Introduction to Industrial Organization

Demand Estimation

Jian-Da Zhu

National Taiwan University

October 24, 2019

Outline

- General Idea of demand estimation
- Logit model
- Nested Logit model
- Hedonic model

Introduction

- The recent empirical literature use the market data to estimate the demand.
- Why is it so important?
  - In the monopoly market, we know:

    $$\frac{P^* - MC(q(P^*))}{P^*} = -\frac{1}{\epsilon(P^*)},$$

    where ϵ is the elasticity of the demand.
  - MC can not be observed, but we can obtain $\epsilon(P^*)$ by estimation.
  - Therefore, MC could be recovered after estimation.
  - We can do the similar work in the oligopoly case.

Introduction

- After the demand estimation, we can further do the following analysis:
  - Estimating market power, markup, and marginal cost.
  - Simulate the merger
  - Welfare gains from new products
  - Policy evaluation
  - Estimating the supply side issue, including the dynamic game model.

General Idea of Demand Estimation

Demand Estimation

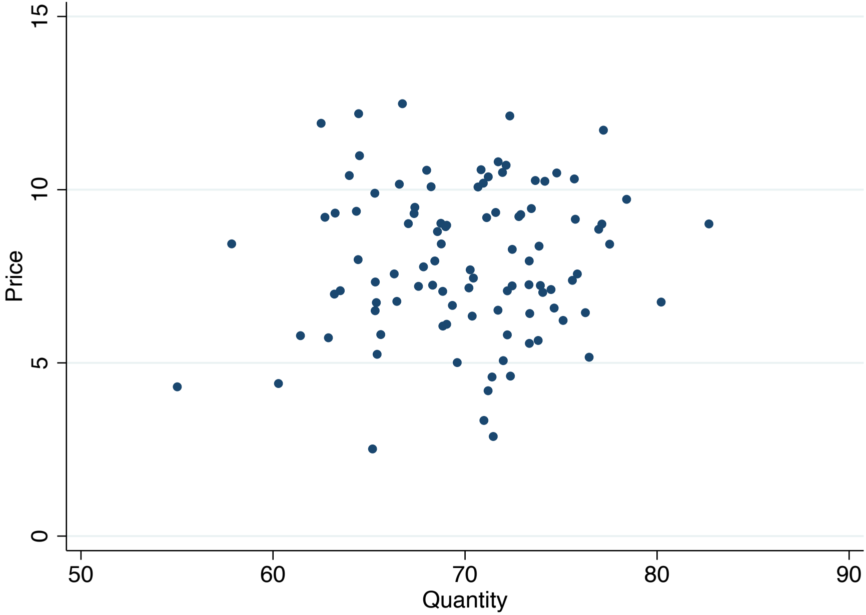
- Let's start from a homogeneous product. Assume that we can observe the time series data $\{P_t, Q_t\}_{t=1}^{T}$.
- It seems that we can estimate the following equation:

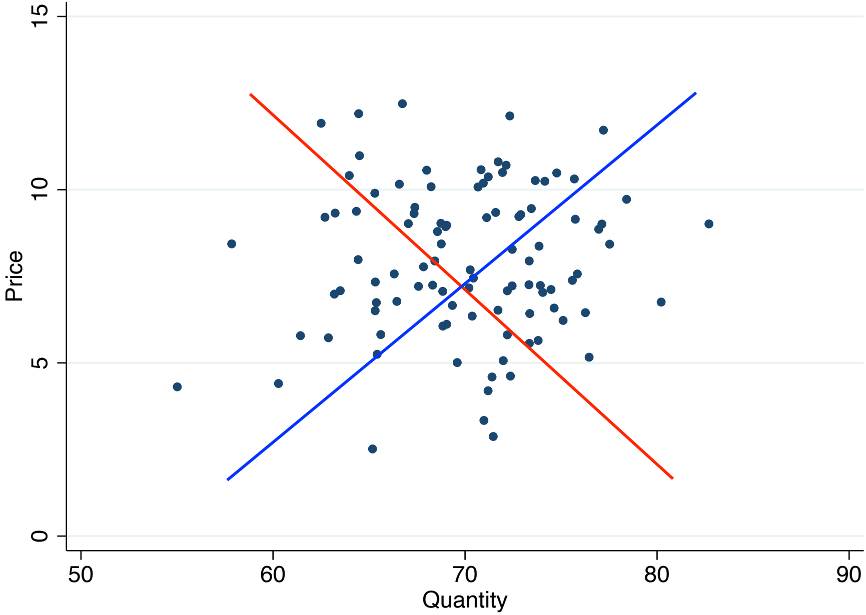
$$\log(Q_t) = \beta_0 + \beta_1 \log(P_t) + \epsilon_t.$$

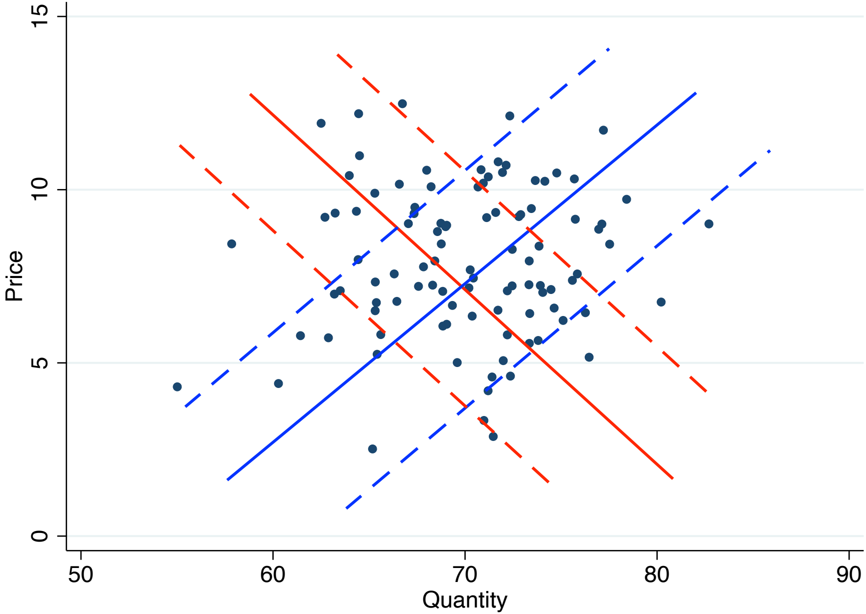
- The coefficient can be interpreted as the elasticity.

$$\beta_1 = \frac{\Delta \log(Q_t)}{\Delta \log(P_t)} = \frac{\frac{\Delta Q_t}{Q_t}}{\frac{\Delta P_t}{P_t}}$$

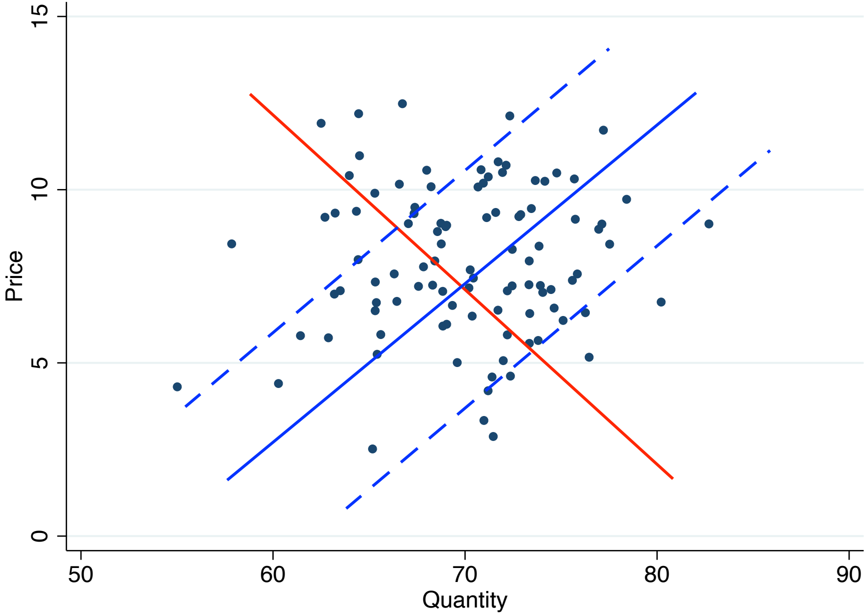
- However, is this the demand function or the supply function?


Price
15
10
5
0
50
60
70
80
90
Quantity


15
10
5
0
Price
50
60
70
80
90
Quantity



15
10
5
0
Price
50
60
70
80
90
Quantity


Price
Quantity
0
5
10
15
50
60
70
80
90

Demand Estimation

- First, we should control the demand:

$$\log(Q_t) = \beta_0 + \beta_1 \log(P_t) + \beta_2 X_t + \epsilon_t,$$

  where X_t should be some factors which can affect the demand, and ϵ_t is called as an unobserved demand shock.

- Second, we should rely on the instrumental variable Z_t, which can shift the supply curve.
- In Econometrics, the instrumental variable should
  - Cov$(Z_t, \epsilon_t) = 0$.
  - Cov$(Z_t, \log(P_t)) \neq 0$.
- For instance, some cost shifters to affect the production costs.

Two-Stage Least Squares (2SLS) Method

- In the first stage:

$$\log(P_t) = \gamma_0 + \gamma_1 Z_t + \gamma_2 X_t + v_t,$$

where Z_t is the instrumental variable, which can provide the exogenous price variation. We predict the predicted price based on $\widehat{\log(P_t)} = \hat{\gamma}_0 + \hat{\gamma}_1 Z_t + \hat{\gamma}_2 X_t$.

- In the second stage, run the regression as

$$\log(Q_t) = \beta_0 + \beta_1 \widehat{\log(P_t)} + \beta_2 X_t + \epsilon_t.$$

- Then $\hat{\beta}_1$ is called a two-stage least squares estimator, which is consistent and unbiased.

Example: Hard Disk Drive Industry

- This example is from Igami and Uetake (2019).
- They consider units of data storage (measured in bytes) as undifferentiated products.
- They specify a log-linear demand for raw data-storage functionality of HDDs,

$$\log(Q_t) = \alpha_0 + \alpha_1 \log(P_t) + \alpha_2 \log(X_t) + \epsilon_t,$$

  where

  - Q_t: the worlds total HDD shipments in exabytes (EB = 1 billion GB)
  - P_t: the average HDD price per gigabytes (\$/GB).
  - X_t: the PC shipments (in million units) as a demand shifter.

Example: Hard Disk Drive Industry

- Two instruments in this example:
  - the average disk price per gigabyte (\$/GB): Disks are one of the main components of HDDs, and hence their price is an important cost shifter for HDDs.
  - a dummy variable indicating a major supply disruption caused by flood in Thailand in the fourth quarter of 2011.

Example: Hard Disk Drive Industry

TABLE 3
Demand estimates

	(1) OLS	(2) OLS	(3) IV	(4) IV
Log HDD price per GB (α_1)	−1.112	−1.046	−1.054	−1.043
	(0.035)	(0.046)	(0.032)	(0.038)
Log PC shipment (α_2)	–	0.271	–	0.276
	(–)	(0.095)	(–)	(0.086)
Number of observations	83	83	83	83
Adjusted R^2	0.942	0.948	–	–
First-stage regression				
Log disk price per GB	–	–	0.813	0.567
	(–)	(–)	(0.026)	(0.032)
Thai flood dummy	–	–	0.263	0.548
	(–)	(–)	(0.079)	(0.070)
F-value	–	–	585.49	732.12
Adjusted R^2	–	–	0.874	0.946

Notes: Dependent variable is log total HDD (in EB) shipped. We use detrended quantities and prices of HDD to address nonstationarity in the original time series of these variables. Huber–White heteroskedasticity-robust standard errors are in parentheses.

Logit Model

Logit Model

- There are J products in the market. the utility of consumer i purchasing product j:

$$u_{ij} = \delta_j + \epsilon_{ij},$$

  where δ_j is the mean utility for product j, and ϵ_{ij} is the idiosyncratic term.

- We assume that ϵ_{ij} are i.i.d. distributed Type I extreme value across consumers:

$$F(\epsilon) = e^{-e^{-\epsilon}}.$$

- Consumer i decides to buy product j if $u_{ij} > u_{ik},\ \forall k \neq j$

Logit Model

- The probability of consumer i purchasing product j:

$$
\begin{aligned}
& \text{Prob}(u_{ij} > u_{ik},\ \forall k \neq j) \\
= & \text{Prob}(\delta_j + \epsilon_{ij} > \delta_k + \epsilon_{ik},\ \forall k \neq j) \\
= & \text{Prob}(\epsilon_{ij} > \epsilon_{ik} + (\delta_k - \delta_j),\ \forall k \neq j) \\
= & \ldots \\
= & \frac{\exp(\delta_j)}{\sum_{k=0}^{J} \exp(\delta_k)}
\end{aligned}
$$

- We usually normalize the mean utility of outside good ($j = 0$) as zero, so the probability of consumer i purchasing product j:

$$\frac{\exp(\delta_j)}{1 + \sum_{k=1}^{J} \exp(\delta_k)} \equiv s_j,$$

as the market share of product j. Details

Logit Model

- Further assume that

$$\delta_j = X_j\beta - \alpha p_j + \xi_j,$$

  where X_j are characteristics of product j, p_j is the price of product j, and ξ_j is the unobserved effect for product j.

- The share of outside good:

$$s_0 = \frac{1}{1 + \sum_{k=1}^{J} \exp(\delta_k)}.$$

- Therefore, we can have a linear equation:

$$\ln(s_j) - \ln(s_0) = X_j\beta - \alpha p_j + \xi_j.$$

Logit Model

- Because price is endogeneous in this linear equation, we need to use instrumental variables to run 2SLS (Two stage least squares) to obtain the unbiased coefficients.
- We usually use some cost shifters as IVs.
- Note that

$$\frac{\partial s_j}{\partial p_j} = -\alpha s_j(1 - s_j);$$
$$\frac{\partial s_j}{\partial p_k} = \alpha s_j s_k, \ \forall k \neq j.$$

- The cross-price elasticities depends only on market shares and prices:

$$\eta_{jk} = \frac{\partial s_j}{\partial p_k} \frac{p_k}{s_j} = \alpha p_k s_k.$$

Nested Logit Model

Nested Logit Model

- To further improve the logit model, we can use the nested logit model to estimate the demand.
- In the nested logit model, we modify the random term to create the substitution pattern among those products which are similar.
- The choice structure:
  - Choice A:
    - Choice A1
    - Choice A2
  - Choice B:
    - Choice B1
    - Choice B2
    - Choice B3

Example: Aircraft Market

- Irwin and Pavcnik (2004) study the competition between Airbus and Boeing.
- The choice structure for consumers in the market:
  - Narrow-body aircraft: Boeing 737, Boeing 757, Airbus A320.
  - Wide-body aircraft:
    - medium-range: Boeing 767, Airbus A300, Airbus A310.
    - long-range: Boeing 747, Boeing 777, Airbus A330, Airbus A340.

Nested Logit Model

- The market is segmented into several groups $g = 0, 1, 2, ...G$.
- The utility of consumer i purchasing product j:

$$u_{ij} = \delta_j + \sigma\zeta_{ig} + (1-\sigma)\epsilon_{ij},$$

where

  - ϵ_{ij} are i.i.d distributed Type I extreme value
  - ζ_{ig} are common to all products in group g
  - $\sigma \in [0, 1)$ measures the magnitude of market segmentation. If $\sigma = 0$, we go back to the Logit model.

- Define the market share of product j in group g as $s_{j|g}$.
- We do the similar work as that in Logit model, and we can get a nice linear equation

$$\ln(s_j) - \ln(s_0) = X_j\beta - \alpha p_j + \sigma \ln(s_{j|g}) + \xi_j.$$

More Extensions: BLP Model

- The other way to capture the flexible substitution pattern is based on Berry, Levinsohn, and Pakes (1995), which is called BLP model.
- They consider the following utility function:

$$u_{ij} = X_j \beta_i - \alpha_i p_j + \xi_j + \epsilon_{ij},$$

where β_i, and α_i are random coefficients to capture the heterogeneous tastes across individuals.

- Some good reference:
  - Nevo (2000), "A Practitioners Guide to Estimation of Random-Coefficients Logit Models of Demand"
  - Nevo (2001), "Measuring Market Power in the Read-to-Eat Cereal Industry"

Hedonic Model

Hedonic Model

- A regression of prices on product characteristics is called a "hedonic regression".
- For instance, in the housing market, we can run the regression of housing prices on housing characteristics:

$$P_{it} = X_{it}\beta + \epsilon_{it},$$

  where X_{it} includes all the housing characteristics, such as with a parking lot or not, housing age, number of bedrooms, ...
- The regression equation treats these characteristics separately, and it estimates the prices based on those characteristics.
- However, prices are equilibrium prices, so the hedonic coefficients combine the effects of demand and supply side.

Example: Housing Market in Taipei

- Data: housing transaction data from January 1, 2016 to July 31, 2016 in Taipei city.
- To study housing prices in Taipei, we can consider the following model:

$$\begin{aligned} \mathit{price}_i = \beta_0 + \beta_1 \mathit{age}_i + \beta_2 \mathit{age}_i^2 + \beta_3 \mathit{elevator}_i + \beta_4 \mathit{floor}_i \\ + \beta_5 \mathit{bedroom}_i + \beta_6 \mathit{livingroom}_i + \beta_7 \mathit{bathroom}_i + u_i, \end{aligned}$$

  where

  - *price* is the transaction price per unit in 10 thousand dollars.
  - *age* housing age in years.
  - *elevator* is a dummy for elevators
  - *floor* is the transaction floor
  - *bedroom*, *livingroom*, and *bathroom* are the number of bedrooms, living rooms, and bathrooms.

Estimation Results

VARIABLES	(1) *price*	(2) *price*	(3) *price*	(4) *price*	(5) *price*	(6) log(*price*)
age	-0.356***	-0.178***	-0.147***	-0.140***	-0.964***	-0.00216***
	[0.0171]	[0.0221]	[0.0225]	[0.0223]	[0.0632]	[0.000367]
age^2					0.0219***	
					[0.00157]	
elevator		9.344***	8.305***	7.581***	11.20***	0.146***
		[0.756]	[0.769]	[0.766]	[0.794]	[0.0126]
floor			0.427***	0.474***	0.384***	0.00821***
			[0.0638]	[0.0635]	[0.0626]	[0.00104]
bedroom				-1.996***	-1.859***	-0.0325***
				[0.347]	[0.340]	[0.00570]
livingroom				-1.918***	-1.717***	-0.0232***
				[0.500]	[0.490]	[0.00821]
bathroom				1.697***	2.008***	0.0259***
				[0.469]	[0.460]	[0.00770]
Constant	66.28***	55.90***	53.54***	59.45***	58.97***	4.008***
	[0.416]	[0.934]	[0.995]	[1.255]	[1.230]	[0.0206]
Observations	4,616	4,616	4,616	4,616	4,616	4,616
R-squared	0.086	0.115	0.124	0.140	0.175	0.154

Standard errors in brackets, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Other Applications

- Besides the housing characteristics, we can also use the hedonic model to quantify
  - quality of environment: air pollution, water pollution, noise
  - amenities: parks, schools, metro
  - risk: earthquake risk map
- You also can combine the difference-in-differences method or regression discontinuity method with the hedonic model.
  - School district effect (regression discontinuity)
  - Earthquake risk (difference-in-differences)

Price Index

- We can use hedonic model to calculate the price index.
- First, in period t, the following equation is estimated:

$$P_{it} = X_{it}\beta_t + \epsilon_{it},$$

- In period $t + 1$, we use last year's estimated coefficients to predict the prices:

$$\hat{P}_{it+1} = X_{it+1}\hat{\beta}_t.$$

- Calculate the price change percentage.
- Alternatively, put year dummies in regression and use the dummy coefficients to observe the price change.

Homework 5

- Pick up a market, and find a dataset of product information, including prices and some important characteristics.
- Run the hedonic model and interpret the coefficients. Does the result make sense to you?
- Reference for the datasets:
  - `https://www.kaggle.com/datasets`
  - If you can read Mandarin, you can check the real estate dataset in Taiwan: `https://plvr.land.moi.gov.tw/DownloadOpenData`

Appendix: Normalization I

- Here is an example to illustrate why we need to normalize one product as the outside good ($j = 0$).
- Assume that the mean utility:

$$\delta_j = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2},$$

where x_{j1}, and x_{j2} are two characteristics for product j.

- The market share for product j:

$$s_j = \frac{\exp(\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2})}{\sum_{k=0}^{J} \exp(\beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2})}$$

- Consider another set of parameters: $\{\tilde{\beta}_0, \beta_1, \beta_2\}$, where $\tilde{\beta}_0 = \beta_0 + c$, and c is a constant.

Appendix: Normalization II

- The market share for product j, based on the other set of parameters $\{\tilde{\beta}_0, \beta_1, \beta_2\}$:

$$
\begin{aligned}
\tilde{s}_j &= \frac{\exp(\tilde{\beta}_0 + \beta_1 x_{j1} + \beta_2 x_{j2})}{\sum_{k=0}^{J} \exp(\tilde{\beta}_0 + \beta_1 x_{k1} + \beta_2 x_{k2})} \\
&= \frac{\exp(c + \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2})}{\sum_{k=0}^{J} \exp(c + \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2})} \\
&= \frac{\exp(c) \exp(\beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2})}{\exp(c) \sum_{k=0}^{J} \exp(\beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2})} = s_j
\end{aligned}
$$

- Two sets of parameters can predict the same market share for product j, so the coefficient are not identified in this model. Back