Coal Production

Team-The Best

Tuesday, October 11, 2016

# Assignment:

Building off the data handling that we did in class, your group assignment is to estimate a supply or demand function, assuming no endogenaity, meaning that you can use OLS.

1. Look at the primer (<http://users.stat.umn.edu/~sandy/alr4ed/links/alrprimer.pdf>) if you need some help.
2. Convert the prices to real with the CPI. Again quandl has what you need.
3. Create a model, or multiple models, with the lm function. You can get as complicated as you like with the model, time trends, decades, log and square transformations.
4. Interpret each of the parameters in your model being very clear about units, e.g. a one dollar increase in the price per MBTU results in an ...

# Cleaning up the data:

## Load necessary Libraries

library(dplyr)  
library(stringr)  
library(Quandl)  
library(zoo)  
Quandl.api\_key("DzG-XQvJrL7g8BoTG8oA")

## Load necessary datatables (note: https would not work with knit)

Coal <-read.csv("http://www.eia.gov/totalenergy/data/browser/csv.cfm?tbl=T06.01")  
CoalPrice <-Quandl("EPI/152")  
CPI <-Quandl("RATEINF/CPI\_USA",collapse="annual")  
NaturalGasPrice<-Quandl("ODA/PNGASUS\_USD",collapse="annual")

## Prepare tables to be joined:

### 1) coal

Coal$Year<-substr(as.character(Coal$YYYYMM),1,4)  
Coal$Month<-substr(as.character(Coal$YYYYMM),5,6)  
CoalProduction<-subset(Coal, Coal$Month==13 & Coal$MSN=="CLPRPUS")  
ReducedCoalProd<-subset(CoalProduction,,c("Value" ,"Year"))  
names(ReducedCoalProd)<-c("CoalShortTons", "Year")  
ReducedCoalProd$CoalShortTons<-as.numeric(as.character(ReducedCoalProd$CoalShortTons))

### 2) coal price (I believe the data is in constant 2000 dollars and doesnt need adjusted)

CoalPrice$Year<-substr(as.character(CoalPrice$Year),1,4)

### 3) CPI

CPI$Year<-substr(as.character(CPI$Date),1,4)  
compyearValue<-CPI$Value[CPI$Year==2016]  
CPI$multiple<-compyearValue/CPI$Value  
ReducedCPI<-subset(CPI,,c( "Year", "multiple"))

### 4)Natural Gas ( The date had to be converted to 2016$ using CPI)

NaturalGasPrice$Year<-substr(as.character(NaturalGasPrice$Date),1,4)  
NaturalGasPrice<-inner\_join(NaturalGasPrice,ReducedCPI)

## Joining by: "Year"

NaturalGasPrice$price2016<-NaturalGasPrice$Value\*NaturalGasPrice$multiple  
NaturalGasPriceReduced<-subset(NaturalGasPrice,,c("Year", "price2016"))  
names(NaturalGasPriceReduced)<-c( "Year","NGprice2016" )

## Join the tables and add log columns:

### Coal (short Tons) vs. Coal Price (2010 $) Full Years

CoalProdCoalPrice<-inner\_join(ReducedCoalProd,CoalPrice)

## Joining by: "Year"

names(CoalProdCoalPrice)<-c( "CoalShortTons","Year","coalPrice2010")  
CoalProdCoalPrice$logCoalProd<-log(CoalProdCoalPrice$CoalShortTons)  
CoalProdCoalPrice$logCoalPrice<-log(CoalProdCoalPrice$coalPrice2010)  
CoalProdCoalPrice$DateFactor<-ifelse(CoalProdCoalPrice$Year>1975,1,0)  
CoalProdCoalPrice$DateFactor<-as.factor(CoalProdCoalPrice$DateFactor)

### Coal (short Tons) vs. Coal Price (2010 $) post 1975

CoalProdCoalPrice\_post1975<-subset(CoalProdCoalPrice,CoalProdCoalPrice$Year>1975)

### Coal (short Tons) vs. Natural Gas Price (2016 $)

CoalProdNGPrice<-inner\_join(ReducedCoalProd,NaturalGasPriceReduced)

## Joining by: "Year"

CoalProdNGPrice$logCoalProd<-log(CoalProdNGPrice$CoalShortTons)  
CoalProdNGPrice$logNGPrice<-log(CoalProdNGPrice$NGprice2016)

## Cleaned up Tables:

### Coal (short Tons) vs. Coal Price (2010 $)

CoalProdCoalPrice

## CoalShortTons Year coalPrice2010 logCoalProd logCoalPrice DateFactor  
## 1 480570.0 1949 32.05 13.08273 3.467297 0  
## 2 560388.0 1950 31.40 13.23638 3.446808 0  
## 3 576335.0 1951 29.86 13.26444 3.396520 0  
## 4 507424.0 1952 29.24 13.13710 3.375538 0  
## 5 488239.0 1953 28.67 13.09856 3.355851 0  
## 6 420789.0 1954 26.12 12.94989 3.262701 0  
## 7 490838.0 1955 25.02 13.10387 3.219676 0  
## 8 529774.0 1956 25.83 13.18021 3.251537 0  
## 9 518042.0 1957 26.35 13.15781 3.271468 0  
## 10 431617.0 1958 24.73 12.97529 3.208017 0  
## 11 432677.0 1959 23.85 12.97775 3.171784 0  
## 12 434329.0 1960 22.96 12.98156 3.133754 0  
## 13 420423.0 1961 22.23 12.94902 3.101443 0  
## 14 439043.0 1962 21.42 12.99235 3.064325 0  
## 15 477195.0 1963 20.87 13.07568 3.038313 0  
## 16 504182.0 1964 20.79 13.13069 3.034472 0  
## 17 526954.0 1965 20.19 13.17487 3.005187 0  
## 18 546822.0 1966 19.93 13.21188 2.992226 0  
## 19 564882.0 1967 19.63 13.24437 2.977059 0  
## 20 556706.0 1968 19.07 13.22979 2.948116 0  
## 21 570978.0 1969 19.43 13.25511 2.966818 0  
## 22 612661.0 1970 23.03 13.32557 3.136798 0  
## 23 560919.0 1971 24.73 13.23733 3.208017 0  
## 24 602492.0 1972 25.59 13.30883 3.242202 0  
## 25 598568.0 1973 26.97 13.30230 3.294725 0  
## 26 610023.0 1974 45.56 13.32125 3.819030 0  
## 27 654641.0 1975 50.92 13.39184 3.930256 0  
## 28 684913.0 1976 48.66 13.43705 3.884857 1  
## 29 697205.0 1977 46.66 13.45483 3.842887 1  
## 30 670164.0 1978 47.77 13.41528 3.866398 1  
## 31 781134.0 1979 47.93 13.56850 3.869742 1  
## 32 829700.0 1980 45.61 13.62882 3.820127 1  
## 33 823775.0 1981 44.66 13.62165 3.799078 1  
## 34 838112.0 1982 43.44 13.63891 3.771381 1  
## 35 782091.0 1983 39.84 13.56973 3.684871 1  
## 36 895920.8 1984 37.85 13.70561 3.633631 1  
## 37 883638.1 1985 36.15 13.69180 3.587677 1  
## 38 890314.7 1986 33.39 13.69933 3.508256 1  
## 39 918762.2 1987 31.52 13.73078 3.450622 1  
## 40 950265.3 1988 29.16 13.76450 3.372798 1  
## 41 980728.8 1989 27.78 13.79605 3.324316 1  
## 42 1029075.5 1990 26.67 13.84417 3.283539 1  
## 43 995983.9 1991 25.45 13.81149 3.236716 1  
## 44 997544.9 1992 24.34 13.81305 3.192121 1  
## 45 945424.3 1993 22.46 13.75939 3.111736 1  
## 46 1033504.3 1994 21.50 13.84847 3.068053 1  
## 47 1032973.8 1995 20.44 13.84795 3.017494 1  
## 48 1063855.5 1996 19.71 13.87741 2.981126 1  
## 49 1089931.8 1997 19.01 13.90163 2.944965 1  
## 50 1117535.2 1998 18.32 13.92664 2.907993 1  
## 51 1100431.4 1999 16.99 13.91121 2.832625 1  
## 52 1073611.6 2000 16.78 13.88654 2.820188 1  
## 53 1127688.8 2001 16.97 13.93568 2.831447 1  
## 54 1094283.0 2002 17.27 13.90561 2.848971 1  
## 55 1071752.6 2003 16.84 13.88481 2.823757 1  
## 56 1112098.9 2004 18.27 13.92176 2.905260 1  
## 57 1131498.1 2005 19.43 13.93905 2.966818 1

### Coal (short Tons) vs. Coal Price (2010 $) post 1975

CoalProdCoalPrice\_post1975

## CoalShortTons Year coalPrice2010 logCoalProd logCoalPrice DateFactor  
## 28 684913.0 1976 48.66 13.43705 3.884857 1  
## 29 697205.0 1977 46.66 13.45483 3.842887 1  
## 30 670164.0 1978 47.77 13.41528 3.866398 1  
## 31 781134.0 1979 47.93 13.56850 3.869742 1  
## 32 829700.0 1980 45.61 13.62882 3.820127 1  
## 33 823775.0 1981 44.66 13.62165 3.799078 1  
## 34 838112.0 1982 43.44 13.63891 3.771381 1  
## 35 782091.0 1983 39.84 13.56973 3.684871 1  
## 36 895920.8 1984 37.85 13.70561 3.633631 1  
## 37 883638.1 1985 36.15 13.69180 3.587677 1  
## 38 890314.7 1986 33.39 13.69933 3.508256 1  
## 39 918762.2 1987 31.52 13.73078 3.450622 1  
## 40 950265.3 1988 29.16 13.76450 3.372798 1  
## 41 980728.8 1989 27.78 13.79605 3.324316 1  
## 42 1029075.5 1990 26.67 13.84417 3.283539 1  
## 43 995983.9 1991 25.45 13.81149 3.236716 1  
## 44 997544.9 1992 24.34 13.81305 3.192121 1  
## 45 945424.3 1993 22.46 13.75939 3.111736 1  
## 46 1033504.3 1994 21.50 13.84847 3.068053 1  
## 47 1032973.8 1995 20.44 13.84795 3.017494 1  
## 48 1063855.5 1996 19.71 13.87741 2.981126 1  
## 49 1089931.8 1997 19.01 13.90163 2.944965 1  
## 50 1117535.2 1998 18.32 13.92664 2.907993 1  
## 51 1100431.4 1999 16.99 13.91121 2.832625 1  
## 52 1073611.6 2000 16.78 13.88654 2.820188 1  
## 53 1127688.8 2001 16.97 13.93568 2.831447 1  
## 54 1094283.0 2002 17.27 13.90561 2.848971 1  
## 55 1071752.6 2003 16.84 13.88481 2.823757 1  
## 56 1112098.9 2004 18.27 13.92176 2.905260 1  
## 57 1131498.1 2005 19.43 13.93905 2.966818 1

### Coal (short Tons) vs. Natural Gas Price (2016 $)

CoalProdNGPrice

## CoalShortTons Year NGprice2016 logCoalProd logNGPrice  
## 1 995983.9 1991 3.231168 13.81149 1.1728436  
## 2 997544.9 1992 3.751128 13.81305 1.3220567  
## 3 945424.3 1993 3.502115 13.75939 1.2533671  
## 4 1033504.3 1994 2.719049 13.84847 1.0002820  
## 5 1032973.8 1995 4.189430 13.84795 1.4325647  
## 6 1063855.5 1996 5.785939 13.87741 1.7554307  
## 7 1089931.8 1997 3.479154 13.90163 1.2467891  
## 8 1117535.2 1998 2.527560 13.92664 0.9272546  
## 9 1100431.4 1999 3.363069 13.91121 1.2128540  
## 10 1073611.6 2000 12.333335 13.88654 2.5123057  
## 11 1127688.8 2001 3.298609 13.93568 1.1935009  
## 12 1094283.0 2002 6.324222 13.90561 1.8443870  
## 13 1071752.6 2003 8.024077 13.88481 2.0824467  
## 14 1112098.9 2004 8.327970 13.92176 2.1196198  
## 15 1131498.1 2005 15.971197 13.93905 2.7707869  
## 16 1162749.7 2006 8.032412 13.96630 2.0834848  
## 17 1146635.3 2007 8.176131 13.95234 2.1012191  
## 18 1171808.7 2008 6.690743 13.97406 1.9007249  
## 19 1074923.4 2009 5.966981 13.88776 1.7862411  
## 20 1084368.1 2010 4.670271 13.89651 1.5412170  
## 21 1095627.5 2011 3.372574 13.90684 1.2156762  
## 22 1016458.4 2012 3.503683 13.83184 1.2538146  
## 23 984841.8 2013 4.330309 13.80024 1.4656388  
## 24 1000048.8 2014 3.518243 13.81556 1.2579619  
## 25 896977.4 2015 1.955133 13.70679 0.6704581

# Model and Results

## 1) Model log(coal Production)~log(coal Price) (all years)

The first model was simply to look to see if there was a statistically significant elasticity between coal production and coal for the year 1949 to 2005

### regression code

logcoal\_logCoalPrice<-lm(CoalProdCoalPrice$logCoalProd~CoalProdCoalPrice$logCoalPrice)

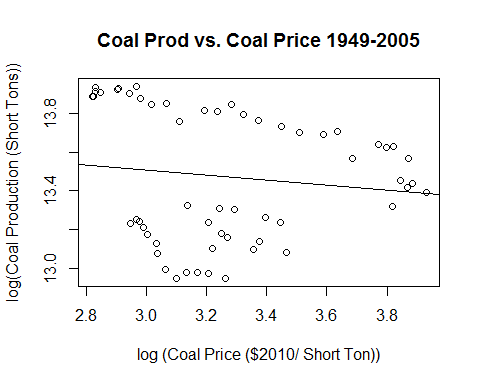
### summary results

summary(logcoal\_logCoalPrice)

##   
## Call:  
## lm(formula = CoalProdCoalPrice$logCoalProd ~ CoalProdCoalPrice$logCoalPrice)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.54716 -0.29831 0.04139 0.32851 0.42560   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13.8941 0.4516 30.764 <2e-16 \*\*\*  
## CoalProdCoalPrice$logCoalPrice -0.1283 0.1374 -0.934 0.354   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3341 on 55 degrees of freedom  
## Multiple R-squared: 0.01562, Adjusted R-squared: -0.002282   
## F-statistic: 0.8725 on 1 and 55 DF, p-value: 0.3543

### plot

plot(CoalProdCoalPrice$logCoalPrice,CoalProdCoalPrice$logCoalProd,main="Coal Prod vs. Coal Price 1949-2005", xlab="log (Coal Price ($2010/ Short Ton))",ylab="log(Coal Production (Short Tons))" )  
 abline(logcoal\_logCoalPrice)



### interpretation

Hypothesis test: 1. H0: There is no significant difference between coal production and coal price in 2010 real dollars. 2. H1: There is a significant difference between coal production and coal price in 2010 real dollars.

The statistical data summmary indicates a high P-value of 0.3543, which means that changes in coal production are not associated with changes in coal price in real 2010 dollars. A R-squared of 0.01562 and an adjusted R-squared of -0.002282 incidcate that this particular model explains none of the variability of the response data around its mean.

Therefore, there is no significant relationship between coal production and coal price in 2010 real dollars. However examination of the plot reveals that there might be two different sets , and possibly two different elasticities

## 2) Model log(coal Production)~log(coal Price) post 1975

SInce the first model over all the years did nbot show statistical significance, we divided the historical data into two sets -pre and post 1975.

### regression code

logcoal\_logCoalPrice\_1975<-lm(CoalProdCoalPrice\_post1975$logCoalProd~CoalProdCoalPrice\_post1975$logCoalPrice)

### summary results

summary(logcoal\_logCoalPrice\_1975)

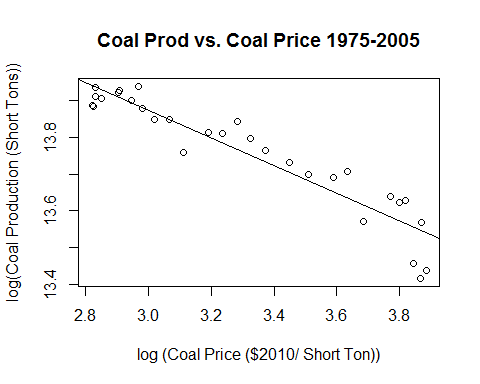
##   
## Call:  
## lm(formula = CoalProdCoalPrice\_post1975$logCoalProd ~ CoalProdCoalPrice\_post1975$logCoalPrice)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.13102 -0.02555 0.01214 0.03807 0.07764   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 15.00723 0.08734 171.8  
## CoalProdCoalPrice\_post1975$logCoalPrice -0.37785 0.02624 -14.4  
## Pr(>|t|)   
## (Intercept) < 2e-16 \*\*\*  
## CoalProdCoalPrice\_post1975$logCoalPrice 1.81e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05485 on 28 degrees of freedom  
## Multiple R-squared: 0.881, Adjusted R-squared: 0.8768   
## F-statistic: 207.3 on 1 and 28 DF, p-value: 1.805e-14

confint(logcoal\_logCoalPrice\_1975)

## 2.5 % 97.5 %  
## (Intercept) 14.8283235 15.1861304  
## CoalProdCoalPrice\_post1975$logCoalPrice -0.4316067 -0.3241011

### plot

plot(CoalProdCoalPrice\_post1975$logCoalPrice,CoalProdCoalPrice\_post1975$logCoalProd,main="Coal Prod vs. Coal Price 1975-2005", xlab="log (Coal Price ($2010/ Short Ton))",ylab="log(Coal Production (Short Tons))" )  
 abline(logcoal\_logCoalPrice\_1975)



### interpretation

Hypothesis test: 1. H0: There is no significant difference between coal production and coal price in 1975 dollars. 2. H1: There is a significant difference between coal production and coal price in 1975 dollars.

The statistical data summmary indicates a low P-value of 1.805e-14, which means that changes in coal production are indeed associated with changes in coal price in 2010 dollars for the year 1975-2005. A R-squared of 0.881 and an adjusted R-squared of 0.8768 incidcate that this particular model explains most of the variability of the response data around its mean. SInce this is a log-log model the elasticity is defined as the slope of log coal price. It was calculated as -.37785 with a 95% confidence surrounding it of [-.43, -.32]. Therefore, a one dollar increase in coal price per MBTU results in 7.8419 MBTU (or 0.37785 Short tons) decrease in coal production due to OPEC inception and its cartel effect in 1975.

There is a significant difference between coal production and coal price in 2010 dollars for the years 1975-2005.

## 3) Model log(coal Production)~log(Natural Gas Price) (Henry Hud)

This model wants to test the theory that coal production is somehow related to the price of natural gas defined for the Henry Hubfor the year 1991-2015 in 2016 $.

### Regression Code

logcoal\_logNgPrice<-lm(CoalProdNGPrice$logCoalProd~CoalProdNGPrice$logNGPrice)

### Summary Results

summary(logcoal\_logNgPrice)

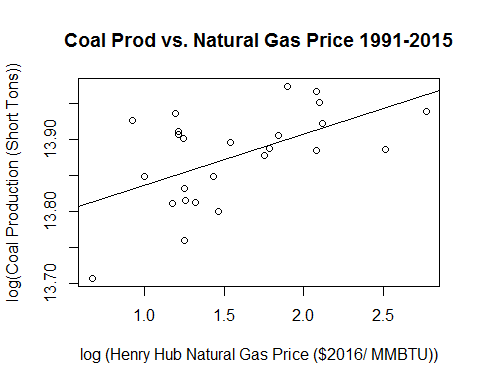
##   
## Call:  
## lm(formula = CoalProdNGPrice$logCoalProd ~ CoalProdNGPrice$logNGPrice)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.105830 -0.036930 -0.004368 0.047940 0.095720   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13.76484 0.03651 376.968 < 2e-16 \*\*\*  
## CoalProdNGPrice$logNGPrice 0.07126 0.02222 3.207 0.00391 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05576 on 23 degrees of freedom  
## Multiple R-squared: 0.309, Adjusted R-squared: 0.279   
## F-statistic: 10.29 on 1 and 23 DF, p-value: 0.00391

confint(logcoal\_logNgPrice)

## 2.5 % 97.5 %  
## (Intercept) 13.68930263 13.8403752  
## CoalProdNGPrice$logNGPrice 0.02529871 0.1172227

### Plot

plot(CoalProdNGPrice$logNGPrice,CoalProdNGPrice$logCoalProd,main="Coal Prod vs. Natural Gas Price 1991-2015", xlab="log (Henry Hub Natural Gas Price ($2016/ MMBTU))",ylab="log(Coal Production (Short Tons))")  
abline(logcoal\_logNgPrice)



### Intrepretation

Hypothesis Test: 1. H0: There is no significant difference between coal production and natural gas price in 2016 dollars. 2. H1: There is a significant difference between coal production and natural gas price in 2016 dollars.

The statistical data summmary indicates a low P-value of 0.00391, which means that changes in coal production are indeed associated with changes in natural gas price in 2016 dollars. A R-squared of 0.309 and an adjusted R-squared of 0.279 incidcate that this particular model explains there is a variability of the response data around its mean there are other factors to consider. SInce this is a log-log model the elasticity is calculated as .071 with a 95% confidence interval of [.025, .11] Therefore, a one dollar increase in natural gas price in 2016 dollars per MBTU results in 1.4789 MBTU (or 0.07126 Short tons) increase in coal production.

Based on the statistical data summary and observations, we can see that natural gas prices have been steadily decreasing since the mid 2005 due to prominent fracking activities, which results in a postive relationship between natural gas prices and coal production. Addtionally, if there is a spike in natural gas prices, then coal production will also increase due to the demand for substitue (coal) increases.

There is a significant difference between coal production and natural gas price in 2016 dollars.