

Coal Assignment Commentary

James Woods

10/12/2016

General Comments on Assignment

Good first attempt working with R. Picking up a new programming language in far too little time is *very* hard.

As always, there is room for improvement. I just wanted to make sure you had at least the start of the capability to do an empirical paper if you chose to.

General Comments on Each Group

- ▶ Team The Best – Best of show with hypotheses tests and organization. Very clear that someone had programming and analysis experience. R code had an accent.
- ▶ Green Team – Great work with the background stories that go with the data. Made it very clear that you can't just put in data and get answers out, you need to know the physical data generating process.
- ▶ Trailblazers – Would not run. Always check this before you turn it in.
- ▶ Working Title – Caught the OPEC I and OPEC II events.
- ▶ Econ Nerds – Journeyman effort (That is praise and a good thing). Tried an elasticity formulation.

Check this link to see what others did

<https://github.com/woodsjam/Course-Energy-Economics/tree/master/CoalAssignment>

Trailblazers

- ▶ Start with a clean environment and put every step in chunks.
- ▶ Do not put code directly into the command line.
- ▶ Beware if you see this in your document with nothing above it.

View(RealPrices)

summary(RealPrices)

Working Title

Here is the interesting part. That gives OPEC I and OPEC II events.

```
CoalRegressionA <- lm(Price ~ ProductionShortTon  
+I(Year<1973), data = CoalMarket)
```

```
CoalRegressionB <- lm(Price ~ ProductionShortTon  
+I(Year>1978), data = CoalMarket)
```

Possible improvement is to do it in one step.

```
CoalRegression$OPEC<- "Pre"
```

```
CoalRegressionOPEC[CoalRegressionYear> 1978] <-  
"PostOPEC"
```

```
CoalRegressionOPEC[CoalRegressionYear>= 1973 &  
CoalRegression$Year<= 1978] <- "OPEC"
```

```
CombinedRegression <- lm(Price ~ ProductionShortTon  
+OPEC , data = CoalMarket)
```

Econ Nerds

- ▶ Explain a little more. You left the elasticity form without explanation. Remember that when you write a report or paper, you must provide the story and explanation of the results.
 - ▶ The table is for the experts
 - ▶ The text is for those that are not but must act.
- ▶ Good comments about fit between Lin-Lin and Log-Log model.
- ▶ By the way, left hand side log alone is commonly used to cope with heteroskedasticity.

Econ Nerds: Technical

- ▶ Did this interest thing. They converted quantities from K Ton to Ton.
- ▶ Be careful. Data should be on similar numerical scale 100s to 100s, but 100000000s to 10s can be dangerous.
 - ▶ Parameter estimates can get lost in the precision.
 - ▶ There can be numerical problems in estimation:
 - ▶ Usually only with large difference in scale
 - ▶ or with many observations
 - ▶ Programs with poorly implemented OLS routines.

Green Team

The key to good stats analysis is making sure the physical data generating process matches up with the statistical data generating process. Green team had great notes on why they made their decisions.

Green Team: Why the low in the 1950s?

Several market changes drove down demand (to its 1954 low), including railroads fuel-switching from coal to diesel, home heating and cooking switching from coal to natural gas, and a general trend of industrial production shifting from coal-fed manufacturing processes to electricity-fed production. (Andrew Needham, *Power Lines: Phoenix and the Making of the Modern Southwest*, Princeton University Press, 2014)

Green Team: What was behind the steady increase in production, sustained since the 1950s?

The uneven but steady increase in coal production, after the lows of the 1950s, was largely driven by an equally steady increase in demand, particularly from the expansion of coal-fired power plants—a trend that was sustained into the first decade of the 21st century. The same increase in production was also enabled by significant technology changes associated with mining—especially those technologies associated with surface mining and mechanical extraction (e.g. conveyor belts and mechanical loaders). (G.E. Harding, “American Coal Production and Use,” *Economic Geography*, Vol. 22, No. 1 (Jan. 1946)).

Also good use of graphs to see what was wrong with specification.

Taking a first glance at our Residuals vs Fitted plot graph, we can tell that this type of graph can only be fairly useful because of the upside U shape of the data points. Also, that the relationship between our two variables is not as linear as we would have hoped. As we can tell the lower the price of coal the lower the variance in our error terms. Where as we move out to the larger price points the large the variance in error terms gets and the more heteroskedasticity we see.

Team the Best

- ▶ Study this one.
- ▶ A few notes on the 'accent'
 - ▶ Load libraries first is usually not necessary in R because of lazy loading.
 - ▶ Logs and the like can be made on the fly in regressions and don't need a new column before hand.
 - ▶ use the data = parameter in the lm function.
 - ▶ Great model
- ▶ Hypotheses tests and interpretation were great.
 - ▶ Really appreciated that.
 - ▶ Beware of R^2 focus. There is a really great paper on the use and abuse of R^2 from the 70s but it is hard to find because the guy's name is 'A. Buse'