Panel Econometrics Exam: Exercise 3

Chris Vennemann

August, 9 2021

```r
library(haven)
library(plm)
library(lmtest)
```

```r
exam = read_dta("exam.dta")
```

---

1. Estimate an OLS regression. Interpret the results (coefficients, significance). What are possible problems with this specification?

```r
pool = plm(y~x3, effect = "individual", model = "pooling", data = exam)
summary(pool)
```

```
## Pooling Model
##
## Call:
## plm(formula = y ~ x3, data = exam, effect = "individual", model = "pooling")
##
## Balanced Panel: n = 7, T = 10, N = 70
##
## Residuals:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -9.59e+09 -1.71e+09  6.81e+07  0.00e+00  1.48e+09  7.21e+09
##
## Coefficients:
##               Estimate Std. Error t-value  Pr(>|t|)
## (Intercept) 1774092859  410539362  4.3214 5.178e-05 ***
## x3            93166933  252554257  0.3689    0.7133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    6.2729e+20
## Residual Sum of Squares: 6.2604e+20
## R-Squared:      0.0019973
## Adj. R-Squared: -0.012679
## F-statistic: 0.136086 on 1 and 68 DF, p-value: 0.71335
```

The estimated coefficient is $x_3 = 93166933$. It is positive but not statistically different from zero. Therefore, based on the pooling model, x_3 does not affect the dependent variable y.

The pooled OLS specification neglects the time dimension of the panel and basically treats the data as if they were 70 independent observations. However, with panel data the observations are generally not independent because the error term is often serially correlated within individuals. Also, the unit-specific effects are excluded in the simple OLS model. Therefore, it does not account for possible unobserved time-invariant heterogeneity among the countries. As a consequence, if the model with unit-specific effects is the true model, pooled OLS yields biased estimates of x_3 because of omitted variables.

---

2. Estimate a LSDV model without intercept but with individual effects. State the model equations and interpret the results (coefficients, significance).

```
lsdv = lm(y~-1 +x3 + factor(country), data = exam)
summary(lsdv)
```

```
##
## Call:
## lm(formula = y ~ -1 + x3 + factor(country), data = exam)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -8.477e+09 -1.580e+09  4.827e+08  1.501e+09  5.560e+09
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## x3               2.064e+08  3.723e+08   0.554 0.581290
## factor(country)1 1.538e+09  9.788e+08   1.571 0.121183
## factor(country)2 2.206e+08  9.168e+08   0.241 0.810681
## factor(country)3 1.236e+09  9.376e+08   1.318 0.192384
## factor(country)4 3.603e+09  9.193e+08   3.919 0.000225 ***
## factor(country)5 7.139e+08  9.169e+08   0.779 0.439201
## factor(country)6 3.487e+09  9.259e+08   3.766 0.000371 ***
## factor(country)7 1.017e+09  1.544e+09   0.658 0.512767
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.899e+09 on 62 degrees of freedom
## Multiple R-squared:  0.398,  Adjusted R-squared:  0.3204
## F-statistic: 5.124 on 8 and 62 DF,  p-value: 6.401e-05
```

The least squares dummy variables model specification is given by

$$y_{it} = \beta x_{3,it} + \sum_{i=1}^{7} \delta_i D_i + u_{it}, \qquad i = 1, 2, ..., 7; \qquad t = 1, 2, ..., 10$$

where the D_i equal 1 for the i^{th} unit and zero otherwise.

Again, the estimate on x_3 is positive but insignificant suggesting that x_3 does not affect y. The dummy variables for country 4 and 6 are positive and significantly different from zero meaning that —controlling for x_3—unobserved time-invariant variables induce significantly higher realizations of y for countries 4 and 6 (high and significant unit-specific intercepts).

---

3. Estimate the random effects model with the Swamy-Arora variance estimation. State the model equations and report/interpret the results (coefficients, p-values/significance).

```
random = plm(y~x3, effect = "individual", model = "random", random.method = "swar",
             data = exam)
summary(random)
```

```
## Oneway (individual) effect Random Effect Model
```

```
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = y ~ x3, data = exam, effect = "individual", model = "random",
##     random.method = "swar")
##
## Balanced Panel: n = 7, T = 10, N = 70
##
## Effects:
##                     var   std.dev share
## idiosyncratic 8.404e+18 2.899e+09 0.872
## individual    1.232e+18 1.110e+09 0.128
## theta: 0.3632
##
## Residuals:
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -9.17e+09 -1.39e+09  2.64e+08  0.00e+00  1.56e+09  6.63e+09
##
## Coefficients:
##               Estimate Std. Error z-value Pr(>|z|)
## (Intercept) 1741207562  585733269  2.9727 0.002952 **
## x3           136331931  296117678  0.4604 0.645231
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    5.6567e+20
## Residual Sum of Squares: 5.6391e+20
## R-Squared:      0.0031075
## Adj. R-Squared: -0.011553
## Chisq: 0.211966 on 1 DF, p-value: 0.64523
```

The random effects model equation is:

$$y_{it} = \alpha + \beta x_{3,it} + \mu_i + u_{it}, \qquad \mu_i \sim IID(0, \sigma_\mu^2)$$

The estimated coefficient for x_3 is positive but insignificant. Hence, x_3 does not have an impact on y.

---

4. What are the main differences in the underlying assumptions of the fixed and the random effects model?

In contrast to the fixed effects model, where the μ_i are parameters to be estimated, the random effects model assumes that they are random variables which are i.i.d. with zero mean and variance σ_μ^2 and independent of the other error component $u_{it} \sim IID(0, \sigma_u^2)$. Another assumption of the random effects model is that x_3 is independent of both μ_i and u_{it}, whereas in the fixed effects model x_3 is assumed to be correlated with μ_i. Due to the additional assumptions the variance-covariance structure of the error term is different between both models. In the random effects model the variance-covariance matrix is block-diagonal with $\sigma_\mu^2 + \sigma_u^2$ on the main diagonal and σ_μ^2 on the off-diagonal elements of each block.

---

5. Use an F-test to test whether the fixed effects model is a superior choice over the pooled OLS model. Report and interpret your results.

```
pFtest(y~x3, effect = "individual", data = exam)
```

```
##
##  F test for individual effects
```

```
##
## data:  y ~ x3
## F = 2.0819, df1 = 6, df2 = 62, p-value = 0.06808
## alternative hypothesis: significant effects
```

Based on the 5% significance level, the null hypothesis of no significant unit-specific effects should not be rejected because p-value $= 0.06808 > 0.05$.

The null of this F-test is that the unit-specific effects are jointly zero. Therefore, it is based on the difference in RSS between the model where the null restriction is active (pooling model) and the the unrestricted model (LSDV/within model). Given that two of the unit fixed effects were highly significant in the LSDV model, the improved RSS of the LSDV model (compared to the pooling RSS) results in a F-statistic whose corresponding p-value is close to but not quite below the 0.05 benchmark.

---

6. Conduct the Hausmann test. Report and interpret your results.

```
phtest(y~x3, effect = "individual", data = exam)
```

```
##
##  Hausman Test
##
## data:  y ~ x3
## chisq = 0.096436, df = 1, p-value = 0.7561
## alternative hypothesis: one model is inconsistent
```

The null hypothesis is $H_0 : E(e_{it}|x_{3,it}) = 0$ with $e_{it} = \mu_i + u_{it}$. The fixed effects estimator is consistent irrespective of whether H_0 is true or not because the within transformation rules the fixed effects out either way. The random effects estimator, however, is inconsistent when H_0 is not true but more efficient than fixed effects when H_0 is true. Given that the test statistic of the Hausman test is based on the difference of the fixed and random effects estimates, a high test statistic suggests that the null is not true. In this case one should opt to use the fixed effects estimator due to its consistency. Is the test statistic low, then this indicates that the null is true and the random effects estimator should be used due to its advantage in efficiency.

Since p-value $= 0.7561 > 0.05$, the null hypothesis should not be rejected. This means that it is better to use random effects as it is more efficient than fixed effects under the null hypothesis.

---

7. Test if the random effects model is superior to a simple pooled OLS regression.

```
plmtest(pool, type = "bp")
```

```
##
##  Lagrange Multiplier Test - (Breusch-Pagan) for balanced panels
##
## data:  y ~ x3
## chisq = 1.717, df = 1, p-value = 0.1901
## alternative hypothesis: significant effects
```

The null should not be rejected because p-value $= 0.1901 > 0.05$. Then, given that $H_0 : \sigma_\mu^2 = 0$, the pooled OLS model, where all countries have a common intercept (i.e. when the μ_i do not vary across countries), fits the data better than the random effects model.

---

8. Test for heteroskedasticity. Calculate the standard errors for the fixed effects model using the heteroskedasticity-consistent covariance matrix.

```r
bptest(pool, data = exam)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  pool
## BP = 3.7515, df = 1, p-value = 0.05276
```

Because p-value $= 0.05276 > 0.05$ the null hypothesis (homosekdasticity) should not be rejected based on the 5% significance level.

```r
### Estimate a fixed effects model
within = plm(y~x3, effect = "individual", model = "within", data = exam)
coeftest(within)
```

```
##
## t test of coefficients:
##
##     Estimate Std. Error t value Pr(>|t|)
## x3 206402047  372288407  0.5544   0.5813
```

```r
coeftest(within, vcov. = vcovHC, method="white1")
```

```
##
## t test of coefficients:
##
##     Estimate Std. Error t value Pr(>|t|)
## x3 206402047  236871280  0.8714   0.3869
```

```r
### One may use method="arellano" to be on the safe side (accounts also for serial correlation);
```
