Grouped data, lagged variables

Christopher A. Sims
Princeton University
sims@princeton.edu

May 9, 2018

The simplest case

$$y_{it} = \rho y_{i,t-1} + \nu_i + \varepsilon_{it} \ : .$$

Assume $\varepsilon_{it} \sim N(0, \sigma^2)$, uncorrelated with $\{y_{is}, s < t\}$, i.e. that ε_{it} is the innovation in y_{it}.

- Looks like what we've already discussed, just with lagged y playing the role of X_{it}.

- $y_{i,t-1}$ is predetermined, not exogenous. If T is large, we know that OLS within groups is consistent and asymptotically has the usual limiting covariance matrix.

- If T is small while M (number of groups) is large, "fixed effects" estimators of coefficients on an exogenous X are consistent and have the usual distribution theory, despite the estimators of ν_i remaining noisy as $M \to \infty$.

- But with X not strictly exogenous, and in particular with lagged y as explanatory variable, "fixed effects" estimation is no longer consistent as $M \to \infty$ with T fixed.

Why is "fixed effects" inconsistent?

- Fixed effects is just the within regression. That is, it amounts to estimating the model using deviations from group means for all variables:

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)\beta + \varepsilon_{it} - \bar{\varepsilon}_i$$

- With strictly exogenous X, X_{it} is uncorrelated with ε_{is} at any lead or lag, so in this within regression, the right-hand-side variable is uncorrelated with the residual.

- When $X_{it} = y_{i,t-1}$, though, the model implies that ε_{is} for $s < t$ are correlated with y_{it}, so the residual in the transformed regression *is* correlated with the right-hand-side variable.

How to proceed: likelihood-based approach

- The model provides only conditional densities for $y_{it} \mid y_{i,t-s}, s > 0$, $t = 2, \dots, T$. Our data also includes M observations on $y_{i,1}$. A likelihood function requires a model for $y_{i1}, i = 1, \dots, M$.

- The short-cut used in single time series — conditioning on the initial observation — doesn't work here, because here the density for the initial observation does not become dominated by the rest of the likelihood as sample size increases.

One possibility: assume stationarity

Then

$$y_{1i} \mid \nu_i \sim N\left(\frac{\nu_i}{1-\rho}, \frac{\sigma^2}{1-\rho^2}\right) \ .$$

- We still need a distribution for ν_i. Assuming a flat prior on ν_i doesn't work, because that implies high variance across groups for the group means of y_{it}, and the data may not be consistent with that.

- What does work is to assume ν_i are drawn from, say, $N(\mu_\nu, \sigma_\nu^2)$. The data can then, through the likelihood, inform us about μ_ν and σ_ν^2.

Stationarity often unattractive in panel data

- For example, i might index individuals, and t might index "Months unemployed between ages 16-20, 21-25, 26-30, etc.

- In this case we can assume

$$\begin{bmatrix} y_{i1} \\ \nu_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_y \\ \mu_\nu \end{bmatrix}, \Sigma_{y\nu} \right) .$$

- This introduces six new parameters to the model. The stationarity assumption is a special case that introduces only two new parameters (the mean and variance of the ν_i distribution).

A pitfall

- If this approach is applied to a collection of time series — for example the time path of log GDP for some set of countries — it is important that the data not be "preprocessed" in certain ways.
- Obviously if means have been removed, it makes no sense to treat μ_y as a free parameter.
- But a less obvious mistake is to use data in "index" form, where to make the data more easily comparable in a graph, they have been normalized to make all the country values the same at some date.
- Then, e.g., if $\rho = 1$, $\nu_i = 0$, the original model implies that the conditional expectation of y_{it} given y_{i1} is y_{i1}, whereas if t is the index normalization date, this will clearly not be true.

Another approach: IV

- Fixed effects gets rid of the ν_i term by taking deviations from group means. Can also do this by first-differencing the data:

$$y_{it} - y_{i,t-1} = \Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta \varepsilon_{it}$$

- Of course OLS on this equation would not work. (Why?)

- What might be an instrument?

Using lagged y's as instruments

- $y_{i,t-2}$ or $\Delta y_{i,t-2}$ are uncorrelated with $\Delta\varepsilon_{it}$, passing the first test of an instrument.

- But are they correlatehd with the "included endogenous" variable?

- If the model is correct and $\rho = 1$,

$$y_{it} = \sum_{s=0}^{t-1} \varepsilon_{i,t-s} + y_{i0} \,.$$

Thus $\Delta y_{i,t-1}$ is uncorrelated with all values of y_s and Δy_s for $s < t-1$. In this case lagged values of y and Δy are useless as instruments. Even if ρ is just close to one, the part of Δy_{t-1} that can be explained with lagged values of y and Δy will be small, so the instruments will be weak.

Recap of dynamic panel regression

With constants c_i varying across groups, short time series, model

$$y_{it} = c_i + y_{i,t-1}\rho + \varepsilon_{it}\ ,$$

we can write the likelihood for all the observables $\{y_{i0}, \ldots, y_{iT}\}$ as

$$\prod_{i=1}^{N} q(c_i, y_{i0}) \prod_{t=1}^{T} p(y_{it} \mid c_i, y_{i,t-1})\ .$$

We use the assumption that data are independent across i and that dependence of y_{it} on the past is entirely through $y_{i,t-1}$.

Recap of dynamic panel regression

$$\prod_{i=1}^{N} q(c_i, y_{i0}) \prod_{t=1}^{T} p(y_{it} \mid c_i, y_{i,t-1}) \, .$$

We use the assumption that data are independent across i and that dependence of y_{it} on the past is entirely through $y_{i,t-1}$. And the usual assumption is

$$p(y_{i,t} \mid c_i, y_{i,t-1}) = \frac{1}{\sigma} \phi \left(\frac{y_{it} - \rho y_{i,t-1}}{\sigma} \right) \, .$$

Why an exogenous variable makes things more complicated

- With no x_t, we can use the model for the conditional distributions, so the only complication is specifying the marginal joint distribution of c_i, y_{i0}.
- This approach would also work for x's indexed only by i, though of course their effects are not identified if we also allow unconstrained group constants c_i and do not assume them uncorrelated with the x_i's.

Why an exogenous variable makes things more complicated

- But for x's indexed i, t, the model provides only a distribution for

$$y_{it} \mid c_i, \{y_{i,t-s-1}, x_{i,t-s}, s = 0, \ldots, \infty\} \ .$$

- Even if x_{it} is strictly exogenous (uncorrelated with ε_{it} at all leads and lags), to follow the strategy we used without x's would require forming a joint distribution for $y_{i0}, c_i, \{x_{is}, s = 1, \ldots, T\}$.

- Giving this an arbitrary (but probably joint normal) density q to form the likelihood conditional on $x_{i1}, \ldots, x_{iT}$ could work well only if T and the dimension of x_{it} are small relative to the number of groups.

Modeling the x's

- Any approach to estimating the model with x's involves making assumptions on their distribution and their oint distribution with c_i and y_{i0}.

- A straightforward approach is to extend the dynamic model to include x_{it}.

- New model:

$$\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} = \begin{bmatrix} c_{iy} \\ c_{ix} \end{bmatrix} + \rho \begin{bmatrix} y_{i,t-1} \\ x_{i,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{iyt} \\ \varepsilon_{ixt} \end{bmatrix} .$$

A panel VAR

We can give the system a form that looks like the original single equation model by using the notation

$$z_{it} = \begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} , \qquad z_{it} = c_i + \rho z_{i,t-1} + \varepsilon_{it} \, .$$

Now the strategies we discussed for the single-equation model with no exognous variables can be applied here — using the stationary unconditional distribution for z_{i0}, postulating an unconstrained joint normal distribution for c_i and z_{i0}, etc., though of course if z is very long, implementing these strategies may be computationally demanding.