Comments on "Two Problems of Partial Identification with Panel Data" by Charles Manski

Manuel Arellano
Cambridge, July 8, 2006

1. Summary

- Two problems of partial identification are considered: 1) Panel distribution with missing data, and 2) state dependence in binary choice. I will focus on state dependence.
- Take a random sample of binary sequences $(y_{j1}, ..., y_{jT})$.Unit j chooses 1 or 0 in t. This choice may depend on the choice in $t-1$. The purpose is to measure this dependence.
- The problem is cast into the framework of potential outcomes:

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

  The treatment is y_{jt-1} and the potential outcomes are $y_{jt}(1), y_{jt}(0)$. The causal effect for person j is $y_{jt}(1) - y_{jt}(0)$.
- There is no personal state dependence if $y_{jt}(1) - y_{jt}(0) = 0$ with probability one.
- There is no population state dependence if the average treatment effect vanishes

$$E[y_{jt}(1) - y_{jt}(0)] = 0.$$

- Absence of personal state dependence is untestable, but since $E[y_{jt}(1)]$ and $E[y_{jt}(0)]$ are set identifiable, the hypothesis of no population state dependence may be testable.

Outline

- I will relate the potential outcome formulation to some of the themes in the dynamic panel data literature. I make three points:
  (a) Time-varying covariates as exclusion restrictions.
  (b) Fixed effects.
  (c) What if T is not fixed?

2. Treatments that are outcomes

- State dependence is a structural notion, not a treatment effect notion.
- Because $(y_{j1}, ..., y_{jT})$ is a sequence of outcomes it is difficult to imagine a conceptual experiment that would justify a non-structural treatment-effects formulation.
- One could assign initial conditions randomly and regard the rest of the time series as a vector of outcomes, but this does not seem to be the intention here.

3. Structural state dependence

- The potential outcome representation may describe a structural decision rule.
- Results hold conditional on covariates. The paper considers identification regions using a conditional stationarity assumption.
- I wish to discuss an aspect of the identifying content of time-varying covariates.
- I consider a nonparametric partial adjustment structural model that exploits exclusion restrictions in a time-varying strictly exogenous covariate $x_j^T = (x_{j1}, ..., x_{jT})$.
- The idea is that $\Pr[y_t(s)]$ is conditional on x_j^T, but we would expect $\Pr[y_t(s)]$ to be more sensitive to x_{jt} than to $x's$ from other periods. A drastic but convenient implementation of this notion is:

$$\Pr[y_t(s) \mid x_{j1}, ..., x_{jT}] = \Pr[y_t(s) \mid x_{jt}]$$

- So, using x_{jt-1}, we have the instrumental-variable assumption

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

- As an example, think of y_{jt} as smoking status, and suppose that cigarette prices x_{jt-1} and x_{jt} 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 LATE literature (Imbens & Angrist, 1994).
- Using a potential outcome formulation for y_{jt-1}:
$$y_{jt-1} = \begin{cases} y_{jt-1}^{[1]} & \text{if } x_{jt-1} = 1 \\ y_{jt-1}^{[0]} & \text{if } x_{jt-1} = 0, \end{cases}$$
we can distinguish between compliers (those induced to quit smoking by changing x_{jt-1} from 0 to 1: $y_{jt-1}^{[0]} - y_{jt-1}^{[1]} = 1$), stayers, and defiers ($y_{jt-1}^{[0]} - y_{jt-1}^{[1]} = -1$).
- If we rule out defiers, the distributions of $y_{jt}(0)$ and $y_{jt}(1)$ for compliers are point identified:
$$P\left[y_{jt}(s) \mid y_{jt-1}^{[0]} - y_{jt-1}^{[1]} = 1, x_{jt}\right] \quad (s = 0, 1).$$
Given this, we can get measures of state dependence (addiction) and price effects on smoking.

Exogeneity

- An alternative conditional exogeneity assumption is
$$\{y_{jt}(0), y_{jt}(1)\} \perp y_{jt-1} \mid x_{jt}.$$
This is a strong assumption because y_{jt-1} is not randomly assigned.
- A linear version of this is the standard partial adjustment model without serial correlation.

Fixed effects

- The previous discussion can be thought as being conditional on time-invariant covariates. The panel literature has emphasized the case where the results hold conditional on a time-invariant unobserved effect α_j:
$$\{y_{jt}(0), y_{jt}(1)\} \perp y_{jt-1} \mid x_{jt}, \alpha_j \tag{1}$$
or
$$\{y_{jt}(0), y_{jt}(1)\} \perp x_{jt-1} \mid x_{jt}, \alpha_j$$
which allows for "fixed-effects endogeneity" of y_{jt-1} or x_{jt-1}.

- In situations of this kind, we only have fixed-T point identification for particular objects in certain models. An example of (1) is
$$y_{jt}(s) = 1(\gamma s + \alpha_j + v_{jt} \geq 0)$$
where $-v_{jt}$ are iid across j and t, independent of α_j, with *cdf* F.

- The average treatment effect in this case is
$$\phi \equiv E[y_{jt}(1) - y_{jt}(0)] = E_{\alpha_j}[F(\gamma + \alpha_j) - F(\alpha_j)].$$

- 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.

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 j persistence is the first-order autocorrelation:

$$\begin{aligned} \rho_j &= \mathcal{P}_j\left(y_{jt}=1 \mid y_{jt-1}=1\right)-\mathcal{P}_j\left(y_{jt}=1 \mid y_{jt-1}=0\right) \\ &= \operatorname*{plim}_{T\to\infty}\left(\frac{1}{T_1}\sum_{y_{jt-1}=1} y_{jt}-\frac{1}{T_0}\sum_{y_{jt-1}=0} y_{jt}\right) \end{aligned}$$

  where $T_1=\sum_{t=2}^{T} y_{jt-1}$ and $T_0=\sum_{t=2}^{T}\left(1-y_{jt-1}\right)$.

- A time-series average of causal effects is:

$$r_j=\operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left[y_{jt}(1)-y_{jt}(0)\right].$$

- In general ρ_j and r_j are different concepts.

- Note that

$$y_{jt}=\left[y_{jt}(1)-y_{jt}(0)\right]y_{jt-1}+y_{jt}(0),$$

  so that we have $r_j=\rho_j$ if $\left[y_{jt}(0), y_{jt}(1)\right]$ are independent of y_{jt-1} over time.

- For example, this is true for the binary autoregressive model with fixed effects.

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

$$\begin{aligned}\pi_t &= \Pr\left(y_{jt} = 1 \mid y_{jt-1} = 1\right) - \Pr\left(y_{jt} = 1 \mid y_{jt-1} = 0\right) \\ &= \operatorname*{plim}_{n\to\infty}\left(\frac{1}{n_1}\sum_{y_{jt-1}=1} y_{jt} - \frac{1}{n_0}\sum_{y_{jt-1}=0} y_{jt}\right)\end{aligned}$$

  where $n_1 = \sum_{j=1}^{n} y_{jt-1}$ and $n_0 = \sum_{j=1}^{n}\left(1 - y_{jt-1}\right)$.

- If ρ_j and π_t are constant for all j and t, they coincide, but not otherwise.

- The microeconometric literature on "genuine versus spurious" state dependence has been concerned with approximating summary measures of ρ_j from short panels.

- This may still be a descriptive pursuit, although $E\left(\rho_j\right)$ 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.

- The fact that ρ_j 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.

Concluding remarks

- 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, 2006) are a promising avenue for progress.
- 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.