ECN431
Lab session 2
Empirical market analysis

Aud 12
Friday, February 2nd 2018

Before the Lab

- Obtain the datasets and the example R notebook from the GitHub repository of the course.

- Try to work through the exercises on your own prior to the lab.

Purpose of this lab

Work with applying instrumental variables (IV) methods to real-world data, in particular for the purpose of estimating competitive supply and demand, and learn about the electricity market.

Learning goals

- Improve the skill of presenting data for own and other's understanding

- Understand what the standard IV method does

- Be able to perform IV estimates using R

- Know several ways of assessing instrumental variables

- Know the basic features of the Nordic electricity market

- Be able to estimate supply and demand in competitive markets

- Be able to explain, present and identify critical assumptions in your analysis

Practical information

The example R notebook (which you either pulled to your fork from the main repository or downloaded from GitHub) is opened in RStudio. This file contains several commands that you will need in order to answer the questions in these exercises, in addition to information on new commands. Remember: If/when you are uncertain about the purpose, functionality and extensions to/options for some of the commands, remember that you can get documentation for the command in RStudio by pressing F1. You can also try to google your particular problem.

Exercises

Warming up at the Fulton fish market

Use the data set in *fishmarket.dta* for this exercise. The Fulton fish market resides in New York, and was located by Fulton Street in the Financial District, Lower Manhattan for 193 years (until 2005, when it moved to Hunts Point in the Bronx). The sellers at the market buy their fish from suppliers, typically having the fish transported to the market by truck. Buyers on this market typically request large quantities of fish (restaurants and other businesses requiring fish).

The data contain daily quantities and prices for sales of Whiting for one particular seller from December 8th 1991 to May 8th 1992.[1] The market is open from Monday to Friday, and prices are to a large degree determined by the supply of Whiting on a given day. Difficult conditions at sea, such as waves and strong wind, will reduce catch. You will apply this insight for the purpose of estimating demand.

Present and understand your data

1. Make a table of summary statistics for price, quantity, and indicators for stormy and mixed weather.

2. Create the following two graphs: i) Quantity sold per date, and ii) Price per unit per date. Choose a suitable graph type.

3. Regress quantity on price, and comment. Would you interpret the price coefficient as the price sensitivity of demand? Why or why not?

Assessing and understanding IV

4. Create a graph showing price over time, clearly indicating periods where weather at sea is stormy or mixed (preferably so one can separate the periods where it is stormy or mixed). Comment on the relationship between price and weather conditions.

5. Regress price on stormy weather and quantity on stormy weather. What is the interpretation of the coefficients? Make a table of the average quantity and price

---

[1]The data was collected by Kathryn Graddy for her PhD dissertation at Princeton University. Her analysis and further description of the data can be read in the article "Testing for imperfect competition at the Fulton fish market", published in the RAND Journal of Economics, Vol. 26, No. 1 (pp. 75–92), 1995.

according to whether the sea is stormy or not. Calculate the average change in price and quantity from stormy weather. Also calculate the ratio $\Delta q/\Delta p$ (ratio of the changes).

6. Regress quantity on price, using the indicator for stormy weather to instrument price. Explain the relation to the ratio of changes you calculated previously.[2] Discuss the validity of this instrument with your group – are there any potential issues you could test with the data you have?[3]

7. Regress quantity on indicators for day of week. Do the same for stormy weather. Now perform the previous IV estimates including these indicators (make sure also report the first-stage regression). Discuss the changes in the regression results with your group.

8. Extra:

   Estimating supply and demand

   (a) So far, we have been interested in the demand side. Day of week seems to shift demand, particularly Tuesday and Wednesday. Create an indicator that takes the value 1 if the day is either Tuesday or Wednesdays. Regress i) price and ii) quantity on stormy weather and the new indicator for low-demand days. Discuss estimating supply either as the quantity supplied as a function of price, or as the inverse supply schedule (price as a function of quantity), in both cases using low-demand period as an instrument for demand.

   (b) Do the IV regression of price on stormy weather and quantity, using low-demand days as an instrument for quantity. Discuss the findings, particularly the validity of the instrument (assess relevance from the first stage) and why or why not you think this supply function seems reasonable. If we believe this supply curve is the right one, what does it tell us about supply in this market?

   (c) In a price-quantity diagram, draw the supply function both when the weather is stormy and not, and demand during regular and low-demand days (use the estimated coefficients). Try to use R commands for drawing functions to make the graphs.

Nord pool spot: The electricity wholesale market

Electricity markets are special for several reasons. One is the typically massive investments required to have an operative grid for transmission of electricity from where it is produced, to where it is consumed. Another is that supply and demand must be

---

[2]If you're stuck, review the lecture slides for instrumental variables, slide 25 and 27.

[3]The relevance criterion means that the instrument must affect the endogenous variable. In practice, the this is usually assessed by the instrument having a significant coefficient in the first stage (regression of the endogenous variable on the instrument and controls), as well as the F-statistic for the first stage regression having a value of at least 10 (the F-statistic is the basis of a test for joint statistical significance of all regressors). This is purely a rule of thumb, and higher values are considered better, though in certain cases a lower F-statistic could be used if one has few observations, includes many insignificant controls, and the instrument can otherwise be shown to have a strong correlation with the endogenous variable (though be wary if you see/get low F-statistics, especially if combined with marginal significance of the coefficient on the instrument).

equated in each instant, otherwise the frequency (in the oscillations of the alternating current (AC)) in the grid will fluctuate, potentially damaging electrical apparatus or even leading to transmission halting completely (blackouts). The operation of modern electric power distribution is a feat of engineering (and to some extent economics). In the Nordic countries, the grid and production is separated, where the grid is subject to natural monopoly regulation, while production is "freely" competing.[4]

Nord Pool is an electricity market, jointly owned by the Nordic and Baltic transmission system operators. The integrated market covers Denmark, Norway, Sweden, Finland, Belarus, Latvia and Lithuania. Most notably, they operate the day-ahead and the intraday market, both designed as auctions. The day-ahead market sets the plan for production and consumption for the upcoming day, setting prices in the process. The intraday market is an adjustment market where actors can buy and sell production after the day-ahead market has closed, but before the hour of production. There is also a market for financial products related to the Nordic and Baltic power market (power derivatives), operated by NASDAQ OMX Commodities Europe. In addition, large firms might have bilateral contracts for electricity supply with producers. In this exercise, we will study the day-ahead market, which is crucial for the formation of the price of electricity, also in the other markets.

In the day-ahead market, sellers and buyers of power in the whole market area submit their bid schedules, The bids can be set very flexibly, and a bidder can submit bids for both selling and buying (typically turning from buyer to seller at a certain price). The typical actors are power production units on the seller side, and industrial firms with power-intensive production and electricity retailers (selling to consumers and other firms) on the buyer side. The bidders submit separate bids for every hour of the day in which they wish to take part in the market. The day before, bids for each of the 24 hours of the next day are submitted with a deadline at noon.

The auction is a uniform price action, where there will be one price for every unit exchanged in a given hour. The price is chosen such that quantity sold is equal to quantity purchased, where buying bids with the highest submitted price and selling bids with the lowest submitted price are activated first. This is done for the market as a whole in a first step, resulting in what is referred to as the "system price" and "turnover at system price". Due to transmission constraints between different areas, this result might not be feasible, which will result in prices in some price areas deviating from the system price to make local demand and supply satisfy the transmission constraints.[5]

The single largest power source in the market is hydro, accounting for about 50% of demanded energy in the Nordic countries in years with normal rain and snow, while nuclear and different types of thermal power (coal, oil, gas, and biofuels) accounts for around 25% each, while wind power has around 3% share after an increase in Danish production the last two years.

Due to the local climatic conditions of the Nordic area, there is a clear relationship between energy consumption and the season, driven by temperature. Temperatures within the variation we see in the region will not have an effect on power production.[6] In the

---

[4]Production is also subject to a wide range of regulations, which you can read more about from other sources.

[5]There are many ways to set the area prices such that transmission constraints are satisfied, though the algorithm employed searches for the solution which can be thought of as minimizing the shadow prices on the transmission constraints in the problem of maximizing market surplus (maximizing the sum of difference between purchasing and selling bids).

[6]It's conceivable that thermal power plants might be subject to regulation during periods of extreme

data set *elmarket.csv*, you will have hourly system prices and turnover at system prices, in addition to rich data on temperatures in Norway and Sweden which you will use to estimate the supply curve.

Hydro power production is heavily dependent on access to rain and melt water to fill up the reservoirs, and it is expected that the amount of water in the reservoirs will affect prices. However, the problem of a producer is not as simple as looking at the amount of water in the reservoir to make a decision on how to set up a bid schedule. If we think about production cost in hydro, the physical costs of producing one unit is basically zero. The true marginal cost comes from something called the *shadow value of water*, an alternative cost which can be thought of in terms of the price one could get by saving the water today, and rather producing tomorrow.[7] This alternative cost arises because the access to water for production is not without limit, thus not driving prices to zero. In the data, you will also have the power potential of the stored water at weekly levels, and attempt to use this as a supply shifter when estimating demand.

Data

- time – date and hour of day for the observation

- price – the system price from the day-ahead market in EUR/MWh (Euro per 1000 KWh) for each hour

- volume – the turnover at system price from the day-ahead market in MWh for each hour

- hydrores – the total production possible in TWh with the water stored in hydro power reservoirs in the Nordic countries. Measured weekly, on Monday of each week (repeated values from Monday 0 hours to Sunday 23 hours).

- Arendal–Uppsala - air temperature in Celsius for several Norwegian and Swedish cities, measured hourly.

Tasks

Data wrangling

1. Generate a population-weighted average temperature for each date and hour for the Norwegian areas and the Swedish areas separately, i.e., a Norwegian average temperature and a Swedish average temperature. Use the population figures quoted in the two files *norpop.csv* and *swepop.csv*.

Present and understand your data

2. Create a table of descriptive statistics for price, volume and magazine capacity, another for temperatures in the Norwegian areas, and a third for temperatures in the Swedish areas, and output them to a document.

---

heat, as has been the case in continental Europe, due to waste water of high temperature potentially damaging river ecosystems. As far as I'm aware, this has not been a problem in the Nordics.

[7]This is an example of a dynamic optimization problem.

3. Create the following graphs: i) Average hourly volume per week, ii) Volume weighted average price per week, iii) Average temperature in Norway and Sweden per week, and iv) Hydro reservoir capacity per week (this variable is recorded weekly, so just take the average). Choose the type of graph you think is best for conveying the information. Maybe some of the graphs would make sense to display in the same figure?

4. Create the following graphs: i) Average volume per hour of day, ii) Average volume weighted price per hour of day, iii) Average temperature in Norway and Sweden per hour of day, and iv) volume weighted price per hour of day against average volume per hour of day.[8] Choose the type of graph you think is best for conveying the information. Do you choose the same as for the previous plots?

5. Discuss the relationship between electricity use, prices, temperatures, and hydro reservoir reserves. Note in particular the pattern over time, and throughout the day. Discuss in terms of demand side and supply side factors.

6. To understand better what is driving electricity use (particularly the pattern we see in volumes over time), make a plot featuring the original series of volumes, as well as a residualized version after regressions of temperature on i) temperature in Norway, ii) temperature in Norway and indicators for month, and iii) temperature in Norway, indicators for month and indicators for year. Add back the average volume to the three residualized versions to make the graphs more comparable.[9] Discuss what this figure tells you.

7. To understand what drives hydro reserves, do the same as above; make a plot featuring the original series, as well as residualized versions, after controlling for the same variables as above. How you would think about hydro reserves being determined after this, particularly comparing the raw series of prices and the pattern of hydro reserves with different controls.

Estimating supply and demand

8. You want to estimate the supply curve. Perform an IV-regression of volume on prices, using temperature as instrument, choosing which controls to add. Could hour of day be used as an instrument? Discuss the instrument, the first stage and the results. Also calculate the elasticity of supply with respect to price (at the average volume and price in the sample).

9. To estimate the demand curve, perform an IV-regression of volume on prices, using hydro reserves as instrument. Choose controls to add to the regression. Discuss the instrument, the first stage and the results. Also calculate the elasticity of demand with respect to price (at the average volume and price in the sample). Output both the estimated demand and supply functions side by side in one table.

Assessing instruments

---

[8]The last one has 24 points in total – the volume weighted price for the first hour (average across all days in the sample) against the average volume for the first hour, and so forth.

[9]Try to make a plot featuring the original series in it's own graph above, while the three residualized versions are shown together in one graph below.

10. To assess the relationship between volume and temperature, graph volume for different *bins* of temperature in Norway using `ggplot` and `stat_summary_bin` (use the argument `bins = x` to specify the creation of x bins, for instance 100).[10] Does the relationship look linear? Also use indicators for each hour as a control (regression on dummies for each value of hour). Discuss the difference. Do the same for price. Feel free to check if the variables in logarithm looks better.

11. Perform the following, relatively complex operations, related to what is called a Visual Instrumental Variables (VIV) plot:

    (a) Generate categories/bins of temperature in Norway in steps of one from the whole degree below the minimum to the whole degree above the maximum.[11] Make graphs showing the average volume and price within each bin, as well as the number of observations within each bin. Also plot the average volume and price within each bin against each other.[12]

    (b) Do the same as above, but first residualize temperature, volume and price with respect to any controls you believe should be included in the supply equation. Add back the average value of each variable to the residuals to make them comparable to the original variables.

    (c) Make the same graphs as before, only this time using bins of hydro reserves. Since hydro reserves are only reported weekly, there are only slightly more than 200 effective observations of hydro reserves in the sample, so try with about 13 bins.

    (d) Do the same as above, but first residualize hydro reserves, volume and price with respect to any controls you believe should be included in the *demand* equation. Add back averages.

12. Extra:

    (a) *Evalutate the approximation:* Discuss and try to check for which range of outcomes you believe the results are good. Can you assess the amount of variation the instrument effectively creates in the outcomes we're interested in here, for example from some of the graphs you've created?

    Apply the results

    (b) The government currently collects a tax of roughly 18 EUR/MWh on electric power.[13] What is the average effect on prices and volumes if this is increased to 20 EUR/MWh? What is the reduction in consumer surplus and producer surplus, and what is the deadweight loss?

---

[10]This is a type of discretization, where we choose a range of x-axis values, say, temperatures between -3 and -2, -2 and -1, etc. and calculate the average y-axis value within this range, say, the average temperature between -3 and -2 (and so forth). The command `stat_summary_bin` automates this.

[11]You can use `stat_summary_bin` with the argument `binwidth`, to achieve this.

[12]This last graph is known as the VIV plot, and can be used to assess the instrument in several ways – both how "noisy" the relationship between the two variables are when "filtered" through the instrument, and at least speculate about possible functional form issues (if we trust the instrument). The plots for the dependent variable and the suspected endogenous regressor against bins of the instrument correspond to the reduced form and the first stage, and can be used for similar assessments.

[13]This applies in Norway, and there are reductions for several types of actors, but you can disregard these additional layers of complexity for this task.

(c) Statnett, together with a partner on the German side, has started construction of a new subsea cable to Germany named *NordLink*. The maximum transfer capacity of the cable is 1400 MW. Assume that the cable will increase electricity exports to Germany by 1200 MWh in every hour.[14] Calculate the effect on prices and domestic production (and/or domestic consumption). How much does domestic producers gain, and how much does domestic consumers lose from this cable?

(d) *Further research:* Go to the webpages of Nordpool spot, and download one of sheets containing data on the market cross for a representative day of the sample. Use the bid curves in a suitable hour (e.g., at noon) to calculate the (arc) elasticity of demand and supply in an area around the equilibrium price. How does your finding here match up with the implied elasticity from your estimates?

---

[14]This is an unrealistically high number, and only serves as an example. The net transfer over the first year of operation is calculated to be 0, and realistic gain/loss calculations would be more complicated than what I'm asking for here.