Project\_Environment\_data

Frances Lin

5/2019

## I. INTRODUCTION

## II. DATA DESCRIPTION

## III. SUMMARY STATISTICS AND GRAPHICS

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

library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

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

library(GGally)

## Loading required package: ggplot2

##   
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':  
##   
## nasa

library(ggplot2)

# Load data   
WBData <- read.csv("https://raw.githubusercontent.com/franceslinyc/Regression-Analysis-of-Environment-data-in-R-2019/master/worldbank.csv")  
str(WBData)

## 'data.frame': 264 obs. of 6 variables:  
## $ Country.Code : Factor w/ 264 levels "ABW","AFG","AGO",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ CO2.per.capita : num 8.408 0.299 1.291 1.979 5.832 ...  
## $ GDP.per.capita : num 25528 625 5413 4579 42295 ...  
## $ Energy.use.per.capita : num NA NA 545 808 NA ...  
## $ Electric.power.per.capita: num NA NA 312 2309 NA ...  
## $ GNI.per.capita : num 24500 640 5010 4540 NA ...

#View(WBData)  
  
# Attach data set   
attach(WBData)  
  
# Log Transform all Y and Xs   
logCO2.per.capita <- log(CO2.per.capita)  
logGDP.per.capita <- log(GDP.per.capita)  
logEnergy.use.per.capita <- log(Energy.use.per.capita)  
logElectric.power.per.capita <- log(Electric.power.per.capita)  
logGNI.per.capita <- log(GNI.per.capita)  
  
# Create a new data frame: log   
logWBData <- data.frame(logCO2.per.capita, logGDP.per.capita,   
 logEnergy.use.per.capita, logElectric.power.per.capita,   
 logGNI.per.capita)   
str(logWBData)

## 'data.frame': 264 obs. of 5 variables:  
## $ logCO2.per.capita : num 2.129 -1.206 0.256 0.682 1.763 ...  
## $ logGDP.per.capita : num 10.15 6.44 8.6 8.43 10.65 ...  
## $ logEnergy.use.per.capita : num NA NA 6.3 6.7 NA ...  
## $ logElectric.power.per.capita: num NA NA 5.74 7.74 NA ...  
## $ logGNI.per.capita : num 10.11 6.46 8.52 8.42 NA ...

#View(logWBData)  
  
# Attach data set   
attach(logWBData)

## The following objects are masked \_by\_ .GlobalEnv:  
##   
## logCO2.per.capita, logElectric.power.per.capita,  
## logEnergy.use.per.capita, logGDP.per.capita, logGNI.per.capita

# Descriptive Statistics: original scale, log   
summary(CO2.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.04449 0.88165 3.15330 4.87489 6.36518 45.42324 14

summary(GDP.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 273.5 2036.9 6429.0 16261.6 17061.3 185152.5 15

summary(Energy.use.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 60.73 682.88 1538.26 2489.09 2995.52 18562.67 89

summary(Electric.power.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 38.97 776.77 2514.42 3965.61 5112.80 53832.48 82

summary(WBData$GNI.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 260 1888 6150 13786 14678 104540 28

summary(logCO2.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -3.1126 -0.1260 1.1480 0.8144 1.8508 3.8160 14

summary(logGDP.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 5.611 7.619 8.769 8.778 9.745 12.129 15

summary(logEnergy.use.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 4.106 6.526 7.338 7.342 8.005 9.829 89

summary(logElectric.power.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 3.663 6.655 7.830 7.561 8.539 10.894 82

summary(logGNI.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 5.561 7.543 8.724 8.686 9.594 11.557 28

# Pairwise Scatterplots using ggplot2  
ggpairs(WBData[, -1]) + ggtitle("Pairwise Scatterplots")

## Warning: Removed 14 rows containing non-finite values (stat\_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 90 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 83 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 32 rows containing missing values

## Warning: Removed 25 rows containing missing values (geom\_point).

## Warning: Removed 15 rows containing non-finite values (stat\_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 93 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 86 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 28 rows containing missing values

## Warning: Removed 90 rows containing missing values (geom\_point).

## Warning: Removed 93 rows containing missing values (geom\_point).

## Warning: Removed 89 rows containing non-finite values (stat\_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 89 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 94 rows containing missing values

## Warning: Removed 83 rows containing missing values (geom\_point).

## Warning: Removed 86 rows containing missing values (geom\_point).

## Warning: Removed 89 rows containing missing values (geom\_point).

## Warning: Removed 82 rows containing non-finite values (stat\_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 87 rows containing missing values

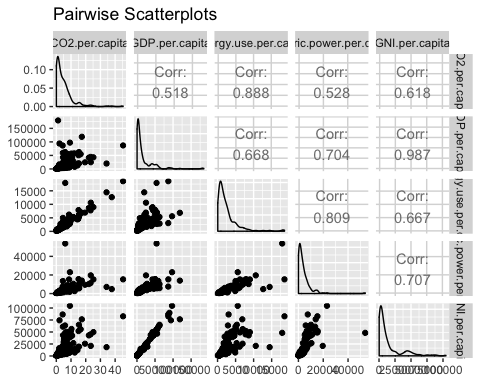
## Warning: Removed 32 rows containing missing values (geom\_point).

## Warning: Removed 28 rows containing missing values (geom\_point).

## Warning: Removed 94 rows containing missing values (geom\_point).

## Warning: Removed 87 rows containing missing values (geom\_point).

## Warning: Removed 28 rows containing non-finite values (stat\_density).



# Pairwise Scatterplots using ggplot2: log  
ggpairs(logWBData) + ggtitle("Pairwise Scatterplots: Log")

## Warning: Removed 14 rows containing non-finite values (stat\_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 25 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 90 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 83 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 32 rows containing missing values

## Warning: Removed 25 rows containing missing values (geom\_point).

## Warning: Removed 15 rows containing non-finite values (stat\_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 93 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 86 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 28 rows containing missing values

## Warning: Removed 90 rows containing missing values (geom\_point).

## Warning: Removed 93 rows containing missing values (geom\_point).

## Warning: Removed 89 rows containing non-finite values (stat\_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 89 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 94 rows containing missing values

## Warning: Removed 83 rows containing missing values (geom\_point).

## Warning: Removed 86 rows containing missing values (geom\_point).

## Warning: Removed 89 rows containing missing values (geom\_point).

## Warning: Removed 82 rows containing non-finite values (stat\_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =  
## "pearson", : Removed 87 rows containing missing values

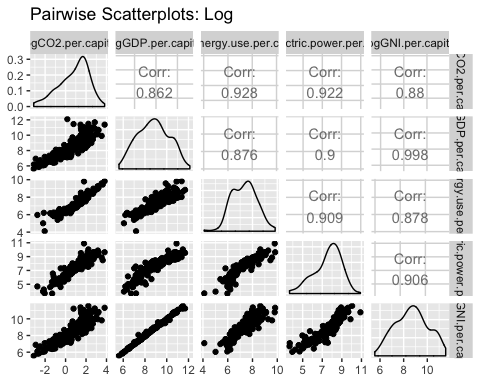
## Warning: Removed 32 rows containing missing values (geom\_point).

## Warning: Removed 28 rows containing missing values (geom\_point).

## Warning: Removed 94 rows containing missing values (geom\_point).

## Warning: Removed 87 rows containing missing values (geom\_point).

## Warning: Removed 28 rows containing non-finite values (stat\_density).



## IV. ANALYSIS

## V. RESULTS AND CONCLUSIONS

# Part I analysis   
# Fit multiple linear regression model: log, w/ interaction   
lmMInteract <- lm(logCO2.per.capita~logGDP.per.capita\*  
 logEnergy.use.per.capita\*logElectric.power.per.capita)  
summary(lmMInteract)

##   
## Call:  
## lm(formula = logCO2.per.capita ~ logGDP.per.capita \* logEnergy.use.per.capita \*   
## logElectric.power.per.capita)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.06249 -0.14786 0.01945 0.19081 0.90025   
##   
## Coefficients:  
## Estimate  
## (Intercept) 10.02204  
## logGDP.per.capita -1.80515  
## logEnergy.use.per.capita -4.14913  
## logElectric.power.per.capita -0.51888  
## logGDP.per.capita:logEnergy.use.per.capita 0.56539  
## logGDP.per.capita:logElectric.power.per.capita 0.06119  
## logEnergy.use.per.capita:logElectric.power.per.capita 0.44001  
## logGDP.per.capita:logEnergy.use.per.capita:logElectric.power.per.capita -0.04805  
## Std. Error  
## (Intercept) 5.73797  
## logGDP.per.capita 0.76596  
## logEnergy.use.per.capita 0.93012  
## logElectric.power.per.capita 0.68630  
## logGDP.per.capita:logEnergy.use.per.capita 0.11867  
## logGDP.per.capita:logElectric.power.per.capita 0.08311  
## logEnergy.use.per.capita:logElectric.power.per.capita 0.09860  
## logGDP.per.capita:logEnergy.use.per.capita:logElectric.power.per.capita 0.01152  
## t value  
## (Intercept) 1.747  
## logGDP.per.capita -2.357  
## logEnergy.use.per.capita -4.461  
## logElectric.power.per.capita -0.756  
## logGDP.per.capita:logEnergy.use.per.capita 4.764  
## logGDP.per.capita:logElectric.power.per.capita 0.736  
## logEnergy.use.per.capita:logElectric.power.per.capita 4.462  
## logGDP.per.capita:logEnergy.use.per.capita:logElectric.power.per.capita -4.170  
## Pr(>|t|)  
## (Intercept) 0.0826  
## logGDP.per.capita 0.0196  
## logEnergy.use.per.capita 1.52e-05  
## logElectric.power.per.capita 0.4507  
## logGDP.per.capita:logEnergy.use.per.capita 4.18e-06  
## logGDP.per.capita:logElectric.power.per.capita 0.4627  
## logEnergy.use.per.capita:logElectric.power.per.capita 1.51e-05  
## logGDP.per.capita:logEnergy.use.per.capita:logElectric.power.per.capita 4.96e-05  
##   
## (Intercept) .   
## logGDP.per.capita \*   
## logEnergy.use.per.capita \*\*\*  
## logElectric.power.per.capita   
## logGDP.per.capita:logEnergy.use.per.capita \*\*\*  
## logGDP.per.capita:logElectric.power.per.capita   
## logEnergy.use.per.capita:logElectric.power.per.capita \*\*\*  
## logGDP.per.capita:logEnergy.use.per.capita:logElectric.power.per.capita \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2985 on 162 degrees of freedom  
## (94 observations deleted due to missingness)  
## Multiple R-squared: 0.9453, Adjusted R-squared: 0.9429   
## F-statistic: 399.9 on 7 and 162 DF, p-value: < 2.2e-16

#plot(lmMInteract)

# Part II analysis   
lmEKC1Log <- lm(logCO2.per.capita~logGDP.per.capita)  
lmEKC2Log <- lm(logCO2.per.capita~logGDP.per.capita+I(logGDP.per.capita^2))  
lmEKC3Log <- lm(logCO2.per.capita~logGDP.per.capita+I(logGDP.per.capita^2)++I(logGDP.per.capita^3))  
summary(lmEKC1Log)

##   
## Call:  
## lm(formula = logCO2.per.capita ~ logGDP.per.capita)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5688 -0.4138 -0.0549 0.4938 1.7258   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.9441 0.2977 -23.32 <2e-16 \*\*\*  
## logGDP.per.capita 0.8831 0.0337 26.21 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7295 on 237 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.7434, Adjusted R-squared: 0.7424   
## F-statistic: 686.8 on 1 and 237 DF, p-value: < 2.2e-16

summary(lmEKC2Log)

##   
## Call:  
## lm(formula = logCO2.per.capita ~ logGDP.per.capita + I(logGDP.per.capita^2))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.85561 -0.38919 0.00961 0.39823 1.62870   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -20.76289 1.37655 -15.08 <2e-16 \*\*\*  
## logGDP.per.capita 4.12514 0.31889 12.94 <2e-16 \*\*\*  
## I(logGDP.per.capita^2) -0.18528 0.01815 -10.21 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6089 on 236 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.822, Adjusted R-squared: 0.8205   
## F-statistic: 545 on 2 and 236 DF, p-value: < 2.2e-16

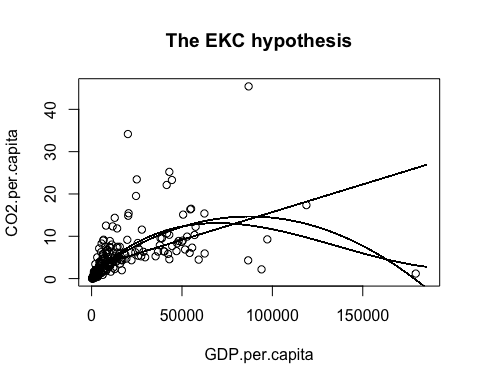
summary(lmEKC3Log)

##   
## Call:  
## lm(formula = logCO2.per.capita ~ logGDP.per.capita + I(logGDP.per.capita^2) +   
## +I(logGDP.per.capita^3))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.80582 -0.42280 -0.00259 0.39312 1.79842   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.91437 7.34058 -0.533 0.5944   
## logGDP.per.capita -1.87669 2.58871 -0.725 0.4692   
## I(logGDP.per.capita^2) 0.51313 0.29953 1.713 0.0880 .  
## I(logGDP.per.capita^3) -0.02658 0.01138 -2.336 0.0203 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6032 on 235 degrees of freedom  
## (25 observations deleted due to missingness)  
## Multiple R-squared: 0.826, Adjusted R-squared: 0.8238   
## F-statistic: 372 on 3 and 235 DF, p-value: < 2.2e-16

# Fit linear, quadratic, and cubic reg model: test the EKC hypotheis  
par(mfrow=c(1, 1))  
plot(CO2.per.capita~GDP.per.capita,   
 main="The EKC hypothesis")   
  
lmEKC1 <- lm(CO2.per.capita~GDP.per.capita)  
summary(GDP.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 273.5 2036.9 6429.0 16261.6 17061.3 185152.5 15

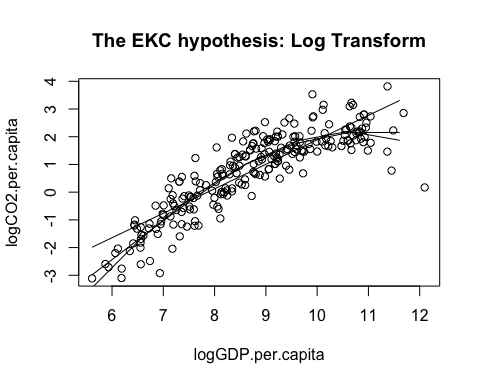
Xnew <- seq(from=273.5, to=185152.5)   
Yhat1 <- predict(lmEKC1, list(GDP.per.capita=Xnew))   
lines(Yhat1~Xnew)  
  
lmEKC2 <- lm(CO2.per.capita~GDP.per.capita+I(GDP.per.capita^2))  
Xnew <- seq(from=273.5, to=185152.5)   
Yhat2 <- predict(lmEKC2, list(GDP.per.capita=Xnew))   
lines(Yhat2~Xnew)  
  
lmEKC3 <- lm(CO2.per.capita~GDP.per.capita+I(GDP.per.capita^2)++I(GDP.per.capita^3))  
Xnew <- seq(from=273.5, to=185152.5)  
Yhat3 <- predict(lmEKC3, list(GDP.per.capita=Xnew))   
lines(Yhat3~Xnew)



#Fit linear, quadratic, and cubic reg model: test the EKC hypotheis, log   
par(mfrow=c(1, 1))  
plot(logCO2.per.capita~logGDP.per.capita,   
 main="The EKC hypothesis: Log Transform")   
  
lmEKC1Log<-lm(logCO2.per.capita~logGDP.per.capita)  
summary(logGDP.per.capita)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 5.611 7.619 8.769 8.778 9.745 12.129 15

Xnew <- seq(from=5.611, to=12.129)   
Yhat1 <- predict(lmEKC1Log, list(logGDP.per.capita=Xnew))   
lines(Yhat1~Xnew)  
  
lmEKC2Log<-lm(logCO2.per.capita~logGDP.per.capita+I(logGDP.per.capita^2))  
Xnew <- seq(from=5.611, to=12.129)   
Yhat2 <- predict(lmEKC2Log, list(logGDP.per.capita=Xnew))   
lines(Yhat2~Xnew)  
  
lmEKC3Log<-lm(logCO2.per.capita~logGDP.per.capita+I(logGDP.per.capita^2)++I(logGDP.per.capita^3))  
Xnew <- seq(from=5.611, to=12.129)   
Yhat3 <- predict(lmEKC3Log, list(logGDP.per.capita=Xnew))   
lines(Yhat3~Xnew)



## REFERENCES

## R CODE

## SUPPLEMENTS