the data. They 'borrowed' this approach, but expanded it by using it not merely as a simple statistical improvement to econometric methodology. The difference between Hildreth/Hoch and Mundlak/Balestra-Nerlove is the purpose behind the use of the fixed- and random-effects models. Indeed, this reference to Fisher's methodology was clearly for all a way to reintroduce concerns on unobservable variables. But in the case of Hildreth and Hock, the concern was still methodological; in the case of Mundlak and Balestra-Nerlove, the concern was more about the economic nature of the latent variables and the relationship among them. Mundlak and Balestra-Nerlove sought in the error term information about economic phenomena, in particular in relation to the individual level of heterogeneity.

The contributions by Eisenhart (1947), Hildreth (1950) and Scheffé (1956, 1959) were always mentioned by the pioneers of panel econometrics and thus constitute fundamental reference points for the emergence of panel data econometrics. Eisenhart (1947) started his introduction to a special issue of Biometrics on variance analysis by acknowledging the contribution of Fisher to analyzing idiosyncrasy among specific phenomena as agricultural or biological. He then questioned the relevance of the use of this particular methodology for variance analysis. He writes (1947, p. 3-5):

*"Turning now to my assignment, I am obliged at the outset to draw attention to the fact that analysis of variance can be, and is, used to provide solutions to problems of two fundamentally different types.*

*These two distinct classes of problems:*

*—Class I: Detection and Estimation of Fixed (Constant) Relations Among the Means of Sub-Sets of the Universe of Objects Concerned. This class includes all the usual problems of estimating, and testing to determine whether to infer the existence of, true differences among 'treatment' means, among 'variety' means, and, under certain conditions, among 'place' means... the analysis-of-variance are the least-squares solutions. The cardinal contribution of analysis of variance to the actual procedure is the analysis-of-variance table devised by R. A. Fisher...*

*—Class II: Detection and Estimation of Components of (Random) Variation Associated with a Composite Population. This class includes all problems of estimating, and testing to determine whether to infer the existence of components of variance ascribable to random deviation of the characteristics of individuals of a particular*

*generic type from the mean values of these characteristics in the 'population' of all individuals of that generic type, etc. In a sense, this is the true analysis of variance. . . Problems of this class received considerably less attention in the literature of analysis of variance than have problems of Class I. . .*

*. . . the mathematical models appropriate to problems of Class I differ from the mathematical models appropriate to problems of Class II and, consequently, so do the questions to be answered by the data.*

Eisenhart took over the debate on the treatment of experimental versus non-experimental data that Fisher's (1925) publication opened. *"The answer [whether the parameters of interest specify fixed relations or components of random variation] depends in part, however, upon how the observations were obtained; on the extent to which the experimental procedure employed sampled the respective variables at random. This generally provides the clue."* (Eisenhart, 1947, p. 19).

In the econometric literature, the idea of distinguishing between fixed and random effects was adopted by Hildreth (1950). Hildreth's contribution should be seen as an early attempt to identify heterogeneity among individuals and periods, including constant vectors as fixed variables to enlighten the presence of 'latent' individual effects and then to introduce and to combine fixed-effect and random-effect modeling. His purpose was clearly to tackle heterogeneity, mainly the possible omitted variables, that is, issues to consider when dealing with unobserved phenomena: *"It may be believed that there are unobserved individual characteristics which cause individuals to act differently and which are persistent over time. There may be observed influences that affect individuals in pretty much the same way but change over time"* (Hildreth, 1950, p. 2).

Indeed, the question of heterogeneity had already been raised by Marschak and Andrews (1944) on estimating a production function. Later, Hildreth (1949, 1950) showed the path to be taken and brought back identification issues to the forefront of the debate. Hoch (1955, 1957, 1958, 1962) attempted to address the estimation and identification issues raised by Marschak and Andrews (1944) by combining time-series and cross-section data to capture the heterogeneity among economic determinants. This is the beginning of numerous papers founded on the same idea, especially the seminal papers by Mundlak (1961, 1963, 1964) and Balestra and Nerlove (1966).

## 2.3 Fixed vs Random Effects and the Hausman test

Despite the classic paper by Balestra and Nerlove (1966) and many theoretical econometrics papers over the past six decades, the *random effects* model has not aged well in empirical panel data econometrics. The standard *Hausman* (1978) *test* has been one of the durable tests in this literature and it has been used more often to reject than not reject the efficiency of the random effects model based on a contrast between the *fixed effects* and *random effects* estimators. This has been taken by researchers as an acceptance of the fixed effects estimator despite several warnings that Hausman's test is a specification test which tests the null of no correlation between the disturbances and the regressors and does not shed any light regarding the alternative which could be due to unknown misspecification. If the null is rejected, it signals misspecification. It does not accept the fixed effects estimator as thousands of empirical studies claim, see for example Glick and Rose (2002) for an application to a Gravity model examining whether common currency enhances trade. The authors reject random effects based on a Hausman test and settle on fixed effects. Amini, Delgado, Henderson and Parmeter (2012) and Baltagi (2024) clearly emphasized that endogeneity, dynamics, functional form are just some examples of misspecification that may lead to rejection of the null using the Hausman test. The Hausman's (1978) specification test has been so influential in econometrics and generated a lot of research in economics. Google scholar has more than 29,000 citations as of October 2025.[2] This test has aged well.

Going back to *fixed versus random* estimators, if there is a popularity contest between these two estimators, fixed effects won it. It is one of the most used estimators in empirical panel data. Its advantages is that it wipes out the possible correlation between the individual effects and the regressors. This is done by eliminating the individual time-invariant effects. Its disadvantages is that it wipes out any time-invariant regressor in the process and may 'throw the baby with the bath water' as many of these time invariant variables are important for policy decisions. Examples are the wage gap between males and females, blacks and non-blacks and the effect of education on earnings in a Mincer wage equation, see Cornwell and Rupert (1988). Also, the effect of common language, distance, common colonizer on trade, see Glick and Rose (2002). Fixed effects is seen as a non-restrictive estimator which assumes that all the regressors are correlated with the individual effects, *a correlated random effects (CRE)* model, see Mundlak (1978), whereas random effects is seen to be restrictive, assuming that non of the regressors is correlated with the individual effects. Mundlak's classic paper showed that adding a reduced form equation relating the individual effects to all the time-averaged regressors plus a stochastic random disturbance that is time-invariant, produces the fixed effects estimator when the random effects estimator is applied to this augmented regression. The test for the non-significance of the averaged regressors coefficients yields a Hausman test based on the contrast between the fixed effects and between estimators.[3] Not rejecting this null yields to non-rejection of the random effects model. Rejecting the null is an indication of correlated random effects rendering the rejection of the random effects estimator. Chamberlain (1982) also related the individual effects to all the regressors, but unlike Mundlak, he did that at every point in time. This introduces $Tk$ parameters in the reduced form of the individual effects where $T$ is the time dimension of the panel and $k$ is the number of regressors. Chamberlain showed that this produces testable restrictions for the fixed effects estimator, but unfortunately, testing these restrictions are not commonly used when settling on the fixed effects estimator. Chamberlain suggested minimum chi-squared estimation in which the

[2] All Google citations mentioned in this chapter are dated October 2025 unless otherwise indicated.

[3] It is interesting to note that Mundlak's (1978) paper is in the same journal *Econometrica* as the Hausman's (1978) paper, same year, same volume.

minimand produces the test statistic. However, this minimum chi-squared test does not perform well as $T$ gets large, as the number of testable restrictions increase, see Baltagi, Bresson and Pirotte (2009). Only a handful of empirical applications used this method despite its simplification by Angrist and Newey (1991) who showed that this test can be obtained using fixed effects residuals, see the details in a recent survey by Baltagi and Wansbeek (2025) on the Chamberlain and Mundlak models. The Mundlak reduced form model aged well but the Chamberlain reduced form model did not.[4]

Instead of this '*all* or *nothing*' correlation among the regressors and the individual effects, Hausman and Taylor (1981) suggested a model where some of the explanatory variables are correlated with the individual effects but others are not. Hausman and Taylor distinguish between time varying and time invariant variables. They also select which variables among these are correlated with the individual effects and which are not. This choice of variables leads to a Hausman test for over-identification. They develop an instrumtental variable random effects estimator that uses the within and between variation of the time-varying exogenous variables to instrument for time invariant endogenous variables. The endogenous time-varying regressors are instrumented with their within variation which subtracts the time-averaged means of these regressor hence wiping out the correlation with the individual effects. No outside instruments are needed and the advantage is that one gets consistent estimates of the time-invariant variables coefficients which were wiped out by fixed effects. A Hausman test based on this Hausman-Taylor estimator versus fixed effects gives an over-identification test which tests the validity of the choice of endogenous and exogenous regressors with the individual effects. This is applied by Cornwell and Rupert (1988) to a Mincer wage equation and by Egger and Pfaffermayr (2004) to a study of the determinant of exports and foreign direct investment (FDI) in a three-factor New Trade Theory model. Also, Serlenga and Shin (2007) to study the effect of common language on trade using a gravity equation of bilateral trade flows among European countries over the period 1960–2001. Fixed, Random, and Hausman and Taylor estimators as well as the Hausman test are available in Stata which helps increase their application, see Baltagi, Bresson and Pirotte (2003). These are essential tools that will continue to be used in empirical panel data applications.

Historically, fixed versus random effects have been debated in the statistics and biometrics literature and this spilled over to econometrics. The classic literature calls it fixed effects because the individual effects were thought of as fixed parameters estimated using least squares with dummy variables for each individual. Using the Frisch-Waugh-Lovell theorem one can show that least squares dummy variables is equivalent to the within estimator (where within is deviation of the regressor from it time mean), see Baltagi (2021) for details. The modern econometrics literature argues that fixed effects are not parameters but *correlated random effects* as pioneered by Mundlak (1978). Whether fixed or random individual effects, these are wiped out by the within transformation for the static panel data regression models and by differencing for the dynamic panel data models, so who cares? This is why fixed

[4] Chamberlain credits Mundlak for the *correlated random effects* model, see the ET interview with Professor Chamberlain by Graham, Hirano and Imbens (2023, pp. 12-13).

effects is appealing to economists. If they are random and endogenous one gets rid of them. However, it is important to note that the random effects estimator is more efficient than the fixed effects estimator when the regressors are not correlated. This is evident because, the random effects uses the between variation in the regressors and does not discard it, as the within estimator. The random effects estimator yields a generalized least squares estimator that is a weighted combination of the fixed and between estimators weighted by the inverses of their variance-covariance matrices. Ignoring the between variation cannot be efficient, but if this between variation is correlated with the random individual effects, the resulting random effects estimator is not consistent, while the fixed effects within estimator is consistent. Fixed effects users should check the within variation and see whether it wipes out a lot of the variation in the data, especially the policy variable they are interested in. If this within variation is small, then inference based on the fixed effects estimator may be poor. It is also important to note that if $T \to \infty$, then the random effects estimator is asymptotically the same as the fixed effects estimator. If $T$ is fixed or small, as in micro-panel data, the two estimators are different.

## 2.4 Dynamic Panel Data: Estimation and Testing

In time series, the past influences the present, but there's no hidden individual effects causing trouble. While in panel data, the past influences the present and interacts with unobserved individual effects creating endogeneity. Arellano and Bond (1991) hit the panel data literature as well as empirical dynamic panel data researchers like a thunderbolt. It is the most cited paper in economics with 47,000 citations by Google scholar. Despite other papers proposing dynamic panel data estimation like Anderson and Hsiao (1982) (and related papers such as, for example, Holtz-Eakin, Newey & Rosen, 1988; Ahn & Schmidt, 1995, 1997; Arellano & Bover, 1995; Wansbeek & Bekker, 1996; and Crépon, Kramarz & Trognon, 1998), the Arellano and Bond estimator is still the most used for empirical dynamic panel data estimation. It is certainly the biggest success story in panel data econometrics and one that aged well as evidenced by its continuous use and high citation. The paper develops a Generalized Method of Moments (GMM) estimator after first differencing the dynamic panel regression to get rid of the individual effects. Moment conditions are used for estimation utilizing the orthogonality conditions between lagged levels of the endogneous variable and the differenced remainder error. This assumes that the remainder disturbances are not serially correlated (before they were differenced) and failure to find zero serial correlation renders this estimator inconsistent. A two-step GMM procedure is suggested where in the first step the variance-covariance of a moving average with unit roots is used for the differenced disturbances. In the second step, a robust variance-covariance matrix based on the one-step GMM residuals is used to obtain the optimal GMM. A Sargan test for over-identification is an important diagnostic to check the moment conditions used in estimation. These moment conditions increase with $T$ and as a result the Sargan test has weak size and power properties for large $T$ (see Bowsher, 2002). The Arellano-Bond GMM procedure is recommended for large $N$ (the number of individuals) and short $T$. It is well known that fixed effects has bias of order $1/T$ in dynamic panel data estimation (see Nickell, 1981), but if $T$ is large it could still be a viable estimator.

Blundell and Bond (1998) showed that the Arellano and Bond estimator suffers from the weak instrumental variables problem and suggested a system GMM estimator that adds a levels equation to the differenced GMM equation based on a mild stationarity assumption on the initial values. This adds more orthogonal moment conditions based on lagged differences of the dependent variable for the levels equation (see also Ahn & Schmidt, 1995). This system GMM estimator improves the efficiency gains compared to the differenced GMM estimator. The Blundell and Bond (1998) paper has over 35,000 Google citations as of October, 2025. An alternative system GMM procedure is also proposed by Arellano and Bover (1995) stacking a forward orthogonalized transformation of the dynamic model to get rid of the individual effects (instead of first differencing as in Arellano and Bond) on top of an averaged levels equation. This paper has over 27,000 Google citations. It is clear that differenced GMM and system GMM have been a success and have aged really well. This is despite the weak performance of the Sargan test for over-identification (as $T$ gets large) when this diagnostic is in fact crucial for this methodology. There is a huge literature on dynamic panel estimation and testing but none of the proposed estimators (Limited Information Maximum Likelihood, Forward Filtering, etc.) have gained the notoriety or use as the Arellano and Bond and Blundell and Bond estimators.

A comparative analysis of the performance a variety of different estimators available for the estimation of dynamic models can be found in Harris and Mátyás (2004). Overall, the study finds that two important factors that matter from the point of view of consistency are the weakness (or strength) of the instrumental variables and most importantly the size of $T$.

## 2.5 Nonstationary Panels: Panel Unit Root Tests and Panel Cointegration

With large $N$ (number of countries) and large $T$ (length of the time series) macro panels in fact introduced new type of asymptotics in contrast with micro panels where, in general, $T$ is small (see Moon & Phillips, 2000). The fact that $T$ is allowed to increase to infinity in macro panel data, generated two strands of ideas. The first rejected the homogeneity assumption of the regression parameters implicit in the use of a pooled panel regression models in favor of heterogeneous regressions, i.e., one for each country, see Pesaran and Smith (1995), Pesaran, Shin and Smith (1999) and Im, Pesaran and Shin (2003), to mention a few. This literature warns against the use of standard pooled homogeneous panel estimators to estimate dynamic panel data models arguing that they are subject to large potential bias when the parameters are heterogeneous across countries and the regressors are serially correlated. This literature continues to strive with Pesaran and Smith (1995) having 7021 Google citations, Pesaran et al. (1999) having 8,321, and Im et al. (2003) having 21,549 respectively. Certainly a success story.

Another strand of literature applying a time series view point to panels as $T \rightarrow \infty$, is worried about nonstationarity, spurious regressions and cointegration, see, among others, the surveys by Baltagi and Kao (2000), Breitung and Pesaran (2008), Breitung (2015) and Choi (2015).

One of the early panel unit root tests that became a success story with 19,205 Google citations is the Levin, Lin and Chu (2002) (LLC) test. This is a generalization of the augmented Dickey-Fuller (ADF) individual country specific unit root tests to a common panel unit root test. The null hypothesis being that each individual time series contains a unit root against the alternative that each time series is stationary. The test crucially depends upon the independence assumption across cross-sections and is not applicable if cross-sectional correlation is present. It has also been criticized for the restrictive assumption that all cross-sections have or do not have a unit root. Im et al. (2003) (IPS) proposed an alternative testing procedure based on averaging individual unit root ADF test statistics. The null hypothesis now is that each series in the panel contains a unit root, and the alternative hypothesis allows for some (but not all) of the individual series to have unit roots. This paper has also a large number of citations: 21,549.

Breitung (2000) studies the local power of LLC and IPS tests statistics against a sequence of local alternatives and finds that the LLC and IPS tests suffer from a dramatic loss of power if individual specific trends are included. Breitung suggests a test statistic that does not employ a bias adjustment whose power is substantially higher than that of LLC or the IPS tests according to Monte Carlo experiments. Maddala and Wu (1999) and Choi (2001) independently proposed a Fisher type test which combines the p-values from unit root tests for each cross-section to test for unit roots in panel data. Hadri (2000) derives a residual-based Lagrange multiplier (LM) test where the null hypothesis is that there is no unit root in any of the series in the panel against the alternative of a unit root in the panel. These tests are first generation unit root tests (assuming cross-section independence) and are available in standard software packages like Stata and EViews, this increased their application in empirical economics.

Pesaran (2021) suggested a simple test of error cross-sectional dependence (CD) that is applicable to a variety of panel models including stationary and unit root dynamic heterogeneous panels with short $T$ and large $N$. The proposed test is based on an average of pair-wise correlation coefficients of OLS residuals from the individual regressions in the panel rather than their squares as in the Breusch-Pagan LM test, see Baltagi (2021, p. 305). This paper has 14,584 Google citations and is widely used in empirical panel studies to test for cross-sectional dependence. Rejecting cross-sectional independence, means one needs second generation panel unit root tests allowing for common factors which cause cross-section dependence (see Bai & Ng, 2004, as well as Pesaran, 2007 who modified the IPS test to allow for cross-sectional dependence and called it CIPS). There are other second generation panel unit root tests as well, but these two tests are real success stories with the Bai and Ng test garnering 2585 Google citations, whereas the Pesaran (2007) test generated 14,388.

Kao (1999) proposed a panel cointegration tests as an extension of the Engle and Granger (1987) test from individual time series to panels. Pedroni (1999, 2004) also proposed several tests for the null hypothesis of no cointegration in a panel data model that allows for considerable heterogeneity, (see Baltagi, 2021, Chapter 12 for more details). There are other cointegration panel unit root tests available as well, but the Kao and Pedroni tests are the success stories with the Kao test garnering 7,307 Google citations whereas the Pedroni tests 9,359 respectively.

Panel data non-stationary methods are not without their critics. Maddala, Wu and Liu (2000) argued that in their empirical work, panel data unit root tests did not rescue purchasing power parity (PPP). In fact, the results on PPP with panels are mixed depending on the group of countries studied, the period of study and the type of unit root test used. More damaging is their argument by that for PPP, panel data tests are the wrong answer to the low power of unit root tests in single time series. After all, the null hypothesis of a single unit root is different from the null hypothesis of a panel unit root for the PPP hypothesis. Maddala (1999) also argued that panel unit root tests did not help settle the question of growth convergence among countries either. He criticized the LLC panel unit root test arguing that the null may be fine for testing convergence in growth among countries, but the alternative restricts every country to converge at the same rate. Maddala's own words on this: (Lahiri, 1999, p. 760): *"I feel that there is a lot of algebra and less of a discussion of the issues involved and the purpose of using these [unit root] tests"*.

See also Banerjee, Marcellino and Osbat (2005) who criticized existing panel unit root tests for assuming that cross-unit cointegrating relationships among the countries are just not present. They warn that the empirical size of these tests is substantially higher than their nominal size. While cross-sectional dependence received a lot of research attention, empirical papers still mostly use first generation panel unit root tests more often than second generation ones, in fact ignoring any eventual cross-sectional dependence.

## 2.6 Large Heterogeneous Panels

To deal with heterogeneity, let us first remember a very revealing quote from Zvi Griliches: *"One might think that if you go down to the micro level, you are going to control for heterogeneity. But in fact the heterogeneity just changes. There is a sense if you go down and look at bakeries, bakeries are just as different from each other as bakeries as a whole are different from steel plants. You go down and the heterogeneity doesn't go down. There are very few enterprises or anything which are single-dimensional"* (Krueger & Taylor, 2000, p. 185).

Traditional panel data models typically assume a fixed time dimension and posit that, conditional on a set of observable characteristics, any remaining heterogeneity across cross-sectional units can be captured by an additive individual-specific effect, whether treated as fixed or random. A more general framework is provided by linear panel data models with slope heterogeneity (varying coefficients), which allow for differences not only in intercepts but also in the responses of units to explanatory variables.To deal with that, especially in the case of dynamic panel data models, it is often necessary to assume that the number of time series observations is relatively large. This implies that individual equations can be estimated for each unit separately. In short, models, estimation and inference procedures are better suited for large $N$ and $T$ panels. This is becoming more relevant as such panel data sets are increasingly available. Thus, some researchers including Robertson and Symons (1992), Pesaran and Smith (1995), and Pesaran, Smith and Im (1996) have questioned the poolability of the data across heterogeneous units. Instead, they argue in favor of heterogeneous estimates that can be combined to obtain homogeneous estimates if the need arises. Clearly, in panel data sets with $T$ up to 10, traditional homogeneous panel estimators would appear the only viable alternative. If $N$ is small ($N < 10$) 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) and then estimate the parameters using generalized least squares approaches in the spirit of Zellner (1962).

Returning to the distinction between homogeneous and heterogeneous estimators, there exists a large body of literature, although it is predominantly developed in settings that assume no cross-sectional dependence among units. Some examples are Garcia-Ferrer, Highfield, Palm and Zellner (1987), Baltagi and Griffin (1997), Baltagi, Griffin and Xiong (2000), Hoogstrate, Palm and Pfann (2000), Baltagi, Bresson, Griffin and Pirotte (2003), Baltagi, Bresson and Pirotte (2004), Mark and Sul (2012). These studies mostly reach the conclusion that the pooled estimators outperform the heterogeneous ones although some authors such as Hoogstrate et al. (2000) and Mark and Sul (2012) point out to the fact that the dominance of pooled estimators is a result of limited number of time series observations and the low degree of heterogeneity between units. So, if $N$ and $T$ are large and heterogeneity among units is high, heterogeneous estimators should be used for applied works.

A substantial literature has developed formal tests for the null hypothesis of slope parameter homogeneity in panel data models. Swamy (1970) proposed a dispersion-based test comparing individual-specific slope estimates to a pooled estimator under the null of coefficient equality. Then Pesaran et al. (1996) developed Hausman-type tests that compare the fixed effects (FE) estimator with the Mean Group (MG) estimator, exploiting the fact that both are consistent under homogeneity but diverge under heterogeneity. Further developments include Phillips and Sul (2003), who proposed G-tests (different type of Hausman Tests) to test coefficient homogeneity under cross-dependence when $N$ is small relative to $T$, and Pesaran and Yamagata (2008), who introduced standardized dispersion statistics derived from Swamy's framework that remain valid in large panels where both the cross-sectional and time dimensions may be sizable. Blomquist and Westerlund (2013) further extended testing procedures to more general dynamic settings (Heteroskedasticity and Autocorrelation Consistent (HAC) extension). More recently, Campello, Galvão and Juhl (2019) proposed a method to quantify the magnitude of bias in the fixed effects estimator induced by slope heterogeneity. In practice (macro panels), Pesaran and Yamagata's (2008) testing approaches are often used, accumulating over 7,000 Google citations as of June 2026.

On the estimation approaches, Swamy (1970) established the theoretical foundations of Random Coefficient Models (RCMs) by allowing regression parameters to vary across cross-sectional units and testing their homogeneity. RCMs laid the foundation for heterogeneous panel estimators, i.e., Mean Group (MG), Pooled Mean Group (PMG), Common Correlated Effects (CCE), and Bayesian heterogeneous panels. In 1995, Pesaran and Smith proposed the Mean Group (MG) estimator for dynamic heterogeneous panels, demonstrating that pooled estimators become inconsistent in the presence of slope heterogeneity. Then, Pesaran et al. (1999) extended this approach through the Pooled Mean Group (PMG) estimator, which combines heterogeneous short-run dynamics with homogeneous long-run relationships. Subsequent contributions by Pesaran and Zhao (1999), and by Hsiao, Pesaran and Tahmiscioglu (1999), addressed finite-sample bias, dynamic random coefficients and Bayesian estimation issues in heterogeneous panels. Finally, Pesaran (2006) incorporated unobserved common factors and cross-sectional dependence[5] through the Common Correlated Effects (CCE) estimator, thereby extending heterogeneous panel methods to more realistic macroeconomic environments characterized by global shocks and interdependencies. The CCE estimator is straightforward to implement compared with the PMG estimator or factor models, as it only requires augmenting the regressions with cross-sectional averages to account for cross-sectional dependence. Eberhardt and Teal (2010) developed the Augmented Mean Group (AMG) estimator as an alternative to the CCE estimator.[6]

## 2.7 Cross-Sectional Dependence

Depending on its strength and nature, omitting or misspecifying cross-dependence (CD) can lead to misleading results (see Moscone & Tosetti, 2009). CD may arise from local interactions that generate spatial spillovers or from common factors affecting multiple units (see Chudik, Pesaran & Tosetti, 2011; Sarafidis & Wansbeek, 2012; and Bailey, Holly & Pesaran, 2016). With increasingly rich datasets, panels with large N and large T have become more common, creating new possibilities and challenges for how CD is characterized. Several papers distinguish between weak CD (WCD) and strong CD (SCD). In terms of their implications for the statistical properties of conventional panel estimators, the two forms of CD differ substantially. Sarafidis and Wansbeek (2012) note that the literature lacks a unique definition of these terms. The factor and spatial econometric frameworks are largely complementary, with the former being more appropriate for capturing SCD and the latter generally relying on WCD, following the definition in Chudik et al. (2011).

A significant body of literature exists on CD testing in large panels which follows two separate strands, depending on whether the cross-sectional units are ordered. Typically, this is the case when the observations are spatial or linked to social or

[5] For more details, see next section.

[6] For implementation details, see Blackburne and Frank (2007) for the MG and PMG estimators and Eberhardt (2012) for the MG, CCE and AMG estimators.

economic networks (i.e., *via* spatial or network matrices, see Moran, 1948, 1950; Cliff & Ord, 1981; Anselin, 1988; Anselin & Bera, 1998, Pinkse, 1999, 2004; Haining, 2003; Baltagi, Song, Jung & Koh, 2007; and Kelejian & Prucha, 2001). These approaches depend mainly on the choice of spatial/network matrices, and may not be appropriate for several types of panels in economics and finance where space is not a natural metric for modelling of CD. Pesaran (2004, 2021) developed a general diagnostic framework for testing cross-sectional dependence in panel data models. He addresses the limitations of traditional dependence tests, such as the Breusch–Pagan Lagrange Multiplier (LM) test, which exhibits substantial size distortion when $N$ is large relative to $T$, a common situation encountered in applied studies. He introduced a cross-sectional dependence tests, based on the average of pairwise correlation coefficients of regression residuals. The proposed statistic is applicable to a wide range of panel data models, including fixed effects, random effects, and heterogeneous panels, and remains valid when both the cross-sectional dimension ($N$) and the time dimension ($T$) are large. Pesaran, Ullah and Yamagata (2008) focused on a bias-adjusted version of the Breusch–Pagan LM test for detecting cross-sectional dependence in panel data models. They derive the exact mean and variance of the LM statistic under the null hypothesis of cross-sectional independence and use these results to construct bias-adjusted test statistics with improved finite-sample properties assuming strictly exogenous regressors and normal errors. They underline that the proposed bias-adjusted LM test remains consistent even in situations where the earlier CD test proposed by Pesaran (2004) may become inconsistent. They also note that the bias-adjusted LM tests are less robust than the CD tests in the presence of non-normal errors or weakly exogenous regressors. A generalization of the bias-adjusted LM tests, which capture spatial patterns, is also proposed. Baltagi, Feng and Kao (2012) derive the limiting distribution of the scaled version of LM test proposed by Pesaran (2004) but applied to a fixed effects model. They find that this LM test exhibits an asymptotic bias which is related to the number of cross-sectional units $N$ and the number of time periods $T$. Therefore, a bias-corrected LM test is proposed and the simulation results show that it successfully controls for size distortions as $N$ gets large relative to $T$. Overall, Pesaran's (2004, 2021) CD tests are the most popular ones in practice, whereas the LM tests are mostly used in classical panel settings. Pesaran's (2021) article has over 16,000 Google citations as of June 2026.

Several estimators have been proposed to deal with cross-sectional dependence, including the popular spatial methods (Anselin, 1988; Anselin & Bera, 1998; Anselin, Florax & Rey, 2004; and Lee, 2024, and factor models (Bai & Ng, 2002; Pesaran, 2006; Kapetanios & Pesaran, 2007; Bai, 2003, 2009; and Song, 2013). The literature dealing with spatial interactions and common factors simultaneously is scarce, some exceptions being Bai (2009), Pesaran and Tosetti (2011), Bailey et al. (2016), Shi and Lee (2018) and Kuersteiner and Prucha (2020). Among the CD estimation methods, Pesaran (2006) introduces a widely used panel data model in which cross-sectional heterogeneity is captured through a factor structure in the errors, with the unobserved factors also influencing the observed regressors. A key advantage of this approach is its simplicity relative to principal-components-based methods. (see Bai & Ng, 2002; Kapetanios & Pesaran, 2007; Bai, 2003, 2009; and Song, 2013). He proposed the CCE estimation method, which approximates the linear combinations of the unobserved factors using cross-sectional averages of the dependent and explanatory variables, and then estimates standard panel regressions augmented with these averages. An important advantage of this approach is that it delivers consistent estimates under a wide range of conditions, including serial correlation in the errors, unit roots in the factors, and possible contemporaneous dependence between the observed regressors and the unobserved factors. In 2011, Pesaran and Tosetti showed that when both common factors and spatial correlation are present, the CCE estimator continues to provide slope coefficient estimates with desirable properties—consistency and asymptotic normality.

## 2.8 Discrete, Qualitative and Limited Dependent Variables in Panels

In many economic applications, the variables of interest are discrete, qualitative, or subject to truncation or censoring. The models used in these settings are typically highly nonlinear (see Chapter 10 of this book). In panel data, the presence—primarily—of individual heterogeneity complicates matters substantially. This difficulty is closely related to the incidental-parameter problem in fixed-effects models, see Neyman and Scott (1948), Lancaster (2000), for which no general solution currently exists and is treated on a case by case basis. Early extensions of panel data methods to nonlinear models largely emerged in the late 1970s and early 1980s.

For qualitative dependent variables, static binary choice models—probit and logit—provide the standard nonlinear panel data framework. Key early contributions include Chamberlain's (1980) development of the fixed effects model (see Chapter 10, Section 10.9.9), as well as the random effects model proposed by Heckman and Willis (1975, 1977), followed by the efficient computational algorithm introduced by Butler and Moffit (1982) (see Chapter 10, Section 10.7.3). Numerous extensions and less parameterized alternatives have been proposed, including Manski (1975), Honoré and Lewbel (2002), Hahn and Newey (2004), Hahn and Newey (2007), Fernández-Val (2009), Hahn and Kuersteiner (2011), and Honoré and Kesina (2017). For limited dependent variable models, see Heckman and MaCurdy (1980), Honoré (1992), Hajivassiliou (1993), Keane (1993), Hajivassiliou and Ruud (1994), Honoré and Powell (1994, 2005), and Honoré and Kesina (2017).

Although the foundations for dynamic discrete choice models were established early by Heckman (1978, 1981), there was limited subsequent development until the 1990s. Research during that period focused primarily on limited dependent variable and sample selection models, see Honoré (1993), Kyriazidou (1997), Hajivassiliou and McFadden (1998), Honoré and Kyriazidou (2000), Wooldridge (1995) and Wooldridge (2005) to mention but a few.

Overall, the problem of identification and estimation is challenging especially when $N$ is large, and $T$ is small. The major point is to identify the coefficients of the observed regressors in the presence of fixed effects that have an unrestricted joint distribution with the regressors. A huge literature has been devoted to this problem, including classical results on binary choice models by Manski (1987), Honoré and Kyriazidou (2000), Honoré and Lewbel (2002), and Chamberlain (2010). More recent advances have been made by Davezies, D'Haultfœuille and Mugnier (2023), Khan, Ponomareva and Tamer (2023) and Honoré and Weidner (2025). When $N$ and $T$ tend to infinity at specified rates, Hahn and Newey (2004), Bai (2009), Dhaene and Jochmans (2015), Fernández-Val and Weidner (2016), among others, had developed general analytical and numerical methods for addressing the incidental parameters problem. Last, it should be noted that work on interactive fixed effects has emerged in recent years, see Chen, Fernández-Val and Weidner (2021), Ando, Bai and Li (2022) and Gao, Liu, Peng and Yan (2023).

To conclude, most contributions focus on micro-panels ($N$ large, $T$ small), and the econometric techniques involved are often complex. Without a general solution to the incidental-parameter problem, each model demands its own specialized treatment, limiting the broader applicability of these methods. This likely explains the modest citation numbers for many of the references mentioned. When both $N$ and $T$ are large, nonlinear panel estimation is still an open and largely unexplored field.

## 2.9 Multi-dimensional Panels

The analysis of flows such as trade, migration, travel, and transport gave rise to the literature on three-dimensional panel data models. Early contributions in the 1970s and 1980s concentrated on international trade, with exporting and importing countries representing the two cross-sectional dimensions (see, for example, Aitken, 1973, and Bergstrand, 1985, Christenson, 1994, and Oguledo & MacPhee, 1994, and Chapter 20 for more details). Mátyás (1997, 1998) played a key role in incorporating these models into the mainstream panel data framework.

The data revolution has led to a rapid expansion in the size and scope of multidimensional data sets, extending their use to numerous areas of economics (see, for example, in Mátyás, 2017, 2023). Data on housing markets, migration flows, prices, and linked relationships such as doctor–patient and firm–worker matches have become increasingly common in empirical analyses. This development has raised new challenges concerning asymptotic theory, the specification of individual, temporal, and cross-sectional heterogeneity, and the treatment of dynamic relationships. For example, it is often unrealistic to assume a common lag structure across all individuals. This makes model specification and selection particularly important in multidimensional settings. Moreover, the increasing number of dimensions complicates the treatment of heterogeneity, dynamics, and asymptotic theory. As Marc Nerlove said: *"Perhaps the most important difference between multi-dimensional and two-dimensional panels is how the number of available effects is specified in the models. Also, the many dimensions in which asymptotic expansions of the distribution of estimates may be considered."* (Nerlove, 2023, p. vi). Furthermore, many contemporary multidimensional data sets cannot be regarded as random samples from an underlying population. Instead, they often constitute essentially complete records, or snapshots, of a given economic or social process over a particular period. As Jeffrey Wooldridge noted (Zhao, 2022, Q. 3): *"There's no sampling, but what you do have is that you have potential outcomes in both a control state and a treated state. Then, all the uncertainty comes through assigning units to the control or the treated".* This observation has far-reaching implications for statistical inference. It raises the question of whether, and to what extent, conventional inferential tools—such as hypothesis testing and confidence intervals—remain appropriate, and how their outcomes should be interpreted when uncertainty stems from assignment mechanisms rather than random sampling.

Panel data econometrics is also gradually converging with newer fields, such as network analysis, as researchers seek more effective ways to exploit the richness of large multidimensional data sets.

At the same time, we are confronted with a dilemma that is common throughout econometrics. The growing complexity of models is often accompanied by a loss of transparency and interpretability. While increasingly sophisticated methods may provide a more realistic representation of economic behaviour, they can also reduce the practical usefulness of empirical results for economic analysis and policy evaluation. The successful future development of multidimensional panel data econometrics will therefore depend on maintaining an appropriate balance between realism and tractability. Overly complex models risk becoming methodological cul-de-sacs rather than useful tools for understanding economic phenomena.

The lessons that can be drawn from the rapid emergence of multidimensional panel data, as well as the challenges that lie ahead, may be summarized as follows:

1. For the 'traditional' multi-D panel data sets, where all dimensions are of 'reasonable' size, rather than in the realm of 'big data'—for example, country- or region-level trade data and the associated gravity models—the conventional fixed-effects (and, in some cases, random-effects) specifications, together with the corresponding projection-based estimation methods, have generally performed well and stood the test of time.
2. More recently, however, data sets have emerged in which one or more dimensions contain such a large number of observations that effective asymptotics become a practical reality (see also, Chapters 3 and 4). In such settings:
   - The traditional fixed-effects approach, which assigns a separate parameter to each individual unit and relies on the associated orthogonal projections, becomes increasingly difficult to justify. Explicitly or implicitly, it may require the estimation of an effectively infinite number of parameters. Consequently, this approach does not age well in such environments. Fixed effects must therefore be specified more parsimoniously, involving only a limited and economically meaningful number of parameters.
   - The conventional random-effects, or error-components, approach may also be open to question. When the amount of information available is effectively unlimited, it becomes less clear what source of uncertainty the stochastic specification is intended to capture and formalize.

- Model selection is another area where substantial challenges remain.. The criteria and procedures reviewed in Chapter 12 were developed for considerably smaller and less complex settings and are not readily applicable to modern multidimensional data sets. Moreover, many traditional indicators of model performance can become difficult to interpret. For example, in most empirical cases the $R^2$ is close to zero, which may or may not be troubling. This area certainly needs further exploration.
- Conventional significance testing may also become of limited relevance in this setting. If a consistent estimator is available in a dimension whose size tends to infinity, estimation uncertainty effectively disappears, as the variance of the estimator converges to zero. In such cases, the focus should shift away from statistical significance toward the magnitude, economic importance, and interpretation of the parameter itself.
- The principal challenge is therefore to formalize individual and temporal heterogeneity in a manner that remains parsimonious, interpretable, and economically meaningful, while still capturing the essential features of the underlying heterogeneity. The objective is to strike an appropriate balance between flexibility and tractability, avoiding specifications that generate an unmanageable number of parameters without sacrificing the ability to represent important differences across individuals and over time.
- Dynamics constitute another important challenge. It is highly implausible that all individuals, across all dimensions of a multidimensional panel, share the same dynamic structure or, for example, the same autoregressive behaviour. Imposing homogeneous dynamics may therefore lead to serious misspecification. At the same time, allowing unrestricted heterogeneity can quickly produce an unmanageable number of parameters. Consequently, the specification of dynamic relationships requires particularly careful consideration, balancing flexibility with parsimony and interpretability.
- An additional issue in dynamic models is that in these data set frequently $T$ is small (and unbalanced) and we know that the usual IV/GMM estimators for such data perform poorly. Future research should focus on working out orthogonality conditions for the GMM approach that work well when at least one (or several) individual dimension is (are) large, while the time dimension is (very) small.
- Certain linked data sets, such as doctor–patient, firm–worker, or teacher–student data, may involve multiple time dimensions, with different units being observed over distinct and potentially overlapping time horizons. How best to model and exploit such structures remains largely an open question. The development of appropriate econometric methods for handling multiple time dimensions in multidimensional panel data therefore represents another important avenue for future research.

## 2.10 Some Further Topics, General Lessons, and Looking Ahead

### 2.10.1 Further Topics

There are several other areas of panel data econometrics that have produced important methodological advances and have substantial practical relevance. One such area concerns incomplete panels. In empirical work, perfectly balanced panel data sets are the exception rather than the rule, as missing observations, entry and exit of units, and measurement problems are commonplace. As a result, the analysis of incomplete and unbalanced panels continues to be of considerable importance both in theory and in practice. In this literature, the main focus is on whether this data is randomly versus non-randomly missing. The findings are usually that one should not worry when the data is randomly missing. This is not necessarily true for the unbalanced error component model, see Chapter 9 of Baltagi (2021) for a textbook treatment of the subject. Serious problems of non-randomly missing data are studied extensively in the literatureand is too numerous to survey here. This includes the literature on sample selection, attrition, studied by Gronau (1974), Heckman (1976, 1979), Hausman and Wise (1979), among others. Biørn (1981), Deaton (1985); Verbeek and Nijman (1996), and Nijman, Verbeek and van Soes (1991) are important contributions in this literature. Data confidentiality, however, may reopen some related issues (see Chapter 15 for more details).

Pseudo-panels are also used in applied work, particularly when confidentiality constraints prevent researchers from accessing true panel data. The foundations of the pseudo-panel approach were laid by
citeAHeckman-Robb1985, Deaton (1985), Moffitt (1993), and Verbeek (2008). Despite their limitations, pseudo-panels often provide a practical and valuable alternative to genuine panel data sets.

System(s) of equations on panels have been out of fashion for a while, but the basic contributions to this topic can be traced back to the work of Baltagi (1981); Balestra and Krishnakumar (1987); and Krishnakumar (1988).

### 2.10.2 Main Lessons

The most important evergreen question in panel data econometrics—and even more so in large and multidimensional panels—concerns the treatment of individual and time heterogeneity. This issue can never be regarded as definitively resolved because each new generation of data sets brings with it new forms of heterogeneity and new modelling challenges. At the same time, our growing understanding of economic behaviour reinforces the need to account for an ever wider range of observable and unobservable individual differences. Consequently, the search for flexible yet parsimonious ways of modelling heterogeneity will continue to occupy a central place in both the theory and practice of panel data econometrics.

More generally, the development of panel data econometrics has highlighted the critical role of identification and the various strategies used to achieve it. This remains one of the most enduring challenges in econometrics. As models, data sets, and empirical questions have become more sophisticated, so too has the need for credible identification. Indeed, many of the major advances in panel data econometrics can be viewed as attempts to exploit the richness of panel structures to obtain more convincing identification of the parameters of interest.

### 2.10.3 Looking Ahead

Let us conclude by mentioning several areas that may illuminate the road ahead. The interaction and eventual integration of panel data econometrics with machine learning (already taking shape; see Chapter˜12), network analysis (see Chapter˜16), and artificial intelligence seem both inevitable and promising. Together, these developments are likely to provide entirely new ways of modelling, interpreting, and exploiting the information contained in increasingly complex panel data sets. At the same time, multidimensional panels raise theoretical questions that cannot always be addressed by existing asymptotic frameworks. New forms of asymptotic and semiasymptotic analysis will be needed to accommodate data structures in which some dimensions are extremely large while others remain limited. Finally, several important issues concerning aggregation and disaggregation remain unresolved. The development of methods capable of linking micro-level behaviour to aggregate outcomes, and vice versa, is particularly relevant for policy analysis and forecasting. Progress in these areas will play an important role in determining the future direction of panel data econometrics.

## 2.11 Conclusions

Overall, panel data methods and models have been profoundly influenced and transformed by their own success. Increasingly large data sets—often described as ‘big data’—are frequently observed in the form of multidimensional panels, creating both new opportunities and new challenges. As a result, panel data econometrics has become a continuously evolving field, with new methods and modelling approaches emerging in response to the data revolution. Size itself presents a challenge, not merely because of computational demands, but because larger data sets tend to reveal increasingly complex forms of heterogeneity. This calls for new ideas, innovative models, and fresh methodological approaches. Some traditional frameworks, such as the error-components or random-effects approach, have become less prominent, while established treatments of issues such as nonstationarity, unit roots, and cointegration are being re-examined and, in some cases, fundamentally questioned in these new settings. At the same time, flexibility, parsimony, and ease of interpretation have emerged as key criteria guiding methodological progress. Sophistication alone is not sufficient; econometric models must also remain useful tools for understanding economic behaviour and informing policy. Ultimately, the relentless pressure exerted by new and ever richer data sets continues to shape the evolution of the field. The message for panel data econometrics is clear: adapt to the changing data environment or risk becoming obsolete. This is perhaps the most important lesson of this chapter.

## References

Ahn, S. C. & Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, *68*, 5-27.

Ahn, S. C. & Schmidt, P. (1997). Efficient estimation of dynamic panel data models: Alternative assumptions and simplified estimation. *Journal of Econometrics*, *76*, 309-321.

Airy, G. B. (1861). *On the algebraical and numerical theory of errors of observations and the combination of observations*. Cambridge and London: Macmillan & Co.

Aitken, N. D. (1973). The effect of the EEC and EFTA on European trade: A temporal cross-section analysis. *American Economic Review*, *63*, 881-892.

Amini, S., Delgado, M., Henderson, D. J. & Parmeter, C. F. (2012). Fixed vs random: The Hausman test four decades later. In B. H. Baltagi, R. Carter Hill, W. K. Newey & H. L. White (Eds.), *Essays in honor of jerry hausman* (Vol. 29, p. 479-513). Emerald Group Publishing Limited, Bingley.

Anderson, T. W. & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, *18*, 47-82.

Ando, T., Bai, J. & Li, K. (2022). Bayesian and maximum likelihood analysis of large-scale panel choice models with unobserved heterogeneity. *Journal of Econometrics*, *230*, 20-38.

Angrist, J. D. & Newey, W. K. (1991). Over-identification tests in earnings functions with fixed effects. *Journal of Business and Economic Statistics*, *9*, 317-323.

Anselin, L. (Ed.). (1988). *Spatial econometrics: Methods and models*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Anselin, L. & Bera, A. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. E. Giles (Eds.), *Handbook of applied economic statistics* (p. 237-289). New York: Marcel Dekker.

Anselin, L., Florax, R. J. G. M. & Rey, S. J. (2004). *Advances in spatial econometrics. methodology, tools and applications*. Berlin: Springer-Verlag.

Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, *58*, 277-297.

Arellano, M. & Bover, O. (1995). Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics*, *68*, 29-51.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, *71*, 135-171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, *77*, 1229-1279.

Bai, J. & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*, 191-221.

Bai, J. & Ng, S. (2004). A PANIC attack on unit roots and cointegration. *Econometrica*, *72*, 1127-1177.

Bailey, N., Holly, S. & Pesaran, M. H. (2016). A two-stage approach to spatiotemporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics*, *31*, 249-280.

Balestra, P. & Krishnakumar, J. (1987). Full information estimations of a system of simultaneous equations with error component structure. *Econometric Theory*, *3*, 223-246.

Balestra, P. & Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*, *34*, 585-612.

Baltagi, B. H. (1981). Simultaneous equations with error components. *Journal of Econometrics*, *17*, 189-200.

Baltagi, B. H. (1995). *Econometric analysis of panel data*. Wiley.

Baltagi, B. H. (2021). *Econometric analysis of panel data* (6th ed.). Springer.

Baltagi, B. H. (2024). Hausman's specification test for panel data: Practical tips. In C. Parmeter, M. Tsionas & H. Wang (Eds.), *Essays in honor of subal kumbhakar* (Vol. 46, p. 13-24). Emerald Publishing Limited.

Baltagi, B. H., Bresson, G., Griffin, J. M. & Pirotte, A. (2003). Homogeneous, heterogeneous or shrinkage estimators? some empirical evidence from french regional gasoline consumption. *Empirical Economics*, *28*, 795-811.

Baltagi, B. H., Bresson, G. & Pirotte, A. (2003). Fixed effects, random effects or Hausman-Taylor? A pretest estimator. *Economics Letters*, *79*, 361-369.

Baltagi, B. H., Bresson, G. & Pirotte, A. (2004). Tobin q: Forecast performance for hierarchical Bayes, shrinkage, heterogeneous and homogeneous panel data estimators. *Empirical Economics*, *29*, 107-113.

Baltagi, B. H., Bresson, G. & Pirotte, A. (2009). Testing the fixed effects restrictions? A Monte Carlo study of Chamberlain's minimum Chi-squared test. *Statistics and Probability Letters*, *79*, 1358-1362.

Baltagi, B. H., Feng, Q. & Kao, C. (2012). A lagrange multiplier test for cross-sectional dependence in a fixed effects panel data model. *Journal of Econometrics*, *170*, 164-177.

Baltagi, B. H. & Griffin, J. M. (1997). Pooled estimators v.s. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics*, *77*, 303-327.

Baltagi, B. H., Griffin, J. M. & Xiong, W. (2000). To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics*, *82*, 117-126.

Baltagi, B. H. & Kao, C. (2000). Nonstationary panels, cointegration in panels and dynamic panels: A survey. In B. H. Baltagi, T. B. Fomby & R. Carter Hill (Eds.), *Nonstationary panels, panel cointegration, and dynamic panels* (Vol. 15, p. 7-51). Emerald Publishing Limited.

Baltagi, B. H., Song, S. H., Jung, B. C. & Koh, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics*, *140*, 5-51.

Baltagi, B. H. & Wansbeek, T. (2025). The basics of the Mundlak and Chamberlain projections. In B. H. Baltagi & L. Mátyás (Eds.), *Seven decades of econometrics and beyond: A tribute to the life and work of Marc Nerlove* (Vol. 57, p. 395-415). Emerald Publishing Limited.

Banerjee, A., Marcellino, M. & Osbat, C. (2005). Testing for PPP: Should we use panel methods? *Empirical Economics*, *30*, 77-91.

Bergstrand, J. H. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *Review of Economics and Statistics*, *67*, 474-481.

Biørn, E. (1981). Estimating economic relations from incomplete cross-section/time-series data. *Journal of Econometrics*, *16*, 221-236.

Blackburne, E. F. & Frank, M. W. (2007). Estimation of nonstationary heterogeneous panels. *Stata Journal*, *7*, 197-208.

Blomquist, J. & Westerlund, J. (2013). Testing slope homogeneity in large panels with serial correlation. *Economics Letters*, *121*, 374-378.

Blundell, R. & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*, 115-143.

Bowsher, C. G. (2002). On testing overidentifying restrictions in dynamic panel data models. *Economics Letters*, *77*, 211-220.

Breitung, J. (2000). The local power of some unit root tests for panel data. In B. H. Baltagi, T. B. Fomby & R. Carter Hill (Eds.), *Nonstationary panels, panel cointegration, and dynamic panels* (Vol. 15, p. 161-177). Emerald Publishing Limited.

Breitung, J. (2015). The analysis of macroeconomic panel data. In B. H. Baltagi (Ed.), *Oxford handbook of panel data* (p. 453-492). Oxford University Press.

Breitung, J. & Pesaran, M. H. (2008). Unit roots and cointegration in panels. In L. Mátyás & P. Sevestre (Eds.), *The econometrics of panel data: Fundamentals and recent developments in theory and practice* (p. 279-322). Springer.

Butler, J. S. & Moffit, R. (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, *50*, 761-764.

Campello, M., Galvão, A. F. & Juhl, T. (2019). Testing for slope heterogeneity bias in panel data models. *Journal of Business & Economic Statistics*, *37*, 749-760.

Chamberlain, G. (1980). Analysis with qualitative data. *Review of Economic Studies*, *47*, 225-238.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics*, *18*, 5-46.

Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica*, *78*, 159-168.

Chauvenet, W. (1863). *A manual of theoretical and practical astronomy: embracing the general problems of spherical astronomy, the special applications to nautical astronomy, and the theory and use of fixed and portable astronomical instruments, with an appendix on the method of least squares*. Philadelphia: J. B. Lippincott & Co.

Chen, M., Fernández-Val, I. & Weidner, M. (2021). Nonlinear factor models for network and panel data. *Journal of Econometrics*, *220*, 296-324.

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, *20*, 249-272.

Choi, I. (2015). Panel cointegration. In B. H. Baltagi (Ed.), *Oxford handbook of panel data* (p. 46-75). Oxford University Press.

Christenson, B. (1994). World trade in apparel: An analysis of trade flows using the gravity model. *International Regional Science Review*, *17*, 151-166.

Chudik, A., Pesaran, M. H. & Tosetti, E. (2011). Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal*, *14*, C45-C90.

Cliff, A. & Ord, J. K. (1981). *Spatial processes: Models and applications*. London: Pion.

Cornwell, C. & Rupert, P. (1988). Efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics*, *3*, 149-155.

Crépon, B., Kramarz, F. & Trognon, A. (1998). Parameters of interest, nuissance parameter and orthogonality condition: An application to autoregressive error components models. *Journal of Econometrics*, *82*, 135-156.

Daniels, H. E. (1939). The estimation of components of variance. *Supplement to the Journal of the Royal Statistical Society*, *6*, 186-197.

Davezies, L., D'Haultfœuille, X. & Mugnier, M. (2023). Fixed effects binary choice models with three or more periods. *Quantitative Economics*, *14*, 1105-1132.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, *30*, 109-126.

Dhaene, G. & Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies*, *82*, 991-1030.

Dupont-Kieffer, A. & Pirotte, A. (2011). The early years of panel data econometrics. *History of Political Economy*, *43*, 258–282.

Eberhardt, M. (2012). Estimating panel time-series models with heterogeneous slopes. *Stata Journal*, *12*, 61-71.

Eberhardt, M. & Teal, F. (2010). *Productivity analysis in global manufacturing production.* Economics Series Working Papers, 515, University of Oxford, Department of Economics.

Egger, P. & Pfaffermayr, M. (2004). Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics*, *19*, 227-246.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics*, *3*, 1-21.

Engle, R. F. & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, *55*, 251-276.

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, *150*, 71-75.

Fernández-Val, I. & Weidner, M. (2016). Individual and time effects in nonlinear panel models with large $N$, $T$. *Journal of Econometrics*, *192*, 291-312.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Gao, J., Liu, F., Peng, B. & Yan, Y. (2023). Binary response models for heterogeneous panel data with interactive fixed effects. *Journal of Econometrics*, *235*, 1654-1679.

Garcia-Ferrer, A., Highfield, R. A., Palm, F. C. & Zellner, A. (1987). Macroeconomic forecasting using pooled international data. *Journal of Business & Economic Statistics*, *5*, 53-67.

Gauss, C. F. (1809). *Theoria motus corporum celestium*. Perthes und Besser. (Translation by C. H. Davis in Theory of Motion of Heavenly Bodies, New York: Dover, 1963)

Glick, R. & Rose, A. K. (2002). Does a currency union affect trade? The time series evidence. *European Economic Review*, *46*, 1125–1151.

Graham, B., Hirano, K. & Imbens, G. W. (2023). The ET interview: Professor Gary Chamberlain. *Econometric Theory*, *39*, 1-26.

Gronau, R. (1974). Wage comparisons–A selectivity bias. *Journal of Political Economy*, *82*, 1119-1143.

Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *Econometrics Journal*, *3*, 148-161.

Hahn, J. & Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, *27*, 1152-1191.

Hahn, J. & Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, *77*, 1295-1313.

Hahn, J. & Newey, W. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics*, *140*, 503-528.

Haining, A. (2003). *Spatial data analysis: Theory and practice*. Cambridge: Cambridge University Press.

Hajivassiliou, V. A. (1993). Simulation estimation methods for limited dependent variable models. In G. S. Maddala, C. R. Rao & H. D. Vinod (Eds.), *Handbook of statistics, vol. 11* (p. 519-543). Amsterdam: North-Holland.

Hajivassiliou, V. A. & McFadden, D. (1998). The method of simulated scores for the estimation of ldv models. *Econometrica*, *66*, 863-896.

Hajivassiliou, V. A. & Ruud, P. A. (1994). Classical estimation methods using simulation. In R. Engle & D. McFadden (Eds.), *Handbook of econometrics, vol. 4* (p. 2383-2441). Amsterdam: North-Holland.

Harris, M. H. & Mátyás, L. (2004). A comparative analysis of different IV and GMM estimators of dynamic panel data models. *International Statistical Review*, *72*, 397-408.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*, 1251-1271.

Hausman, J. A. & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, *49*, 1377-1398.

Hausman, J. A. & Wise, D. A. (1979). Attrition bias in experimental and panel data: The Gary income maintenance experiment. *Econometrica*, *47*, 455-473.

Heckman, J. J. (1978). Simple statistical models for discrete panel data developed and applied toi test of the hypothesis of true test dependence against the hypothesis of spurious state dependence. *Annales de l'INSEE*, *30/31*, 227-269.

Heckman, J. J. (1981). Statistical models for discrete panel data. In C. F. Manski & D. McFadden (Eds.), *Structural analysis of discrete data with econometric applications* (p. 114-178). Cambridge, MA: MIT Press.

Heckman, J. J. & MaCurdy, T. E. (1980). A life cycle model of female labour supply. *Review of Economic Studies*, *47*, 47-74.

Heckman, J. J. & Willis, R. (1975). Estimation of a stochastic model of reproduction: An econometric approach. In N. Terleckyj (Ed.), *Household production and consumption* (p. 99-138). New York: National Bureau of Economic Research.

Heckman, J. J. & Willis, R. (1977). A beta logistic model for the analysis of sequential labor force participation by married women. *Journal of Political Economy*, *85*, 27-58.

Hildreth, C. (1949). *Preliminary considerations regarding time series and/or cross section studies.* Cowles Commission Discussion Paper: Statistics No. 333.

Hildreth, C. (1950). *Combining cross section data and time series.* Cowles Commission Discussion Paper, No. 347.

Hoch, I. (1955). Estimation of production function parameters and testing for efficiency. *Econometrica*, *23*, 325-326.

Hoch, I. (1957). *Estimation of agricultural resource productivities combining time series and cross-section data.* PhD diss., University of Chicago.

Hoch, I. (1958). Simultaneous equations bias in the context of cobb-douglas production function. *Econometrica*, *26*, 566-578.

Hoch, I. (1962). Estimation of production function parameters combining time series and cross-section data. *Econometrica*, *30*, 34-53.

Holtz-Eakin, D., Newey, W. & Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica*, *56*, 1371–1395.

Honoré, B. E. (1992). Trimmed lad and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica*, *60*, 533-565.

Honoré, B. E. (1993). Orthogonality conditions for tobit models with fixed effects and lagged dependent variables. *Journal of Econometrics*, *59*, 35-61.

Honoré, B. E. & Kesina, M. (2017). Estimation of some nonlinear panel data models with both time-varying and time-invariant explanatory variables. *Business and Economic Statistics*, *35*, 543-558.

Honoré, B. E. & Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, *68*, 839-874.

Honoré, B. E. & Lewbel, A. (2002). Semiparametric binary choice panel data models without strictly exogenous regressors. *Econometrica*, *70*, 2053-2063.

Honoré, B. E. & Powell, J. (2005). Pairwise difference estimation of nonlinear models. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference* *for econometric models. essays in honor of Thomas Rothenberg* (p. 520-553). Cambridge University Press.

Honoré, B. E. & Powell, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*, *64*, 241-278.

Honoré, B. E. & Weidner, M. (2025). Moment conditions for dynamic panel logit models with fixed effects. *Review of Economic Studies*, *92*, 3112-3137.

Hoogstrate, A. J., Palm, F. C. & Pfann, G. A. (2000). Pooling in dynamic paneldata models: An application to forecasting gdp growth rates. *Journal of Business & Economic Statistics*, *18*, 274-283.

Hsiao, C. (1986). *Analysis of panel data*. Cambridge University Press.

Hsiao, C., Pesaran, M. H. & Tahmiscioglu, A. K. (1999). Bayes estimation of short-run coefficients in dynamic panel data models. In C. Hsiao, K. Lahiri, L.-F. Lee & M. H. Pesaran (Eds.), *Analysis of panels and limited dependent variables: A volume in honour of G. S. Maddala* (p. 268-296). Cambridge University Press.

Im, K. S., Pesaran, M. H. & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, *115*, 53-74.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of Econometrics*, *90*, 1-44.

Kapetanios, G. & Pesaran, M. H. (2007). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. In G. Phillips & E. Tzavalis (Eds.), *The refinement of econometric estimation and test procedures: Finite sample and asymptotic analysis* (p. 239-281). Cambridge: Cambridge University Press.

Keane, M. (1993). Simulation estimation methods for panel data limited dependent variable models. In G. S. Maddala, C. R. Rao & H. D. Vinod (Eds.), *Handbook of statistics, vol. 11* (p. 545-571). Amsterdam: North-Holland.

Kelejian, H. H. & Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, *104*, 219-257.

Khan, S., Ponomareva, M. & Tamer, E. (2023). Identification of dynamic binary response models. *Journal of Econometrics*, *237*, 105515.

Krishnakumar, J. (1988). *Estimation of simultaneous equation models with error components structure*. Berlin: Springer Verlag.

Krueger, A. B. & Taylor, T. (2000). An interview with Zvi Griliches. *Journal of Economic Perspectives*, *14*, 171–189.

Kuersteiner, G. M. & Prucha, I. R. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, *88*, 2109-2146.

Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica*, *65*, 1335-1364.

Lahiri, K. (1999). Econometric Theory interview with G.S. Maddala. *Econometric Theory*, *15*, 753–776.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, *95*, 391-413.

Lee, L.-F. (2024). *Spatial econometrics. spatial autoregressive models*. World Scientific Series on Econometrics and Statistics, Vol. 1, World Scientific.

Legendre, A. M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. Courcier.

Levin, A., Lin, C. F. & Chu, C. (2002). Unit root test in panel data: Asymptotic and finite sample properties. *Journal of Econometrics*, *108*, 1-24.

Maddala, G. S. (1999). On the use of panel data methods with cross country data. *Annales D'Économie et de Statistique*, *55-56*, 429-448.

Maddala, G. S. & Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics*, *61*, 631-652.

Maddala, G. S., Wu, S. & Liu, P. (2000). Do panel data rescue purchasing power parity (PPP) theory? In J. Krishnakumar & E. Ronchetti (Eds.), *Panel data econometrics: Future directions* (p. 35-51). North-Holland.

Manski, C. (1975). The maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, *3*, 205-228.

Manski, C. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, *55*, 357-362.

Mark, N. C. & Sul, D. (2012). When are pooled panel-data regression forecasts of exchange rates more accurate than the time-series regression forecasts? In J. James, I. W. Marsh & L. Sarno (Eds.), *Handbook of exchange rates* (p. 265-281). Wiley.

Marschak, J. & Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica*, *12*, 143-205.

Mátyás, L. (1997). Proper econometric specification of the gravity model. *The World Economy*, *20*, 363-368.

Mátyás, L. (1998). The gravity model: Some econometric considerations. *The World Economy*, *21*, 397-401.

Mátyás, L. (Ed.). (2017). *The econometrics of multi-dimensional panels* (1st ed.). Springer.

Mátyás, L. (Ed.). (2023). *The econometrics of multi-dimensional panels* (2nd ed.). Springer.

Mátyás, L. & Sevestre, P. (Eds.). (1992). *The econometrics of panel data*. Kluwer Academic Publishers.

Moffitt, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics*, *59*, 99–123.

Moon, H. R. & Phillips, P. C. B. (2000). Estimation of autoregressive roots near unity using panel data. *Econometric Theory*, *16*, 927-997.

Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society*, *Series B, 10*, 243-251.

Moran, P. A. P. (1950). Notes on continous stochastic phenomena. *Biometrika*, *37*, 17-23.

Moscone, F. & Tosetti, E. (2009). A review and comparison of tests of cross-section independence in panels. *Journal of Economic Surveys*, *23*, 528-561.

Mundlak, Y. (1961). Empirical production function free of management bias. *American Journal of Agricultural Economics*, *43*, 44-56.

Mundlak, Y. (1963). Estimation of production and behavioral functions from a combination of cross-section and time-series data. In C. F. Christ et al. (Eds.), *In measurement in economics: Studies in mathematical economics and econometrics in memory studies in mathematical economics and econometrics in memory of Yehuda Grunfeld* (p. 138-166). Stanford: Stanford University Press.

Mundlak, Y. (1964). Transcendental multiproduct production functions. *International Economic Review*, *5*, 273-284.

Mundlak, Y. (1978). On the pooling of time series and cross-section data. *Econometrica*, *46*, 69-85.

Nerlove, M. (2014). Individual heterogeneity and state dependence: From George Biddell Airy to James Joseph Heckman. *Œconomia*, *4-3*, 281-320.

Nerlove, M. (2023). Foreword. In L. Mátyás (Ed.), *The econometrics of multi-dimensional panels* (2nd ed.). Springer.

Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, *16*, 1-32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, *49*, 1417-1426.

Nijman, T. E., Verbeek, M. J. & van Soes, A. (1991). The efficiency of rotating-panel designs in an analysis-of-variance model. *Journal of Econometrics*, *49*, 373-399.

Oguledo, V. I. & MacPhee, C. R. (1994). Gravity models: A a reformulation and an application to discriminatory trade agreements. *Applied Economics*, *26*, 107-120.

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, *61*, 653-678.

Pedroni, P. (2004). Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric Theory*, *20*, 597-625.

Pesaran, M. H. (2004). *General diagnostic tests for cross section dependence in panels.* IZA DP No. 1240.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, *74*, 967-1012.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross section dependence. *Journal of Applied Econometrics*, *27*, 265-312.

Pesaran, M. H. (2021). General diagnostic tests for cross-sectional dependence in panels. *Empirical Economics*, *60*, 13-50.

Pesaran, M. H., Shin, Y. & Smith, R. (1999). Pooled mean group estimation of dynamic heterogeneousp panels. *Journal of the American Statistical Association*, *94*, 621-634.

Pesaran, M. H. & Smith, R. (1995). Estimating long-run relationships from dynamic heterogenous panels. *Journal of Econometrics*, *68*, 79-113.

Pesaran, M. H., Smith, R. & Im, K. S. (1996). Dynamic linear models for heterogeneous panels. In L. Mátyás & P. Sevestre (Eds.), *The econometrics of panel data:*

*A handbook of the theory with applications* (p. 145-195). Kluwer Academic Publications.

Pesaran, M. H. & Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, *161*, 182-202.

Pesaran, M. H., Ullah, A. & Yamagata, T. (2008). A bias-adjusted lm test of error cross-section independence. *Econometrics Journal*, *11*, 105-127.

Pesaran, M. H. & Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, *142*, 50-93.

Pesaran, M. H. & Zhao, Z. (1999). Bias reduction in estimating long-run relationships from dynamic heterogenous panels. In C. Hsiao, K. Lahiri, L.-F. Lee & M. H. Pesaran (Eds.), *Analysis of panels and limited dependent variables: A volume in honour of G. S. Maddala* (p. 297-322). Cambridge University Press.

Phillips, P. C. B. & Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dependenc. *Econometrics Journal*, *6*, 217-259.

Pinkse, J. (1999). *Asymptotic properties of Moran and related tests and testing for spatial correlation in probit models.* University of British Columbia.

Pinkse, J. (2004). Moran–flavored tests with nuisance parameters: examples. In L. Anselin, R. J. G. M. Florax & S. J. Rey (Eds.), *Advances in spatial econometrics* (p. 67-77). Springer.

Robertson, D. & Symons, J. (1992). Some strange properties of panel data estimators. *Journal of Applied Econometrics*, *7*, 175-189.

Sarafidis, V. & Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews*, *31*, 483-531.

Scheffé, H. (1956). Alternative models for the analysis of variance. *The Annals of Mathematical Statistics*, *27*, 251-271.

Scheffé, H. (1959). *The analysis of variance*. Wiley.

Serlenga, L. & Shin, Y. (2007). Gravity models of intra-EU trade: Application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, *22*, 361-381.

Shi, W. & Lee, L.-F. (2018). A spatial panel data model with time varying endogenous weights matrices and common factors. *Regional Science and Urban Economics*, *72*, 6-34.

Song, M. (2013). *Asymptotic theory for dynamic heterogeneous panels with crosssectional dependence and its applications.* Unpublished Manuscript, Department of Economics, Columbia University.

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, *38*, 311-323.

Verbeek, M. (2008). Pseudo-panels and repeated cross-sections. In L. Mátyás & P. Sevestre (Eds.), *The econometrics of panel data.* Springer.

Verbeek, M. & Nijman, T. E. (1996). Incomplete panels and selection bias. In (p. 449-490). Kluwer Academic Publishers.

Wansbeek, T. & Bekker, P. (1996). On IV, GMM and ML in a dynamic panel data model. *Economic Letters*, *51*, 145-152.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, *68*, 115–132.

Wooldridge, J. M. (2005). imple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, *20*, 39–54.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and test for aggregation bias. *Journal of the American Statistical Association*, *57*, 348-368.

Zhao, W. (2022). *The current and future of econometrics - the AMA interview with prof. Jeffrey Wooldridge.* SciEcon-AMA. Retrieved from https://medium.com/sciecon-ama/the-current-and-future-of-econometrics-ed30569e7edd

# Index