PSC 504: Panel data and fixed effects

Matthew Blackwell

3/07/2013

Identification in panel data

Ignorability conditional on unit

- Sometimes we do not have all of the possible confounders for the effect on the treatment on the outcome, so it's not plausible that ignorability holds. Another way to think about this: even conditional on the covariates, the treatment isn't quite randomly assigned. Today we're going to start thinking about how to estimate effects when the usual ignorability assumption doesn't hold. Generally, we will look to other sources of variation to identify our effect.

- When we observe the same units over time, we have new ways to identify the effect of the treatment on the outcome. Note that simply having panel data does not identify an effect, but it does allow us to rely on different identifying assumptions.

- Generally, we talk about **panel data** and **time-series cross-sectional data** in political science. Panel data usually refers to situations where the number of time periods is quite short and the number of units quite high. The NES panel is like this: 2000 respondent asked questions at various points in time over the course of an election (or multiple elections). TSCS data, on the other hand, has fewer units and many time periods. The usual application is something like U.S. states over time or Western European countries over time. For the most part, the issues of causality are the same for these two types of data, so I will refer to them both as panel data.

- The basic idea is that ignorability doesn't hold, conditional on the observed covariates, $Y_{it}(a) \not\perp\!\!\!\perp A_{it}|X_{it}$, but ignorability might hold conditional on some unobserved, time-constant, variable:

$$Y_{it}(a) \perp\!\!\!\perp A_{it}|X_{it}, U_i.$$

- This type of ignorability gives us an insight: within units, effects are identified. This is because, even if U_i is unobserved, it is held constant within a unit. Thus, by performing analyses within the units, we can control for this unobserved heterogeneity.

Fixed effects estimators

Basic linear fixed-effects model

- Let's say we have units $i = 1, \ldots, N$ and time periods $t = 1, \ldots, T$ with $T \geq 2$, with Y_{it} be the outcome for unit i in period t. Let A_{it} be the treatment, similarly defined. We have a set of covariates in each period, as well, X_{it}, which we define as being "prior" to A_{it}. Thus, we have something like this:

$$\begin{tikzcd} & X_t \arrow[dl] \arrow[dr] & \\ A_t \arrow[rr] & & Y_t \end{tikzcd}$$

- Let U_i be an unobserved unit effects that are functions of observed and unobserved baseline confounders, W_i and η_i, respectively. All of these variables are causally prior to $t = 1$.

- We need some notation to indicate the history of some variables up to time t. We will write these as $\underline{A}_{it} = (A_1, \ldots, A_t)$.

- The typical way that we write a fixed effect model is as a linear regression:

$$Y_{it} = X_{it}'\beta + \tau(\underline{A}_{it}) + U_i + \varepsilon_{it}$$

- A couple of notes: first, we now write τ as a function since, there are many possible effects that could be produced by the treatment history. Imagine the case with just two time periods. Now, even with a binary treatment in each time period, there are four possible combinations of the treatment history: (0,0), (1,0), (0,1), and (1,1). Thus, while we usually have $\tau(\underline{A}_{it}) = \tau A_{it}$, we might also have $\tau(\underline{A}_{it}) = \tau_1 A_{it} + \tau_2 A_{i,t-1}$.

- Of couse, with the above regression formula, the key assumptions will be on the relationship between U_i and ε_{it}. When we do not have any lagged dependent variables in X_{it}, the current practice is to rely on what is called a **strict exogeneity** assumption:

$$E[\varepsilon_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = 0$$

- This combining this with the above regression, we get the following conditional expectation function:

$$E[Y_{it} | \underline{X}_{iT}, \underline{A}_{iT}, U_i] = X_{it}'\beta + \tau(\underline{A}_{it}) + U_i$$

- Once we fix one of the unobserved unit effects at 0 (or fix the mean at 0), we treat the rest of the unit terms as constants and β and $\tau(\underline{a}_t)$ are identified. If we have the usual effects, $\tau(\underline{A}_{it}) = \tau A_{it}$, then the typical way to estimate this fixed effect model is using the so-called "within" estimator:

$$(Y_{it} - \bar{Y}_i) = (X_{it} - \bar{X}_i)'\beta + \tau(A_{it} - \bar{A}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

- Here, the $\bar{Y}_i$ refers to the mean of that variable for unit across time. The logic here is fairly straightforward: since the unobserve effect is constant over time, subtracting off the mean also subtracts that unobserved effect. This also demonstrates why the assumption of the fixed effects being time-constant is so important. Otherwise, there would still be a residual function of them caught up in the error term above.

- Note that first differencing has the same effect as the within estimator: since the unit effects are time constant, it doesn't matter if we subtract off the value from the last period or the mean from all periods.

- The informal proof for this is that under strict exogeneity, the mean-differenced errors are uncorrelated with the treatment or regressors from **any** time period. Thus, the mean-differenced treatment and covariates must also be uncorrelated with the mean-differenced errors. Thus, we can identify all the effects.

Lagged dependent variables

- The above strict exogeneity assumption is quite strong. Let's think about the relationship between economic interdependence between countries ($A_{it} = 1$ if a county dyad is interdependent in period t) and conflict severity between countries. Then the strict exogeneity assumption implies that a shock to the conflict severity is uncorrelated with future values of conflict severity, economic interdendence and any covariate we include in the model. Thus, this assumption rules out the possibility of lagged dependent variables.

- To see why this is the case, imagine that $Y_{i,t-1}$ was included in the covariates, X_{it}. Under strict exogeneity, the error must be uncorrelated with past and future values of X_{it}. Thus, ε_{it} must be uncorrelated with $X_{i,t+1}$, which by construction includes Y_{it}. But of course they cannot be uncorrelated because the error is a component of that value of the dependent variable!

- A weaker assumption that is used in these cases is that of **sequential exogeneity**, which assumes that the errors are only uncorrelated with past values. That is, we assume:

$$E[\varepsilon_{it}|\underline{X}_{it}, \underline{A}_{it}, U_i] = 0.$$

- Unfortunately, this assumption alone does not identify a fixed effect regression like the strict exogeneity does. This is because the mean-differenced (or first-differenced) LDV will in general be correlated with the mean- or first-differenced error. In an extremely simple case (the only covariate is a LDV), you can see this:

$$(Y_{it} - Y_{i,t-1}) = \beta(Y_{i,t-1} - Y_{i,t-2}) + \tau(A_{it} - A_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

- Obviously, the $Y_{i,t-1}$ induces correlation with the $\varepsilon_{i,t-1}$. In order to estimate these types of models (often called dynamic panel models), you need to use an instrument. Fortunately, they are generally given by sequential exogeneity, combined with a few assumptions on the time-series (no serial correlation and no correlation between the initial outcome and future errors). Under these conditions, the level of the two-period lag, $Y_{i,t-2}$, is correlated with the first-differenced LDV, $(Y_{i,t-1} - Y_{i,t-2})$, but uncorrelated with the first-differenced error $(\varepsilon_{it} - \varepsilon_{i,t-1})$. Thus, we can use this variable as an instrument for the LDV in order to identify the parameters of the model.

- The IV approach to dynamic panel models is quite robust, with a large literature finding new/better sets of instruments to use. Unfortunately, they all share similar weaknesses: the linear modeling assumption and the strong assumptions on the error terms. This approach also focuses on a very specific type of effect: the contemporaneous effect. More general effects are not generally identified using this approach.

Treatment effects in fixed effects causal models

Basic effects

- When moving from static to dynamic treatments, we need to adjust our notions a bit. Now, we have the potential for more general effects. Let's refer to $\underline{a}_T$ as a possible treatment regime/history. It can take a value in the basis set $M_T \subseteq \{0, 1\}^T$. We will assume that the potential outcomes for some unit depend on the regime for that unit only, which is equivalent to SUTVA for units. Importantly, this does not assume SUTVA for time periods within units. This allows us to write the potential outcome under $\underline{a}_T$ as: $\underline{Y}_i(\underline{a}_T) = (Y_{i1}(\underline{a}_T), \dots, Y_{iT}(\underline{a}_T))$.

- We also can make a "no-anticipation" assumption: namely, that the potential outcome in period t only depends on the values of the treatment up to that point. So, if we have $\underline{a}_T = (\underline{a}_t, \underline{a}_{j>t})$ and $\underline{a}_t^* = (\underline{a}_t, \underline{a}_{j>t}^*)$, then it must be the case that $Y_{it}(\underline{a}_T) = Y_{it}(\underline{a}_T^*)$. This allows us to write the potential outcomes at time t as $Y_{it}(\underline{a}_t)$.
- Another complication here is that the time-varying regressors might also be affected by the treatment, so that we have $X_{it}(\underline{a}_{t-1})$.
- In general, we will be interested in average treatment effects:

$$E[Y_t(\underline{a}_t) - Y_t(\underline{a}_t^*)].$$

Or we might be interested in these effects, conditional on an observed regime:

$$E[Y_t(\underline{a}_t) - Y_t(\underline{a}_t^*)|\underline{A}_t \in B_t],$$

where B_t is set of regimes.

Fixed effects causal models

- To make progress, we will develop a causal model that are similar in spirit to the fixed effects regressions above. Namely, we will write a linear model for the potential outcome:

$$Y_{it}(\underline{a}_t) = X_{it}'(\underline{a}_{t-1})\beta_c + \tau_c(\underline{a}_t) + U_i + \varepsilon_{it}(\underline{a}_t)$$

- A few things have changed here. First, we write the coefficients with subscripts to indicate that these parameters may be different than the fixed effects regressions above. This is because we want to investigate when the fixed effects regressions will recover the causal parameters. Second, now there is a potential error, $\varepsilon_{it}(\underline{a}_t)$, which is the error that occurs when the treatment regime is $\underline{a}_t$.
- In order to make progress, we need to make some assumptions about how the errors relate to the covariates and the unit effects. This is similar to the assumptions we had to make about the relationship between the error and covariates with simple regression. And these assumptions will fall into two categories very similar to the fixed effects regression assumptions. First, we can have that the error is **strictly mean independent**:

$$E[\varepsilon_{it}(\underline{a}_t)|\{X_{i,j}(\underline{a}_{j-1})\}_{j=1}^T, U_i] = 0$$

- With the strictly mean independence, we have that the error is (mean) independent of the value of the time-varying covariates under the regime of interest and the unit effects. Again, this would rule out the dependent variable affecting future values of the time-varying covariates. Under this assumption, we can write the CEF of the potential outcome as follows:

$$E[Y_{it}(\underline{a}_t)|\{X_{i,j}(\underline{a}_{j-1})\}_{j=1}^T, U_i] = X_{it}'(\underline{a}_{1-t})\beta_c + \tau_c(\underline{a}_t) + U_i$$

- The second possible assumption weakens that to simply **sequentially mean independent**:

$$E[\varepsilon_{it}(\underline{a}_t)|\{X_{i,j}(\underline{a}_{j-1})\}_{j=1}^T, U_i] = 0$$

- This allows us to write the CEF as follows:

$$E[Y_{it}(\underline{a}_t)|\{X_{i,j}(\underline{a}_{j-1})\}_{j=1}^t, U_i] = X_{it}'(\underline{a}_{1-t})\beta_c + \tau_c(\underline{a}_t) + U_i$$

- We have made a lot of progress here. We have seen that the mean of the outcome conditional on the treatment regime and the mean of the potential outcome under that regime have similar forms. But, as always, it might not be the case that the mean of the potential outcomes under a treatment regime is the same as the mean among those who actually followed that treatment regime.

Treatment effects under the causal models

- We can combine the above causal models with the definition of the effects to simplify the defintion of the effects:

$$E[Y_t(\underline{a}_t) - Y_t(\underline{a}_t^*)] = E[\tau_c(\underline{a}_t) - \tau_c(\underline{a}_t^*) + (X_t(\underline{a}_{t-1}) - X_t(\underline{a}_{t-1}^*))'\beta_c + \varepsilon_{it}(\underline{a}_t) - \varepsilon_{it}(\underline{a}_t^*)]$$

- Note that in general, the effect of the regime on the outcome depends on how the regime affects the time-varying regressors. Unless of course, we have $X_t(\underline{a}_{t-1}) = X_t(\underline{a}_{t-1}^*)$, which happens if either the treatment has no effect on X_t or the two regimes agree up to $t-1$: $\underline{a}_{t-1} = \underline{a}_{t-1}^*$.

- There one other effect we might be interested in, the **contemporaneous effect**, which is the average effect of treatment in time t on the outcome in the same time period. That is, we are interested in the potential outcome if we were to keep your observed treatment regime and set the last period to be a_t. We would define the contemporaneous effect as $E[Y_t(\underline{A}_{t-1}, 1) - Y_t(\underline{A}_{t-1}, 0)]$, where now the expectation also averages across the previous treatment history. If each treatment status is available at each time period, then the this definition would be:

$$E\left[\sum_{M_{t-1}} [(\tau_c(\underline{a}_{t-1}, 1) - \tau_c(\underline{a}_{t-1}, 0)] \Pr[\underline{A}_{t-1} = \underline{a}_{t-1}]\right]$$

- Thus, the way we compute the contemporaneous effect of treatment is to estimate the effect of treatment in time t for each treatment regime up to time $t-1$. Then, we simply average those effect by the distribution of the those treatment regimes.

Indentification of treatment effects from fixed effect regression models

- In order to identify the effects from the last section, we need to make assumptions on the relationship between the treatment and the potential outcomes. We can give sufficient conditions for identification from the usual fixed-effect regressions.

- Condition 1: **sequential randomization**:

$$\left\{X_j(\underline{a}_{j-1}), Y_j(\underline{a}_j)\right\}_{j=t+1}^{T} \perp\!\!\!\perp A_t | \underline{X}_t, \underline{A}_{t-1} = \underline{a}_{t-1}, U$$

- This sequential randomization assumption implies the following:

$$E[Y_t|\underline{X}_t, \underline{A}_t = \underline{a}_t, U] = E[Y_t(\underline{a}_t)|\{X_m(\underline{a}_{m-1})\}_{m=1}^{t}, U]$$

- Further, if the above sequential mean independence assumption holds on the causal model, then we have: $E[\varepsilon_t(\underline{a}_t)|\underline{A}_{t-1} = \underline{a}_{t-1}] = 0$. Thus, when these two conditions hold, we have $\beta = \beta_c$ and $\tau(\underline{a}_t) = \tau_c(\underline{a}_t)$.

- One thing to note here is that the conditions on the potential error are conditional on observed treatment history, $\underline{A}_{t-1}$. Thus, even if there was no effect of the treatment on the time-varying regressors, we would not be able to identify effects of treamtent regimes that differed before t.

- Condition 2: **complete randomization**, which is sequential randomization plus:

$$Y_t(\underline{a}_t) \perp\!\!\!\perp A_{t+1}, \ldots, A_T | \{X_m(\underline{a}_{m-1})\}_{m=1}^t, \underline{A}_t = \underline{a}_t, U$$

- This (much stronger) condition implies:

$$E[Y_t|\underline{X}_T, \underline{A}_T = \underline{a}_T, U] = E[Y_t(\underline{a}_t)|\{X_m(\underline{a}_{m-1})\}_{m=1}^T, U]$$

- If we have, in addition, the strict mean independence from above, then we get that $E[\varepsilon_t(\underline{a}_t)|\underline{A}_t = \underline{a}_t] = 0$. Thus, the usual fixed effects regression identifies the causal parameters, τ_c and β_c. Note that the two conditions combined basically require a completely randomized treatment.

- If either of these conditions holds, we can identify those parameters and use them to identify the effect of a treatment in time t, conditional on the treatment histories being the same up to $t-1$. Thus, this also implies we can identify the contemporaneous effects as well.

- Why is this? Remember the formula for the effect:

$$E[Y_t(\underline{a}_t) - Y_t(\underline{a}_t^*)] = E[\tau_c(\underline{a}_t) - \tau_c(\underline{a}_t^*) + (X_t(\underline{a}_{t-1}) - X_t(\underline{a}_{t-1}^*))'\beta_c + \varepsilon_{it}(\underline{a}_t) - \varepsilon_{it}(\underline{a}_t^*)]$$

- While it is true that we can identify τ_c and β_c and under the assumptions the error terms have 0 mean, we have not identified the effect of treatment on the time-varying confounders, $(X_t(\underline{a}_{t-1}) - X_t(\underline{a}_{t-1}^*))$. We could put restrictions on the effect, such as the it only depends on the number of previously treated periods or that there is "transience" in the sense that $X_t(\underline{a}_t) = X_t(\underline{a}_t^*)$ if $\underline{a}_t$ and $\underline{a}_t^*$ agree for the last k periods.

What do fixed effects estimate?

- What this tells us is that with a properly specified fixed-effect model, the best we can hope to estimate is a version of the contemporaneous effect of treatment. That is, we can identify the effect of the last period, conditional on the past being the same.

- What fixed effects models cannot give us are "cumulative effects" unless we are willing to make even stronger assumptions on the data.

- Many times in applied work, folks assume there are no cumulative effects (either explicitly or implicitly) and so their fixed effects models are "fine" (as long as randomization within unit, linearity, and strict exogeneity hold). This might be better than nothing, though.

- What do we do? Well, we'll look at this again in a different context in a few weeks when we talk about dynamic causal inference.