ARE 213 Problem Set 2A

Becky Cardinali, Yuen Ho, Sara Johns, and Jacob Lefler

Due 10/26/2020

Question 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.

(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.

Define $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ and $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$. Since $T = 2$, we have $\bar{Y}_i = \frac{Y_{i1}+Y_{i2}}{2}$ and $\bar{X}_i = \frac{X_{i1}+X_{i2}}{2}$.

Let $\ddot{Y}_{it} = Y_{it} - \bar{Y}_i$, $\ddot{X}_{it} = X_{it} - \bar{X}_i$, and $\ddot{\epsilon}_{it} = \epsilon_{it} - \bar{\epsilon}_i$.

The within estimator comes from running the following regression: $\ddot{Y}_{it} = \ddot{X}_{it}'\beta + \ddot{\epsilon}_{it}$. So the within estimator is given by $\hat{\beta}_W = (\ddot{X}_{it}'\ddot{X}_{it})^{-1}(\ddot{X}_{it}'\ddot{Y}_{it})$. As we know from Lecture Notes 3Ai (pg 12-13), the within estimator gives us estimates of β that are numerically identical to those produced by the FE estimator. So we have

$$\hat{\beta}_{FE} = \hat{\beta}_W = (\ddot{X}_{it}'\ddot{X}_{it})^{-1}(\ddot{X}_{it}'\ddot{Y}_{it})$$

Since $T = 2$ we can rewrite this as

$$\hat{\beta}_{FE} = \left[\begin{pmatrix}\ddot{X}_{i1}' & \ddot{X}_{i2}'\end{pmatrix}\begin{pmatrix}\ddot{X}_{i1} \\ \ddot{X}_{i2}\end{pmatrix}\right]^{-1}\begin{pmatrix}\ddot{X}_{i1}' & \ddot{X}_{i2}'\end{pmatrix}\begin{pmatrix}\ddot{Y}_{i1} \\ \ddot{Y}_{i2}\end{pmatrix} = \left(\ddot{X}_{i1}'\ddot{X}_{i1} + \ddot{X}_{i2}'\ddot{X}_{i2}\right)^{-1}\left(\ddot{X}_{i1}'\ddot{Y}_{i1} + \ddot{X}_{i2}'\ddot{Y}_{i2}\right)$$

Define $\Delta Y_{it} = Y_{it} - Y_{it-1}$, $\Delta X_{it} = X_{it} - X_{it-1}$, $\Delta\epsilon_{it} = \epsilon_{it} - \epsilon_{it-1}$. Then the regression $\Delta Y_{it} = \Delta X_{it}'\beta + \Delta\epsilon_{it}$ using data from time periods $2, ..., T$ yields the first differences estimator $\hat{\beta}_{FD} = (\Delta X_{it}'\Delta X_{it})^{-1}(\Delta X_{it}'\Delta Y_{it})$. Since we have $T = 2$, $\Delta Y_{it} = Y_{i2} - Y_{i1}$, $\Delta X_{it} = X_{i2} - X_{i1}$, $\Delta\epsilon_{it} = \epsilon_{i2} - \epsilon_{i1}$

From the definition above, we have $\ddot{X}_{i1} = X_{i1} - \bar{X}_i$. Substituting in for $\bar{X}_i$ and then using the definition of ΔX_{it} gives us

$$\ddot{X}_{i1} = X_{i1} - \frac{X_{i1}+X_{i2}}{2} = \frac{X_{i1}-X_{i2}}{2} = \frac{-\Delta X_{it}}{2}$$

Similarly,

$$\ddot{X}_{i2} = X_{i2} - \frac{X_{i1}+X_{i2}}{2} = \frac{X_{i2}-X_{i1}}{2} = \frac{\Delta X_{it}}{2}$$

$$\ddot{Y}_{i1} = Y_{i1} - \frac{Y_{i1}+Y_{i2}}{2} = \frac{Y_{i1}-Y_{i2}}{2} = \frac{-\Delta Y_{it}}{2}$$

$$\ddot{Y}_{i2} = Y_{i2} - \frac{Y_{i1}+Y_{i2}}{2} = \frac{Y_{i2}-Y_{i1}}{2} = \frac{\Delta Y_{it}}{2}$$

Substituting these values into $\hat{\beta}_{FE}$ gives

$$\hat{\beta}_{FE} = \left(\frac{-\Delta X_{it}'}{2}\frac{-\Delta X_{it}}{2} + \frac{\Delta X_{it}'}{2}\frac{\Delta X_{it}}{2}\right)^{-1}\left(\frac{-\Delta X_{it}'}{2}\frac{-\Delta Y_{it}}{2} + \frac{\Delta X_{it}'}{2}\frac{\Delta Y_{it}}{2}\right)$$
$$= \left(\frac{\Delta X_{it}'\Delta X_{it}}{4} + \frac{\Delta X_{it}'\Delta X_{it}}{4}\right)^{-1}\left(\frac{\Delta X_{it}'\Delta Y_{it}}{4} + \frac{\Delta X_{it}'\Delta Y_{it}}{4}\right)$$
$$= \left(\frac{\Delta X_{it}'\Delta X_{it}}{2}\right)^{-1}\left(\frac{\Delta X_{it}'\Delta Y_{it}}{2}\right) = (\frac{1}{2})^{-1}(\Delta X_{it}'\Delta X_{it})^{-1}(\frac{1}{2})(\Delta X_{it}'\Delta Y_{it})$$
$$= (\Delta X_{it}'\Delta X_{it})^{-1}(\Delta X_{it}'\Delta Y_{it}) = \hat{\beta}_{FD}$$

So for $T = 2$, $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical.

(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.

Ignoring any degree of freedom corrections, $Var(\hat{\beta}_{FE}) = \hat{\sigma}^2(\ddot{X}_{it}'\ddot{X}_{it})^{-1}$ where $\hat{\sigma}^2$ is the sum of squared residuals.

For $\hat{\beta}_{FE}$ and $T = 1$,

$$\hat{\ddot{\epsilon}}_{i1} = \ddot{Y}_{i1} - \ddot{X}_{i1}\hat{\beta}_{FE} = \frac{-\Delta Y_{it}}{2} - \frac{-\Delta X_{it}}{2}\hat{\beta}_{FE}$$

And since we showed in part (a) that $\hat{\beta}_{FE} = \hat{\beta}_{FD}$

$$= \frac{-\Delta Y_{it}}{2} - \frac{-\Delta X_{it}}{2}\hat{\beta}_{FD} = (-\frac{1}{2})(\Delta Y_{it} - \Delta X_{it}\hat{\beta}_{FD}) = (-\frac{1}{2})\Delta\hat{\epsilon}_{it}$$

So for $T = 1$, $(\hat{\ddot{\epsilon}}_{i1})^2 = \frac{1}{4}(\Delta\hat{\epsilon}_{it})^2$

Similarly, for $\hat{\beta}_{FE}$ and $T = 2$,

$$\hat{\ddot{\epsilon}}_{i2} = \ddot{Y}_{i2} - \ddot{X}_{i2}\hat{\beta}_{FE} = \frac{\Delta Y_{it}}{2} - \frac{\Delta X_{it}}{2}\hat{\beta}_{FE}$$

And since we showed in part (a) that $\hat{\beta}_{FE} = \hat{\beta}_{FD}$

$$= \frac{\Delta Y_{it}}{2} - \frac{\Delta X_{it}}{2}\hat{\beta}_{FD} = (\frac{1}{2})(\Delta Y_{it} - \Delta X_{it}\hat{\beta}_{FD}) = (\frac{1}{2})\Delta\hat{\epsilon}_{it}$$

So for $T = 2$, $(\hat{\ddot{\epsilon}}_{i2})^2 = \frac{1}{4}(\Delta\hat{\epsilon}_{it})^2$

Therefore, the sum of squared residuals $\hat{\sigma}^2$ for $\hat{\beta}_{FE}$ is $(\hat{\ddot{\epsilon}}_{i1})^2 + (\hat{\ddot{\epsilon}}_{i2})^2 = \frac{1}{2}(\Delta\hat{\epsilon}_{it})^2$

Subsituting in for $\hat{\sigma}^2$ and $\ddot{X}_{it}$, we have

$$Var(\hat{\beta}_{FE}) = \hat{\sigma}^2(\ddot{X}_{it}'\ddot{X}_{it})^{-1} = \frac{1}{2}(\Delta\hat{\epsilon}_{it})^2\left[\begin{pmatrix}\ddot{X}_{i1}' & \ddot{X}_{i2}'\end{pmatrix}\begin{pmatrix}\ddot{X}_{i1} \\ \ddot{X}_{i2}\end{pmatrix}\right]^{-1}$$
$$= \frac{1}{2}(\Delta\hat{\epsilon}_{it})^2\left(\ddot{X}_{i1}'\ddot{X}_{i1} + \ddot{X}_{i2}'\ddot{X}_{i2}\right)^{-1}$$
$$= \frac{1}{2}(\Delta\hat{\epsilon}_{it})^2\left(\frac{-\Delta X_{it}'}{2}\frac{-\Delta X_{it}}{2} + \frac{\Delta X_{it}'}{2}\frac{\Delta X_{it}}{2}\right)^{-1}$$
$$= \frac{1}{2}(\Delta\hat{\epsilon}_{it})^2\left(\frac{\Delta X_{it}'\Delta X_{it}}{4} + \frac{\Delta X_{it}'\Delta X_{it}}{4}\right)^{-1} = \frac{1}{2}(\Delta\hat{\epsilon}_{it})^2\left(\frac{\Delta X_{it}'\Delta X_{it}}{2}\right)^{-1}$$
$$= \frac{1}{2}(\Delta\hat{\epsilon}_{it})^2(\frac{1}{2})^{-1}(\Delta X_{it}'\Delta X_{it})^{-1} = (\Delta\hat{\epsilon}_{it})^2(\Delta X_{it}'\Delta X_{it})^{-1} = Var(\hat{\beta}_{FD})$$

since $(\Delta\hat{\epsilon}_{it})^2$ is the sum of squared residuals for $\hat{\beta}_{FD}$ because there are only 2 time periods, so $\Delta\epsilon_{it} = \epsilon_{i2} - \epsilon_{i1}$ is the only residual to square and include in the sum. We found that the variances are equal, and the standard error is just the square root of the variance.

So for $T = 2$ and ignoring any degree of freedom corrections, the standard errors of $\hat{\beta}_{FE}$ and $\hat{\beta}_{FD}$ are numerically identical.

Question 2

(a) We want to show that the fixed effects (FEs) estimator of β in $y_{it} = \alpha + x_{it}\beta + \mu_i + \lambda_t + \epsilon_{it}$ can be obtained from two within transformations on this model. We show this by applying the Frisch-Waugh-Lovell (FWL) theorem.

First, we let R be the matrix of unit dummy variables, similar to the one defined in the lecture notes (the only difference being that we drop one dummy from R to avoid perfect colinearity; for example, let $R_1 = \mathbf{0}$). Further define Q to be a matrix of time fixed effects, so $Q_{it,s} = \mathbf{1}(t = s)$. (Note that for each observation it we clearly have $R_{it} = R_i$ and $Q_{it} = Q_t$.) The fixed effects estimator would involve regressing y_{it} on X_{it}, R_i and Q_t.

The FWL theorem, however, says that we can obtain the same coefficients from performing three simpler regressions. The first step is to regress y_{it}, X_{it} and R_{it} on Q_{it} and obtain the residuals. This amounts to subtracting off the mean of each variable within each time period t. Define $\tilde{y}_{it} = y_{it} - \bar{y}_t$ where $\bar{y}_t := \frac{1}{T}\sum_t y_{it}$, and define $\tilde{X}_{it}$ and $\tilde{R}_{it}$ analogously.

Note that

$$\tilde{R}_{it} = R_{it} - \begin{bmatrix} 0 \\ 1/N \\ \vdots \\ 1/N \end{bmatrix} = R_{it} - \bar{R}_t$$

In the second step, we regress $\tilde{y}_{it}$ and $\tilde{X}_{it}$ on $\tilde{R}_{it}$ and obtain the residuals, $\ddot{y}_{it}$ and $\ddot{X}_{it}$. The FWL theorem then guarantees that a regression of $\ddot{y}_{it}$ on $\ddot{X}_{it}$ will yield the fixed effects estimate of β. In what sense is the second regression a within estimator? Note that $\bar{R}_t$ is a constant, so we are substracting a constant off of a set of variables. We can recall from introductory econometrics that substracting a constant from a variable will not change the estimated relationship between the dependent and independent variables.

There is a key difference between this second regression and a one-step within estimator: we had to drop the first dummy variable for unit 1 (or whichever unit we chose). Thus, this second regression is equivalent to regressing $\tilde{y}_{it}$ and $\tilde{X}_{it}$ on R_{it} and a constant.

(b)

In (a) we could have made two minor changes and obtained the same estimate of β after two "within" transformations. First, we need to retain the dummy variable for the first unit in R and dropped the dummy for the first time period (or some other time period) in Q to avoid colinearity. And second, we change the order of the regressions. Now we first regress y_{it}, X_{it} and Q_{it} on R_{it} and obtain the residuals $\tilde{y}_{it}$, $\tilde{X}_{it}$, and $\tilde{Q}_{it}$. Here we find that $\tilde{Q}_{it} = Q_{it} - \bar{Q}_i$ where in fact $\bar{Q}_i$ is constant across units i. The same logic implies that the second application of the FWL theorem can also be considered a within transformation, and of course our estimate of β will be unchanged since the FWL theorem implies that we can perform these sets of regressions in any order (and the estimate is robust to dropping any of the dummy variables since the dummies remain, in a sense, saturated).

Intuition: a two-way fixed effects estimator takes out the variation common to all units within a time period or to all observations of a unit. It should not matter what order we take out these components if the remaining variation is the same: we compare obserations of one unit but we ignore variation from time trends. Both paths lead to this final analysis.

(c)

In the case of unbalanced panel data, we no longer have the same expression for $\tilde{R}_{it}$ after regressing y_{it}, X_{it} and R_{it} on Q_{it}. Start by defining A_t to be the set of units i that are observed at time t. Then define N_t to be the cardinality of A_t (the number of observations observed at time t). We find that

$$\tilde{R}_{it} = R_{it} - \bar{R}_t$$

but now $\bar{R}_t$ is not constant across t. Instead,

$$\bar{R}_t = \frac{1}{N_t}\sum_{i \in A_t} R_{it}$$

Note, however, that this does not affect the logic from part (a). In particular, in the second within transformation, we are regressing on $R_{it} - \bar{R}_t$ and $\bar{R}_t$ is constant across i which is the relevent dimension. Thus, the unbalancedness does not affect our ability to calculate the fixed effects estimator with two within transformations.

Question 3

(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
# create y variable
traffic[, ln_fat_pc := log((fatalities/population))]
# log covariates
traffic[,ln_unemploy := log(unemploy)]
traffic[,ln_totalvmt := log(totalvmt)]
traffic[,ln_precip := log(precip)]
traffic[,ln_snow := log(snow32+0.01)] # to avoid NA from zeroes
# create dummies for FEs (to be used later)
traffic <- dummy_cols(traffic, select_columns = c("year", "state"))

# Pooled bivariate OLS
biv <- feols(ln_fat_pc ~ primary, data=traffic)
summary(biv, se="standard")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Standard-errors: Standard
##              Estimate Std. Error   t value  Pr(>|t|)
## (Intercept) -1.703100   0.010885 -156.4700 < 2.2e-16 ***
## primary     -0.143884   0.025839   -5.5684  3.21e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -353.50   Adj. R2: 0.02596
```

A naive bivariate OLS estimate suggests that having primary belt laws is associated with a 14.39% decrease in traffic fatalities per capita, ceteris paribus. Of course, we are likely omitting both observed and unobserved variables that affect selection into treatment (having primary belt laws), so this basic bivariate model simply gives us some descriptive information.

```r
# Pooled bivariate OLS with year fixed effects
biv_yfe <- feols(ln_fat_pc ~ primary, fixef = "year", data=traffic)
summary(biv_yfe, se = "standard")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Fixed-effects: year: 23
## Standard-errors: Standard
##          Estimate Std. Error t value Pr(>|t|)
## primary -0.079675   0.026249 -3.0353 0.002459 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -304.10   Adj. R2: 0.08992
##                         R2-Within: 0.00828
```

When we add year fixed effects to our pooled OLS model, we find that having primary belt laws is associated with a 7.97% decrease in traffic fatalities per capita, ceteris paribus. That is, some of the relationship between primary belt laws and traffic fatalities that we found in our bivariate model above can actually be explained by aggregate trends over time, which we control for when we include year fixed effects.

```r
# Pooled bivariate OLS with fixed effects and covariates
biv_yfe_cov <- feols(ln_fat_pc ~ primary + secondary + college +
                       beer + ln_unemploy + ln_totalvmt + ln_precip +
                       ln_snow + rural_speed + urban_speed, fixef = "year", data=traffic)
summary(biv_yfe_cov, se = "standard")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Fixed-effects: year: 23
## Standard-errors: Standard
##              Estimate Std. Error    t value  Pr(>|t|)
## primary      0.041908   0.025928   1.616300  0.106315
## secondary    0.055491   0.021486   2.582700  0.009933 **
## college     -3.005100   0.175500 -17.123000 < 2.2e-16 ***
## beer         0.195430   0.029526   6.618800  5.66e-11 ***
## ln_unemploy -0.022197   0.026950  -0.823636  0.410326
## ln_totalvmt -0.068970   0.007972  -8.652000 < 2.2e-16 ***
## ln_precip   -0.082981   0.015512  -5.349300  1.07e-07 ***
## ln_snow     -0.069961   0.003603 -19.418000 < 2.2e-16 ***
## rural_speed  0.020434   0.002426   8.422500 < 2.2e-16 ***
## urban_speed  0.002012   0.001791   1.123400  0.261501
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 223.22   Adj. R2: 0.64007
##                        R2-Within: 0.61098
```

Next we add covariates to our model, specifically secondary belt laws, percent college grads, per capita beer consumption, rural interstate speed limit, urban interstate speed limit, and logs of the unemployment rate, total vehicle miles travelled, percipitation, and snowfall. We decided to log those variables because they have a positively skewed distribution.

With these additional covariates, our results indicate that having primary belt laws is associated with a 4.19% increase in traffic fatalities per capita, ceteris paribus, which is not statistically significant. It seems unlikely that having primary belt laws would increase traffic fatalities, so it seems possible that our model suffers from omitted variable bias. For example, there may be state-specific, time-invariant characteristics which affect both the likelihood of having primary belt laws and traffic fatalities per capita.

(b) Ignore omitted variables bias issues for the moment. Do you think the standard errors from above are right? Compute the Huber-White heteroskedasticity robust standard errors. Do they change much? Compute the clustered standard errors that are robust to within-state correlation. Do this using both the canned command and manually using the formulas we learned in class. Do the standard errors change much? Are you surprised? Interpret.

Our OLS standard errors above are only correct if we assume homoskedasticity and independence across observations. However, it seems likely that observations in different years within a given state will be positively correlated, even if we assume state-specific factors are independent across states. To account for heteroskedasticity we can calculate the Huber-White heteroskedasticity robust standard errors. To account for within-state correlation, we can calculate clustered standard errors at the state-level.

```r
# package command - heteroskedastic robust standard errors
summary(biv, se = "white")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
```

```
## Standard-errors: White
##             Estimate Std. Error   t value  Pr(>|t|)
## (Intercept) -1.703100   0.010825 -157.3400 < 2.2e-16 ***
## primary     -0.143884   0.026356   -5.4593  5.88e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -353.50   Adj. R2: 0.02596
```

```r
summary(biv_yfe, se = "white")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Fixed-effects: year: 23
## Standard-errors: White
##         Estimate Std. Error t value Pr(>|t|)
## primary -0.079675   0.027778 -2.8683 0.004206 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -304.10   Adj. R2: 0.08992
##                         R2-Within: 0.00828
```

```r
summary(biv_yfe_cov, se = "white")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Fixed-effects: year: 23
## Standard-errors: White
##             Estimate Std. Error    t value  Pr(>|t|)
## primary      0.041908   0.022780   1.839700  0.066087 .
## secondary    0.055491   0.020067   2.765300  0.005783 **
## college     -3.005100   0.170602 -17.615000 < 2.2e-16 ***
## beer         0.195430   0.026649   7.333600  4.35e-13 ***
## ln_unemploy -0.022197   0.027119  -0.818521  0.413238
## ln_totalvmt -0.068970   0.009711  -7.102000  2.21e-12 ***
## ln_precip   -0.082981   0.017153  -4.837800   1.5e-06 ***
## ln_snow     -0.069961   0.003297 -21.222000 < 2.2e-16 ***
## rural_speed  0.020434   0.002398   8.520100 < 2.2e-16 ***
## urban_speed  0.002012   0.001783   1.128000  0.259571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 223.22   Adj. R2: 0.64007
##                        R2-Within: 0.61098
```

```r
# package command - clustered standard errors
summary(biv, cluster  = traffic$state)
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Standard-errors: Clustered
##             Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) -1.703100   0.045801 -37.1850 < 2.2e-16 ***
## primary     -0.143884   0.093883  -1.5326  0.125659
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -353.50   Adj. R2: 0.02596
```

```r
summary(biv_yfe, cluster = traffic$state)
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Fixed-effects: year: 23
## Standard-errors: Clustered
##          Estimate Std. Error   t value Pr(>|t|)
## primary -0.079675   0.106207 -0.750186 0.453303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -304.10   Adj. R2: 0.08992
##                         R2-Within: 0.00828
```

```r
summary(biv_yfe_cov, cluster = traffic$state)
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Fixed-effects: year: 23
## Standard-errors: Clustered
##              Estimate Std. Error   t value  Pr(>|t|)
## primary      0.041908   0.045876  0.913493  0.361184
## secondary    0.055491   0.039344  1.410400  0.158699
## college     -3.005100   0.504346 -5.958500  3.43e-09 ***
## beer         0.195430   0.079580  2.455800  0.014213 *
## ln_unemploy -0.022197   0.068344 -0.324786  0.745405
## ln_totalvmt -0.068970   0.037429 -1.842700  0.065642 .
## ln_precip   -0.082981   0.059004 -1.406400  0.159902
## ln_snow     -0.069961   0.008267 -8.463200 < 2.2e-16 ***
## rural_speed  0.020434   0.004472  4.569500  5.44e-06 ***
## urban_speed  0.002012   0.003406  0.590567  0.554933
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 223.22   Adj. R2: 0.64007
##                        R2-Within: 0.61098
```

Using the "canned" commands for calculating White-robust standard errors and clustered standard errors we find that our White-robust standard errors are generally slightly larger than our unadjusted OLS standard errors and that our clustered standard errors are larger than both the robust and unadjusted standard errors. We expect our clustered standard errors to be larger than our unclustered SEs the more we have intra-cluster correlation of regressors, the more observations we have per cluster, and the more correlated the errors are within-cluster. Intuitively, if errors are positively correlated within cluster then an additional observation in the cluster no longer provides a completely independent piece of new information.

```r
# Manually calculate White-robust and clustered standard errors

# Manually obtain beta OLS matrix
calc.beta <- function(xmat, ymat) {
  (solve(t(xmat)%*%xmat)) %*% (t(xmat)%*%ymat)
}

white_middle <- function(xmat, ymat, beta) {
  residsq <- diag(as.vector((ymat - xmat %*% beta)^2))
  mid <- (t(xmat)%*%residsq%*%xmat)
  return(mid)
}
```

```r
robust.se <- function(xmat, middle, adj) {

  var.robust <- adj * (solve(t(xmat)%*%xmat) %*% middle %*% solve(t(xmat)%*%xmat))

  se <- sqrt(diag(var.robust))

  return(se)
}

cluster_middle <- function(i, beta, DT, yvar, xvars) {

  state.xmat <- as.matrix(cbind(1,select(DT[state == i,], xvars)))
  state.ymat <- as.matrix(select(DT[state == i,], yvar))

  resid <- as.vector(state.ymat - state.xmat %*% beta)

  middle.term <- t(state.xmat) %*% resid %*% t(resid) %*% state.xmat

  return(middle.term)
}

# List of our variables for the three regressions
biv_var <- c("primary")
biv_yfe_var <- c("primary", colnames(traffic[,year_1982:year_2003]))
biv_yfe_cov_var <- c("primary", "secondary", "college", "beer",
                     "ln_unemploy", "ln_totalvmt", "ln_precip",
                     "ln_snow", "rural_speed", "urban_speed", colnames(traffic[,year_1982:year_2003]))

# Run regression
xmat_biv <- as.matrix(cbind(1,select(traffic, all_of(biv_var))))
xmat_biv_yfe <- as.matrix(cbind(1, select(traffic, all_of(biv_yfe_var))))
xmat_biv_yfe_cov <- as.matrix(cbind(1, select(traffic, all_of(biv_yfe_cov_var))))
ymat <- as.matrix(select(traffic, ln_fat_pc))

beta_biv <- calc.beta(xmat_biv, ymat)
beta_biv_yfe <- calc.beta(xmat_biv_yfe, ymat)
beta_biv_yfe_cov <- calc.beta(xmat_biv_yfe_cov, ymat)

# Manually calculate White robust SEs
# get middle terms
w_mid_biv <- white_middle(xmat_biv, ymat, beta_biv)
w_mid_biv_yfe <- white_middle(xmat_biv_yfe, ymat, beta_biv_yfe)
w_mid_biv_yfe_cov <- white_middle(xmat_biv_yfe_cov, ymat, beta_biv_yfe_cov)
# adjustment factor (so that it matches HC1)
white_biv_adj <- nrow(xmat_biv)/(nrow(xmat_biv)- ncol(xmat_biv))
white_biv_yfe_adj <- nrow(xmat_biv_yfe)/(nrow(xmat_biv_yfe)- ncol(xmat_biv_yfe))
white_biv_yfe_cov_adj <- nrow(xmat_biv_yfe_cov)/(nrow(xmat_biv_yfe_cov)- ncol(xmat_biv_yfe_cov))
# get standard errors
# bivariate
white_biv <- robust.se(xmat_biv, w_mid_biv, white_biv_adj)
white_biv
```

```
##         V1    primary
## 0.01082450 0.02635596
```

```r
# bivariate with year FEs
white_biv_yfe <- robust.se(xmat_biv_yfe, w_mid_biv_yfe, white_biv_yfe_adj)
white_biv_yfe[1:2]
```

```
##         V1    primary 
## 0.04432273 0.02777826
```

```r
# year FEs and covariates
white_biv_yfe_cov <- robust.se(xmat_biv_yfe_cov, w_mid_biv_yfe_cov, white_biv_yfe_cov_adj)
white_biv_yfe_cov[1:11]
```

```
##          V1     primary   secondary     college        beer ln_unemploy 
## 0.169961214 0.022779839 0.020067061 0.170602484 0.026648578 0.027118598 
## ln_totalvmt   ln_precip     ln_snow rural_speed urban_speed 
## 0.009711377 0.017152613 0.003296679 0.002398350 0.001783336
```

```r
# Clustered by state
states <- as.vector(unique(traffic[,state])) # list of states
# adjustment factor (from section)
cl_biv_adj <- (length(states)/(length(states)-1))*((nrow(xmat_biv)-1)/(nrow(xmat_biv)- ncol(xmat_biv)))
cl_biv_yfe_adj <- (length(states)/(length(states)-1))*((nrow(xmat_biv_yfe)-1)/(nrow(xmat_biv_yfe)- ncol(xmat_
cl_biv_yfe_cov_adj <- (length(states)/(length(states)-1))*((nrow(xmat_biv_yfe_cov)-1)/(nrow(xmat_biv_yfe_cov)

# middle terms
cl_mid_biv_terms <- lapply(states, cluster_middle, beta = beta_biv, DT = traffic,
                           yvar="ln_fat_pc", xvars=biv_var)
cl_mid_biv <- Reduce('+', cl_mid_biv_terms)

cl_mid_biv_yfe_terms <- lapply(states, cluster_middle, beta = beta_biv_yfe, DT = traffic,
                               yvar="ln_fat_pc", xvars=biv_yfe_var)
cl_mid_biv_yfe <- Reduce('+', cl_mid_biv_yfe_terms)

cl_mid_biv_yfe_cov_terms <- lapply(states, cluster_middle, beta = beta_biv_yfe_cov, DT = traffic,
                                   yvar="ln_fat_pc", xvars=biv_yfe_cov_var)
cl_mid_biv_yfe_cov <- Reduce('+', cl_mid_biv_yfe_cov_terms)
```

```r
# get standard errors
# bivariate
cl_biv <- robust.se(xmat_biv, cl_mid_biv, cl_biv_adj)
cl_biv
```

```
##         V1    primary 
## 0.04580143 0.09388303
```

```r
# bivariate with year FEs
cl_biv_yfe <- robust.se(xmat_biv_yfe, cl_mid_biv_yfe, cl_biv_yfe_adj)
cl_biv_yfe[1:2]
```

```
##         V1    primary 
## 0.04476217 0.10620747
```

```r
# year FEs and covariates
cl_biv_yfe_cov <- robust.se(xmat_biv_yfe_cov, cl_mid_biv_yfe_cov, cl_biv_yfe_cov_adj)
cl_biv_yfe_cov[1:11]
```

```
##          V1     primary   secondary     college        beer ln_unemploy 
## 0.515289301 0.045876099 0.039343758 0.504346199 0.079580243 0.068344038 
## ln_totalvmt   ln_precip     ln_snow rural_speed urban_speed 
## 0.037428642 0.059003902 0.008266521 0.004471839 0.003406197
```

When we calculate the White-robust and clustered standard errors by hand they are the same as the standard errors calculated by the "canned" commands above. We added adjustments to ensure this. The White standard errors in the canned command are 'HC1' and use an adjustment factor of $\frac{N}{N-K}$. The clustered standard errors use an adjustment factor (also in the section notes) of $\frac{C}{C-1}\frac{N-1}{N-K}$.

(c) Compute the between estimator, both with and without covariates. Under what conditions will this give an unbiased estimate of the effect of primary seat belt laws on fatalities per capita? Do you believe those conditions are met? Are you concerned about the standard errors in this case?

```
# c - between estimator with and without covariates
traffic_bet <- traffic[, lapply(.SD, mean), by = "state"] # get means by state

between <- feols(ln_fat_pc ~ primary, data=traffic_bet)
summary(between, se="standard")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 49
## Standard-errors: Standard
##              Estimate Std. Error    t value  Pr(>|t|)
## (Intercept) -1.716000   0.051659 -33.218000 < 2.2e-16 ***
## primary     -0.071216   0.154766  -0.460156  0.647526
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -10.52   Adj. R2: -0.0167
```

```
between_cov <- feols(ln_fat_pc ~ primary + secondary + college +
                       beer + ln_unemploy + ln_totalvmt + ln_precip +
                       ln_snow + rural_speed + urban_speed, data=traffic_bet)
summary(between_cov, se = "standard")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 49
## Standard-errors: Standard
##              Estimate Std. Error   t value Pr(>|t|)
## (Intercept) -4.809700   1.116600 -4.307500 0.000112 ***
## primary      0.321236   0.174393  1.842000 0.073285 .
## secondary    0.233456   0.164156  1.422200 0.163135
## college     -2.097700   0.718585 -2.919200 0.005869 **
## beer         0.112346   0.111615  1.006500 0.320518
## ln_unemploy  0.116680   0.134577  0.867012 0.391378
## ln_totalvmt -0.122812   0.032914 -3.731300 0.000621 ***
## ln_precip    0.058401   0.077797  0.750683 0.457467
## ln_snow     -0.069262   0.016324 -4.242900 0.000136 ***
## rural_speed  0.067399   0.017554  3.839500 0.000453 ***
## urban_speed -0.002351   0.012843 -0.183063 0.855722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 30.89   Adj. R2: 0.76812
```

Demeaning our data by state allows us to estimate the between estimator. Without covariates, our between estimator implies that having primary belt laws is associated with a 7.12% decrease in traffic fatalities per capita, ceteris paribus. With covariates, our between estimator implies that having primary belt laws is associated with a 32.12% increase in traffic fatalities per capita, ceteris paribus.

Under the strict exogeneity assumption, our between estimator gives us an unbiased estimate of the effect of primary seat belt laws on fatalities per capita, that is under the strict exogeneity assumption, $E[\ddot{\epsilon}_{it}|\ddot{X}_{it}] = 0$ where $\ddot{X}_{it} = X_{it} - \bar{X}_i$, and $\ddot{\epsilon}_{it} = \epsilon_{it} - \bar{\epsilon}_i$. The strict exogeneity assumption implies that the error term is uncorrelated with all past, present, and future values of the control variables, which is quite a strong assumption. In our model, this assumption is unlikely to hold if, for example, past snowfall affects current traffic fatalities per capita because in areas with lower past snowfall drivers are less experienced with driving safely during adverse weather conditions and are more likely to get into serious traffic accidents as a result.

Note that our OLS standard errors from the between estimator are incorrect because $\ddot{\epsilon}_{it} = \epsilon_{it} - \bar{\epsilon}_i$ will be correlated across different observations within the same unit. However, we can obtain the correct standard errors by multiplying our OLS standard errors by $\sqrt{T/(T-1)}$.

(d) Compute the RE 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?

```r
# d - random effects estimator
random <- plm(ln_fat_pc ~ primary + secondary + college +
                beer + ln_unemploy + ln_totalvmt + ln_precip +
                ln_snow + rural_speed + urban_speed, data=traffic, model="random")
summary(random)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = ln_fat_pc ~ primary + secondary + college + beer +
##     ln_unemploy + ln_totalvmt + ln_precip + ln_snow + rural_speed +
##     urban_speed, data = traffic, model = "random")
##
## Balanced Panel: n = 49, T = 23, N = 1127
##
## Effects:
##                    var  std.dev share
## idiosyncratic 0.008127 0.090151 0.279
## individual    0.021041 0.145054 0.721
## theta: 0.8715
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.3721113 -0.0612727  0.0058319  0.0665328  0.3316814
##
## Coefficients:
##                Estimate  Std. Error  z-value  Pr(>|z|)
## (Intercept) -1.11123728  0.20996847  -5.2924 1.207e-07 ***
## primary     -0.13636132  0.01488872  -9.1587 < 2.2e-16 ***
## secondary   -0.05599048  0.01043423  -5.3660 8.048e-08 ***
## college     -1.49468923  0.17199787  -8.6902 < 2.2e-16 ***
## beer         0.76046185  0.03768054  20.1818 < 2.2e-16 ***
## ln_unemploy -0.16002704  0.01453137 -11.0125 < 2.2e-16 ***
## ln_totalvmt -0.06931491  0.02031308  -3.4123 0.0006441 ***
## ln_precip   -0.06959186  0.01745034  -3.9880 6.663e-05 ***
## ln_snow     -0.00533264  0.00297623  -1.7917 0.0731739 .
## rural_speed -0.00528299  0.00096709  -5.4628 4.687e-08 ***
## urban_speed  0.00263347  0.00085253   3.0890 0.0020084 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    26.797
## Residual Sum of Squares: 10.631
## R-Squared:      0.60326
## Adj. R-Squared: 0.5997
## Chisq: 1696.92 on 10 DF, p-value: < 2.22e-16
```

Our random effects estimator implies that having primary belt laws is associated with a 13.64% decrease in traffic fatalities per capita, ceteris paribus. Under the strict exogeneity assumption and the uncorrelated effects assumptions, that is the state-specific effect is uncorrelated with the regressors, the RE estimator is consistent. However if the uncorrelated effects assumption is violated, then our RE estimator will be biased. The benefit of the RE estimator is that if we can model the heteroskedasticity correctly it will be more efficient than pooled OLS. However, in order for the RE error structure to be correct, we need to assume that all residuals within a cluster are equally correlated with each other, which may not hold in practice. If this assumption does not hold, our RE estimator will still be consistent but may no longer be more efficient than OLS.

(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?)

As discussed above, the standard errors from our RE estimator above are only correct if we assume that all residuals within a cluster are equally correlated with each other, that is if we assume that the correlation between different observations for the same unit is always the same, regardless of how far apart in time they are. If this assumption is violated then our RE standard errors will be wrong. Indeed, when we compute the clustered SEs below they are generally larger than our RE standard errors, implying that our assumption that all residuals within a cluster are equally correlated with each other likely does not hold.

```
# e - clustered SEs
coeftest(random, vcovHC(random, type="HC1", cluster="group"))
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) -1.1112373  0.3784728  -2.9361 0.0033917 **
## primary     -0.1363613  0.0281395  -4.8459 1.438e-06 ***
## secondary   -0.0559905  0.0179337  -3.1221 0.0018419 **
## college     -1.4946892  0.2963614  -5.0435 5.332e-07 ***
## beer         0.7604618  0.0647113  11.7516 < 2.2e-16 ***
## ln_unemploy -0.1600270  0.0139590 -11.4641 < 2.2e-16 ***
## ln_totalvmt -0.0693149  0.0339647  -2.0408 0.0415067 *
## ln_precip   -0.0695919  0.0230450  -3.0198 0.0025867 **
## ln_snow     -0.0053326  0.0038798  -1.3745 0.1695734
## rural_speed -0.0052830  0.0014466  -3.6520 0.0002723 ***
## urban_speed  0.0026335  0.0013847   1.9018 0.0574488 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(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?

```r
fixed <- plm(ln_fat_pc ~ primary, data = traffic, effect="twoways",
             index = c("state", "year"), model = "within")
summary(fixed)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = ln_fat_pc ~ primary, data = traffic, effect = "twoways",
##     model = "within", index = c("state", "year"))
##
## Balanced Panel: n = 49, T = 23, N = 1127
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.3426419 -0.0590258 -0.0010778  0.0682636  0.3400593
##
## Coefficients:
##          Estimate Std. Error t-value  Pr(>|t|)
## primary -0.092618   0.013771 -6.7258 2.857e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    12.282
## Residual Sum of Squares: 11.777
## R-Squared:      0.041115
## Adj. R-Squared: -0.023417
## F-statistic: 45.2362 on 1 and 1055 DF, p-value: 2.8572e-11
```

```r
coeftest(fixed, vcov = vcovHC(fixed, type = "HC1", cluster = "group"))
```

```
##
## t test of coefficients:
##
##          Estimate Std. Error t value Pr(>|t|)
## primary -0.092618   0.031647 -2.9266 0.003501 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Our FE estimator implies that having primary belt laws is associated with 9.26% decrease in traffic fatalities per capita, ceteris paribus. When we compute clustered standard errors they are almost double the size of our normal standard errors. As discussed in part b above, clustered SEs will be larger than unclustered SEs the more we have intra-cluster correlation of regressors, the more observations we have per cluster, and the more correlated the errors are within-cluster.

(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
fixed_cov <- plm(ln_fat_pc ~ primary + secondary + college +
                beer + ln_unemploy + ln_totalvmt + ln_precip +
                ln_snow + rural_speed + urban_speed, data = traffic,
                effect="twoways", index = c("state", "year"), model = "within")
summary(fixed_cov)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = ln_fat_pc ~ primary + secondary + college + beer +
##     ln_unemploy + ln_totalvmt + ln_precip + ln_snow + rural_speed +
##     urban_speed, data = traffic, effect = "twoways", model = "within",
##     index = c("state", "year"))
##
## Balanced Panel: n = 49, T = 23, N = 1127
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -0.29243199 -0.04959348  0.00065118  0.04672937  0.34027723
##
## Coefficients:
##                Estimate  Std. Error  t-value  Pr(>|t|)
## primary     -8.9751e-02  1.4036e-02  -6.3941  2.43e-10 ***
## secondary   -1.6321e-02  1.0092e-02  -1.6173  0.106122
## college     -2.5094e-01  2.2644e-01  -1.1082  0.268022
## beer         6.8154e-01  3.7717e-02  18.0699 < 2.2e-16 ***
## ln_unemploy -2.0015e-01  1.6769e-02 -11.9356 < 2.2e-16 ***
## ln_totalvmt  8.3462e-03  3.9555e-02   0.2110  0.832925
## ln_precip   -4.5401e-02  1.7461e-02  -2.6002  0.009449 **
## ln_snow      1.1063e-05  2.7815e-03   0.0040  0.996827
## rural_speed  2.0607e-03  1.2009e-03   1.7159  0.086470 .
## urban_speed  1.3897e-03  8.5212e-04   1.6309  0.103210
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    12.282
## Residual Sum of Squares: 7.0794
## R-Squared:      0.42358
## Adj. R-Squared: 0.3795
## F-statistic: 76.8662 on 10 and 1046 DF, p-value: < 2.22e-16
```

```r
coeftest(fixed_cov, vcov = vcovHC(fixed, type = "HC1", cluster = "group"))
```

```
##
## t test of coefficients:
##
##          Estimate Std. Error t value Pr(>|t|)
## primary -0.089751   0.031647  -2.836 0.004657 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When we add our full set of covariates to the FE estimator we find that having primary belt laws is associated with 8.98% decrease in traffic fatalities per capita, ceteris paribus, which is statistically significant at the 1% level. These results are more stable than our OLS estimates in part a above because the sign is now in the expected direction and the effect is measured much more precisely. Thus it seems likely that our earlier OLS model suffered from omitted variable bias, namely from omitting state-specific effects, which we control for in our FE model.

(h) Estimate a first-difference 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?

The coefficient on primary seat belt laws is increasing in magnitude with the time period of differencing, that is our first differences estimator has the smallest and our longest differences estimator has the largest magnitude. Our first differences and five year differences estimators are statistically significant, at the 10% and 1% level respectively, while our long differences estimator is not statistically significant. This makes sense since our long differences estimator only uses one observation per state. The FE estimator is most similar to the five year differences estimator, which is what we would expect since intuitively the FE estimator is an intermediary between the two extremes of the first differencing and long differencing estimators, combining the entire range of different differencing estimators together.

```r
## First Differences Estimator
# Create first-differenced variables
setkey(traffic, state)
# lists of differenced vars
cols <- c("ln_fat_pc", "primary", "secondary", "college", "beer",
          "ln_unemploy", "ln_totalvmt", "ln_precip", "ln_snow", "rural_speed", "urban_speed")
fd_cols <- c("ln_fat_pc_fd", "primary_fd", "secondary_fd", "college_fd",
             "beer_fd", "ln_unemploy_fd", "ln_totalvmt_fd", "ln_precip_fd",
             "ln_snow_fd", "rural_speed_fd", "urban_speed_fd")
fived_cols <- c("ln_fat_pc_fived", "primary_fived", "secondary_fived",
                "college_fived", "beer_fived", "ln_unemploy_fived", "ln_totalvmt_fived",
                "ln_precip_fived", "ln_snow_fived", "rural_speed_fived", "urban_speed_fived")
ld_cols <- c("ln_fat_pc_ld", "primary_ld", "secondary_ld",
             "college_ld", "beer_ld", "ln_unemploy_ld", "ln_totalvmt_ld",
             "ln_precip_ld", "ln_snow_ld", "rural_speed_ld", "urban_speed_ld")

traffic[,c(fd_cols) := lapply(.SD, function(x){c(NA, diff(x, lag = 1))}), .SDcols = c(cols), by = state]
traffic[,c(fived_cols) := lapply(.SD, function(x){c(NA, NA, NA, NA, NA, diff(x, lag = 5))}), .SDcols = c(cols
traffic[,c(ld_cols) := lapply(.SD, function(x){c(NA, NA, NA, NA, NA, NA, NA, NA,
                                                 NA, NA, NA, NA, NA, NA, NA, NA, NA,
                                                 NA, NA, NA, NA, NA, diff(x, lag = 22))}), .SDcols = c(cols),

# Estimate first-differenced estimator with year FEs and covariates
fd <- feols(ln_fat_pc_fd ~ primary_fd + secondary_fd + college_fd +
              beer_fd + ln_unemploy_fd + ln_totalvmt_fd + ln_precip_fd +
              ln_snow_fd + rural_speed_fd + urban_speed_fd,
            fixef = "year", data = traffic[year > 1981])
summary(fd, cluster=traffic[year > 1981]$state)
```

```
## OLS estimation, Dep. Var.: ln_fat_pc_fd
## Observations: 1,078
## Fixed-effects: year: 22
## Standard-errors: Clustered
##                 Estimate Std. Error t value    Pr(>|t|)
## primary_fd     -0.050136   0.019629 -2.5541 1.07860e-02 *
## secondary_fd   -0.038531   0.015473 -2.4902 1.29210e-02 *
## college_fd     -0.662941   0.448095 -1.4795 1.39317e-01
## beer_fd         0.272697   0.075363  3.6184 3.11000e-04 ***
## ln_unemploy_fd -0.106864   0.024481 -4.3652 1.40000e-05 ***
## ln_totalvmt_fd  0.247552   0.094553  2.6181 8.96900e-03 **
## ln_precip_fd   -0.082502   0.015343 -5.3773 9.32000e-08 ***
## ln_snow_fd      0.003329   0.002976  1.1188 2.63498e-01
## rural_speed_fd -0.004051   0.001416 -2.8612 4.30500e-03 **
## urban_speed_fd  0.002455   0.001133  2.1665 3.04980e-02 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 1,158.44   Adj. R2: 0.19898
##                          R2-Within: 0.09274
```

```r
# Estimate 5-year differences estimator with year FEs and covariates
fived <- feols(ln_fat_pc_fived ~ primary_fived + secondary_fived +
                 college_fived + beer_fived + ln_unemploy_fived + ln_totalvmt_fived +
                 ln_precip_fived + ln_snow_fived + rural_speed_fived + urban_speed_fived,
               fixef = "year", data = traffic[year > 1985])
summary(fived, cluster=traffic[year > 1985]$state)
```

```
## OLS estimation, Dep. Var.: ln_fat_pc_fived
## Observations: 882
## Fixed-effects: year: 18
## Standard-errors: Clustered
##                    Estimate Std. Error   t value    Pr(>|t|)
## primary_fived     -0.085650   0.020831 -4.111600 4.30000e-05 ***
## secondary_fived   -0.039367   0.014271 -2.758500 5.93100e-03 **
## college_fived     -0.392394   0.325505 -1.205500 2.28347e-01
## beer_fived         0.565468   0.082656  6.841200 1.50000e-11 ***
## ln_unemploy_fived -0.172223   0.024616 -6.996400 5.31000e-12 ***
## ln_totalvmt_fived  0.088942   0.085919  1.035200 3.00879e-01
## ln_precip_fived   -0.043433   0.025216 -1.722500 8.53450e-02 .
## ln_snow_fived     -0.002079   0.003417 -0.608374 5.43101e-01
## rural_speed_fived -0.000412   0.001367 -0.301517 7.63094e-01
## urban_speed_fived  0.002761   0.001135  2.433100 1.51740e-02 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 734.71   Adj. R2: 0.48457
##                        R2-Within: 0.33853
```

```r
# Estimate long differences estimator with covariates
ld <- feols(ln_fat_pc_ld ~ primary_ld + secondary_ld +
              college_ld + beer_ld + ln_unemploy_ld + ln_totalvmt_ld + ln_precip_ld +
              ln_snow_ld + rural_speed_ld + urban_speed_ld -1,
            data = traffic[year == 2003])
summary(ld, se = "white")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc_ld
## Observations: 49
## Standard-errors: White
##                 Estimate Std. Error   t value   Pr(>|t|)
## primary_ld     -0.156923   0.129024 -1.216200 0.23120600
## secondary_ld   -0.039544   0.147763 -0.267620 0.79040200
## college_ld     -1.433800   0.793693 -1.806400 0.07857000 .
## beer_ld         0.727984   0.138760  5.246400 0.00000574 ***
## ln_unemploy_ld -0.310630   0.085531 -3.631800 0.00080900 ***
## ln_totalvmt_ld -0.177645   0.147634 -1.203300 0.23612000
## ln_precip_ld   -0.002839   0.123936 -0.022906 0.98184200
## ln_snow_ld      0.038380   0.028621  1.341000 0.18769100
## rural_speed_ld -0.005879   0.008466 -0.694412 0.49154000
## urban_speed_ld  0.001885   0.004718  0.399522 0.69168800
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 27.64   Adj. R2: 0.51102
```

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

The first differences estimator uses more data than the 5-year or long differences estimator, since for any differencing estimator we can only use periods $s+1, ..., T$ for estimation, where s indicates the lag period for differencing. As such, the FD estimator is able to take advantage of high frequency variation in our covariates over time for estimation. This is appealing from an identification perspective since we can control for unobserved trends, making it less likely that unobserved trends are driving our results. The 5-year or long differences estimates, in contrast, could be biased by unobserved trends that differ across units.

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

The 5-year differences and long differences estimates are less sensitive to measurement error than the first differencing estimator. To see why the first differencing estimator is more sensitive to measurement error, note that measurement error (such as classical measurement error) in the covariates typically has less serial correlation than the true data generating process. Thus, sudden changes in the regressors are more likely to be noise than true signal. Unfortunately, a first differences estimator will accentuate any problems with measurement error, as it uses the shortest-run (highest frequency) variation in the regressors for estimation. Relatedly, a first differences estimator is less able to pick up on longer term effects. In contrast, a five year or long differences estimator is less sensitive to measurement error, because it uses longer-run (lower frequency) variation in the regressors, and is better able to pick up on longer-term trends.