# In-Class Lab 15

ECON 4223 (Prof. Tyler Ransom, U of Oklahoma)

November 17, 2020

The purpose of this in-class lab is to use R to practice with panel data estimation methods. To get credit, upload your .R script to the appropriate place on Canvas.

## For starters

Open up a new R script (named `ICL15_XYZ.R`, where `XYZ` are your initials) and add the usual "preamble" to the top:

```
library(tidyverse)
library(wooldridge)
library(broom)
library(magrittr)
library(clubSandwich)
library(modelsummary)
library(estimatr)
library(plm)    # You may need to install this package
```

### Load the data

Our data set will be a panel of wages for 545 men. Load the data from the `wooldridge` package, format `year` to be a factor, and rename the variable `nr` to something more descriptive (`id`):

```
df <- as_tibble(wagepan)
df %<>% mutate(year=as.factor(year))
df %<>% rename(id = nr)
```

### Summary statistics for panel data

It is important to know what your panel looks like. How many units? How many time periods? The easiest way to do this is the `pdim()` function in the `plm` package.

```
pdim(df)
```

It is also helpful to "convert" the data to a cross-section of within-unit averages. Let's do this for some of the key variables of our analysis.

```
df.within <- df %>% select(id,year,educ,married,union,rur) %>%
             group_by(id) %>%
             summarize(
                 mean.edu = mean(educ),
                 var.edu  = var(educ),
                mean.marr = mean(married),
```

```r
                 var.marr = var(married),
               mean.union = mean(union),
                var.union = var(union),
               mean.rural = mean(rur),
                var.rural = var(rur)
             )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
df.within %>% datasummary_skim()
```

1. Is there any within-person variance in the `educ` variable? What about `married`, `union`, and `rural`?
2. What does it mean for the married, union, or rural variables to have a positive within-person variance?
3. Why is it important to know if a variable has positive within-person variance?

## Pooled OLS, Random Effects, and Fixed Effects Models

Now estimate the following model using various options of the `plm()` function:

$$
\log(wage_{it}) = \beta_0 + \beta_1 educ_i + \beta_2 black_i + \beta_3 hisp_i + \beta_4 exper_{it} + \beta_5 exper_{it}^2 + \beta_6 married_{it} + \\
\beta_7 union_{it} + \beta_8 rur_{it} + \sum_t \beta_{9,t} year_{it} + a_i + u_{it}
$$

### Pooled OLS

The pooled OLS model can be run from the `lm_robust()` function as follows.

```r
est.pols <- lm_robust(lwage ~ educ + black + hisp + exper + I(exper^2) + married + union + rur + year,
                data = df, clusters=id)
```

4. Interpret the coefficient on $\beta_7$ in the pooled OLS model

### Random effects

RE uses the `plm()` function as follows.

```r
est.re <- plm(lwage ~ educ + black + hisp + exper + I(exper^2) + married + union + rur + year,
              data = df, index = c("id","year"), model = "random")
```

```
## Warning in Ops.pseries(y, bX): indexes of pseries have same length but not same
## content: result was assigned first operand's index

## Warning in Ops.pseries(y, bX): indexes of pseries have same length but not same
## content: result was assigned first operand's index

## Warning in Ops.pseries(y, bX): indexes of pseries have same length but not same
## content: result was assigned first operand's index
```

5. What is the estimate of $\theta$ in the RE model? (Hint: check `est.re$ercomp$theta`) What does this tell you about what you expect the random effects estimates to be relative to the fixed effects estimates?

### Fixed effects

FE also come from the `lm_robust()` function:

```
est.fe <- lm_robust(lwage ~ I(exper^2) + married + union + rur + year,
              data = df, fixed_effects = ~id)
```

6. Explain why we cannot estimate coefficients on `educ`, `black`, `hisp`, or `exper` in the fixed effects model. (Note: the reason for not being able to estimate `exper` is more nuanced)

## Clustered standard errors

As discussed in class, the most appropriate standard errors account for within-person serial correlation and are robust to heteroskedasticity. We already used these for pooled OLS and fixed effects, but not for random effects.

```
clust.re <- coef_test(est.re, vcov = "CR1", cluster = "individual")
clust.re.SE <- clust.re$SE
names(clust.re.SE) <- names(est.re$coefficients)
```

We can put these all in one `modelsummary` table:

```
modelsummary(list("POLS"=est.pols,"RE"=est.re,"FE"=est.fe),
             statistic_override=list(sqrt(diag(est.pols$vcov)),clust.re.SE,sqrt(diag(est.fe$vcov))),
             output="latex")
```

| | POLS | RE | FE |
|---|---|---|---|
| (Intercept) | 0.165 | 0.036 | |
| | (0.163) | (0.161) | |
| educ | 0.087 | 0.091 | |
| | (0.011) | (0.011) | |
| black | -0.149 | -0.141 | |
| | (0.051) | (0.051) | |
| hisp | -0.016 | 0.016 | |
| | (0.040) | (0.040) | |
| exper | 0.069 | 0.106 | |
| | (0.019) | (0.016) | |
| I(exper^2) | -0.002 | -0.005 | -0.005 |
| | (0.001) | (0.001) | (0.001) |
| married | 0.126 | 0.065 | 0.047 |
| | (0.026) | (0.019) | (0.018) |
| union | 0.182 | 0.107 | 0.079 |
| | (0.028) | (0.021) | (0.019) |
| rur | -0.138 | -0.023 | 0.049 |
| | (0.035) | (0.031) | (0.031) |
| year1981 | 0.053 | 0.040 | 0.152 |
| | (0.028) | (0.028) | (0.027) |
| year1982 | 0.055 | 0.030 | 0.254 |
| | (0.037) | (0.035) | (0.028) |
| year1983 | 0.049 | 0.019 | 0.357 |
| | (0.046) | (0.044) | (0.032) |
| year1984 | 0.074 | 0.041 | 0.494 |
| | (0.057) | (0.055) | (0.040) |
| year1985 | 0.090 | 0.056 | 0.622 |
| | (0.066) | (0.064) | (0.048) |
| year1986 | 0.120 | 0.089 | 0.771 |
| | (0.076) | (0.075) | (0.059) |
| year1987 | 0.151 | 0.132 | 0.931 |
| | (0.085) | (0.085) | (0.070) |
| Num.Obs. | 4360 | 4360 | 4360 |
| R2 | 0.199 | 0.181 | 0.621 |
| R2 Adj. | 0.196 | 0.178 | 0.566 |
| se_type | CR2 | | HC2 |

7. Interpret the coefficient on `union` across the three models.
8. Comment on the comparability of the pooled OLS and RE estimators when within-person serial correlation has been properly addressed.