ECON 493 Term Paper Proposal

I want to model and forecast the price of crude oil in Canada using the Canadian crude benchmark Western Canadian Select (WCS). I have monthly data of WCS prices ranging from 2009 to the present.

I plan to either use a dynamic regression model with ARIMA errors, or just a regular ARIMA model. For the DRM, the regressors I have collected are WTI prices, which is the benchmark for American light crude and by which most people use to price WCS, total amount of crude exported to the US, which accounts for about 99% of Canada’s crude exports, and transportation amounts by crude and rail. Capacity constraints are also a giant factor to pricing – this month’s WCS prices face a huge long lasting negative shock because certain refineries in the US have shut down for maintenance that last for weeks, leaving our pipelines full and our export crude nowhere to go. All this data comes from a variety of sources: namely, Statistics Canada, US Energy Information Agency, and the National Energy Board.

The explanatory variable data for a DRM with ARIMA comes with its own problems. Firstly, the data had to be gathered from different sources and therefore has a lot of consistency issues. For example, there is a lot of missing data; the National Energy Board only presents exports of crude with rail starting from 2012, and Statistics Canada only reports its pipeline export data starting from 2016 – this leaves me with about 30 points of data. StatsCan also does not report pipeline export data solely on crude oil, but a variety of material. I have also not been able to find any data pertaining to transportation costs, a big factor in the pricing of WCS, other than an annual average from 2014. A solution to this problem would be to use the average and transform the data into total monthly transportation costs and use that as a proxy; however, I don’t see how transportation costs wouldn’t fluctuate at least annually, so this data might be inaccurate.

Another problem with the data is that a large factor in the pricing of WCS is capacity constraints. The current market is facing a huge long lasting negative shock because several refineries in the US have shut down for maintenance, leaving our pipelines filled and our exports nowhere to go. First problem with this is that the data on constraints is very limited; either our exports are moving or they are not. To model this variable, I would either have to use a proxy or a dummy, and modelling monthly capacity of a pipeline is not very accurate data. So I’d have to use a dummy with something along the lines of “did exports stop moving” – given how these shocks only last 2-3 weeks until the market rights itself or the refineries go back into production, this dummy isn’t very precise either.

The biggest problem among these is the missing data. With them, I’m missing more than 2/3 of the data, and if I exclude the regressors, I’ll not only have bias in my estimators, but I’ll be missing a lot of the explanatory power that I know have powerful contributions to the predictors. Given that the data has such significant flaws, I might have to stick with a regular ARIMA model, and just forecast the pricing with just WCS data. For example, I could just model the capacity constraints as a shock in the data. Further testing of the dynamic model and comparison of the two types of models are required.

The WCS prices are non-stationary, so some data transformation will be necessary.

R

> library(ggplot2)

> wcs=data.frame(X493tp\_data2)

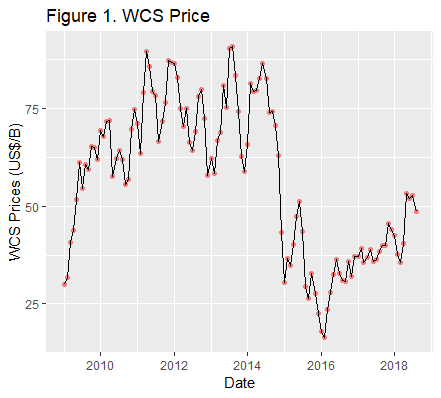
> wcsg = ggplot(wcs,aes(Date,WCS))+geom\_point(color="lightcoral")+geom\_line(color="grey0")

> wcsg+

+ ylab("WCS Prices (US$/B)")+

+ xlab("Date")+

+ ggtitle("Figure 1. WCS Price")



> summary(wcs)

Date WCS WTI..US..b.

Min. :2009-01-01 00:00:00 Min. :16.30 Min. : 30.32

1st Qu.:2011-05-24 06:00:00 1st Qu.:38.30 1st Qu.: 50.82

Median :2013-10-16 12:00:00 Median :60.05 Median : 75.68

Mean :2013-10-16 02:04:08 Mean :56.79 Mean : 73.31

3rd Qu.:2016-03-08 18:00:00 3rd Qu.:72.83 3rd Qu.: 94.60

Max. :2018-08-01 00:00:00 Max. :90.97 Max. :110.04

crude.pipeline.exports..b. export.by.rail..bbl.

Min. : 77105168 Min. : 290635

1st Qu.: 86243561 1st Qu.:2711110

Median : 90894447 Median :3716146

Mean : 89999285 Mean :3562688

3rd Qu.: 93881577 3rd Qu.:4548439

Max. :100285508 Max. :6405338

NA's :85 NA's :37

US.imports.from.Canada..thousands.of.b.

Min. : 52971

1st Qu.: 67886

Median : 81009

Mean : 82657

3rd Qu.: 99027

Max. :119871

NA's :1

Data references (complete MLA sourcing on final paper)

* <https://economicdashboard.alberta.ca/OilPrice>
* <https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MCRIMUSCA1&f=M>
* Statistics Canada. Table 25-10-0056-01 Canadian pipeline transport of oil and other liquid petroleum products, monthly (cubic metres)
  + <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2510005601>
* <http://www.aopl.org/resources/pipeline-basics/about-pipelines/>
* <https://www.neb-one.gc.ca/nrg/sttstc/crdlndptrlmprdct/stt/cndncrdlxprtsrl-eng.html>