

Panel Data Econometrics – Christian Schluter (Aix Marseille School of Economics)

Empirical work is challenging. Usually, we cannot measure or observe everything that is relevant for the determination of outcomes. Such unobservable heterogeneity then poses serious problems for the researcher if it is correlated with control variables (the omitted variables problem). Overcoming this problem using an instrumental variables strategy is often not feasible in practice since credible instruments are very difficult to find. However, if we observe the same individual over several periods, such panel data can offer a solution. We will develop and put into practice empirical methods for estimation and inference that exploit such a panel structure. After reviewing the classic approaches, we then proceed to discuss some important papers from the theoretical and applied econometrics literatures. An important setting in which panel data methods have become very popular is the estimation of causal policy effects of natural experiments. We will discuss the challenges and limitations of such difference-in-difference strategies. Finally, we turn to dynamic panel data models which require an altogether different estimation approach.

All methods will be illustrated with real-world data using \mathbb{R} , and several papers in applied labour economics will be discussed and replicated.

Part I. Classic Panel Data Methods.

Summary. Fixed effects (FE) modelling and estimation. Challenges to identification. The FE (“within”) estimator. First-differencing (FD) estimation. Random effects (RE) modelling. The dummy variable FE estimator. Hausman-type tests. FE/RE instrumental variable estimation.

Empirical applications in \mathbb{R} : Replication of Ruhm, C.J. (1996), “Alcohol policies and highway vehicle fatalities”, *Journal of Health Economics*.

Further illustrations (estimation of fixed effects):

Card, D., J. Heining, and P. Kline. (2013) “Workplace Heterogeneity and the Rise of West German Wage Inequality,” *The Quarterly Journal of Economics*.

Part II. Difference-in-difference (DiD) estimation and the estimation of causal policy effects (natural experiments).

Researchers are often interested in measuring the causal effect of a policy intervention. In situations in which some groups are exposed to the new policy but others are not (often referred to as a “natural” experiment), and individuals are observed before and after the policy intervention, panel data methods might become available.

Summary, Part II.1. The potential outcomes model. The causal treatment effect. Identification of causal policy effects. The DiD estimator. The need for placebo tests.

Illustrations: (i) Duflo, E. (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", *American Economic Review*. (ii) The Mariel boat lift that did and did not happen (Card and Krueger, "empirical strategies in labor economics", *Handbook of Labor Economics*).

Summary, Part II.2, challenges for statistical inference. While DiD is a popular estimation method in situations in which the researcher disposes over a panel, drawing inference and testing is challenging. We consider several such situations which arise when errors are correlated within a group or across time.

Moulton (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units", *The Review of Economics and Statistics*.

Donald and Lang (2007) "Inference with Difference-in-Differences and Other Panel Data", *The Review of Economics and Statistics*.

Bertrand, Duflo, Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?", *The Quarterly Journal of Economics*.

Part III. Dynamic panel data models.

These are models in which outcomes at time t depend on outcomes at time $t-1$ (e.g. employment at a firm). In this setting FE/FD estimators are invalid, thus requiring a different estimation strategy.

Summary. The (GMM) estimator of Arellano and Bond (1991). The weak instruments problem, consequences for estimation. The Blundell and Bond (1998) estimator.

Empirical applications in R: (i) Replication of Arellano, M. and S. Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *The Review of Economic Studies*. (ii) Replication of Blundell, R. and S. Bond (1998), "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*.