Exercise Sheet 1: Estimation of Linear Panel Data Models

– Draft Solution –

Question 1

Consider the matrices $\mathbf{J}_T = \boldsymbol{\iota}_T(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T)^{-1}\boldsymbol{\iota}_T'$, $\mathbf{Q}_T = \mathbf{I}_T - \mathbf{J}_T$, $\mathbf{J} = \mathbf{I}_N \otimes \mathbf{J}_T$, and $\mathbf{Q} = \mathbf{I}_N \otimes \mathbf{Q}$, where $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones.

(a) Show that $\mathbf{J}_T'\mathbf{J}_T = \mathbf{J}_T$, $\mathbf{Q}_T'\mathbf{Q}_T = \mathbf{Q}_T$, $\mathbf{J}_T\boldsymbol{\iota}_T = \boldsymbol{\iota}_T$, $\mathbf{Q}_T\boldsymbol{\iota}_T = \mathbf{0}$, and $\mathbf{Q}_T\mathbf{J}_T = \mathbf{0}$.

Answer:

$\mathbf{J}_T\mathbf{J}_T = \boldsymbol{\iota}_T\left(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T\right)^{-1}\boldsymbol{\iota}_T'\boldsymbol{\iota}_T\left(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T\right)^{-1}\boldsymbol{\iota}_T' = \boldsymbol{\iota}_T\left(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T\right)^{-1}\boldsymbol{\iota}_T' = \mathbf{J}_T$

$\mathbf{Q}_T'\mathbf{Q}_T = \left(\mathbf{I}_T - \mathbf{J}_T\right)'\left(\mathbf{I}_T - \mathbf{J}_T\right) = \mathbf{I}_T - \mathbf{J}_T - \mathbf{J}_T + \mathbf{J}_T\mathbf{J}_T = \mathbf{I}_T - 2\mathbf{J}_T + \mathbf{J}_T = \mathbf{I}_T - \mathbf{J}_T = \mathbf{Q}_T$

$\mathbf{J}_T\boldsymbol{\iota}_T = \boldsymbol{\iota}_T\left(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T\right)^{-1}\boldsymbol{\iota}_T'\boldsymbol{\iota}_T = \boldsymbol{\iota}_T$

$\mathbf{Q}_T\boldsymbol{\iota}_T = \left(\mathbf{I}_T - \mathbf{J}_T\right)\boldsymbol{\iota}_T = \boldsymbol{\iota}_T - \mathbf{J}_T\boldsymbol{\iota}_T = \boldsymbol{\iota}_T - \boldsymbol{\iota}_T = \mathbf{0}$

$\mathbf{Q}_T\mathbf{J}_T = \mathbf{Q}_T\boldsymbol{\iota}_T\left(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T\right)^{-1}\boldsymbol{\iota}_T' = \mathbf{0}$ because $\mathbf{Q}_T\boldsymbol{\iota}_T = \mathbf{0}$

(b) Show that the within-transformed regressors $\ddot{\mathbf{X}}_i$ and the between-transformed regressors $\bar{\mathbf{X}}_i$ are orthogonal to each other.

Answer:

$\ddot{\mathbf{X}}_i'\bar{\mathbf{X}}_i = \left(\mathbf{Q}_T\mathbf{X}_i\right)'\left(\mathbf{J}_T\mathbf{X}_i\right) = \mathbf{X}_i'\mathbf{Q}_T'\mathbf{J}_T\mathbf{X}_i = \mathbf{0}$ because $\mathbf{Q}_T'\mathbf{J}_T = \mathbf{0}$

(c) Show that $\mathbf{J}'\mathbf{J} = \mathbf{J}$, $\mathbf{Q}'\mathbf{Q} = \mathbf{Q}$, and $\mathbf{Q}'\mathbf{J} = \mathbf{0}$.

Answer:

$\mathbf{J}'\mathbf{J} = \left(\mathbf{I}_N \otimes \mathbf{J}_T\right)'\left(\mathbf{I}_N \otimes \mathbf{J}_T\right) = \left(\mathbf{I}_N \otimes \mathbf{J}_T'\mathbf{J}_T\right) = \mathbf{I}_N \otimes \mathbf{J}_T = \mathbf{J}$
$\mathbf{Q}'\mathbf{Q} = \left(\mathbf{I}_N \otimes \mathbf{Q}_T\right)'\left(\mathbf{I}_N \otimes \mathbf{Q}_T\right) = \left(\mathbf{I}_N \otimes \mathbf{Q}_T'\mathbf{Q}_T\right) = \mathbf{I}_N \otimes \mathbf{Q}_T = \mathbf{Q}$
$\mathbf{Q}'\mathbf{J} = \left(\mathbf{I}_N \otimes \mathbf{Q}_T\right)'\left(\mathbf{I}_N \otimes \mathbf{J}_T\right) = \left(\mathbf{I}_N \otimes \mathbf{Q}_T'\mathbf{J}_T\right) = \mathbf{I}_N \otimes \mathbf{0} = \mathbf{0}$

Question 2

Consider the error components model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)\mathbf{c} + \mathbf{u}$$

with $\boldsymbol{\Omega} = \sigma_c^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \sigma_u^2 \mathbf{I}_T$.

(a) Show that the FE estimator is equivalent to the OLS estimator applied to the within-transformed equation.

Answer:

FE estimator: $\hat{\boldsymbol{\beta}}_{FE} = \left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right)^{-1} \ddot{\mathbf{X}}'\ddot{\mathbf{y}}$

Within–transformed equation: $\mathbf{Q}\mathbf{y} = \mathbf{Q}\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{v} \Leftrightarrow \ddot{\mathbf{y}} = \ddot{\mathbf{X}}\boldsymbol{\beta} + \ddot{\mathbf{v}}$

OLS applied to this equation: $\hat{\boldsymbol{\beta}} = \left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right)^{-1} \ddot{\mathbf{X}}'\ddot{\mathbf{y}} = \hat{\boldsymbol{\beta}}_{FE}$

(b) Show that the between estimator is equivalent to the OLS estimator applied to the between-transformed equation.

Answer:

Between estimator: $\hat{\boldsymbol{\beta}}_B = \left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1} \bar{\mathbf{X}}'\bar{\mathbf{y}}$

Between-transformed equation: $\mathbf{J}\mathbf{y} = \mathbf{J}\mathbf{X}\boldsymbol{\beta} + \mathbf{J}\mathbf{v} \Leftrightarrow \bar{\mathbf{y}} = \bar{\mathbf{X}}\boldsymbol{\beta} + \bar{\mathbf{v}}$

OLS applied to this equation: $\hat{\boldsymbol{\beta}} = \left(\bar{\mathbf{X}}'\bar{\mathbf{X}}\right)^{-1} \bar{\mathbf{X}}'\bar{\mathbf{y}} = \hat{\boldsymbol{\beta}}_B$

(c) Show that the RE estimator erroneously applied to the within-transformed equation,

$$\hat{\boldsymbol{\beta}}_{err} = \left(\ddot{\mathbf{X}}'[\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}}^{-1}]\ddot{\mathbf{X}}\right)^{-1} \ddot{\mathbf{X}}'[\mathbf{I}_N \otimes \hat{\boldsymbol{\Omega}}^{-1}]\ddot{\mathbf{y}},$$

is identical to the FE estimator.

Answer:

Recall that $\ddot{\mathbf{X}} = \mathbf{Q}\mathbf{X} = (\mathbf{I}_N \otimes \mathbf{Q}_T)\,\mathbf{X}$ and $\ddot{\mathbf{y}} = \mathbf{Q}\mathbf{y} = (\mathbf{I}_N \otimes \mathbf{Q}_T)\,\mathbf{y}$. Also recall that

$\hat{\sigma}_u^2 \hat{\boldsymbol{\Omega}}^{-1} = \mathbf{Q}_T + \hat{\phi}^2 \mathbf{J}_T$. Hence:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{err} &= \left(\mathbf{X}'\mathbf{Q}\left[\mathbf{I}_N \otimes \left(\mathbf{Q}_T + \hat{\phi}^2 \mathbf{J}_T\right)\right]\mathbf{Q}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Q}\left[\mathbf{I}_N \otimes \left(\mathbf{Q}_T + \hat{\phi}^2 \mathbf{J}_T\right)\right]\mathbf{Q}\mathbf{y} \\
&= \left(\mathbf{X}'\left(\mathbf{I}_N \otimes \mathbf{Q}_T\right)\left[\mathbf{I}_N \otimes \left(\mathbf{Q}_T + \hat{\phi}^2 \mathbf{J}_T\right)\right]\left(\mathbf{I}_N \otimes \mathbf{Q}_T\right)\mathbf{X}\right)^{-1} \\
&\quad \mathbf{X}'\left(\mathbf{I}_N \otimes \mathbf{Q}_T\right)\left[\mathbf{I}_N \otimes \left(\mathbf{Q}_T + \hat{\phi}^2 \mathbf{J}_T\right)\right]\left(\mathbf{I}_N \otimes \mathbf{Q}_T\right)\mathbf{y} \\
&= \left(\mathbf{X}'\left[\mathbf{I}_N \otimes \mathbf{Q}_T\left(\mathbf{Q}_T + \hat{\phi}^2 \mathbf{J}_T\right)\mathbf{Q}_T\right]\mathbf{X}\right)^{-1} \mathbf{X}'\left[\mathbf{I}_N \otimes \mathbf{Q}_T\left(\mathbf{Q}_T + \hat{\phi}^2 \mathbf{J}_T\right)\mathbf{Q}_T\right]\mathbf{y}
\end{aligned}$$

Due to

$$\mathbf{Q}_T\left(\mathbf{Q}_T + \hat{\phi}^2 \mathbf{J}_T\right)\mathbf{Q}_T = \mathbf{Q}_T\mathbf{Q}_T\mathbf{Q}_T + \hat{\phi}^2 \mathbf{Q}_T\mathbf{J}_T\mathbf{Q}_T = \mathbf{Q}_T + \mathbf{0} = \mathbf{Q}_T,$$

we can further simplify to

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{err} &= \left(\mathbf{X}'\left(\mathbf{I}_N \otimes \mathbf{Q}_T\right)\mathbf{X}\right)^{-1} \mathbf{X}'\left(\mathbf{I}_N \otimes \mathbf{Q}_T\right)\mathbf{y} \\
&= \left(\mathbf{X}'\mathbf{Q}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Q}\mathbf{y} = \left(\left(\mathbf{Q}\mathbf{X}\right)'\left(\mathbf{Q}\mathbf{X}\right)\right)^{-1}\left(\mathbf{Q}\mathbf{X}\right)'\left(\mathbf{Q}\mathbf{y}\right) \\
&= \left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right)^{-1} \ddot{\mathbf{X}}'\ddot{\mathbf{y}} = \hat{\boldsymbol{\beta}}_{FE}.
\end{aligned}$$

(d) What is the result of the FE estimator applied to the between-transformed equation? Explain.

Answer:

The between-transformed equation reads:

$$\mathbf{J}_{\mathbf{y}} = \mathbf{J}\mathbf{X}\boldsymbol{\beta} + \mathbf{J}\mathbf{v}.$$

The FE estimator applied to this equation is equivalent to OLS applied to its within-transformation

$$\mathbf{Q}\mathbf{J}\mathbf{y} = \mathbf{Q}\mathbf{J}\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{J}\mathbf{v}.$$

Since $\mathbf{Q}\mathbf{J} = \mathbf{0}$, this equation has solely zero observations. Hence, the FE estimator cannot be used here. The intuitive explanation is simple: A within (FE) estimator uses only the deviations from time averages. But the between-transformed equation includes only time averages and thus no deviations from these averages. Hence, applying the within-transformation to the between-transformed equation yields only zero observations.

Question 3

Acemoglu et al. (2008) analyze the effect of income on democracy.[1] They use a large country panel for 1960-2000 sampled at five-year intervals. Their baseline specification is

$$dem_{it} = \beta_1 dem_{i,t-1} + \beta_2 inc_{i,t-1} + \mu_t + c_i + u_{it}, \tag{1}$$

where dem_{it} denotes the democracy score of country i in period t (measured as the Freedom House Political Rights Index and scaled so that it is between zero and one, with one corresponding to the most democratic set of institutions), inc_{it} denotes log income per capita (in constant 1990 US dollars), and μ_t is a full set of year dummies.

(a) A pooled OLS regression of dem_{it} on inc_{it} yields the results presented below. The variable "code_numeric" is a country identifier used when computing robust s.e.'s.

```
. reg dem inc if sample==1, cluster(code_numeric)

Linear regression                                      Number of obs =     960
                                                       F(  1,    151) =  414.02
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4488
                                                       Root MSE      =  .26857

                             (Std. Err. adjusted for 152 clusters in code_numeric)
------------------------------------------------------------------------------
             |               Robust
         dem |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         inc |   .2310104   .0113533    20.35   0.000     .2085785    .2534422
       _cons |  -1.339442   .0975336   -13.73   0.000    -1.532149   -1.146735
------------------------------------------------------------------------------
```

i. Interpret the estimated coefficient assuming the relationship is causal.

   Answer: Note that inc_{it} is measured in logs, hence the coefficient is a semi-elasticity. Assuming the relationship is causal, the estimated coefficient of 0.23 means that on average an increase of a country's per capita income by 100 percent raises its democracy score by 0.23.

ii. Is the relationship quantitatively relevant? To answer this question, compare two groups of countries. Group 1 countries had a democracy score of 1 and an average log per capita income of 9.57 in the year 2000 (this includes quite a few countries including the EU member states). Group 2 countries had a democracy score of 0.5 and an average log per capita income of 7.85 (this includes countries like Albania, Burkina Faso, Kuwait, Paraguay, Turkey, and Ukraine). By how much could have the latter countries, according to the estimated model, closed the democracy gap by fully catching up economically?

Answer: Based on the estimated relationship we can use

$$\Delta \widehat{dem}_{it} = 0.23 \Delta inc_{it}.$$

Hence, according to the model, an increase in income of $9.57 - 7.85 = 1.72$ would cause on average an increase in the democracy score by $0.23 \times 1.72 \approx 0.4$ to 0.9 index points. Thus, if group 2 countries had caught up economically, they would have almost reached the democracy level of group 1 countries.

iii. Discuss whether the estimated relationship should be interpreted as being causal.

Answer: There are good reason to suspect the estimated relationship may not be causal. In particular, inc_{it} may be endogenous and thus the OLS estimator may be biased. In the tutorial you discuss several reasons why this may be the case here. Another concern is reverse causality — it is even possible that the causality runs the opposite way: democratizing a country may cause prosperity to rise. Another concern is that unobserved country-specific factors affecting both income and democracy may lead to estimation bias. Acemoglu et al. (2008) argue as follows: "The major source of potential bias in a regression of democracy on income per capita is country-specific, historical factors influencing both political and economic development. If these omitted characteristics are, to a first approximation, time-invariant, the inclusion of fixed effects will remove them and this source of bias. Consider, for example, the comparison of the United States and Colombia. The United States is both richer and more democratic, so a simple cross-country comparison ... would suggest that higher per capita income causes democracy. The idea of fixed effects is to move beyond this comparison and investigate the 'within-country variation,' that is, to ask whether Colombia is more likely to *become* (relatively) democratic as it *becomes* (relatively) richer."

(b) A pooled OLS estimation of (1) yields (robust s.e.'s in brackets below the estimates)

$$\widehat{dem}_{it} = \underset{(0.035)}{0.706} dem_{i,t-1} + \underset{(0.010)}{0.072} inc_{i,t-1} + \hat{\mu}_t, \tag{2}$$

i. Discuss the pros and cons of adding a full set of time dummies.

Answer: Pro: Time dummies control for aggregate (=world) developments that affect all countries. Neglecting them may induce a spurious correlation between the regressors and the disturbance. For example, there might be a general tendency towards democracy over time unrelated to income, and technological progress unrelated to democracy may lead to rising income. Not controlling for these unrelated common trends may lead to the spurious finding that income and democracy correlate. Con: if these trends are related because on world average rising income leads to more democracy, we are "controlling out" this aggregate causal relationship. (A good robustness check would thus be to leave the time dummies out. It turns out that the estimation results remain largely unchanged.)

ii. Why may it be sensible to include income with a lag?

Answer: If income affects democracy, then certainly with a long lag because political institutions are persistent and change takes time. (Note that the data are sampled at five-year intervals, so this is a lag of five years which seems to be the minimum frequency that might be relevant.)

iii. What is the rationale behind including the lagged democracy score as a regressor?

Answer: The lagged democracy score is included as a regressor to capture persistence in democracy and also potentially mean-reverting dynamics, i.e., the potential tendency of the democracy score to return to some equilibrium value.

iv. Compute the short-term and long-term effects of an increase in income by 100 percent.

Answer: The short-term (=five year) effect is conditional on the past level of democracy, $dem_{i,t-1}$, and is $0.072 \times 1.0 = 0.072$. Hence, given the past level of democracy, an increase in income by 100 percent leads on average to an increase in the democracy score within five years by 0.072 which appears to be a moderate step. The long-term effect effect is reached when the democracy adjustment process comes to an end. As you may know from Econometrics II, this is the case when dem does not change any more and thus $dem = dem_{it} = dem_{i,t-1}$. It implies the long-run relationship (neglecting the time dummies)

$$\widehat{dem} = 0.706\,\widehat{dem} + 0.072\,inc \quad \Rightarrow \quad \widehat{dem} = \frac{0.072}{1-0.706}\,inc \approx 0.245\,inc.$$

Hence, in the long term (possibly after many decades) an increase in income by 100 percent raises the democracy score on average by $0.245 \times 1.0 = 0.245$. This is almost 25 percent of the full range (recall $0 \leq dem_{it} \leq 1$) and thus a large step.

(c) A fixed effects estimation of (1) yields the results presented below.

```
. xtreg dem L.dem L.inc yr3-yr10 if sample==1, fe vce(robust)

Fixed-effects (within) regression               Number of obs      =       945
Group variable: code_numeric                    Number of groups   =       150

R-sq:  within  = 0.2417                         Obs per group: min =         1
       between = 0.8845                                        avg =       6.3
       overall = 0.6772                                        max =         9

                                                F(10,149)          =     16.06
corr(u_i, Xb)  = 0.7546                         Prob > F           =    0.0000

                         (Std. Err. adjusted for 150 clusters in code_numeric)
------------------------------------------------------------------------------
             |               Robust
         dem |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         dem |
         L1. |   .3786284   .0466924     8.11   0.000     .2863636    .4708931
             |
         inc |
         L1. |    .010415   .0316728     0.33   0.743    -.0521709    .0730009
             |
         yr3 |   -.044566   .0338314    -1.32   0.190    -.1114174    .0222853
         yr4 |  -.0744071   .0301114    -2.47   0.015    -.1339076   -.0149067
         yr5 |  -.1781914   .0311613    -5.72   0.000    -.2397666   -.1166163
         yr6 |   -.133589   .0286265    -4.67   0.000    -.1901554   -.0770226
         yr7 |  -.0731129   .0288599    -2.53   0.012    -.1301405   -.0160854
         yr8 |  -.0780685   .0253412    -3.08   0.002    -.1281431   -.0279939
         yr9 |  -.0432207   .0195065    -2.22   0.028    -.0817659   -.0046756
        yr10 |  -.0028151   .0190529    -0.15   0.883    -.0404639    .0348337
       _cons |   .3343984   .2696243     1.24   0.217    -.1983827    .8671796
-------------+----------------------------------------------------------------
     sigma_u |  .20460922
     sigma_e |  .18004117
         rho |   .5636116   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

i. Interpret the overall, within, and between R^2.

<u>Answer:</u> The overall R^2 of 0.68 means that the model explains two thirds of the variation in y_{it} (the correlation between $\hat{y}_{it}$ and y_{it} is 0.68). The within R^2 of 0.24 means that the model explains roughly a quarter of the within variation of y_{it} (the correlation between $\hat{\ddot{y}}_{it}$ and $\ddot{y}_{it}$ is 0.24). The between R^2 of 0.88 means that the model explains almost 90 percent of the between variation of y_{it} (the correlation between $\hat{\bar{y}}_t$ and $\bar{y}_t$ is 0.24). Hence, the model is much better suited to explain differences in democracy between countries than changes of democracy of countries over time. (In fact, a FE estimator that only includes $inc_{i,t-1}$ as a regressor has an R^2 of 0.02. Hence, most of the within variation in democracy is unrelated to income.)

ii. Which of the POLS assumptions seems to be violated?

Answer: The regression output says the sample correlation of $\hat{c}_i$ and $\mathbf{X}_i\hat{\boldsymbol{\beta}}$ is 0.75. Even though it does not include a significance statement, it indicates that the POLS assumption $\mathrm{E}[\mathbf{x}_{it}'c_i] = \mathbf{0}$ is violated.

iii. Compute the short-term and long-term effects of an increase in income by 100 percent.

Answer: The short-term (=five year) effect conditional on the past level of democracy, $dem_{i,t-1}$, and is $0.010 \times 1.0 = 0.01$. Hence, given the past level of democracy, an increase in income by 100 percent leads on average to an increase in the democracy score within five years that is hardly measurable. The long-term effect effect is

$$\Delta\widehat{dem} = \frac{0.010}{1 - 0.379}\Delta inc = 0.016 \times 1.0 \approx 0.016$$

and again very small.

iv. Give a potential explanation for why the estimation result differs so much from the POLS results (2).

Answer: The major difference between POLS and FE is that the latter allows arbitrary correlation between regressors and unobserved individual effect c_i, while the former does not. Given that the sample correlation of $\hat{c}_i$ and $\mathbf{X}_i\hat{\boldsymbol{\beta}}$ is 0.75, there is strong evidence for a violation of the POLS assumption. This suggests that there is a positive correlation between $inc_{i,t-1}$ and c_i which biases the POLS estimator upwards. The underlying reason are most probably the country-specific, historical factors influencing both political and economic development Acemoglu et al. (2008) have in mind (see above and have a look at the paper).

v. Are you confident that the FE estimation results are valid?

Answer: While the FE estimator is robust to correlation between $inc_{i,t-1}$ and c_i, it is not robust to correlation between $inc_{i,t-1}$ and u_{it}. For example, there might be reverse causality running from democracy to income. This may still bias the results.[2] In addition, the lagged endogenous variable violates the strict exogeneity assumption. Therefore, even the FE estimator is inconsistent here.

[2]In addition, Cervellati et al. (2014, AER) find that the effect of income on democracy is negative for one group of countries and positive for another group while being zero on average over all countries.

Question 4

Lundberg and Rose (2002) estimate the effect of the number of kids on fathers' labor supply and wage.[3] They consider the following two specifications:

$$y_{it} = \beta_1 MARR_{it} + \beta_2 NKID04_{it} + \beta_3 DKID5_{it} + \sum_j \beta_{age,j} DAGE_{j,it} + \sum_k \beta_{year,k} DYEAR_{k,it} + \sum_l \beta_{educ,l} DEDUC_{l,it} + c_i + u_{it} \quad (3)$$

and

$$y_{it} = \beta_1 MARR_{it} + \sum_{m=1}^{4} \beta_{nkid,m} DKID_{m,it} + \beta_3 DKID5_{it} + \sum_j \beta_{age,j} DAGE_{j,it} + \sum_k \beta_{year,k} DYEAR_{k,it} + \sum_l \beta_{educ,l} DEDUC_{l,it} + c_i + u_{it}, \quad (4)$$

where y_{it} is the outcome variable (either the log of the real hourly wage rate, or annual hours of work), $MARR_{it}$ is a marriage dummy (1=married), $NKID04_{it}$ is the number of kids if the man has four children or less and zero otherwise, $DKID5_{it}$ is a dummy variable for five or more children (1=at least five kids), and $DKID_{m,it}$ is a dummy variable indicating whether the man has exactly m kids. In addition, $DAGE_{j,it}$ is a series of dummy variables for each year of age of the individual, $DYEAR_{k,it}$ is a series of dummy variables representing the year of the observation, and $DEDUC_{l,it}$ is a series of dummy variables indicating the number of years of education.

Tab. 1: Estimation results taken from Lundberg and Rose, 2002, p. 260

TABLE 5A.—THE EFFECT OF MARRIAGE AND CHILDREN ON ANNUAL HOURS WORKED (ENTIRE SAMPLE) (N = 26205)

	(1) OLS	(2) OLS	(3) OLS	(4) FE	(5) FE	(6) FE
Married	200.679	160.945	148.516	115.325	111.264	103.686
	(24.560)	(24.645)	(24.892)	(16.327)	(16.335)	(16.470)
Number of children (0 if none or >4)		45.86			38.416	
		(10.245)			(7.266)	
(Exactly) one child			68.297			82.023
			(22.983)			(14.849)
(Exactly) two children			138.562			108.165
			(25.595)			(17.729)
(Exactly) three children			138.922			113.230
			(34.375)			(24.544)
(Exactly) four children			126.268			152.212
			(66.625)			(36.551)
More than four children		−57.497	−34.916		38.074	49.624
		(133.137)	(132.643)		(62.147)	(62.319)
Two children − one child			*70.265*			*26.142*
			(24.215)			*(13.554)*
Three children − two children			*0.360*			*5.065*
			(30)			*(17.907)*
Four children − three children			*−12.654*			*38.982*
			(63.27)			*(31.111)*
R^2	0.04	0.04	0.04	0.45	0.45	0.45

Additional regressors include dummy variables for year of observation, years of education, and age. Standard errors in parentheses.

(a) Why is it potentially important to control for age, year, and education effects? For each group of dummies give an example why leaving them out can lead to inconsistent estimates of the effect of the number of kids on fathers' labor supply or wage.

Answer:

These covariates are likely to be correlated with both, the outcome "Annual Hours Worked" and the explanatory variables of interest. As a consequence, omitting age, year, and education effects would induce omitted-variable (OV) bias, i.e. it would render the explanatory variables of interest endogenous.

Age: The average number of children almost certainly increases with age. At the same time, Annual Hours Worked most likely is a function of age. Please note that including age-dummies (and not just a continuous age-variable) accounts for nonlinearities in this relationship, e.g. Hours worked increasing in the early years of a career, but decreasing in the later years (inverse U-shape). If excluded, number of children might just pick up the effects of age (OV-Bias).

Year: Year-dummies account for cyclical effects which affect all observations alike. For instance, in recession years, Annual Hours Worked will decrease, on average. At the same time, number of children might increase or decrease, on average. (On the one hand, the opportunity costs of reproduction might decrease, if there is a shortage in labor demand. On the other hand, uncertainty about the future could lead to the average individual discounting the value of children more strongly.) The one way or the other, cyclical effects should always be accounted for in panel estimations. NOTE: year dummies cannot be simultaneously included with the age variable(s) in the FE regression. Within individual, the variance of age is constant over time (every year, individuals get one year older). Thus, year dummies are perfectly collinear to the age dummies. The authors decided to drop the year dummies from their FE regressions.

Education: Again, education might relate to both, an average individual's number of children AND his/her hours worked. One might expect a negative relationship between educational attainment and number of children. At the same time, educational attainment might be positively correlated with Annual hours worked. If this was the case, the estimates on the number of children would be downward-biased, if education would not be controlled for.

(b) What kinds of variables are captured in c_i? Give a few relevant examples.

Answer:

c_i captures unobserved, individual-level variables which are related to the outcome (Annual Hours Worked) and do not change over time (c_i is time-invariant). For instance, this might relate to an individual's motivation / joy of working / leisure time preferences, his/her productivity/ability/skills, or just accuracy in reporting working hours.

Please note again that c_i captures these effects only to the degree that they are time-constant, i.e. they do not change over the period of analysis. [Hardly any observable characteristic is perfectly constant over an observation's life cycle. But it may be fairly stable, particularly within a given time-series]. The more relevant question is whether these effects are also related to the regressors, i.e. number of children.

(c) Explain how you would estimate this model taking the assumptions of the various estimators into account.

Answer:

The crucial assumptions are on the relationship between the c_i and the x_i's. Assuming the former to be correlated with the latter would ask for a FE or a FD model. In this specific case, one would opt for a FE model, as long as there is no reason to assume that the unobserved errors follow a random walk. Assuming $cov(x, c) = 0$ would allow for POLS or RE. Given that we know about the presence of c_i, one would in this case opt for a RE model, which is more efficient.

In the concretes example, it is most likely that $cov(x, c) \neq 0$. For instance, unobserved time preference are likely to affect both an individual's desire to have children and the time s/he spends on the job. Accordingly, a FE model seems to be most appropriate.

(d) Discuss the strict exogeneity assumption for the number of kids. Why may it fail?

Answer:

Very generally, the exogeneity assumption $E(x_{it}' u_{it}) = 0$ for $t = 1, \ldots, T$ may fail in the given example, since the idiosyncratic error is most likely correlated with the explanatory variables. The FE estimator only helps to account time-invariant OV bias. However, time-variant unobserved factors affecting both the number of children and the Hours worked in a given year are still not accounted for. Thus, estimates should be interpreted with caution.

Even if we were willing to accept $E(x_{it}' u_{it}) = 0$ for $t = 1, \ldots, T$, the strict exogeneity assumption $E(x_{is}' u_{it}) = 0$ for $s, t = 1, \ldots, T$ is still likely to be violated in the given example. It is plausible to assume serial autocorrelation in the errors. The decision to have $n + s$ children naturally depends on the decision to have n children. Since the estimation results suggest that having children per se affects the outcome, u_{it} must relate to i_{it+1}. To account for that serial autocorrelation, robust standard errors should be used.

(e) Table 1 reports some of their regression results using both pooled OLS and FE estimation.

i. Explain why the R^2 is (so much) higher for the FE estimator than for pooled OLS. Hint: Read Chapter 10.5.3 in the textbook about the least squares dummy variables (LSDV) estimator and conjecture that Lundberg and Rose used this estimator.

Answer:

With the LSDV estimator, the authors include $1 * N$ additional variables (i.e. 2,243 dummies) into the regression. Thus, it is not too surprising that the FE-model in this particular specification explains much more of the variation in the data than the simple OLS. Looking at the adjusted R^2 would be more meaningful, but still the increase in the R^2 would rather be a side-effect of the particular way in which OV-bias is accounted for (i.e. by including a mass of dummy variables).

ii. (*) Prove that the LSDV estimator yields, in a standard error components model, the same estimator of $\boldsymbol{\beta}$ as the FE estimator. Hint: Have a look at the Frisch-Waugh-Lovell-Theorem at Wikipedia and show that the transformation matrix $\mathbf{M}_{\mathbf{X}_1}$ used there is identical to our $\mathbf{I}_N \otimes \mathbf{Q}_T$ matrix.

Answer: — not discussed in the tutorial —

iii. Which of the estimation methods do you trust more?

Answer:

Even if the OLS assumptions held, RE would be the more efficient estimator. However, given that the OLS estimates almost certainly suffer from an OV-bias, the RE estimates would be biased as well. I would put much more confidence in the FE estimates.

iv. Now interpret the FE results. Compared to an unmarried man without kids, how many hours per year does a man work more who (1) is married and has one kid, (2) is married an has four kids (3), is married and has six kids? Compare the results of the two specifications. Discuss.

Answer:

Focus on Column 6 (FE) and compare to column 3 (OLS).

(1) In the FE: This guy works on average 103.686+82.023 = 185.709 hours more per year holding everything else constant. In the OLS: This guy works on average 148.686+68.297 = 216.983 hours more per year holding everything else constant.

(2) In the FE: This guy works on average 103.686+152.212 = 255.898 hours more per year holding everything else constant. In the OLS: This guy works on average 148.686+126.268 = 274.954 hours more per year holding everything else constant.

(3) In the FE: This guy works on average 103.686+49.624 = 153.31 hours more per year holding everything else constant. In the OLS: This guy works on average 148.686-34.916 = 113.77 hours more per year holding everything else constant.

Discussion: Please note that the effect result from comparison with an unmarried guy without kids. Only the second child gives a significant markup over the first child, additional children do not make a significant difference in statistical terms. It would be interesting to see the number of children interacted with marriage.

There seems to be upward-bias in the OLS estimates of “married”. OV positively affect both, the decision to marry and to work more (e.g. self-selection into marriage and job were one has to work much).

The OLS estimates on children seem to be downward-biased. OV negatively affects the probability to having children, but positively affects the hours worked (e.g. self-selection into working much but not having children).

Overall, OLS tends to over-estimate the aggregate effect but for a large number of children.

Both estimators consistently imply nonlinearities in the relationship between number of children and hours worked.