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.

(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 freedome corrections. You are not clustering the standard errors in this problem.

Question 2

Question 3

A -

```
# a- Pooled bivariate OLS , yr FE, All covariates

# 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"))

# bivariate OLS
biv <- feols(ln_fat_pc ~ primary + secondary, 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) -1.625000   0.016033 -101.3500 < 2.2e-16 ***
```

```
## primary     -0.222018   0.028046   -7.9162  5.84e-15 ***
## secondary   -0.140641   0.021510   -6.5383  9.42e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -332.47   Adj. R2: 0.06081
```

```
biv_yfe <- feols(ln_fat_pc ~ primary + secondary, 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   -0.086378   0.037159 -2.324500 0.020278 *
## secondary -0.008271   0.032443 -0.254946 0.798812
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -304.06   Adj. R2: 0.08915
##                         R2-Within: 0.00834
```

```
biv_yfe_cov <- feols(ln_fat_pc ~ primary + secondary + college +
                       beer + ln_unemploy + ln_totalvmt + ln_precip +
                       ln_snow + rural_speed + urban_speed, fixef = c("year", "state"), data=traffic)
summary(biv_yfe_cov, se = "standard")
```

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Fixed-effects: year: 23,  state: 49
## Standard-errors: Standard
##              Estimate Std. Error    t value  Pr(>|t|)
## primary     -0.089751   0.014036  -6.394100  2.43e-10 ***
## secondary   -0.016321   0.010092  -1.617300  0.106122
## college     -0.250942   0.226437  -1.108200  0.268022
## beer         0.681538   0.037717  18.070000 < 2.2e-16 ***
## ln_unemploy -0.200147   0.016769 -11.936000 < 2.2e-16 ***
## ln_totalvmt  0.008346   0.039555   0.211004  0.832925
## ln_precip   -0.045401   0.017461  -2.600200  0.009449 **
## ln_snow      0.000011   0.002782   0.003977  0.996827
## rural_speed  0.002061   0.001201   1.715900   0.08647 .
## urban_speed  0.001390   0.000852   1.630900   0.10321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 1,257.87   Adj. R2: 0.93998
##                          R2-Within: 0.42358
```

B -

```
# b - white robust and clustered

# package command - heteroskedastic
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) -1.625000   0.015258 -106.5000 < 2.2e-16 ***
## primary     -0.222018   0.028474   -7.7972  1.44e-14 ***
## secondary   -0.140641   0.021134   -6.6546  4.43e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -332.47   Adj. R2: 0.06081
```

```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   -0.086378   0.038634 -2.235800 0.025563 *
## secondary -0.008271   0.031955 -0.258838 0.795808
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -304.06   Adj. R2: 0.08915
##                         R2-Within: 0.00834
```

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

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Fixed-effects: year: 23,  state: 49
## Standard-errors: White
##              Estimate Std. Error    t value  Pr(>|t|)
## primary     -0.089751   0.014062  -6.382700  2.61e-10 ***
## secondary   -0.016321   0.010466  -1.559500  0.119184
## college     -0.250942   0.256157  -0.979641   0.32749
## beer         0.681538   0.038584  17.664000 < 2.2e-16 ***
## ln_unemploy -0.200147   0.017147 -11.672000 < 2.2e-16 ***
## ln_totalvmt  0.008346   0.048469   0.172197  0.863316
## ln_precip   -0.045401   0.017550  -2.587000  0.009817 **
## ln_snow      0.000011   0.002700   0.004097  0.996732
## rural_speed  0.002061   0.001255   1.641400  0.101023
## urban_speed  0.001390   0.000852   1.631000  0.103186
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 1,257.87   Adj. R2: 0.93998
##                          R2-Within: 0.42358
```

```r
# package command - cluster
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) -1.625000   0.046411 -35.0140 < 2.2e-16 ***
## primary     -0.222018   0.090393  -2.4561  0.014195 *
## secondary   -0.140641   0.035964  -3.9107   9.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -332.47   Adj. R2: 0.06081
```

```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   -0.086378   0.134724 -0.641150 0.521559
## secondary -0.008271   0.079979 -0.103418 0.917650
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: -304.06   Adj. R2: 0.08915
##                         R2-Within: 0.00834
```

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

```
## OLS estimation, Dep. Var.: ln_fat_pc
## Observations: 1,127
## Fixed-effects: year: 23,  state: 49
## Standard-errors: Clustered
##              Estimate Std. Error   t value  Pr(>|t|)
## primary     -0.089751   0.027702 -3.239800  0.001234 **
## secondary   -0.016321   0.019374 -0.842441  0.399733
## college     -0.250942   0.487674 -0.514570  0.606962
## beer         0.681538   0.072175  9.442900 < 2.2e-16 ***
## ln_unemploy -0.200147   0.022316 -8.968900 < 2.2e-16 ***
## ln_totalvmt  0.008346   0.093534  0.089232  0.928914
## ln_precip   -0.045401   0.020494 -2.215400  0.026951 *
## ln_snow      0.000011   0.003193  0.003465  0.997236
## rural_speed  0.002061   0.002015  1.022400  0.306813
## urban_speed  0.001390   0.001552  0.895658  0.370641
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Log-likelihood: 1,257.87   Adj. R2: 0.93998
##                          R2-Within: 0.42358
```

```r
# write own commands

# will need to get beta matrix manually
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)
}
```

```r
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", "secondary")
biv_yfe_var <- c("primary", "secondary", 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]),
                     colnames(traffic[,state_2:state_99]))

# 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)

# White robust
# 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_yfe <- robust.se(xmat_biv_yfe, w_mid_biv_yfe)
white_biv_yfe_cov <- robust.se(xmat_biv_yfe_cov, w_mid_biv_yfe_cov)

# 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_yfe <- robust.se(xmat_biv_yfe, cl_mid_biv_yfe)
cl_biv_yfe_cov <- robust.se(xmat_biv_yfe_cov, cl_mid_biv_yfe_cov)
```