Homework 4

Wei Ye*
ECON 7910- Econometrics I

Due on Oct 14, 2021

1 Question 1 – 4.11

Solution:

a) With KWW and IQ as proxy of ability, $\beta_7 = 0.049837$. However, with only IQ as proxy of ability, $\beta_7 = 0.0544106$, which is a little increase. For specific information, see Table (1)

b) Since p-value is 0.0003181 only, thus, we can't reject null hypothesis.

c) No, it will not disappear. $AME = -0.1304$, and corresponding p value is 0.0011.

d) From the table 2, the interaction term $educ(iq - 100)$, aka, $educ : iq\_diff$ whose p value is high, which means not significant. However, $educ : kww\_educ$ is significant. The conclusion the interaction of educ and kww difference can somehow positively contribute to the log wage.

R codes as below:

```
          #This file is for homework 4-Econometrics
############################################
##All the codes are written by Wei Ye #####
############################################
library(tidyverse)
library(stargazer)
library(car)# In order to use linearHypothesis function
library(margins)

#Import data from csv file
NLS80 <- read_csv('nls80.csv')
head(NLS80,n=5)
```

*1st year PhD student in Economics Department at Fordham University. Email: wye22@fordham.edu

```
#Question 4.11(a)
with_iq_kww<-lm(lwage~exper+tenure+married+south+urban+black+educ+kww+iq ,data=
summary(with_iq_kww)
with_iq_only<- lm(lwage~exper+tenure+married+south+urban+black+educ+iq ,data=N
summary(with_iq_only)
stargazer(with_iq_kww,with_iq_only ,title = 'Compare Different Proxy of Ability
#Question 4.11 (b)
linearHypothesis(with_iq_kww,c('iq=0','kww=0'),white.adjust = 'hc1')

#Question 4.11 (c)
summary(margins(with_iq_kww, variables = 'black'))

#Question 4.11 (d)
NLS80<- NLS80%>%
  mutate(mean_kww=mean(kww))
NLS80<-NLS80%>%
  mutate(iq_diff=iq-100,
         kww_diff=kww-mean_kww)
with_all_terms_required<- lm(lwage~exper+tenure+married+south+urban+black+educ
summary(with_all_terms_required)
stargazer(with_iq_kww,with_iq_only ,with_all_terms_required ,titile='Regression
## 4.11 DONE!
```

2 Question – 4.12

Solution: From the table 3, adding the variable union, while negelacting the variable lag scrap, union can positively heavily contribute to log scrap. However, if put lag term in the regression, the effect of untion is negect and unsigficant as well.

R codes as below:

```
    ##########4.12
train<- read_csv('jtrain1.csv')
head(train)
new_var_with_lag <- lm(lscrap~grant+lscrap_1+union,data=train)
summary(new_var_with_lag)
new_var_without_lag<- lm(lscrap~grant+union,data=train)
summary(new_var_without_lag)
stargazer(new_var_with_lag ,new_var_without_lag , title="4.12 Regression Results"
```

3 Question – 4.13

Solution

a) See the table 4

b) From the Regression table 4, the log crime rate in 1987 is heavily attributed to the rate in 86, which means it's positively autocorrelated. If we add previous year's crime rate into regression model, the effect of lprbpris and lavgsen are reverse, and lprbconv would be from significant to insignificant. Lprbarr is also dcreasing.

c) $Fstatistic = 31.478$

d) studentized Breusch-Pagan test, the $BP = 10.155$. Unsure about (d), not sure about the specific ideas of codes. Check later.

```
    ###4.13
crime1 <- read_csv('cornwell.csv')
head(crime1,n=5)
crime2 <- crime1%>%
  filter(year==87)
logcrm_87 <- lm(lcrmrte~lprbarr+lprbconv+lprbpris+lavgsen,data=crime2)
summary(logcrm_87)
crime3 <- crime1 %>%
  filter(year==86)
logcrm_joint <- lm(lcrmrte~lprbarr+lprbarr+lprbconv+lprbpris+lavgsen+
                        crime3$lcrmrte,
                      data=crime2)
summary(logcrm_joint)
stargazer(logcrm_87,logcrm_joint, title = "4.13 part (a) and (b) Regression Re
library(lmtest)
bptest(logcrm_joint)
```

4 Question – 4.14

Solution:

a) From Regression result the coefficient of attend is 0.008163, which means increasing attending of classes will increase final grades. P value is 0.000228, obviously significant.

b) Not exactly! Because we can't control fresh year or sophomore year of students. Students with good standing may attend more classes, so there is homogeneity in the model, which means we can't make a conclustion of causality for the model directly.

c) The coefficient of attend will decrease to 0.005225, but still positive and significant effect even with lower t value. As expected, prior GPA and ACT score have significant and positive effect to final grades.

d) Significant level of frosh becomes from sigficance to insignificance. But soph will be significant under 90% confidence interval from insignificance.

e) From table (6), squares terms are both very significant. The coefficient of atndrte is unchanged.

f) The coefficient of atndrte will decrease and become insignificant, and the square term of attend will also insignificant. These two change remind us that we shouldn't add non-linear term of attend rate into our regression model.

Codes associated with this question as below:

```
    ### 4.14
#part(a)
attend1 <- read_csv('attend.csv')
head(attend1,n=5)
fgrade_1 <- lm(stndfnl~atndrte+frosh+soph,data=attend1)
summary(fgrade_1)
#part(c)
fgrade_2 <- lm(stndfnl~atndrte+frosh+soph+ACT+priGPA,data=attend1)
summary(fgrade_2)
stargazer(fgrade_1,fgrade_2,title="4.14 Regression Results")
#part(e)
attend2 <- attend1%>%
  mutate(priGPAsq=priGPA^2,
         ACTsq=ACT^2)
fgrade_3 <- lm(stndfnl~atndrte+frosh+soph+priGPAsq+ACTsq,data=attend2)
summary(fgrade_3)
stargazer(fgrade_3,title="4.14 Part(e) Regression Results")
#part(f)
attend3 <- attend2%>%
  mutate(atndrtesq=atndrte^2)
fgrade_4 <- lm(stndfnl~atndrte+frosh+soph+priGPAsq+ACTsq+atndrtesq,data=attenc
summary(fgrade_4)
stargazer(fgrade_4,title = "4.14 part(f) Regression Results")
```

Appendix

Table 1: Compare Different Proxy of Ability in 4.11

	Dependent variable:	
	lwage	
	(1)	(2)
exper	0.013***	0.014***
	(0.003)	(0.003)
tenure	0.011***	0.011***
	(0.002)	(0.002)
married	0.192***	0.200***
	(0.039)	(0.039)
south	−0.082***	−0.080***
	(0.026)	(0.026)
urban	0.176***	0.182***
	(0.027)	(0.027)
black	−0.130***	−0.143***
	(0.040)	(0.039)
educ	0.050***	0.054***
	(0.007)	(0.007)
kww	0.004**	
	(0.002)	
iq	0.003***	0.004***
	(0.001)	(0.001)
Constant	5.176***	5.176***
	(0.128)	(0.128)
Observations	935	935
R^2	0.266	0.263
Adjusted R^2	0.259	0.256
Residual Std. Error	0.363 (df = 925)	0.363 (df = 926)
F Statistic	37.284*** (df = 9; 925)	41.265*** (df = 8; 926)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2

	Dependent variable:		
	lwage		
	(1)	(2)	(3)
exper	0.013***	0.014***	0.012***
	(0.003)	(0.003)	(0.003)
tenure	0.011***	0.011***	0.011***
	(0.002)	(0.002)	(0.002)
married	0.192***	0.200***	0.198***
	(0.039)	(0.039)	(0.039)
south	−0.082***	−0.080***	−0.081***
	(0.026)	(0.026)	(0.026)
urban	0.176***	0.182***	0.178***
	(0.027)	(0.027)	(0.027)
black	−0.130***	−0.143***	−0.138***
	(0.040)	(0.039)	(0.040)
educ	0.050***	0.054***	0.045***
	(0.007)	(0.007)	(0.008)
kww	0.004**		−0.025**
	(0.002)		(0.011)
iq	0.003***	0.004***	0.005
	(0.001)	(0.001)	(0.006)
educ:iq_diff			−0.0001
			(0.0004)
educ:kww_diff			0.002***
			(0.001)
Constant	5.176***	5.176***	6.080***
	(0.128)	(0.128)	(0.561)
Observations	935	935	935
R^2	0.266	0.263	0.273
Adjusted R^2	0.259	0.256	0.264
Residual Std. Error	0.363 (df = 925)	0.363 (df = 926)	0.361 (df = 923)
F Statistic	37.284*** (df = 9; 925)	41.265*** (df = 8; 926)	31.478*** (df = 11; 923)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: 4.12 Regression Results

	Dependent variable:	
	lscrap	
	(1)	(2)
grant	−0.119	0.107
	(0.121)	(0.302)
lscrap_1	0.878^{***}	
	(0.037)	
union	0.069	0.536^{**}
	(0.115)	(0.246)
Constant	-0.150^{**}	0.196
	(0.074)	(0.152)
Observations	108	162
R^2	0.854	0.030
Adjusted R^2	0.849	0.017
Residual Std. Error	0.553 (df = 104)	1.473 (df = 159)
F Statistic	202.265^{***} (df = 3; 104)	2.430^{*} (df = 2; 159)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 4: 4.13 part (a) and (b) Regression Results

	Dependent variable:	
	lcrmrte	
	(1)	(2)
lprbarr	−0.724***	−0.185***
	(0.115)	(0.063)
lprbconv	−0.473***	−0.039
	(0.083)	(0.047)
lprbpris	0.160	−0.127
	(0.206)	(0.099)
lavgsen	0.076	−0.152*
	(0.163)	(0.078)
lcrmrte86		0.780***
		(0.045)
Constant	−4.868***	−0.767**
	(0.432)	(0.313)
Observations	90	90
R^2	0.416	0.871
Adjusted R^2	0.389	0.864
Residual Std. Error	0.429 (df = 85)	0.203 (df = 84)
F Statistic	15.152*** (df = 4; 85)	113.903*** (df = 5; 84)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 5: 4.14 Regression Results

	Dependent variable:	
	stndfnl	
	(1)	(2)
atndrte	0.008***	0.005**
	(0.002)	(0.002)
frosh	−0.290**	−0.049
	(0.116)	(0.108)
soph	−0.118	−0.160*
	(0.099)	(0.090)
ACT		0.084***
		(0.011)
priGPA		0.427***
		(0.082)
Constant	−0.502**	−3.297***
	(0.196)	(0.309)
Observations	680	680
R^2	0.029	0.206
Adjusted R^2	0.025	0.200
Residual Std. Error	0.977 (df = 676)	0.885 (df = 674)
F Statistic	6.739*** (df = 3; 676)	34.928*** (df = 5; 674)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 6: 4.14 Part(e) Regression Results

	Dependent variable:
	stndfnl
atndrte	0.005**
	(0.002)
frosh	−0.051
	(0.107)
soph	−0.166*
	(0.089)
priGPAsq	0.087***
	(0.015)
ACTsq	0.002***
	(0.0002)
Constant	−1.812***
	(0.231)
Observations	680
R^2	0.218
Adjusted R^2	0.212
Residual Std. Error	0.878 (df = 674)
F Statistic	37.505*** (df = 5; 674)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7: 4.14 part(f) Regression Results

	Dependent variable:
	stndfnl
atndrte	0.001
	(0.011)
frosh	−0.052
	(0.107)
soph	-0.167^{*}
	(0.089)
priGPAsq	0.087^{***}
	(0.015)
ACTsq	0.002^{***}
	(0.0002)
atndrtesq	0.00003
	(0.0001)
Constant	-1.704^{***}
	(0.407)
Observations	680
R^2	0.218
Adjusted R^2	0.211
Residual Std. Error	0.879 (df = 673)
F Statistic	31.230^{***} (df = 6; 673)
Note:	*p<0.1; **p<0.05; ***p<0.01