FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

So far we have assumed that treatment assignment is unconfounded given measured covariates. In many instances this assumption is not reasonable.

Economists argue that if measurements were taken on multiple units in a cluster and all these units shared a common unobserved variable U, then U, even though it is not measured, could be taken into account using standard models such as regression.

FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

e.g. in panel (repeated measures) data, individuals are observed at multiple points in time. Each individual is a cluster, and U measures unobserved variables that are unobserved and constant over time.

Economists have used such data to study the effect of marriage on men's earnings. The question is whether more productive men are more likely to be married or whether marriage increases productivity. Measured covariates include education, family background, occupation and industry, and union status. But productivity itself is not measured.

Suppose that U represents unmeasured aspects of productivity, and U is constant within units. Suppose also that earnings Y_{it} for unit $i = 1, ..., n$ in period $t = 1, ..., T \geq 2$ depends on time varying regressors $\mathbf{X}_{it}$ and marital status Z_{it}(1 if currently married, 0 otherwise), U_i and random errors ε_{it}.

FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

To keep matters simple and to mirror the empirical literature, suppose

$$Y_{it} = \alpha + \mathbf{X}_{it}'\boldsymbol{\beta} + \tau(\bar{Z}_{it}) + U_i + \varepsilon_{it},$$

where $\bar{Z}_{it} = (Z_{i1}, ..., Z_{it})$, and we assume either

$$E(\varepsilon_{it} \mid \bar{\mathbf{X}}_{iT}, \bar{Z}_{iT}, U_i) = 0,$$

$i = 1, ..., n,\ t = 1, ..., T$, where $\bar{\mathbf{X}}_{iT} = (\mathbf{X}_{i1}, ..., \mathbf{X}_{iT})$, $\bar{Z}_{iT} = (Z_{i1}, ..., Z_{iT})$, an assumption Chamberlain (1982) calls strict exogeneity, or sequential exogeneity (Chamberlain 1992): for $i = 1, ..., n,\ t = 1, ..., T$,

$$E(\varepsilon_{it} \mid \bar{\mathbf{X}}_{it}, \bar{Z}_{it}, U_i) = 0,$$

where $\bar{\mathbf{X}}_{it} = \mathbf{X}_{i1}, ..., \mathbf{X}_{it}$.
The former assumption is unreasonable when $\mathbf{X}_{it}$ contains lagged values $Y_{i,t-k}$, in which case sequential exogeneity is more reasonable.

FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

If U_i in

$$Y_{it} = \alpha + \mathbf{X}_{it}'\boldsymbol{\beta} + \tau(\bar{Z}_{it}) + U_i + \varepsilon_{it}$$

is treated as a mean 0 random variable with $U_i \perp\!\!\!\perp \mathbf{X}_{it}, \bar{Z}_{it}$, we obtain a random effects model.

This assumption is often unreasonable. e.g. in the marriage example, it would not be reasonable to assume unmeasured aspects of productivity are independent of education and other covariates.

Economists often prefer to treat U_i as a "fixed effect" (Mundlak, 1978). In this case, by including dummy variables for each unit i, U_i can be treated as a parameter to be estimated as above or differenced out in several ways e.g.

$$Y_{it} - Y_{i1} = (\mathbf{X}_{it} - \mathbf{X}_{i1})'\boldsymbol{\beta} + \tau(\bar{Z}_{it}) - \tau(\bar{Z}_{i1}) + (\varepsilon_{it} - \varepsilon_{i1}),$$

or $Y_{it} - \bar{Y}_i$, where $\bar{Y}_i = T^{-1}\sum_{t=1}^{T} Y_{it}$.

FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

Under strict exogeneity, consistent estimates of $\boldsymbol{\beta}$ and $\tau(\bar{Z}_{it})$ can be obtained using ordinary least squares (OLS). Typically, $\tau(\bar{Z}_{it}) = \tau Z_{it}$ and the estimate $\hat{\tau}$ of τ is simply interpreted as an effect, the exact nature of which is not stated.

Estimation is more complicated under sequential exogeneity and will not be discussed here (Wooldridge, 2010).

FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

Throughout this course, we emphasize the importance of using potential outcomes to define unit and average effects, then stating conditions under which these are identified, followed by estimation.

In the approach above, the regression model

$$Y_{it} = \alpha + \mathbf{X}_{it}'\boldsymbol{\beta} + \tau(\bar{Z}_{it}) + U_i + \varepsilon_{it}$$

or

$$Y_{it} - Y_{i1} = (\mathbf{X}_{it} - \mathbf{X}_{i1})'\boldsymbol{\beta} + \tau(\bar{Z}_{it}) - \tau(\bar{Z}_{i1}) + (\varepsilon_{it} - \varepsilon_{i1}),$$

is the starting point, and the parameter $\tau(Z_{it})$, a model by-product, is simply interpreted as a treatment effect.

This is problematic. We reformulate this problem within the framework of longitudinal causal inference using potential outcomes (Sobel, 2012).

FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

In each period t, we assume the time varying covariates are first observed, followed by treatment, then by the response.

An important point of great consequence, not considered in the previous approach, is the dependence of the time varying covariates on prior treatment assignments.

We define time varying covariates $\mathbf{X}_{it}(\bar{z}_{t-1})$ and potential responses $Y_{it}(\bar{z}_t)$, where $\bar{z}_t = (z_1, ...z_t)$ is the vector of treatment assignments up through and including period t.

For the regression model previously considered

$$Y_{it} = \alpha + \mathbf{X}_{it}'\boldsymbol{\beta} + \tau(\bar{Z}_{it}) + U_i + \varepsilon_{it},$$

we consider the causal model:

$$Y_{it}(\bar{z}_t) = \alpha^{(c)} + \mathbf{X}_{it}'(\bar{z}_{t-1})\boldsymbol{\beta}^{(c)} + \tau^{(c)}(\bar{z}_t) + U_i + \varepsilon_{it}(\bar{z}_t),$$

where, for $i = 1, \ldots, n$, $t = 1, \ldots, T$, either

$$E(\varepsilon_{it}(\bar{z}_t) \mid \bar{\mathbf{X}}_{iT}, U_i) = 0, \tag{1}$$

or

$$E(\varepsilon_{it}(\bar{z}_t) \mid \bar{\mathbf{X}}_{it}, U_i) = 0. \tag{2}$$

We write the parameters with the $^{(c)}$ to distinguish these from the analogous parameters of the regression model; in general, the two sets of parameters are not identical. Assumption (1) is analogous to strict exogeneity, and assumption (2) to sequential exogeneity.

Under either assumption

$$E(Y_t(\bar{z}_t) - Y_t(\bar{z}_t^*)) = E(\mathbf{X}_{it}(\bar{z}_{t-1}) - \mathbf{X}_{it}(\bar{z}_{t-1}^*))'\boldsymbol{\beta}^{(c)} + \tau^{(c)}(\bar{z}_t) - \tau^{(c)}(\bar{z}_t^*).$$

This makes evident that comparison between different treatment regimens depends not only on terms $\tau(\bar{z}_t)$, but also on the effect of different treatment histories on the time varying confounders.

To estimate the effect of $\bar{z}_t$ vs. $\bar{z}_t^*$, it is necessary to model the effect of treatment history on the time varying confounders if $E(\mathbf{X}_{it}(\bar{z}_{t-1}) - \mathbf{X}_{it}(\bar{z}_{t-1}^*))'\boldsymbol{\beta} \neq 0$: $\tau^{(c)}(\bar{z}_t) - \tau^{(c)}(\bar{z}_t^*)$ is the effect of $\bar{z}_t$ vs. $\bar{z}_t^*$ only if $E(\mathbf{X}_{it}(\bar{z}_{t-1}) - \mathbf{X}_{it}(\bar{z}_{t-1}^*))'\boldsymbol{\beta} = 0$, as would be the case if treatment does not affect the time varying confounders. However, this assumption is unreasonable.

Consider assignments with $\mathbf{X}_{it}(\bar{z}_{t-1}) = \mathbf{X}_{it}(\bar{z}^*_{t-1})$.

If $\bar{z}_{t-1} = \bar{z}^*_{t-1}$, $z_t = 1, z^*_t = 0$, and $\tau^{(c)}(\bar{z}_t) = \tau^{(c)} z_t$ then $\tau^{(c)}$ is then the average effect of treatment in period 1 and in periods $t = 2, ..., T$ when $\bar{z}_{t-1} = \bar{z}^*_{t-1}$.

Presumably, $\tau^{(c)}$ is what empirical researchers are attempting to estimate when $\tau(\bar{Z}_{it}) = \tau Z_{it}$.

FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

Sobel (2012) gives conditions under which consistent estimates of $\tau^{(c)}(\bar{z}_t)$, hence valid estimates of treatment effects when $\mathbf{X}_{it}(\bar{z}_{t-1}) = \mathbf{X}_{it}(\bar{z}^*_{t-1})$ can be obtained using regression. To justify the use of estimating $Y_{it} = \alpha + \mathbf{X}'_{it}\boldsymbol{\beta} + \tau(\bar{Z}_{it}) + U_i + \varepsilon_{it}$, assuming sequential exogeneity, a sequential randomization assumption is required.

The condition is stronger than the module on longitudinal causal inference, where we assumed period t treatment assignment was independent of future potential outcomes given previous outcomes, confounders, and treatment history.

We now assume period t treatment assignment independent of future potential outcomes and potential confounders, given previous outcomes, confounders, treatment history, and the unobserved confounder U_i.

FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

The vast majority of analyses using fixed effects models use OLS, which gives consistent estimates of model parameters under $Y_{it} = \alpha + \mathbf{X}_{it}'\boldsymbol{\beta} + \tau(\bar{Z}_{it}) + U_i + \varepsilon_{it}$, and strict exogeneity.

Here, the identification conditions required to justify using OLS are even stronger, requiring conditional independence of future potential outcomes and the complete assignment history Z_{iT}.

In observational studies, such an assumption is almost never credible, requiring that potential outcomes do not affect subsequent treatment assignments.

FIXED EFFECTS REGRESSIONS IN ECONOMETRICS

Fixed effects models are also used when units i are clustered. e.g. units are husband/wife pairs, siblings in a family, students in classrooms, persons in neighborhoods, etc. Then a fixed effect for the cluster is sometimes postulated to account for a common environment within the cluster that is not measured by observed variables. In this case, the problems above may not arise.

In these cases, however, it is likely that units interfere with one another. Even if a randomized experiment had been conducted, ignoring interference would lead to invalid inferences. Second, in our treatment of interference, the estimands we considered were identified under randomization conditions. As the rationale for using fixed effects models is to control for unmeasured confounding in observational studies, different identification conditions would need to be developed.