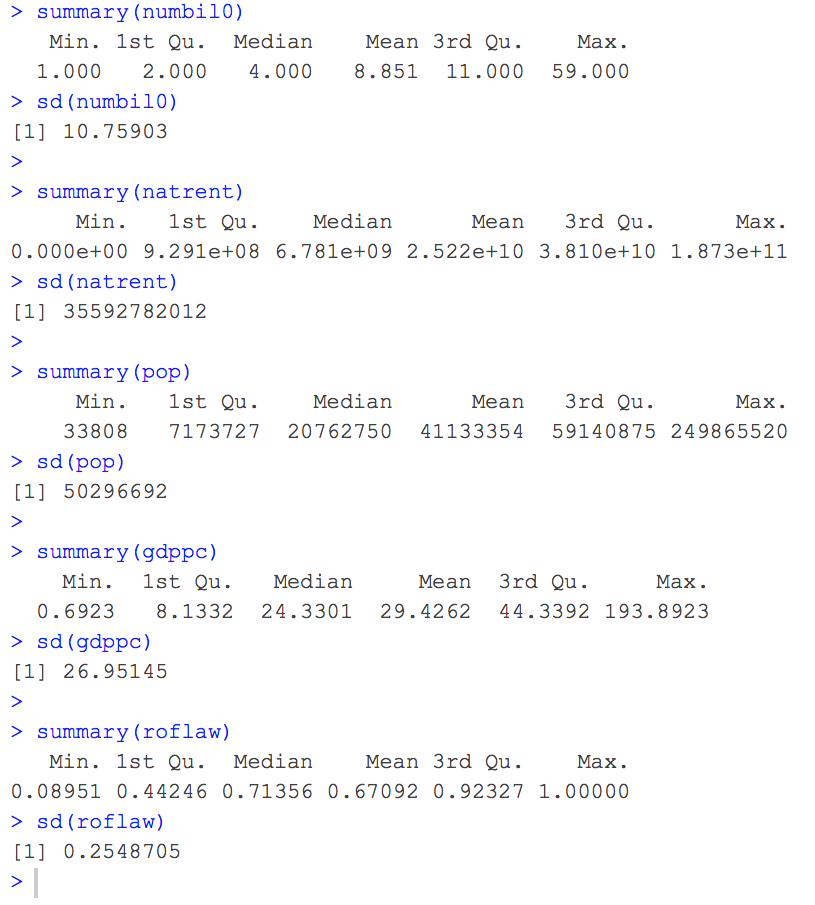
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Student ID Number** | **Tutor** | **Tutorial Day & Time** | **Tutorial Location** |
| Chan Jie Ho | 961948 | Richard Hayes | Monday 3:15 | The Spot 3031 |

**ASSIGNMENT 2**

**Question 1**



A typical country would have approximately 8 to 9 billionaires, with a US$25.22 billion in total resources rent, a population size of 41 133 354 people, each of whom earns a GDP per capita of US$29.4262 thousand, and has a rule of law index of 0.67092, in which there is a standard deviation between countries of approximately 10 to 11 billionaires, US$35 billion in total resources rent, a stand deviation of 50 million people for the population size, approximately US$26.95 thousand in GDP per capita and a rule of law index standard deviation of 0.25. Looking at the means and standard deviations, it should be safe for us to scale the total natural resources rent to be in terms of $10 billion and the population to be in terms of 10 million people.

**Question 2**



We can see a very distinctly similar pattern in all three graphs where not only is there a positive correlation between each pair of variables, but they all have a very significant concentration of countries in the lower left corner of the graphs. However, while we can also see that there is a very strong positive correlation between the number of billionaires and the population as well as between the total natural resources rent and the population of a country, the correlation between the total natural resources rent and the number of billionaires has a relatively weak correlation. This would mean that in the case of a regression between the number of billionaires and the total natural resources rent, there would be a significant positive bias, which is due to the strong correlation between the number of billionaires and the population, and therefore by controlling for this omitted variable bias, the coefficient of the total natural resources rent would decrease and it may be lower than the coefficient for the population. Thus, showing that the total natural resources rent does not contribute to the number of billionaires in a country as much as population does.

**Question 3**



There is no clear linear correlation between the number of billionaires a country has and the GDP per capita which is due to the concentration of countries all along the left side of the graph as well as some countries being all along the bottom of the graph. However, the distribution of the countries in the scatter plots of numbil0 against roflaw and the gdppc against roflaw both have a moderately and strong positive correlation between the variables respectively, and this would mean that the rule of law index is correlated with both the GDP per capita and the number of billionaires a country has, and would therefore have a positive bias on the coefficient of GDP per capita if we do not account for this correlation. This is all true despite the very large standard errors the regression slope between GDP per capita and the number of billionaires, in which the true slope may have a chance of being negative.

**Question 4**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Billionaires and country characteristics | | | | | |
|  | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** |
| natrent | 0.33\* | -0.01 | -0.05 | 0.24\* | 0.22\* |
|  | (0.15) | (0.15) | (0.14) | (0.11) | (0.11) |
| pop |  | 0.70\*\* | 0.90\*\* | 1.06\*\* | 1.08\*\* |
|  |  | (0.13) | (0.15) | (0.14) | (0.14) |
| gdppc |  |  | 0.10\*\* | -0.03\*\* | -0.03\*\* |
|  |  |  | (0.02) | (0.01) | (0.01) |
| roflaw |  |  |  | 21.97\*\* | 22.69\*\* |
|  |  |  |  | (2.27) | (2.29) |
| d2006 |  |  |  |  | 0.86 |
|  |  |  |  |  | (1.75) |
| d2007 |  |  |  |  | 1.93 |
|  |  |  |  |  | (1.72) |
| d2008 |  |  |  |  | 2.59 |
|  |  |  |  |  | (1.83) |
| d2009 |  |  |  |  | 0.44 |
|  |  |  |  |  | (1.66) |
| d2010 |  |  |  |  | 1.42 |
|  |  |  |  |  | (1.72) |
| d2011 |  |  |  |  | 3.39 |
|  |  |  |  |  | (1.83) |
| d2012 |  |  |  |  | 3.87\* |
|  |  |  |  |  | (1.80) |
| d2013 |  |  |  |  | 6.10\*\* |
|  |  |  |  |  | (1.85) |
| Constant | 8.03\*\* | 5.99\*\* | 2.41\*\* | -10.04\*\* | -12.81\*\* |
|  | (0.65) | (0.55) | (0.86) | (1.55) | (1.93) |
| N | 444 | 444 | 444 | 444 | 444 |
| adj. R^2 | 0.01 | 0.10 | 0.15 | 0.28 | 0.30 |
| F | 5.19 | 26.46 | 27.38 | 44.52 | 16.77 |
| Dependent variable is the number of billionaires in the country. natrent is natural resources rents in billions of dollars. pop is total population of a country in millions. gdppc is GDP per capita in thousands of dollars. roflaw is an index between 0 and 1 on the level of the rule of law in a country. dXXXX is a dummy variable for the year XXXX. Heteroskedasticity robust standard errors in parentheses. Statistical significance from two-sided tests of the null of no effect marked as \* for 5% and \*\* for 1% | | | | | |

Looking at the table, we can see in the first column that natrent has a very large significance, due to it being the only regressor. However, once we had added pop, we can see this coefficient decreasing significantly to the point that it become statistically insignificant in both the 1% and 5% level in our regression, which is in line with our findings in Question 2, as this is due to it being positive biased due to the significant correlation between numbil0, natrent, and population, where the magnitudes would imply an increase of one billionaire for every 10 million people increment in a country. This further decreases once we had included gdppc, albeit just a little, but increased significantly when we had roflaw, indicating a very significant negative bias due to a most likely very strong negative correlation between roflaw and natrent, which is expectable as countries rich in natural resources would most not need to rent it from other countries and can instead export it to other countries, thus serving as a form of income to the country which would in turn go into improving their the country in many aspects such as education and other infrastructures, thus increasing their roflaw, and hence, a negative correlation between the two variables.

Not controlling for roflaw also introduces a positive bias in GDP per capita, as shown by the significant decrease in the GDP per capita coefficient, which would imply a positive correlation between the two variables, which could be due to the fact that a higher roflaw would most likely imply better protections against the exploitation of the working class, thus leading to them receiving a higher income and GDP per capita. However, this bias lead to the small but positive coefficient indicating an increase of 0.1 billionaires for every US$1000 increase in GDP per capita to becoming a negative coefficient, implying that an increase of GDP per capita would lead to a decrease in the number of billionaires, which is highly unexpected as we would expect a higher GDP per capita would likely mean an increase in the likelihood of a country producing more billionaires. This may either be due to more omitted variable biases that we have yet to explore, or the slightly unlikely chance that it may be due to the increase in GDP per capita coming from income paid directly from these potential billionaires' pockets, thus leading to some not being able to upgrade to that status.

Looking deeper into the roflaw coefficients, the large coefficients would mean that for every 1-unit change in roflaw, there should be an increase of approximately 22 billionaires, but since roflaw only goes from 0 to 1, it would be better to interpret this as a 0.1-unit change in roflaw, a 10 percentage points change on the rule of law 0-1 scale, would lead to an increase of 2 to 3 billionaires. These significant biases also tie in with the large increases in the adjusted R squared, the largest increase being when we had introduced roflaw into our regression, thus serving as a point to show that the all four variables jointly explain about 28% of the variation in the number of billionaires across countries.

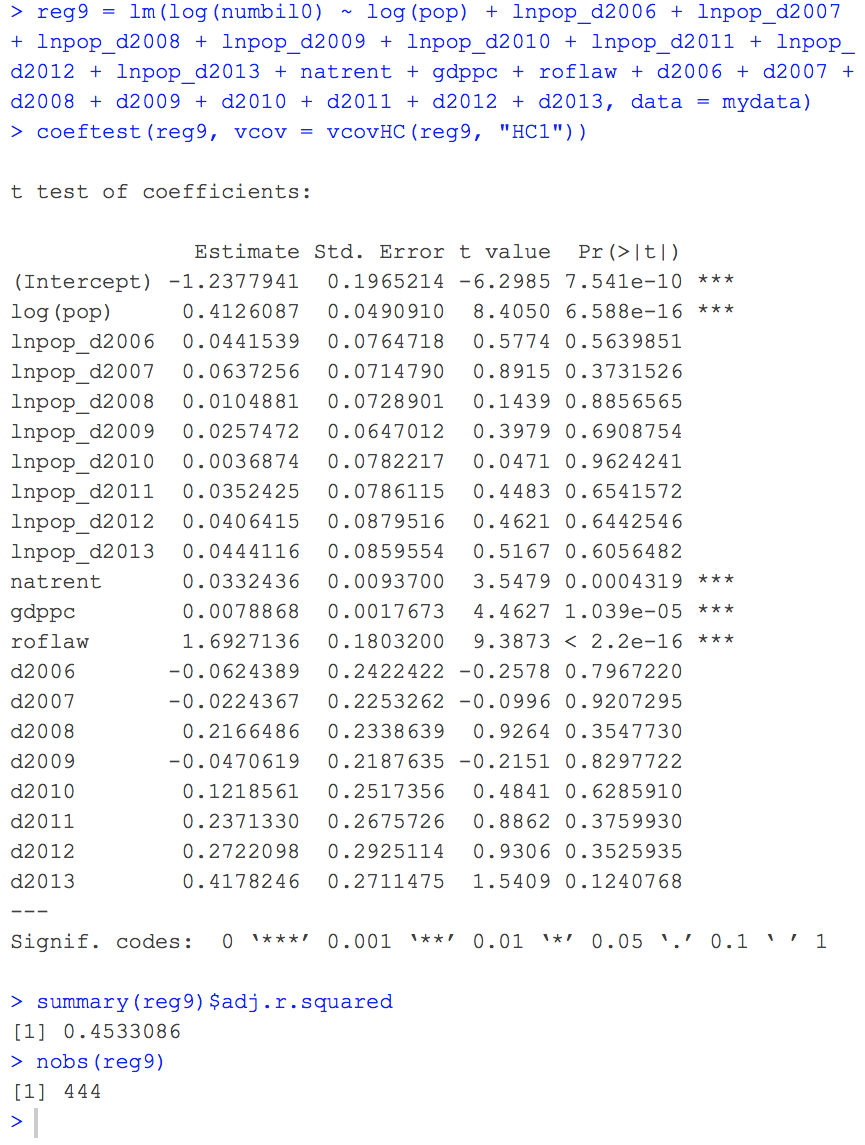
Looking at the last column, the year dummy variables and their coefficient estimates allow us to see how the different years had different "effects" on the number of billionaires in the country relative to the number of billionaires they had in 2005. The coefficient estimate of 6.1 on the 2013 year-dummy, on the other hand, is statistically significant at the 1% level and this means that in 2013, countries had an increase approximate 6 billionaires relative to the number of billionaires they had in 2005, which could be due to there being better economic policies in 2012-2013 compared to 2004-2005 that has major effects on the economy, thus the statistical significance, as compared to other years which were not statistically significant.

**Question 5**

|  |  |  |  |
| --- | --- | --- | --- |
| Billionaires and population size | | | |
|  | **(1)**  **numbil0** | **(2)**  **log(numbil0)** | **(3)**  **log(numbil0)** |
| lnpop | 3.87\*\* |  | 0.44\*\* |
|  | (0.44) |  | (0.03) |
| pop |  | 0.10\*\* |  |
|  |  | (0.01) |  |
| natrent | 0.16 | 0.05\*\* | 0.04\*\* |
|  | (0.13) | (0.01) | (0.01) |
| gdppc | 0.08\*\* | -0.01\*\* | 0.01\*\* |
|  | (0.02) | (0.00) | (0.00) |
| roflaw | 13.78\*\* | 2.46\*\* | 1.62\*\* |
|  | (1.86) | (0.20) | (0.18) |
| Constant | -5.39\*\* | -0.42\*\* | -0.06 |
|  | (1.30) | (0.14) | (0.13) |
| N | 444 | 444 | 444 |
| adj. R^2 | 0.31 | 0.32 | 0.44 |
| F | 51.60 | 52.28 | 89.66 |
| numbil0 is the number of billionaires in the country. natrent is natural resources rents in billions of dollars. pop is total population of a country in millions. gdppc is GDP per capita in thousands of dollars. roflaw is an index between 0 and 1 on the level of the rule of law in a country. Heteroskedasticity robust standard errors in parentheses. Statistical significance from two-sided tests of the null of no effect marked as \* for 5% and \*\* for 1% | | | |

Looking at the first column, the coefficient of lnpop means that a one percent increase (100 thousand people) in pop is associated with a 0.01 x 3.87 = 0.0387 increase in the number of billionaires, while the second column tells us that a 10-million-people-increase in pop is associated with a 100 x 0.1 = 10 percent increase in the number of billionaires. The last column means that a one percent increase in pop is associated with a 0.44 percent increase in the number of billionaires, which implies that the elasticity of the number of billionaires with respect to population is 0.44. All these relationships are statistically significant at the 1% level and this means that they each play a role in explaining the variation in the number of billionaires a country has, and therefore cannot be omitted from our model.

**Question 6**

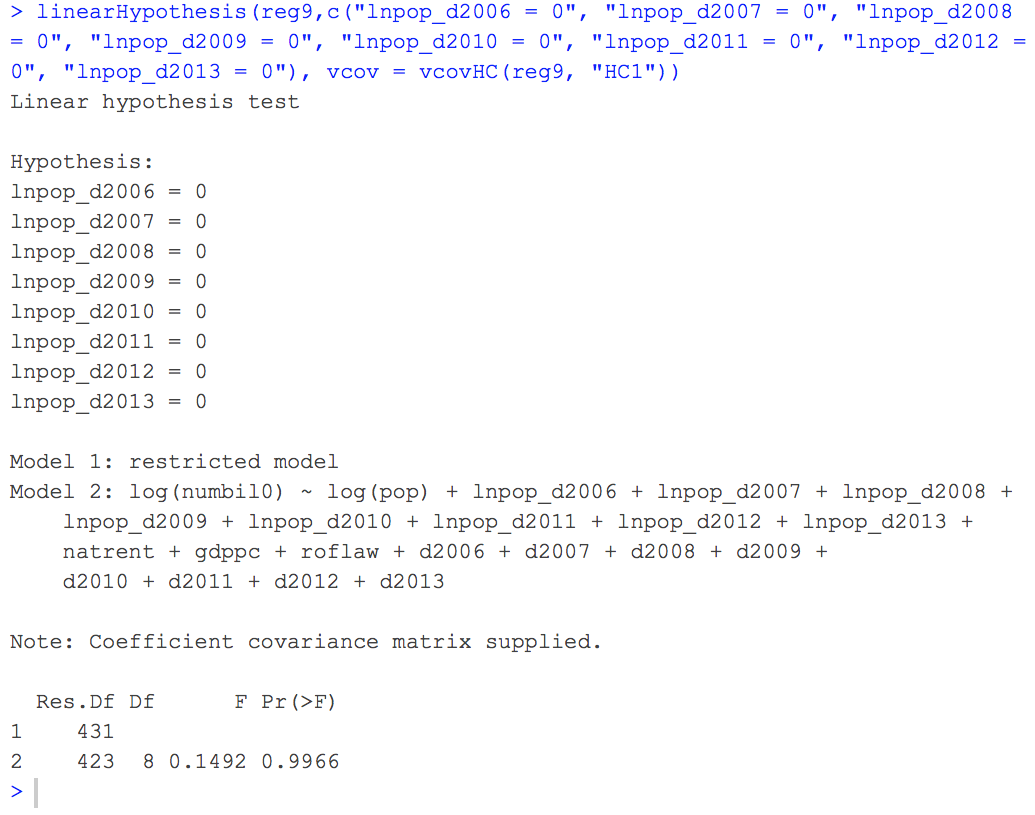


From our results, we can see that a one percent increase in pop is associated with a 0.413 percent increase in the number of billionaires, a relationship that is statistically significant at the 0.1% level. That is, the elasticity of the number of billionaires with respect to population is 0.413. This elasticity can be altered depending on the year as a result of the interaction between log(pop) and the year dummy variables, but they are all statistically insignificant and can therefore be said to not play truly a role in the elasticity of the number of billionaires a country has.

Looking specifically at 2013, we get our base elasticity of the number of billionaires with respect to population as 0.413 and then add an additional 0.044 due to it being 2013 due to the interaction between the population and the years, thus giving us a total elasticity of 0.457. However, only the base elasticity is statistically significant at the 0.01% level while the elasticity increment due to the interaction between the two variables is not. Therefore, we can say that the year is insignificant when it comes to helping us build a model to show the elasticity of the number of billionaires with respect to population. However, the elasticity of the number of billionaires with respect to population is still statistically significant at the 0.01% level and since this number is less than 1, with or without the interaction between the year and the population, we can conclude that the number of billionaires is inelastic with respect to population, meaning that changes in population would result in a less than proportional change in the number of billionaires.

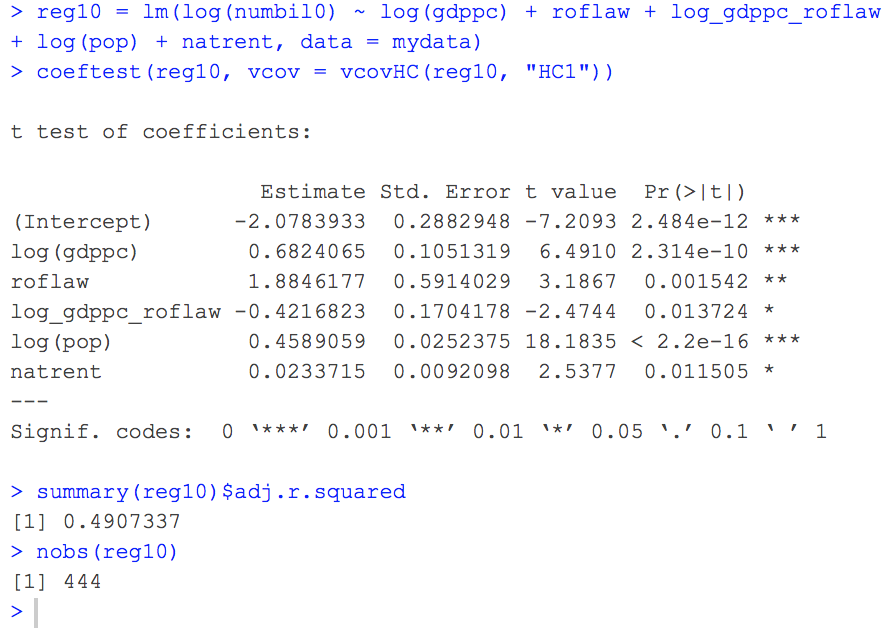
Also, by ensuring that 2005 is the base year and omitting this in our regression, we are able to avoid the dummy variable trap when looking at the impact the different years has on the number of billionaires a country has. We are also able to avoid it by doing it this way when testing the interaction between the years and the elasticity of the number of billionaires with respect to the population.

**Question 7**



Performing this test with a degree of freedom of 8 from the 8 variables that we are testing, we get a very low F-stat of 0.1492, which has a p value of 0.9966, thus meaning that we cannot reject the null that the interaction between the population and the dummy year variables is equal to 0 as a low F-statistic would imply that difference in the adjusted R-squared between the restricted and unrestricted models is very small, and can therefore be interpreted as there being a probability that the increase of the adjusted R squared may just be due to chance. This means that we can conclude that the year does not affect the elasticity of the number of billionaires in a country with respect to population

**Question 8**



Looking at the signs of regression coefficients, we get a positive coefficient for both log(gdppc) and roflaw but a negative coefficient for log(gdppc)\_roflaw, all of which are statistically significant at the 0.1%, 1%, and 5% level respectively. These signs show mean that the elasticity of the number of billionaires with respect to GDP per capita is positive, thus meaning that a positive change in GDP per capita would mean a positive change in the number of billionaires a country has, which decreases when a country has a higher rule of law index, but when we look at the rule of law index individually, we can see a positive relationship between the variable and the number of billionaires. This all means that GDP per capita and the rule of law index both have a positive correlation with the number of billionaires that a country has, but when put together, an increase in the rule of law index would have a diminishing positive effect that the GDP per capita has on the number of billionaires.

**Question 9**

|  |  |  |  |
| --- | --- | --- | --- |
| Roflaw | Elasticity | F-Statistic | P-Value |
| **0.1** | 0.640 | 49.618 | 7.282e-12 |
| **0.2** | 0.598 | 59.116 | 9.882e-14 |
| **0.3** | 0.556 | 69.903 | 8.365e-16 |
| **0.4** | 0.514 | 78.272 | 2.2e-16 |
| **0.5** | 0.472 | 76.582 | 2.2e-16 |
| **0.6** | 0.429 | 61.125 | 4.029e-14 |
| **0.7** | 0.387 | 40.416 | 5.163e-10 |
| **0.8** | 0.345 | 23.887 | 1.436e-06 |
| **0.9** | 0.303 | 13.413 | 0.0002804 |
| **1** | 0.261 | 7.3284 | 0.007053 |

As we can see from the elasticity row that decreases as the rule of law index increases, we can say the rule of law index has a downward effect on the elasticity of the number of billionaires with respect to the GDP per capita. This finding goes in hand with the negative coefficient that we got for log(gdppc)\_roflaw in our previous question, which is due to the relationships between the three variables. This would mean that as the rule of law index increases, the significance of the effect of GDP per capita on the elasticity of the number of billionaires, and this may be due to the fact that an increase of a rule of law index would mean that the country would have stricter legislations on a wide variety of aspects, thus meaning that many businesses would likely have to spend more money to ensure that they are following ethical methods and cannot use unethical methods to generate income. Only the elasticity of the number of billionaires with respect to GDP per capita when the rule of law index is 1 is statistically significant at 1% level, thus meaning that GDP per capita is an important factor to consider when building our model to test the elasticity.

**APPENDIX – R CODE**

# Set working directory using setwd() function

setwd("~/Downloads/ECOM20001/Assignment/Assignment 2")

# Read data from a csv file using read.csv() function

mydata=read.csv(file = "billionaires\_clean2.csv")

# Create variables for each table attribute simplify script

country = mydata$country

year = mydata$year

numbil0 = mydata$numbil0

natrent = mydata$natrent

pop = mydata$pop

gdppc = mydata$gdppc

roflaw = mydata$roflaw

# ==========================================================================================================================

# QUESTION 1

# ----------

# Summary statistics and standard deviation for numbil0, natrent, pop, gdppc, roflaw using the summary() and sd() functions

summary(numbil0)

sd(numbil0)

summary(natrent)

sd(natrent)

summary(pop)

sd(pop)

summary(gdppc)

sd(gdppc)

summary(roflaw)

sd(roflaw)

# Scale natrent as 1 unit increase be increase of $10 billlion

mydata$natrent = natrent / 10000000000

natrent = mydata$natrent

# Scale pop as 1 unit increase be increase of 10 million people

mydata$pop = pop / 10000000

pop = mydata$pop

# ==========================================================================================================================

# QUESTION 2

# ----------

library(ggplot2)

# Scatter plot of numbil0 against natrent

pdf("numbil0 vs natrent.pdf")

ggplot(mydata, aes(y=numbil0, x=natrent)) + # Define the dataset, x and y variables for scatter plot

geom\_point(alpha = .3) + # Allow for shading of the points in the scatter plot to help visualisation

stat\_smooth(method = "lm", formula = y ~ poly(x,1), col="blue") + # Fit a polynomial of DEGREE 2 (QUADRATIC)

ggtitle("Relationship Between numbil0 and natrent") + # Scatter plot title

theme(plot.title = element\_text(hjust = 0.5)) # Center the scatter plot title

dev.off()

# Scatter plot of numbil0 against pop

pdf("numbil0 vs pop.pdf")

ggplot(mydata, aes(y=numbil0, x=pop)) + # Define the dataset, x and y variables for scatter plot

geom\_point(alpha = .3) + # Allow for shading of the points in the scatter plot to help visualisation

stat\_smooth(method = "lm", formula = y ~ poly(x,1), col="blue") + # Fit a polynomial of DEGREE 2 (QUADRATIC)

ggtitle("Relationship Between numbil0 and pop") + # Scatter plot title

theme(plot.title = element\_text(hjust = 0.5)) # Center the scatter plot title

dev.off()

# Scatter plot of natrent against pop

pdf("natrent vs pop.pdf")

ggplot(mydata, aes(y=natrent, x=pop)) + # Define the dataset, x and y variables for scatter plot

geom\_point(alpha = .3) + # Allow for shading of the points in the scatter plot to help visualisation

stat\_smooth(method = "lm", formula = y ~ poly(x,1), col="blue") + # Fit a polynomial of DEGREE 2 (QUADRATIC)

ggtitle("Relationship Between natrent and pop") + # Scatter plot title

theme(plot.title = element\_text(hjust = 0.5)) # Center the scatter plot title

dev.off()

# ==========================================================================================================================

# QUESTION 3

# ----------

# Scatter plot of numbil0 against gdppc

pdf("numbil0 vs gdppc.pdf")

ggplot(mydata, aes(y=numbil0, x=gdppc)) + # Define the dataset, x and y variables for scatter plot

geom\_point(alpha = .3) + # Allow for shading of the points in the scatter plot to help visualisation

stat\_smooth(method = "lm", formula = y ~ poly(x,1), col="blue") + # Fit a polynomial of DEGREE 2 (QUADRATIC)

ggtitle("Relationship Between numbil0 and gdppc") + # Scatter plot title

theme(plot.title = element\_text(hjust = 0.5)) # Center the scatter plot title

dev.off()

# Scatter plot of numbil0 against roflaw

pdf("numbil0 vs roflaw.pdf")

ggplot(mydata, aes(y=numbil0, x=roflaw)) + # Define the dataset, x and y variables for scatter plot

geom\_point(alpha = .3) + # Allow for shading of the points in the scatter plot to help visualisation

stat\_smooth(method = "lm", formula = y ~ poly(x,1), col="blue") + # Fit a polynomial of DEGREE 2 (QUADRATIC)

ggtitle("Relationship Between numbil0 and roflaw") + # Scatter plot title

theme(plot.title = element\_text(hjust = 0.5)) # Center the scatter plot title

dev.off()

# Scatter plot of gdppc against roflaw

pdf("gdppc vs roflaw.pdf")

ggplot(mydata, aes(y=gdppc, x=roflaw)) + # Define the dataset, x and y variables for scatter plot

geom\_point(alpha = .3) + # Allow for shading of the points in the scatter plot to help visualisation

stat\_smooth(method = "lm", formula = y ~ poly(x,1), col="blue") + # Fit a polynomial of DEGREE 2 (QUADRATIC)

ggtitle("Relationship Between gdppc and roflaw") + # Scatter plot title

theme(plot.title = element\_text(hjust = 0.5)) # Center the scatter plot title

dev.off()

# ==========================================================================================================================

# QUESTION 4

# ----------

# Make dummy variables for each of the year from 2005 to 2013 so we can

# avoid the dummy variable trap by making (omitting) 2005 as a base year

d2005=as.numeric(mydata$year==2005)

d2006=as.numeric(mydata$year==2006)

d2007=as.numeric(mydata$year==2007)

d2008=as.numeric(mydata$year==2008)

d2009=as.numeric(mydata$year==2009)

d2010=as.numeric(mydata$year==2010)

d2011=as.numeric(mydata$year==2011)

d2012=as.numeric(mydata$year==2012)

d2013=as.numeric(mydata$year==2013)

# Single linear regression model of the number of billionaires in a country (regressand) based on

# the total natural resources rent (regressor) and save the heteroskedastic-robust standard errors

reg1 = lm(numbil0 ~ natrent, data = mydata)

ct1 = coeftest(reg1, vcov = vcovHC(reg1, "HC1"))

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Total natural resources rent (natrent)

# • Population (pop)

reg2 = lm(numbil0 ~ natrent + pop, data = mydata)

ct2 = coeftest(reg2, vcov = vcovHC(reg2, "HC1"))

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Total natural resources rent (natrent)

# • Population (pop)

# • GDP per capita (gdppc)

reg3 = lm(numbil0 ~ natrent + pop + gdppc, data = mydata)

ct3 = coeftest(reg3, vcov = vcovHC(reg3, "HC1"))

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Total natural resources rent (natrent)

# • Population (pop)

# • GDP per capita (gdppc)

# • Rule of law index (roflaw)

reg4 = lm(numbil0 ~ natrent + pop + gdppc + roflaw, data = mydata)

ct4 = coeftest(reg4, vcov = vcovHC(reg4, "HC1"))

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Total natural resources rent (natrent)

# • Population (pop)

# • GDP per capita (gdppc)

# • Rule of law index (roflaw)

# • Year dummy variables (d20xx)

reg5 = lm(numbil0 ~ natrent + pop + gdppc + roflaw + d2006 + d2007 + d2008 + d2009 + d2010 + d2011 + d2012 + d2013, data = mydata)

ct5 = coeftest(reg5, vcov = vcovHC(reg5, "HC1"))

# Footnote to explain the table of reports (below)

custom\_note = "Dependent variable is the number of billionaires in the country.

natrent is natural resources rents in billions of dollars.

pop is total population of a country in millions.

gdppc is GDP per capita in thousands of dollars.

roflaw is an index between 0 and 1 on the level of the rule of law in a country.

dXXXX is a dummy variable for the year XXXX.

Heteroskedasticity robust standard errors in parentheses.

Statistical significance from two-sided tests of the null of no effect marked as \* for 5% and \*\* for 1%"

library(texreg)

# Print the coefficients of all the linear regression models (compiled in a list using the list() function) in

# a table form using the texreg() function where each column is numbered from (1) through (5) (assigned to

# 'custom.model.names') as to allow comparison between linear regression models as to help determine omitted

# variable biasness as well as performing two-sided t-tests on the coeffients at a 5% and 1% significance level

# using stars to denote the rejection of null of the coefficient being equal to 0 using stars (assigned to 'stars'),

# reporting the goodness of fit statistics at the bottom such as the number of observations, adjusted R^2, and

# the regression F-statistic (assigned to 'include.fstatistic') while not including the rmse (assigned to 'include.rmse')

# and the R^2 (assigned to 'include.rsquared') with the footnote at the bottom of the table (assigned to 'custom.note')

# and the table title (assigned to 'caption' and 'caption.above') and put them all in a file (assigned to 'file')

# then override the homoskedastic standard errors and p-values with the heteroskedastic-robust standard errors (assigned to

# 'override.se' and 'override.pvalues')

htmlreg(list(reg1, reg2, reg3, reg4, reg5),

file = "Billionaires and country characteristics.doc",

caption = "Billionaires and country characteristics",

caption.above = TRUE,

stars = c(0.01, 0.05),

custom.note = custom\_note,

include.rsquared = FALSE,

include.rmse = FALSE,

include.fstatistic=TRUE,

override.se = list(ct1[,2], ct2[,2], ct3[,2], ct4[,2], ct5[,2]),

override.pvalues = list(ct1[,4], ct2[,4], ct3[,4], ct4[,4], ct5[,4]),

reorder.coef = c(2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 1),

custom.model.names = c("(1)", "(2)", "(3)", "(4)","(5)"),

custom.coef.names = c("Constant", "natrent", "pop", "gdppc", "roflaw", "d2006", "d2007", "d2008", "d2009", "d2010", "d2011", "d2012", "d2013"),

custom.gof.names = c("adj. R^2", "N", "F"),

reorder.gof = c(2, 1, 3))

# ==========================================================================================================================

# QUESTION 5

# ----------

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Population (pop)

# • Total natural resources rent (natrent)

# • GDP per capita (gdppc)

# • Rule of law index (roflaw)

reg6 = lm(numbil0 ~ log(pop) + natrent + gdppc + roflaw, data = mydata)

ct6 = coeftest(reg6, vcov = vcovHC(reg6, "HC1"))

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Population (pop)

# • Total natural resources rent (natrent)

# • GDP per capita (gdppc)

# • Rule of law index (roflaw)

reg7 = lm(log(numbil0) ~ pop + natrent + gdppc + roflaw, data = mydata)

ct7 = coeftest(reg7, vcov = vcovHC(reg7, "HC1"))

# Multiple linear regression model of the number of billionaires in a country (regressand) based on (regressors):

# • Population (pop)

# • Total natural resources rent (natrent)

# • GDP per capita (gdppc)

# • Rule of law index (roflaw)

reg8 = lm(log(numbil0) ~ log(pop) + natrent + gdppc + roflaw, data = mydata)

ct8 = coeftest(reg8, vcov = vcovHC(reg8, "HC1"))

# Footnote to explain the table of reports (below)

custom\_note = "numbil0 is the number of billionaires in the country.

natrent is natural resources rents in billions of dollars.

pop is total population of a country in millions.

gdppc is GDP per capita in thousands of dollars.

roflaw is an index between 0 and 1 on the level of the rule of law in a country.

Heteroskedasticity robust standard errors in parentheses.

Statistical significance from two-sided tests of the null of no effect marked as \* for 5% and \*\* for 1%"

# Print the coefficients of all the linear regression models (compiled in a list using the list() function) in

# a table form using the texreg() function where each column is numbered from (1) through (3) (assigned to

# 'custom.model.names') as to allow comparison between linear regression models as to help determine omitted

# variable biasness as well as performing two-sided t-tests on the coeffients at a 5% and 1% significance level

# using stars to denote the rejection of null of the coefficient being equal to 0 using stars (assigned to 'stars'),

# reporting the goodness of fit statistics at the bottom such as the number of observations, adjusted R^2, and

# the regression F-statistic (assigned to 'include.fstatistic') while not including the rmse (assigned to 'include.rmse')

# and the R^2 (assigned to 'include.rsquared') with the footnote at the bottom of the table (assigned to 'custom.note')

# and the table title (assigned to 'caption' and 'caption.above') and put them all in a file (assigned to 'file')

# then override the homoskedastic standard errors and p-values with the heteroskedastic-robust standard errors (assigned to

# 'override.se' and 'override.pvalues')

htmlreg(list(reg6, reg7, reg8),

file = "Billionaires and population size.doc",

caption = "Billionaires and population size",

caption.above = TRUE,

stars = c(0.01, 0.05),

custom.note = custom\_note,

include.rsquared = FALSE,

include.rmse = FALSE,

include.fstatistic=TRUE,

override.se = list(ct6[,2], ct7[,2], ct8[,2]),

override.pvalues = list(ct6[,4], ct7[,4], ct8[,4]),

reorder.coef = c(2, 6, 3, 4, 5, 1),

custom.model.names = c("(1) numbil0", "(2) log(numbil0)", "(3) log(numbil0)"),

custom.coef.names = c("Constant", "lnpop", "natrent", "gdppc", "roflaw", "pop"),

custom.gof.names = c("adj. R^2", "N", "F"),

reorder.gof = c(2, 1, 3))

# ==========================================================================================================================

# QUESTION 6

# ----------

# Create the interaction variables between log(pop) and each of the dummy year variables

lnpop\_d2005 = log(pop) \* d2005

lnpop\_d2006 = log(pop) \* d2006

lnpop\_d2007 = log(pop) \* d2007

lnpop\_d2008 = log(pop) \* d2008

lnpop\_d2009 = log(pop) \* d2009

lnpop\_d2010 = log(pop) \* d2010

lnpop\_d2011 = log(pop) \* d2011

lnpop\_d2012 = log(pop) \* d2012

lnpop\_d2013 = log(pop) \* d2013

# Multiple linear regression model of the log of the number of billionaires in a country (regressand) based on (regressors):

# • Log of population (log(pop))

# • Interaction terms between log(pop) and each of the dummy year variables with 2005 as the base year (omitted) (lnpop\_d20xx)

# • Total natural resources rent (natrent)

# • GDP per capita (gdppc)

# • Rule of law index (roflaw)

# • Dummy year variables with 2005 as the base year (omitted) (d20xx)

reg9 = lm(log(numbil0) ~ log(pop) + lnpop\_d2006 + lnpop\_d2007 + lnpop\_d2008 + lnpop\_d2009 + lnpop\_d2010 + lnpop\_d2011 + lnpop\_d2012 + lnpop\_d2013 + natrent + gdppc + roflaw + d2006 + d2007 + d2008 + d2009 + d2010 + d2011 + d2012 + d2013, data = mydata)

coeftest(reg9, vcov = vcovHC(reg9, "HC1"))

summary(reg9)$adj.r.squared # Report the adjusted R squared

nobs(reg9) # Report the number of observations used

# ==========================================================================================================================

# QUESTION 7

# ----------

# Linear hypothesis test of the above regression with the interaction terms set to 0 to

# test the significance of the partial effects on the elasticity of numbil0 with respect to pop

linearHypothesis(reg9,c("lnpop\_d2006 = 0", "lnpop\_d2007 = 0", "lnpop\_d2008 = 0", "lnpop\_d2009 = 0", "lnpop\_d2010 = 0", "lnpop\_d2011 = 0", "lnpop\_d2012 = 0", "lnpop\_d2013 = 0"), vcov = vcovHC(reg9, "HC1"))

# ==========================================================================================================================

# QUESTION 8

# ----------

# Create interaction variable between log(gdppc) and roflaw

log\_gdppc\_roflaw = log(gdppc) \* roflaw

# Multiple linear regression model of the log of the number of billionaires in a country (regressand) based on (regressors):

# • Log of GDP per capita (log(gdppc))

# • Rule of law index (roflaw)

# • Interaction term between log of GDP per capita and the rule of law index (log\_gdppc\_roflaw)

# • Log of population (log(pop))

# • Total natural resources rent (natrent)

reg10 = lm(log(numbil0) ~ log(gdppc) + roflaw + log\_gdppc\_roflaw + log(pop) + natrent, data = mydata)

coeftest(reg10, vcov = vcovHC(reg10, "HC1"))

summary(reg10)$adj.r.squared # Report the adjusted R squared

nobs(reg10) # Report the number of observations used

# ==========================================================================================================================

# QUESTION 9

# ----------

# Create variables to hold the coefficients of the log(gdppc) and the interaction

# term of log(gdppc) and roflaw that can be retrieved using the summary() function

coef\_log\_gdppc = summary(reg10)$coefficients[2,1]

coef\_log\_gdppc\_roflaw = summary(reg10)$coefficients[4,1] \* 0.1 # Change the 0.1 to test other roflaw values

# Add the two variables with respect to the given roflaw value to get the elasticity

elasticity = coef\_log\_gdppc + coef\_log\_gdppc\_roflaw

elasticity

# Linear hypothesis test of the above regression with the log(gdppc) and the interaction term set to 0 to test the

# significance of the partial effects on the elasticity of numbil0 with respect to gdppc depending on the roflaw value

linearHypothesis(reg10,c("log(gdppc)+ 0.1 \* log\_gdppc\_roflaw=0"),vcov = vcovHC(reg10,"HC1"))

# ==========================================================================================================================