UNIVERSITY OF NEVADA LAS VEGAS

LEE BUSINESS SCHOOL

ECO 772- ECONOMETRICS II

ASSIGNMENT #06

STUDENT NAME: DANIEL EMAASIT

MAJOR: CIVIL ENGINEERING

INSTRUCTOR NAME: Prof. Ian McDonough

DATE SUBMITTED: 05/14/2015

SPRING 2015

Problem 13.5

Suppose that we want to estimate the effect of several variables on annual saving and that we have a panel data set on individuals collected on January 31, 1990, and January 31, 1992. If we include a year dummy for 1992 and use first differencing, can we also include age in the original model? Explain.

No, we cannot include age as an explanatory variable in the original model. Each person in the panel data set is exactly two years older on January 31, 1992. Therefore when we use first differencing, the explanatory variable is constant which violates the multicollinearlity assumption required for estimating the coefficients using Pooled OLS.

Problem C13.7

Use GPA3.RAW for this exercise. The data set is for 366 student-athletes from a large university for fall and spring semesters. [A similar analysis is in Maloney and McCormick (1993), but here we use a true panel data set.] Because you have two terms of data for each student, an unobserved effects model is appropriate. The primary question of interest is this: Do athletes perform more poorly in school during the semester their sport is in season?

Part (i)

Use pooled OLS to estimate a model with term GPA (trmgpa) as the dependent variable. The explanatory variables are spring, sat, hsperc, female, black, white, frstsem, tothrs, crsgpa, and season. Interpret the coefficient on season. Is it statistically significant?

```
# Start by reading the data into R
library(haven)
gpa_data <- read_dta("GPA3.dta")

# Create a panel dataset
library(plm)

## Loading required package: Formula

pdata <- plm.data(gpa_data, index = c("term", "id"))

# Use pooled OLS to estimate the model
gpa_pool <- plm(trmgpa ~ spring + sat + hsperc + female + black + white +
frstsem + tothrs + crsgpa + season, data = pdata, model = "pooling")

# Display the results
summary(gpa_pool)

## Oneway (individual) effect Pooling Model
##
## Call:
## plm(formula = trmgpa ~ spring + sat + hsperc + female + black +
```

```
##     white + frstsem + tothrs + crsgpa + season, data = pdata,
##     model = "pooling")
##
## Balanced Panel: n=2, T=366, N=732
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -1.8500 -0.3310  0.0192  0.3800  1.5800
##
## Coefficients :
##                Estimate  Std. Error t-value  Pr(>|t|)
## (Intercept) -1.75284742  0.34790489 -5.0383 5.943e-07 ***
## spring      -0.05800662  0.04803681 -1.2075   0.22762
## sat          0.00169844  0.00014942 11.3671 < 2.2e-16 ***
## hsperc      -0.00866104  0.00103628 -8.3578 3.280e-16 ***
## female       0.35040133  0.05185242  6.7577 2.894e-11 ***
## black       -0.25414949  0.12292159 -2.0676   0.03904 *
## white       -0.02331462  0.11739542 -0.1986   0.84263
## frstsem     -0.03465848  0.07603448 -0.4558   0.64865
## tothrs      -0.00033894  0.00072672 -0.4664   0.64108
## crsgpa       1.04786548  0.10411440 10.0646 < 2.2e-16 ***
## season      -0.02729036  0.04904604 -0.5564   0.57809
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    420.3
## Residual Sum of Squares: 219.58
## R-Squared      :  0.47756
##       Adj. R-Squared :  0.47039
## F-statistic: 65.907 on 10 and 721 DF, p-value: < 2.22e-16
```

<u>The coefficient on season implies that, other things fixed, an athlete's term GPA is about 0.027 points lower when his/her sport is in season.</u>

Part (ii)

Most of the athletes who play their sport only in the fall are football players. Suppose the ability levels of football players differ systematically from those of other athletes. If ability is not adequately captured by SAT score and high school percentile, explain why the pooled OLS estimators will be biased.

<u>If omitted ability is correlated with season then OLS is biased and inconsistent.</u>

Part (iii)

Now, use the data differenced across the two terms. Which variables drop out? Now, test for an in-season effect.

```r
# Use first differencing
gpa_fd <- plm(trmgpa ~ spring + sat + hsperc + female + black + white +
frstsem + tothrs + crsgpa + season, data = pdata, model = "fd", effect =
```

```r
"individual")

# Display the results
summary(gpa_fd)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = trmgpa ~ spring + sat + hsperc + female + black +
##     white + frstsem + tothrs + crsgpa + season, data = pdata,
##     effect = "individual", model = "fd")
##
## Balanced Panel: n=2, T=366, N=732
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -2.5700 -0.4950  0.0118  0.4880  2.3800
##
## Coefficients :
##                Estimate  Std. Error t-value  Pr(>|t|)
## (intercept)  0.00283380  0.02762909  0.1026   0.91834
## sat          0.00172618  0.00014631 11.7983 < 2.2e-16 ***
## hsperc      -0.00761676  0.00107703 -7.0720 3.621e-12 ***
## female       0.35593860  0.05111870  6.9630 7.516e-12 ***
## black       -0.27967552  0.12347803 -2.2650   0.02381 *
## white       -0.10278133  0.12177902 -0.8440   0.39895
## frstsem     -0.07744396  0.07195517 -1.0763   0.28216
## tothrs      -0.00101322  0.00068801 -1.4727   0.14127
## crsgpa       1.04293533  0.10144695 10.2806 < 2.2e-16 ***
## season      -0.02518295  0.04800135 -0.5246   0.60000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    730.17
## Residual Sum of Squares: 401.21
## R-Squared      :  0.45052
##       Adj. R-Squared :  0.44435
## F-statistic: 65.5929 on 9 and 720 DF, p-value: < 2.22e-16
```

<u>The variables spring drop outs.</u>

Part (iv)

Can you think of one or more potentially important, time-varying variables that have been omitted from the analysis?

<u>One possibility is a measure of course load. If some fraction of student-athletes take a lighter load during the season, then term GPAs may tend to be higher, other things equal.</u>

Problem 13.10

For this exercise, we use JTRAIN.RAW to determine the effect of the job training grant on hours of job training per employee.

Part (i)

Estimate the equation using first differencing. How many firms are used in the estimation? How many total observations would be used if each firm had data on all variables (in particular, hrsemp) for all three time periods?

```
# Start by reading the data in R
train_data <- read_dta("JTRAIN.dta")

# Create a panel dataset
jdata <- plm.data(train_data, indexes = c("year", "fcode"))

# Estimate the equation using first differencing
j_fd <- plm(hrsemp ~ d88 + d89 + grant + grant_1 + lemploy, data = jdata,
model = "fd")

# Display the results
summary(j_fd)
```

```
## Oneway (individual) effect First-Difference Model
##
## Call:
## plm(formula = hrsemp ~ d88 + d89 + grant + grant_1 + lemploy,
##     data = jdata, model = "fd")
##
## Unbalanced Panel: n=3, T=127-134, N=390
##
## Residuals :
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -157.000  -11.800   -0.278   11.300  152.000
##
## Coefficients :
##                Estimate  Std. Error t-value  Pr(>|t|)
## (intercept)  -0.0061314   1.6344981 -0.0038 0.9970089
## grant        16.6353948   5.6176573  2.9613 0.0032546 **
## grant_1     -13.3854662   8.6738848 -1.5432 0.1236101
## lemploy      -4.2604779   1.0978787 -3.8806 0.0001226 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    423940
## Residual Sum of Squares: 395950
## R-Squared      :  0.066029
##       Adj. R-Squared :  0.065346
## F-statistic: 9.02563 on 3 and 383 DF, p-value: 8.6751e-06
```

There are 124 firms with both years of data and three firms with only one year of data used, for a total of 127 firms; 30 firms in the sample have missing information in both years and are not used at all. If we had information for all 157 firms, we would have 314 total observations in estimating the equation.

Part (ii)

Interpret the coefficient on grant and comment on its significance.

The coefficient means that if a firm received a grant for the current year, it trained each worker an average of 16.6 hours more than it would have otherwise.

Part (iii)

Is it surprising that grant21 is insignificant? Explain.

Since a grant last year was used to pay for training last year, it is perhaps not surprising that the grant does not carry over into more training this year.

Part (iv)

Do larger firms train their employees more or less, on average? How big are the differences in training?

The coefficient on the employees variable is not small: a 1% increase in employ decreases hours per employee by only 4.26 hrs

Problem C14.2

Use CRIME4.RAW for this exercise.

Part (i)

Reestimate the unobserved effects model for crime in Example 13.9 but use fixed effects rather than differencing. Are there any notable sign or magnitude changes in the coefficients? What about statistical significance?

```r
# Start by reading the data into R
crime_data <- read_dta("CRIME4.dta")

# Create a panel data frame
c_data <- plm.data(crime_data, indexes = c("year", "county"))

# Estimate the unobserved effects model
c_fe <- plm(lcrmrte ~ d83 + d84 + d85 + d86 + d87 + lprbarr + lprbconv +
lprbpris + lavgsen + lpolpc, data = c_data, model = "within")

# Display the results
summary(c_fe)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = lcrmrte ~ d83 + d84 + d85 + d86 + d87 + lprbarr +
##     lprbconv + lprbpris + lavgsen + lpolpc, data = c_data, model =
"within")
##
## Balanced Panel: n=7, T=90, N=630
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
##  -1.900  -0.187   0.029   0.232   1.310
##
## Coefficients :
##           Estimate Std. Error  t-value  Pr(>|t|)
## lprbarr  -0.719503   0.036766 -19.5699 < 2.2e-16 ***
## lprbconv -0.545659   0.026368 -20.6937 < 2.2e-16 ***
## lprbpris  0.247552   0.067227   3.6823 0.0002513 ***
## lavgsen  -0.086757   0.057920  -1.4979 0.1346769
## lpolpc    0.365989   0.030025  12.1894 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    204.51
## Residual Sum of Squares: 88.736
## R-Squared      :  0.56611
##       Adj. R-Squared :  0.55533
## F-statistic: 161.266 on 5 and 618 DF, p-value: < 2.22e-16
```

There is no intercept because it gets swept away in the time demeaning. The first-difference and fixed effects slope estimates are broadly consistent. The variables that are significant with first differencing are significant in the FE estimation, and the signs are all the same.

Part (ii)

Add the logs of each wage variable in the data set and estimate the model by fixed effects. How does including these variables affect the coefficients on the criminal justice variables in part (i)?

When the nine log wage variables are added and the equation is estimated by fixed effects, very little of importance changes on the criminal justice variables.

Part (iii)

Do the wage variables in part (ii) all have the expected sign? Explain. Are they jointly significant?

Yes, the wage variables are jointly significant.

Problem C14.4

In Example 13.8, we used the unemployment claims data from Papke (1994) to estimate the effect of enterprise zones on unemployment claims. Papke also uses a model that allows each city to have its own time trend:

Part (i)

Show that, when the previous equation is first differenced, we obtain;

Write the equation for times t and t – 1 as;

log(uclmsit) = ai + cit + β1ezit + uit,

log(uclmsi,t-1) = ai + ci(t – 1) + β1ezi,t-1 + ui,t-1

Part (ii)

Estimate the differenced equation by fixed effects. What is the estimate of ez? Is it very different from the estimate obtained in Example 13.8? Is the effect of enterprise zones still statistically significant?

```
# Start by reading the data into R
employ_data <- read_dta("EZUNEM.dta")

# Create a panel data frame
e_data <- plm.data(employ_data, indexes = c("year", "city"))

# Estimate the differenced equation
e_fe <- plm(luclms ~ d82+ d83 + d84 + d85 + d86 + d87+ d88 + ez, data =
e_data, model = "within")

# Display the results
summary(e_fe)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = luclms ~ d82 + d83 + d84 + d85 + d86 + d87 + d88 +
##     ez, data = e_data, model = "within")
##
## Balanced Panel: n=9, T=22, N=198
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -1.0900 -0.3970 -0.0613  0.3400  1.8100
##
## Coefficients :
##     Estimate Std. Error t-value Pr(>|t|)
## ez -0.038708   0.114850  -0.337   0.7365
##
```

```
## Total Sum of Squares:    64.965
## Residual Sum of Squares: 64.926
## R-Squared      :  0.00060385
##       Adj. R-Squared :  0.00057335
## F-statistic: 0.113592 on 1 and 188 DF, p-value: 0.73647
```

The estimated effect of an EZ is small and not statistically significant.

Part (iii)

Add a full set of year dummies to the estimation in part (ii). What happens to the estimate of ez1?

```
# Estimate the differenced equation using a full set of year dummies
e_fe2 <- plm(luclms ~ d81:d88 + ez, data = e_data, model = "within")

# Display the results
summary(e_fe2)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = luclms ~ d81:d88 + ez, data = e_data, model = "within")
##
## Balanced Panel: n=9, T=22, N=198
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -1.0900 -0.3970 -0.0613  0.3400  1.8100
##
## Coefficients :
##     Estimate Std. Error t-value Pr(>|t|)
## ez -0.038708   0.114850  -0.337   0.7365
##
## Total Sum of Squares:    64.965
## Residual Sum of Squares: 64.926
## R-Squared      :  0.00060385
##       Adj. R-Squared :  0.00057335
## F-statistic: 0.113592 on 1 and 188 DF, p-value: 0.73647
```

There is no difference.

Problem C14.6

Add the interaction term unionit⋅t to the equation estimated in Table 14.2 to see if wage growth depends on union status. Estimate the equation by random and fixed effects and compare the results.

```
# Start by reading the data into R
wagePan_data <- read_dta("WAGEPAN.dta")
```

```r
# Create a panel data frame
w_data <- plm.data(wagePan_data, indexes = c("year", "nr"))

# Estimate the equation by random effects
wageRE <- plm(lwage ~ educ + black + hisp + exper + expersq + married + union
+ union, data = w_data, method = "random", random.method = "walhus")

# Estimate the equation by Fixed effects
wageFE <- plm(lwage ~ educ + black + hisp + exper + expersq + married + union
+ union, data = w_data, method = "within")

# Display the resutls
summary(wageRE)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = lwage ~ educ + black + hisp + exper + expersq +
##     married + union + union, data = w_data, random.method = "walhus",
##     method = "random")
##
## Balanced Panel: n=8, T=545, N=4360
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -5.2700 -0.2480  0.0319  0.2950  2.5300
##
## Coefficients :
##             Estimate  Std. Error t-value  Pr(>|t|)
## educ     0.09134979  0.00523738 17.4419 < 2.2e-16 ***
## black   -0.13923422  0.02357956 -5.9049 3.799e-09 ***
## hisp     0.01601951  0.02079714  0.7703  0.441179
## exper    0.06723450  0.01369484  4.9095 9.467e-07 ***
## expersq -0.00241170  0.00081995 -2.9413  0.003286 **
## married  0.10825296  0.01568942  6.8997 5.962e-12 ***
## union    0.18246126  0.01715677 10.6349 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1143.6
## Residual Sum of Squares: 1002.5
## R-Squared      :  0.12337
##       Adj. R-Squared :  0.12295
## F-statistic: 87.3545 on 7 and 4345 DF, p-value: < 2.22e-16

summary(wageFE)

## Oneway (individual) effect Within Model
##
```

```
## Call:
## plm(formula = lwage ~ educ + black + hisp + exper + expersq +
##     married + union + union, data = w_data, method = "within")
##
## Balanced Panel: n=8, T=545, N=4360
##
## Residuals :
##    Min. 1st Qu.  Median 3rd Qu.    Max.
## -5.2700 -0.2480  0.0319  0.2950  2.5300
##
## Coefficients :
##            Estimate  Std. Error t-value  Pr(>|t|)
## educ     0.09134979  0.00523738 17.4419 < 2.2e-16 ***
## black   -0.13923422  0.02357956 -5.9049 3.799e-09 ***
## hisp     0.01601951  0.02079714  0.7703  0.441179
## exper    0.06723450  0.01369484  4.9095 9.467e-07 ***
## expersq -0.00241170  0.00081995 -2.9413  0.003286 **
## married  0.10825296  0.01568942  6.8997 5.962e-12 ***
## union    0.18246126  0.01715677 10.6349 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1143.6
## Residual Sum of Squares: 1002.5
## R-Squared      :  0.12337
##       Adj. R-Squared :  0.12295
## F-statistic: 87.3545 on 7 and 4345 DF, p-value: < 2.22e-16
```

They appear to be the same.