Afternoon Exam 2 - Part 2

Econ 329 Spring 2018

Part 2 (100 points) is an open-note and open-book portion that will require you to use a computer to perform econometric calculations. You will be graded on your ability to generate the proper calculations and interpret them. You will be allowed to use resources like your notes and textbook as well as search for resources on the internet (for instance, if you forget how to accomplish a certain task with the computer - you can Google it!).

You are not allowed to ask anyone for help. This means, but is not limited to, talking, emailing, texting, calling, chatting, or posting questions to the Internet.

You are not allowed to offer help to anyone. This means, but is not limited to, talking, emailing, texting, calling, chatting, or posting answers/information to the Internet or anywhere else that someone could find them.

You should know that any and all suspected violations of this policy will be pursued with the College's Academic Standing Committee via the Associate Dean of Students or the Associate Dean of the College. A formal finding of responsibility can result in disciplinary sanctions including grade of zero in the exam, to a F in this course, to permanent dismissal from the College in the case of repeated and serious offenses.

We will use the `CO2emissionsdata.xlsx` data file for Part 2 of the exam. The `CO2emissionsdata.xlsx` file contains country-level data about Carbon Dioxide emissions per capita (in tons), GDP per capita (in \$), population, life expectancy (in years), the year of observation, the country name, and the continent of the country.

Each question is worth 10 points for a total of 100 points in Part 2.

```r
library(readxl)
examData <- read_excel("~/Documents/2017-18/ECON329/notes/CO2emissionsdata.xlsx")
```

1. Make a table of summary statistics (min, max, mean, median, sd) for the GDP per capita and CO2 emissions per capita variables.

```r
suppressPackageStartupMessages(library(stargazer))
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
stargazer(data.frame(examData %>% select(gdpPercap, CO2percapita)), type = "text", median = TRUE)
```

```
##
## ====================================================================
## Statistic     N    Mean     St. Dev.    Min    Median      Max
## --------------------------------------------------------------------
## gdpPercap    534 9,843.880 11,104.500 241.166 5,237.045 49,357.190
## CO2percapita 534   4.068     5.253     0.002    1.807     35.044
## --------------------------------------------------------------------
```

2. Transform the `gdpPercap` variable so that it is measured in \$1,000s rather than just \$. Update your table of summary statistics and use this transformed variable for the remainder of these questions.

```
examData <- examData %>% mutate(gdpPercap_1000 = gdpPercap/1000)
stargazer(data.frame(examData %>% select(gdpPercap_1000, CO2percapita)), type = "text", median = TRUE)
```

```
##
## =====================================================
## Statistic       N  Mean  St. Dev.  Min  Median  Max
## -----------------------------------------------------
## gdpPercap_1000 534 9.844  11.104  0.241 5.237  49.357
## CO2percapita   534 4.068  5.253   0.002 1.807  35.044
## -----------------------------------------------------
```

3. Estimate and interpret a linear regression of CO2 emissions per capita as a function of GDP per capita.

```
exam.lm1 <- lm(CO2percapita ~ gdpPercap_1000, data = examData)

stargazer(exam.lm1, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                            CO2percapita
## -----------------------------------------------
## gdpPercap_1000               0.365***
##                               (0.013)
##
## Constant                      0.474**
##                               (0.193)
##
## -----------------------------------------------
## Observations                    534
## R2                             0.596
## Adjusted R2                    0.595
## Residual Std. Error      3.344 (df = 532)
## F Statistic          783.873*** (df = 1; 532)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

CO2 emissions per capita is expected to increase about 0.365 tons for every one thousand dollars increase in GDP per capita of a country. The standard error of 0.013 implies that this coefficient is significantly different from zero (our estimate is over 28 standard errors away from zero). The constant term implies that our model predicts that the expected CO2 emissions per capita is 0.474 tons if GDP per capita of the country is 0. The standard error of 0.193 for the constant term implies that this coefficient is 2.5 standard errors away from zero, so there is a significant evidence that the constant term is not zero.

4. Estimate and interpret a log-linear regression of CO2 emissions per capita as a function of GDP per capita.

```
exam.lm2 <- lm(log(CO2percapita) ~ log(gdpPercap_1000), data = examData)
stargazer(exam.lm2, type = "text")
```

```
##
## ===============================================
##                         Dependent variable:
##                     ---------------------------
```

```
##                              log(CO2percapita)
## -----------------------------------------------
## log(gdpPercap_1000)          1.239***
##                               (0.025)
##
## Constant                     -1.608***
##                               (0.051)
##
## -----------------------------------------------
## Observations                    534
## R2                             0.822
## Adjusted R2                    0.822
## Residual Std. Error      0.764 (df = 532)
## F Statistic         2,459.850*** (df = 1; 532)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

CO2 emissions per capita is expected to increase about 1.239 percent for every 1 percent increase in GDP per capita of a country. The standard error of 0.025 implies that this coefficient is significantly different from zero (our estimate is over 49 standard errors away from zero). The constant term implies that our model predicts that the expected CO2 emissions per capita is -1.608 tons if GDP per capita of the country is $1000 (log(gdpPercap_1000) = 0, so gdpPercap_1000 = 1). The standard error of 0.051 for the constant term implies that this coefficient is 31 standard errors away from zero, so there is a significant evidence that the constant term is not zero.

5. What is the predicted change in CO2 emissions per capita from an extra $1,000 in GDP per capita for a country with current GDP per capita of $10,000 for each of your two models?

3.65 tons increase in CO2 emission per cepita for linear model.
12.39 percent increase in CO2 emssion per capita for log-linear model.

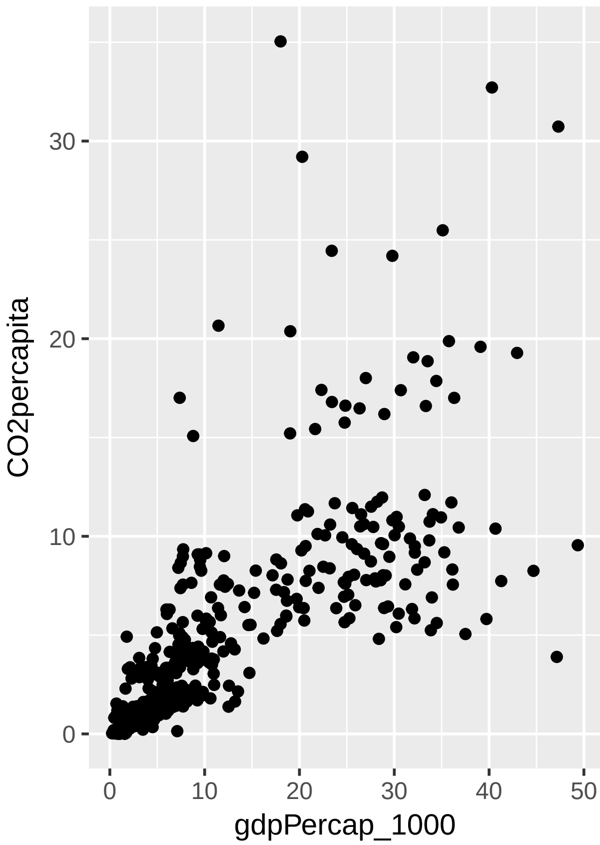
6. Generate two plots of the data with CO2 emissions per capita on the vertical axis and GDP per capita on the horizontal axis. Use linear scales on one graph and log scales on the other.

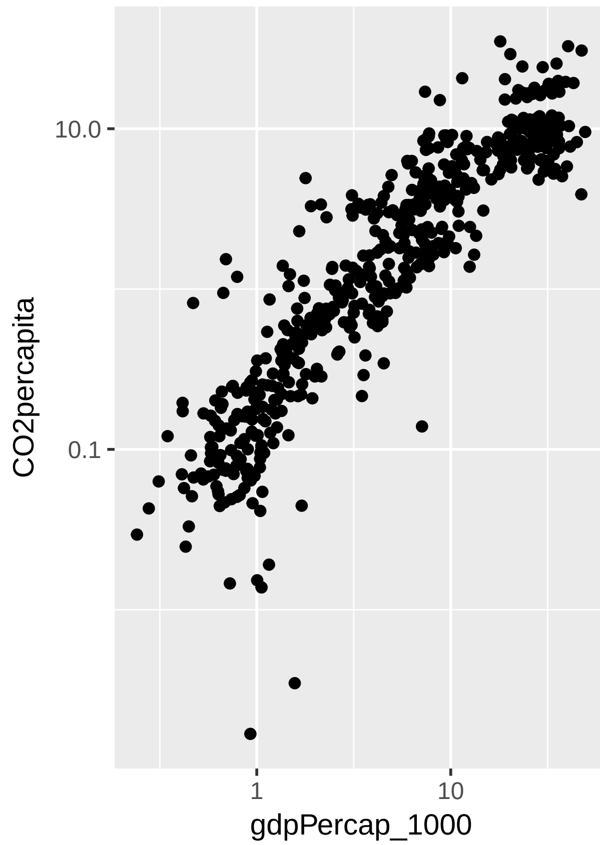
```r
library(ggplot2)
suppressMessages(library(gridExtra))

linear <- examData %>% ggplot(aes(x = gdpPercap_1000, y = CO2percapita)) +
  geom_point()

log <- examData %>% ggplot(aes(x = gdpPercap_1000, y = CO2percapita)) +
  geom_point() +
  scale_x_log10()+
  scale_y_log10()

grid.arrange(linear, log, ncol=2)
```





7. Restrict your sample to only observations in the Americas and Europe. Does it appear that the relationship between GDP per capita and CO2 emissions is the same across these two continents? Perform an appropriate hypothesis test to justify your answer. If there seem to be differences, what are they?

```
europe_america <- examData %>% filter(continent %in% c("Americas","Europe"))
exam.lm3 <- lm(log(CO2percapita) ~ log(gdpPercap_1000)*continent, data = europe_america)
stargazer(exam.lm3, type = "text")
```

```
##
## ================================================================
##                                         Dependent variable:
##                                     ---------------------------
##                                          log(CO2percapita)
## ----------------------------------------------------------------
## log(gdpPercap_1000)                          1.406***
##                                               (0.063)
##
## continentEurope                              2.110***
##                                               (0.217)
##
## log(gdpPercap_1000):continentEurope          -0.807***
##                                               (0.086)
##
## Constant                                     -1.966***
##                                               (0.129)
##
## ----------------------------------------------------------------
```

```
## Observations                                        208
## R2                                                0.820
## Adjusted R2                                       0.818
## Residual Std. Error                         0.434 (df = 204)
## F Statistic                           310.409*** (df = 3; 204)
## ================================================================
## Note:                                 *p<0.1; **p<0.05; ***p<0.01
```

```r
library(car)
```

```
##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

```r
linearHypothesis(exam.lm3, c("continentEurope = 0", "log(gdpPercap_1000):continentEurope = 0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## continentEurope = 0
## log(gdpPercap_1000):continentEurope = 0
##
## Model 1: restricted model
## Model 2: log(CO2percapita) ~ log(gdpPercap_1000) * continent
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    206 56.299
## 2    204 38.385  2    17.915 47.605 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the hypothesis test above (ANOVA test), we can seem that there is a significant difference in relationship between GDP per capita and CO2 emissions depending on the continents (p-value < 2e-16). The coefficients tell us that the positive relationship between GDP per capita and CO2 are smaller in European countries (1.406-0.807 = 0.599) (all else equal), while the estimated CO2 emission of European countries are bigger than that of America (0.114 tones higher) when GDP per capita of a country is $1000 (all else equal).

8. Re-estimate the log-linear model (from #4) while also including a "$(log(gdpPercap))^2$" term in the regression (that is, estimate the log of CO2 emissions per capita as a quadratic function of the log of GDP per capita). Present and interpret the results.

```r
log_gdpPercap_2 <- (log(examData$gdpPercap_1000))^2
exam.lm4 <- lm(log(CO2percapita) ~ log(gdpPercap_1000) + log_gdpPercap_2, data = examData)
stargazer(exam.lm4, type = "text")
```

```
##
## ================================================
##                         Dependent variable:
##                     ---------------------------
##                          log(CO2percapita)
## ------------------------------------------------
## log(gdpPercap_1000)          1.568***
##                               (0.064)
##
## log_gdpPercap_2              -0.110***
```

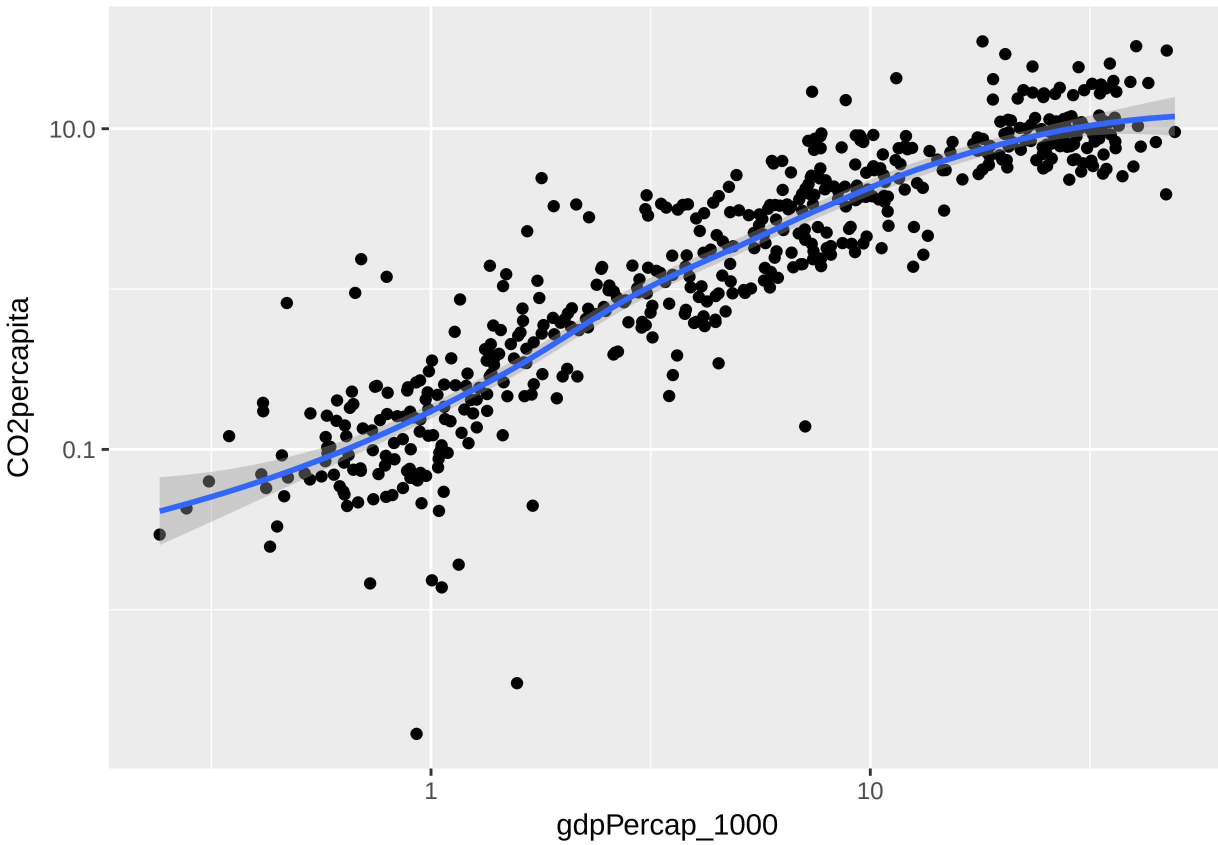
```
##                                (0.020)
##
## Constant                       -1.661***
##                                (0.050)
##
## -----------------------------------------------
## Observations                     534
## R2                              0.832
## Adjusted R2                     0.831
## Residual Std. Error       0.743 (df = 531)
## F Statistic         1,314.643*** (df = 2; 531)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

The positive sign of $log(gdpPercap)$ means that there is a positive relationship between GDP per capita and CO2 emissions per capita, but the negative $(log(gdpPercap))^2$ term means that the positive effect between GDP per capita and CO2 emission per capita only exists until at the certain GDP per capita and that effect starts to decrease after hitting that point.

9. Some economists hypothesize that there is an "upside down u" shaped relationship between income and various pollutants. Do your results (from #8) seem to conform with the idea an upside down u shaped relationship with these variables? Why or why not?

In that sign of the term for $(log(gdpPercap))^2$ is negative, we can confirm there there is an "upside down u" shaped relationship between income and CO2 emissions.

```r
examData %>% ggplot(aes(x = gdpPercap_1000, y = CO2percapita)) +
  geom_point() +
  scale_x_log10()+
  scale_y_log10() +
  geom_smooth(method = "loess")
```



From the plot above we can observe the positive effect seems to decrease as gdp per capita gets larger (left side of the u-shaped parabola)

10. Use your results from #8 to calculate the "turning point" in your regression model (the point at which further income leads to reductions in emissions rather than increases) for this model. Report the value in $.

For the x value (logGdpPercap_1000) of vertex of parabobola = -b/2a if y = ax^2 + bx + c

```
x <- -1.568/(2*-0.110)
x
```

```
## [1] 7.127273
```

GDP per capita = exp(x) = $1,245,476