Problem Set 3

We talked about the role of conditional expectation functions in estimating differences in outcome. Then we considered a specific form of the error: unobserved group level differences. We discussed using fixed effects models as well as, in the case when this effect is uncorrelated, random effects models. We discussed this in the context of a family effects model. This exercise is intended to help you develop the underlying theory of CEF's and an empirical understanding of estimation with fixed effects.

Question 1 - Conditional Expectation Functions

A) **Question:** *Prove that the any outcome Y can be decomposed into two parts, the CEF (i.e $E[Y_i|X_i]$), and a mean-independent idiosyncratic term, ϵ, and that therefore, ϵ is uncorrelated with any function of X.*

Answer:

(i) Let us just define $\epsilon_i \equiv y_i - E[y_i|X_i]$.[1] Then by construction

$$E[\epsilon_i|X] = E[y_i - E[y_i|X_i]|X_i] = E[y_i|X_i] - E[y_i|X_i] = 0$$

(ii) Let $h(X_i)$ be any function of X_i, then

$$E[h(X_i)\epsilon_i] = E[E[h(X_i) \cdot \epsilon_i|X_i]] = E[h(X_i) \cdot E[\epsilon_i|X_i]] = 0$$

B) **Question:** *Now show that the function $X'\beta$, where β is the population regression parameter vector, provides the best (in a minimum MSE sense) linear approximation to $E[Y|X]$, i.e. prove $\beta = argmin_b E(E[Y|X] - X'b)^2$.*

Answer: Any linear approximation of the CEF will be of the form Xb. We search for the best linear approximation minimizing the mean-squared-error, so we want to solve

$$min_b E[(E[y|X] - Xb)^2] \tag{1}$$

It turns out that the solution to

$$min_b E[(y - Xb)^2] \tag{2}$$

will be the same as to the first problem. But the solution to the second problem is OLS. You can see this since

$$\begin{aligned} E(y - Xb)^2 &= E(y - E[y|X] + E[y|X] - Xb)^2 \\ &= E[(y - E[y|X]) + (E[y|X] - Xb)]^2 \\ &= E(y - E[y|X])^2 + E(E[y|X] - Xb)^2 \\ &\quad +2E[(y - E[y|X]) \cdot (E[y|X] - Xb)] \end{aligned}$$

The first bit does not depend on b and hence does not affect the minimization, and the last bit is just $E[\epsilon \cdot h(X)]$, where $h(X)$ is a function of X. But this we showed is 0.

This is a justification for being interested in the true population regression parameter β.

C) **Question:** Explain (in words or math) why the fact that the CEF can be additively separated from the unobserved error is critical for the Conditional Independence Assumption to allow causal interpretations of β?

[1] To understand this, just think of the joint density function of y_i and X_i. What we are doing is that at any X_i we take the expectations over all possible y_is that come with this X_i and then define y_i as this expectation plus some error - all at a given X_i, so at one cut through the joint density function.

Answer: Generally the CEF will take some form $f(\mathbf{X};\beta)$ and we write $y = f(\mathbf{X};\beta) + \epsilon$. By the above property we have that $E[y|X] = E[f(\mathbf{X};\beta)|X]$. If further $f(\mathbf{X};\beta)$ is linear, so $f(\mathbf{X};\beta) = \mathbf{X}\beta$, this allows us to interpret β as causal since

$$\frac{\partial E[y]}{\partial \mathbf{X}} = \frac{\partial f(\mathbf{X};\beta)}{\partial \mathbf{X}} = \frac{\partial \mathbf{X}\beta}{\partial \mathbf{X}} = \beta \tag{3}$$

Note that if the CEF is not linear, this would not work. The second derived property ensures that, if we specify $f(\mathbf{X};\beta) = \mathbf{X}\beta$ correctly (corresponding to **A2+A3Rmi**), β_{OLS} will be an unbiased estimate for β.

Question 2 - Understanding Fixed and Random Effects

Suppose we have a standard linear model of an outcome y_{it}, that is $y_{it} = x_{it}'\beta + \epsilon_{it}$. In this model, suppose that the error term can be decomposed into two parts, a time specific component δ_t and an idiosyncratic, mean zero error term μ_{it}. That is the composite error term is $\epsilon_{it} = \delta_t + \mu_{it}$.

A) **Question:** *If we assume that δ_t is not correlated with $\mathbf{X}$, we can estimate this without worrying about omitted variable bias. To test that it is independent, suppose that we estimated the function with a time fixed effect and without. What do we compare to demonstrate that the time fixed effect is independent of X? Why does that test make sense?*

Answer: First note that whether δ_t and X are correlated or not has implications for $\hat{\beta}_{RE}$ and $\hat{\beta}_{FE}$:

Random Effects If the δ_t and $\mathbf{X}$s are uncorrelated, then Random Effects will be consistent and efficient (it is just GLS). But if δ_t and $\mathbf{X}$ are correlated, then it will be inconsistent.

Fixed Effects This is not the case for a Fixed Effects estimation, since this essentially gets away with δ_t. It is not efficient if δ_t and $\mathbf{X}$ are uncorrelated, but in both cases it will be consistent.

So under the H_0: δ_t and $\mathbf{X}$ are uncorrelated, both estimators are consistent and should give similar results while under the alternative, H_a: δ_t and $\mathbf{X}$ are correlated, FE/OLS is consistent but RE/GLS is not. If we can find the distribution of the difference of the estimators under the null we can define a critical region and perform a hypothesis test.

To develop this test, use the idea is that under the null both FE/OLS and RE/GLS estimators are consistent but only GLS is efficient. To find the distribution of the difference note that the covariance of an efficient estimator with the difference between that efficient estimator and an inefficient estimator is zero[2], or

$$Cov[(\hat{\beta}_{fe} - \hat{\beta}_{re}), \hat{\beta}_{re}] = Cov[\hat{\beta}_{fe}, \hat{\beta}_{re}] - Var[\hat{\beta}_{re}] = 0$$

So then the variance of $[\hat{\beta}_{fe} - \hat{\beta}_{re}]$ is:

$$Var[\hat{\beta}_{fe} - \hat{\beta}_{re}] = Var[\hat{\beta}_{fe}] + Var[\hat{\beta}_{re}] - 2 \cdot Cov[\hat{\beta}_{fe}, \hat{\beta}_{re}] = Var[\hat{\beta}_{fe}] - Var[\hat{\beta}_{re}]$$

Define this difference in variances as Ω. Then we can define the test statistics H,

$$H = [\hat{\beta}_{fe} - \hat{\beta}_{re}]'\Omega^{-1}[\hat{\beta}_{fe} - \hat{\beta}_{re}]$$

This is distributed χ^2 with $K-1$ degrees of freedom. If this quantity is very large (so the different of $\hat{\beta}_{re}$ and $\hat{\beta}_{fe}$ is big) we reject the null.[3] In our case H_0: 'δ_t and $\mathbf{X}$ are independent'.

[2] Surely you were not expected to know this!
[3] If you are interested, this test is a version of the Wu-Hausmann test.

B) **Question:** *Now suppose we find that δ_t is correlated with $\boldsymbol{X}$, and we wish to estimate a model in differences from mean form. What assumptions must we make to ensure that will eliminate the problem of omitted variable bias?*

Answer: To the extent that an omitted variable affects all individuals at a given time, t, we can control for this using a time fixed effect.[4] This works since we are controlling for any factor which is the same across all individuals at time t in the regression when we run

$$y_{it} = X_{it}'\beta + D_t'\delta + \epsilon_{it} \tag{4}$$

where D_t is a matrix of time dummies and δ are the coefficients on these time dummies.

For this to work we certainly need that

- *the omitted variable indeed takes the same value for all individuals i in time t and*
- *the omitted variable indeed has the same effect for all i, so there is no heterogeneity in δ.*

C) **Question:** *We want to test if it is reasonable to assume that β is the same in all time periods. We could do this by running a single regressions. If we had T periods, what could we test to verify if our hypothesis on β is correct?*

Answer: If you want to test whether the effect of x_{it} on y_{it} is the same in all time periods, you can just allow them to be different and then test the linear hypothesis that they are the same.

So you would want to run the regression

$$y_{it} = X_{it}\beta_t + \delta_t + \epsilon_{it} \tag{5}$$

This will give you the same coefficient estimates as running T separate regressions (since you have for each time period a separate intercept and slope coefficient). However, running them jointly allows you to test the H_0: $\beta_1 = \beta_2 = ... = \beta_k$ in the simple fashion we are used to.

Question 3

In lecture we discussed a study on the returns to education which used differences in schooling between identical twins to identify the causal effect of education on earnings. The key assumption of this paper was the family level effects were the only omitted variable and that these effects were homogeneous within a family and additively separable from other variables of interest. They use this fact to then estimate the returns to schooling and find that the standard estimate is significantly bigger than the twins-based estimate.

A) **Question:** *The study authors Ashenfelter and Rouse (1998) present evidence that differences in schooling between twins are uncorrelated with birth order and a range of characteristics such as union status, self-employment, tenure and spouse's education. They therefore argue that between-twins education differences estimates are not biased. What is the assumption that this evidence makes to convince us that the twins-difference estimated effect is unbiased.*

Answer: We saw, that once we have included fixed effects, our estimator will explain differences in the outcome variable **within groups** (i.e. families) with the variation in the explanatory variable **within these groups**. That excludes many potential omitted variables. But we still need to ensure that **within one group** the differences in the explanatory variable are indeed **as good as random**. So, in our example, the reasons why identical twins get different amounts

[4] For a concrete example, think of some data where you are worried about **omitted variables**. For example y_{it} being a variable capturing whether individual i broke his arm in week t and x_{it} being a variable capturing for how many hours he worked. Wanting to find the causal relation between the two, you should be worried about omitted variables. For example if there was snow in week t this might drive up the incidence of broken arms and drive down the number of hours worked - *which would give us a correlation between y_{it} and x_{it} even though there is no causal relation.*

of schooling should have nothing to do with their earnings! Or: We try to show that for our purposes the differences in schooling among identical twins are as good as random.

A potential problem is that there are many things which might drive one of the twins to get more education than the other and drive the earnings as well. We cannot look at all of them. But if we at least find that some of the potential candidates have no correlation with how many years of schooling one or the other twin gets, this gives us some comfort. That is what the authors do - and the question says they find no correlation.

If we *believe/assume* that **all other factors which may be correlated with both schooling choice and earnings and which we could not observe are sufficiently correlated with those which we checked**, then the fact that we do not find anything for those which we observed makes it unlikely that our estimates are biased by any of the unobserved and hence omitted factors.

B) **Question:** *In the paper, the authors estimate the return to schooling including the average level of education in the family as control (but no fixed effects). They find a significant correlation between the average family education and income. What must they assume for the average level of education to deal with omitted variables?*

Answer: It needs to be both that the family effect (of e.g. income) is occurring only through schooling (i.e. not other unobservable family characteristics) and that the effect of average family education of the outcome variable is linear. If the omitted variables work through the average family education, including the latter variable will capture the effects of the omitted variable. However, compared to including fixed effects this specification is less flexible. In particular the authors need to believe that the average level of education influences the outcome in a linear way. If they had included family fixed effects this would not be the case, rather we would estimate for each family a family-specific coefficient.

C) **Question:** *In the table below, the author find different estimate when including average family education (column 2) than when excluding family variable (column 1). What does the change in the coefficient estimate when including family control variables tell you about the bias in the more parsimonious specification in column 1?*

Answer: Suppose we do not include fixed effects, but just control for the average level of education in the family.[5] Then the effect of the own education goes down. This might have two reasons:

- There might be something else which drives both average education and earning up in the family, e.g. parents social network.
- [Or this might as well capture the fact that there are spill-over effects of education on the siblings.]

In any case it shows that the **initially estimated relation between an individual's education and his earnings was overestimated in (1)**.

D) **Question:** *Show that the β estimated from the fixed effects estimator is the same as a difference estimator (i.e. twin-i - twin-average variables). Assume we are estimating a simple regression of $y = \beta s + F\delta + \mu$ where s is schooling and F is a matrix of family level dummy variables and δ is the associated coefficient vector.*

Answer: We can estimate straightforwardly the OLS estimator for the following FE model

$$y_{ij} = \beta s_{ij} + f'_{ij}\delta + \epsilon_{ij}$$

We know that using **partitioned regression results** we can calculate $\hat{\beta}_{OLS}$ equivalently as

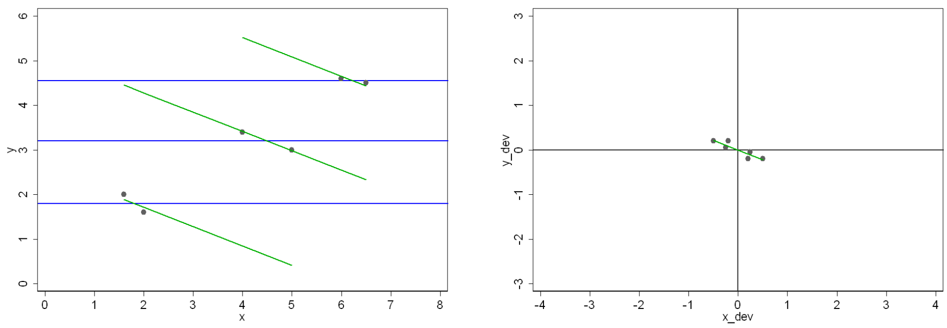
[5] This will never fit the data better than fixed effects.

$$\hat{\beta}_{OLS} = [s'M_F s]^{-1} s'M_F y$$
$$= [(M_F s)'(M_F s)]^{-1}(M_F s)'M_F y$$

where $M_F = I - F(F'F)^{-1}F'$.

You can easily see why this is called the **difference estimator**. Just show that $M_F x$ gives a vector of deviations of x_{ij} from the family mean of x. (We have done something very similar with seasonal fixed effects in Vassilis' part.)

Graphically,



(a) Fixed Effects estimation (b) Within estimation

E) **Question:** *Columns (3) and (5) correct for measurement error. What happens to the coefficients? Why does this make sense?*

Answer: In columns (3) and (5) the authors instrument for explanatory variable (years of schooling). Among other things this takes care of potential measurement error.

The fact that the coefficient estimates indeed increase relative to the respective non-instrumented counterparts, i.e. (2) and (4), might be explained by the fact that previously the coefficient estimates suffered from the attenuation (i.e. downward) bias with is cause by measurement error.[6]

[6] Note that this has pretty much nothing to do with the fixed effects issues.