Market Analysis with Econometrics and Machine Learning

1a Introduction: Bob's Ice Cream Business

Uni Ulm

Prof. Dr. Sebastian Kranz

SoSe 2020

Bob’s Ice Cream Business

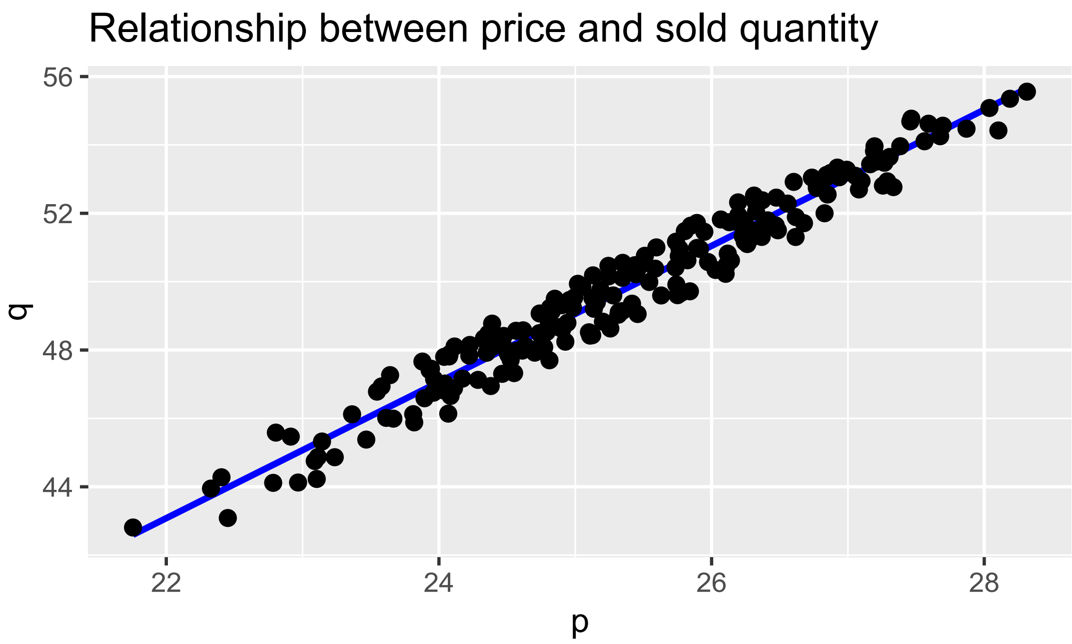
- Bob just started his career as an entrepreneur by taking over a small ice cream truck
- The previous owner, Emma, drove to different locations each day and used a little blackboard, on which every morning she wrote down how much a portion of ice cream shall cost today
- Emma left Bob a small collection of sales data, in which she noted for each day t
  - the price p_t she had set
  - the number of ice portions q_t she sold
  - the wholesale price w_t she had to pay for a big box of ice cream from the wholesale store
- You are a young consultant who shall help Bob to optimize his ice cream business. How would you proceed to find an optimal pricing rule?

The data

```
  t      w       p       q

  1   31.73   26.84   53.05
  2   15.58   26.32   52.33
  3   86.08   26.11   50.50
  4   98.91   23.10   44.23
  5   55.00   27.68   54.25
  6   73.19   27.08   52.70
  7   74.85   23.81   46.13
  8   85.80   26.03   50.34
  9   53.25   25.14   49.21
 10   54.65   24.50   47.91
...     ...     ...     ...
```

Plot of prices and output



A model of the data generating process with a profit-maximizing firm

- We now present a simple model:
  - Downward sloping demand function, with random demand shocks
  - Firm sets profit maximizing prices
  - We can find a positive relationship between prices and output.

Demand Function

- The sales in period t shall be given by the following demand function

$$q_t = a_t - bp_t$$

  - The market size parameter a_t shall be given by

$$a_t = a_0 + \varepsilon_t$$

  - ε_t is a random variable that measures a “demand shock” in period t.
  - Example: Good weather -> higher ε
  - The demand shock shall not be influenced by the price (but the price may depend on ε)
  - b and a_0 are exogenously given numerical parameters with

$$a_0 > 0 \text{ and } b > 0$$

Profit Maximization

- The firm’s profits shall be given by

$$\pi_t = p_t q_t - c_t q_t - F$$

  - $c_t \geq 0$ is a constant marginal cost of production in period t (cost of one portion of ice cream)
  - F are some fixed costs that don't depend on the price.
- Assume the firm knows in each period t its cost c_t, the demand shock ε_t and the parameters a_0 and b in period t and chooses a price p_t that maximizes its expected profits.

Optimal Prices and Corresponding Output

Exercise: Show that the profit-maximizing price and corresponding sold quanitity are given by:

$$p_t^* = \frac{a_0 + \varepsilon_t}{2b} + \frac{c_t}{2}$$

and

$$q_t^* = \frac{a_0 + \varepsilon_t}{2} - \frac{bc_t}{2}$$

Solution

The optimal price p_t^* maximizes the profit function.

$$\begin{aligned} \pi(p_t) &= q_t(p_t)(p - c_t) - F \\ &= (a_t - bp_t)(p_t - c_t) - F \\ &= a_t p - a_t c_t - bp_t^2 + bp_t c_t - F \end{aligned}$$

The first order condition is given by

$$\begin{aligned} \pi'(p_t^*) =& a_t - 2bp_t^* + bc_t = 0 \Leftrightarrow \\ p_t^* =& \frac{a_t}{2b} + \frac{c_t}{2} \end{aligned}$$

If we plug in $a_t = a_0 + \varepsilon_t$, we get

$$p_t^* = \frac{a_0 + \varepsilon_t}{2b} + \frac{c_t}{2}$$

We get the equilibrium quantity as

$$\begin{aligned} q_t^* &= a_0 + \varepsilon_t - bp_t^* \\ &= \frac{a_0 + \varepsilon_t}{2} - \frac{bc_t}{2} \end{aligned}$$

Simulating the model in R

- To get intuition about this economic model, we want to simulate in R.
- Here and in other places I will show you some R code in the class and the steps are roughly described in the slides.
- You will repeat similar steps in more detail in the corresponding RTutor problem sets (here the problem set for Chapter 1a).
  - For that reason you won't get the R code from class. Look at the RTutor problem set instead.

Simulating the model in R

- Open a new R script in RStudio and write a program that numerically simulates the model above
  - Choose some values for the parameters $a_0 > 0$ and $b > 0$ and pick a total number of periods T
  - Assume demand shocks are independently, identically normally distributed with standard deviation σ_ε

$$\varepsilon_t \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

  - Assume costs c_t are uniformely distributed on an interval.
  - Compute the values of q and p for each period and show a scatter plot of both variables.

Analysing the Model with Simulation and Theory

- Is the plotted relationship between prices and quantity just a noisy version of the demand curve?
- Try to find parameter constellations such that prices are positively correlated with sold quantities.
- Why can prices be positively correlated with sold quantities? Try to explain intuitively.
- Study with your simulation how the correlation between q and p changes if you reduce the standard deviation of the demand shocks σ_ε to a very small value.

Insights from the analysis so far

- It is really useful to think about how a data set is created: Form a model about the data generating process!
- A positive correlation between prices and output does not imply that higher prices *cause* higher output.
  - A positive *correlation* can arise if prices are set systematically higher in markets with high demand.
- Here, the observed *relationship* between prices and output is *not* a noisy version of a demand function.
- The demand function describes how, ceteris paribus, higher prices *cause* lower quantities sold.
- To set profit maximizing prices it would be really helpful to estimate such a demand function.

A pricing experiment to find the demand function

- You convinced Bob that the data could have been generated by the model described above. He asks you to estimate the the demand function.
- You propose a “randomized pricing experiment”
  - Bob shall choose for T periods in each period t some randomly selected price p_t and observe the resulting sales q_t

R Simulation of Pricing Experiment

(You do this in the corresponding RTutor problem set)

- Adapt your R code such that prices are now randomly selected.
- Plot the resulting relationship between p and q. Does it look like a noisy version of the demand curve?

Estimating the demand function

- Use the R function lm (stands for linear model) to estimate the following linear regression model of the demand function

$$\begin{aligned} q_t = & \beta_0 & + \beta_1 p_t & & + \varepsilon_t \\ (= & a_0 & + (-b) p_t & & + \varepsilon_t) \end{aligned}$$

  via ordinary least squares (OLS).
  - Compare the true values of β_0 and β_1 with the OLS estimates.
  - Does the OLS estimator seem to be **consistent** in the sense that the estimated coefficients are close to to their true values when we have a very large number T of observations? (We will formally define consistency later)
- Repeat the previous exercise for the case that the data is generated by the model with profit maximizing prices.

What have we done so far?

- **Policy Question:** We want to find a rule to set profit maximizing prices for Bob’s ice cream business.
- **Model for decision:** We discussed how to set profit maximizing prices based on an economic model with a downward sloping “demand function”. To find optimal prices, we need to know / estimate a demand function for ice cream.
- **Data:** We got a historical data set of prices and output. Interestingly, prices are positively correlated with quantity sold.
- **Model of data generating process:** We wrote down and studied a simple model of how the data could have been generated. Similar to our decision model, we assume the model has a downward sloping demand function and prices have been set in a profit maximizing fashion. Importantly, there are also random demand shocks ε (incorporating conditions like weather) that have been known by the price setter.

What have we done so far?

- **Simulation:** To get better intuition about the data generating process, we simulated it in R.
  - We found that the model can indeed generate a positive correlation between equilibrium prices and output, even though the demand function is downward sloping. The reason is that positive demand shocks increase both prices and output.
  - We also simulated an alternative model in which prices are set randomly.
- **Estimation:** We estimated a linear regression of our demand function with the simulated data.
  - When prices were set randomly, it looked as if we had a consistent estimator of the true parameters of the demand function
  - When prices were set in a profit maximizing fashion, our estimator of the demand function was inconsistent.

What will we do next?

1. We will review some basic concepts of econometrics, focusing on the linear regession model

- Key idea in econometrics: there is a true model with random variables that generated the data.
- Econometric tools, like estimation procedures or statistical tests, only make sense if the model satisfies certain conditions.
- We study some important tools and concepts.

2. We will study instrumental variable (IV) estimation

- IV estimation is a very prominent method to consistently estimate coefficients in cases in which OLS estimation does not work.
- IV estimation will allow us to consistently estimate the demand function for our example data set, in which prices have not been set randomly.