ARE 213 PS 2a

S. Sung, H. Husain, T. Woolley, A. Watt

2021-11-08

Contents

Packages

```
library(tidyverse)
library(haven)
library(plm)
library(lmtest)
library(sandwich)
library(stargazer)
```

```r
library(ggplot2)
library(gridExtra)
library(grid)
library(gtable)
library(tinytex)
library(fastDummies)
library(EnvStats)
```

Problem 1

Question 10.3 from Wooldridge: For $T = 2$ consider the standard unobserved effects model:

$$y_{it} = \alpha + x_{it}\beta + c_i + u_{it} \tag{1}$$

Let $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ represent the fixed effects and first differences estimators respectively.

Part (a)

Show that $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical. Hint: it may be easier to write $\hat{\beta}_{FE}$ as the "within estimator" rather than the fixed effects estimator.

Writing $\hat{\beta}_{FE}$ as the within estimator, $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are given by

$$\hat{\beta}_{FD} = (\Delta X'\Delta X)^{-1}(\Delta X'\Delta y) \qquad \text{and} \qquad \hat{\beta}_{FE} = \left(\ddot{X}'\ddot{X}\right)^{-1}\left(\ddot{X}'\ddot{y}\right)$$

Expanding the inner products, we have

$$\hat{\beta}_{FD} = \left(\sum_i \sum_t \Delta X_{it}'\Delta X_{it}\right)^{-1}\left(\sum_i \sum_t \Delta X_{it}'\Delta y_{it}\right)$$

and

$$\hat{\beta}_{FE} = \left(\sum_i \sum_t \ddot{X}_{it}'\ddot{X}_{it}\right)^{-1}\left(\sum_i \sum_t \ddot{X}_{it}'\ddot{y}_{it}\right)$$

Since there are only two periods, $\hat{\beta}_{FD}$ simplifies to

$$\hat{\beta}_{FD} = \left(\sum_i \Delta X_i'\Delta X_i\right)^{-1}\left(\sum_i \Delta X_i'\Delta y_i\right)$$

where

$$\Delta X_i \equiv X_{i2} - X_{i1} \qquad \text{and} \qquad \Delta y_i \equiv y_{i2} - y_{i1}$$

Now we note that

$$\ddot{X}_{i1} = X_{i1} - \frac{1}{2}(X_{i1} + X_{i2}) = \frac{1}{2}(X_{i1} - X_{i2}) = -\frac{1}{2}\Delta X_i$$

and similiarly

$$\ddot{X}_{i2} = \frac{1}{2}\Delta X_i, \quad \ddot{y}_{i1} = -\frac{1}{2}\Delta y_i \quad \ddot{y}_{i2} = \frac{1}{2}\Delta y_i$$

Then, $\hat{\beta}_{FE}$ becomes

$$\begin{aligned}
\hat{\beta}_{FE} &= \left(\sum_i \sum_t \ddot{X}_{it}' \ddot{X}_{it}\right)^{-1} \left(\sum_i \sum_t \ddot{X}_{it}' \ddot{y}_{it}\right) \\
&= \left(\sum_i \frac{1}{4} \Delta X_i' \Delta X_i + \frac{1}{4} \Delta X_i' \Delta X_i\right)^{-1} \left(\sum_i \frac{1}{4} \Delta X_i' \Delta y_i + \frac{1}{4} \Delta X_i' \Delta y_i\right) \\
&= \left(\frac{1}{2} \sum_i \Delta X_i' \Delta X_i\right)^{-1} \left(\frac{1}{2} \sum_i \Delta X_i' \Delta y_i\right) \\
&= \left(\sum_i \Delta X_i' \Delta X_i\right)^{-1} \left(\sum_i \Delta X_i' \Delta y_i\right) \\
&= \hat{\beta}_{FD}
\end{aligned}$$

So $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical.

Part (b)

Show that the standard errors of $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical. If you wish, you may assume that x_{it} is a scalar (i.e. there is only one regressor) and ignore any degree of freedom corrections. You are not clustering the standard errors in this problem.

The standard errors are estimates of the square root of the asymptotic variances of our estimators, so WLOG, we can compare the asymptotic variances. The asymptotic variances of our estimators are

$$\widehat{Avar(\hat{\beta}_{FE})} = \hat{\sigma}^2_{u,FE}\left(\ddot{X}'\ddot{X}\right)^{-1} \qquad \text{and} \qquad \widehat{Avar(\hat{\beta}_{FE})} = \hat{\sigma}^2_{u,FD}\left(\Delta X'\Delta X\right)^{-1}$$

where $\hat{\sigma}^2_{u,FE}$ and $\hat{\sigma}^2_{u,FD}$ are estimated from the residuals of the corresponding regressions and using the correct degrees of freedom:

$$\hat{\sigma}^2_{u,FE} = \frac{\sum_i \sum_t \hat{\ddot{u}}^2_{it}}{N(T-1)-K} = \frac{\sum_i \sum_t \hat{\ddot{u}}^2_{it}}{N-K} \qquad \text{and} \qquad \hat{\sigma}^2_{u,FD} = \frac{\sum_i \widehat{\Delta u_i}^2}{N(T-1)-K} = \frac{\sum_i \widehat{\Delta u_i}^2}{N-K}$$

Let $\hat{\beta} := \hat{\beta}_{FD} = \hat{\beta}_{FE}$. Then, from part (a), we can find the relationship between $\widehat{\Delta u_i}$ and $\hat{\ddot{u}}_{it}$:

$$\begin{aligned}
\hat{\ddot{u}}^2_{it} &= (\ddot{y}_{it} - \ddot{X}_{it}\hat{\beta})^2 \\
&= \left((-1)^t\left(\frac{1}{2}\Delta y_i - \frac{1}{2}\Delta X_i\hat{\beta}\right)\right)^2 \\
&= \frac{1}{4}\left(\Delta y_i - \Delta X_i\hat{\beta}\right)^2 \\
&= \frac{1}{4}\widehat{\Delta u_i}^2
\end{aligned}$$

So the estimated error variances are related by

$$\begin{aligned}
\hat{\sigma}^2_{u,FE} &= \frac{\sum_i \sum_t \hat{\ddot{u}}^2_{it}}{N-K} \\
&= \frac{\sum_i \sum_t \frac{1}{4}\widehat{\Delta u_i}^2}{N-K} \\
&= \frac{\sum_i \frac{1}{2}\widehat{\Delta u_i}^2}{N-K} \\
&= \frac{1}{2}\frac{\sum_i \widehat{\Delta u_i}^2}{N-K} \\
&= \frac{1}{2}\hat{\sigma}^2_{u,FD}
\end{aligned}$$

We know from part (a) that

$$\begin{aligned}
\left(\sum_i \sum_t \ddot{X}'_{it}\ddot{X}_{it}\right)^{-1} &= \left(\frac{1}{2}\sum_i \Delta X'_i \Delta X_i\right)^{-1} \\
&= 2\left(\sum_i \Delta X'_i \Delta X_i\right)^{-1}
\end{aligned}$$

And putting all these together, we have

$$
\begin{aligned}
\widehat{Avar(\hat{\beta}_{FE})} &= \hat{\sigma}^2_{u,FE}\left(\ddot{X}'\ddot{X}\right)^{-1} \\
&= \frac{1}{2}\hat{\sigma}^2_{u,FD}2\left(\sum_i \Delta X_i'\Delta X_i\right)^{-1} \\
&= \hat{\sigma}^2_{u,FD}\left(\sum_i \Delta X_i'\Delta X_i\right)^{-1} \\
&= \widehat{Avar(\hat{\beta}_{FE})}
\end{aligned}
$$

Because the estimates of the asymptotic variances are equal, the standard errors (the square roots) will be equal.

Problem 2

Question 21-3 from Cameron-Trivedi (enhanced): Consider the fixed effects, two-way error component panel data model:

$$y_{it} = \alpha + x_{it}\beta + \mu_i + \lambda_t + \epsilon_{it} \tag{2}$$

Part (a)

Show that the fixed effects estimator of β can be obtained by applying two within (one-way) transformations on this model. The first is the within transformation ignoring the time effects followed by the within transformation ignoring the individual effects. Assume the panel is balanced. (Hint: it may be easier to analyze the fixed effects regression using partitioned regression.)

We want to show that

$$\dddot{y}_{it} = \dddot{x}_{it}\beta + \dddot{e}_{it}$$

Ignoring time effects, we get

$$\begin{aligned}
\ddot{y}_{it} &= y_{it} - \overline{y_i} = y_{it} - \frac{1}{T}\sum_t y_{it} \\
\implies \ddot{y}_{it} &= y_{it} - \frac{1}{T}\sum_t (\alpha + x_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})
\end{aligned}$$

Applying the second within transformation, we get that

$$\begin{aligned}
\dddot{y}_{it} &= \ddot{y}_{it} - \overline{\ddot{y}_t} \\
&= \ddot{y}_{it} - \frac{1}{N}\sum_i \ddot{y}_{it} \\
&= y_{it} - \overline{y_i} - \overline{y_t} + \overline{y} \\
&= \alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it} \\
&\quad - \left(\frac{1}{T}\sum_t(\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\
&\quad - \left(\frac{1}{N}\sum_i(\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\
&\quad + \left(\frac{1}{NT}\sum_i\sum_t(\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\
&= \beta\left(x_{it} - \overline{x_i} - \overline{x_t} + \overline{x}\right) + \epsilon_{it} - \overline{\epsilon_i} - \overline{\epsilon_t} + \overline{\epsilon}
\end{aligned}$$

Since

$$\dddot{x}_{it} = x_{it} - \overline{x_i} - \overline{x_t} + \overline{x}$$

and

$$\dddot{e}_{it} = e_{it} - \overline{e_i} - \overline{e_t} + \overline{e}$$

We get that

$$\dddot{y}_{it} = \dddot{x}_{it}\beta + \dddot{e}_{it}$$

Part (b)

Show that the order of the transformations is unimportant. Give an intuitive explanation for why.

Reversing the order, we can show that we get the same result

Again, we want to show that

$$\dddot{y}_{it} = \dddot{x}_{it}\beta + \dddot{e}_{it}$$

Ignoring individual effects, we get

$$\ddot{y}_{it} = y_{it} - \overline{y_t} = y_{it} - \frac{1}{N}\sum_i y_{it}$$
$$\implies \ddot{y}_{it} = y_{it} - \frac{1}{N}\sum_i (\alpha + x_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})$$

Applying the second within transformation, we get that

$$\begin{aligned}
\dddot{y}_{it} &= \ddot{y}_{it} - \overline{\ddot{y}_{it}} \\
&= \ddot{y}_{it} - \frac{1}{T}\sum_i \ddot{y}_{it} \\
&= y_{it} - \overline{y_t} - \overline{y_i} + \overline{y} \\
&= \alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it} \\
&\quad - \left(\frac{1}{N}\sum_i (\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\
&\quad - \left(\frac{1}{T}\sum_t (\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\
&\quad + \left(\frac{1}{NT}\sum_i \sum_t (\alpha + X_{it}\beta + \mu_i + \lambda_t + \epsilon_{it})\right) \\
&= \beta\left(x_{it} - \overline{x_t} - \overline{x_i} + \overline{x}\right) + \epsilon_{it} - \overline{\epsilon_t} - \overline{\epsilon_i} + \overline{\epsilon}
\end{aligned}$$

Since

$$\dddot{x}_{it} = x_{it} - \overline{x_t} - \overline{x_i} + \overline{x}$$

and

$$\ddot{e}_{it} = e_{it} - \overline{e_t} - \overline{e_i} + \overline{e}$$

We get that

$$\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{e}_{it}$$

Intuitively, the order of the transformations is unimportant because in the end, were still manage to difference out the individual and time effects. There is nothing particular to individual or time effects that would warrant removal in any particular order.

Part (c)

Does your answer to part (a) change if the panel becomes unbalanced (i.e., contains different numbers of observations for each individual i). Why or why not?

Yes – attrition from the panel may be related to the dependent variable and we might get selection bias. To address selection bias caused by non-ignorable attrition, we would need to weight by the inverse propensity score.

Problem 3

We now begin with an actual analysis of the data. The goal here is to determine what effect, if any, primary belt laws have on the log of traffic fatalities per capita (we log the LHS variable because we believe the effect of safety belt laws should be proportional to the overall level of fatalities per capita).

```r
data = read_dta('traffic_safety2.dta') %>%
    filter(state != 99) %>%
    mutate(fatal_per_cap = fatalities / population,
           vmt_per_cap = totalvmt/population)
```

Part (a)

Run pooled bivariate OLS. Interpret. Add year fixed effects. Interpret. Add all covariates that you believe are appropriate. Think carefully about which covariates should be log transformed and which should enter in levels. What happens when you add these covariates? Why?

```r
# Bivariate with lm
reg_a_bivariate_lm = lm(log(fatal_per_cap) ~ primary , data = data)
# Bivariate with plm
reg_a_bivariate_plm = plm(log(fatal_per_cap) ~ primary,
                       data = data,
                       model = "pooling")
# FE with lm
reg_a_yfe_lm = lm(log(fatal_per_cap) ~ primary + factor(year), data = data)
# FE with plm
reg_a_yfe_plm = plm(log(fatal_per_cap) ~ primary,
                    data = data,
                    index = c("state", "year"),  # order matters: unit-var, time-var
                    model = "within",
                    effect = "time")  # only do time

# FE + covars with lm
reg_a_full_lm = lm(log(fatal_per_cap) ~ primary + factor(year) + college + beer
                   + secondary + unemploy + log(vmt_per_cap) + log(precip) + snow32
                   + log(rural_speed) + log(urban_speed), data = data)
# FE + covars with plm
reg_a_full_plm = plm(log(fatal_per_cap) ~ primary + college + beer
                     + secondary + unemploy + log(vmt_per_cap) + log(precip) + snow32
                     + log(rural_speed) + log(urban_speed),
                     data = data,
                     index = c("state", "year"),  # order matters: unit-var, time-var
                     model = "within",
                     effect = "time")  # only do time
```

```r
stargazer(reg_a_bivariate_plm, reg_a_yfe_plm, reg_a_full_plm, type='text')
```

==

Dependent variable:

—————————————————————— log(fatal_per_cap)

(1) (2) (3)

——— primary -0.155*** -0.091*** 0.031
(0.027) (0.028) (0.020)

college -2.387***
(0.146)

beer 0.176***
(0.025)

secondary 0.038**
(0.017)

unemploy 0.017***
(0.004)

log(vmt_per_cap) 1.154***
(0.041)

log(precip) -0.051***
(0.012)

snow32 -0.171***
(0.014)

log(rural_speed) 0.554***
(0.127)

log(urban_speed) 0.182**
(0.091)

Constant -1.705***
(0.011)

Observations 1,104 1,104 1,104 R2 0.029 0.010 0.744 Adjusted R2 0.028 -0.011 0.736 F Statistic 32.799***
(df = 1; 1102) 10.848*** (df = 1; 1080) 311.366*** (df = 10; 1071)
==
Note: *p<0.1;* ***p<0.05;*** p<0.01

primary -0.0936 0.0789 (0.1641) (0.1366)

Covariates No Yes
Observations 48 48
R2 0.0070 0.8598
Adjusted R2 -0.0146 0.8220
=== Note: *p<0.1;* ***p<0.05;*** p<0.01

Part (d)

Compute the Random Effects estimator (including covariates). Under what conditions will this give an unbiased estimate of the effect of primary seat belt laws on fatalities per capita? What are its advantages or disadvantages as compared to pooled OLS?

```
reg_3d_RE_bin <- plm(ln_fat_pc ~ primary,
                     data = data,
                     model = "random")

reg_3d_RE_cov <- plm(ln_fat_pc ~ primary + secondary + college + beer + unemploy
```

```r
                            + ln_vmt_pc + ln_precip + snow32 + ln_rspeed + ln_uspeed,
                            data = data,
                            model = "random")

reg_3d_pooled_bin <- plm(ln_fat_pc ~ primary,
                            data = data,
                            model = "pooling")

reg_3d_pooled_cov <- plm(ln_fat_pc ~ primary + secondary + college + beer + unemploy
                            + ln_vmt_pc + ln_precip + snow32 + ln_rspeed + ln_uspeed,
                            data = data,
                            model = "pooling")

stargazer(reg_3d_RE_bin, reg_3d_RE_cov, reg_3d_pooled_bin, reg_3d_pooled_cov,
          title = "Random Effects vs Pooled",
          dep.var.caption = "Log(Fatality per Population)",
          dep.var.labels.include = FALSE, model.names = FALSE,
          column.labels = c("RE", "RE", "Pooled", "Pooled"),
          keep = c("primary"),
          add.lines=list(c('Covariates', 'No', 'Yes', 'No', 'Yes')),
          font.size = "footnotesize", column.sep.width = "1pt", no.space = TRUE, omit.stat=c("f", "ser")
          type = "text", digits = 4)
```

```
##
## Random Effects vs Pooled
## ==========================================================
##                      Log(Fatality per Population)
##               --------------------------------------------
##                   RE         RE        Pooled     Pooled
##                  (1)        (2)         (3)        (4)
## ----------------------------------------------------------
## primary      -0.2243*** -0.1493*** -0.1549*** -0.1407***
##                (0.0167)   (0.0161)   (0.0270)   (0.0210)
## ----------------------------------------------------------
## Covariates        No         Yes         No         Yes
## Observations   1,104      1,104      1,104      1,104
## R2             0.1411     0.5807     0.0289     0.6743
## Adjusted R2    0.1403     0.5768     0.0280     0.6713
## ==========================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
```

Random Effects model will be unbiased when state-level unobserved heterogeneity is uncorrelated with the independent variables. If this were the case, then RE is more efficient than FE. However, as evidenced by 3a, this assumption does not not appear to hold.

Compared to pooled OLS, random effects assumes that within-state residuals are equally correlated with each other (as opposed to assuming no correlation between them in OLS). Random effects then used these residuals to compute weighted least squares, using the inverse of the variance-covariance matrices as weights. Similar to OLS, RE still assumed that across-state residuals are uncorrelated.

The advantage of using RE is that it is more efficient than pooled OLS in the case when we have reason to think that states have exist time-invariant unobserved characteristics (which islikely the case here). However, when we think those unobservables are correlated with the independent variables, we would be better off using FE.

Part (e)

Do you think the standard errors from RE are right? Compute the clustered standard errors. Are they substantially different? If so, why? (i.e., what assumption(s) are being violated?)

```
reg_3d_RE_cov_cluster <- coeftest(reg_3d_RE_cov, vcov = vcovHC(reg_3d_RE_cov, type = "sss", cluster = "g
reg_3d_pooled_cov_cluster <- coeftest(reg_3d_pooled_cov, vcov = vcovHC(reg_3d_pooled_cov, type = "sss",

stargazer(reg_3d_RE_cov, reg_3d_RE_cov_cluster, reg_3d_pooled_cov, reg_3d_pooled_cov_cluster,
          title = "Regression with Clustered SE",
          dep.var.caption = "Log(Fatality per Population)",
          dep.var.labels.include = FALSE, model.names = FALSE,
          column.labels = c("RE", "RE(Cluster)", "Pooled", "Pool(Cluster)"),
          keep = "primary",
          add.lines=list(c('Covariates', 'Yes', 'Yes', 'Yes', 'Yes')),
          font.size = "footnotesize", column.sep.width = "1pt", no.space = TRUE, omit.stat=c("f", "ser")
          type = "text", digits = 4)
```

```
##
## Regression with Clustered SE
## ============================================================
##                        Log(Fatality per Population)
##                 --------------------------------------------
##                     RE     RE(Cluster)   Pooled   Pool(Cluster)
##                    (1)         (2)         (3)          (4)
## ------------------------------------------------------------
## primary      -0.1493*** -0.1493***  -0.1407***  -0.1407***
##                (0.0161)   (0.0324)    (0.0210)    (0.0490)
## ------------------------------------------------------------
## Covariates        Yes         Yes         Yes          Yes
## Observations     1,104                   1,104
## R2              0.5807                  0.6743
## Adjusted R2     0.5768                  0.6713
## ============================================================
## Note:                             *p<0.1; **p<0.05; ***p<0.01
```

```
# Test for serial correlation
# serialCorrelationTest(reg_3d_RE_cov, test = "rank.von.Neumann",
#     alternative = "two.sided", conf.level = 0.95)

pdwtest(reg_3d_RE_cov, alternative="two.sided")
```

```
##
##  Durbin-Watson test for serial correlation in panel models
##
## data:  ln_fat_pc ~ primary + secondary + college + beer + unemploy +  ...
## DW = 0.90152, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

The standard errors are different with and without clustering because there exists serial correlation in the error term (as evidenced by the Durbin-Watson test above). Clustered standard errors correct for these estimates.

Part (f)

Compute the FE estimator using only primary and year fixed effects as the covariates. Compute the normal standard errors and the clustered standard errors. If they are different, why?

```
reg_a_yfe_plm_cluster <- coeftest(reg_a_yfe_plm, vcov = vcovHC(reg_a_yfe_plm, method = "white1", type =

stargazer(reg_a_yfe_plm, reg_a_yfe_plm_cluster,
          title = "FE with Clustered SE",
          dep.var.caption = "Log(Fatality per Population)",
          dep.var.labels.include = FALSE, model.names = FALSE,
          column.labels = c("FE", "FE(Cluster)"),
          keep = "primary",
          add.lines=list(c('Covariates', 'No', 'No')),
          font.size = "footnotesize", column.sep.width = "1pt", no.space = TRUE, omit.stat=c("f", "ser")
          type = "text", digits = 4)
```

```
##
## FE with Clustered SE
## ===========================================
##              Log(Fatality per Population)
##              ----------------------------
##                    FE        FE(Cluster)
##                   (1)            (2)
## -------------------------------------------
## primary        -0.0906***     -0.0906***
##                 (0.0275)       (0.0296)
## -------------------------------------------
## Covariates         No             No
## Observations     1,104
## R2               0.0099
## Adjusted R2     -0.0111
## ===========================================
## Note:          *p<0.1; **p<0.05; ***p<0.01
```

```
# Test for serial correlation
# serialCorrelationTest(reg_3d_RE_cov, test = "rank.von.Neumann",
#     alternative = "two.sided", conf.level = 0.95)

pdwtest(reg_a_yfe_plm, alternative="two.sided")
```

```
##
##  Durbin-Watson test for serial correlation in panel models
##
## data:  log(fatal_per_cap) ~ primary
## DW = 0.12953, p-value < 2.2e-16
## alternative hypothesis: serial correlation in idiosyncratic errors
```

The estimate of FE with only primary and year FE is computed in 3a. The standard errors are surprisingly similar with and without clustering, though still the SE is larger in the clustered case. The Durbin-Watson test indicates that there is indeed positive autocorrelation and that it is stronger here than under the previous RE model. (Also note that we interpreted the question to be that we were asked to only estimate year FE; not state FE with year FE covariates.)

Part (g)

Add the same range of covariates to the FE estimator that you did to the OLS estimator. Are the FE estimates more or less stable than the OLS estimates? Why?

```r
# OLS + covars with lm
reg_g_ols_lm = lm(log(fatal_per_cap) ~ primary + college + beer
                        + secondary + unemploy + log(vmt_per_cap) + log(precip) + snow32
                   + log(rural_speed) + log(urban_speed), data = data)

stargazer(reg_a_bivariate_lm, reg_g_ols_lm,reg_a_yfe_lm, reg_a_full_lm,
          title = "OLS and FE with covariates",
          dep.var.caption = "Log(Fatality per Population)",
          dep.var.labels.include = FALSE, model.names = FALSE,
          column.labels = c("OLS", "OLS", "FE", "FE"),
          keep = "primary",
          add.lines=list(c('Covariates', 'No', 'Yes', 'No', 'Yes')),
          font.size = "footnotesize", column.sep.width = "1pt", no.space = TRUE, omit.stat=c("f", "ser")
          type = "text", digits = 4)
```

```
##
## OLS and FE with covariates
## =========================================================
##                     Log(Fatality per Population)
##                ------------------------------------------
##                   OLS        OLS         FE         FE
##                   (1)        (2)         (3)        (4)
## ---------------------------------------------------------
## primary      -0.1549*** -0.1407*** -0.0906***  0.0309
##                (0.0270)   (0.0210)   (0.0275)  (0.0204)
## ---------------------------------------------------------
## Covariates       No         Yes         No        Yes
## Observations   1,104      1,104       1,104     1,104
## R2             0.0289     0.6743      0.1072    0.7692
## Adjusted R2    0.0280     0.6713      0.0881    0.7623
## =========================================================
## Note:                         *p<0.1; **p<0.05; ***p<0.01
```

The FE estimates are less stable than the OLS estimates when adding covariates. This is likely due to the fact that the FE estimate without covariates is exploiting variation between states with very different characteristics. Once those characteristics are controlled for with covariates, the effect disappears. On the other hand, OLS is exploiting variation across both time and state. When the covariates are added, it takes some variation away from the state dimension, but is still (perhaps unjustifiably) exploiting the time dimension.

Part (h)

Estimate a first-differences estimator, a 5-year differences estimator, and a long differences estimator, including year fixed effects (when feasible) and the appropriate covariates in each case. Briefly describe the pattern that emerges from the three differencing estimates. Where does the FE estimate fall in this pattern? Are you surprised?

```r
df_diff <- data %>% arrange(state, year)
varlist <- names(df_diff)[3:21]

# First Differences
for (i in varlist) {
df_diff <- df_diff %>%
              group_by(state) %>%
              mutate("{i}_diff" := eval(as.symbol(i)) - dplyr::lag(eval(as.symbol(i))))
}

reg3h_1 = plm(ln_fat_pc_diff ~ primary_diff + college_diff + beer_diff
                     + secondary_diff + unemploy_diff + ln_vmt_pc_diff + ln_precip_diff + snow32_diff
                     + ln_rspeed_diff + ln_uspeed_diff,
                     data = df_diff,
                     index = c("state", "year"),
                     model = "within",
                     effect = "time")

# Five Year Difference Estimator
df_diff <- data %>% arrange(state, year)

for (i in varlist) {
df_diff <- df_diff %>%
              group_by(state) %>%
              mutate("{i}_diff" := eval(as.symbol(i)) - dplyr::lag(eval(as.symbol(i)), n= 2))
}

reg3h_2 = plm(ln_fat_pc_diff ~ primary_diff + college_diff + beer_diff
                     + secondary_diff + unemploy_diff + ln_vmt_pc_diff + ln_precip_diff + snow32_diff
                     + ln_rspeed_diff + ln_uspeed_diff,
                     data = df_diff,
                     index = c("state", "year"),
                     model = "within",
                     effect = "time")


# Long Differences Estimator
df_diff <- data %>% arrange(state, year)

for (i in varlist) {
df_diff <- df_diff %>%
              group_by(state) %>%
              mutate("{i}_diff" := eval(as.symbol(i)) - dplyr::lag(eval(as.symbol(i)), n= 22))
}

reg3h_3 = plm(ln_fat_pc_diff ~ primary_diff + college_diff + beer_diff
                     + secondary_diff + unemploy_diff + ln_vmt_pc_diff + ln_precip_diff + snow32_diff
                     + ln_rspeed_diff + ln_uspeed_diff,
                     data = df_diff,
                     index = c("state", "year"),
                     model = "pooling")

# Table
```

```r
stargazer(reg3h_1, reg3h_2, reg3h_3,
          title = "First, Second, and Long Difference Estimator",
          dep.var.caption = "Differences in Log(Fatality per Population)",
          dep.var.labels.include = FALSE, model.names = FALSE,
          column.labels = c("1st", "2nd", "Long"),
          add.lines=list(c('Time FE', 'Yes', 'Yes', 'No')),
          font.size = "footnotesize", column.sep.width = "1pt", no.space = TRUE, omit.stat=c("f", "ser")
          type = "text", digits = 4)
```

```
##
## First, Second, and Long Difference Estimator
## ============================================================
##                 Differences in Log(Fatality per Population)
##                ---------------------------------------------
##                      1st            2nd            Long
##                      (1)            (2)             (3)
## ------------------------------------------------------------
## primary_diff      -0.0456**      -0.0465***       -0.0910
##                   (0.0225)       (0.0175)        (0.1622)
## college_diff       -0.6400        -0.4678         -0.2613
##                   (0.5299)       (0.3317)        (0.9933)
## beer_diff         0.2648***      0.4116***       0.6559***
##                   (0.0697)       (0.0605)        (0.1725)
## secondary_diff   -0.0377***     -0.0342***        0.0335
##                   (0.0136)       (0.0109)        (0.1624)
## unemploy_diff    -0.0140***     -0.0197***       -0.0450**
##                   (0.0039)       (0.0031)        (0.0168)
## ln_vmt_pc_diff    0.2896***      0.2481***        0.3315
##                   (0.0908)       (0.0720)        (0.2394)
## ln_precip_diff   -0.0815***     -0.0538***       -0.1113
##                   (0.0132)       (0.0146)        (0.1405)
## snow32_diff        0.0127         0.0160         0.1690**
##                   (0.0100)       (0.0106)        (0.0773)
## ln_rspeed_diff    -0.2438**       -0.0222         -0.0345
##                   (0.1158)       (0.0924)        (0.6765)
## ln_uspeed_diff     0.1176         0.1147          0.1655
##                   (0.0888)       (0.0722)        (0.2812)
## Constant                                         -0.4790*
##                                                  (0.2693)
## ------------------------------------------------------------
## Time FE              Yes            Yes             No
## Observations        1,056          1,008            48
## R2                 0.0855         0.1433          0.6480
## Adjusted R2        0.0578         0.1170          0.5528
## ============================================================
## Note:                            *p<0.1; **p<0.05; ***p<0.01
```

Part (i)

Make the case that the first-differences estimate is superior to the 5-year or long differences estimates.

First differences provides the researcher with more observations than the 5-year logn differences and therefore potentially more power.

Part (j)

Make the case that the 5-year or long differences estimates are superior to the first-differences estimate.

Five-year differences could potentially identify a larger effect than first-differences, though with less power (fewer observations). This may also be the strategy a researcher would want to take if they suspect that the treatment to take a while to fully take effect.