

Comments on: Panel Data Analysis—Advantages and Challenges

Manuel Arellano
CEMFI, Madrid
November 2006

This paper provides an impressive, yet compact and easily accessible review of the econometric literature on panel data analysis. Professor Cheng Hsiao has succeeded in surveying, in a coherent manner, classic results as well as more challenging recent developments on nonlinear models, cross-sectional dependence, and long time-series panels. The coverage of topics in the article reflects the breadth of Professor Hsiao's important contributions to panel data econometrics over many years.

In my comments I will focus on two aspects of recent work on nonlinear models. Firstly, I will discuss state dependence and dynamics from a treatment effect perspective. Secondly, I will consider the issue of choice of population framework and its implications for identifiability. Both comments are closely interconnected.

1. State dependence and treatment effects

Take a random sample of binary sequences (y_{i1}, \dots, y_{iT}) . Unit i chooses 1 or 0 in period t . This choice may depend on the choice in $t - 1$. The purpose is to measure this dependence.

The problem can be cast into the framework of potential outcomes:

$$y_{it} = \begin{cases} y_{it}(1) & \text{if } y_{it-1} = 1 \\ y_{it}(0) & \text{if } y_{it-1} = 0. \end{cases}$$

The “treatment” is y_{it-1} , and the potential outcomes are $y_{it}(1), y_{it}(0)$. The causal effect for person i is $y_{it}(1) - y_{it}(0)$. A measure of population state dependence is provided by the average treatment effect $E[y_{it}(1) - y_{it}(0)]$. We may also consider a conditional average given some exogenous variables or covariates. A discussion on identified bounds in this setting is in Manski (2006).

Because (y_{i1}, \dots, y_{iT}) is a sequence of outcomes, it is difficult to imagine a conceptual experiment that would justify a nonstructural treatment-effects formulation. One could assign initial conditions randomly and regard the rest of the time series as a vector of outcomes, but this is not typically the intention when seeking to measure the extent of state dependence. Thus, it is natural to regard the potential outcome representation as describing a structural decision rule.

1.1 Exclusion restrictions

I wish to discuss an aspect of the identifying content of time-varying covariates, which is standard in the context of linear models but has not received attention from a potential outcome perspective. I consider a nonparametric partial adjustment structural model that exploits exclusion restrictions in a time-varying strictly exogenous covariate $x_i^T = (x_{i1}, \dots, x_{iT})$.

The idea is that $\Pr[y_{it}(s)]$ ($s = 0, 1$) is conditional on x_i^T , but we would expect $\Pr[y_{it}(s)]$ to be more sensitive to x_{it} than to x' s from other periods. A drastic but convenient implementation of this notion is:

$$\Pr[y_{it}(s) \mid x_{i1}, \dots, x_{iT}] = \Pr[y_{it}(s) \mid x_{it}].$$

So, using x_{it-1} , we have the instrumental-variable (IV) assumption

$$\{y_{it}(0), y_{it}(1)\} \perp x_{it-1} \mid x_{it}.$$

As an example, think of y_{it} as smoking status and suppose that cigarette prices x_{it-1} and x_{it} are set exogenously. The IV assumption says that, given current prices, (past smoking-induced) potential smoking outcomes are independent of past prices. This is the type of situation discussed in the local average treatment effect (LATE) literature (c.f. Imbens and Angrist, 1994).

Using a potential outcome formulation for y_{it-1} and a binary x_{it-1} :

$$y_{it-1} = \begin{cases} y_{it-1}^{[1]} & \text{if } x_{it-1} = 1 \\ y_{it-1}^{[0]} & \text{if } x_{it-1} = 0, \end{cases}$$

we can distinguish between compliers (those induced to quit smoking by changing x_{it-1} from 0 to 1: $y_{it-1}^{[0]} - y_{it-1}^{[1]} = 1$), stayers, and defiers (those with $y_{it-1}^{[0]} - y_{it-1}^{[1]} = -1$). If we rule out defiers, the distributions of $y_{it}(0)$ and $y_{it}(1)$ for compliers are point identified:

$$\Pr(y_{it}(s) \mid y_{it-1}^{[0]} - y_{it-1}^{[1]} = 1, x_{it}) \quad (s = 0, 1).$$

Given this, we can get measures of state dependence (addiction) and price effects on smoking. Note that we have defined two different sequences of potential outcomes, $y_{it}^{[s]}$ and $y_{it}(s)$.

1.2 Exogeneity

An alternative conditional exogeneity assumption is

$$\{y_{it}(0), y_{it}(1)\} \perp y_{it-1} \mid x_{it}.$$

This is a strong assumption because y_{it-1} is not randomly assigned. A linear version of this is the standard partial adjustment model without serial correlation.

1.3 Fixed effects

The previous discussion can be thought of as being conditional on time-invariant observable covariates. The panel literature has emphasized parametric situations where the results hold conditional on a time-invariant unobserved effect α_i :

$$\{y_{it}(0), y_{it}(1)\} \perp y_{it-1} \mid x_{it}, \alpha_i \tag{1}$$

or

$$\{y_{it}(0), y_{it}(1)\} \perp x_{it-1} \mid x_{it}, \alpha_i,$$

thus allowing for “fixed-effects endogeneity” of y_{it-1} or x_{it-1} .

In situations of this kind, we only have fixed- T point identification for particular objects in certain models. An example of (1) is the binary autoregressive formulation

$$y_{it}(s) = 1(\gamma s + \alpha_i + v_{it} \geq 0) \quad (s = 0, 1) \quad (2)$$

where v_{it} are iid across i and t , independent of α_i , with logit or probit *cdf* F .

The average treatment effect in this case is

$$\phi \equiv E[y_{it}(1) - y_{it}(0)] = E_{\alpha_i}[F(\gamma + \alpha_i) - F(\alpha_i)].$$

There is point identification of γ for logit if $T \geq 4$, but not for probit, although the identified set for γ seems to be small (Honoré and Tamer, 2006). There is set identification for ϕ for both logit and probit.

1.4 Unobserved heterogeneity and identification

Take just one individual time series and think of it as the realization of a well defined, suitably stable, but individual-specific, stochastic process. A descriptive measure of unit’s i persistence is the first-order autocorrelation:

$$\begin{aligned} \rho_i &= \mathcal{P}_i(y_{it} = 1 \mid y_{it-1} = 1) - \mathcal{P}_i(y_{it} = 1 \mid y_{it-1} = 0) \\ &= \text{plim}_{T \rightarrow \infty} \left(\frac{1}{T_1} \sum_{y_{it-1}=1} y_{it} - \frac{1}{T_0} \sum_{y_{it-1}=0} y_{it} \right) \end{aligned}$$

where $T_1 = \sum_{t=2}^T y_{it-1}$ and $T_0 = \sum_{t=2}^T (1 - y_{it-1})$. On the other hand, a time-series average of causal effects is:

$$r_i = \text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T [y_{it}(1) - y_{it}(0)].$$

In general, ρ_i and r_i are different concepts. Note that

$$y_{it} = [y_{it}(1) - y_{it}(0)] y_{it-1} + y_{it}(0),$$

so that we have $r_i = \rho_i$ if $[y_{it}(0), y_{it}(1)]$ are independent of y_{it-1} over time. For example, this is true for the binary fixed-effect autoregressive model (2).

A cross-sectional measure of persistence in a two-period panel is

$$\begin{aligned} \pi_t &= \Pr(y_{it} = 1 \mid y_{it-1} = 1) - \Pr(y_{it} = 1 \mid y_{it-1} = 0) \\ &= \text{plim}_{N \rightarrow \infty} \left(\frac{1}{N_1} \sum_{y_{it-1}=1} y_{it} - \frac{1}{N_0} \sum_{y_{it-1}=0} y_{it} \right) \end{aligned}$$

where $N_1 = \sum_{i=1}^N y_{it-1}$ and $N_0 = \sum_{i=1}^N (1 - y_{it-1})$. If ρ_i and π_t are constant for all i and t , they will coincide, but not otherwise.

The microeconomic literature on “genuine versus spurious” state dependence has been concerned with approximating summary measures of ρ_i from short panels. This may still be a descriptive pursuit, although $E(\rho_i)$ is arguably more informative than π_t because it distinguishes between cross-sectional unobserved heterogeneity and unit-specific time-series persistence.

Even for some of these descriptive objects, we lack point identification under fixed T . However, the fact that ρ_i or cross-sectional functionals of it are not point identified from a fixed- T perspective, reflects a limitation of this perspective when T is statistically informative. I now turn to discuss this problem.

2. Population framework and identification

Fixed T identification may be problematic because it rules out statistical learning from individual time series data. For micro panels of moderate time dimension, approximate solutions to the incidental parameter problem from a time-series perspective (reviewed in Arellano and Hahn, 2007) are a promising avenue for progress.

In this literature three different approaches can be distinguished. One approach is to construct and analytical or numerical bias correction of a fixed effects estimator. A second approach is to consider estimators from bias corrected moment equations. The third one is to consider estimation from a bias corrected objective function relative to some target criterion. The latter is particularly attractive for its simplicity, specially in models with multiple fixed effects.

By way of illustration, suppose a likelihood model for independent data with common parameter θ and a potentially vector-valued individual effect α_i , where the log likelihood for individual i is $\sum_{t=1}^T \ell_{it}(\theta, \alpha_i)$. A modified concentrated likelihood that produces estimates of the common parameters with bias of order $1/T^2$ or less is given by

$$L_M(\theta) = \sum_{i=1}^N \left[\sum_{t=1}^T \ell_{it}(\theta, \hat{\alpha}_i(\theta)) + \frac{1}{2} \ln \det H_i(\theta) - \frac{1}{2} \ln \det \Upsilon_i(\theta) \right]$$

where $\hat{\alpha}_i(\theta)$ is the maximum likelihood estimate of α_i for given θ , $H_i(\theta) = -\sum_{t=1}^T \partial^2 \ell_{it}(\theta, \hat{\alpha}_i(\theta)) / \partial \alpha_i \partial \alpha_i'$, and $\Upsilon_i(\theta) = \sum_{t=1}^T q_{it}(\theta) q_{it}(\theta)'$, where $q_{it}(\theta) = \partial \ell_{it}(\theta, \hat{\alpha}_i(\theta)) / \partial \alpha_i$. Thus, as discussed in Arellano and Hahn (2007), the adjustment depends exclusively on the sample Hessian $H_i(\theta)$ and the sample outer product of score term $\Upsilon_i(\theta)$.

3. Final remarks

In panel data analysis, there is a choice of population framework, which may lead to conflicting identification arrangements. In situations of this kind, there is much to be learned from research on both partial identification and estimability issues.

References

- ARELLANO, M. AND J. HAHN (2007): “Understanding Bias in Nonlinear Panel Models: Some Recent Developments”. In: R. Blundell, W. Newey, and T. Persson (eds.): *Advances in Economics and Econometrics, Ninth World Congress*, Cambridge University Press.
- HONORÉ, B. E. AND E. TAMER (2006): “Bounds on Parameters in Panel Dynamic Discrete Choice Models”, *Econometrica*, 74, 611–629.
- IMBENS, G. W. AND J. ANGRIST (1994): “Identification and Estimation of Local Average Treatment Effects”, *Econometrica*, 62, 467–475.
- MANSKI, C. F. (2006): “Two Problems of Partial Identification with Panel Data”, 13th Conference on Panel Data, Cambridge.