

ECON 6120I Topics in Empirical Industrial Organization

Kohei Kawaguchi

Last updated: 2019-02-24

Contents

1	Syllabus	5
1.1	Instructor Information	5
1.2	General Information	5
1.3	Required Environment	6
1.4	Evaluation	6
1.5	Academic Integrity	7
1.6	Schedule	7
1.7	Course Materials	7
2	Introduction	9
2.1	Structural Estimation and Counterfactual Analysis	9
2.2	Setting Up The Environment	14
3	Production and Cost Function Estimation	21
3.1	Motivations	21
3.2	Analyzing Producer Behaviors	22
3.3	Production Function Estimation	23
3.4	Cost Function Estimation	35
4	Demand Function Estimation	41
4.1	Motivations	41
4.2	Analyzing Consumer Behaviors	42
4.3	Continuous Choice	42
4.4	Discrete Choice	46
5	Assignment 1: Basic Programming in R	65
5.1	Simulate data	65
5.2	Estimate the parameter	67
6	Assignment 2: Production Function Estimation	69
6.1	Simulate data	69
6.2	Estimate the parameters	72

Chapter 1

Syllabus

1.1 Instructor Information

- Instructor:
 - Name: Kohei Kawaguchi.
 - Office: LSK6070, Monday 11:00-12:00.
- All questions related to the class have to be publicly asked on the discussion board of canvas rather than being privately asked in e-mail. The instructor usually does not reply in the evening, weekends, and holidays.

1.2 General Information

1.2.1 Class Time

- Date: Monday.
- Time: 13:30-17:20.
- Venue: CYTG001.

1.2.2 Description

- This is a PhD-level course for empirical industrial organization. This course covers various econometric methods used in industrial organization that is often referred to as the structural estimation approach. These methods have been gradually developed since 1980s in parallel with the modernization of industrial organization based on the game theory and now widely applied in antitrust policy, business strategy, and neighboring fields such as labor economics and international economics.
- This course presumes a good understanding of PhD-level microeconomics and microeconometrics. Participants are expected to understand at least UG-level industrial organization. This course requires participants to write programs mostly in R and sometimes in C++ to implement various econometric methods. In particular, all assignments will involve such a non-trivial programming task. Even though the understanding of these programming languages is not a prerequisite, a sharp learning curve will be required. Some experience in other programming languages will help. Audit without a credit is not admitted for students.

1.2.3 Expectation and Goals

- The goal of this course is to learn and practice econometric methods for empirical industrial organization. The lecture covers the econometric methods that have been developed between 80s and 00s to estimate primitive parameters governing imperfect competition among firms, such as production and cost function estimation, demand function estimation, merger simulation, entry and exit analysis, and dynamic decision models. The lecture also covers various new methods to recover model primitives in certain mechanisms such as auction, matching, network, and bargaining. The emphasis is put on the former group of methods, because they are the basis for other new methods. Participants will not only understand the theoretical background of the methods but also become able to implement these methods from scratches by writing their own programs. I will briefly discuss the latter class of new methods through reading recent papers. The participants will become able to understand and use these new methods.

1.3 Required Environment

- Participants should bring their laptop to the class. We have enough extension codes for students. The laptop should have sufficient RAM (at least $\geq 8\text{GB}$, $\geq 16\text{GB}$ is recommended) and CPU power (at least Core i5 grade, Core i7 grade is recommended). Participants are fully responsible for their hardware issues. Operating System can be arbitrary. The instructor mainly uses OSX High Sierra with iMac (Retina 5K, 27-inch, Late 2015) and Macbook Pro (Retina, 15-inch, Early 2017).
- Please install the following software by the first lecture. Technical issues related to the installment should be resolved by yourself, because it depends on your local environment. If you had an error, copy and paste the error message on a search engine, and find a solution. This solves 99.9% of the problems.
 - R: <https://www.r-project.org/>
 - RStudio: <https://www.rstudio.com/>
 - LaTeX:
 - * MixTex <https://miktex.org/>
 - * TeXLive <https://www.tug.org/texlive/>
 - * MacTeX <http://www.tug.org/mactex/>

1.4 Evaluation

- Assignments (80): In total 8 homework are assigned. Each homework involves the implementation of the methods covered in the class. Each homework has 10 points. The working hour for each homework will be around 10-20 hours.
- Participation (10): Every time a participant asks a question in the class, after the class, during the office hour, or in the canvas. the participant gets one point, up to 10 points. The participant who asked the question writes the name, ID number, his/her question, and my answer in a discussion board on the course website to claim a point.
- Referee report (10): Toward the end of the semester, a paper in industrial organization is randomly assigned to each participant. Each participant writes a critical referee report of the assigned paper in A4 2 pages that consists of the summary, contribution, strong and weak points of the paper.
- Grading is based on the absolute scores: A+ with more than 80 points, A with more than 70 points, A- with more than 60 points, B+ with more than 50 points, B with more than 40 points, B- with more than 30 points and C otherwise.

1.5 Academic Integrity

Without academic integrity, there is no serious learning. Thus you are expected to hold the highest standard of academic integrity in the course. You are encouraged to study and do homework in groups. However, no cheating, plagiarism will be tolerated. Anyone caught cheating, plagiarism will fail the course. Please make sure adhere to the HKUST Academic Honor Code at all time (see <http://www.ust.hk/vpao/integrity/>).

1.6 Schedule

- Introduction to structural estimation, R and RStudio
- Production function estimation I
- Production function estimation II
- Demand function estimation I
- Demand function estimation II
- Merger Analysis
- Entry and Exit
- Single-Agent Dynamics I
- Single-Agent Dynamics II, I change date due to my business trip
- Dynamic Game I
- Dynamic Game II
- Auction
- Other Mechanisms

1.7 Course Materials

1.7.1 R and RStudio

- Golemund, G., 2014, Hands-On Programming with R, O'Reilly.
 - Free online version is available: <https://rstudio-education.github.io/hopr/>.
- Wickham, H., & Golemund, G., 2017, R for Data Science, O'Reilly.
 - Free online version is available: <https://r4ds.had.co.nz/>.
- Boswell, D., & Foucher, T., 2011, The Art of Readable Code: Simple and Practical Techniques for Writing Better Code, O'Reilly.

1.7.2 Handbook Chapters

- Akerberg, D., Benkard, C., Berry, S., & Pakes, A. (2007). “Econometric tools for analyzing market outcomes”. Handbook of econometrics, 6, 4171-4276.
- Athey, S., & Haile, P. A. (2007). “Nonparametric approaches to auctions”. Handbook of Econometrics, 6, 3847-3965.
- Berry, S., & Reiss, P. (2007). “Empirical models of entry and market structure”. Handbook of Industrial Organization, 3, 1845-1886.
- Bresnahan, T. F. (1989). “Empirical studies of industries with market power”. Handbook of Industrial Organization, 2, 1011-1057.
- Hendricks, K., & Porter, R. H. (2007). “An empirical perspective on auctions”. Handbook of Industrial Organization, 3, 2073-2143.
- Matzkin, R. L. (2007). “Nonparametric identification”. Handbook of Econometrics, 6, 5307-5368.
- Newey, W. K., & McFadden, D. (1994). “Large sample estimation and hypothesis testing”. Handbook of Econometrics, 4, 2111-2245.

- Reiss, P. C., & Wolak, F. A. (2007). “Structural econometric modeling: Rationales and examples from industrial organization”. Handbook of Econometrics, 6, 4277-4415.

1.7.3 Books

- Train, K. E. (2009). Discrete Choice Methods with Simulation, Cambridge university press.
- Davis, P., & Garces, E. (2010). Quantitative Techniques for Competition and Antitrust Analysis, Princeton University Press.
- Tirole, J. (1988). The Theory of Industrial Organization, The MIT Press.

1.7.4 Papers

- The list of important papers are occasionally given during the course.

Chapter 2

Introduction

2.1 Structural Estimation and Counterfactual Analysis

2.1.1 Example

- Igami [2017] “Estimating the Innovator’s Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981-1998”.
- **Question:**
- Does “Innovator’s Dilemma” [Christensen, 1997] or the delay of innovation among incumbents exist?
- Christensen argued that old winners tend to lag behind entrants even when introducing a new technology is not too difficult, with a case study of the HDD industry.
 - Apple’s smartphone vs. Nokia’s feature phones.
 - Amazon vs. Borders.
 - Kodak’s digital camera.
- If it exists, what is the reason for that?
- How do we empirically answer this question?
- **Hypotheses:**
- Identify potentially competing hypotheses to explain the phenomenon.
 1. Cannibalization: Because of cannibalization, the benefits of introducing a new product are smaller for incumbents than for entrants.
 2. Different costs: The incumbents may have higher costs for innovation due to the organizational inertia, but at the same time they may have some cost advantage due to accumulated R&D and better financial access.
 3. Preemption: The incumbents have additional incentive for innovation to preempt potential rivals.
 4. Institutional environment: The impacts of the three components differ across different institutional contexts such as the rules governing patents and market size.
- Casual empiricists pick up their favorite factors to make up a story.
- Serious empiricists should try to separate the contributions of each factor from data.
- To do so, the author develops an economic model that explicitly incorporates the above mentioned factors, while keeping the model parameters flexible enough to let the data tell the sign and size of the effects of each factor on innovation.



FIG. 1.—The incumbent-entrant innovation gap. The top panel plots the timing of the initial shipment of 3.5-inch HDDs separately for incumbents (i.e., firms already active in the 5.25-inch generation) and entrants (i.e., firms that appeared for the first time as producers of 3.5-inch HDDs). The bottom panel expresses similar numbers in terms of the fractions of all firms that could have innovated. See the text for details.

Figure 2.1: Figure 1 of Igam (2017)

- **Economic model:**

- The time is discrete with finite horizon $t = 1, \dots, T$.
- In each year, there is a finite number of firms indexed by i .
- Each firm is in one of the technological states:

$$s_{it} \in \{\text{old only, both, new only, potential entrant}\}, \quad (2.1)$$

where the first two states are for incumbents (stick to the old technology or start using the new technology) and the last two states are for actual and potential entrants (enter with the new technology or stay outside the market).

- In each year:
 - Pre-innovation incumbent ($s_{it} = \text{old}$): exit or innovate by paying a sunk cost κ^{inc} (to be $s_{i,t+1} = \text{both}$).
 - Post-innovation incumbent ($s_{it} = \text{both}$): exit or stay to be both.
 - Potential entrant ($s_{it} = \text{potential entrant}$): give up entry or enter with the new technology by paying a sunk cost κ^{net} (to be $s_{i,t+1} = \text{new}$).
 - Actual entrant ($s_{it} = \text{new}$): exit or stay to be new.
- Given the industry state $s_t = \{s_{it}\}_i$, the product market competition opens and the profit of firm i , $\pi_t(s_{it}, s_{-it})$, is realized for each active firm.
- As the product market competition closes:
 - Pre-innovation incumbents draw private cost shocks and make decisions: a_t^{pre} .
 - Observing this, post-innovation incumbents draw private cost shocks and make decisions: a_t^{post} .
 - Observing this, actual entrants draw private cost shocks and make decisions: a_t^{act} .
 - Observing this, potential entrants draw private cost shocks and make decisions: a_t^{pot} .
- This is a dynamic game. The equilibrium is defined by the concept of **Markov-perfect equilibrium** [Maskin and Tirole, 1988].
- The representation of the competing theories in the model:
 - The existence of cannibalization is represented by the assumption that an incumbent maximizes the joint profits of old and new technology products.
 - The size of cannibalization is captured by the shape of profit function.
 - The difference in the cost of innovation is captured by the difference in the sunk costs of innovation.
 - The preemptive incentive for incumbents are embodied in the dynamic optimization problem for each incumbent.
- **Econometric model:**
- The author then turns the economic model into an econometric model.
- This amounts to specify which part of the economic model is observed/known and which part is unobserved/unknown.
- The author collects the data set of the HDD industry during 1977-99.
- Based on the data, the author specify the identities of active firms and their products and the technologies embodied in the products in each year to code their **state variables**.
- Moreover, by tracking the change in the state, the author code their **action variables**.
- Thus, the state and action variables, s_t and a_t . These are the **observables**.
- The author does not observe:

- Profit function $\pi_t(\cdot)$.
- Sunk cost of innovation for pre-innovation incumbents κ^{inc} .
- Sunk cost of entry for potential entrants κ^{net} .
- Private cost shocks.
- These are the **unobservables**.
- Among the unobservables, the profit function and sunk costs are the **parameter of interests** and the private cost shocks are **nuisance parameters** in the sense only the knowledge about the distribution of the latter is demanded.
- **Identification:**
- Can we infer the unobservables from the observables and the restrictions on the distribution of observable by the economic theory?
- The profit function is identified from estimating the demand function for each firm's product, and estimating the cost function for each firm from using their price setting behavior.
- The sunk costs of innovation are identified from the conditional probability of innovation across various states. If the cost is low, the probability should be high.
- **Estimation:**
- The identification established that in principle we can uncover the parameters of interests from observables under the restrictions of economic theory.
- Finally, we apply a statistical method to the econometric model and infer the parameters of interest.
- **Counterfactual analysis:**
- If we can uncover the parameters of interest, we can conduct **comparative statics**: study the change in the endogenous variables when the exogenous variables including the model parameters are set different. In the current framework, this exercise is often called the **counterfactual analysis**.
- What if there was no cannibalization?:
 - An incumbents separately maximizes the profit from old technology and new technology instead of jointly maximizing the profits. Solve the model under this new assumption everything else being equal.
 - Free of cannibalization concerns, 8.95 incumbents start producing new HDDs in the first 10 years, compared with 6.30 in the baseline.
 - The cumulative numbers of innovators among incumbents and entrants differ only by 2.8 compared with 6.45 in the baseline.
 - Thus cannibalization can explain a significant part of the incumbent-entrant innovation gap.
- What if there was no preemption?:
 - A potential entrant ignores the incumbents' innovations upon making entry decisions.
 - Without the preemptive motives, only 6.02 incumbents would innovate in the first 10 years, compared with 6.30 in the baseline.
 - The cumulative incumbent-entrant innovation gap widen to 8.91 compared with 6.45 in the baseline.
- The sunk cost of entry is smaller for incumbents than for entrants in the baseline.
- **Interpretations and policy/managerial implication:**
- Despite the cost advantage and the preemptive motives, the speed of innovation is slower among incumbents due to the strong cannibalization effect.
- Incumbents that attempt to avoid the “innovator's dilemma” should separate the decision makings between old and new sections inside the organization so that it can avoid the concern for cannibalization.

2.1.2 Recap

- The structural approach in empirical industrial organization consists of the following components:
 1. Research question.
 2. Competing hypotheses.
 3. Economic model.
 4. Econometric model
 5. Identification.
 6. Data collection.
 7. Data cleaning.
 8. Estimation.
 9. Counterfactual analysis.
 10. Coding.
 11. Interpretations and policy/managerial implications.
- The goal of this course is to be familiar with the standard methodology to complete this process.
- The methodology covered in this class is mostly developed to analyze the standard framework to dynamic or oligopoly competition.
- The policy implications are centered around competition policies.
- But the basic idea can be extend to different class of situations such as auction, matching, voting, contract, marketing, and so on.
- Note that the depth of the research question and the relevance of the policy/managerial implications are the most important part of the research.
- Focusing on the methodology in this class is to minimize the time to allocate to less important issues and maximize the attention and time to the most valuable part in the future research.
- Given a research question, what kind of data is necessary to answer the question?
- Given data, what kind of research questions can you address? Which question can be credibly answered? Which question can be an over-stretch?
- Given a research question and data, what is the best way to answer the question? What type of problem can you avoid using the method? What is the limitation of your approach? How will you defend the possible referee comments?
- Given a result, what kinds of interpretation can you credibly derive? What kinds of interpretation can be contested by potential opponents? What kinds of contribution can you claim?
- To address these issues is **necessary** to publish a paper and it is **necessary** to be familiar with the methodology to do so.

2.1.3 Historical Remark

- The words **reduced-form** and **structural-form** date back to the literature of estimation of simultaneous equations in macroeconomics [Hsiao, 1983].
- Let y be the vector of observed endogenous variables, x be the vector of observed exogenous variables, and ϵ be the vector of unobserved exogenous variables.
- The equilibrium condition for y on x and ϵ is often written as:

$$Ay + Bx = \Sigma\epsilon. \quad (2.2)$$

- These equations **implicitly** determine the vector of endogenous variables y .
- If A is invertible, we can solve the equations for y to obtain:

$$y = -A^{-1}Bx + A^{-1}\Sigma\epsilon. \quad (2.3)$$

- These equations **explicitly** determine the vector of endogenous variables y .
- Equation (2.2) is the **structural-form** and (2.3) is the **reduced-form**.

- If y and x are observed and x is of full column rank, then $A^{-1}B$ and $A^{-1}\Sigma A^{-1}$ will be estimated by regression for (2.3). But this does not mean that A , B and Σ are separately estimated.
- This was the traditional identification problems.
- Thus, reduced-form does not mean either of:
 - Regression analysis;
 - Statistical analysis free from economic assumptions.
- Recent development in this line of literature of identification is found in Matzkin [2007].
- In econometrics, the idea of imposing restrictions from economic theories seems to have been formalized by the work of Manski [1994] and Matzkin [1994].

2.2 Setting Up The Environment

- Assume that R, RStudio and LaTeX are all installed in the local computer.

2.2.1 RStudio Project

- The assignments should be conducted inside a project folder for this course.
- File > New Project...> New Directory > New Directory > R Package using RcppEigen.
- Name the directory ECON6120I and place in your favorite location.
- You can open this project from the upper right menu of RStudio or by double clicking the ECON6120I.Rproj file in the ECON6120I directory.
- This navigates you to the root directory of the project.
- In the root directory, make folders named:
 - assignment.
 - input.
 - output.
 - figuretable.
- We will store R functions in R folder, C/C++ functions in src folder, and data in input folder, data generated from the code in output, and figures and tables in figuretable folder.
- Open src/Makevars and erase the content. Then, write: `PKG_CPPFLAGS = -w -std=c++11 -O3`
- Open src/Makevars.win and erase the content. Then, write: `PKG_CPPFLAGS = -w -std=c++11`

2.2.2 Basic Programming in R

- File > New File > R Script to open Untitled file.
- Ctrl (Cmd) + S to save it with test.R in assignment folder.
- In the console, type and push enter:

```
1 + 1
```

```
## [1] 2
```

```
100:130
```

```
## [1] 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116
```

```
## [18] 117 118 119 120 121 122 123 124 125 126 127 128 129 130
```

- This is the interactive way of using R functionalities.
- In test.R, write:

```
1 + 1
```

- Then, save the file and push Run.
- Alternatively, place the mouse over the `1 + 1` line in test.R file.

- Then, **Ctrl** (**Cmd**) + **Enter** to run the line.
- In this way, we can write procedures in the file and send to the console to run.
- There are functions to conduct basic calculations:

```
1 + 2
```

```
## [1] 3
```

```
2 * 3
```

```
## [1] 6
```

```
4 - 1
```

```
## [1] 3
```

```
6 / 2
```

```
## [1] 3
```

```
2^3
```

```
## [1] 8
```

- We can define objects and assign values to them.

```
a <- 1
```

```
a
```

```
## [1] 1
```

```
a + 2
```

```
## [1] 3
```

- In addition to scalar object, we can define a vector by:

```
2:10
```

```
## [1] 2 3 4 5 6 7 8 9 10
```

```
3:20
```

```
## [1] 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
```

```
c(2, 3, 5, 9, 10)
```

```
## [1] 2 3 5 9 10
```

```
seq(1, 10, 2)
```

```
## [1] 1 3 5 7 9
```

- `seq` is a function with initial value, end values, and the increment value.
- By typing `seq` in the `help`, we can read the manual page of the function.
- `seq {base}` means that this function is named `seq` and is contained in the library called `base`.
- Some libraries are automatically called when the R is launched, but some are not.
- Some libraries are even not installed.
- We can install a library from a repository called `CRAN`.

```
install.packages("ggplot2")
```

- To use the package, we have to load by:

```
library(ggplot2)
```

- Use `qplot` function in `ggplot2` library to draw a scatter plot.

```
x <- c(-1, -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.7, 1)
y <- x^3
qplot(x, y)
```



Instead of loading a package by `library`, you can directly call it as:

```
ggplot2::qplot(x, y)
```




- We can write own functions.

```
roll <- function(n) {
  die <- 1:6
  dice <- sample(die, size = n, replace = TRUE)
  y <- sum(dice)
  return(y)
}
roll(1)
```

```
## [1] 5
```

```
roll(2)
```

```
## [1] 9
```

```
roll(10)
```

```
## [1] 27
```

```
roll(10)
```

```
## [1] 41
```

- We can `set.seed` to obtain the same realization of random variables.

```
set.seed(1)
roll(10)
```

```
## [1] 38
```

```
set.seed(1)
roll(10)
```

```
## [1] 38
```

- When a variable used in a function is not given as its argument, the function calls the variable in the global environment:

```
y <- 1
plus_1 <- function(x) {
  return(x + y)
}
plus_1(1)
```

```
## [1] 2
```

```
plus_1(2)
```

```
## [1] 3
```

- However, you should NOT do this. All variables used in a function should be given as its arguments:

```
y <- 1
plus_2 <- function(x, y) {
  return(x + y)
}
plus_2(2, 3)
```

```
## [1] 5
```

```
plus_2(2, 4)
```

```
## [1] 6
```

- The best practice is to use `findGlobals` function in `codetools` to check global variables in a function:

```
library(codetools)
findGlobals(plus_1)
```

```
## [1] "{"      "+"      "return" "y"
```

```
findGlobals(plus_2)
```

```
## [1] "{"      "+"      "return"
```

- This function returns the list of global variables used in a function. If this returns a global variable other than the system global variables, you should include it as the argument of the function.
- You can write the functions in the files with executing codes.
- But I recommend you to separate files for writing functions and executing codes.
- File > New File > R Rscript and name it as `functions.R` and save to R folder.
- Cut the function you wrote and paste it in `functions.R`.
- There are two ways of calling a function in `functions.R` from `test.R`.
- One way is to use `source` function.

```
source("R/functions.R")
```

- When this line is read, the codes in the file are executed.
- The other way is to bundle functions as a package and load it.

- Choose **Build > Clean and Rebuild**.
- This compiles files in **src** folder and bundle functions in **R** folder and build a package named **ECON6120I**.
- Now, the functions in **R** folder and **src** folder can be used by loading the package by:

```
library(ECON6120I)
```

- Best practice:
 1. Write functions in the scratch file.
 2. As the functions are tested, move them to **R/functions.R**.
 3. Clean and rebuild and load them as a package.

2.2.3 Reproducible Reports using Rmarkdown

- Reporting in empirical studies involves:
 1. Writing texts;
 2. Writing formulas;
 3. Writing and implementing programs;
 4. Demonstrating the results with figures and tables.
- Moreover, this has to be done in a **reproducible** manner: Whoever can reproduce the output from the scratch.
- “Whoever” includes yourself in the future. Because the revision process of structural papers is usually lengthy, you often have to remember the content few weeks or few months later. It is inefficient if you cannot recall what you have done.
- We use **Rmarkdown** to achieve this goal.
- This assumes that you have LaTeX installed.
- Install package **Rmarkdown**:

```
install.packages("rmarkdown")
```

- **File > New File > R Markdown... > HTML with title Test**.
- Save it in **assignment** folder with name **test.Rmd**.
- From **Knit** tab, choose **Knit to HTML**.
- This outputs the content to html file.
- You can also choose **Knit to PDF** from **Knit** tab to obtain output in pdf file.
- Reports should be knit to pdf to submit.
- But you can use html output while writing a report because html is lighter to compile.
- Refer to the help page for further information.

Chapter 3

Production and Cost Function Estimation

3.1 Motivations

- Estimating **production and cost functions** of producers is the cornerstone of economic analysis.
- Estimating the functions includes to separate the contribution of observed inputs and the other factors, which is often referred to as the **productivity**.
- “What determines productivity?” [Syverson, 2011]-type research questions naturally follow.
- The methods covered in this chapter are widely used across different fields.
- Some of them are variants from the standard methods.

3.1.1 IO

- Olley and Pakes [1996]:
 - How much did the deregulation in the U.S. telecommunication industry, in particular the divestiture of AT&T in 1984, spurred the productivity growth of the incumbent, facilitated entries, and increased the aggregate productivity?
 - To do so, the authors estimate the plant-level production functions and productivity in the telecommunication industry.
- Doraszelski and Jaumandreu [2013]:
 - What is the role of R&D in determining the differences in productivity across firms and the evolution of firm-level productivity over time?
 - To do so, the authors estimate the firm-level production functions and productivity of Spanish manufacturing firms during 1990s in which the transition probability of a productivity is a function of the R&D activities.

3.1.2 Development

- Hsieh and Klenow [2009]:
 - How large is the misallocation of inputs across manufacturing firms in China and India compared to the U.S? How will the aggregate productivity of China and India change if the degree of misallocation is reduced to the U.S. level?
 - To do so, the authors measure the revenue productivity of firms, which should be the same across firms within an industry if there were no distortion, and the measurement of the revenue productivity requires to estimate the production function.

- Gennaioli et al. [2013]:
 - What are the determinants of regional growth? Do geographic, institutional, cultural, and human capital factors explain the difference across regions?
 - To do so, the authors construct the data set that covers 74% of the world’s surface and 97% of its GDP and estimate the production function in which the above mentioned factors could affect the productivity.

3.1.3 Trade

- Haskel et al. [2007]:
 - Are there spillovers from FDI to domestic firms?
 - To do so, the authors estimate the plant-level production function of the U.K. manufacturing firms during 1973 and 1992 and study how the foreign presence in the U.K. affected the productivity.
- De Loecker [2011]:
 - Does the removal of trade barriers induces efficiency gain for producers?
 - To do so, the author estimate the production functions of Belgian textile industry during 1994-2002 in which the degree of trade protection can affect the productivity level.

3.1.4 Management

- Bloom and Van Reenen [2007]:
 - How do management practices affect the firm productivity?
 - To do so, the authors first estimate the production function and productivity of manufacturing firms in developed countries, and then study how the independently measured management practices of the firms affect the estimated productivity.
- Braguinsky et al. [2015]:
 - How do changes in ownership affect the productivity and profitability of firms?
 - To do so, the authors estimate the production function for various outputs including the physical output, return on capital and labor, and the utilization rate, price level, using the cotton spinners data in Japan during 1896 and 1920.

3.1.5 Education

- Cunha et al. [2010]:
 - How do childhood and schooling interventions “produce” the cognitive and non-cognitive skills of children?
 - To do so, the authors estimate the mapping from childhood and schooling interventions to children’s cognitive and non-cognitive skills, the “production function” of childhood environment and education.

3.2 Analyzing Producer Behaviors

- There are several levels of parameters that govern the behavior of firms:
- **Production function**
 - Add factor market structure.
 - Add cost minimization.
- → **Cost function**
 - Add product market structure.

- Add profit maximization.
- → **Supply function (Pricing function)**
 - Combine cost and supply (pricing) functions.
- → **Profit function**
- Which parameter to identify?
- Primitive enough to be invariant to relevant policy changes.
 - e.g. If you conduct a policy experiment that changes the factor market structure, identifying cost functions is not enough.
- As reduced-form as possible among such specifications.
 - A reduced-form parameter usually can be rationalized by a class of underlying structural parameters and institutional assumptions. Thus, the analysis becomes robust to some misspecifications.
 - e.g. A non-parametric function $C(q, w)$ can represent a cost function of a producer who is not necessarily minimizing the cost. If we derive a cost function from a production function and a factor market structure, then the cost function cannot represent such a non-optimization behavior.

3.3 Production Function Estimation

3.3.1 Cobb-Douglas Specification as a Benchmark

- Most of the following argument carries over to a general model.
- For firm $j = 1, \dots, J$ and time $t = 1, \dots, T$, we observe output Y_{jt} , labor L_{jt} , and capital K_{jt} .
- We consider an asymptotic of $J \rightarrow \infty$ for a fixed T .
- Assume Cobb-Douglas production function:

$$Y_{jt} = A_{jt} L_{jt}^{\beta_l} K_{jt}^{\beta_k}, \quad (3.1)$$

where A_{jt} is firm j and time t specific unobserved heterogeneity in the model.

- Taking the logarithm gives:

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \epsilon_{jt}, \quad (3.2)$$

where lowercase symbols represent natural logs of variables and $\ln(A_{jt}) = \beta_0 + \epsilon_{jt}$.

- This can be regarded as a first-order log-linear approximation of a production function.
- Linear regression model! May OLS work?

3.3.2 Potential Bias I: Endogeneity

- ϵ_{jt} contains everything that cannot be explained by the observed inputs: better capital may be employed, a worker may have obtained better skills, etc.
- When the manager of a firm makes an input choice, she should have some information about the realization of ϵ_{jt} .
- Thus, the input choice can be correlated with ϵ_{jt} ; for example under static optimization of L_{jt} given K_{jt} :

$$L_{jt} = \left[\frac{p_{jt}}{w_{jt}} \beta_l \exp^{\beta_0 + \epsilon_{jt}} K_{jt}^{\beta_k} \right]^{\frac{1}{1-\beta_l}}. \quad (3.3)$$

- In this case, OLS estimator for β_l is *positively* biased, because when ϵ_{jt} is high, l_{jt} is high and thus the increase in output caused by ϵ_{jt} is captured as if caused by the increase in labor input.

- The endogeneity problem was already recognized by Marschak and Andrews [1944].

3.3.3 Potential Bias II: Selection

- Firms freely enter and exit market.
- Therefore, a firm that had low ϵ_{jt} is likely to exit.
- However, if firms have high capital K_{jt} , it can stay in the market even if the realization of ϵ_{jt} is very low.
- Therefore, conditional on being in the market, there is a *negative* correlation between the capital K_{jt} and ϵ_{jt} .
- This problem occurs even if the choice of K_{jt} itself is not a function of ϵ_{jt} .

3.3.4 How to Resolve Endogeneity Bias?

- Temporarily abstract away from entry and exit.
 - The data is balanced.
1. Panel data.
 2. First-order condition for inputs.
 3. Instrumental variable.
 4. Olley-Pakes approach and its followers/critics.
- Griliches and Mairesse [1998] is a good survey of the history up to Olley-Pakes approach.
 - Akerberg et al. [2015] also offer a good survey and clarify problems and implicit assumptions in Olley-Pakes approach.

3.3.5 Panel Data

- Assume that $\epsilon_{jt} = \mu_j + \eta_{jt}$, where η_{jt} is uncorrelated with input choices up to period t :

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \mu_j + \eta_{jt}. \quad (3.4)$$

- Then, by differentiating period t and $t - 1$ equations, we get:

$$y_{jt} - y_{j,t-1} = \beta_l(l_{jt} - l_{j,t-1}) + \beta_k(k_{jt} - k_{j,t-1}) + (\eta_{jt} - \eta_{j,t-1}). \quad (3.5)$$

- Then, because $\eta_{jt} - \eta_{j,t-1}$ is uncorrelated either with $l_{jt} - l_{j,t-1}$ or $k_{jt} - k_{j,t-1}$, we can identify the parameter.
- Problem:
 - Restrictive heterogeneity.
 - When there are measurement errors, fixed-effect estimator can generate higher biases than OLS estimator, because measurement errors more likely to survive first-difference and within-transformation.

3.3.6 First-Order Condition for Inputs

- Use the first-order condition for inputs as the moment condition [McElroy, 1987].
- Closely related to the cost function estimation literature.
- Need to specify the factor market structure and the nature of the optimization problem for a firm.
- Recently being center of attention again as one of the solutions to the “collinearity problem” discussed below.

3.3.7 Instrumental Variable

- Borrow the idea from the first-order condition approach that the input choices are affected by some exogenous variables.
- If we have instrumental variables that affect inputs but are uncorrelated with errors ϵ_{jt} , then we can identify the parameter by an instrumental variable method.
- One candidate for the instrumental variables: **input prices**.
- Input price affect input decision.
- Input price is not correlated with ϵ_{jt} if the factor product market is competitive and ϵ_{jt} is an idiosyncratic shock to a firm.
- Problems:
 - Input prices often lack cross-sectional variation.
 - Cross-sectional variation is often due to unobserved input quality.
- Another candidate for the instrumental variables: **lagged inputs**.
- If ϵ_{jt} does not have auto-correlation, lagged inputs are not correlated with the current shock.
- If there are adjustment costs for inputs, then lagged inputs are correlated with the current inputs.
- Problem:
 - If ϵ_{jt} has auto-correlation, all lagged inputs are correlated with the errors: For example, if ϵ_{jt} is AR(1), $\epsilon_{jt} = \alpha\epsilon_{j,t-1} + \nu_{j,t-1} = \dots \alpha^l \epsilon_{j,t-l} + \nu_{j,t-l} + \dots, \alpha^{l-1} \nu_{j,t-l}$ for any l .

3.3.8 Olley-Pakes Approach

- Exploit restrictions from the economic theory [Olley and Pakes, 1996].
- Write $\epsilon_{jt} = \omega_{jt} + \eta_{jt}$, where ω_{jt} is an anticipated shock and η_{jt} is an ex-post shock.
- Inputs are correlated with ω_{jt} but not with η_{jt}
- The model is written as:

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt} + \eta_{jt}. \quad (3.6)$$

- OP use economic theory to derive a valid proxy for the anticipated shock ω_{jt} .

3.3.9 Assumption I: Information Set

- The firm's information set at t , I_{jt} , includes current and past productivity shocks $\{\omega_{j\tau}\}_{\tau=0}^t$ but does not include future productivity shocks $\{\omega_{j\tau}\}_{\tau=t+1}^{\infty}$.
- The transitory shocks η_{jt} satisfy $\mathbb{E}\{\eta_{jt}|I_{jt}\} = 0$.

3.3.10 Assumption II: First Order Markov

- Productivity shocks evolve according to the distribution:

$$p(\omega_{j,t+1}|I_{jt}) = p(\omega_{j,t+1}|\omega_{jt}), \quad (3.7)$$

and the distribution is known to firms and stochastically increasing in ω_{jt} .

- Then:

$$\omega_{jt} = \mathbb{E}\{\omega_{jt}|\omega_{j,t-1}\} + \nu_{jt}, \quad (3.8)$$

and:

$$\mathbb{E}\{\nu_{jt}|I_{j,t-1}\} = 0, \quad (3.9)$$

by construction.

3.3.11 Assumption III: Timing of Input Choices

- Firms accumulate capital according to:

$$k_{jt} = \kappa(k_{j,t-1}, i_{j,t-1}), \quad (3.10)$$

where investment $i_{j,t-1}$ is chosen in period $t - 1$.

- Labor input l_{jt} is non-dynamic and chosen at t .
- This assumption characterizes and distinguishes labor and capital.
- Intuitively, it takes a full period for new capital to be ordered, delivered, and installed.

3.3.12 Assumption IV: Scalar Unobservable

- Firms' investment decisions are given by:

$$i_{jt} = f_t(k_{jt}, \omega_{jt}). \quad (3.11)$$

- This assumption places strong implicit restrictions on additional firm-specific unobservables.
 - No **across firm** unobserved heterogeneity in adjustment cost of capital, in demand and labor market conditions, or in other parts of the production function.
 - Okay with **across time** unobserved heterogeneity.

3.3.13 Assumption IV: Strict Monotonicity

- The investment policy function $f_t(k_{jt}, \omega_{jt})$ is strictly increasing in ω_{jt} .
- This holds if the realization of higher ω_{jt} implies higher expectation for future productivity (Assumption III) and if the marginal product of capital is increasing in the expectation for future productivity.
- To verify the latter condition in a given game is often not easy.

3.3.14 Two-step Approach: The First Step

- Insert $\omega_{jt} = h(k_{jt}, i_{jt})$ to the original equation to get:

$$\begin{aligned} y_{jt} &= \beta_l l_{jt} + \underbrace{\beta_0 + \beta_k k_{jt} + h(k_{jt}, i_{jt})}_{\text{unknown function of } k_{jt} \text{ and } i_{jt}} + \eta_{jt} \\ &\equiv \beta_l l_{jt} + \phi(k_{jt}, i_{jt}) + \eta_{jt}. \end{aligned} \quad (3.12)$$

- This is a **partially linear model**: see Ichimura and Todd [2007] for reference.

- Because l_{jt}, k_{jt} and i_{jt} are uncorrelated with η_{jt} , we can identify β_l and $\phi(\cdot)$ by exploiting the moment condition:

$$\begin{aligned}\mathbb{E}\{\eta_{jt}|l_{jt}, k_{jt}, i_{jt}\} &= 0 \\ \Leftrightarrow \mathbb{E}\{y_{jt} - \beta_l l_{jt} - \phi(k_{jt}, i_{jt})|l_{jt}, k_{jt}, i_{jt}\} &= 0.\end{aligned}\tag{3.13}$$

if there is enough variation in l_{jt}, k_{jt} and i_{jt} .

- This “if there is enough variation” part is actually problematic. Discuss later.
- Let β_l^0 and ϕ^0 be the identified true parameters.

3.3.15 Two-step Approach: The Second Step

- Note that:

$$\omega_{jt} \equiv \phi(k_{jt}, i_{jt}) - \beta_0 - \beta_k k_{jt}.\tag{3.14}$$

- Therefore, we have:

$$\begin{aligned}y_{jt} - \beta_l^0 l_{jt} &= \beta_0 + \beta_k k_{jt} + \omega_{jt} + \eta_{jt} \\ &= \beta_0 + \beta_k k_{jt} + g(\omega_{j,t-1}) + \nu_{jt} + \eta_{jt} \\ &= \beta_0 + \beta_k k_{jt} + g[\phi^0(k_{j,t-1}, i_{j,t-1}) - (\beta_0 + \beta_k k_{j,t-1})] + \nu_{jt} + \eta_{jt}.\end{aligned}\tag{3.15}$$

- ν_{jt} and η_{jt} are independent of the covariates.
- This is a **multiple-index model** with indices $\beta_0 + \beta_1 k_{jt}$ and $\beta_0 + \beta_1 k_{j,t-1}$ where parameters of two indices are restricted to be the same: see Ichimura and Todd [2007] for reference.
- We can identify β_0, β_k and g by exploiting the moment condition:

$$\begin{aligned}\mathbb{E}\{\nu_{jt} + \eta_{jt}|k_{jt}, k_{j,t-1}, i_{j,t-1}\} &= 0 \\ \Leftrightarrow \mathbb{E}\{\nu_{jt} + \eta_{jt}|k_{jt}, k_{j,t-1}, i_{j,t-1}\} &= 0.\end{aligned}\tag{3.16}$$

3.3.16 Identification of the Anticipated Shocks

- If ϕ, β_0, β_k are identified, then ω_{jt} is also identified by:

$$\omega_{jt} \equiv \phi(k_{jt}, i_{jt}) - \beta_0 - \beta_k k_{jt}.\tag{3.17}$$

3.3.17 Two-Step Estimation of Olley and Pakes [1996].

- **First step:** Estimate β_L and ϕ in :

$$y_{jt} = \beta_l l_{jt} + \phi(k_{jt}, i_{jt}) + \eta_{jt}.\tag{3.18}$$

by approximating ϕ with some basis functions, say, polynomials or splines:

$$\begin{aligned}y_{jt} &= \beta_l l_{jt} + \sum_{p=1}^P \gamma_p \phi_p(k_{jt}, i_{jt}) + \left[\phi(k_{jt}, i_{jt}) - \sum_{n=1}^N \gamma_n \phi_n(k_{jt}, i_{jt}) \right] + \eta_{jt} \\ &= \beta_l l_{jt} + \sum_{p=1}^P \gamma_p \phi_p(k_{jt}, i_{jt}) + \tilde{\eta}_{jt}\end{aligned}\tag{3.19}$$

where $P \rightarrow \infty$ when the sample size goes to infinity.

- e.g. second-order polynomial approximation:

$$\begin{aligned}\phi_1(k_{jt}, i_{jt}) &= k_{jt}, \phi_2(k_{jt}, i_{jt}) = i_{jt} \\ \phi_3(k_{jt}, i_{jt}) &= k_{jt}^2, \phi_4(k_{jt}, i_{jt}) = i_{jt}^2 \\ \phi_5(k_{jt}, i_{jt}) &= k_{jt}i_{jt}.\end{aligned}\tag{3.20}$$

- Once the basis functions are fixed, estimation is the same as the linear model.
- But the inference (the computation of the standard deviation) is difference, because of the approximation error.
- See Chen [2007] for reference.
- Let $\hat{\beta}_l$ and $\hat{\phi}$ be the estimates from the first step.
- **Second step:** Estimate β_0 , β_k , and g in:

$$\begin{aligned}y_{jt} - \hat{\beta}_l l_{jt} &= \beta_0 + \beta_k k_{jt} + g[\hat{\phi}(k_{j,t-1}, i_{j,t-1}) - (\beta_0 + \beta_k k_{j,t-1})] + \nu_{jt} + \eta_{jt} \\ &\quad + [\beta_l - \hat{\beta}_l] l_{jt} \\ &\quad + \left\{ g[\phi(k_{j,t-1}, i_{j,t-1}) - (\beta_0 + \beta_k k_{j,t-1})] - g[\hat{\phi}(k_{j,t-1}, i_{j,t-1}) - (\beta_0 + \beta_k k_{j,t-1})] \right\} \\ &= \beta_0 + \beta_k k_{jt} + g[\hat{\phi}(k_{j,t-1}, i_{j,t-1}) - (\beta_0 + \beta_k k_{j,t-1})] + \nu_{jt} + \tilde{\eta}_{jt}\end{aligned}\tag{3.21}$$

by approximating g by some basis functions, say, polynomials or splines.

3.3.18 From An Economic Models to An Econometric Model

- Starting from economic model with some unobserved heterogeneity, we reach some reduced-form model.
- If the resulting model belongs to a class of econometric models whose identification and estimation are established, we can simply apply the existing methods.

3.3.19 How to Resolve Selection Bias

- Use propensity score to correct selection bias: Ahn and Powell [1993].
- At the beginning of period t , after observing ω_{jt} , firm j decides whether to continue the business ($\chi_{jt} = 1$) or exit ($\chi_{jt} = 0$).
- Assume that the difference between continuation and exit values is strictly increasing in ω_{jt} .
- Then, there is a threshold $\underline{\omega}(k_{jt})$ such that:

$$\chi_{jt} = \begin{cases} 1 & \text{if } \omega_{jt} \geq \underline{\omega}(k_{jt}) \\ 0 & \text{otherwise.} \end{cases}\tag{3.22}$$

- We can only observe firms that satisfy $\chi_{jt} = 1$.

3.3.20 Correction in the First Step

- In the first step, we need no correction because:

$$\begin{aligned}\mathbb{E}\{y_{jt} | l_{jt}, k_{jt}, i_{jt}, \chi_{jt} = 1\} \\ &= \beta_l l_{jt} + \phi(k_{jt}, i_{jt}) + \mathbb{E}\{\eta_{jt} | \chi_{jt} = 1\} \\ &= \beta_l l_{jt} + \phi(k_{jt}, i_{jt}).\end{aligned}\tag{3.23}$$

- Ex-post shock η_{jt} is independent of continuation/exit decision. Therefore, we can identify β_l and $\phi(\cdot)$ as in the previous case.

3.3.21 Correction in the Second Step I: The Source of Bias

- One the other hand, we need correction in the second step, because:

$$\begin{aligned}
& \mathbb{E}\{y_{jt} - \beta_l^0 l_{jt} | k_{jt}, i_{jt}, k_{j,t-1}, l_{j,t-1}, \chi_{jt} = 1\} \\
&= \beta_0 + \beta_k k_{jt} + g[\phi^0(k_{jt}, i_{jt}) - (\beta_0 + \beta_k k_{jt})] \\
&+ \mathbb{E}\{\nu_{jt} + \eta_{jt} | k_{jt}, i_{jt}, k_{j,t-1}, l_{j,t-1}, \chi_{jt} = 1\} \\
&= \beta_0 + \beta_k k_{jt} + g[\phi^0(k_{j,t-1}, i_{j,t-1}) - (\beta_0 + \beta_k k_{j,t-1})] \\
&+ \mathbb{E}\{\nu_{jt} | k_{jt}, i_{jt}, k_{j,t-1}, l_{j,t-1}, \chi_{jt} = 1\}.
\end{aligned} \tag{3.24}$$

and

$$\mathbb{E}\{\nu_{jt} | k_{jt}, i_{jt}, k_{j,t-1}, l_{j,t-1}, \chi_{jt} = 1\} \neq 0, \tag{3.25}$$

since anticipated shock matters continuation/exit decision in period t .

3.3.22 Correction in the Second Step II: Conditional Exit Probability

- Let's see that the conditional expectation:

$$\begin{aligned}
& \mathbb{E}\{\omega_{jt} | k_{jt}, i_{jt}, k_{j,t-1}, l_{j,t-1}, \chi_{jt} = 1\} \\
&= \mathbb{E}\{\omega_{jt} | k_{jt}, i_{jt}, k_{j,t-1}, l_{j,t-1}, \omega_{jt} \geq \underline{\omega}(k_{jt})\} \\
&= \int_{\underline{\omega}(k_{jt})} \omega_{jt} \frac{p(\omega_{jt} | \omega_{j,t-1})}{\int_{\underline{\omega}(k_{jt})} p(\omega | \omega_{j,t-1}) d\omega} d\omega_{jt} \\
&\equiv \tilde{g}(\omega_{j,t-1}, \underline{\omega}(k_{jt})),
\end{aligned} \tag{3.26}$$

is a function of $\omega_{j,t-1}$ and $\underline{\omega}(k_{jt})$.

3.3.23 Correction in the Second Step III: Invertibility in Threshold

- The propensity of continuation conditional on observed information up to period $t - 1$:

$$\begin{aligned}
P_{jt} &\equiv \mathbb{P}\{\chi_{jt} = 1 | \mathcal{I}_{j,t-1}\} \\
&= \mathbb{P}\{\omega_{jt} \geq \underline{\omega}(k_{jt}) | \mathcal{I}_{j,t-1}\} \\
&= \mathbb{P}\{g(\omega_{j,t-1}) + \nu_{jt} \geq \underline{\omega}[(1 - \delta)k_{j,t-1} + i_{j,t-1}] | \mathcal{I}_{j,t-1}\} \\
&= \mathbb{P}\{\chi_{jt} = 1 | i_{j,t-1}, k_{j,t-1}\}.
\end{aligned} \tag{3.27}$$

- \rightarrow It suffices to condition on $i_{j,t-1}, k_{j,t-1}$.
- We also have:

$$P_{jt} = \mathbb{P}\{\chi_{jt} = 1 | \omega_{j,t-1}, \underline{\omega}(k_{jt})\}, \tag{3.28}$$

and it is invertible in $\underline{\omega}(k_{jt})$, that is,

$$\underline{\omega}(k_{jt}) \equiv \psi(P_{jt}, \omega_{j,t-1}). \tag{3.29}$$

3.3.24 Correction in the Second Step IV: Controlling the Threshold

- Now, he have:

$$\begin{aligned}
& \mathbb{E}\{y_{jt} - \beta_l^0 l_{jt} | k_{jt}, i_{jt}, k_{j,t-1}, l_{j,t-1}, \chi_{jt} = 1\} \\
&= \beta_0 + \beta_k k_{jt} + \mathbb{E}\{\omega_{jt} | k_{jt}, i_{jt}, k_{j,t-1}, l_{j,t-1}, \chi_{jt} = 1\} \\
&= \beta_0 + \beta_k k_{jt} + \tilde{g}(\omega_{j,t-1}, \underline{\omega}(k_{jt})) \\
&= \beta_0 + \beta_k k_{jt} + \tilde{g}(\omega_{j,t-1}, \psi(P_{jt}, \omega_{j,t-1})) \\
&\equiv \beta_0 + \beta_k k_{jt} + \tilde{g}(\omega_{j,t-1}, P_{jt}) \\
&= \beta_0 + \beta_k k_{jt} + \tilde{\phi}^0(k_{j,t-1}, i_{j,t-1}) - (\beta_0 + \beta_k k_{j,t-1}), P_{jt}.
\end{aligned} \tag{3.30}$$

- At the end, the only difference is to include P_{jt} as a covariate.
- P_{jt} is a **known** function of $i_{j,t-1}$ and $k_{j,t-1}$.
- Even if we condition on $P_{jt} = p$, there are still many combinations of $i_{j,t-1}$ and $k_{j,t-1}$ that gives $P_{jt} = p$.
- With this remaining variation, we can identify β_0 , β_k , and \tilde{g} by the same argument as the case without selection, for each $P_{jt} = p$.

3.3.25 Three Step Estimation of Olley and Pakes [1996]

- **Zero step:** Estimate the propensity score:

$$P_{jt} = 1\{\chi_{jt} = 1 | i_{j,t-1}, k_{j,t-1}\}, \tag{3.31}$$

by a kernel estimator.

- Insert the resulting estimates \hat{P}_{jt} into the first and second steps.

3.3.26 Zero Investment Problem

- One of the key assumptions in OP method was invertibility between anticipated shock and investment:

$$\omega_{jt} = i^{-1}(k_{jt}, i_{jt}) \equiv h(k_{jt}, i_{jt}). \tag{3.32}$$

- However, in micro data, zero investment is a rule rather than exceptions.
- Then, the invertibility does not hold globally: there are some region of the anticipated shock in which the investment takes value zero.

3.3.27 Tackle Zero Investment Problem I: Discard Some Data

- Discard a data (j, t) such that $i_{j,t-1} = 0$.
- Use a data (j, t) such that $i_{j,t-1} > 0$.
- Then, invertibility recovers on this selected sample.
- This **does not** cause bias in the estimator because ν_{jt} in :

$$\beta_0 + \beta_l k_{jt} + g[\phi^0(k_{j,t-1}, i_{j,t-1}) - (\beta_0 + \beta_k k_{j,t-1})] + \nu_{jt} + \eta_{jt}, \tag{3.33}$$

is independent of the event up to $t - 1$, including $i_{j,t-1}$.

- However, this **does** cause information loss. The loss is high if the proportion of the sample such that $i_{j,t-1} = 0$ is high.

3.3.28 Tackle Zero Investment Problem II: Use Another Proxy

- Investment is just a possible proxy for the anticipated shock.
- Intermediate inputs can be used as proxies as well [Levinsohn and Petrin, 2003].
- The problem is that these intermediate inputs are included in the gross production function, whereas investment is excluded.
- Let m_{jt} be the log material input, and assume that the production function takes the form of:

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \beta_m m_{jt} + \omega_{jt} + \eta_{jt}. \quad (3.34)$$

- In addition, assume that the **optimal policy function** for m_{jt} is strictly monotonic in the ex-ante shock, and hence is invertible:

$$m_{jt} = m(k_{jt}, \omega_{jt}) \Leftrightarrow \omega_{jt} = m^{-1}(m_{jt}, k_{jt}) \equiv h(m_{jt}, k_{jt}). \quad (3.35)$$

- **First step:**

$$\begin{aligned} y_{jt} &= \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \beta_m m_{jt} + h(m_{jt}, k_{jt}) + \eta_{jt} \\ &= \beta_l l_{jt} + \phi(m_{jt}, k_{jt}) + \eta_{jt}. \end{aligned} \quad (3.36)$$

- We can identify β_l and ϕ by exploiting the moment condition:

$$\begin{aligned} \mathbb{E}\{\eta_{jt} | l_{jt}, m_{jt}, k_{jt}, i_{jt}\} &= 0 \\ \Leftrightarrow \mathbb{E}\{y_{jt} - \beta_0 - \beta_l l_{jt} - \phi(m_{jt}, k_{jt}) | l_{jt}, m_{jt}, k_{jt}\} &= 0, \end{aligned} \quad (3.37)$$

if there is enough variation in l_{jt}, m_{jt}, k_{jt} .

- **Second step:**

$$\begin{aligned} y_{jt} - \beta_l^0 l_{jt} &= \beta_k k_{jt} + \beta_m m_{jt} + g[\phi^0(m_{j,t-1}, k_{j,t-1}) - \beta_k k_{j,t-1} - \beta_m m_{j,t-1}] \\ &\quad + \nu_{jt} + \eta_{jt}. \end{aligned} \quad (3.38)$$

- We can identify β_k , β_m , and g by exploiting the moment condition:

$$\mathbb{E}\{\nu_{jt} + \eta_{jt} | k_{jt}, m_{j,t-1}, k_{j,t-1}\} = 0. \quad (3.39)$$

- Because m_{jt} is correlated with ν_{jt} , the moment should not condition on m_{jt} .
- The identification of β_m comes from $\beta_m m_{j,t-1}$.

3.3.29 One-step Estimation of Olley and Pakes [1996] and Levinsohn and Petrin [2003]

- Levinsohn and Petrin [2003] can be estimated in the similar two-step method.
- We can jointly estimate the parameters in first and second steps to improve the efficiency [Wooldridge, 2009].
- We estimate under the assumptions of Olley and Pakes [1996]:

$$y_{jt} = \beta_0 + \beta_1 l_{jt} + \beta_k k_{jt} + \omega_{jt} + \eta_{jt}. \quad (3.40)$$

- The first step exploits the following moment:

$$\mathbb{E}\{\eta_{jt} | l_{jt}, k_{jt}, i_{jt}\} = 0, \quad (3.41)$$

that is:

$$\mathbb{E}\{y_{jt} - \beta_1 l_{jt} - \beta_0 - \beta_k k_{jt} - \omega(k_{jt}, i_{jt}) | l_{jt}, k_{jt}, i_{jt}\} = 0. \quad (3.42)$$

- We can reinforce the moment condition as:

$$\mathbb{E}\{\eta_{jt} | l_{jt}, k_{jt}, i_{jt}, \dots, l_{j1}, k_{j1}, i_{j1}\} = 0 \quad (3.43)$$

if we assume that lagged inputs are correlated with the current inputs and η_{jt} is independent.

- The second step exploits the following moment:

$$\mathbb{E}\{\nu_{jt} | k_{jt}, i_{j,t-1}, l_{j,t-1}\} = 0, \quad (3.44)$$

that is:

$$\mathbb{E}\{y_{jt} - \beta_0 - \beta_1 l_{jt} - \beta_k k_{jt} - g[\omega(k_{j,t-1}, i_{j,t-1})] | k_{jt}, i_{j,t-1}, l_{j,t-1}\} = 0. \quad (3.45)$$

- We can reinforce the moment condition as:

$$\mathbb{E}\{\nu_{jt} | k_{jt}, i_{j,t-1}, l_{j,t-1}, \dots, k_{j1}, i_{j1}, l_{j1}\} = 0, \quad (3.46)$$

if we assume that lagged input are correlated with the current inputs and $\nu_{jt} + \eta_{jt}$ are independent.

- We can construct a GMM estimator based on equations (3.42) and (3.45).
- The one-step estimator can be more efficient but can be computationally heavier than the two-step estimator.

3.3.30 Scalar Unobservable Problem: Finite-order Markov Process

- Borrow the idea of using the first-order condition to resolve the collinearity problem [Gandhi et al., 2017].
- We have assumed that anticipated shocks follow a first-order Markov process:

$$\omega_{jt} = g(\omega_{j,t-1}) + \nu_{jt}. \quad (3.47)$$

- However, it may be true that it has more than one lags, for example:

$$\omega_{jt} = g(\omega_{j,t-1}, \omega_{j,t-2}) + \nu_{jt}. \quad (3.48)$$

- Then, we need proxies as many as the number of unobservables:

$$\begin{pmatrix} i_{jt} \\ m_{jt} \end{pmatrix} = \Gamma(k_{jt}, \omega_{jt}, \omega_{j,t-1}), \quad (3.49)$$

such that the policy function for the proxies is a bijection in $(\omega_{jt}, \omega_{j,t-1})$.

- Then, we can have:

$$\omega_{jt} = \Gamma_1^{-1}(k_{jt}, i_{jt}, m_{jt}). \quad (3.50)$$

- The reminder goes as in the standard OP method.

3.3.31 Scalar Unobservable Problem: Demand and Productivity Shocks

- There may be a demand shock μ_{jt} that also follows first-order Markov process.
- Then, the policy function depend both on μ_{jt} and ω_{jt} .
- We again need proxies as many as the number of unobservable.
- Suppose that we can observe the price of the firm p_{jt} .

- Inverting the policy function:

$$\begin{pmatrix} i_{jt} \\ p_{jt} \end{pmatrix} = \Gamma(k_{jt}, \omega_{jt}, \mu_{jt}). \quad (3.51)$$

yields:

$$\omega_{jt} = \Gamma_1^{-1}(k_{jt}, i_{jt}, p_{jt}). \quad (3.52)$$

- If ω_{jt} only depends on $\omega_{j,t-1}$ but not on $\mu_{j,t-1}$, then the second step of the modified OP method is to estimate:

$$\begin{aligned} y_{jt} - \hat{\beta}_l l_{jt} &= \beta_0 + \beta_k k_{jt} \\ &+ g(\omega_{j,t-1}) + \nu_{jt} + \eta_{jt} \\ &= \beta_0 + \beta_k k_{jt} \\ &+ g(\hat{\phi}_{j,t-1} - \beta_0 - \beta_k k_{j,t-1}) + \nu_{jt} + \eta_{jt}. \end{aligned} \quad (3.53)$$

- It goes as in the standard OP method.
- If ω_{jt} depends both on $\omega_{j,t-1}$ and $\mu_{j,t-1}$, the second step regression equation will be:

$$\begin{aligned} y_{jt} - \hat{\beta}_l l_{jt} &= \beta_0 + \beta_k k_{jt} \\ &+ g(\omega_{j,t-1}, \mu_{j,t-1}) + \nu_{jt} + \eta_{jt} \\ &= \beta_0 + \beta_k k_{jt} \\ &+ g(\hat{\phi}_{j,t-1} - \beta_0 - \beta_k k_{j,t-1}, \mu_{j,t-1}) + \nu_{jt} + \eta_{jt}. \end{aligned} \quad (3.54)$$

- We still have to control $\mu_{j,t-1}$ in the second step.
- Invert the policy function for $\mu_{j,t-1}$ to get:

$$\mu_{j,t-1} = \Gamma_2^{-1}(k_{j,t-1}, i_{j,t-1}, p_{j,t-1}), \quad (3.55)$$

and plug it into the second step regression equation to get:

$$\begin{aligned} y_{jt} - \hat{\beta}_l l_{jt} &= \beta_0 + \beta_k k_{jt} \\ &+ g(\hat{\phi}_{j,t-1} - \beta_0 - \beta_k k_{j,t-1}, \Gamma_2^{-1}(k_{j,t-1}, i_{j,t-1}, p_{j,t-1})) + \nu_{jt} + \eta_{jt}. \end{aligned} \quad (3.56)$$

- The parameters β_0 and β_k **cannot** be identified only with this observation, because Γ_2^{-1} is **unknown non-parametric** function: it can mean any function of $(k_{j,t-1}, i_{j,t-1}, p_{j,t-1})$.
- To estimate such a model, we jointly estimate the demand function along with the production function.
- At this point, we do not investigate it further because we have not yet learned how to estimate the demand function.
- For now just keep in mind that:
 - There has to be as many proxies as the dimension of the unobservable state variables.
 - It is okay that the unobservable state variable includes a demand shock.
 - It can be problematic when the unobservable demand shock affect the evolution of the anticipated productivity shock.

3.3.32 Collinearity Problem

- The collinearity problem is formally pointed out by Akerberg et al. [2015].
- This paper is finally published in 2015, but has been circulated since 2005.

- We assumed that k_{jt} and ω_{jt} are state variables.
- Then the policy function for labor input should take the form of:

$$l_{jt} = l(k_{jt}, \omega_{jt}). \quad (3.57)$$

- However, because $\omega_{jt} = h(i_{jt}, k_{jt})$, we have:

$$l_{jt} = l[k_{jt}, h(i_{jt}, k_{jt})] = \tilde{l}(i_{jt}, k_{jt}). \quad (3.58)$$

- Therefore, in the first stage, we encounter a multicollinearity problem:

$$\begin{aligned} y_{jt} &= \beta_0 + \beta_l \tilde{l}(i_{jt}, k_{jt}) + \phi(i_{jt}, k_{jt}) + \eta_{jt} \\ &\equiv \tilde{\phi}(i_{jt}, k_{jt}). \end{aligned} \quad (3.59)$$

- Thus, β_l cannot be identified in the first step.
- The second step becomes:

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + g[\tilde{\phi}(i_{j,t-1}, k_{j,t-1}) - \beta_0 - \beta_l l_{j,t-1} - \beta_k k_{j,t-1}] + \nu_{jt} + \eta_{jt} \quad (3.60)$$

- Because l_{jt} is correlated with ν_{jt} , moment can only condition on $l_{j,t-1}$.
- However, conditioning on $k_{j,t-1}$ and $i_{j,t-1}$, again there is no remaining variation in $l_{j,t-1}$.
- Therefore, β_l cannot be identified either in the second step.
- β_l **cannot be identified!**

3.3.33 Tackle Collinearity Problem: Peculiar Assumptions

- To make Olley-Pakes/Levinsohn-Petrin approach workable, we need peculiar data generating process for l_{jt} .
 - Consider Levinsohn-Petrin framework.
1. There is an optimization error in l_{jt} .
 - If it is not i.i.d over time, it becomes a state variable and enters to the policy for m_{jt} , violating the scalar unobserved heterogeneity assumption of m_{jt} .
 - If there is an optimization error for m_{jt} , this again violates the scalar unobserved heterogeneity assumption.
 2. k_{jt} is realized, ω_{jt} is observed, m_{jt} and i_{jt} are determined, a new i.i.d. unexpected shock is observed, l_{jt} is determined, and η_{jt} is observed.
 - If it is not i.i.d over time, it becomes a state variable and enters to the policy for m_{jt} , violating the scalar unobserved heterogeneity assumption.
 3. k_{jt} is realized, an unexpected shock is observed, l_{jt} is determined, ω_{jt} is observed, m_{jt} and i_{jt} are determined, and η_{jt} is observed (Akerberg [2016] recommends this assumption).
 - In this case, the unexpected shock can be serially correlated, because it suffices to know k_{jt} , i_{jt} , l_{jt} to decide m_{jt} . It does not have to predict the future unexpected shock based on the realization of the current shock because m_{jt} is a static decision.
 - This changes the optimal policy function of m_{jt} (3.35) to:

$$m_{jt} = m(k_{jt}, \omega_{jt}, l_{jt}). \quad (3.61)$$

- The first step:

$$\begin{aligned} y_{jt} &= \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + h(k_{jt}, m_{jt}, l_{jt}) + \eta_{jt} \\ &= \psi(k_{jt}, m_{jt}, l_{jt}) + \eta_{jt} \\ &\Rightarrow \mathbb{E}\{y_{jt} - \psi(k_{jt}, m_{jt}, l_{jt}) | k_{jt}, m_{jt}, l_{jt}\} = 0. \end{aligned} \quad (3.62)$$

- The second step:

$$\begin{aligned}
y_{jt} &= \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + g[\psi(k_{j,t-1}, m_{j,t-1}, l_{j,t-1}) - \beta_0 - \beta_l l_{j,t-1} - \beta_k k_{j,t-1}] + \nu_{jt} + \eta_{jt} \\
&\Rightarrow \mathbb{E}\{y_{jt} - \beta_0 - \beta_l l_{jt} - \beta_k k_{jt} - g[\psi(k_{j,t-1}, m_{j,t-1}, l_{j,t-1}) - \beta_0 - \beta_l l_{j,t-1} - \beta_k k_{j,t-1}] | k_{j,t-1}, i_{j,t-1}, l_{j,t-1}, m_{j,t-1}\}
\end{aligned} \tag{3.63}$$

- m_{jt} has to be excluded from the production function, i.e., it has to be a value-added production function. Otherwise, $\beta_m m_{jt}$ and $\beta_m m_{j,t-1}$ appear in the second step. Because m_{jt} is correlated with ν_{jt} , the only hope is to vary $m_{j,t-1}$. But there is no additional variation in $m_{j,t-1}$ conditional on $k_{j,t-1}$, $i_{j,t-1}$, and $l_{j,t-1}$.

3.3.34 Tackle Collinearity Problem: Share Regression

- How to avoid the peculiar assumptions on shocks and timing of decisions?
- How to identify gross production function avoiding the third assumption by Akerberg et al. [2015]?
- Return to the old literature using the first-order condition.
- Let w_t be wage and p_t be the product price.
- Assume that the factor market is competitive.
- Then, the first-order condition for profit maximization with respect to L_{jt} is:

$$\begin{aligned}
P_t F_L(L_{jt}, K_{jt}) e^{\omega_{jt}} \mathbb{E} e^{\eta_{jt}} &= w_t \\
\Leftrightarrow \frac{P_t F_L(L_{jt}, K_{jt}) e^{\omega_{jt}} \mathbb{E} e^{\eta_{jt}}}{F(L_{jt}, K_{jt})} &= \frac{w_t}{F(L_{jt}, K_{jt})} \\
\Leftrightarrow \frac{F_L(L_{jt}, K_{jt}) L_{jt}}{F(L_{jt}, K_{jt}) e^{\eta_{jt}}} &= \frac{w_t L_{jt}}{\underbrace{P_t F(L_{jt}, K_{jt}) e^{\omega_{jt}} e^{\eta_{jt}}}_{Y_{jt}}},
\end{aligned} \tag{3.64}$$

where the right hand side is expenditure share to the labor, which is observed.

- Furthermore, on the left hand side, we only have η_{jt} , which is independent of inputs.
- Let s_{jt} be the log of expenditure share to the labor, and take a log of the previous equation gives:

$$\begin{aligned}
s_{jt} &= \log[F_L(L_{jt}, K_{jt}) L_{jt} \mathbb{E} e^{\eta_{jt}} / F(L_{jt}, K_{jt})] - \eta_{jt} \\
&= \log(\beta_l) + \ln \mathbb{E} e^{\eta_{jt}} - \eta_{jt}.
\end{aligned} \tag{3.65}$$

- Remember that the coefficient in the Cobb-Douglas function is equal to the expenditure share.
- In general, share regression provides additional variation to identify the elasticity of anticipated production with respect to the labor. Then we can follow the standard OP method to recover other parameters.

3.4 Cost Function Estimation

3.4.1 Cost Function: Duality

- Given a function $y = F(x)$ such that:

- Add factor market structure.
- Add cost minimization.
- → There exists a unique **cost function** $c = C(y, p)$:
 - **Positivity**: positive for positive input prices and a positive.
 - **Homogeneity**: homogeneous of degree one in the input prices.
 - **Monotonicity**: increasing in the input prices and in the level of output.
 - **Concavity**: concave in the input prices.
- Given a function $c = C(y, p)$ such that:
 - **Positivity**: positive for positive input prices and a positive.
 - **Homogeneity**: homogeneous of degree one in the input prices.
 - **Monotonicity**: increasing in the input prices and in the level of output.
 - **Concavity**: concave in the input prices.
- → There exists a unique production function $F(x)$ that yields $C(y, p)$ as a solution to the cost minimization problem:

$$C(y, p) = \min_x p'x \text{ s.t. } F(x) \geq y. \quad (3.66)$$

- If the latter condition holds, the function C is said to be **integrable**.
- It is rare that you can find a closed-form cost function of a production function.
- It makes sense to start from cost function.
- The duality ensures that there is a one-to-one mapping between a class of cost function and a class of production function.
- If you accept competitive factor markets and cost minimization, identifying a cost function is equivalent to identifying a production function.
- We used this idea in the last slides to identify the parameters regarding static decision variables.
- See Jorgenson [1986] for the literature in this topic up to the mid 80s.

3.4.2 Translog Cost Function

- One of the popular specifications:

$$\begin{aligned} \ln c = & \alpha_0 + \alpha'_p \ln p + \alpha_y \ln y + \frac{1}{2} \ln p' B_{pp} \ln p \\ & + \ln p' \beta_{py} \ln y + \frac{1}{2} \beta_{yy} (\ln y)^2. \end{aligned} \quad (3.67)$$

- It assumes that the first and second order elasticities are constant.
- A second-order (log) Taylor approximation of a general cost function.

3.4.3 Translog Cost Function: Integrability

- Translog cost function is known to be integrable if the following conditions hold:
- **Homogeneity**: the cost shares and the cost flexibility are homogeneity of degree zero: $B_{pp}1 = 0$, $\beta'_{py}1 = 0$.
- **Cost exhaustion**: the sum of cost shares is equal to unity: $\alpha'_p1 = 1$, $B'_{pp}1 = 0$, $\beta'_{py}1 = 0$.

- **Symmetry:** the matrix of share elasticities, biases of scale, and the cost flexibility elasticity is symmetric:

$$\begin{pmatrix} B_{pp} & \beta_{py} \\ \beta'_{py} & \beta_{yy} \end{pmatrix} = \begin{pmatrix} B_{pp} & \beta_{py} \\ \beta'_{py} & \beta_{yy} \end{pmatrix}' . \quad (3.68)$$

- **Monotonicity:** The matrix of share elasticities $B_{pp} + vv' - \text{diag}(v)$ is positive semi-definite.

3.4.4 Two Approaches

1. Cost data approach.
 - Use accounting cost data.
 - It does not depend on behavioral assumption.
 - One can impose restrictions of assuming cost minimization.
 - The accounting cost data may not represent economic cost.
2. Revealed preference approach.
 - Assume decision problem for firms.
 - Assume profit maximization.
 - Reveal the costs from firm's equilibrium strategy.
 - It depends on structural assumptions.
 - It reveals the cost as perceived by firms.

3.4.5 Cost Data Approach

- Estimating a cost function using cost data from accounting data.
- McElroy [1987] is one of the most flexible and robust frameworks.
- The approach is somewhat getting less popular in IO researchers.
- Recently, the approach is not popular among IO researchers.
- I *conjecture* one of the reasons for this is that IO researchers believe cost data taken from accounting information does not capture all the costs firms face.
- However, it is good to know the classical literature because it sometimes gives a new insight.
- cf. Byrne et al. [2015] : Propose a novel method to combine accounting cost data to estimate demand and cost function jointly without using instrumental variable approach.

3.4.6 Revealed Preference Approach

- Another approach is to **reveal** the marginal cost from firm's price/quantity setting behavior assuming it is maximizing profit.
 - A parameter affects economic agent's action.
 - Therefore, economic agent's action **reveals** the information about the parameter.
 - See Bresnahan [1981] and Bresnahan [1989] for reference.
- We have shown that the assumption on the factor market and cost function minimization gives restriction on the cost parameters.
- We may further assume the product market structure and profit maximization to identify cost parameters.
- Example: In a competitive market, the equilibrium price is equal to the marginal cost. Therefore, the marginal cost is identified from prices.
- What if the competition is imperfect?

3.4.7 Single-product Monopolist

- This approach requires researcher to specify the decision problem of a firm.
- Assume that the firm is a single-product monopolist.
- Let $D(p)$ be the demand function.
- Let $C(q)$ be the cost function.
- Temporarily, assume that we **know** the demand function.
- We learn how to estimate demand functions in coming weeks.
- The only unknown parameter is the cost function.
- The monopolist solves:

$$\max_p D(p)p - C(D(p)). \quad (3.69)$$

- The first-order condition w.r.t. p for profit maximization is:

$$\begin{aligned} D(p) + pD'(p) - C'(D(p))D'(p) &= 0. \\ \Leftrightarrow C'(D(p)) &= \underbrace{\frac{D(p) + pD'(p)}{D'(p)}}_{p \text{ is observed and } D(p) \text{ is known.}} \end{aligned} \quad (3.70)$$

- This identifies the marginal cost *at the equilibrium quantity*.
- To trace out the entire marginal cost function, you need a demand shifter Z that changes the equilibrium: $D(p, Z)$.

$$C'(D(p, z)) = \frac{D(p, z) + pD'(p, z)}{D'(p, z)} \quad (3.71)$$

- This identifies the marginal cost function *at the equilibrium quantity when $Z = z$* .
- If the equilibrium quantities cover the domain of the marginal cost function when the demand shifter Z moves around, then it identifies the entire marginal cost function.

3.4.8 Unobserved Heterogeneity in the Cost Function

- Previously we did not consider any unobserved heterogeneity in the cost function.
- Now suppose that the cost function is given by:

$$C(q) = \tilde{C}(q) + q\epsilon + \mu, \quad (3.72)$$

and ϵ and μ are not observed.

- Moreover, because it includes anticipated shocks, it is likely to be correlated with input decisions and hence the output.

- The first-order condition w.r.t. p for profit maximization is:

$$\begin{aligned} D(p, z) + pD'(p, z) - [\tilde{C}'(D(p, z)) + \epsilon]D'(p, z) &= 0. \\ \Leftrightarrow \tilde{C}'(D(p, z)) &= \frac{D(p, z) + pD'(p, z)}{D'(p, z)} - \epsilon. \end{aligned} \quad (3.73)$$

- Take the expectation conditional on $Z = z$:

$$\tilde{C}'(D(p, z)) = \frac{D(p, z) + pD'(p, z)}{D'(p, z)} - \mathbb{E}\{\epsilon|Z = z\}. \quad (3.74)$$

- If Z and ϵ is independent, then the last term becomes zero and we can follow the same argument as before to trace out the marginal cost function.

3.4.9 Multi-product Monopolist Case

- Demand for good j is $D_j(p)$ given a price vector p .
- Cost for producing a vector of good q is $C(q)$.
- Demand function is **known** but cost function is not known.
- The monopolist solves:

$$\max_p \sum_{j=1}^J p_j D_j(p) - C(D_1(p), \dots, D_J(p)). \quad (3.75)$$

- The first-order condition w.r.t. p_i for profit maximization is:

$$\begin{aligned} D_i(p) + p_i \sum_{j=1}^J \frac{\partial D_j(p)}{\partial p_i} &= \sum_{j=1}^J \frac{\partial C(D_1(p), \dots, D_J(p))}{\partial q_j} \frac{\partial D_j(p)}{\partial p_i}. \\ &= \begin{pmatrix} \frac{\partial D_1(p)}{\partial p_i} & \dots & \frac{\partial D_J(p)}{\partial p_i} \end{pmatrix} \begin{pmatrix} \frac{\partial C(D_1(p), \dots, D_J(p))}{\partial q_1} \\ \vdots \\ \frac{\partial C(D_1(p), \dots, D_J(p))}{\partial q_J} \end{pmatrix} \end{aligned} \quad (3.76)$$

- Summing up, the first-order condition w.r.t. p is summarized as:

$$\begin{aligned} \begin{pmatrix} D_1(p) + p_1 \sum_{j=1}^J \frac{\partial D_j(p)}{\partial p_1} \\ \vdots \\ D_J(p) + p_J \sum_{j=1}^J \frac{\partial D_j(p)}{\partial p_J} \end{pmatrix} &= \begin{pmatrix} \frac{\partial D_1(p)}{\partial p_1} & \dots & \frac{\partial D_J(p)}{\partial p_1} \\ \vdots & & \vdots \\ \frac{\partial D_1(p)}{\partial p_J} & \dots & \frac{\partial D_J(p)}{\partial p_J} \end{pmatrix} \begin{pmatrix} \frac{\partial C(D_1(p), \dots, D_J(p))}{\partial q_1} \\ \vdots \\ \frac{\partial C(D_1(p), \dots, D_J(p))}{\partial q_J} \end{pmatrix} \\ \Leftrightarrow \begin{pmatrix} \frac{\partial C(D_1(p), \dots, D_J(p))}{\partial q_1} \\ \vdots \\ \frac{\partial C(D_1(p), \dots, D_J(p))}{\partial q_J} \end{pmatrix} &= \underbrace{\begin{pmatrix} \frac{\partial D_1(p)}{\partial p_1} & \dots & \frac{\partial D_J(p)}{\partial p_1} \\ \vdots & & \vdots \\ \frac{\partial D_1(p)}{\partial p_J} & \dots & \frac{\partial D_J(p)}{\partial p_J} \end{pmatrix}^{-1}}_{p \text{ is observed and } D(p)\text{s are known.}} \begin{pmatrix} D_1(p) + p_1 \sum_{j=1}^J \frac{\partial D_j(p)}{\partial p_1} \\ \vdots \\ D_J(p) + p_J \sum_{j=1}^J \frac{\partial D_j(p)}{\partial p_J} \end{pmatrix}. \end{aligned} \quad (3.77)$$

- Hence, the cost function is identified.
- Including unobserved heterogeneity in the cost function causes the same problem as in the previous case.

3.4.10 Oligopoly

- There are firm $j = 1, \dots, J$ and they sell product $j = 1, \dots, J$, that is, firm = product (for simplicity).
- Consider a price setting game. When the price vector is p , demand for product j is given by $D_j(p)$.
- The cost function for firm j is $C_j(q_j)$.
- Given other firms' price p_{-j} , firm j solves:

$$\max_{p_j} D_j(p)p_j - C_j(D_j(p)). \quad (3.78)$$

- The first-order condition w.r.t. p_j for profit maximization is:

$$\begin{aligned} D_j(p) + \frac{\partial D_j(p)}{\partial p_j} p_j &= \frac{\partial C_j(D_j(p))}{\partial q_j} \frac{\partial D_j(p)}{\partial p_j}. \\ \frac{\partial C_j(D_j(p))}{\partial q_j} &= \underbrace{\frac{\partial D_j(p)^{-1}}{\partial p_j} \left[D_j(p) + \frac{\partial D_j(p)}{\partial p_j} p_j \right]}_{p \text{ is observed and } D_j(p) \text{ is known}}. \end{aligned} \quad (3.79)$$

- In Nash equilibrium, these equations jointly hold for all firms $j = 1, \dots, J$.
- Including unobserved heterogeneity in the cost function causes the same problem as in the previous case.

Chapter 4

Demand Function Estimation

4.1 Motivations

- From demand function and utility maximization assumption, we can reveal the preference of the decision maker.
- Thus, estimating demand function is necessary for **evaluating the consumer welfare**.
- In IO, estimating the **price elasticity of demand** is specifically important, because it determines the **market power** of a monopolist and the size of the dead-weight loss.
- In macroeconomics, estimating demand is in important to determine the **price level**, because the price level is the minimum expenditure for a consumer to achieve the certain level of utility.
- In marketing, estimating demand is necessary to design the optimal pricing, advertising, and all the other marketing interventions.
- In principle, the theory can be applied to whatever decisions other than the consumer choice.
- Nevo [2000]:
 - How do the hypothetical mergers in the ready-to-eat cereal industry affect the market price, markup, and consumer surplus?
 - To do so, the authors estimate the demand for ready-to-eat cereals and the cost functions for each product. Then, the authors conduct counterfactual simulations of mergers to quantify the effects.
- Chung and Alcácer [2002]:
 - To what extent do firms go abroad to access technology available in other locations?
 - To study this issue, the authors estimate the firms' locational choice when going abroad.
- Rysman [2004]:
 - In Yellow pages, how do consumers evaluate the advertisement on it, and how do advertisers value consumer usage?
 - To study this, the author simultaneously estimate the consumer demand for usage of a directory, advertiser demand for advertising, and a publisher's first-order condition.
- Gentzow [2004]:
 - Are online and print newspapers substitutes or complements?
 - To study this, the author estimate a demand function in which online and print newspapers can be either substitutes or complements.
- Bayer et al. [2007]:

- How is the preference of people for schools and neighborhoods? How is this capitalized into housing prices?
- To do so, the authors estimate the discrete choice of residents over locations. To deal with the endogeneity between the neighborhood and the unobserved attributes of the location, the authors use the discontinuity at the school attendance zone.
- Archak et al. [2011]:
 - How does the information embedded in product reviews the consumer choice?
 - To study this, the authors estimate the discrete choice model of consumers in which the text information from the product reviews are included as the product attributes.
- Holmes [2011]:
 - Wal-Mart maintain high store density. How large is the economy of density and the sales cannibalization?
 - To study this, the author first estimate the demand function across neighborhood Wal-Mart to capture the sales cannibalization, and then estimate the cost structure from their entry and exit behaviors.
- Handbury and Weinstein [2014]:
 - Urban and rural areas differ in available products. How does the price difference change if the heterogeneity in the product availability is incorporated?
 - To do so, the authors estimate the demand function at each location, and the construct the spatial price index based on the available products at each location.

4.2 Analyzing Consumer Behaviors

- **Alternative set.**
- **Utility function.**
 - Add system of **choice sets**.
 - Add utility maximization.
- → **Demand function.**
- In case of producer behavior, there was a chance to directly observe the output of the most primitive function, the production function.
- In case of consumer behavior, we never directly observe the output of the most primitive function, the utility function.
- We can at most identify demand functions.
- **Revealed preference theory:**
 - Samuelson [1938], Houthakker [1950], Richter [1966], Afriat [1967], Varian [1982].
 - If the demand function is derived from a preference by maximizing the preference, the demand function should satisfy some restrictions.
 - If the assumption is true, we can recover part of the preference from the demand function.

4.3 Continuous Choice

- The alternative set \mathcal{X} is a subset of \mathbb{R}^J .
- The utility function u is rational, monotone, and continuous on \mathcal{X} .

- The choice sets are given by a system of **linear budget set**:

$$\mathcal{B}(p, x) = \{q \in \mathcal{X} : p \cdot q \leq w\}.$$

- If choice sets are non-linear, the following duality approach needs to be modified.

4.3.1 Duality between Utility and Expenditure Functions

- It is rather a special case that we can derive a closed form solution to a utility maximization problem.
- We can use the first-order conditions as moment conditions for identification.

$$\frac{\partial u(q)}{\partial q_i} = \lambda p_i, i = 1, \dots, J. \quad (4.1)$$

- The derivation of a demand function from the identified utility function in general require a numerical simulation, which can be bothering.
- As well as the duality between production and cost functions, we have the same duality theorem for utility and expenditure functions.
- There is a one-to-one mapping between a class of utility functions and a class of expenditure functions.
- Therefore, it is okay to start from an expenditure function.
- It is rare that we can recover the utility function associated with an expenditure function in a closed form. But it is not often required for analysis.
- Moreover, we can easily derive other important functions from the expenditure functions.
- Let p be the price vector and u be the target utility level.
- Let $u(q)$ be a utility function.
- An expenditure function associated with the utility function is defined by:

$$e(u, p) = \min_q p \cdot q, u(q) \geq u. \quad (4.2)$$

- Let x be the total expenditure such that:

$$x = e(u, p). \quad (4.3)$$

- We can start the analysis by specifying this function instead of the utility function.

4.3.2 Deriving Other Functions

- It is easy to derive other functions from an expenditure function.
- **Indirect utility function**: invert the expenditure function to get:

$$u = e^{-1}(p, x) \equiv v(p, x). \quad (4.4)$$

- **Hicksian demand function**: apply Shepard's lemma:

$$q_i = \frac{\partial e(u, p)}{\partial p_i} \equiv h_i(u, p). \quad (4.5)$$

- **Marshallian demand function**: insert Hicksian demand function to the expenditure function:

$$q_i = h_i(v(p, x), p) \equiv d_i(p, x). \quad (4.6)$$

4.3.3 Starting from an Indirect Utility Function

- It is almost equivalent to start from an indirect utility function.
- An indirect utility function with the utility function is defined by:

$$v(p, x) \equiv \max_q u(q), p'q \leq x. \quad (4.7)$$

- We can derive Marshallian demand function by Roy's identity:

$$q_i = \frac{-\partial v(p, x)/\partial p_i}{\partial v(p, x)/\partial x} \equiv d_i(p, x). \quad (4.8)$$

4.3.4 Expenditure Share Equation

- Let's start from an expenditure function $e(p, x)$.
- By Shepard's lemma, we have:

$$\frac{\partial \ln e(u, p)}{\partial \ln p_i} = \frac{\partial e(u, p)}{\partial p_i} \frac{p_i}{e(u, p)} = \frac{p_i q_i}{x} \equiv w_i. \quad (4.9)$$

- We call this an expenditure share equation.
- The estimation is based on the share equations.

4.3.5 Almost Ideal Demand System (AIDS)

- Based on Deaton and Muellbauer [1980].
- See Angus Deaton and John Muellbauer [1980] for further reference.
- Consider an expenditure function that satisfies the following useful conditions:
 - It allows aggregation (this motivation is less important in recent days).
 - It gives an arbitrary first-order approximation to any demand system.
 - It can satisfy the restrictions of utility maximization.
 - It can be used to test the restrictions of utility maximization.

4.3.6 PIGLOG Class

- PIGLOG (price-independent generalized logarithmic) class [Muellbauer, 1976].

$$\ln e(u, p) = (1 - u) \ln a(p) + u \ln b(p), \quad (4.10)$$

where $a(p)$ and $b(p)$ are arbitrary linear homogeneous concave functions.

- Consider households that differ in total income.
- PIGLOG form ensures that the aggregate demand can be written in the same form where the total income is replaced with the sum of household total income.
- The derivatives should be given free parameters for the model to be an arbitrary first-order approximation to any demand system.
- In AIDS, we specify $a(p)$ and $b(p)$ as:

$$\begin{aligned} \ln a(p) &\equiv a_0 + \sum_k \alpha_k \ln p_k + \frac{1}{2} \sum_k \sum_j \gamma_{kj}^* \ln p_k \ln p_j \\ \ln b(p) &\equiv \ln a(p) + \beta_0 \prod_k p_k^{\beta_k}. \end{aligned} \quad (4.11)$$

4.3.7 Derive the Share Equation I

- By Roy's identify, we can derive the associated share equation:

$$w_i \equiv \frac{\partial \ln e(u, p)}{\partial \ln p_i} = \alpha_i + \sum_j \gamma_{ij} + \beta_i u \beta_0 \prod_k p_k^{\beta_k}, \quad (4.12)$$

where

$$\gamma_{ij} = \frac{1}{2}(\gamma_{ij}^* + \gamma_{ji}^*). \quad (4.13)$$

4.3.8 Derive the Share Equation II

- Insert indirect utility function $u = v(p, x)$ to this to get:

$$w_i = \alpha_i + \sum_j \gamma_{ij} \ln p_j + \beta_i \ln \frac{x}{P}, \quad (4.14)$$

where

$$\ln P \equiv \alpha_0 + \sum_k \alpha_k \ln p_k + \frac{1}{2} \sum_j \sum_k \gamma_{kj} \ln p_k \ln p_j. \quad (4.15)$$

- P is a price index associated with the given preference.
- With the specification of Richard Stone [1954], it becomes $\ln P = \sum_j x_j \ln p_j$.
- It can be used as an approximation.

4.3.9 Specify the Detail II

- It can satisfy the restrictions of utility maximization.
- It can be used to test the restrictions of utility maximization.
 - $\sum_j x_j = 1$:

$$\sum_j \alpha_j = 1, \sum_j \gamma_{jk} = 0, \sum_j \beta_j = 0. \quad (4.16)$$

- $e(u, p)$ is linear homogeneous in p :

$$\sum_j \gamma_{ij} = 0. \quad (4.17)$$

- Symmetry:

$$\gamma_{ij} = \gamma_{ji}. \quad (4.18)$$

4.3.10 Estimation

- We can estimate parameters based on the share equations.
- If we use aggregate data, the aggregate error term is correlated with the price vector.
- Therefore, we need at least as many instrumental variables as the dimension of the price vector.
- With valid instrumental variables, we can estimate the model with GMM.
- If we use household-level data, the household-specific errors controlling for aggregate errors will not be correlated with the price vector if the price is determined in a competitive market.

4.3.11 From Product Space Approach to Characteristics Space Approach

- The framework up to here is called **product space approach** because the utility has been defined over a product space.
- When there are J goods, there are J^2 parameters for prices.
- One way to resolve this issue is to introduce a priori knowledge about the preference.
- For example, we can introduce a priori segmentation with separability.
- It is hard to evaluate the effect of introducing new product.
- Again, we have to a priori decide which segment/product is similar to the new product.
- This leads us to the **characteristics space approach** [Lancaster, 1966, Muth, 1966]:
 - Consumption is an activity in which goods are inputs and in which the output is a collection of characteristics.
 - Utility ranks collections of characteristics and only to rank collections of goods indirectly through the characteristics that they possesses.
 - There are $k = 1, \dots, K$ activities.
 - The activity y requires to consume $x = Ay$ products.
 - The activity y generates $z = By$ characteristics.
 - The budget constraint is $p \cdot x \leq 1$.
 - The utility is defined over the characteristics $u(z)$.
 - The consumer's problem is:

$$\max_y u(z)$$

s.t.

$$p \cdot x \leq 1, x = Ay, z = By, x, y, z \geq 0.$$

- Then, only the dimension of characteristics matters and the value of new products can be evaluated by the contribution to the production of characteristics.
- The early application includes Rosen [1974], Muellbauer [1974], Gorman [1980].
- The nonparametric analysis based on the reveals preference is Blow et al. [2008].

4.3.12 From Continuous Choice Approach to Discrete Choice Approach

- The aggregate demand is a collection of choice across consumers and within consumers over time.
- It makes sense to model individual choices and then aggregate rather than directly modeling the aggregate demand.
- The resulting aggregate demand will satisfy restrictions that are consistent with the underlying consumer choice model.
- If there is an interaction across choices, the aggregation is not trivial.
- This is especially true when aggregating choices within consumers.
- For now, assume that each choice is independent.

4.4 Discrete Choice

4.4.1 Discrete Choice Approach

- Let $u(q, z_i)$ be the utility of a consumer over $J + 1$ dimensional consumption bundle q characterized by consumer characteristics z_i .
- The consumer solves:

$$V(p, y_i, z_i) = \max_q u(q, z_i), \text{ s.t. } p'q \leq y_i. \quad (4.19)$$

- Alternative 0 is an **outside good**.

- Normalize $p_0 = 1$.
- We call alternatives $j = 1, \dots, J$ **inside goods**.
- The choice space is restricted on:

$$Q = \{q : q_0 \in [0, M], q_j \in \{0, 1\}, j = 1, \dots, J, q_j q_k = 0, \forall j \neq k, j, k > 0, M < \infty\}. \quad (4.20)$$

4.4.2 Discrete Choice Approach

- The budget constraint reduces to:

$$\begin{cases} q_0 + p_j q_j = y & \text{if } q_j = 1, j > 0 \\ q_0 = y & \text{otherwise.} \end{cases} \quad (4.21)$$

- Hence,

$$q_0 = y - \sum_{j=1}^J p_j q_j. \quad (4.22)$$

4.4.3 Discrete Choice Approach

- The utility maximization problem can be written as:

$$V(p, y_i, z_i) = \max_{j=0,1,\dots,J} v_j(p_j, y_i, z_i), \quad (4.23)$$

where

$$\begin{aligned} & v_j(p_j, y_i, z_i) \\ &= \begin{cases} u(y_i - p_j, 0, \dots, \underbrace{1}_{q_j}, \dots, 0, z_i) & \text{if } j > 0, \\ u(y_i, 0, \dots, 0, z_i) & \text{if } j = 0, \end{cases} \end{aligned} \quad (4.24)$$

is called the **choice-specific indirect utility**.

4.4.4 Characteristics Space Approach

- Preference is defined over the characteristics of alternatives, x_j :
- Car: vehicle, engine power, model-year, car maker, etc.
- PC: CPU power, number of cores, memory, HDD volume, etc.
- The choice-specific indirect utility is a function of the characteristics of the alternative:

$$\begin{aligned} v_j(p_j, y_i, z_i) &= u(y_i - p_j, 0, \dots, \underbrace{1}_{q_j}, \dots, 0, z_i) \\ &= u^*(y_i - p_j, x_j, z_i) \\ &\equiv v(p_j, x_j, y_i, z_i). \end{aligned} \quad (4.25)$$

4.4.5 Weak Separability and Income Effect

- We usually focus on a particular product category such as cars, PCs, cereals, detergents, and so on.
- Assume that the preference is separable between the category in question and other categories.
- Then, call the category in question inside goods.
- The outside good captures the income effect of the choice in this category.

4.4.6 Weak Separability and Income Effect

- Thus, how the preference for the outside good is modeled determines how the individual income affects the choice.

$$\begin{aligned} u^*(y_i - p_j, x_j, z_i) &= u^{**}(x_j, z_i) + \alpha(y_i - p_j). \\ u^*(y_i - p_j, x_j, z_i) &= u^{**}(x_j, z_i) + \alpha \ln(y_i - p_j). \end{aligned} \quad (4.26)$$

- In the first example, the income level does not affect the choice because the term αy_i is common and constant across choices (there is no income effect).
- We often do not observe income of a consumer, y_i .
- Remember that the price of a product enters because we here consider **indirect** utility function.

4.4.7 Utility Function Normalization

- The **location** of utility function is often normalized by setting:

$$u(y^*, 0, \dots, 0, z^*) = 0, \quad (4.27)$$

for certain choice of (y^*, z^*) .

4.4.8 Aggregation of the Individual Demand

- Let $q_j(p, x, y_i, z_i) = \{q_j(p, x, y_i, z_i)\}_{j=0, \dots, J}$ be the demand function of consumer i , that is:

$$q_j(p, x, y_i, z_i) = 1 \Leftrightarrow j = \operatorname{argmax}_{j=0, 1, \dots, J} v(p_j, x_j, y_i, z_i). \quad (4.28)$$

- Let $f(y, z)$ be the joint distribution of the income and other consumer characteristics.
- The aggregate demand for good j is:

$$D_j(p, x_j; f) \equiv N \int q_j(p, x_j, y, z) f(y, z) dy dz, \quad (4.29)$$

where N is the population.

4.4.9 Horizontal Product Differentiation

- **horizontal product differentiation:** consumers do not agree on the ranking of the choices.
- There are two convenience stores $j = 1, 2$ on a street $[0, 1]$.
- Let $z_i = L_i$, the location of consumer i and $x_j = L_j$, the location of the choice on a street $[0, 1]$ with $L_1 < L_2$.
- A consumer has a preference such that:

$$v_{ij} \equiv v(p_j, x_j, y_i, z_i) \equiv s - t|L_i - L_j| - p_j. \quad (4.30)$$

4.4.10 Horizontal Product Differentiation

- Suppose that the prices are low enough that entire consumers on the street are willing to buy either from the stores.
- Consumer i buys from store 1 if and only if:

$$\begin{aligned}
 v(p_1, x_1, y_i, z_i) &\geq v(p_2, x_2, y_i, z_i) \\
 \Leftrightarrow s - t|L_i - L_1| - p_1 &\geq s - t|L_i - L_2| - p_2 \\
 \Leftrightarrow L_i &\leq \frac{p_2 - p_1}{2t} + \frac{L_1 + L_2}{2} \equiv \bar{L}_1(p_1, p_2).
 \end{aligned} \tag{4.31}$$

- Let $f(L_i)$ be $U[0, 1]$. Then, the aggregate demand for store 1 is:

$$\sigma_1(p, L_1, L_2) = N \int_0^{\bar{L}_1(p_1, p_2)} dL_i = N\bar{L}_1(p_1, p_2). \tag{4.32}$$

4.4.11 Vertical Product Differentiation

- **Vertical product differentiation:** Consumers agree on the ranking of the choices. Consumers can have different willingness to pay.
- Bresnahan [1987] analyzed automobile demand with this framework.
- There are J goods and consumer i has a utility such as:

$$v_{ij} \equiv v(p_j, x_j, y_i, z_i) = z_i x_j - p_j, \tag{4.33}$$

where x_j is a quality of product j and z_i is the consumer's willingness to pay for the quality with $x_j < w_{j+1}$.

- Consumers' problem is:

$$\max\{0, z_i x_1 - p_1, \dots, z_i x_J - p_J\}. \tag{4.34}$$

4.4.12 Vertical Product Differentiation

- Consumer i prefers good $j + 1$ to good j if and only if:

$$\begin{aligned}
 v(p_{j+1}, w_{j+1}, y_i, z_i) &\geq v(p_j, x_j, y_i, z_i) \\
 \Leftrightarrow z_i w_{j+1} - p_{j+1} &\geq z_i x_j - p_j \\
 \Leftrightarrow z_i &\geq \frac{p_{j+1} - p_j}{w_{j+1} - x_j} \equiv \Delta_j.
 \end{aligned} \tag{4.35}$$

- So consumer i purchases good j if and only if $z_i \in [\Delta_{j-1}, \Delta_j)$ and buys nothing if:

$$z_i \leq \Delta_0 \equiv \min\{p_1/x_1, \dots, p_J/x_J\}. \tag{4.36}$$

- Letting $F(z)$ be the distribution function of z , the aggregate demand for good j is:

$$\sigma_j(p, x, z) = N[F(\Delta_j) - F(\Delta_{j-1})]. \tag{4.37}$$

4.4.13 Econometric Models

- So far there was no econometrics.
- Next we define what are observable and unobservable, and what are known and unknown.
- Then consider how to identify and estimate the model.

4.4.14 Multinomial Logit Model: Preference Shock

- This originates at Mcfadden [1974].
- See Train [2009] for reference.
- Suppose that there is some unobservable component in consumer characteristics.
- In reality, consumers choice change somewhat randomly.
- Let's capture such a **preference shock** by consider the following model:

$$v(p_j, x_j, y_i, z_i) + \epsilon_{ij}, \quad (4.38)$$

where I assumed that there is no other z_i for exposition simplicity, with some random vector:

$$\epsilon_i \equiv (\epsilon_{i0}, \dots, \epsilon_{iJ})' \sim G. \quad (4.39)$$

- At this point, G can be any distribution and the shocks can be dependent across j within i .
- p, x, y_i, z_i are **observed** but ϵ_{ij} are **unobserved**.
- When the realization of the preference shock is given, the consumer choice is:

$$q_j(p, x, y_i, z_i, \epsilon_i) \equiv 1\{j = \operatorname{argmax}_{k=0, \dots, J} v(p_k, x_k, y_i, z_i) + \epsilon_{ik}\}$$

for $k = 0, \dots, J$.

- The **choice probability** as observed by econometrician is:

$$\sigma_j(p, x, y_i, z_i) \equiv \int q_j(p, x, y_i, z_i, \epsilon_i) dG(\epsilon_i).$$

4.4.15 Multinomial Logit Model: Distributional Assumption

- Now assume the followings:
- ϵ_{ij} are independent across j : $G(\epsilon_i) = \prod_{j=0, \dots, J} G_j(\epsilon_{ij})$.
- ϵ_{ij} are identical across j : $G_j(\epsilon_{ij}) = \bar{G}(\epsilon_{ij})$.
- \bar{G} is a type-I extreme value.
- \rightarrow The density $g(\epsilon_{ij}) = \exp[-\exp(-\epsilon_{ij}) - \epsilon_{ij}]$.
- This is called the (homoskedastic) **multinomial logit model**.
- Setting the variance of ϵ_{ij} at 1 for some j is a **scale** normalization.
- By dropping some of the assumptions, we can have heteroskedastic multinomial logit model, generalized extreme value model, and so on.
- Another popular distribution assumption is to assume a multivariate normal distribution of ϵ_i . This case is called the **multinomial probit model**.

4.4.16 Multinomial Logit Model: Choice Probability

- The **choice probability** of consumer i of good j is:

$$\begin{aligned} \sigma_j(p, x, y_i, z_i) &\equiv \mathbb{P}\{j = \operatorname{argmax}_{k=0, 1, \dots, J} v(p_k, x_k, y_i, z_i) + \epsilon_{ik}\} \\ &= \mathbb{P}\{v(p_j, x_j, y_i, z_i) - v(p_k, x_k, y_i, z_i) \geq \epsilon_{ik} - \epsilon_{ij}, \forall k \neq j\} \\ &= \dots \text{after some algebra: leave as an exercise...} \\ &= \frac{\exp[v(p_j, x_j, y_i, z_i)]}{\sum_{k=0}^J \exp[v(p_k, x_k, y_i, z_i)]}. \end{aligned} \quad (4.40)$$

- For example, if:

$$v(p_k, x_k, y_i, z_i) = \beta'_i x_k + \alpha_i(y_i - p_k), \quad (4.41)$$

$$\begin{pmatrix} \beta_i \\ \alpha_i \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \alpha_0 \end{pmatrix} + \begin{pmatrix} \Gamma \\ \pi' \end{pmatrix} z_i. \quad (4.42)$$

- Then, we have:

$$\begin{aligned} \sigma_j(p, x, y_i, z_i) &= \frac{\exp[\beta'_i x_j + \alpha_i(y_i - p_j)]}{\sum_{k=0}^J \exp[\beta'_i x_k + \alpha_i(y_i - p_k)]} \\ &= \frac{\exp[\beta'_i x_j - \alpha_i p_j]}{\sum_{k=0}^J \exp[\beta'_i x_k - \alpha_i p_k]} \end{aligned} \quad (4.43)$$

- If we normalize the characteristics vector so that $w_0 = 0$ holds for the outside option, it becomes:

$$\sigma_j(p, x, y_i, z_i) = \frac{\exp[\beta'_i x_j - \alpha_i p_j]}{1 + \sum_{k=1}^J \exp[\beta'_i x_k - \alpha_i p_k]}$$

4.4.17 Multinomial Logit Model: Inclusive Value

- The expected utility for consumer i before the preference shocks are drawn under multinomial logit model is given by:

$$\begin{aligned} &\mathbb{E}\left\{\max_{j=0, \dots, J} v(p_j, x_j, y_i, z_i) + \epsilon_{ij}\right\} \\ &= \dots \text{after some algebra: leave as an exercise...} \\ &= \ln \left\{ \sum_{j=0}^J \exp[v(p_j, x_j, y_i, z_i)] \right\} + \text{constant}. \end{aligned} \quad (4.44)$$

- This is sometimes called the **inclusive value** of the choice set.

4.4.18 Maximum Likelihood Estimation of Multinomial Logit Model

- Suppose we observe a sequence of income y_i , consumer characteristics z_i , choice q_i , product characteristics x_j and price p_j .
- $q_i = (q_{i0}, \dots, q_{iJ})'$ and $q_{ij} = 1$ if j is chosen and 0 otherwise.
- The parameter of interest is the mean indirect utility function v .
- Then the log likelihood of $\{q_i\}_{i=1}^N$ conditional on $\{y_i, z_i\}_{i=1}^N$ and $\{x_j, p_j\}_{j=1}^J$ is:

$$\begin{aligned} l(v; q, y, z, w) &= \sum_{i=1}^N \ln \mathbb{P}\{q_i = q(p, x, y_i, z_i) | p, x, y_i, z_i\} \\ &= \sum_{i=1}^N \log \left\{ \prod_{j=0}^J \sigma_j(p, x, y_i, z_i)^{q_{ij}} \right\} \\ &= \sum_{i=1}^N \sum_{j=0}^J \log \sigma_j(p, x, y_i, z_i)^{q_{ij}}. \end{aligned} \quad (4.45)$$

- We can estimate the parameters by finding parameters that maximize the log likelihood.

4.4.19 Nonlinear Least Square Estimation of Multinomial Logit Model

- The multinomial logit model can be estimated by nonlinear least square method as well.
- Suppose that the share of product j among consumers with characteristics z and income y was:

$$\sigma_j(p, x, y, z). \quad (4.46)$$

- Note that:

$$\begin{aligned} \ln \sigma_j(p, x, y, z) &= \ln \left\{ \frac{\exp[v(p_j, x_j, y, z)]}{\sum_{k=0}^J \exp[v(p_k, x_k, y, z)]} \right\} \\ &= v(p_j, x_j, y, z) - \ln \left\{ \sum_{k=0}^J \exp[v(p_k, x_k, y, z)] \right\}. \end{aligned} \quad (4.47)$$

- Moreover, because of the location normalization of the utility function,

$$\sigma_0(p, x, y, z) = \frac{1}{\sum_{k=0}^J \exp[v(p_k, x_k, y, z)]}. \quad (4.48)$$

- Hence,

$$\ln \sigma_j(p, x, y, z) - \ln \sigma_0(p, x, y, z) = v(p_j, x_j, y, z). \quad (4.49)$$

- The left-hand variables are observed in the data.
- Let $s_j(y, z)$ be the share of product j among consumers with characteristics z and income y **in the data**.
- This can be calculated from the consumer-level data.
- More importantly, if there is the total sales data for each demographic, we can use this approach.
- Then, we can estimate the parameter by NLLS such that:

$$\min_{(y,z)} \sum_{j=1}^J \{ \ln[s_j(y, z)/s_0(y, z)] - v(p_j, x_j, y, z) \}^2. \quad (4.50)$$

- If v is linear in parameter, it is the ordinal least squares:

$$v(p_j, x_j, y_m) = \beta'_i x_j - \alpha_i p_j. \quad (4.51)$$

$$\ln[s_j(y, z)/s_0(y, z)] = \beta'_i x_j - \alpha_i p_j. \quad (4.52)$$

4.4.20 IIA Problem

- Multinomial logit problem is intuitive and easy to implement.
- However, there are several problems in the model.
- The most important problem is the **independence of irrelevant alternatives (IIA)** problem.
- Notice that:

$$\frac{\sigma_j(p, x, y, z)}{\sigma_k(p, x, y, z)} = \frac{\exp[v(p_j, x_j, y, z)]}{\exp[v(p_k, x_k, y, z)]}. \quad (4.53)$$

- The ratio of choice probabilities between two alternatives depend only on the mean indirect utility of these two alternatives and **independent of irrelevant alternatives (IIA)**.
- Why is this a problem?

4.4.21 Blue Bus and Red Bus Problem

- Suppose that you can go to a town by bus or by train.
- Half of commuters use a bus and the other half use a train.
- The existing bus was blue. Now, the county introduced a red bus, which is identical to the existing blue bus.
- No one take care of the color of bus. So the mean indirect utility of blue bus and red bus are equal.
- What is the new share across blue bus, red bus, and train?
- IIA \rightarrow share of blue bus = share of train.
- Buses are identical \rightarrow share of blue bus = share of red bus.
- Therefore, shares have to be 1/3, respectively.
- But shouldn't it be that train keeps half share and bus have half share in total?

4.4.22 Restrictive Price Elasticity

- IIA property restrict price elasticities in an unfavorable manner.
- This is a serious problem because the main purpose for us to estimate demand functions is to identify the price elasticity.
- Let $v(p_j, x_j, y, z) = \beta'_z x_j - \alpha_z p_j$. Then, we have:

$$e_{jk} = \begin{cases} -\alpha p_j (1 - \sigma_j(p, x, y, z)) & \text{if } k = j \\ \alpha p_k \sigma_k(p_k, x_k, y, z) & \text{if } k \neq j. \end{cases} \quad (4.54)$$

- The price elasticity is completely determined by the existing choice probabilities of the relevant alternatives.
- Suppose that there are coca cola, Pepsi cola, and a coffee.
- The shares were 1/2, 1/6, 1/3, respectively.
- Suppose that the price of coca cola increased.
- We expect that they instead purchase Pepsi cola because Pepsi cola is more similar to coca cola than coffee.
- However, according to the previous result, twice more consumers substitute to coffee rather than to Pepsi cola.

4.4.23 Monotonic Inclusive Value

- Suppose that there is a good whose mean indirect utility is v .
- The inclusive value for this choice set is $\ln[1 + \exp(v)]$.
- Suppose that we put J same goods on the shelf and consumer can choose any of them.
- The inclusive value is $\ln[1 + J \exp(v)]$.
- We just added the same goods. But the expected utility of consumer increases monotonically in the number of alternatives.

4.4.24 The Source of the Problem

- **The source of the problem is that there is no correlation in the preference shock across products.**
- When the preference shock to coca cola is high, the preference shock to Pepsi cola should be high, while the preference shock to coffee should be relatively independent.
- Because the expected value of the maximum of the preference shocks increases according to the number of alternatives, the inclusive value becomes increasing in the number of alternatives.

- However, the preference shocks should be the same for the same good. Then, the the expected value of the maximum of the preference shock should not increase even if we add the same products on the shelf.

4.4.25 Correlation in Preference Shocks

- Therefore, the preference shock should be such that: preference shocks between two alternative should be more correlated when they are closer in the characteristics space.
- So we have to allow the covariance matrix of the preference shock to be free parameters.
- If we allow flexible covariance matrix, the curse of dimensionality in the number of alternatives comes back: The dimensionality of the covariance matrix is J^2 .
- Another way is to remove ϵ_{ij} : it is called a **pure characteristics model** [Berry and Pakes, 2007].
- But the pure characteristics model is computationally not straightforward.
- We explore the way of introducing mild correlation across similar products in the preference shocks.

4.4.26 Observed and Unobserved Consumer Heterogeneity

- Consider beverage demand and let $x_j = \text{carbonated}_j$ and $z_i = \text{teenager}_i$.
- Suppose that the mean indirect utility is:

$$v(p_j, x_j, y_i, z_i) = \beta_i(\text{carbonated})_j - \alpha_i p_j, \quad (4.55)$$

$$\beta_i = 0.1 + 0.2 \cdot (\text{carbonated})_i. \quad (4.56)$$

- The mean utility of a carbonated drink for a teenager is 0.3 but only 0.1 for others.
- When coca cola was not available, teenager will substitute more to Pepsi cola than non-teenagers.
- IIA holds at the market-segment level but not at the market level.
- How to avoid IIA at the market-segment level?: Introduce unobserved consumer heterogeneity.
- Suppose that the mean indirect utility is:

$$\beta_i = 0.1 + 0.2 \cdot (\text{carbonated})_i + \nu_i. \quad (4.57)$$

- Consumers with high ν_i values carbonated drinks more than those with low ν_i values.
- When coca cola was not available, consumers with high ν_i will substitute more to Pepsi cola than those with low ν_i values.
- IIA holds at the market-segment- ν level but not at the market-segment level.
- In the above example, “ $0.2 \cdot (\text{carbonated})_i$ ” captures the consumer heterogeneity by observed characteristics and “ ν_i ” by unobserved characteristics.

4.4.27 Mixed Logit Model

- Suppose that the mean indirect utility is:

$$v(p_j, x_j, y_i, z_i, \beta_i, \alpha_i) = \beta_i' x_j - \alpha_i p_j, \quad (4.58)$$

with

$$(\beta_i, \alpha_i) \sim f(\beta_i, \alpha_i | y_i, z_i). \quad (4.59)$$

- If ϵ_{ij} is drawn i.i.d. from type-I extreme value distribution, the choice probability of good j by consumer i conditional on p, x, y_i, z_i is:

$$s_j(p, x, y_i, z_i) = \int_{\beta_i, \alpha_i} \frac{\exp[v(p_j, x_j, y_i, z_i, \beta_i, \alpha_i)]}{\sum_{k=0}^J \exp[v(p_k, x_k, y_i, z_i, \beta_i, \alpha_i)]} f(\beta_i, \alpha_i | y_i, z_i) d\beta_i d\alpha_i. \quad (4.60)$$

- This is called the **mixed-logit model**.
- If the distribution of ϵ_{ij} is different, it is no longer mixed logit.
- Conditional on (β_i, α_i) the choice probability is written in the same way with the multinomial logit model.
- β_i, α_i are marginal out, because econometrician does not observe them.

4.4.28 Mixed Logit Model : Parametric Assumptions

- It is often assumed that:

$$v(p_j, x_j, y_i, z_i, \beta_i, \alpha_i) = \beta_i' x_j - \alpha_i p_j. \quad (4.61)$$

- McFadden and Train [2000] showed that any discrete choice models that are consistent with the random utility maximization can be arbitrarily closely approximated by this class of mixed-logit model.
- The distribution of β_i and α_i is often assumed to be:

$$\begin{aligned} \beta_i &= \beta_0 + \Gamma z_i + \Sigma \nu_i, \\ \alpha_i &= \alpha_0 + \pi' z_i + \Omega v_i, \end{aligned} \quad (4.62)$$

where ν_i and v_i are i.i.d. standard normal random vectors.

4.4.29 Mixed Logit Model: IIA

- There is no IIA at the market-segment level:

$$\frac{\sigma_j(p, x, y, z)}{\sigma_l(p, x, y_i, z_i)} = \frac{\int_{\beta_i, \alpha_i} \frac{\exp[v(p_j, x_j, y_i, z_i, \beta_i, \alpha_i)]}{\sum_{k=0}^J \exp[v(p_k, x_k, y_i, z_i, \beta_i, \alpha_i)]} f(\beta_i, \alpha_i | y_i) d\beta_i d\alpha_i}{\int_{\beta_i, \alpha_i} \frac{\exp[v(p_l, x_l, y_i, z_i, \beta_i, \alpha_i)]}{\sum_{k=0}^J \exp[v(p_k, x_k, y_i, z_i, \beta_i, \alpha_i)]} f(\beta_i, \alpha_i | y_i) d\beta_i d\alpha_i}. \quad (4.63)$$

- The share ratio depends on the price and characteristics of all the other products.

4.4.30 Mixed Logit Model: Price Elasticities

- Let:

$$v(p_j, x_j, y_i, z_i, \beta_i, \alpha_i) = \beta_i' x_j - \alpha_i p_j. \quad (4.64)$$

- The price elasticities of the choice probabilities conditional on p, x, y_i, z_i is:

$$e_{jk} = \begin{cases} -\frac{p_j}{\sigma_j} \int \alpha_i \sigma_{ij} (1 - \sigma_{ij}) f(\beta_i, \alpha_i | y_i, z_i) d\beta_i d\alpha_i & \text{if } j = k \\ \frac{p_k}{\sigma_j} \int \alpha_i \sigma_{ij} \sigma_{ik} f(\beta_i, \alpha_i | y_i, z_i) d\beta_i d\alpha_i & \text{otherwise,} \end{cases} \quad (4.65)$$

where

$$\sigma_{ij} = \frac{\exp(\beta_i' x_j - \alpha_i p_j)}{\sum_{k=0}^J \exp(\beta_i' x_k - \alpha_i p_k)}. \quad (4.66)$$

- The price elasticity depends on the density of unobserved consumer types.

4.4.31 Simulated Maximum Likelihood Estimation of the Mixed Logit Model

- The choice probability of the mixed logit model is an integration of the multinomial logit choice probability.
- This is not derived analytically in general.
- We can use simulation to evaluate the choice probability:
- Draw R values of β and α , $\{\beta^r, \alpha^r\}_{r=1}^R$.
- Compute the multinomial choice probabilities associated with (β^r, α^r) for each $r = 1, \dots, R$.
- Approximate the choice probability with the mean of the simulated multinomial choice share:

$$\sigma_j(p, x, y_i, z_i) \approx \hat{\sigma}_j(p, x, y_i, z_i) \equiv \frac{1}{R} \sum_{r=1}^R \frac{\exp[v(p_j, x_j, y_i, z_i, \beta^r, \alpha^r)]}{\sum_{k=0}^J \exp[v(p_k, x_k, y_i, z_i, \beta^r, \alpha^r)]}. \quad (4.67)$$

- This is one of the numerical integration: **Monte Carlo integration**.
- Another approach is to use **quadrature**. See Judd [1998] for reference.

4.4.32 Simulated Maximum Likelihood Estimation of the Mixed Logit Model

- There are $t = 1, \dots, T$ markets and there $i = 1, \dots, N$ consumers in each market.
- Let \mathcal{J}_t be the set of products that are available in market t .
- Suppose that we observe income y_{it} , characteristics z_{it} , and choice q_{it} for each consumer in a market.
- Suppose that we observe product characteristics x_{jt} and price p_{jt} of each product in each market.
- The simulated conditional log likelihood is:

$$\begin{aligned} & \sum_{i=1}^N \sum_{t=1}^T \ln \mathbb{P}\{q_{it} \\ & = q(p_t, w_t, y_{it}, z_{it}) | p_t, w_t, y_{it}, z_{it}\} \approx \sum_{i=1}^N \ln \left\{ \prod_{j \in \mathcal{J}_t \cup \{0\}} \hat{\sigma}_j(p_t, w_t, y_{it}, z_{it})^{q_{itj}} \right\}. \end{aligned} \quad (4.68)$$

- We find parameters that maximize the simulated conditional log likelihood.

4.4.33 Simulated Non-linear Least Square Estimation of the Mixed Logit Model

- Suppose that we only know the sales or share at the market-segment level.
- That is, we only observe the share of product j in market t among consumers of characteristics z and income y , $s_{jt}(y, z)$.
- Then we can estimate the parameter by:

$$\min \sum_{t=1}^T \sum_{j \in \mathcal{J}_t \cup \{0\}} \sum_{(y, z) \in \mathcal{Y} \times \mathcal{Z}} \{s_{jt}(y, z) - \hat{\sigma}_j(p_t, w_t, y, z)\}^2. \quad (4.69)$$

4.4.34 Nested Logit Model: A Special Case of Mixed Logit Model

- Let w_{j1}, \dots, w_{jG} be the indicator of product category, i.e., w_{jg} takes value 1 if good j belong to category g and 0 otherwise.
- e.g., car category = {Sports, Luxury, Large, Midsize, Small}.
- We have:

$$v(p, x_j, y_i, z_i) = \beta' x_j - \alpha_i p_j + \sum_{g=1}^G \zeta_{ig} w_{jg} + \epsilon_{ij}. \quad (4.70)$$

- If ζ_{it} takes high value, the consumer attaches higher value to the category.
- When a product in category g was not available, consumers with high ζ_{ig} will substitute more to the other products in the same category than consumers with low ζ_{ig} .

4.4.35 Nested Logit Model: Distributional Assumption

- Let

$$\varepsilon_{ij} \equiv \sum_{g=1}^G \zeta_{ig} w_{jg} + \epsilon_{ij}. \quad (4.71)$$

- Under certain distributional assumption on ζ_{ig} and ϵ_{ij} , the term ε_{ij} have a cumulative distribution [I forgot the reference! Let me know if you find it]:

$$F(\varepsilon_i) = \exp \left\{ - \sum_{g=1}^G \left(\sum_{j \in \text{category } g} \exp[-\varepsilon_{ij}/\lambda_g] \right)^{\lambda_g} \right\}. \quad (4.72)$$

4.4.36 Nested Logit Model: Choice Probability

- Under this distributional assumption, the choice probability is:

$$\sigma_j(p, x, y_i, z_i) = \frac{\exp[v(p, x_j, y_i, z_i)/\lambda_g] \left(\sum_{k \in \text{category } g} \exp[v(p, x_k, y_i, z_i)/\lambda_g] \right)^{\lambda_g - 1}}{\sum_{g=1}^G \left(\sum_{k \in \text{category } g} \exp[v(p, x_k, y_i, z_i)/\lambda_g] \right)^{\lambda_g}}, \quad (4.73)$$

if good j belongs to category g .

- The higher $\lambda_g \in [0, 1]$ implies lower correlation within category g .
- $\lambda_g = 1$ for all g coincides with the multinomial logit model.

4.4.37 Nested Logit Model: Decomposition of the Choice Probability

- The choice probability can be decomposed into two parts:

$$\sigma_j(p, x, y_i, z_i) = \frac{\exp[v(p, x_j, y_i, z_i)/\lambda_g]}{\sum_{k \in \text{category } g} \exp[v(p, x_k, y_i, z_i)/\lambda_g]} \frac{\sum_{k \in \text{category } g} \exp[v(p, x_k, y_i, z_i)/\lambda_g]^{\lambda_g}}{\sum_{g=1}^G \left(\sum_{k \in \text{category } g} \exp[v(p, x_k, y_i, z_i)/\lambda_g] \right)^{\lambda_g}}. \quad (4.74)$$

- Letting:

$$I_g(p, x, y_i, z_i) \equiv \log \sum_{k \in \text{category } g} \exp[v(p, x_k, y_i, z_i)/\lambda_g],$$

we have:

$$\sigma_j(p, x, y_i, z_i) = \frac{\exp[v(p, x_j, y_i, z_i)/\lambda_g]}{\sum_{k \in \text{category } g} \exp[v(p, x_k, y_i, z_i)/\lambda_g]} \frac{\exp[\lambda_g I_g(p, x, y_i, z_i)]}{\sum_{g=1}^G \exp[\lambda_g I_g(p, x, y_i, z_i)]}. \quad (4.75)$$

- The second first term can be interpreted as the probability of choosing product j conditional on choosing category g and the second term as the probability of choosing category g .

4.4.38 Discrete Choice Model with Unobserved Fixed Effects

- We have assumed that good j is characterized by a vector of observed characteristics x_j .
- Can econometrician observe all the relevant characteristics of the products in the choice set? Maybe no. For example, econometrician may not observe brand values that are created by advertisement and recognized by consumers.
- Such unobserved product characteristics is likely to be correlated with the price.
- This can cause **endogeneity problems**.
- In the following, we consider the situation where only market-segment level share data is available.
- Because we can construct the market-share level data from individual choice level data, all the arguments should go through with the individual choice level data.

4.4.39 Unobserved Fixed Effects in Multinomial Logit Model

- To fix the idea, let's revisit the multinomial logit model.
- For now, we do not consider either observed or unobserved consumer heterogeneity.
- Including observed heterogeneity is straightforward.
- We discuss how to include unobserved heterogeneity in the subsequent sections.
- Suppose that the indirect utility function of good j for consumer i in market t is:

$$\beta'x_{jt} - \alpha p_{jt} - \xi_{jt} + \epsilon_{ik}, \quad (4.76)$$

- ϵ_{ik} is i.i.d. Type-I extreme value.
- ξ_{jt} is the *unobserved product-market-specific fixed effect* of product j in market t , which can be correlated with p_{jt} .
- We hold the assumption that x_{jt} is uncorrelated with ξ_{jt} .
- The choice probability of good j for this consumer and hence the choice share in this market is:

$$\sigma_j(p_t, w_t, \xi_t) = \frac{\exp(\beta'x_j - \alpha p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^J \exp(\beta'x_k - \alpha p_{kt} + \xi_{kt})}. \quad (4.77)$$

- How to deal with the endogeneity between p_{jt} and ξ_{jt} ?

4.4.40 Instrumental Variables and Inversion

- Suppose that we have a vector of instrumental variables w_{jt} such that:

$$\mathbb{E}\{\xi_{jt}|w_{jt}\} = 0. \quad (4.78)$$

- In a liner model, we **invert** the model for the unobserved fixed effects:

$$\xi_{jt} = y_{jt} - \beta'x_{jt}, \quad (4.79)$$

- Notice that the unobserved fixed effect is written as a function of parameters and data.
- Then we exploit the moment condition by:

$$\begin{aligned} \mathbb{E}\{\xi_{jt}|w_{jt}\} &= 0, \\ \Rightarrow \mathbb{E}\{\xi_{jt}w_{jt}\} &= 0, \\ \Leftrightarrow \mathbb{E}\{(y_{jt} - \beta'x_{jt})w_{jt}\} &= 0 \end{aligned} \quad (4.80)$$

- We can estimate β by finding the value that makes the sample analogue of the above expectation zero.

4.4.41 Inversion in Multinomial Logit Model

- Can we invert the multinomial model for ξ_{jt} ?
- We have:

$$\begin{aligned} \ln[\sigma_{jt}(p_t, x_t, \xi_t)/\sigma_{0t}(p_t, x_t, \xi_t)] &= \beta'x_j - \alpha p_{jt} + \xi_{jt} \\ \Leftrightarrow \xi_{jt} &= \ln[\sigma_{jt}(p_t, x_t, \xi_t)/\sigma_{0t}(p_t, x_t, \xi_t)] - [\beta'x_j - \alpha p_{jt}]. \end{aligned} \quad (4.81)$$

- Therefore, the moment condition can be written as:

$$\begin{aligned} \mathbb{E}\{\xi_{jt}|w_{jt}\} &= 0, \\ \Rightarrow \mathbb{E}\{\xi_{jt}w_{jt}\} &= 0, \\ \Leftrightarrow \mathbb{E}\{(\ln[\sigma_{jt}(p_t, x_t, \xi_t)/\sigma_{0t}(p_t, x_t, \xi_t)] - [\beta'x_j - \alpha p_{jt}])w_{jt}\} &= 0. \end{aligned} \quad (4.82)$$

- We can evaluate the sample analogue of the expectation by replacing the theoretical choice probability σ with the observed share s .
- At the end, it is no different from the linear model where the dependent variable is $\ln s_{jt}/s_{0t}$.

4.4.42 Market-invariant Product-specific Fixed Effects

- Furthermore, if you can assume $\xi_{jt} = \xi_j$, then

$$\ln[\sigma_j(p_t, x_t, \xi_t)/\sigma_0(p_t, x_t, \xi_t)] = \beta'x_{jt} - \alpha p_{jt} + \xi_j. \quad (4.83)$$

- This is nothing but a linear regression on x_j and p_{jt} with product-specific unobserved fixed effect.
- This can be estimated by a within-estimator.
- This specification is a good starting point: we better start with the simplest specification and use the estimate as the initial guess for the following specifications.

4.4.43 Unobserved Consumer Heterogeneity and Unobserved Fixed Effects in Mixed-logit Model

- So far we abstracted away from the unobserved consumer heterogeneity.
- Next, suppose that the indirect utility function of good j for consumer i in market t is:

$$\beta'_i x_{jt} - \alpha_i p_{jt} - \xi_{jt} + \epsilon_{ik}, \quad (4.84)$$

where ϵ_{ik} is i.i.d. Type-I extreme value.

- The coefficient are drawn according to:

$$\begin{aligned} \beta_{it} &= \beta_0 + \Sigma \nu_{it}, \\ \alpha_{it} &= \alpha_0 + \Omega v_{it}, \end{aligned} \quad (4.85)$$

- ν_i are i.i.d. standard normal random variables.
- Then the indirect utility of good j for consumer i in market t is written as:

$$\underbrace{\beta'_0 x_{jt} - \alpha_0 p_{jt} + \xi_{jt}}_{\text{(conditional) mean}} + \underbrace{\nu'_{it} \Sigma x_{jt} - v'_{it} \Omega p_{jt}}_{\text{deviation from the mean}} \quad (4.86)$$

- We refer to β_0, α_0 as **linear parameters** and Σ, Ω as **non-linear parameters**, because of the reason I explain in the subsequent section.
- Let θ_1 be the linear parameters and θ_2 the non-linear parameters and let $\theta = (\theta'_1, \theta'_2)'$.

4.4.44 Unobserved Fixed Effects in Mixed-logit Model

- The choice share of good j in market t is:

$$\begin{aligned} & \sigma_j(p_t, x_t, \xi_t, y, z; \theta) \\ &= \int \frac{\exp[\beta'_0 x_{jt} - \alpha_0 p_{jt} + \xi_{jt} + \nu'_{it} \Sigma x_{jt} - v'_{it} \Omega p_{jt}]}{1 + \sum_{k \in \mathcal{J}_t} \exp[\exp[\beta'_0 x_{kt} - \alpha_0 p_{kt} + \xi_{kt} + \nu'_{it} \Sigma x_{kt} - v'_{it} \Omega p_{kt}]]} f(\nu, v) d\nu dv. \end{aligned} \quad (4.87)$$

- How can we represent ξ_{jt} as a function of parameters of interest to exploit the moment condition?

4.4.45 Representing ξ_{jt} as a Function of Parameters of Interest

- Let s_{jt} be the share of product j in market t .
- The following system of equations implicitly determines ξ_{jt} as a function of parameters of interest:

$$s_{jt} = \sigma_j(p_t, x_t, \xi_t; \theta). \quad (4.88)$$

- Let $\xi_{jt}(\theta)$ is the solution to the system of equations above given parameter θ .
- If it exists, it is the unobserved heterogeneity as a function of parameters and data.
- Does this solution exist?
- Is it unique?
- Is there efficient method to find the solution?

4.4.46 Summarizing the Conditional Mean Term

- Now, let δ_{jt} be the conditional mean term in the indirect utility:

$$\delta_{jt} \equiv \beta'_0 x_{jt} - \alpha_0 p_{jt} + \xi_{jt}. \quad (4.89)$$

- I call it the average utility of the product in the market.
- Then, the choice share of product j in market t is written as:

$$\begin{aligned} & \sigma_{jt}(\delta_t, \theta_2) \\ & \equiv \int \frac{\exp\left(\delta_{jt} + \nu' \Sigma x_{jt} - v' \Omega p_{jt}\right)}{1 + \sum_{k \in \mathcal{J}_t} \exp\left(\delta_{kt} + \nu' \Sigma x_{kt} - v' \Omega p_{kt}\right)} f(\nu, v) d\nu dv, \end{aligned} \quad (4.90)$$

for $j = 1, \dots, J, t = 1, \dots, T$.

4.4.47 Contraction Mapping for δ_t .

- Now, fix θ_2 and define an operator T such that:

$$T_t(\delta_t) = \delta_t + \underbrace{\ln s_{jt}}_{\text{data}} - \underbrace{\ln \sigma_{jt}(\delta_t, \theta_2)}_{\text{model}}. \quad (4.91)$$

- Let $\delta_t^{(0)} = (\delta_{1t}^{(0)}, \dots, \delta_{Jt}^{(0)})'$ be an arbitrary starting vector of average utility of products in a market.

- Using the operator above, we update $\delta_t^{(r)}$ by:

$$\delta_t^{(r+1)} = T_t(\delta_t^{(r)}) = \delta_t^{(r)} + \ln s_{jt} - \ln \sigma_{jt}(\delta_t^{(r)}, \theta_2), \quad (4.92)$$

for $r = 0, 1, \dots$.

- Berry et al. [1995] proved that T_t as specified above is a **contraction mapping with modulus less than one**.
- This means that:
 1. T_t has a unique fixed point;
 2. For arbitrary $\delta_t^{(r)}$, $\lim_{r \rightarrow \infty} T_t^r(\delta_t^{(0)})$ is the unique fixed point.
- The fixed point of T_t is δ_t^* such that $\delta_t^* = T_t(\delta_t^*)$, i.e.,

$$\begin{aligned} \delta_t^* &= \delta_t^* + \ln s_{jt} - \ln \sigma_{jt}(\delta_t^*, \theta_2), \\ &\Leftrightarrow s_{jt} = \sigma_{jt}(\delta_t^*, \theta_2). \end{aligned} \quad (4.93)$$

- So, the fixed point δ_t^* is the conditional mean indirect utility that solves the equality given non-linear parameter θ_2 .
- Moreover, the solution is unique.
- Moreover, it can be found by iterating the operator.
- Let $\delta_t(\theta_2)$ be the solution to this equation, i.e., the limit of this operation.
- The above result is useful because it ensures the inversion and provides the algorithm to find the solution.
- The invertibility itself holds under more general settings [Berry et al., 2013].

4.4.48 Solving for $\xi_{jt}(\theta)$

- We defined the average utility as:

$$\delta_{jt} = \beta'_0 x_{jt} - \alpha_0 p_{jt} + \xi_{jt}. \quad (4.94)$$

- Hence, if we set:

$$\xi_{jt}(\theta) \equiv \delta_t(\theta_2) - \left[\sum_{l=1}^L \beta_l w_{jl} - \alpha_0 p_{jt} \right], \quad (4.95)$$

the $\xi_{jt}(\theta)$ solves the equality:

$$s_{jt} = \sigma_j(p_t, x, \xi_t; \theta). \quad (4.96)$$

4.4.49 Solving for $\xi_{jt}(\theta)$: Summary

- In summary, ξ_{jt} that solves the equality exists and unique, and can be computed by:
- Fix $\theta = \{\theta_1, \theta_2\}$.
- Fix arbitrary starting value $\delta_t^{(0)}$ for $t = 1, \dots, T$.
- Let $\delta_t(\theta_2)$ be the limit of $T_t^r(\delta_t^{(0)})$ for $r = 0, 1, \dots$ for each $t = 1, \dots, T$.
- Stop the iteration if $|\delta_t(\theta_2)^{(r+1)} - \delta_t(\theta_2)^{(r)}|$ is below a threshold.
- Let $\xi_{jt}(\theta)$ be such that:

$$\xi_{jt}(\theta) = \delta_{jt}(\theta_2) - \beta'_0 x_{jt} - \alpha_0 p_{jt}. \quad (4.97)$$

- Then we can evaluate the moment at θ by:

$$\mathbb{E}\{\xi_{jt}(\theta) | w_{jt}\} = 0. \quad (4.98)$$

- We run this algorithm every time we evaluate the moment condition at a parameter value.

4.4.50 GMM Objective Function

- Find θ that solves:

$$\min_{\theta} \xi(\theta)' W \Phi^{-1} W' \xi(\theta), \quad (4.99)$$

where Φ is a weight matrix,

$$\xi(\theta) = \begin{pmatrix} \xi_{11}(\theta) \\ \vdots \\ \xi_{J_1 1}(\theta) \\ \vdots \\ \xi_{1T} \\ \vdots \\ \xi_{J_T T} \end{pmatrix}, W = \begin{pmatrix} w'_{11} \\ \vdots \\ w'_{J_1 1} \\ \vdots \\ w'_{1T} \\ \vdots \\ w'_{J_T T} \end{pmatrix}. \quad (4.100)$$

- There are $J \rightarrow \infty$ and $T \rightarrow \infty$ asymptotics. Either is fine to consistently estimate the parameters.
- $w_{jt} = (x'_{jt}, w^*_{jt})'$ where w^*_{jt} is an excluded instrument that is relevant to p_{jt} .

4.4.51 Estimating Linear Parameters

- The first-order condition for θ_1 is:

$$\theta_1 = (X_1' W \Phi^{-1} W' X_1)^{-1} X_1' W \Phi^{-1} W' \delta(\theta_2), \quad (4.101)$$

where

$$X_1 = \begin{pmatrix} x'_{11} & -p_{11} \\ \vdots & \vdots \\ x'_{J_1 1} & -p_{J_1 1} \\ \vdots & \vdots \\ x'_{1T} & -p_{1T} \\ \vdots & \vdots \\ x'_{J_T T} & -p_{J_T T} \end{pmatrix}, \delta(\theta_2) = \begin{pmatrix} \delta_1(\theta_2) \\ \vdots \\ \delta_T(\theta_2) \end{pmatrix} \quad (4.102)$$

- If θ_2 is given, the optimal θ_1 is computed by the above formula.
- \rightarrow We only have to search over θ_2 .
- This is the reason why we called θ_1 linear parameters and θ_2 non-linear parameters.

4.4.52 BLP Algorithm

- Find θ_2 that maximizes the GMM objective function.
- To do so:
 1. Pick up θ_2 .
 2. Compute $\delta(\theta_2)$ by the fixed-point algorithm.
 3. Compute associated θ_1 by the formula:

$$\theta_1 = (X_1' W \Phi^{-1} W' X_1)^{-1} X_1' W \Phi^{-1} W' \delta(\theta_2), \quad (4.103)$$

4. Compute $\xi(\theta)$ from the above $\delta(\theta_2)$ and θ_1 .
5. Evaluate the GMM objective function with the $\xi(\theta)$.

4.4.53 Mathematical Program with Equilibrium Constraints (MPEC)

- In the BLP algorithm, for each parameter θ , find $\xi(\theta)$ that solve:

$$s = \sigma(p, x, \xi; \theta) \quad (4.104)$$

by the fixed-point algorithm and then evaluate the GMM objective function.

- This inner loop takes time if the stopping criterion is tight.
- If the stopping criterion is loose, the loop may stop earlier but the error may be unacceptably large.
- Dubé et al. [2012] suggest to minimize the GMM objective function with the above equation as the constraints.

$$\min_{\theta} \xi(\theta)' W \Phi^{-1} W' \xi(\theta) \text{ s.t. } s = \sigma(p, x, \xi; \theta). \quad (4.105)$$

- To enjoy the benefit of this approach, we have to analytically derive the gradient and hessian of the objective function and the constraints, which are anyway needed if we estimate the standard error with the plug-in method.
- If the problem is of small scale, BLP algorithm will be fast enough and easier to implement.
- If the problem is of large scale, you may better use the MPEC approach.

4.4.54 Instrumental Variables

- The remaining problem is how to choose the excluded instrumental variable w_{jt}^* for each product/market.
- **Cost shifters:**
 - Traditional instruments.
- **Hausman-type IV** [Hausman et al., 1994]:
 - Assume that demand shocks are independent across markets, whereas the cost shocks are correlated.
 - The latter will be true if the product is produced by the same manufacturer.
 - Then, the price of the same product in the other markets $p_{j,-t}$ will be valid instruments for the price of the product in a given market, p_{jt} .
- **BLP-type IV** [Berry et al., 1995]:
 - In oligopoly, the price of a good in a market depends on the market structure, i.e., what kind of products are available in the market.
 - For example, if there are similar products in the market, the price will tend to be lower.
 - Then, the product characteristics of other products in the market, will be valid instrument for the price of goods in a given market, p_{jt} .
 - If there are multi-product firms, whether the other good is owned by the same company will also affect the price.
 - Specifically, Berry et al. [1995] use:

$$\sum_{k \neq j \in \mathcal{J}_t \cap \mathcal{F}_f} x_{kt}, \quad (4.106)$$

$$\sum_{k \neq j \in \mathcal{J}_t \setminus \mathcal{F}_f} x_{kt}. \quad (4.107)$$

- f is the firm that owns product j and \mathcal{F}_f is the set of products firm f owns.
- **Differentiation IV** [Gandhi and Houde, 2015]:
 - Let $d_{jkt} = d(x_{jt}, x_{kt})$ be some distance between product characteristics.
 - They showed that under certain conditions the optimal BLP-type IV is a function $d_{-jt}\{d_{jkt}\}_{k \neq j \in \mathcal{J}_t}$.
 - The suggest to use the moments of d_{-jt} as the excluded instrument variables.

- Weak instruments problem of BLP-type IV:
 - Armstrong [2016] argued that estimates based on BLP-type IV may be inconsistent when $J \times \infty$ asymptotics is considered, because then the market approaches the competitive market and the correlation between the markup and the product characteristics of the rivals disappear.
 - Specifically, the estimator is inconsistent if all of the following conditions are met:
 1. $J \rightarrow \infty$ but T is fixed;
 2. The demand/cost functions are such that the correlation between markups and characteristics of other products decreases quickly enough as $J \rightarrow \infty$.
 3. There is no cost instruments or other sources of identification.

Chapter 5

Assignment 1: Basic Programming in R

The deadline is the **start time of February 18 class**.

Report the following results in html format using R markdown. In other words, replicate this document. You write functions in a separate R file and put in R folder in the project folder. Build the project as a package and load it from the R markdown file. The execution code should be written in R markdown file.

You submit:

- R file containing functions.
- R markdown file containing your answers and executing codes.
- HTML report generated from the R markdown.

5.1 Simulate data

Consider to simulate data from the following model and estimate the parameters from the simulated data.

$$y_{ij} = 1\{j = \operatorname{argmax}_{k=1,2} \beta x_k + \epsilon_{ik}\},$$

where ϵ_{ik} follows i.i.d. type-I extreme value distribution, $\beta = 0.2$, $x_1 = 0$ and $x_2 = 1$.

1. To simulate data, first make a data frame as follows:

```
## # A tibble: 2,000 x 3
##       i     k     x
##   <int> <int> <dbl>
## 1     1     1     0
## 2     1     2     1
## 3     2     1     0
## 4     2     2     1
## 5     3     1     0
## 6     3     2     1
## 7     4     1     0
## 8     4     2     1
## 9     5     1     0
## 10    5     2     1
## # ... with 1,990 more rows
```

2. Second, draw type-I extreme value random variables. Set the seed at 1. You can use `evd` package to draw the variables. You should get exactly the same realization if the seed is correctly set.

```
## # A tibble: 2,000 x 4
##       i     k     x     e
##   <int> <int> <dbl>  <dbl>
## 1     1     1     0  0.281
## 2     1     2     1 -0.167
## 3     2     1     0  1.93
## 4     2     2     1  1.97
## 5     3     1     0  0.830
## 6     3     2     1 -1.06
## 7     4     1     0 -0.207
## 8     4     2     1  0.617
## 9     5     1     0  0.0444
## 10    5     2     1  1.92
## # ... with 1,990 more rows
```

3. Third, compute the latent value of each option to obtain the following data frame:

```
## # A tibble: 2,000 x 5
##       i     k     x     e latent
##   <int> <int> <dbl>  <dbl>  <dbl>
## 1     1     1     0  0.281  0.281
## 2     1     2     1 -0.167  0.0331
## 3     2     1     0  1.93   1.93
## 4     2     2     1  1.97   2.17
## 5     3     1     0  0.830  0.830
## 6     3     2     1 -1.06  -0.863
## 7     4     1     0 -0.207 -0.207
## 8     4     2     1  0.617  0.817
## 9     5     1     0  0.0444 0.0444
## 10    5     2     1  1.92   2.12
## # ... with 1,990 more rows
```

4. Finally, compute y by comparing the latent values of $k = 1, 2$ for each i to obtain the following result:

```
## # A tibble: 2,000 x 6
##       i     k     x     e latent  y
##   <int> <int> <dbl>  <dbl>  <dbl> <dbl>
## 1     1     1     0  0.281  0.281    1
## 2     1     2     1 -0.167  0.0331    0
## 3     2     1     0  1.93   1.93    0
## 4     2     2     1  1.97   2.17    1
## 5     3     1     0  0.830  0.830    1
## 6     3     2     1 -1.06  -0.863    0
## 7     4     1     0 -0.207 -0.207    0
## 8     4     2     1  0.617  0.817    1
## 9     5     1     0  0.0444 0.0444    0
## 10    5     2     1  1.92   2.12    1
## # ... with 1,990 more rows
```

5.2 Estimate the parameter

Now you generated simulated data. Suppose you observe x_k and y_{ik} for each i and k and estimate β by a maximum likelihood estimator. The likelihood for i to choose k ($y_{ik} = 1$) can be shown to be:

$$p_{ik}(\beta) = \frac{\exp(\beta x_k)}{\exp(\beta x_1) + \exp(\beta x_2)}.$$

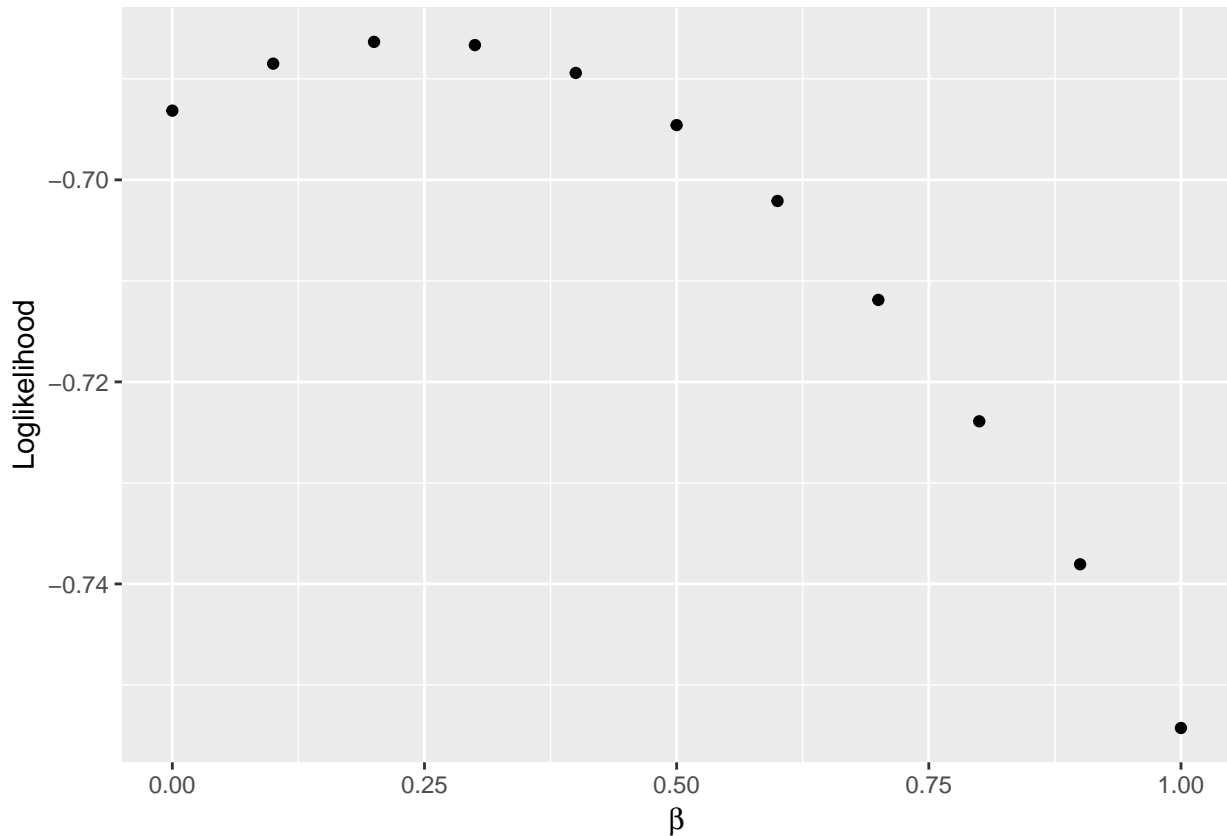
Then, the likelihood of observing $\{y_{ik}\}_{i,k}$ is:

$$L(\beta) = \prod_{i=1}^{1000} p_{i1}(\beta)^{y_{i1}} [1 - p_{i1}(\beta)]^{1-y_{i1}},$$

and the log likelihood is:

$$l(\beta) = \sum_{i=1}^{1000} \{y_{i1} \log p_{i1}(\beta) + (1 - y_{i1}) \log [1 - p_{i1}(\beta)]\}.$$

1. Write a function to compute the likelihood for a given β and data and name the function `loglikelihood_A1`.
2. Compute the value of log likelihood for $\beta = 0, 0.1, \dots, 1$ and plot the result using `ggplot2` packages. You can use `latex2exp` package to use LaTeX math symbol in the label:



1. Find and report β that maximizes the log likelihood for the simulated data. You can use `optim` function to achieve this. You will use `Brent` method and set the lower bound at -1 and upper bound at 1 for the parameter search.

```
## $par
## [1] 0.2371046
##
## $value
## [1] -0.6861689
##
## $counts
## function gradient
##      NA      NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

Chapter 6

Assignment 2: Production Function Estimation

The deadline is the **February 28 1:30pm**.

6.1 Simulate data

Consider the following production and investment process for $j = 1, \dots, 1000$ firms across $t = 1, \dots, 10$ periods.

The log production function is of the form:

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt} + \eta_{jt},$$

where ω_{jt} is an anticipated shock and η_{jt} is an ex post shock.

The anticipated shocks evolve as:

$$\omega_{jt} = \alpha \omega_{j,t-1} + \nu_{jt},$$

where ν_{jt} is an i.i.d. normal random variable with mean 0 and standard deviation σ_ν . The ex post shock is an i.i.d. normal random variable with mean 0 and standard deviation σ_η .

The product price the same across firms and normalized at 1. The price is normalized at 1. The wage w_t is constant at 0.5.

Finally, the capital accumulate according to:

$$K_{j,t+1} = (1 - \delta)K_{jt} + I_{jt}.$$

We set the parameters as follows:

parameter	variable	value
β_0	beta_0	1
β_l	beta_l	0.2
β_k	beta_k	0.7
α	alpha	0.7
σ_η	sigma_eta	0.2
σ_ν	sigma_nu	0.5
σ_w	sigma_w	0.1
δ	delta	0.05

1. Define the parameter variables as above.
2. Write a function that returns the log output given l_{jt} , k_{jt} , ω_{jt} , and η_{jt} under the given parameter values according to the above production function and name it `log_production(l, k, omega, eta, beta_0, beta_1, beta_k)`.

Suppose that the labor is determined after ω_{jt} is observed, but before η_{jt} is observed, given the log capital level k_{jt} .

3. Derive the optimal log labor as a function of ω_{jt} , η_{jt} , k_{jt} , and wage. Write a function to return the optimal log labor given the variables and parameters and name it `log_labor_choice(k, wage, omega, beta_0, beta_1, beta_k, sigma_eta)`.

As discussed in the class, if there is no additional variation in labor, the coefficient on the labor β_l is not identified. Thus, if we generate labor choice from the previous function, β_l will not be identified from the simulated data. To see this, we write a modified version of the previous function in which ω_{jt} is replaced with $\omega_{jt} + \iota_{jt}$, where ι_{jt} is an optimization error that follows an i.i.d. normal distribution with mean 0 and standard deviation 0.05. That is, the manager of the firm perceives as if the shock is $\omega_{jt} + \iota_{jt}$, even though the true shock is ω_{jt} .

4. Modify the previous function by including ι_{jt} as an additional input and name it `log_labor_choice_error(k, wage, omega, beta_0, beta_1, beta_k, iota, sigma_eta)`.

Consider an investment process such that:

$$I_{jt} = (\delta + \gamma\omega_{jt})K_{jt},$$

where I_{jt} and K_{jt} are investment and capital in level. Set $\gamma = 0.1$, i.e., the investment is strictly increasing in ω_{jt} . The investment function should be derived by solving the dynamic problem of a firm. But here, we just specify it in a reduced-form.

5. Define variable γ and assign it the value. Write a function that returns the investment given K_{jt} , ω_{jt} , and parameter values, according to the previous equation, and name it `investment_choice(k, omega, gamma, delta)`.

Simulate the data first using the labor choice without optimization error and second using the labor choice with optimization error. To do so, we specify the initial values for the state variables k_{j1} and ω_{j1} as follows.

6. Draw k_{j1} from an i.i.d. normal distribution with mean 1 and standard deviation 0.5. Draw ω_{j1} from its stationary distribution (check the stationary distribution of AR(1) process). Draw a wage. Before simulating the rest of the data, set the seed at 1.

```
## # A tibble: 1,000 x 5
##       j     t     k  omega  wage
##   <int> <dbl> <dbl>   <dbl> <dbl>
## 1     1     1  0.687  0.795   0.5
## 2     2     2  1.09   0.779   0.5
## 3     3     3  0.582 -0.610   0.5
## 4     4     4  1.80   0.148   0.5
## 5     5     5  1.16   0.0486  0.5
## 6     6     6  0.590 -1.16    0.5
## 7     7     7  1.24   0.568   0.5
## 8     8     8  1.37  -1.34    0.5
## 9     9     9  1.29  -0.873   0.5
## 10    10    1  0.847  0.699   0.5
## # ... with 990 more rows
```

7. Draw optimization error ι_{jt} and compute the labor and investment choice of period 1. For labor choice, compute both types of labor choices.

```
## # A tibble: 1,000 x 9
##       j      t      k  omega wage      iota      l l_error      I
##   <int> <dbl> <dbl>  <dbl> <dbl>  <dbl> <dbl>  <dbl>  <dbl>
## 1     1     1  0.687  0.795   0.5 -0.0443  1.72  1.67  0.257
## 2     2     2  1.109  0.779   0.5 -0.0961  2.06  1.94  0.381
## 3     3     3  0.582 -0.610   0.5  0.0810 -0.123 -0.0218 -0.0196
## 4     4     4  1.180  0.148   0.5  0.0260  1.89  1.92  0.391
## 5     5     5  1.116  0.0486  0.5 -0.00279 1.21  1.21  0.176
## 6     6     6  0.590 -1.16    0.5  0.0348 -0.809 -0.766 -0.120
## 7     7     7  1.124  0.568   0.5  0.00268 1.93  1.93  0.370
## 8     8     8  1.137 -1.34    0.5 -0.0655 -0.346 -0.428 -0.330
## 9     9     9  1.129 -0.873   0.5 -0.106   0.165  0.0327 -0.135
## 10    10    10  0.847  0.699   0.5 -0.0104  1.74  1.73  0.280
## # ... with 990 more rows
```

8. Draw ex post shock and compute the output according to the production function for both labor without optimization error and with optimization error. Name the output without optimization error y and the one with optimization error y_error .

```
## # A tibble: 1,000 x 12
##       j      t      k  omega wage      iota      l l_error      I      eta
##   <int> <dbl> <dbl>  <dbl> <dbl>  <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1     1     1  0.687  0.795   0.5 -0.0443  1.72  1.67  0.257  0.148
## 2     2     2  1.109  0.779   0.5 -0.0961  2.06  1.94  0.381  0.0773
## 3     3     3  0.582 -0.610   0.5  0.0810 -0.123 -0.0218 -0.0196 0.259
## 4     4     4  1.180  0.148   0.5  0.0260  1.89  1.92  0.391 -0.161
## 5     5     5  1.116  0.0486  0.5 -0.00279 1.21  1.21  0.176 -0.321
## 6     6     6  0.590 -1.16    0.5  0.0348 -0.809 -0.766 -0.120  0.187
## 7     7     7  1.124  0.568   0.5  0.00268 1.93  1.93  0.370  0.361
## 8     8     8  1.137 -1.34    0.5 -0.0655 -0.346 -0.428 -0.330 -0.0113
## 9     9     9  1.129 -0.873   0.5 -0.106   0.165  0.0327 -0.135  0.377
## 10    10    10  0.847  0.699   0.5 -0.0104  1.74  1.73  0.280  0.316
## # ... with 990 more rows, and 2 more variables: y <dbl>, y_error <dbl>
```

9. Repeat this procedure for $t = 1, \dots, 10$ by updating the capital and anticipated shocks, and name the resulting data frame df_T .

```
## # A tibble: 10,000 x 13
##       j      t      k  omega wage      iota      l l_error      I      eta
##   <int> <dbl> <dbl>  <dbl> <dbl>  <dbl> <dbl>  <dbl>  <dbl>  <dbl>
## 1     1     1  0.687  0.795   0.5 -0.0443  1.72  1.67  0.257  0.148
## 2     2     2  1.109  0.779   0.5 -0.0961  2.06  1.94  0.381  0.0773
## 3     3     3  0.582 -0.610   0.5  0.0810 -0.123 -0.0218 -0.0196 0.259
## 4     4     4  1.180  0.148   0.5  0.0260  1.89  1.92  0.391 -0.161
## 5     5     5  1.116  0.0486  0.5 -0.00279 1.21  1.21  0.176 -0.321
## 6     6     6  0.590 -1.16    0.5  0.0348 -0.809 -0.766 -0.120  0.187
## 7     7     7  1.124  0.568   0.5  0.00268 1.93  1.93  0.370  0.361
## 8     8     8  1.137 -1.34    0.5 -0.0655 -0.346 -0.428 -0.330 -0.0113
## 9     9     9  1.129 -0.873   0.5 -0.106   0.165  0.0327 -0.135  0.377
## 10    10    10  0.847  0.699   0.5 -0.0104  1.74  1.73  0.280  0.316
## # ... with 9,990 more rows, and 3 more variables: y <dbl>, y_error <dbl>,
## #   nu <dbl>
```

10. Check the simulated data by making summary table.

	N	Mean	Sd	Min	Max
j	10000	500.5000000	288.6894251	1.0000000	1000.0000000
t	10000	5.5000000	2.8724249	1.0000000	10.0000000
k	10000	0.9797900	0.5838949	-1.2822534	3.2332312
omega	10000	-0.0055826	0.7025102	-2.5894171	2.6281307
wage	10000	0.5000000	0.0000000	0.5000000	0.5000000
iota	10000	-0.0000696	0.0502883	-0.1841453	0.1715419
l	10000	0.9799746	1.0965108	-3.3281023	4.9679634
l_error	10000	0.9798876	1.0971595	-3.3765433	4.9520674
I	10000	0.1793502	0.3006526	-1.2722627	3.2975332
eta	10000	0.0015825	0.2001539	-0.7650371	0.7455922
y	10000	1.8778479	1.1171035	-2.4680251	6.1228291
y_error	10000	1.8778305	1.1169266	-2.4777133	6.1196499
nu	10000	-0.0021155	0.4984324	-2.1513907	1.8253882

6.2 Estimate the parameters

For now, use the labor choice with optimization error.

1. First, simply regress y_{jt} on l_{jt} and k_{jt} using the least square method. This is likely to give an upwardly biased estimates on β_l and β_k . Why is it?

```
##
## Call:
## lm(formula = y_error ~ l_error + k, data = df_T)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.73002 -0.14117 -0.00071  0.13743  0.87983
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.892542   0.004058  219.966  <2e-16 ***
## l_error      0.997913   0.002396  416.454  <2e-16 ***
## k            0.007599   0.004503   1.688   0.0915 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2068 on 9997 degrees of freedom
## Multiple R-squared:  0.9657, Adjusted R-squared:  0.9657
## F-statistic: 1.408e+05 on 2 and 9997 DF,  p-value: < 2.2e-16
```

2. Second, take within-transformation on y_{jt} , l_{jt} , and k_{jt} and let Δy_{jt} , Δl_{jt} , and Δk_{jt} denote them. Then, regress Δy_{jt} on Δl_{jt} , and Δk_{jt} by the least squares method.

```
##
## Call:
## lm(formula = dy_error ~ -1 + dl_error + dk, data = df_T_within)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.72450 -0.13285 -0.00244  0.12931  0.77657
##
## Coefficients:
```



```
##           Estimate Std. Error t value Pr(>|t|)
## dl_error  0.9910916  0.0029548 335.413  <2e-16 ***
## dk        -0.0009029  0.0127539  -0.071   0.944
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1961 on 9998 degrees of freedom
## Multiple R-squared:  0.9184, Adjusted R-squared:  0.9184
## F-statistic: 5.629e+04 on 2 and 9998 DF,  p-value: < 2.2e-16
```

Next, we attempt to estimate the parameters using Olley-Pakes method. Estimate the first-step model of Olley-Pakes method:

$$y_{jt} = \beta_0 + \beta_1 l_{jt} + \phi(k_{jt}, I_{jt}) + \eta_{jt},$$

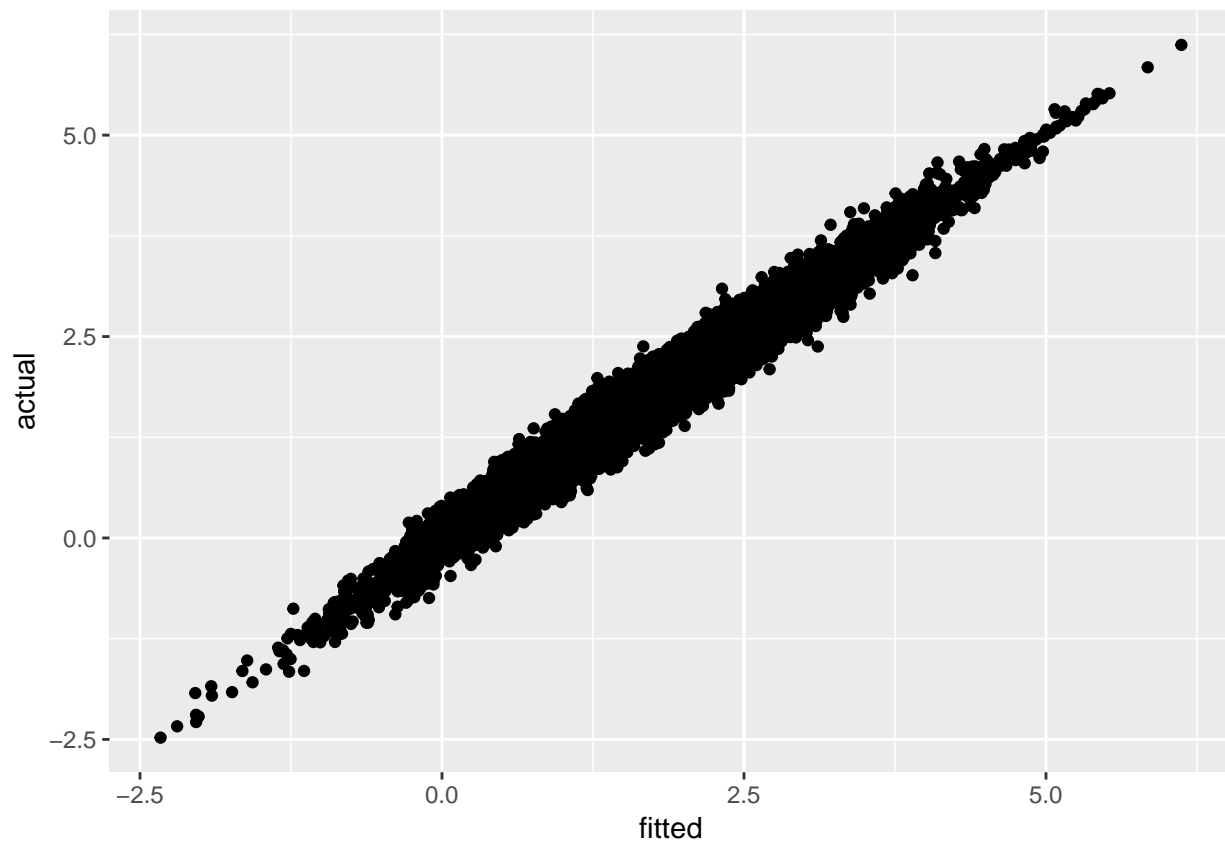
by approximating ϕ_t by a kernel function.

Remark that ϕ in general depends on observed and unobserved state variables. For this reason, in theory, ϕ should be estimated for each period. In this exercise, we assume ϕ is common across periods because we know that there is no unobserved state variables in the true data generating process. Moreover, we do not include w_t because we know that it is time -invariant. Do not forget to consider them in the actual data analysis.

You can use `npplreg` function of `np` package to estimate a partially linear model with a multivariate kernel. You first use `npplregbw` to obtain the optimal band width and then use `npplreg` to estimate the model under the optimal bandwidth. The computation of the optimal bandwidth is time consuming.

3. Return the summary of the first stage estimation and plot the fitted values against the data points.

```
##
## Partially Linear Model
## Regression data: 10000 training points, in 5 variable(s)
## With 3 linear parametric regressor(s), 2 nonparametric regressor(s)
##
##           y(z)
## Bandwidth(s): 0.07355058 0.01435558
##
##           x(z)
## Bandwidth(s): 0.03908594 0.01191551
##              0.01397428 3.74952954
##              0.83529394 0.00398329
##
##           l_error      k      I
## Coefficient(s): 0.2485295 2.355522 5.345144
##
## Kernel Regression Estimator: Local-Constant
## Bandwidth Type: Fixed
##
## Residual standard error: 0.1934585
## R-squared: 0.970064
##
## Continuous Kernel Type: Second-Order Gaussian
## No. Continuous Explanatory Vars.: 2
```



3. Check that β_l is not identified with the data without optimization error. Estimate the first stage model of Olley-Pakes with the labor choice without optimization error and report the result.

```
##
## Partially Linear Model
## Regression data: 10000 training points, in 5 variable(s)
## With 3 linear parametric regressor(s), 2 nonparametric regressor(s)
##
##           y(z)
## Bandwidth(s): 0.07347226 0.01437256
##
##           x(z)
## Bandwidth(s): 0.02960021 0.009986945
##               0.01397428 3.749529542
##               0.83529394 0.003983290
##
##           l       k       I
## Coefficient(s): 1.180628 2.034182 0.7805077
##
## Kernel Regression Estimator: Local-Constant
## Bandwidth Type: Fixed
##
## Residual standard error: 0.1932285
## R-squared: 0.970116
##
## Continuous Kernel Type: Second-Order Gaussian
## No. Continuous Explanatory Vars.: 2
```

Then, we estimate the second stage model of Olley-Pakes method:

$$y_{jt} - \hat{\beta}_l l_{jt} = \beta_0 + \beta_k k_{jt} + \alpha[\hat{\phi}(k_{j,t-1}, I_{j,t-1}) - \beta_0 - \beta_k k_{jt}] + \nu_{jt} + \eta_{jt}.$$

In this model, we do not have to non-parametrically estimate the conditional expectation of ω_{jt} on $\omega_{j,t-1}$, because we know that the anticipated shock follows an AR(1) process. Remark that we in general have to non-parametrically estimate it.

The model is non-linear in parameters, because of the term $\alpha\beta_0$ and $\alpha\beta_k$. We estimate α , β_0 , and β_k by a GMM estimator. The moment is:

$$g_{JT}(\alpha, \beta_0, \beta_k) \equiv \frac{1}{JT} \sum_{j=1}^J \sum_{t=1}^T \{y_{jt} - \hat{\beta}_l l_{jt} - \beta_0 - \beta_k k_{jt} - \alpha[\hat{\phi}(k_{j,t-1}, I_{j,t-1}) - \beta_0 - \beta_k k_{jt}]\} \begin{bmatrix} k_{jt} \\ k_{j,t-1} \\ I_{j,t-1} \end{bmatrix}.$$

4. Using the estimates in the first step, compute:

$$y_{jt} - \hat{\beta}_l l_{jt},$$

and:

$$\hat{\phi}(k_{j,t-1}, I_{j,t-1}),$$

for each j and t and save it as a data frame names `df_T_1st`.

```
## # A tibble: 10,000 x 4
##       j      t y_error_tilde phi_t_1
##   <int> <dbl>      <dbl>    <dbl>
## 1     1     1      2.34      NA
## 2     1     2      1.37     2.21
## 3     1     3      0.621    1.49
## 4     1     4      0.447    0.882
## 5     1     5      0.878    0.611
## 6     1     6      1.62     0.926
## 7     1     7      0.558    1.40
## 8     1     8      0.684    0.439
## 9     1     9      0.939    0.520
## 10    1    10      1.49     0.836
## # ... with 9,990 more rows
```

3. Compute a function that returns the value of $g_{JT}(\alpha, \beta_0, \beta_k)$ given parameter values, data, and `df_T_1st`, and name it `moment_OP_2nd`. Show the values of the moments evaluated at the true parameters.

```
## [1] -0.018507303 -0.019038229 -0.003867714
```

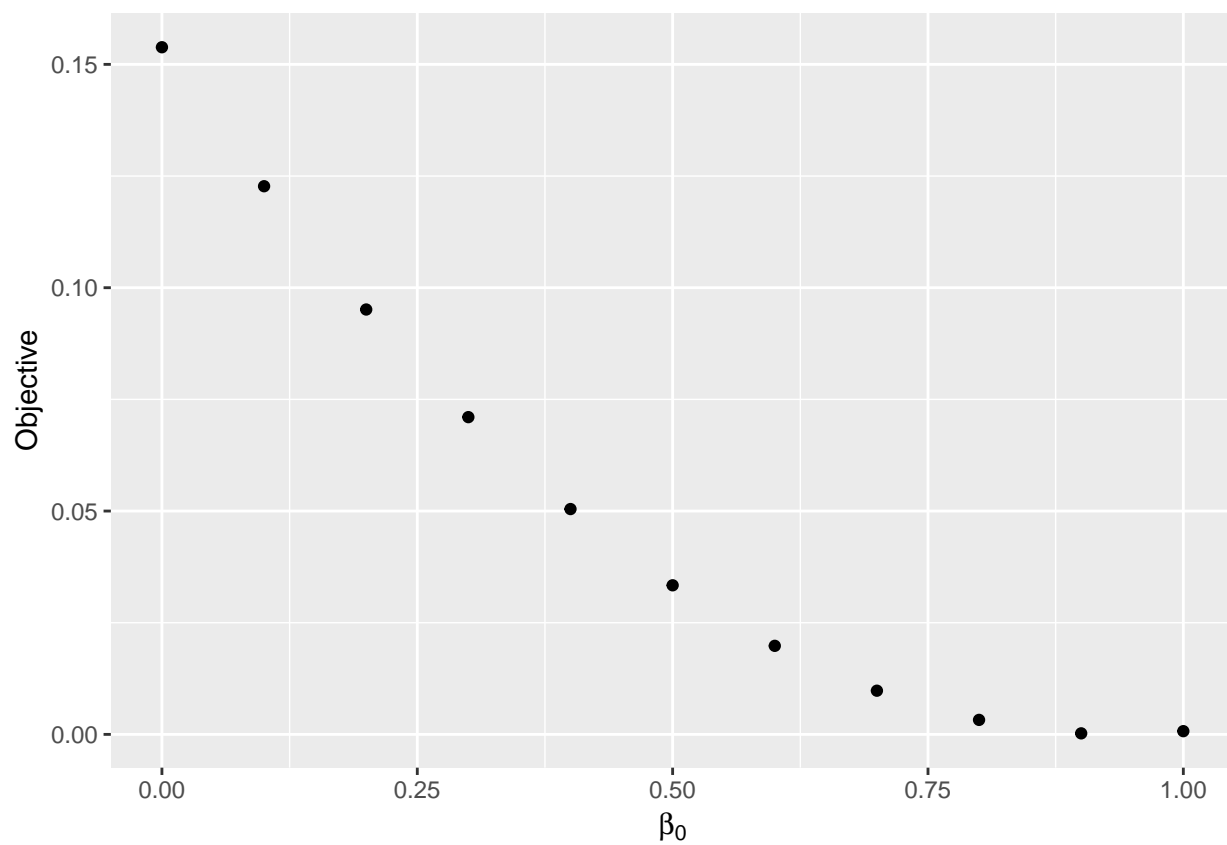
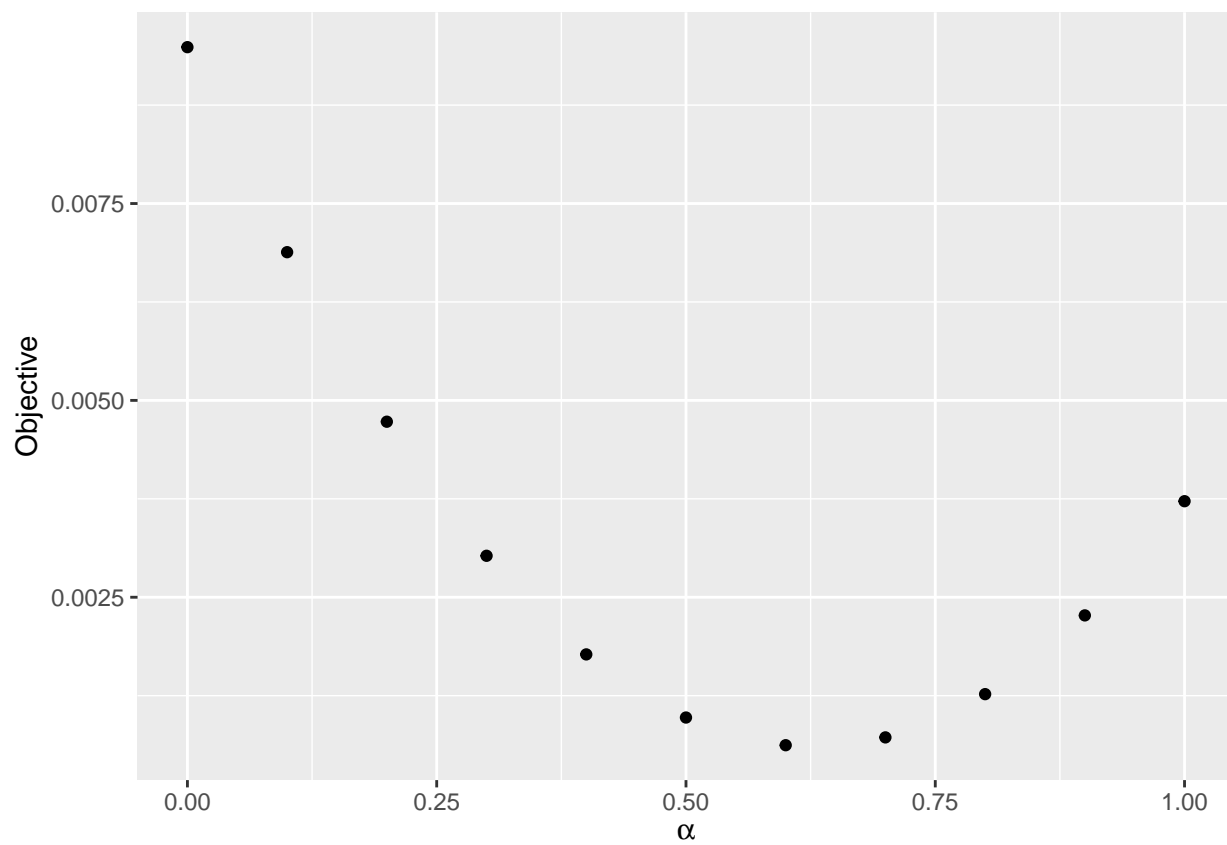
Based on the moment, we can define the objective function of a generalized method of moments estimator with a weighting matrix W as:

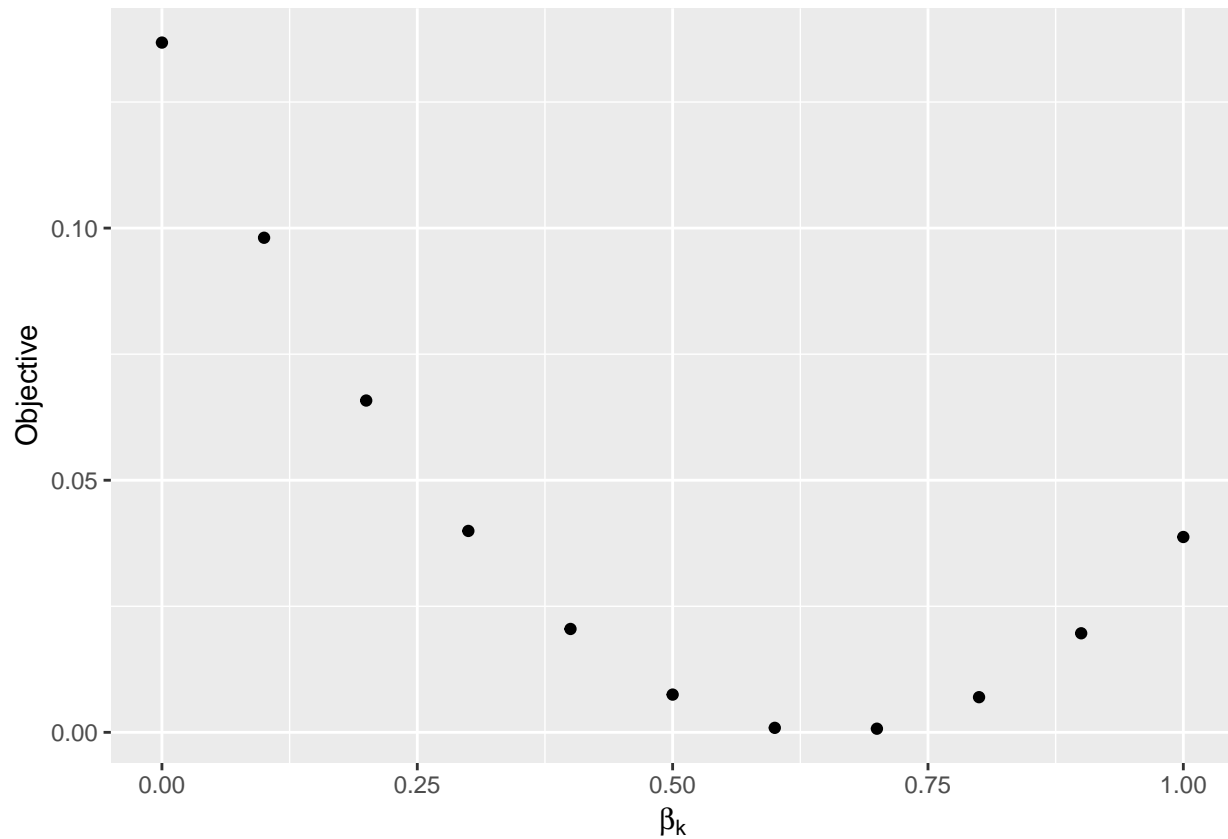
$$Q_{JT}(\alpha, \beta_0, \beta_k) \equiv g_{JT}(\alpha, \beta_0, \beta_k)' W g_{JT}(\alpha, \beta_0, \beta_k).$$

4. Write a function that returns the value of $Q_{JT}(\alpha, \beta_0, \beta_k)$ given the vector of parameter values, data, and `df_T_1st`, and name it `objective_OP_2nd`. Setting W at the identity matrix, show the value of the objective function evaluated at the true parameters.

```
##           [,1]
## [1,] 0.0007199336
```

5. Draw the graph of the objective function when one of α , β_0 , and β_k are changed from 0 to 1 by 0.1 while the others are set at the true value. Is the objective function minimized at around the true value?





6. Find the parameters that minimize the objective function using `optim`. You may use L-BFGS-B method to solve it.

```
## $par
## [1] 0.7020260 0.9766308 0.6693945
##
## $value
## [1] 1.994601e-07
##
## $counts
## function gradient
##      10      10
##
## $convergence
## [1] 0
##
## $message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
```


Bibliography

- Daniel A Akerberg. Timing Assumptions and Efficiency: Empirical Evidence in a Production Function Context. Technical report, 2016. URL <http://economics.mit.edu/files/12328>.
- Daniel A. Akerberg, Kevin Caves, and Garth Frazer. Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6):2411–2451, nov 2015. ISSN 0012-9682. doi: 10.3982/ECTA13408. URL <https://www.econometricsociety.org/doi/10.3982/ECTA13408>.
- S N Afriat. The Construction of Utility Functions from Expenditure Data. *International Economic Review*, 8(1), 1967. URL <https://about.jstor.org/terms>.
- Hyungtaik Ahn and James L. Powell. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1-2):3–29, jul 1993. ISSN 0304-4076. doi: 10.1016/0304-4076(93)90111-H. URL <https://www.sciencedirect.com/science/article/pii/030440769390111H>.
- Angus Deaton and John Muellbauer. *Economics and Consumer Behavior*. Cambridge University Press, New York, 1980. URL <http://www.library.fa.ru/files/Deaton-Economics.pdf>.
- Nikolay Archak, Anindya Ghose, and Panagiotis G. Ipeirotis. Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, 57(8):1485–1509, 2011. ISSN 0025-1909. doi: 10.1287/mnsc.1110.1370. URL <http://pubsonline.informs.orghttp://www.informs.org>.
- Timothy B Armstrong. Large Market Asymptotics for Differentiated Product Demand Estimators with Economic Models of Supply. *Econometrica*, 84(5):1961–1980, 2016. doi: 10.3982/ECTA10600. URL <https://www.econometricsociety.org/system/files/ecta1632.pdf>.
- Patrick Bayer, Fernando Ferreira, Robert Mcmillan, Joseph Altonji, Pat Bajari, Steve Berry, Sandra Black, David Card, Ken Chay, David Cutler, Hanming Fang, David Figlio, Edward Glaeser, David Lee, Steven Levitt, Enrico Moretti, Tom Nechyba, Jesse Rothstein, Kim Rueben, Holger Sieg, Chris Taber, and Chris Timmins. A Unified Framework for Measuring Preferences for Schools and Neighborhoods. Technical Report 4, 2007. URL <http://www.journals.uchicago.edu/t-and-c>.
- Steven Berry and Ariel Pakes. The Pure Characteristics Demand Model. *International Economic Review*, 48(4):1193–1225, dec 2007. doi: 10.1111/j.1468-2354.2007.00459.x. URL <http://doi.wiley.com/10.1111/j.1468-2354.2007.00459.x>.
- Steven Berry, James Levinsohn, and Ariel Pakes. Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841–890, 1995. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.4697-10.2011.
- Steven Berry, Amit Gandhi, and Philip Haile. Connected Substitutes and Invertibility of Demand. *Econometrica*, 81(5):2087–2111, sep 2013. ISSN 0012-9682. doi: 10.3982/ECTA10135. URL <http://doi.wiley.com/10.3982/ECTA10135>.
- N. Bloom and J. Van Reenen. Measuring and Explaining Management Practices Across Firms and Countries. *The Quarterly Journal of Economics*, 122(4):1351–1408, nov 2007. ISSN 0033-5533. doi: 10.1162/qjec.2007.122.4.1351. URL <https://academic.oup.com/qje/article-lookup/doi/10.1162/qjec.2007.122.4.1351>.

- Laura Blow, Martin Browning, and Ian Crawford. Revealed Preference Analysis of Characteristics Models. Technical report, 2008. URL <https://academic.oup.com/restud/article-abstract/75/2/371/1621743>.
- Serguey Braguinsky, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson. Acquisitions, Productivity, and Profitability: Evidence from the Japanese Cotton Spinning Industry. *American Economic Review*, 105(7): 2086–2119, jul 2015. ISSN 0002-8282. doi: 10.1257/aer.20140150. URL <http://pubs.aeaweb.org/doi/10.1257/aer.20140150>.
- Timothy F. Bresnahan. Departures from marginal-cost pricing in the American automobile industry: Estimates for 1977–1978. *Journal of Econometrics*, 17(2):201–227, nov 1981. ISSN 0304-4076. doi: 10.1016/0304-4076(81)90027-0. URL <https://www.sciencedirect.com/science/article/pii/0304407681900270>.
- Timothy F. Bresnahan. The Empirical Renaissance in Industrial Economics. *The Journal of Industrial Economics*, 35(4):457–482, 1987. URL <https://www.jstor.org/stable/pdf/2098583.pdf?refreqid=excelsior%7B3A9bdf36ab7f36cc46c27d533a749d00f3>.
- Timothy F. Bresnahan. Chapter 17 Empirical studies of industries with market power. In *Handbook of Industrial Organization*, volume 2, pages 1011–1057. Elsevier, jan 1989. ISBN 9780444704351. doi: 10.1016/S1573-448X(89)02005-4. URL <https://www.sciencedirect.com/science/article/pii/S1573448X89020054>.
- David P. Byrne, Susumu Imai, Neelam Jain, Vasilis Sarafidis, and Masayuki Hirukawa. Identification and Estimation of Differentiated Products Models using Market Size and Cost Data. Technical report, 2015. URL https://www.city.ac.uk/{_}{_}data/assets/pdf/{_}file/0019/286012/blpcost15_{_}final2_{_}Byrne_{_}Imai_{_}Jain_{_}Sarafidis_{_}Hirukawa.pdf.
- Xiaohong Chen. Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models. *Handbook of Econometrics*, 6:5549–5632, jan 2007. ISSN 1573-4412. doi: 10.1016/S1573-4412(07)06076-X. URL <https://www.sciencedirect.com/science/article/pii/S157344120706076X>.
- Clayton M. Christensen. *The innovator’s dilemma : when new technologies cause great firms to fail*. HarperBusiness, New York, 1997. ISBN 142219602X. URL https://books.google.co.jp/books?hl=ja&lr=lang_{_}ja{%7C}lang_{_}en{%7C}id=3JnBAGAAQBAJ{%7C}oi=fnd{%7C}pg=PR4{%7C}dq=innovator{%7C}27s+dilemma{%7C}ots=G4y61Qdl_{_}2{%7C}sig=Yy7KYncJzItvyAfACoZITQcg49s{#}v=onepage{%7C}q=innovator’sdilemma{%7C}f=false.
- Wilbur Chung and Juan Alcácer. Knowledge Seeking and Location Choice of Foreign Direct Investment in the United States. *Management Science*, 48(12), 2002. doi: 10.1287/mnsc.48.12.1534.440.
- Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, 78(3):883–931, may 2010. ISSN 0012-9682. doi: 10.3982/ECTA6551. URL <http://doi.wiley.com/10.3982/ECTA6551>.
- Jan De Loecker. Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica*, 79(5):1407–1451, sep 2011. ISSN 0012-9682. doi: 10.3982/ECTA7617. URL <http://doi.wiley.com/10.3982/ECTA7617>.
- Angus Deaton and John Muellbauer. An Almost Ideal Demand System. Technical Report 3, 1980. URL <https://www.jstor.org/stable/pdf/1805222.pdf?refreqid=excelsior%7B3Aad7f36a181a724d6a87257a36b2a9d20>.
- U. Doraszelski and J. Jaumandreu. R&D and Productivity: Estimating Endogenous Productivity. *The Review of Economic Studies*, 80(4):1338–1383, oct 2013. ISSN 0034-6527. doi: 10.1093/restud/rdt011. URL <https://academic.oup.com/restud/article-lookup/doi/10.1093/restud/rdt011>.
- Jean-Pierre Dubé, Jeremy T. Fox, and Che-Lin Su. Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation. *Econometrica*, 80(5):2231–2267, 2012. doi: 10.3982/ECTA8585. URL <http://www.econometricsociety.org>.
- Amit Gandhi and Jean-François Houde. Measuring Substitution Patterns in Differentiated Products Industries. 2015. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.718.5758{%7C}rep=rep1{%7C}type=pdf>.

- Amit Gandhi, Salvador Navarro, David Rivers, Dan Akerberg, Richard Blundell, Juan Esteban Carranza, Allan Collard-Wexler, Ulrich Do-Raszelski, Steven Durlauf, Jeremy Fox, Silvia Goncalves, Phil Haile, Jinyong Hahn, Joel Horowitz, Jean-Francois Houde, Lance Lochner, Aureo De Paula, Amil Petrin, Mark Roberts, Nicolas Roys, Todd Stinebrickner, Chad Syverson, Chris Taber, and Quang Vuong. On the Identification of Gross Output Production Functions. Technical report, 2017. URL http://publish.uwo.ca/~drivers2/research/GNR_{_}5{_{_}17{_{_}17.pdf.
- Nicola Gennaioli, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. Human Capital and Regional Development. *The Quarterly Journal of Economics*, 128(1):105–164, feb 2013. doi: 10.1093/qje/qjs050. URL <https://academic.oup.com/qje/article-lookup/doi/10.1093/qje/qjs050>.
- Matthew Gentzow. Valuing new goods in a model with complementarity : Online newspapers. Technical Report 3, 2004. URL <https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.97.3.713>.
- W M Gorman. A Possible Procedure for Analysing Quality Differentials in the Egg Market. Technical Report 5, 1980. URL <https://www.jstor.org/stable/pdf/2296916.pdf>.
- Zvi Griliches and Jacques Mairesse. Production Functions: The Search for Identification. In Steinar Strom, editor, *Econometrics and economic theory in the twentieth century: The Ragnar Frisch Centennial Symposium*. Cambridge University Press, Cambridge, MA, 1998. doi: 10.3386/w5067. URL <http://www.nber.org/papers/w5067.pdf>.
- Jessie Handbury and David E Weinstein. Goods Prices and Availability in Cities. *The Review of Economic Studies*, 82(1):258–296, 2014. ISSN 0034-6527. doi: 10.1093/restud/rdu033. URL <https://academic.oup.com/restud/article-abstract/82/1/258/1547368>.
- Jonathan E Haskel, Sonia C Pereira, and Matthew J Slaughter. Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms? *Review of Economics and Statistics*, 89(3):482–496, aug 2007. ISSN 0034-6535. doi: 10.1162/rest.89.3.482. URL <http://www.mitpressjournals.org/doi/10.1162/rest.89.3.482>.
- Hausman, Leonard, and Zona. Competitive Analysis with Differentiated Products. *Annales d'Économie et de Statistique*, (34):159–180, 1994. ISSN 0769489X. doi: 10.2307/20075951. URL <https://www.jstor.org/stable/pdf/20075951.pdf?refreqid=excelsior%3A753ebfc8c3ff6bb8b67698d0971f0922>.
- B Y T Homas J H Holmes. The Diffusion of Wal-Mart and Economies of Density. *Econometrica*, 79(1): 253–302, 2011. ISSN 0012-9682. doi: 10.3982/ecta7699.
- H S Houthakker. Revealed Preference and the Utility Function Author. *Economica*, 17(66):159–174, 1950. URL <https://www.jstor.org/stable/pdf/2549382.pdf>.
- Cheng Hsiao. Chapter 4 Identification. *Handbook of Econometrics*, 1:223–283, jan 1983. ISSN 1573-4412. doi: 10.1016/S1573-4412(83)01008-9. URL <https://www.sciencedirect.com/science/article/pii/S1573441283010089>.
- Chang-Tai Hsieh and Peter J. Klenow. Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4):1403–1448, nov 2009. doi: 10.1162/qjec.2009.124.4.1403. URL <https://academic.oup.com/qje/article-lookup/doi/10.1162/qjec.2009.124.4.1403>.
- Hidehiko Ichimura and Petra E. Todd. Chapter 74 Implementing Nonparametric and Semiparametric Estimators. *Handbook of Econometrics*, 6:5369–5468, jan 2007. ISSN 1573-4412. doi: 10.1016/S1573-4412(07)06074-6. URL <https://www.sciencedirect.com/science/article/pii/S1573441207060746>.
- Mitsuru Igami. Estimating the Innovator’s Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry, 1981–1998. *Journal of Political Economy*, 125(3):798–847, jun 2017. ISSN 0022-3808. doi: 10.1086/691524. URL <https://www.journals.uchicago.edu/doi/10.1086/691524>.
- Dale W. Jorgenson. Chapter 31 Econometric methods for modeling producer behavior. In *Handbook of Econometrics*, volume 3, pages 1841–1915. Elsevier, jan 1986. ISBN 9780444861870. doi: 10.1016/S1573-4412(86)03011-8. URL <https://www.sciencedirect.com/science/article/pii/S1573441286030118>.

- Kenneth L. Judd. *Numerical methods in economics*. MIT Press, 1998. ISBN 0262100711.
- Kelvin J. Lancaster. A New Approach to Consumer Theory. *Journal of Political Economy*, 74(2):132–157, apr 1966. ISSN 0022-3808. doi: 10.1086/259131. URL <https://www.journals.uchicago.edu/doi/10.1086/259131>.
- James Levinsohn and Amil Petrin. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2):317–341, apr 2003. ISSN 0034-6527. doi: 10.1111/1467-937X.00246. URL <https://academic.oup.com/restud/article-lookup/doi/10.1111/1467-937X.00246>.
- Charles F. Manski. Chapter 43 Analog estimation of econometric models. *Handbook of Econometrics*, 4:2559–2582, jan 1994. ISSN 1573-4412. doi: 10.1016/S1573-4412(05)80012-1. URL <https://www.sciencedirect.com/science/article/pii/S1573441205800121>.
- Jacob Marschak and William H Andrews. Random Simultaneous Equations and the Theory of Production Author. Technical Report 4, 1944. URL <https://www.jstor.org/stable/pdf/1905432.pdf?refreqid=excelsior{%}3A226527735e61586bdd17a4d63a8e7ade>.
- Eric Maskin and Jean Tirole. A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs. Technical Report 3, 1988. URL <https://www.jstor.org/stable/1911700?origin=crossref>.
- Rosa L. Matzkin. Chapter 42 Restrictions of economic theory in nonparametric methods. *Handbook of Econometrics*, 4:2523–2558, jan 1994. ISSN 1573-4412. doi: 10.1016/S1573-4412(05)80011-X. URL <https://www.sciencedirect.com/science/article/pii/S157344120580011X>.
- Rosa L. Matzkin. Chapter 73 Nonparametric identification. *Handbook of Econometrics*, 6:5307–5368, jan 2007. ISSN 1573-4412. doi: 10.1016/S1573-4412(07)06073-4. URL <https://www.sciencedirect.com/science/article/pii/S1573441207060734>.
- Marjorie B. McElroy. Additive General Error Models for Production, Cost, and Derived Demand or Share Systems. *Journal of Political Economy*, 95(4):737–757, aug 1987. ISSN 0022-3808. doi: 10.1086/261483. URL <https://www.journals.uchicago.edu/doi/10.1086/261483>.
- Daniel McFadden and Kenneth Train. Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5):447–470, 2000. ISSN 08837252. doi: 10.1002/1099-1255(200009/10)15:5<447::aid-jae570>3.3.co;2-t. URL <http://pages.stern.nyu.edu/~wgreene/DiscreteChoice/Readings/McFadden-Train.pdf>.
- Daniel L Mcfadden. Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka, editor, *Frontiers in Econometrics*. Academic Press, New York, 1974.
- John Muellbauer. Household Production Theory, Quality, and the "Hedonic Technique". *The American Economic Review*, 64(6):977–994, 1974. URL <https://www.jstor.org/stable/pdf/1815246.pdf?refreqid=excelsior{%}3A9afcc6ea46e4307e27b0ae4d468c4a11>.
- John Muellbauer. Community Preferences and the Representative Consumer Author. *Econometrica*, 44(5):979–999, 1976. URL <https://www.jstor.org/stable/pdf/1911540.pdf?refreqid=excelsior{%}3A831bc54594c508f1d145bf7d37ab90a9>.
- Richard F Muth. Household Production and Consumer Demand Functions. Technical Report 3, 1966. URL <https://www.jstor.org/stable/pdf/1909778.pdf?refreqid=excelsior{%}3A722a90f6b2005c7cc4e0e7b90f83f10b>.
- Aviv Nevo. Mergers with differentiated products: the case of the ready-to-eat cereal industry. Technical Report 3, 2000. URL <https://about.jstor.org/terms>.
- G Steven Olley and Ariel Pakes. The Dynamics of Productivity in the Telecommunications Equipment Industry. Technical Report 6, 1996.

- Richard Stone. *The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom 1920-1938, vol. 1*. Cambridge University Press, Cambridge, 1954. URL <https://biblioteca.ufm.edu/library/index.php?title=51780{&}query=@title=Special:GSMSearchPage@process=@field1=encabezamiento@value1=REINOUNIDO@mode=advanced{&}recnum=6>.
- Marcel K Richter. Revealed Preference Theory. *Econometrica*, 34(3), 1966. URL <https://about.jstor.org/terms>.
- Sherwin Rosen. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1):34–55, 1974. URL <http://www.journals.uchicago.edu/t-and-c>.
- Marc Rysman. Competition Between Networks: A Study of the Market for Yellow Pages. Technical report, 2004. URL <https://academic.oup.com/restud/article-abstract/71/2/483/1590045>.
- P A Samuelson. A Note on the Pure Theory of Consumer's Behaviour. *Economica*, 5(17):61–71, 1938. URL <https://about.jstor.org/terms>.
- Chad Syverson. What Determines Productivity? *Journal of Economic Literature*, 49(2):326–365, jun 2011. ISSN 0022-0515. doi: 10.1257/jel.49.2.326. URL <http://pubs.aeaweb.org/doi/10.1257/jel.49.2.326>.
- Kenneth Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, 2009. URL <https://eml.berkeley.edu/books/train1201.pdf>.
- Hal R Varian. The Nonparametric Approach to Demand Analysis. *Econometrica*, 50(4):945–973, 1982. URL <https://www.jstor.org/stable/pdf/1912771.pdf>.
- Jeffrey M Wooldridge. On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104:112–114, 2009. doi: 10.1016/j.econlet.2009.04.026. URL https://ac.els-cdn.com/S0165176509001487/1-s2.0-S0165176509001487-main.pdf?{_}tid=de86c166-4bc8-4227-96e2-26ca915d7f2f{&}acdnat=1548981733{_{}}720cc65c80b173f5ecdbd9db838e1d33.