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

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

$$= \left(\frac{\Delta X_{it}' \Delta X_{it}}{4} + \frac{\Delta X_{it}' \Delta X_{it}}{4} \right)^{-1} \left(\frac{\Delta X_{it}' \Delta Y_{it}}{4} + \frac{\Delta X_{it}' \Delta Y_{it}}{4} \right)$$

$$= \left(\frac{\Delta X_{it}' \Delta X_{it}}{2} \right)^{-1} \left(\frac{\Delta X_{it}' \Delta Y_{it}}{2} \right) = (\tfrac{1}{2})^{-1} (\Delta X_{it}' \Delta X_{it})^{-1} (\tfrac{1}{2})(\Delta X_{it}' \Delta Y_{it})$$

$$= (\Delta X_{it}' \Delta X_{it})^{-1} (\Delta X_{it}' \Delta Y_{it}) = \hat{\beta}_{FD}$$

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 $\hat{\beta}_{FE} = \hat{\beta}_{FD}$

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

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

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 $\hat{\beta}_{FE} = \hat{\beta}_{FD}$

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

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

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

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

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

$$= \frac{1}{2} (\Delta \hat{\epsilon}_{it})^2 \left(\ddot{X}_{i1}' \ddot{X}_{i1} + \ddot{X}_{i2}' \ddot{X}_{i2} \right)^{-1}$$

$$= \frac{1}{2} (\Delta \hat{\epsilon}_{it})^2 \left(\frac{-\Delta X_{it}'}{2} \frac{-\Delta X_{it}}{2} + \frac{\Delta X_{it}'}{2} \frac{\Delta X_{it}}{2} \right)^{-1}$$

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

$$= \frac{1}{2} (\Delta \hat{\epsilon}_{it})^2 (\tfrac{1}{2})^{-1} (\Delta X_{it}' \Delta X_{it})^{-1} = (\Delta \hat{\epsilon}_{it})^2 (\Delta X_{it}' \Delta X_{it})^{-1} = Var(\hat{\beta}_{FD})$$

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

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?

```
# 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.

```
# 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.

```
# 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.

```
# 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
```

```
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
```

```
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
```

```
# 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
```

```
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)
}


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

  var.robust <- 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)
```

```r
    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)
# get standard errors
white_biv <- robust.se(xmat_biv, w_mid_biv)
white_biv
```

```
##         V1    primary
## 0.01081489 0.02633257
```

```r
white_biv_yfe <- robust.se(xmat_biv_yfe, w_mid_biv_yfe)
white_biv_yfe[1:3]
```

```
##         V1    primary  year_1982
## 0.04384825 0.02748089 0.06084987
```

```r
white_biv_yfe_cov <- robust.se(xmat_biv_yfe_cov, w_mid_biv_yfe_cov)
white_biv_yfe_cov[1:11]
```

```
##          V1     primary   secondary     college        beer ln_unemploy
## 0.167454386 0.022443850 0.019771083 0.168086198 0.026255527 0.026718615
## ln_totalvmt   ln_precip     ln_snow rural_speed urban_speed
## 0.009568140 0.016899622 0.003248055 0.002362976 0.001757033
```

```r
# Clustered by state
states <- as.vector(unique(traffic[,state]))

cl_mid_biv_terms <- mclapply(states, cluster_middle, beta = beta_biv, DT = traffic,
                             yvar="ln_fat_pc", xvars=biv_var, mc.cores = core.num)
cl_mid_biv <- Reduce('+', cl_mid_biv_terms)

cl_mid_biv_yfe_terms <- mclapply(states, cluster_middle, beta = beta_biv_yfe, DT = traffic,
                                 yvar="ln_fat_pc", xvars=biv_yfe_var, mc.cores = core.num)
cl_mid_biv_yfe <- Reduce('+', cl_mid_biv_yfe_terms)

cl_mid_biv_yfe_cov_terms <- mclapply(states, cluster_middle, beta = beta_biv_yfe_cov, DT = traffic,
```

```
                                    yvar="ln_fat_pc", xvars=biv_yfe_cov_var, mc.cores = core.num)
cl_mid_biv_yfe_cov <- Reduce('+', cl_mid_biv_yfe_cov_terms)

cl_biv <- robust.se(xmat_biv, cl_mid_biv)
cl_biv
```

```
##         V1    primary
## 0.04531152 0.09287883
```

```
cl_biv_yfe <- robust.se(xmat_biv_yfe, cl_mid_biv_yfe)
cl_biv_yfe[1:3]
```

```
##         V1    primary  year_1982
## 0.04384825 0.10403901 0.01579344
```

```
cl_biv_yfe_cov <- robust.se(xmat_biv_yfe_cov, cl_mid_biv_yfe_cov)
cl_biv_yfe_cov[1:11]
```

```
##          V1      primary    secondary      college         beer ln_unemploy
## 0.502704959 0.044755718 0.038382908 0.492029109 0.077636743 0.066674947
## ln_totalvmt    ln_precip      ln_snow rural_speed urban_speed
## 0.036514564 0.057562915 0.008064637 0.004362628 0.003323011
```

When we calculate the White-robust and clustered standard errors by hand they are quite similar to the standard errors calculated by the "canned" commands above.

(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 can 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="sss", cluster="group"))
```

```
##
## t test of coefficients:
##
##                Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) -1.1112373  0.3822252  -2.9073  0.0037179 **
## primary     -0.1363613  0.0284185  -4.7983  1.817e-06 ***
## secondary   -0.0559905  0.0181115  -3.0914  0.0020413 **
## college     -1.4946892  0.2992997  -4.9940  6.860e-07 ***
## beer         0.7604618  0.0653529  11.6362  < 2.2e-16 ***
## ln_unemploy -0.1600270  0.0140974 -11.3516  < 2.2e-16 ***
## ln_totalvmt -0.0693149  0.0343015  -2.0208  0.0435436 *
## ln_precip   -0.0695919  0.0232735  -2.9902  0.0028493 **
## ln_snow     -0.0053326  0.0039183  -1.3610  0.1737976
## rural_speed -0.0052830  0.0014609  -3.6162  0.0003124 ***
```

```
## urban_speed  0.0026335  0.0013984    1.8832 0.0599368 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```