

Book review: Analysis of panel data: Second Edition, Cheng Hsiao, Cambridge University Press, Cambridge, United Kingdom, 2003, ISBN 0-521-81855-9, 382 pages, £21.95

By: David C. Ribar

Ribar, David. "Book Review: *Analysis of Panel Data*, 2nd Edition by Cheng Hsiao," *International Journal of Forecasting* 20:1 (2004), 142-3. doi:10.1016/j.ijforecast.2003.11.004

Made available courtesy of Elsevier:

http://www.elsevier.com/wps/find/journaldescription.cws_home/505555/description

*****Reprinted with permission. No further reproduction is authorized without written permission from Elsevier. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document.*****

Article:

Professor Cheng Hsiao has revised and updated his highly regarded 1986 monograph on the Analysis of Panel Data. Empirical analyses of panel data—data in which a cross-section of people, firms, countries or other observational units are followed over time—have become increasingly common in social science research over the last two decades. A large part of this growth is undoubtedly due to books, like Professor Hsiao's first edition, that painstakingly cataloged and examined methods for estimating panel models. The growth also stems from the increased availability of longitudinal data sets, the implementation of panel estimation routines into commercial software packages, and the development of new and more robust statistical methods. Professor Hsiao has contributed to many of these methodological developments and discusses them in his book. The structure of the second edition mostly follows that from the first. The book begins with an introduction that describes the advantages and challenges associated with using panel data. In the second chapter, the book reviews basic analysis-of-variance and analysis-of-covariance tests that can be performed with these data. The third chapter describes linear regression models with variable intercepts—the standard fixed and random effects regression models. The fourth chapter considers the complications that arise when the explanatory variables in a linear regression model include lags of the dependent variable. The fifth chapter describes simultaneous equation models and alternative identification strategies. The sixth chapter describes models in which the intercepts and coefficients vary across cross-section units. Discrete choice (mainly binary choice) models are presented in the seventh chapter. A new eighth chapter describes models with truncated, censored and selective dependent variables. Problems of incomplete data are discussed in the ninth chapter. A new chapter with miscellaneous topics follows, and a summary chapter concludes the volume. The strong point of the original monograph was its detailed treatment of different estimation methods. The second edition builds on this strength. For example, the substantially revised fourth chapter on dynamic models gives complete descriptions of alternative maximum likelihood, generalized least squares, instrumental variable and generalized method of moments (GMM) approaches for estimating random-effects specifications. For each technique, Professor Hsiao derives estimation equations and covariance matrices; he also discusses the properties of the estimators under different conditions. In the same chapter, he provides similarly comprehensive descriptions of the transformed likelihood, minimum distance and GMM approaches for estimating single equation and vector autoregressive models with fixed effects. Descriptions of several individual approaches as well as the entire section on VAR models are new to the second edition. Time series econometricians and forecasters will find many of the methods to be useful.

The sixth chapter on variable-coefficient models is the most strongly oriented toward forecasting applications. The original version of this chapter described methods for estimating and predicting coefficients that vary over time and over cross-section units. The chapter has been expanded and now also discusses dynamic random-coefficient models and mixed specifications of fixed- and random-coefficient models. A Bayesian method for estimating mixed models is presented along with an example of how this method has been applied to analyze and forecast electricity demand.

Material in the miscellaneous topics chapter covering long panel data sets and panel unit root tests will be of interest to time series analysts. The section on long panels discusses asymptotics when data sets become large in both their cross-section and time-series dimensions. The unit root section describes how time series can be combined to generate more powerful tests. The miscellaneous topics chapter also includes an overview of simulation estimators. These methods are useful in nonlinear models, such as dynamic binary choice and limited dependent variable specifications, with general patterns of serial correlation in the error terms. The chapter also describes special estimators for addressing errors-in-variables problems.

The new chapter on limited dependent variable models is also a significant addition, though possibly of less use to forecasters. The chapter briefly discusses random-effects estimators for tobit, censored regression and selectivity models; the random-effects specifications can be used to predict outcomes. The emphasis in the chapter, however, is on semi-parametric, fixed-effects methods. These methods provide robust estimates of subsets of parameters—but not all of the parameters—in the relevant models. The methods are tremendously valuable in examining empirical relationships and testing hypotheses but cannot be used to predict or forecast general outcomes.

If there is a drawback to the book, it is the scant attention that is given to random-effects estimators in the later chapters on non-linear models. The book points to computational complexity and restrictive assumptions as reasons for avoiding these estimators. While these are certainly issues to be considered, the book is unnecessarily discouraging. Faster computers have lowered some of the hurdles associated with these methods, as has the availability of commercial software, such as the aML package, that implements the relevant procedures. With respect to the restrictive assumptions, [Mroz, 1999] has shown that a flexible procedure for accounting for the distribution of random effects can be readily applied to a large family of models. Finally, there are many types of analyses, including forecasting, for which random-effects methods are essentially the only game in town. Random-effects methods are thoroughly discussed in the early chapters of the book that deal with linear models. It is surprising that they do not receive the same attention in the later chapters.

There are some other weaknesses from the original monograph that have not been addressed in the second edition. At several points, the book continues to rely on dated examples. It is disappointing that fresher examples were not incorporated into some of these sections. The examples involving multiple methods also generally do not include formal specification tests. Thus, they show that results can be sensitive to alternative procedures but do not indicate whether particular estimates can be rejected. Another broad criticism is that the writing is often impenetrable. The methods in this book should be part of every applied researcher's tool kit. Clearer and more intuitive descriptions would help to bring the methods to a wider audience.

These are minor flaws, however, in an otherwise excellent volume. Researchers will find that the insights that they gain from working through the book's tougher sections are well worth the effort. The book remains an indispensable and comprehensive reference for panel estimation methods.

References

Mroz, 1999. T.A. Mroz, Discrete factor approximations in simultaneous equation models: Estimating the impact of a dummy endogenous variable on a continuous outcome. *Journal of Econometrics* 92 (1999), pp. 233–274