What does the Arellano-Bond estimator do?

Class Notes

Manuel Arellano

September 20, 2018

Summary Causal questions in economics typically cannot be answered experimentally. Rather, when interested in quantifying the effect of a policy or some other variable on an outcome, one often resorts to calculations under some form of conditional exogeneity. That is, one compares differences in outcomes within groups of units that are as similar as possible, in hopes that policy variation within such comparison groups is as good as random.

Policies not only vary across units but also over time. One can then use repeated observations on the same unit (panel data) as a reference group. A popular implementation of this notion is a ("fixed effects") method in which the comparison group consists of units that share the same exposure to the policy on average.

However, future policy exposure is likely to depend on past outcomes, even if the current policy level remains independent of present and future potential outcomes. In such situation units with the same average policy level (over a small number of periods) are not a valid comparison group for causal calculations. Instead, one would like to rely on a comparison group of units that share the same exposure to the policy in the past but not necessarily in the future.

The Arellano-Bond method is a simple implementation of this notion. It proceeds by first taking the data in deviations with respect to future means to remove fixed effects. Next, each policy deviation is projected period-by-period on the available history of past policy, so as to extract the exogenous variation in the policy deviations. Finally, the Arellano-Bond estimate is obtained as the coefficient in a pooled regression of the forward deviations of outcomes on the predicted policy deviations.

Summing-up, the Arellano-Bond estimator exploits time patterns in panel data to estimate the economic response to a change in a policy or other variable, while controlling for permanent unobserved confounding variation. An estimate is formed under a weak form of sequential exogeneity in which, net of individual effects and other covariates, current policy exposure is independent of present and future potential outcomes but not of past outcomes. An important special case in which exogeneity can only be sequential by construction is when the interest is in the effect of past outcomes on current outcomes.

Details More specifically, given panel data on an outcome y_{it} and a policy x_{it} for $i = 1, ..., N$ units observed $t = 1, ..., T$ time periods (ignoring other covariates), consider the fixed effects model

$$y_{it} = \beta x_{it} + \alpha_i + v_{it}$$

where x_{it} may be correlated with the unobservable fixed effect α_i. If x_{it} is uncorrelated with all the errors $v_{i1}, ..., v_{iT}$, the fixed effects method teases out an effect of the policy on the outcome by relying on individual variation in the policy over time relative to the individual-specific average policy exposure $\overline{x}_i = (x_{i1} + ... + x_{iT}) / T$; for example, by regressing y_{it} on x_{it} and $\overline{x}_i$.

Convenient for our purposes, the fixed effects estimator is also OLS in the following equation where fixed effects have been removed by taking the data in deviations with respect to future means:

$$y_{it}^* = \beta x_{it}^* + v_{it}^*$$

where x_{it}^* is the forward deviation $x_{it} - (x_{i,t+1} + ... + x_{iT}) / (T - t)$ scaled to equalize variances, etc. However, if x_{it} is only uncorrelated with $v_{it}, ..., v_{iT}$, OLS does not work because x_{it}^* and v_{it}^* are correlated. Instead, each policy deviation can be projected period-by-period on the available history of past policy, so as to extract their exogenous variation. Thus, we get a sequence of $T - 1$ first-stage OLS fitted values; for example, if $T = 4$ we get:

$$\begin{aligned}
\widehat{x}_{i1}^* &= \widehat{\pi}_{11} x_{i1} \\
\widehat{x}_{i2}^* &= \widehat{\pi}_{21} x_{i1} + \widehat{\pi}_{22} x_{i2} \\
\widehat{x}_{i3}^* &= \widehat{\pi}_{31} x_{i1} + \widehat{\pi}_{32} x_{i2} + \widehat{\pi}_{33} x_{i3}.
\end{aligned}$$

Finally, the Arellano-Bond estimate is obtained as the OLS coefficient in the pooled regression:

$$y_{it}^* = \beta \widehat{x}_{it}^* + error_{it}.$$

In the language of potential outcomes, our model deals with potential outcomes of the form $y_{it}(x) = \beta x + \alpha_i + v_{it}$, which are either taken to be mean independent of the full sequence of actual policies $(x_{i1}, ..., x_{iT})$ given α_i, or in the weaker sequentially exogenous case, only mean independent of past and present policies $(x_{i1}, ..., x_{it})$ given α_i.