MKT 440 Pricing Analytics

Takeaki Sunada[1]

[1]Simon Business School

University of Rochester

What is pricing analytics?

- Pricing analytics = "apply data analytic tools to help finding profit-maximizing prices".
- But how does a data help us setting prices? i.e. What do we need to measure from a data, in order to set the correct price?

What do we need to know to price a product?

- Consider a simple case where your firm produces a single product, and you want to set its price.
- Econ001 tells us that your profit can be expressed as
-
-

What do we need to know to price a product?

- Consider a simple case where your firm produces a single product, and you want to set its price.
- Econ001 tells us that your profit can be expressed as

$$\textbf{Profit} = \textbf{Revenue} - \textbf{Cost}.$$

- Moreover, revenue is nothing but

What do we need to know to price a product?

- Consider a simple case where your firm produces a single product, and you want to set its price.
- Econ001 tells us that your profit can be expressed as

$$\textbf{Profit} = \textbf{Revenue - Cost.}$$

- Moreover, revenue is nothing but

$$\textbf{Revenue} = \textbf{Price} \times \textbf{Quantity sold}.$$

- So at a very high level, we need to know "quantity sold" and "cost".

What do we need to know to price a product?

- Often, cost has little to do with data analytics.
- Cost may be a complicated object, but as a manager you likely know exactly what each element is. Hence in this course, we take the cost side as known.

What do we need to know to price a product?

Quantity sold.

- Hence in order to set prices, we need to know quantity sold *at each price we could set*.
- Something like:
  1. A can of coke, priced at one dollar, sells one million units.
  2. If it is 1.5 dollars, it sells 600,000 units.
- Then we can calculate the revenue for every possible price, compare profits (don't forget the cost!) and pick the price that maximizes it.

We learn how to estimate demand using data.

- The objective of this course is hence stated as such - "use data analytics to estimate *the quantity sold at every possible price we could set*".
- I call this "relationship between the price set and the quantity sold" **demand**.
- The term "demand" hence refers to an entire price-quantity relationship. "Knowing demand" means "we know how many units we can sell at every single price point".

Wait... how about targeted pricing?

- So far we have assumed that we set one price to everyone.
- We know that segmentation (targeted pricing) can be an effective pricing tool.
- In that case,

$$\textbf{Revenue} = \textbf{Price (segment 1)} \times \textbf{Quantity sold (segment 1)} + \textbf{Price (segment 2)} \times \textbf{Quantity sold (segment 2)} + \ldots.$$

- Now we want to recover demand *for each segment*. But the main idea remains the same.

Wait... how about multi-product firm?

- You may set prices of multiple products.
- In that case,

$$\textbf{Revenue} = \textbf{Price (product 1)} \times \textbf{Quantity sold (product 1)}$$
$$+ \textbf{Price (product 2)} \times \textbf{Quantity sold (product 2)} + \ldots.$$

- Now we want to recover demand *for each product*. Note that demand for product 1 may also depend on price of product 2 (and vice versa). But the main idea remains the same.

We learn how to estimate demand using data

So, we need to know the demand to set the right price.

We learn how to estimate demand using data

Next: how does a data help us estimate (predict) demand?

We learn how to estimate demand using data.

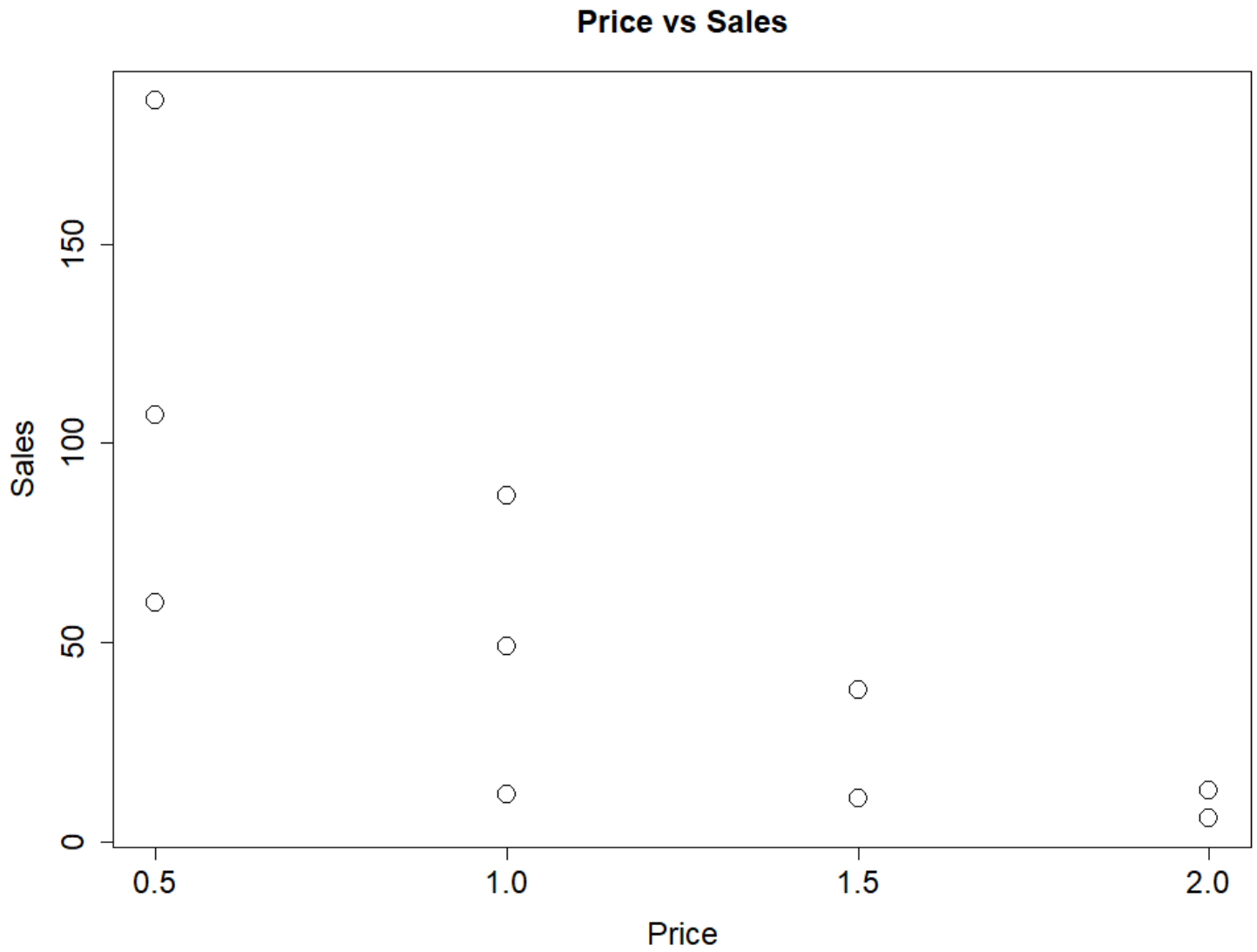
- Suppose that we are offering a single product, and we have data on its past sales and prices.
- In the data, we see "quantity sold" at each of the "prices" we set in the past - don't we know the "demand" already? Why do we need to estimate it?

A typical structure of sales data

	A	B	C
1	month	sales	price
2	1	11	1.5
3	2	8	1
4	3	107	0.5
5	4	7	1.5
6	5	13	2
7	6	186	0.5
8	7	12	1
9	8	60	0.5
10	9	6	2
11	10	49	1

simulate_example_data

Descriptive analysis



Data and model

- In general, data, in itself, is utterly incomplete.
- Data never lets you see the quantity sold at *every* possible price. It only tells you sales at *some* specific prices.
- Even at observed points, the data is noisy, imperfect and subject to selection. Even when we observe e.g. $P = 1$ was associated with $Q = 50$ in one time in the past, it doesn't mean that if we set $P = 1$ tomorrow the sales will be $Q = 50$.
- This is why I have been calling our task as "estimating" demand. We need to "fill in" the data imperfection.

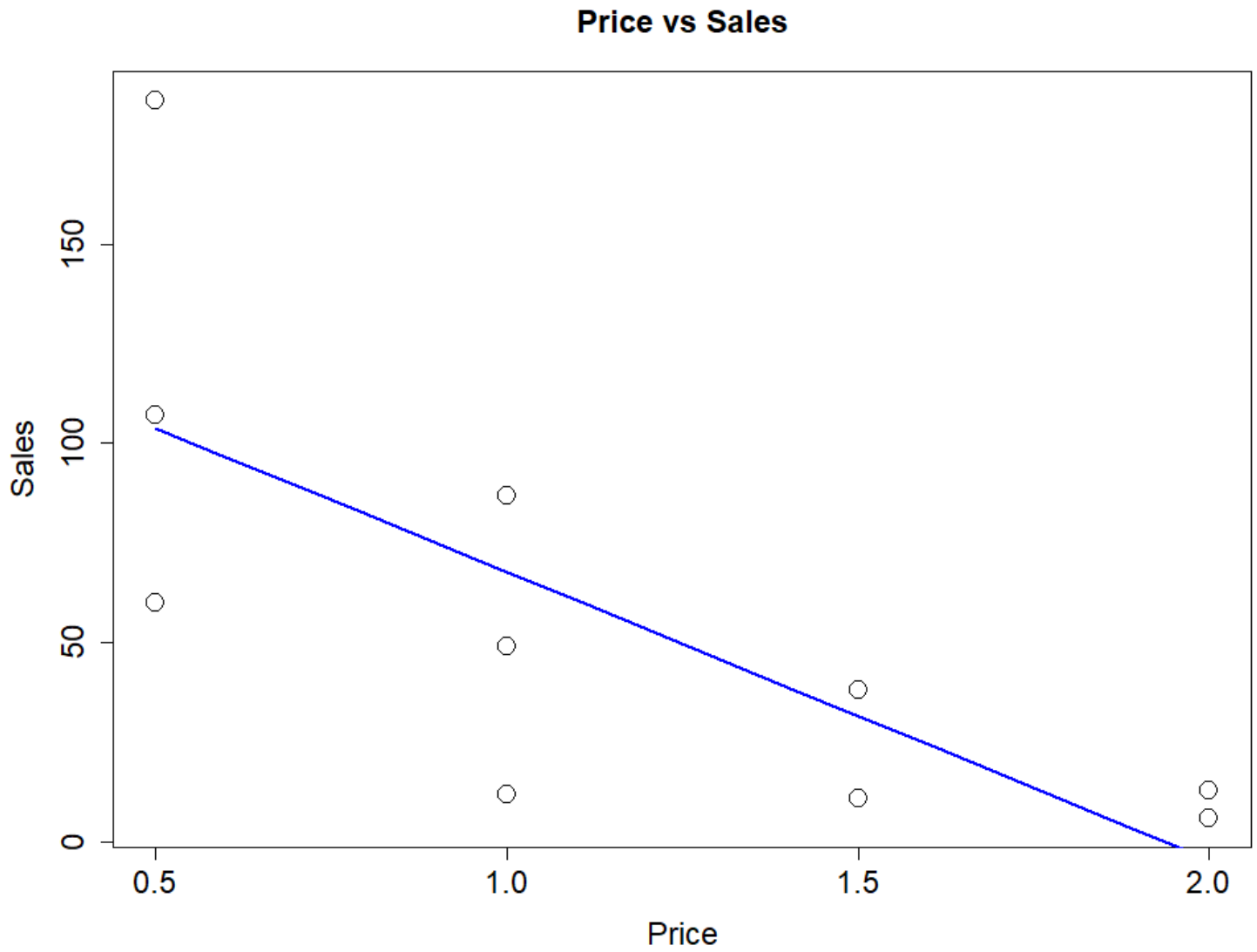
A regression model

- Think about a way to fill in the gap that we already know - a regression.
- By regressing Q on P, we have a fitted value of Q.

$$Q = \beta_0 + \beta_1 P + \epsilon$$

- Let's estimate this by OLS using the data we have.

A regression model



Data and model

- A regression line, defined by $Q = \beta_0 + \beta_1 P + \epsilon$, estimated with observed sales data, now allows us to estimate demand at prices not in the data.
- Moreover, at price points observed in the data, it reduces idiosyncratic noise involved in each observation and takes us closer to the "true" sales associated with the price.

Data and model

- We call by "model", the set of restrictions we impose on the relationship between observed variables (in this case, Q and P) to make up for the data incompleteness.
- In the regression case, the data is missing "observation of Q at some P", and the regression imposes the assumption that "there is a linear relationship between Q and P" to supplement it.
- In that sense, a regression is a "model". We just applied a regression model to a sales data to recover demand.

Data analytics is all about combining data and model

- Data analytics, including price analytics we study, is all about supplementing missing data with an appropriate model (assumptions) to obtain insights we want.
- In other words, the model puts restrictions on "how the missing part should look like" and thereby supplements observations.
- As the restrictions are merely our assumptions, no guarantee that they are correct. However, as there are many ways to impose restrictions (i.e. many different models), there are "more right" ones and "less right" ones.
- Which one is "more right" depends on many factors...

Data analytics is all about combining data and model

- What model (restrictions) we want is determined by
  1. nature of the environment we study
  2. types of data we have
  3. managerial objectives

What environment we study

- Sometimes, assuming linear demand and use regression may just work fine.
- But what if you want to study demand for a house? Consumers buy at most one unit at a time. What does "linearity" mean?
- What model to use (i.e. what assumptions to impose) hence depends on the product you study.

What data we have

- At virtually any firm, you should have access to a data on past history of sales and prices. We call it "sales data".
- If you are lucky, you may have access to a richer data, tracking individual consumer's purchase history. This type of data is called "choice data".
- Choice data with long window may let you see dynamics of repeat purchases. Other times, you may have access to consumer demographics.
- In order to fully extract information that your data have, you need to choose the right model that fits the data structure best.

What managerial questions we are after

- Simple questions (e.g. set one price for a single product) require a simple model and less data.
- Difficult questions (e.g. how to segment market based on consumers' latent types) require a rich data and a more complicated model.
- If you use a very simple model to study a complicated question, you don't get very far.
- If you use a super-complicated model to study a simple question, it would probably do, but may take too much time. Also you are likely imposing "unnecessary" assumptions.

Data analytics is all about combining data and model

- In coming weeks, we cover various kinds of
  - models available (specifically, regression models and various choice models)
  - data sets that you may possibly encounter in the future (sales data, choice data with long panel and short panel, consumer demographic data, etc)
  - managerial objectives regarding pricing (setting uniform price, segmentation, etc)
- and you learn how to combine them to get the best outcome.

Or you could say you learn how to build a right model

- Once in a firm, you may have no control about managerial objectives (your boss tells you what you do).
- The data set may also be just given to you - you just do the analysis using an existing data set.
- From that point of view, our objective can be rephrased as "how to build a best model for a given question and data limitation".

Summary of course objectives

- We study how to recover demand by combining modeling assumptions with available data.
- Data is incomplete, so we need an appropriate set of assumptions (models) to fill in the gap.
- The reliability of estimated demand rests on our ability to use the right model to study the right data set with the right managerial question.

Goals

- In weeks 1 and 2, we study "sales data" using "regression models". Sales data is never an ideal data. Nevertheless you will likely have to work with it given its universal availability.
- Starting week 3, we move on to "choice data", which records individual consumer's history of choices. We develop "choice models" to extract full information off of it. In particular, we discuss ways to find market segmentation.
- At the end of the course, we aim that *you are able to utilize a data on consumers' past purchase history to find latent segments in the market, and price/position products accordingly.*.

Prequalifications

- Basic R skills are required (levels equivalent to GBA462R or GBA436/GBA436R – e.g. load an excel sheet, run a regression and interpret results without issues).
- Concepts from MKT 414 (Pricing Policy) will also be very helpful (though you can complete this course without it).
- If you'd like a reference book for R, my recommendations are available on blackboard.

Course organization: lecture

- In-class sessions are mostly lecture. This course puts more emphasis on quantitative analysis - we don't do much case discussions. We instead learn skills that are applicable to any cases.
- In general, decks of lecture slides are organized topic by topic, rather than session by session. We may cover one deck of slides over multiple sessions, or multiple slide decks within a single session.

Course organization: projects

- You will practice actual implementations through projects. Much of the assignments is built around applying the analytic tools covered in the class to working with data.
- You never learn computational skills unless you implement them by yourself.
- Hence take-home projects are not "tests". Rather they are the purpose of the class. Expect to spend some amount of time on them.

Course organization: online video

- Online content mainly covers advanced materials and project solutions. Watching videos is not necessary to understand the materials in the following week, but of course you are expected to watch all videos by the end of the course.

Grading

- 25% on three in-class quizzes
  - Dates: Jan 31 (Tue), Feb 14 (Tue) and Feb 21 (Tue). We use the first 10-15 minutes of the session. Please bring your laptop.
  - Two short-answer questions (three sentences max) in each quiz - totaling six questions. Each question count toward 5% of your final grade, excluding your lowest-grade question.
  - General range of coverage is "previous two weeks". If there's any exception I will announce it in the class in the previous week.
  - The quizzes are open book, but you don't have much time to check the materials anyway, so expect that they are essentially closed book.

Grading

- 30% on two projects
  - The first covers sales data and is due on Jan 27 (Fri). The second covers choice data and is due on Feb 10 (Fri). Both projects are already posted on Blackboard.
  - For projects, please form a team of 2-5 people by next week (target date: Jan 17). You can mix with evening session participants.
- 35% on final exam.
  - Location and time TBD (during final exam week). It is about 2-hour long and 50-50 mix of conceptual questions and R coding questions.

Grading

- 10% on attendance
  - We will use Qwickly check in. The first date of attendance recording is Jan 17 (Tue).
  - I will post the PIN on the screen at the beginning of the session (perhaps also at the end too, TBD).
  - You will receive full attendance score if you miss up to one session. If you are absent from more than one session, your score will be reduced proportionally.
  - Valid leave of absence is allowed without penalty (job interview, doctor's appointment, etc., ideally with proof). In particular, If you must stay home due to COVID-related restrictions, don't come to class.

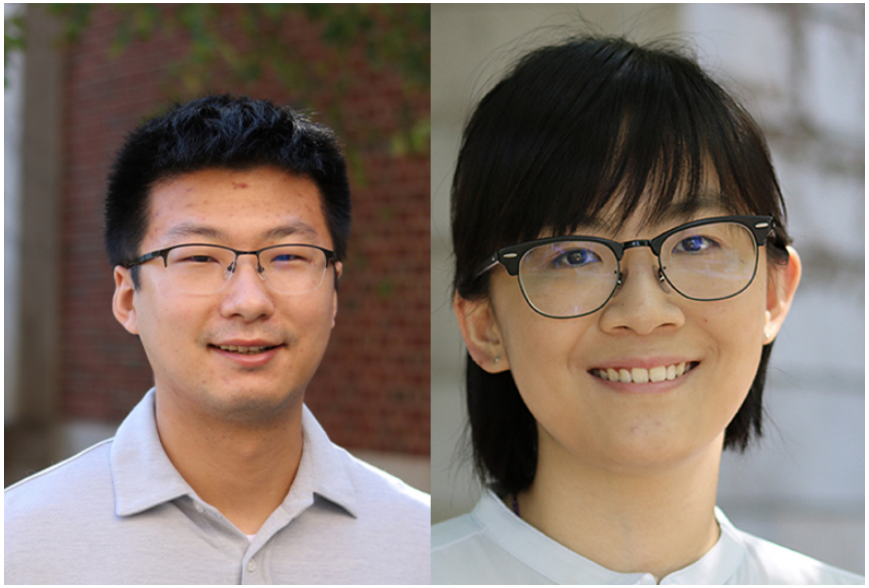
Grading

- Modified grading formula for remote participants:
  - All quizzes and the final exam will take place online. But that's the only difference from the in-person participants: you will take all quizzes and the exam at the same time and see the exact same questions as the in-person participants.
  - Attendance grades won't apply to remote participants. Instead, we will inflate your weight for projects and the final exam by 5% (so your weight is 25% quiz, 35% projects and 40% final).
- Same grading curve for everyone:
  - All MS, MBA, offline and online participants are graded on the same curve.

Office hours

- My office hours: Tue, 12:30 pm - 2:20 pm at Carol Simon Hall 3-110R and on Zoom (same as class Zoom link - link on blackboard).

TAs



- Xiaojie Li: Office hours Fri 2:00 - 3:30pm at Gleason G124 and on Zoom
- Ziyao Tang: Office hours Wed 4:30 - 6:00pm, at Gleason G124 and on Zoom
- All office hours start from week 2 (not offered this week).

TAs

- There will be two more TAs in charge of logistics (details TBD).
- TAs' email address is available on the syllabus and on blackboard.
- We will offer extra office hours near project due dates - see syllabus and blackboard announcement about a week before the deadline.

Miscellaneous

- More details about each content presented here is available in the syllabus. Note that the course schedule listed therein is tentative - we may go ahead of (or behind) the schedule as we move on. I will update the syllabus if there are any changes.
- All lecture materials, class recordings and announcements are posted on Blackboard. Please check it regularly.