Name	Student ID Number	Tutor	Tutorial Day & Time	Tutorial Location
Kim Seang CHY	**998008**	**Charles Siriban**	**Fri 12:00pm**	**The Spot 3014**

Subject Code: ECOM20001	**Subject Name:** Econometric 1
Assignment Name: Assignment 2	
Due Date and Time: 8am, Monday, 05 May 2019	

1. Summary Result

Table 1: Data Summary of World From 2005 to 2013

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
year	2,009	2.59	2,005	2,007	2,011	2,013
numbil	7.31	38.27	0	0	2	469
gdppc	11,491.36	17,042.97	162.81	1,347.51	12,567.03	113,731.60
fullprivproc	13.90	34.66	0.00	0.09	8.22	184.33
gattwto08	29.06	21.56	0	11	46	60
topint08	33.91	12.03	0.00	27.00	42.10	62.40
rintr	8.05	9.87	−28.86	3.91	10.98	51.52
roflaw	−0.10	0.99	−1.95	−0.76	0.66	1.95
nrrents	9.79	13.87	0.00	1.23	11.78	77.29
Mpop	50.68	164.35	0.18	4.0	35.65	1,360

Mpop is an adjustment of the population (pop) to a per million unit.

Looking at the data sets, we can see a typical country in the world in the combined sample of 2005 to 2013 will have 7.31 billionaires, GDP per capita of \$11,491.36, a population of 50.68 million and \$13.9 billion from privatisation proceeds in the economy. We can also see a typical country in 2008 has 29.06 years in the General Agreement on Tariffs and Trade or World Trade Organisation, with a top marginal income tax rate of 33.91%, real interest rate of 8.05%, natural resources rents as 9.79% of the GDP and -0.1 in rule of law index. The data middle of the data time frame was the year 2009.

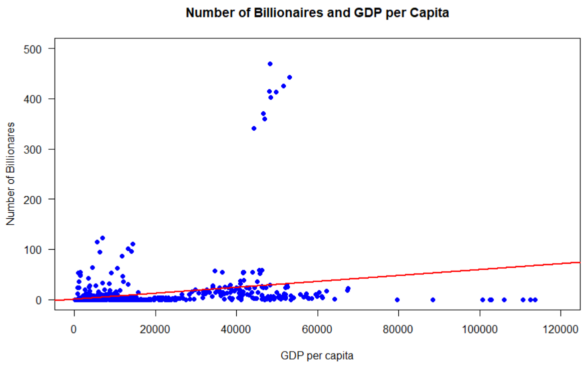
2. First 20 Observations after Sorting From Highest to Lowest Number of Billionaires

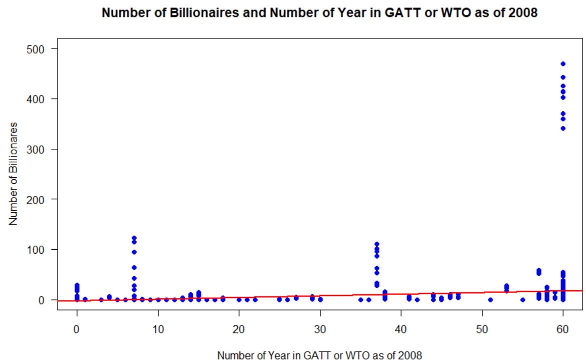
Table 2: Country from the Sorted Data

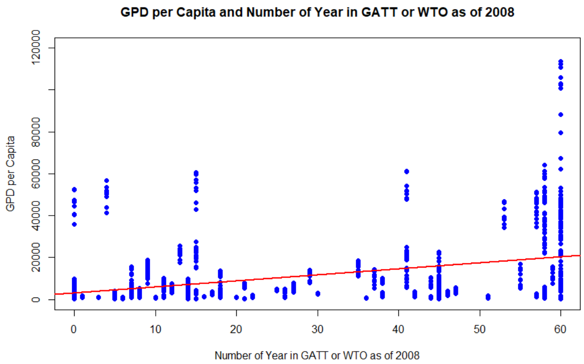
Year	United States	Russian Federation	China	Germany
2005	1	0	0	0
2006	1	0	0	0
2007	1	0	0	0
2008	1	1	0	1
2009	1	0	0	0
2010	1	1	1	0
2011	1	1	1	0
2012	1	1	1	0
2013	1	1	1	1
Total	9	5	4	2

Looking at the sorted data, the countries in the first 20 observations are United States, Russian Federation, China and Germany. Additionally, we can see United Stated taking all top 9 ranking positions in the observation while also having an abnormal amount of billionaire with a minimum of 341 when compared 121 in China in 2013 in the rank 10 of top 20 observation.

3. Graphing Scatter Plot







4. Regression Result

$$\widehat{Numbil}_i = \underset{(1.3468)}{0.488} + \underset{(0.0001)}{0.0006}\, Gdppc_i, R^2 = 0.0699, SER = 36.93$$

Looking at the scatter plots in question 3, we can see there is a positive relationship between the number of billionaires and GDP per capita, a positive relationship between the number of billion and number of year in GATT or WTO. These relationships imply an indirect positive relationship between GDP per capita and the number of years in GATT or WTO as seen in the scatter plots.

Additionally, looking at the linear regression result between Number of Billionaire and GDP per capita, it is implied the number of billionaires increased by 1 for every $1,667 in GDP per capita. However, due to the positive indirect relationship *gattwto08*, this implied the current coefficient of *gdppc* in the regression equation will be higher than its' true coefficient, indicating a positive bias in the regression.

5. Constant and Dummy Variable Traps

Constant

$$\widehat{Numbil}_i = \underset{(1.157)}{7.313}, R^2 = N/A, SER = N/A$$

Separated Dummy Variables From d2005 to d2013

$$\begin{aligned}\widehat{Numbil}_i = \; & \underset{(3.475)}{5.074}\, d2005_i + \underset{(3.475)}{5.713}\, d2006_i + \underset{(3.475)}{6.754}\, d2007_i + \underset{(3.489)}{8.099}\, d2008_i + \\ & \underset{(3.489)}{5.752}\, d2009_i + \underset{(3.489)}{7.24}\, d2010_i + \underset{(3.489)}{8.57}\, d2011_i + \underset{(3.475)}{8.623}\, d2012_i + \\ & \underset{(3.475)}{9.992}\, d2013_i, \qquad R^2 = 0.037, SER = 38.38\end{aligned}$$

Running *numbil* with the sum of d2008 to d2013 as a constant will create an invalid or false regression equation. This problem can be fixed by separating all the dummy variables with no perfect multicollinearity and dropping related dummy variables.

6. Regression Table

Note: The three table below are Reg(1) to Reg(8) are a split of the original single. Population (*pop*) have been scaled to per million unit(*Mpop*) for an easier interpretation.

	Dependent variable:	
	Billionaire in a Country From 2005 to 2013	
	Reg(1)	Reg(2)
	(1)	(2)
Population per Million	0.0712***	0.0739***
	(0.0139)	(0.0136)
GDP per capita		0.0006***
		(0.0001)
Constant	3.7032***	−3.5949***
	(0.6586)	(1.1572)
Observations	1,094	1,094
R^2	0.0935	0.1704
Adjusted R^2	0.0927	0.1689
Residual Std. Error	36.4578 (df = 1092)	34.8931 (df = 1091)
F Statistic	112.6449*** (df = 1; 1092)	112.0548*** (df = 2; 1091)
Note:		*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:			
	Billionaire in a Country From 2005 to 2013			
	Reg(3)	Reg(4)	Reg(5)	Reg(6)
	(1)	(2)	(3)	(4)
Population per Million	0.0723***	0.0817***	0.0827***	0.0830***
	(0.0133)	(0.0189)	(0.0191)	(0.0192)
GDP per capita	0.0006***	0.0007***	0.0007***	0.0007***
	(0.0001)	(0.0002)	(0.0002)	(0.0002)
Number of Year in GATT or WTO as 2008	0.1227***	0.1426***	0.1626***	0.1573***
	(0.0321)	(0.0448)	(0.0489)	(0.0487)
Privatisation Proceed in the economy		−0.1111	−0.0996	−0.1003
		(0.0855)	(0.0830)	(0.0832)
Top Marginal income Tax as 2008			−0.2693***	−0.2683***
			(0.0757)	(0.0756)
Real Interest Rate as 2008				0.0451*
				(0.0250)
Rule of Law Index as 2008	−6.4273***	−6.9363***	0.6249	0.3053
	(1.6863)	(2.0632)	(1.3308)	(1.3247)
Constant	−6.4273***	−6.9363***	0.6249	0.3053
	(1.7941)	(1.7959)	(3.3352)	(3.4274)
Observations	1,094	1,094	1,094	1,094
R^2	0.1745	0.1811	0.1865	0.1867
Adjusted R^2	0.1722	0.1781	0.1828	0.1822
Residual Std. Error	34.8226 (df = 1090)	34.6988 (df = 1089)	34.6001 (df = 1088)	34.6134 (df = 1087)
F Statistic	76.8123*** (df = 3; 1090)	60.2199*** (df = 4; 1089)	49.8955*** (df = 5; 1088)	41.5754*** (df = 6; 1087)
Note:				*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:	
	Billionaire in a Country From 2005 to 2013	
	Reg(7)	Reg(8)
	(1)	(2)
Population Per Million	0.0831***	0.0830***
	(0.0192)	(0.0193)
GDP per capita	0.0007***	0.0007***
	(0.0002)	(0.0002)
Number of Year in GATT or WTO as 2008	0.1568***	0.1571***
	(0.0484)	(0.0485)
Privatisation Proceed in the economy	−0.0988	−0.0983
	(0.0831)	(0.0835)
Top Marginal income Tax as 2008	−0.2773***	−0.2767***
	(0.0775)	(0.0778)
Real Interest Rate as 2008	0.0501**	0.0502**
	(0.0220)	(0.0219)
Rule of Law Index as 2008	0.3848	0.4537
	(1.4301)	(1.4268)
Natural Resources Rents as	0.054483*	0.055657*
	(0.0283)	(0.0290)
Year 2005		−1.8362
		(4.3061)
Year 2006		−1.8356
		(4.4592)
Year 2007		−1.9130
		(4.6942)
Year 2008		−1.6539
		(5.0310)
Year 2009		−2.9530
		(4.3902)
Year 2010		−1.9223

		(4.5975)
Year 2011		−1.5027
		(4.7097)
Year 2012		−1.1188
		(4.7275)
Constant	0.1044	1.7619
	(1.2709)	(3.4761)
Observations	1,094	1,094
R^2	0.1870	0.1873
Adjusted R^2	0.1810	0.1753
Residual Std. Error	34.6387 (df = 1085)	34.7591 (df = 1077)
F Statistic	31.1877*** (df = 8; 1085)	15.5169*** (df = 16; 1077)
Note:		*p<0.1; **p<0.05; ***p<0.01

7. Regression Analysis

· While not reflected in the table above due rounding, there was indeed a slight upward bias in *gdppc* of $5.6 \times 10^{-5}(6.23 \times 10^{-4} - 5.67 \times 10^{-4})$ as stated in question 4 when comparing Reg(2) and Reg(3) regression analysis.

· Looking at Reg(2) to Reg(8), the coefficient on *gdppc* started to "settle down" in from the Reg(5) to the Reg(8) where it is vary less than (1×10^{-5}), when compare Reg(2) to Reg(4) where it is vary by at least 3×10^{-5}). Therefore, this indicate *topint08* is the last of the omitted variable that create a significant bias in *gdppc*.

· Looking Reg(8) in the last table, the coefficients with a statistically significant of 5% $(\alpha = 0.05)$ are *Mpop*, *gdppc*, *gattwto08*, *topint08* and *rintr*.

· The statistically significant coefficient on *Mpop* with a standard deviation of 164.35 million , implied an increase in the number of billionaire in a country 2005 to 2013 by 0 (0.0830) per 1 million increase in population or by 13 ($164.35 \times 0.0830 = 13.64$) per 163.35(one standard deviation) million increase in population.

· The statistically significant coefficient on *gdppc* with a standard deviation of $\$17, 042.97$, implied an increase in the number of billionaire a country 2005 to 2013 by 0 (0.0007) per 1 dollar increase in GDP per capita or by 12 ($170,042.97 \times 7.19 \times 10^{-4} = 12.25$) per $17, 042.97(one standard deviation) in GDP per capita.

· The statistically significant coefficient on *gattwto08* with a standard deviation of 21.56 years, implied an increase in the number of billionaire a country 2005 to 2013 by 0 (0.1571) per 1 year in GATT or WTO or by 3 ($21.56 \times 0.1571 = 3.39$) per 21.56 years(one standard deviation) in GATT or WTO.

· The statistically significant coefficient on *topint08* with a standard deviation of 12.02 %, implied an increase in the number of billionaire a country 2005 to 2013 by 0 (0.2767) per 1% decrease top marginal income tax rate or by 3 (12.02 -0.2767 = 3.33) per 13.87%(one standard deviation) decrease in top marginal income tax rate.

· The statistically significant coefficient on *rintr* with a standard deviation of 9.87 %, implied an increase in the number of billionaire a country 2005 to 2013 by 0 (0.0502) per 1% increase in the real interest rate or still by 0 ($9.87 \times 0.0502 = 0.50$) per 9.87%(one standard deviation) increase in the real interest rate.

8. Comparing R^2 and Adjusted R^2 between Reg(5) and Reg(8)

· In multiple linear regression, adding more independent variables will increase adjusted R^2, if the new regressors have explanatory power and will decrease adjusted R^2 if the new regressor does not have explanatory power. On the other hand, adding more independent variables will always increase non-adjusted R^2, even if the new variable does not have explanatory power.

· The adjusted R^2 punishment justified the Reg(5) have a higher adjusted R^2 then the Reg(8), despite the Reg(5) having a smaller R^2 and less independent variables than the Reg(8).

9. Unnecessary Variables

The last regression to add an independent variable with a statistically significant coefficient of 5% for the number of billionaires is Reg(5), which help to explain the decreasing trend in adjusted R^2 from Reg(5) to Reg(8). Therefore, we will reject Reg(6) to Reg(8) in the table from question 6 to reduce bias in the number of billionaires regression model.

10. Joint Hypotheses testing

· Testing joint null that the coefficients on B_*roflaw* equal to 0 and B_*nrrents* equal to 0: has F-statistics = 1.91 with df1 = 2 and df2 = 1077 and p-value = 0.1482. Therefore, we failed to reject the null hypotheses at 5% level of statistical significance that natural resources rents as a % of GDP in 2008 and/or rule of index in 2008 has no impact on the number of billionaires in a country from 2005 to 2013.
· Testing joint null that the coefficients on B_*topint08* equal to B_*rintr*: has F-statistics = 15.026 with df1 = 1 and df2 = 1077 and p-value = 0.0001124. Therefore, we reject the null hypotheses at 5% level of statistical significance that top marginal income tax rate in 2008 has the same impact as real interest rate in 2008 on the number of billionaires in a country from 2005 to 2013.

```
#Appendix


#Pre Question

setwd("C:/Users/Kim CHY/Desktop/Assginment 2")  #Setting Working Dirctory
df <- read.csv(file="as2_billions.csv")         #Read File
options(scipen=0)                               #Preferred regular number over scientific notation


#Installing Stargazer Package
install.packages("stargazer")
install.packages("AER")

#loading the library
library(AER)
library(stargazer)


#Rescale variable to population to per million units
df$Mpop <- (df$pop/1000000)



#Question 1: Data Summary


#Getting Summary in R
summary(df)

#Creating a pre-formart table for LaTex Omitting  Number of observation
stargazer(df, title = "Data Summary of the World From 2005 to 2013",
          omit.summary.stat = c("n"), digits= 2)



#Question 2: Top 20

#Creating sorted data of the first 20 observation in Working dirctory
write.csv(data.frame(df[order(-df$numbil), ][1:20, 1:12]), file = "df_sorted.csv")

#Assign the sorted data
df_sorted <- read.csv(file = "df_sorted.csv")

#Counting and Listing country and year in the new data set
table(df_sorted$country)
table(df_sorted$year)




#Question 3: Graph Ploting
#Plotting numbil vs gdppc with linear regression line and Remove Excess variables
y <- df$numbil
x <- df$gdppc

plot(x, y,
     main="Number of Billionaires and GDP per Capita",
     xlab="GDP per capita", ylab="Number of Billionaires",
     pch=19, las=1, col="blue", ylim = c(0,500), xlim = c(0,120000))
abline(lm(y ~ x, data=df), col="red", lwd=2)
remove(x, y)

#Plotting numbil vs gattwto08 with linear regression line and Remove Excess variables
y <- df$numbil
x <- df$gattwto08

plot(x, y,
     main="Number of Billionaires and Number of Year in GATT or WTO as of 2008",
     xlab="Number of Year in GATT or WTO as of 2008", ylab="Number of Billionaires",
     pch=19, las=1, col="blue",  ylim = c(0,500))
abline(lm(y ~ x, data=df), col="red", lwd=2)
remove(x, y)


#plotting gdppc vs gattawto with Linear Regression line and Remove Excess variables
y <- df$gdppc
x <- df$gattwto08

plot(x, y,
     main="GPD per Capita and Number of Year in GATT or WTO as of 2008",
     xlab="Number of Year in GATT or WTO as of 2008", ylab="GPD per Capita",
     pch=19, col="blue", ylim = c(0,120000))
abline(lm(y ~ x, data=df), col="red", lwd=2)
remove(x, y)
```

```
#All Image are exported Manually


#Question 4:

reg_numbil_gdppc <- lm(df$numbil~df$gdppc, data = df)
summary(reg_numbil_gdppc)


#Question 5



#Create Dummy variable for 2005 to 2013 with Column Name 1 to 9 temporary
number = c(2005:2013)
df <- cbind( df , sapply( number , function(x) as.numeric( df$year == x ) ) )

#Rename the Column Name to d2005 to d2013
new_year = c("d2005","d2006","d2007","d2008","d2009","d2010","d2011","d2012","d2013")
colnames(df)[13:21] <- new_year

#To Run the code above sucessfully the Column index must be correct. In this case, no adjustment to the existing data
frame

#Remove Excess variables
remove(number, new_year)



#create a constant data in the data frame with the sum of d2005 to d2013 for each row temporary
df$d_constant <- df$d2005+df$d2006+df$d2007+df$d2008+df$d2009+df$d2011+df$d2010+df$d2012+df$d2013
summary(df$d_constant)

#Creating regression with the constant and printing the regression result
reg_constant <- lm(df$numbil~df$d_constant, data=df)
summary(reg_constant)

#Creating regression with d2005 to d2013 as separatedummy variable the and printing the regression result
reg_d2005_to_d2013 <- lm(df$numbil~df$d2005+df$d2006+df$d2007+df$d2008+df$d2009+df$d2010+df$d2011+df$d2012, data=df)
summary(reg_d2005_to_d2013)



#Question 6:

# Creating Reg (1) with following variable: pop
reg_1 <- lm(numbil~Mpop , data=df)

#Creating Reg (2) with following variables: pop, gdppc
reg_2 <- lm(numbil~Mpop+gdppc , data=df)

#Creating Reg (3) with following variable: pop, gdppc, gattwto08
reg_3 <- lm(numbil~Mpop+gdppc+gattwto08, data=df)

#Creating Reg (4) with following variable: pop, gdppc, gattwto08, fullprivproc
reg_4 <- lm(numbil~Mpop+gdppc+gattwto08+fullprivproc, data=df)

#Creating Reg (5) with following variable: pop, gdppc, gattwto08, fullprivproc, topint08
reg_5 <- lm(numbil~Mpop+gdppc+gattwto08+fullprivproc+topint08, data=df)

#Creating Reg (6) with following variable: pop, gdppc, gattwto08, fullprivproc, topint08, rint
reg_6 <- lm(numbil~Mpop+gdppc+gattwto08+fullprivproc+topint08+rintr, data=df)

#Creating Reg (7) with following variable: pop, gdppc, gattwto08, fullprivproc, topint08, rint, roflaw, nrrents
reg_7 <- lm(numbil~Mpop+gdppc+gattwto08+fullprivproc+topint08+rintr+roflaw+nrrents, data=df)

#Creating Reg (8) with following variable: pop, gdppc, gattwto08, fullprivproc, topint08, rint, roflaw, nrrents,d2006,
d2007, d2008, d2009, d2010, d2011, d2012, d2013
reg_8 <-
lm(numbil~Mpop+gdppc+gattwto08+fullprivproc+topint08+rintr+roflaw+nrrents+d2005+d2006+d2007+d2008+d2009+d2010+d2011+d2012,
 data=df)




#Saved the robust standard errors for reg(1)
cov_1=vcovHC(reg_1, type = "HC1")
se_1=sqrt(diag(cov_1))

#Saved the robust standard errors for reg(2)
cov_2=vcovHC(reg_2, type = "HC1")
se_2=sqrt(diag(cov_2))

#Saved the robust standard errors for reg(3)
cov_3=vcovHC(reg_3, type = "HC1")
se_3=sqrt(diag(cov_3))
```

```
#Saved the robust standard errors for reg(4)
cov_4=vcovHC(reg_4, type = "HC1")
se_4=sqrt(diag(cov_4))


#Saved the robust standard errors for reg(5)
cov_5=vcovHC(reg_5, type = "HC1")
se_5=sqrt(diag(cov_5))


#Saved the robust standard errors for reg(6)
cov_6=vcovHC(reg_6, type = "HC1")
se_6=sqrt(diag(cov_6))

#Saved the robust standard errors for reg(7)
cov_7=vcovHC(reg_7, type = "HC1")
se_7=sqrt(diag(cov_7))

#Saved the robust standard errors for reg(8)
cov_8=vcovHC(reg_8, type = "HC1")
se_8=sqrt(diag(cov_8))


#Producing Table for Reg (1) and Reg(2) in LaTex
stargazer(reg_1, reg_2,
          digits=4, se=list(se_1, se_2),
          column.labels = c("Reg(1)", "Reg(2)"),
          dep.var.labels=c("Billionaire in a Country From 2005 to 2013"),
          covariate.labels=
            c("Population per Million",
              "GDP per capita"))


#Producing Table For Reg(3) to Reg(6) in LaTex

stargazer(reg_3, reg_4, reg_5, reg_6,
          digits=4, se=list(se_3, se_4, se_5, se_6),
          column.labels = c("Reg(3)", "Reg(4)", "Reg(5)", "Reg(6)"),
          dep.var.labels=c("Billionaire in a Country From 2005 to 2013"),
          covariate.labels=
            c("Population per Million",
              "GDP per capita",
              "Number of Year in GATT or WTO as 2008",
              "Privatisation Proceed in the economy",
              "Top Marginal income Tax as 2008",
              "Real Interest Rate as 2008",
              "Rule of Law Index as 2008",
              "Natural Resources Rents as % in GDP in 2008"))


#Producing Table For Reg(7) and Reg(8) in LaTex
stargazer(reg_7, reg_8,
          digits=4, se=list(se_7, se_8),
          column.labels = c("Reg(7)", "Reg(8)"),
          dep.var.labels=c("Billionaire in a Country From 2005 to 2013"),
          covariate.labels=
            c("Population Per Million",
              "GDP per capita",
              "Number of Year in GATT or WTO as 2008",
              "Privatisation Proceed in the economy",
              "Top Marginal income Tax as 2008",
              "Real Interest Rate as 2008",
              "Rule of Law Index as 2008",
              "Natural Resources Rents as % in GDP in 2008",
              "Year 2005",
              "Year 2006",
              "Year 2007",
              "Year 2008",
              "Year 2009",
              "Year 2010",
              "Year 2011",
              "Year 2012"))


#Question 7:

#Producing Regression table in text and output them in a file name regression.txt for Reg(1) to Reg(8)
stargazer(reg_1, reg_2,reg_3, reg_4, reg_5, reg_6,reg_7, reg_8, type = "text",
          digits=6, se=list(se_1, se_2, se_3, se_4, se_5, se_6, se_7, se_8),
          column.labels = c("Reg(1)","Reg(2)","Reg(3)","Reg(4)","Reg(5)","Reg(6)","Reg(7)", "Reg(8)"),
          dep.var.labels=c("Billionaire in a Country From 2005 to 2013"),
          covariate.labels=
            c("Population per Million",
              "GDP per capita",
              "Number of Year in GATT or WTO as 2008",
              "Privatisation Proceed in the economy",
```

```
                 "Top Marginal income Tax as 2008",
                 "Real Interest Rate as 2008",
                 "Rule of Law Index as 2008",
                 "Natural Resources Rents as % in GDP in 2008",
                 "Year 2005",
                 "Year 2006",
                 "Year 2007",
                 "Year 2008",
                 "Year 2009",
                 "Year 2010",
                 "Year 2011",
                 "Year 2012",
                 "Year 2013"), out = "regression.txt")


# Printing reg(8) results with robust standard errors
coeftest(reg_8, vcov = vcovHC(reg_8, "HC0"))

#Printing the current data summary
stargazer(df, type="text")




#Question 10

#Testing for H0:B_roflaw=0 and B_nrrents=0, H0:B_roflaw =! 0 or B_nrrents =! 0
linearHypothesis(reg_8,c("roflaw=0","nrrents=0"),vcov = vcovHC(reg_8, "HC1"))
summary(reg_8)

#Testing for H0:B_topint08 = B_rintr, H1: B_topint08 =! B_rintr
linearHypothesis(reg_8,c("topint08=rintr"),vcov = vcovHC(reg_8, "HC1"))
summary(reg_8)
```