Empirical Industrial Organization & Consumer Choice

1a Introduction: Data, Model, Simulation and Estimation

Summer 2016

Prof. Sebastian Kranz, Uni Ulm

6. April 2016

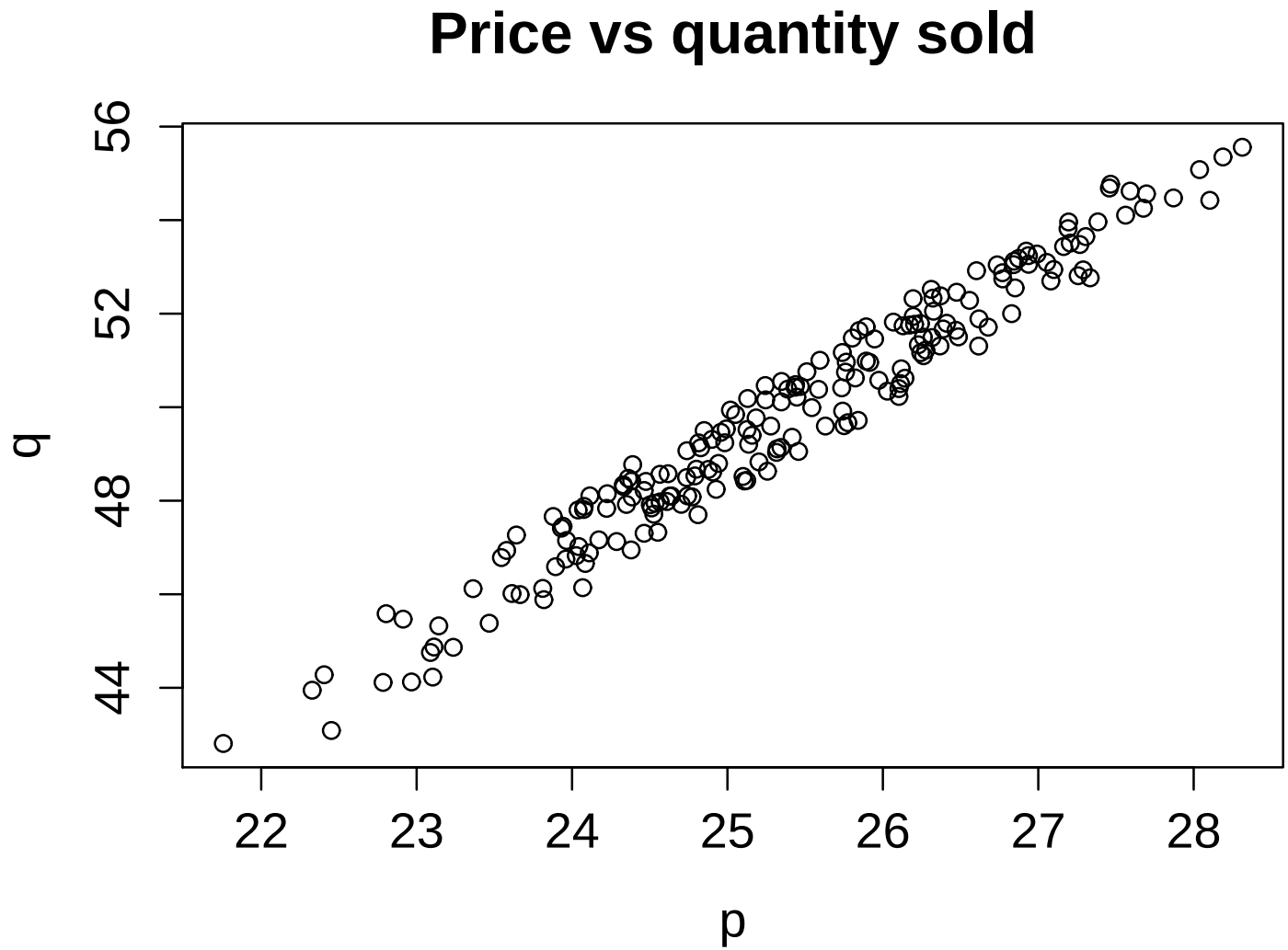
Bob's Ice Cream Business

- 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 ball 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 balls 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



- *Start a new script in RStudio. Load the data set "ice cream sales.csv" (available from MOODLE) into a data frame and replicate the plot of prices vs output.*
- Puzzled, Bob shows you the price-output diagram and asks you for advice:
  - *Do you think Emma understood the ice-cream market and set her prices in a profit maximizing fashion?*
  - *What explanations do you have for the shown positive relationship between prices and sales?*
  - *How would you proceed now to find a rule for setting profit-maximizing prices?*

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

- Assume the sales in period t are 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. The size of the demand shock shall not be influenced by the price.
  - *Which real-world factors could determine the size of the demand shock?*
  - b and a_0 are exogenously given numerical parameters with

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

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

- The firm's profits shall be given by

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

  - $c_t \geq 0$ is a constant marginal cost of production in period t (cost of one ball of ice cream)
- Assume the firm knows in each period t the demand shock ε_t and the parameters $\bar{a}$ and b sales in period t and chooses a price p_t that maximizes its expected profits.
- *Find a formula for the chosen profit-maximizing price p_t and the equilibrium output q_t*
- *Could this model in principle have generated the price-output relationship plotted by Bob?*

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 $[c_{min}, c_{max}]$
  - 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 increase the standard deviation of the demand shocks σ_ε? Do you have an intuition for the result?*

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 curve.
- The demand function describes how, ceteris paribus, higher prices *cause* lower quantities sold. For our goal to set profit maximizing prices, it is 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
- *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*

$$q_t = \beta_0 + \beta_1 p_t + \varepsilon_t$$

  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.