Business Analytics - 2023/24

Microeconomic Foundations for Pricing Management and Discrete Choice Models

Prof. Paolo Brandimarte
Dip. di Scienze Matematiche – Politecnico di Torino
e-mail: paolo.brandimarte@polito.it
URL: staff.polito.it/paolo.brandimarte

This version: August 22, 2024

- M. Bierlaire, V. Lurkin. *Introduction to Disaggregate Demand Models*. INFORMS TutORials in Operations Research. https://doi.org/10.1287/educ.2017.0169

- H.A. Kuyumcu, I. Popescu. *Deterministic price–inventory management for substitutable products.* Journal of Revenue and Pricing Management, vol. 4, no. 4, 2006, pp. 354–366.

- R. Phillips. Why are prices set the way they are? In: Ö. Özer and R. Phillips (eds). *The Oxford Handbook of Pricing Management*. Oxford University Press, 2014.

- R. Phillips. *Pricing and Revenue Optimization* (2nd ed.). Stanford University Press, 2021.

- O. Shy. *How to Price: A Guide to Pricing Techniques and Yield Management*. Cambridge University Press, 2008.

- H. Simon. *Confessions of the Pricing Man: How Price Affects Everything*. Springer, 2015.

- H. Simon, M. Fassnacht. *Price Management: Strategy, Analysis, Decision, Implementation*. Springer, 2019.

- G. van Ryzin. *Future of Demand Management: Models of Demand*. Journal of Revenue and Pricing Management, vol. 4, no. 2, 2005, pp. 204–210/

Profit is driven by four basic factors, and a simple model is

$$\text{Profit} = (\text{Price} - \text{VariableCost}) \times \text{Volume} - \text{FixedCost}.$$

Consider the base case:

- Price: €100 per item.
- Volume: 1 million items.
- Variable cost: €60 per item.
- Fixed cost: €30 million

Then, profit is

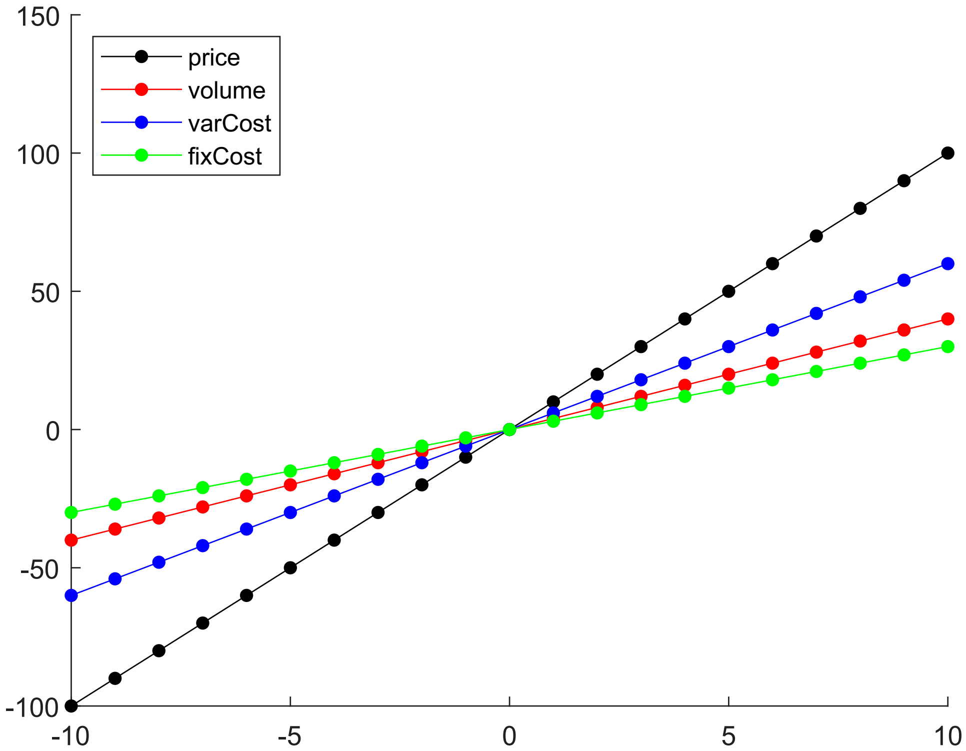
$$(100 - 60) \times 10^6 - 30 \times 10^6 = \text{€}10 \text{ million}$$

Now, what would be impact of a change in each factor, assuming that the other ones do not change?

Let us consider percentage improvements in the range from -10% to 10% in the factors (an improvement in costs is actually a percentage reduction), and their prcentage impact on profit.

A 10% increase in price results in doubled profit. On the contrary a 10% reduction in variable cost implies a profit improvement of 50%.

The other two factors have less impact. The picture is symmetric with respect to variations



It is worth noting that changing price is essentially costless and instantaneous, unlike investments in advertisements or cost reductions.

However, the picture is not realistic, since price does affect volume (and cost need not be an affine function of volume). We need to consider interactions between all factors. Nevertheless, let us use this crude model to check our intuition.

For instance. Assume that price is cut by 20%, with respect to the above base case. What is the required increase in volume, in order to keep profit at the same level?

To find the new volume V_{new} we should solve

$$10 = (80 - 60)V_{\text{new}} - 30 \quad \Rightarrow \quad V_{\text{new}} = 2 \text{ million items},$$

i.e., a 100% increase in volume is required to compensate a reduction of price by 20%.

Going the other way around, we may ask which reduction in volume we may sustain after a 20% increase in price:

$$10 = (120 - 60)V_{\text{new}} - 30 \quad \Rightarrow \quad V_{\text{new}} = \frac{2}{3} \times 1 \text{ million items},$$

i.e., we might afford losing one third of volume.

Clearly, we need a model of the relationship between price and volume (demand). We also need to be aware of the possible causal chains:

- Price → Volume → Revenue → Profit
- Price → Volume → Cost → Profit
- Price → Competitors' prices → Market share → Volume → Revenue → Profit
- Price of manufacturer → Price of retailer → Volume → Revenue → Profit

It is commonly stated that a firm's objective is profit maximization. Is it true that profit maximization boils down to cost minimization? The previous rough-cut analysis shows that this need not be true.

If profit is revenue minus cost, we have another possibility. Revenue and yield management strategies have become common since their introduction in the airline industry.

What are the possible pricing management strategies?

- The seemingly obvious pricing strategy is cost-based: Assess the cost of an item, and apply a markup. This simple approach disregards the reaction of both consumers and competitors.

- Another idea, quite common in economics, is that price arises from the equilibrium of supply and demand. This leads to the idea of demand-based pricing, which may be tackled by simple models in the case of a monopoly.

- A further ingredient is the role of competition. In the classical microeconomics literature, this may be tackled by game theory models. Game theory models may also be used to model the interaction between a firm and customer, rather than among firms.

The form of competition and interaction are actually many and diverse:

- competition among firms (oligopoly vs. monopoly);
- interactions between actors along the supply chain (e.g., manufacturers and retailers);
- strategic customer behavior (do not sell unused capacity last minute).

As an example, which kind of competition must an airline face?

- Direct competition among alternative airlines.
- Competition among different transportation modes: aircraft vs. train.
- Indirect competition: traveling by aircraft vs. meeting online.

We may need a shift of paradigm:

From *Design–Build–Price* to *Price–Design–Build*

A sensible strategy of price positioning must be selected:

luxury, premium, medium, low, ultra-low

Since pricing psychology is relevant, the proper strategy depends on the key type of product (or service) attributes:

functional, emotional, symbolic, ethical.

We must choose the proper model to tackle different settings:

- repeated vs. single purchase;
- diapers vs. chocolate;
- business-to-business (B2B) vs. business-to-consumer (B2C).

Alternative mechanisms

There are a variety of ways to set and communicate prices, including direct negotiations and auctions.

Setting a price is not the only way to improve profit/revenue.

- bundling and tying, i.e., quantity discounts, packages, and ancillaries; sometimes, unbundling is pursued (e.g., Ryanair etc.);
- using coupons;
- customized subscriptions;
- managing the availability of different classes (quantity-based revenue management is common in the travel industry);
- overbooking to protect against no-shows and manage non-storable capacity;
- risk sharing and contract/incentives design.

Part 1

Pricing and demand functions

An introductory model: Linear demand–price relationship

Demand is a random variable, influenced by a multitude of factors, including (but not limited to) price.

As a starting point let us introduce a simple model based on a **linear demand function** linking price and demand:

$$d(p) = \alpha - \beta p. \tag{1}$$

A seemingly obvious assumption is $\beta > 0$, but counterexamples are provided by luxury goods and by cases in which price works as a signal of quality.

The intercept α is demand when $p = 0$, a price point where the model makes little sense. In other models, demand goes to infinity when $p \to 0$, but if the consumer population is finite, we may interpret α as the population size.

Since, demand cannot be negative, the model makes sense in the price range

$$p \in [0, p_{\mathrm{lim}}],$$

where we define the **limit price**

$$p_{\mathrm{lim}} \doteq \frac{\alpha}{\beta}$$

as the largest price for which we have a positive demand.

We may write the model in a possibly better way as

$$d(p) = \big(\alpha - \beta p\big)^+,$$

where $(x)^+ \equiv \max\{0, x\}$.

In microeconomics, it is also common to invert the demand function to yield the market price when a quantity q is produced and offered on the market. By inverting Eq. (1), we find the **inverse demand function**

$$p(q) = a - bq,$$

where $a = \alpha/\beta$ and $b = 1/\beta$.

This comes in handy when studying quantity-based competition in naive microeconomics, but it is much less natural when we deal with proper statistical modeling for pricing management.

Nevertheless, there are markets in which price arises from complicated mechanisms (e.g., auctions) that are affected by total availability.

Revenue and profit maximization

A natural objective that we could consider is profit maximization. Profit in its simplest form is revenue minus cost, which may be expressed as a function of price or as a function of quantity.

The latter choice looks more natural to express cost:

$$\pi(q) = p(q) \cdot q - c(q) = r(q) - c(q),$$

where the cost function $c(q)$ is a generic function of quantity, which may account for fixed charges, as well as economies or diseconomies of scale, and we introduce the revenue function $r(q) \doteq p(q) \cdot q$.

Assuming that all of the functions are differentiable and that the overall profit function is concave, we apply the first-order optimality condition:

$$\pi'(q^*) = r'(q^*) - c'(q^*) = 0 \qquad \Rightarrow \qquad r'(q^*) = c'(q^*),$$

i.e., the optimal quantity is such that marginal revenue and marginal cost are the same.

If marginal revenue is larger than marginal cost, the revenue benefit from increased production outweighs the increase in cost, so that we should increase q. If marginal cost is larger than marginal revenue, the cost reduction benefit from reduced production outweighs the reduction in revenue, so that we should decrease q.

In the case of linear inverse demand function and constant variable cost, we have a concave quadratic function,

$$\pi(q) = (a - bq) \cdot q - cq = (a - c) \cdot q - bq^2,$$

and the first-order optimality condition,

$$\pi'(q) = (a - c) - 2bq, \qquad \text{yields} \qquad q^* = \frac{a - c}{2b}.$$

Note that profit maximization is *not* equivalent to cost minimization.

Cost minimization can make sense when a given demand must be satisfied anyway and we have to choose among different production technologies or different production plans over time.

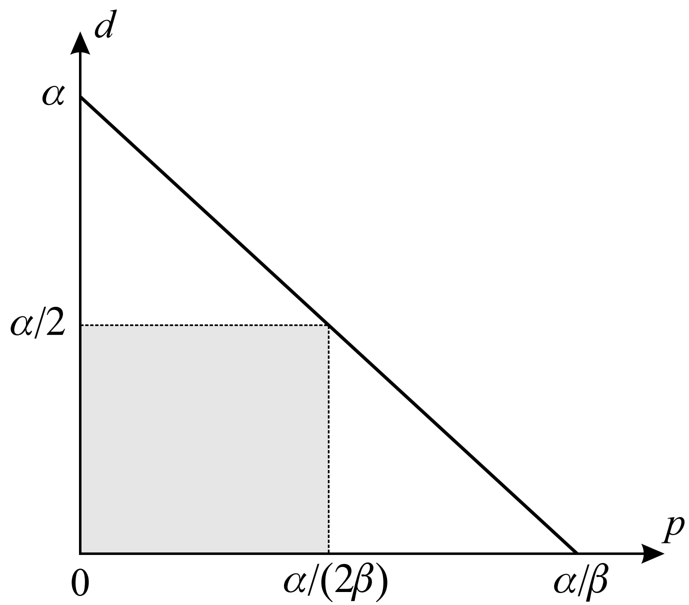
Sometimes, the marginal cost of one more unit is negligible (airline industry, web services) or the production cost is a sunk cost (markdown management). Then, revenue maximization may make sense, and it may be naturally expressed as a function of price:

$$\max r(p) = p \cdot d(p) = p \cdot \big(\alpha - \beta p\big) = \alpha p - \beta p^2.$$

The optimal solution is

$$p^* = \frac{\alpha}{2\beta} \qquad \Rightarrow \qquad d^* = \frac{\alpha}{2}, \quad r^* = \frac{\alpha^2}{4\beta}.$$

The solution may be interpreted geometrically, since the optimal price is the midpoint of the price range and maximizes the area of the rectangle depicted below:



Estimating the model

In naive microeconomics, deterministic models are used, but even a trivial linear demand model like

$$d(p) = \alpha - \beta p$$

involves unknown parameters that must be estimated.

Even if we assume that the linear model is just fine, there is a thorny issue: the *cost* of the estimation procedure. We should learn online, rather than offline, and the proper planning of experiments is crucial.

Which prices should we use to effectively learn the parameters? We should focus on the slope β, as the intercept α is just a mathematical construct. We know that the standard estimation error of slope is

$$\mathsf{SE}_\beta = \frac{\sigma_\epsilon}{\sqrt{\sum_{k=1}^{n} (p_k - \overline{p})^2}},$$

where σ_ϵ is the (unknown) standard deviation of errors.

This tells that the more spread out the experimented prices are, the better our learning. Unfortunately, small and large prices may harm revenue. So there is tradeoff between what we learn and the cost of learning it.

Furthermore, there are additional issues in practice:

- Commercial prices. Not every price can be applied.
- Large price swings emphasize nonlinearity.
- The role of time. In fashion items, time is essential. We have little time to learn.
- Confounding issues. Imagine that we apply a price on Wednesday and a reduce price on Saturday. Can we attribute the change in demand to the price reduction? Experiment planning in marketing may be difficult.
- Operational and execution issues: What about writing labels at a physical store?
- Excessive experimentation may lead to consumer dissatisfaction, as well as inducing strategic behavior.

The limits of linear demand models

- The assumption of constant slope (price sensitivity) is not quite sensible.
- It relies on an unrealistic assumption about willingness-to-pay.
- We assume that slope is negative. However, in the case of luxury goods and when price is a quality signal, we may find locally different behavior.
- It is a static model. Introducing time is essential in dynamic markdown strategies.
- We do not consider the role of capacity (see, e.g., peak-load pricing).
- We consider selling a single good or service (see tying and bundling, as well as assortment management).
- We do not consider consumer choice among alternative options, which may not only differ in price.
- We do not consider behavioral issues and information asymmetry. Psychology of pricing may be important.

The cost function

It is common to consider a fixed and a variable cost component and the simplest form of variable cost is the linear one:

$$c(q) = F + cq.$$

Note that a different cost function is

$$c(q) = \begin{cases} F + cq & \text{if } q > 0 \\ 0 & \text{if } q = 0 \end{cases}$$

which may be written as $c(q) = F \cdot \delta(q) + cq$, where $\delta(q) = 1$, if $q > 0$, 0 otherwise.

To avoid ambiguity, we talk of a fixed cost in the first case and a fixed charge in the second one. We should note that this may be a matter of time scale and hierarchical decision level (at the proper time scale, all costs are variable; see semivariable costs).

Given a cost function, we define the following concepts:

- Marginal cost $c'(q)$
- Average cost $c(q)/q$

Clearly, in order to define the marginal cost we have to assume differentiability. Apart from fixed charges, a discontinuity may be introduced by all-unit quantity discounts, and a kinky point by incremental quantity discounts.

When we must account for discrete production, so that q takes integer values, we may consider the difference $c(q+1) - c(q)$ as the marginal cost.

The marginal cost itself may be an increasing or decreasing function. When the marginal cost is decreasing, assuming differentiability again, it means that the second-order derivative $c''(q)$ is negative, so that the function is concave. This models an economy of scale, i.e., the more we produce, the larger the gain in efficiency.

On the contrary, a convex cost function, featuring an increasing marginal cost, models a diseconomy of scales.

Demand functions

There is a wide variety of demand models:

- We may model aggregate demand of the whole market, possibly a selected channel or part of the market, or demand of an individual.
- We may consider repeated purchases or not. This depends on the nature of the good/service (stock vs. flow goods), and the time horizon.
- We may consider uncertainty or not.

There are several possible input factors (time, past sales, price of competitors or alternative items). Moreover, the output value may be:

- A real variable (aggregate demand, possibly approximating an integer but large value)
- An integer variable (the amount purchased by an individual consumer)
- A 0/1 variable (the consumer buys or not)
- A categorical variable or a vector (which bundle or portfolio of items is purchased)

We have already considered, mainly for illustrative purposes, the linear demand function

$$d(p) = \left(\alpha - \beta p\right)^{+}.$$

Of course, one should question the validity of such a simple model:

- How can it be justified from an economic view point? We should investigate a microeconomic foundation for a demand model.

- Is it an empirically validated model, or is it contradicted by common sense and actual consumer behavior?

Assuming differentiability, an obvious feature of any demand function is its slope $d'(p)$, which by common sense is supposed to be negative, as it is in the linear demand model.

Indeed, it is usually negative, but there are exceptions:

- Signal of quality

- Luxury goods and conspicuous purchasing

Hence, there may be price ranges for which slope is positive, a fact that is confirmed by empirical investigations (see examples in Simon, 2015).

Demand functions: Point elasticity

Since the value of slope depends on the scale, i.e., the units we use to measure demand, a possibly better measure is price elasticity of demand:

$$\frac{\delta d/d}{\delta p/p} = \frac{\delta d}{\delta p} \cdot \frac{p}{d},$$

which, taking limits for $\delta p \to 0$ yields the **point price elasticity**:

$$\epsilon(p) = -\frac{d'(p)p}{d(p)}, \tag{2}$$

where the minus sign is convenient as slope is typically negative.

Demand functions may be classified as follows:

- Elastic, when $\epsilon(p) > 1$ (perfectly elastic if it is ∞)
- Unit elastic when $\epsilon(p) = 1$
- Inelastic when $\epsilon(p) < 1$ (perfectly inelastic if it is 0)

The nature of demand functions may depend on the specific point we consider.

Example: Linear demand function

Note that elasticity should not be confused with slope, as the following example shows.

The linear demand model features a constant slope, which is not realistic, but this does not imply a constant (point) elasticity. The elasticity is, for a linear demand function,

$$\epsilon(p) = \frac{\beta p}{\alpha - \beta p},$$

which grows from zero to $+\infty$ on the range $[0, p_{\mathrm{lim}}]$. Elasticity is 1 for

$$p = \frac{\alpha}{2\beta} = p^*,$$

i.e., the midpoint of the price range. It is inelastic for $p < p^*$, and elastic for $p > p^*$. In the limit, when $p \to p_{\mathrm{lim}} = \alpha/\beta$, it becomes perfectly elastic: For a little decrease in price, demand jumps from $d = 0$ to $d > 0$.

Elasticity is not constant for a linear demand function. Can we find a demand function such that elasticity is constant?

Example: Constant elasticity demand function

If we set

$$\frac{d'(p)p}{d(p)} = -\epsilon,$$

we obtain an ordinary differential equation,

$$d'(p) = -\frac{\epsilon \cdot d(p)}{p}.$$

which gives

$$d(p) = c \cdot p^{-\epsilon},$$

where $c = d(1)$.

Technical note. To solve the above differential equation, let us frame it in the following form, by separation of variables:

$$\frac{dy}{y} + \epsilon \frac{dx}{x} = 0,$$

where $y \equiv d(p)$ and $x \equiv p$. Straightforward integration yields

$$\begin{aligned} & \int^{y} \frac{dz}{z} + \epsilon \int^{x} \frac{dz}{z} = 0 \\ \Rightarrow \quad & \log y + \epsilon \log x = K, \qquad \text{where } K \text{ is an integration constant} \\ \Rightarrow \quad & \log \left(y \cdot x^{\epsilon} \right) = K \\ \Rightarrow \quad & y = c x^{-\epsilon}, \qquad \text{where we set } c = e^{K}. \end{aligned}$$

Thus, we find

$$d(p) = c \cdot p^{-\epsilon},$$

but we need an additional condition to find the constant c. If we set $p = 1$,

$$d(1) = c \cdot 1^{-\epsilon} = c.$$

It is also possible to derive a link between constant-elasticity and linear demand by a suitable data transformation, based on taking logs of data:

$$d = cp^{-\epsilon} \quad \Rightarrow \quad \log d = \log c - \epsilon \cdot \log p.$$

Hence, the usual tools of linear regression may be used to estimate the model. We also note that taking logs is a good way to simplify a model based on multiplicative, rather than additive effects.

To further understand the point, we observe that, in terms of increments

$$\frac{d \log x}{dx} = \frac{1}{x} \quad \Rightarrow \quad \delta \log x \approx \frac{\delta x}{x},$$

for $x > 0$. Therefore, we may rewrite elasticity as

$$\frac{\delta d/d}{\delta p/p} \approx \frac{d \log d(p)}{d \log p}.$$

The constant elasticity demand function results in a simple expression for revenue:

$$R(p) = p \cdot cp^{-\epsilon} = cp^{1-\epsilon}.$$

Thus, revenue is constant for unit elasticity, it is increasing for inelastic demand, and it is decreasing for elastic demand. Thus, a reduction in price will increase revenue only if demand is elastic.

Interpreting demand models: willingness-to-pay

How can we compare different demand models? Is the linear demand function a sensible model?

A linear model may be (at best) a local approximation, but we need to understand in more depth the economic rationale behind such a model.

This means that we should understand what really drives demand.

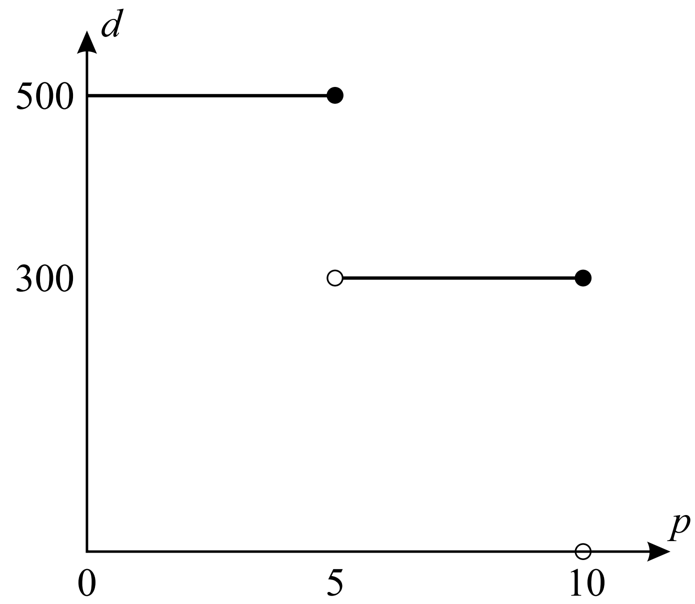
A basic mechanism behind demand models is the concept of **reserve price** or **willingness-to-pay**: Each consumer is characterized by the maximum price at which she is willing to buy a good.

As a simple example, consider the following table:

	Students	Nonstudents
Willingness-to-pay	€ 5	€ 10
Market size	200	300

This hypothetical example describes the willingness-to-pay for some kind of show, if the population can be segmented into two groups, students and nonstudents, featuring a different willingness-to-pay.

This would result in a piecewise constant demand function, as shown in the figure.



Now imagine a more differentiated population of N potential consumers, indexed by $k = 1, \ldots, N$, each one featuring a unique willingness-to-pay $p_k^\circ = p_{\mathsf{max}} \cdot \frac{k}{N}$. The index $k = 1$ gives a price, corresponding to the consumer who is less willing to pay, such that everyone will buy and demand is N. For $k = N$, only one customer is buying.

Demand is related to k by

$$d\big(p_k\big) = N - k + 1, \qquad k = 1, \ldots, N,$$

which results in a staircase decreasing function.

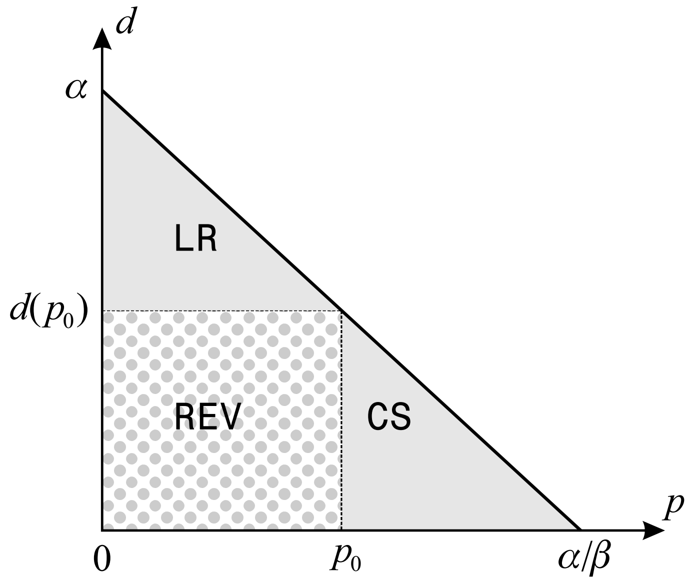
Intuition suggests that, if we take a continuous limit, i.e., we assume a uniformly distributed willingness-to-pay, we shall find a linear demand function. Alternative demand functions may be defined by assuming different distributions of the reserve price, and by investigating the sensitivity of demand with respect to price.

Let us consider a consumer who is willing to pay €100 for a good. If the price is €70, we say that the consumer surplus was the difference between her willingness-to-pay and the paid price, in this case, €30. When a single price is applied, some consumer surplus will be obtained by the set of consumers buying the good.

Geometrically, this can be visualized as the triangle CS depicted in the figure, with area

$$\begin{aligned}\mathsf{CS} &= \frac{1}{2}\cdot\left(\frac{\alpha}{\beta}-p_0\right)\cdot(\alpha-\beta p_0)\\ &= \frac{1}{2\beta}\cdot(\alpha-\beta p_0)^2,\end{aligned}$$

where p_0 is the chosen price.



Note that consumer surplus is zero when $p_0 = p_{\mathsf{lim}} = \alpha/\beta$, and it is $\alpha^2/(2\beta)$, the total triangle area, when $p_0 = 0$. In the figure, we also depict the collected revenue as the rectangular area REV.

In microeconomics, consumer surplus is a rough measure of consumer utility from a trade.

Looking the other way around, consumer surplus is, in a sense, lost revenue for the firm.

The same applies to the triangle LR, which is lost revenue from consumer not buying, because p_0 exceeds their willingness-to-pay.

Actually, we should also consider cost and the fact that, usually, a firm does not want to sell below cost.

But if we only consider revenue, we see that having to choose a single price leads to lost on both sides, consumer surplus (missed opportunity for a larger profit) and lost sales.

Ideally, the firm would like to apply a different price to each individual, corresponding to her willingness-to-pay (assuming marginal cost is zero).

Of course, setting an individual price is not practical, not to mention downright illegal. But how can a firm extract, at least partially, consumer surplus?

The key is price discrimination based on consumer segmentation, and it is best illustrated by a simple example (borrowed from Shy, 2008, p. 6).

Example of price discrimination

Consider again the willingness-to-pay in the previous table.

If no discrimination is possible, i.e., we have to apply the same price to the whole population, there are two possible choices:

- Option 1: sell at €10, with revenue

$$10 \times 300 = €3000.$$

- Option 2: sell at €5, with revenue

$$5 \times (200 + 300) = €2500.$$

Hence, it is better to ask for the larger price, preventing students to pay the ticket for the show. If we can discriminate and apply different prices, profit is increased:

$$5 \times 200 + 10 \times 300 = €4000.$$

Clearly, such a policy is feasible if we can identify students, e.g., by requiring a student badge, and prevent arbitrage (a student buying the ticket for €5 and selling it to a nonstudent for €10), which would have a **dilution** effect on revenue.

By applying price discrimination, the firm is able to extract consumer surplus. On the other hand, a subpopulation that would be excluded from the trade can participate.

In general, we may not find a win–win solution.

Furthermore, it may not be possible to discriminate in a clean way. There are different kinds of discrimination:

- Complete discrimination means that each consumer is charged a specific price.

- Direct segmentation means that identifiable segments are charged a different price.

- Indirect segmentation means that different variations of product/service are offered, and consumers make a choice. OK with non-identifiable characteristics. E.g.: Advance purchase and booking; discount coupons.

Formalizing willingness-to-pay and consumer surplus

After introducing these concepts informally, let us introduce a more formal framework.

Willingness-to-pay is the maximum price (reserve price) that a consumer is willing to pay for an item. The distribution of willingness-to-pay over the consumer population shapes the demand function.

Let us consider here a continuous model and define the function $w(x)$ as the distribution of willingness-to-pay.

This plays a similar role to the density of a random variable, in the sense that

$$\int_{p_1}^{p_2} w(x)\,dx$$

is the fraction of population with willingness to pay between p_1 and p_2. Hence,

$$d(p) = D_{\mathsf{max}} \int_{p}^{+\infty} w(x)\,dx,$$

where $D_{\mathsf{max}} = d(0)$ is the maximum demand achievable (assuming a finite population). Since

$$d'(p) = -D_{\mathsf{max}} \cdot w(p),$$

we may relate demand function and willingness-to-pay by

$$w(p) = -\frac{d'(p)}{D_{\mathsf{max}}}. \tag{3}$$

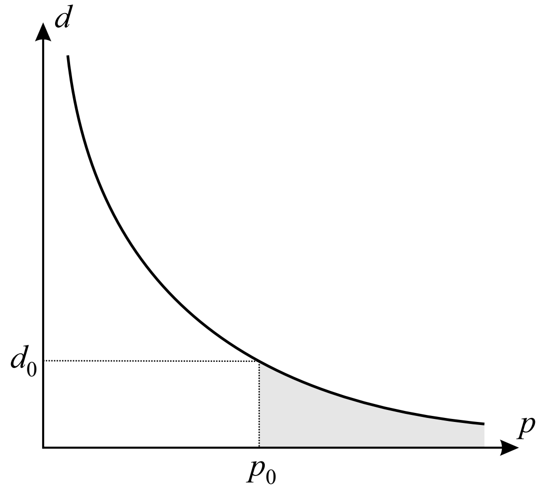
If a consumer willing to pay p_1 for a unit of a good, but she happens to pay $p_0 < p_1$, she is said to enjoy a consumer surplus $p_1 - p_0$. Given the function $w(x)$, we may integrate consumer surplus at price p_0 over the population buying, and define the overall (net) consumer surplus

$$S\big(p_0\big) = D_{\mathsf{max}} \int_{p_0}^{+\infty} w(x) \cdot (x - p_0)\, dx.$$

Using Eq. (3) we may write (using integration by parts):

$$\begin{aligned} S\big(p_0\big) &= -\int_{p_0}^{+\infty} d'(x) \cdot (x - p_0)\, dx = -\int_{p_0}^{+\infty} d'(x) x\, dx + p_0 \int_{p_0}^{+\infty} d'(x)\, dx \\ &= -\left[d(x) \cdot x \Big|_{p_0}^{+\infty} - \int_{p_0}^{+\infty} d(x)\, dx \right] - p_0 d\big(p_0\big) = \int_{p_0}^{+\infty} d(x)\, dx. \end{aligned}$$

Thus, consumer surplus is the area under the demand curve, to the right of the applied price p_0:



Example: Consumer surplus under linear demand and cost

Let us consider an item with unit production cost €3 and an aggregate demand function

$$d(p) = (1000 - 100p)^+ .$$

Note that the limit price, resulting in no demand, is

$$p_{\mathsf{lim}} = \frac{1000}{100} = €10.$$

Then, the optimal price is

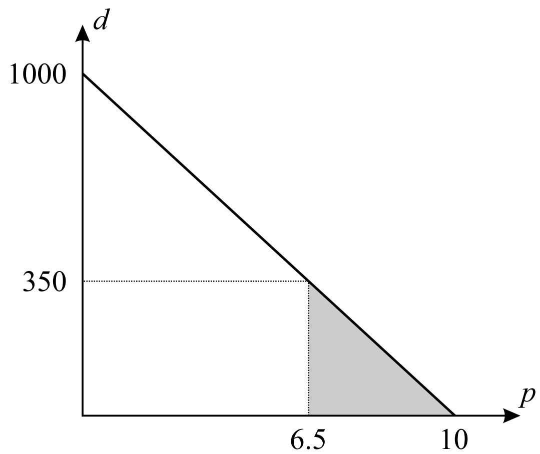
$$p^* = \frac{1000 + 100 \times 3}{2 \times 100} = 6.5,$$

with demand, profit, and consumer surplus given by

$$\begin{aligned} d\big(p^*\big) &= 1000 - 100 \times 6.5 = 350, \\ \pi^* &= 350 \times (6.5 - 3) = €1225, \\ S\big(p^*\big) &= \frac{(10 - 6.5) \times 350}{2} = 612.5, \end{aligned}$$

respectively.

Consumer surplus is the shaded area, and it may be interpreted as the opportunity loss that a firm incurs because of its inability to ask the reserve price from each consumer.



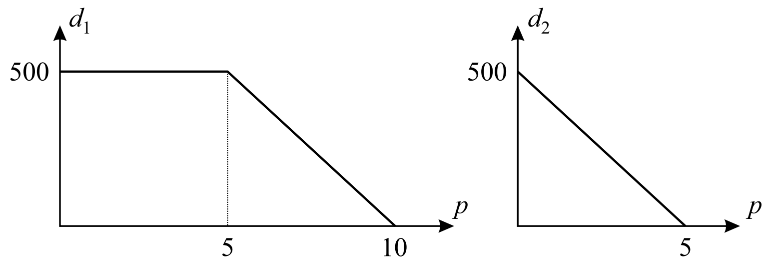
A natural objective for a firm is to "extract" consumer surplus, at least partially, by engaging in some form of price discrimination. If the firm is able to apply price discrimination, it will improve its profit. Intuition would suggest that this necessarily happens at the expense of consumers.

The following example shows that this need not be the case, if consumers with smaller willingness-to-pay are able to buy at a smaller price.

Example: Consumer surplus and price discrimination

Let us consider the same demand function of the previous example, but imagine that it results from the aggregation of two populations with respective demand functions

$$d_1(p) = \min\left\{500,\ (1000 - 100p)^+\right\},$$
$$d_2(p) = (500 - 100p)^+ .$$



This means that population 1 consists of consumers with willingness-to-pay larger than €5, whereas population 2 includes consumers with willingness-to-pay smaller than €5.

If the firm can discriminate, without incurring in arbitrage or dilution, the optimal price for population 1 would still be the same as before, just like demand and profit:

$$p_1^* = €6.5, \qquad d_1\left(p_1^*\right) = 350, \qquad \pi_1^* = €1225.$$

However, now the firm could sell items to population 2, with optimal price, demand and profit given by

$$p_2^* = \frac{500 + 100 \times 3}{2 \times 100} = €4,$$
$$d_2(p_2^*) = 500 - 100 \times 4 = 100,$$
$$\pi_2^* = 100 \times (4 - 3) = €100.$$

Being able to sell above cost to population 2, without reducing revenue from population 1, the firm earn an additional profit, so that total profit now is

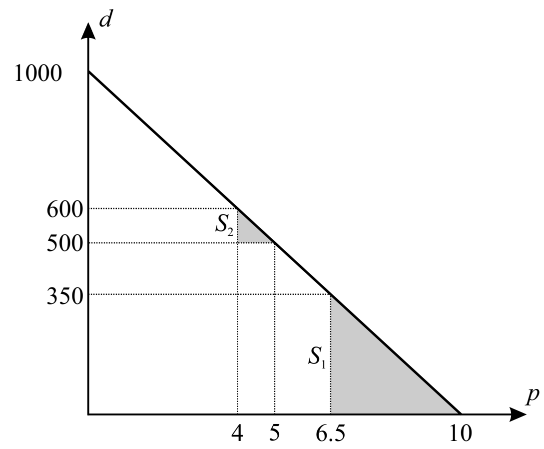
$$\pi^* = \pi_1^* + \pi_2^* = €1325.$$

The firm benefits from segmentation, but consumers do, too.

There is an additional consumer surplus, since 100 consumers are now able to buy at a price which is smaller than their reserve price:

$$S_2\big(p_2^*\big) = \frac{100 \times (5 - 4)}{2} = 50.$$

The total surplus consists of the two shaded triangles.



The logit function

Integrating willingness-to-pay density, much like we do with a PDF to find a CDF, gives a demand function.

A linear demand model does not rely on a sensible distribution of willingness-to-pay, and it is reasonable to guess a peaked function, with a maximum (mode) concentrated around the most typical willingness to pay (reserve price).

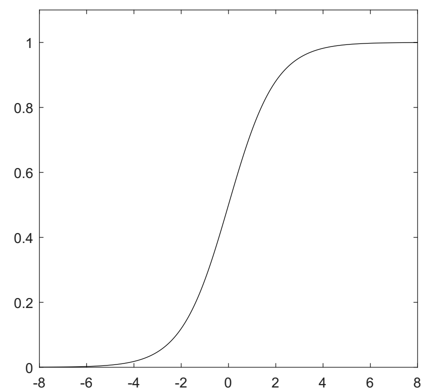
In the statistics and marketing literature, there are two functions that are used for this and other purposes, leading to probit and logit models.

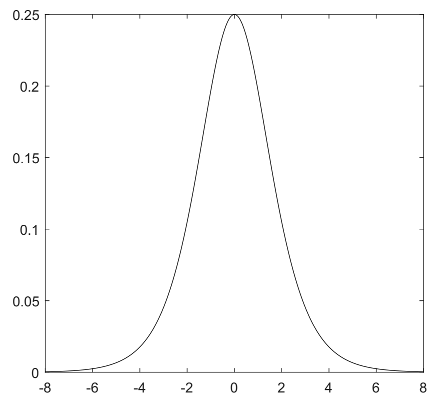
Probit models are obtained starting from the CDF of a standard normal, whereas logit models are built on the basis of the logistic (sigmoid) function.

The logistic function is defined as

$$L(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x},$$

which features a classical S-shape.





We observe that $L(0) = 0.5$ and that the logit function has a symmetry property,

$$L(-x) = 1 - L(x).$$

In the figure, we also show the derivative

$$L'(x) = \frac{e^x}{\left(1 + e^x\right)^2} = L(x) \cdot \left(1 - L(x)\right).$$

The derivative has a symmetry property $L'(x) = L'(-x)$, and its maximum occurs at $x = 0$.

Since demand functions should be decreasing, we should flip the x-axis from left to right and introduce sensible scaling and shift factors in the logit function, which yields the following (decreasing) logit demand function:

$$d(p) = \frac{ce^{-(\alpha+\beta p)}}{1 + e^{-(\alpha+\beta p)}},$$

where $\beta, c > 0$.

The larger β, the larger the price sensitivity, and the function is steepest for

$$\alpha + \beta p^\circ = 0 \quad \Rightarrow \quad p^\circ = -\alpha/\beta,$$

which suggests $\alpha < 0$. This is the value at which the density of willingness-to-pay is largest, and it may be considered as a "market price."

The corresponding willingness to pay is

$$w(x) = -\frac{d'(p)}{d(0)} = \frac{Ke^{-(\alpha+\beta x)}}{\left[1 + e^{-(\alpha+\beta x)}\right]^2},$$

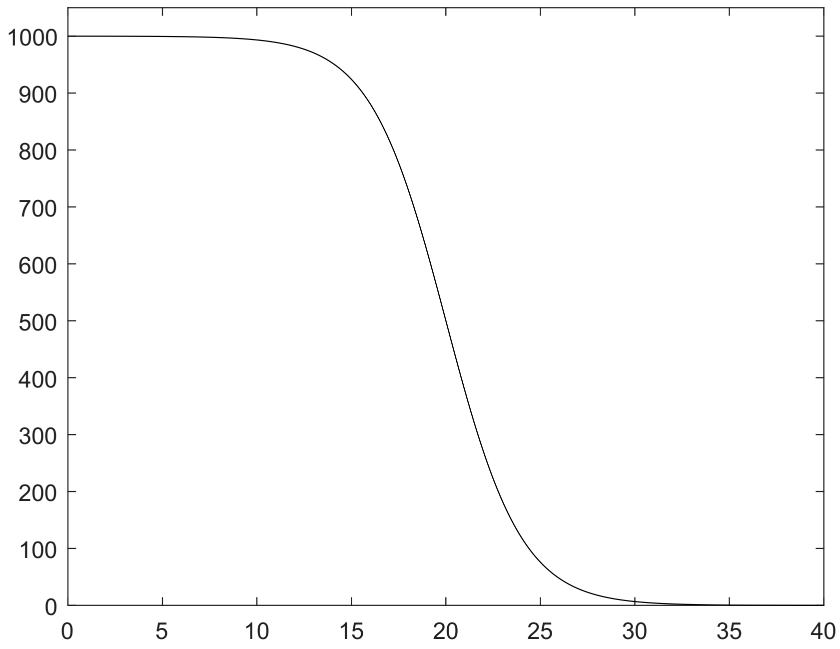
where $K = \beta c / d(0)$.

Here, we show the case for $c = 1000$, $p^\circ = 20$, and $\beta = 0.5$. In this case,

$$\alpha = -p^\circ \beta = -20 \times 0.5 = -10,$$

and the demand for $p = 0$ is

$$\begin{aligned} d(0) &= \frac{1000 \cdot e^{-\alpha}}{1 + e^{-\alpha}} \\ &= \frac{1000 \cdot e^{-10}}{1 + e^{-10}} = 999.9546 \approx c. \end{aligned}$$



A hybrid model

Different models, as we shall see, may be built based on quantities or prices. Here we consider an instructive numerical example, borrowed from (Kuyumcu and Popescu, 2006), showing that prices and quantities need not be mutually exclusive decision variables.

Let us consider two items following the linear demand model

$$d_1 = 500 - p_1 - 5p_2,$$
$$d_2 = 10 - 0.01p_1 - 0.05p_2.$$

Inspection of signs shows that the two goods are complements, rather than substitutes (decreasing p_2 increases d_1). In the original paper, the two goods are regular and meeting rooms at a hotel, which are obviously not substitutes. This helps in understanding the numerical values of the parameters.

In practice, there may be capacity limitations. In this case, there is a limited room availability, 250 and 6, respectively. Thus, we need to model demand explicitly.

To find optimal prices, we could consider the following model:

$$\begin{aligned}
\max \quad & p_1 d_1 + p_2 d_2 \\
\text{s.t.} \quad & d_1 = 500 - p_1 - 5p_2, \\
& d_2 = 10 - 0.01p_1 - 0.05p_2 \\
& p_1, p_2, d_1, d_2 \geq 0 \\
& d_1 \leq 250,\ d_2 \leq 6.
\end{aligned}$$

However, this is not a quite sensible model. To see this, let us consider its solution, in terms of optimal prices:

$$p_1^* = 400,\ p_2^* = 0.$$

These prices imply demands $d_1^* = 100$ and $d_2^* = 6$. We notice that meeting rooms are given for free, which might be a sensible strategy to boost sales of the complementary good.

Indeed, in some extreme cases, the price of a good could also be negative, in order to boost sales of the complementary good. This may look weird but, in fact, some goods may be sold below cost in order to get a hold of revenue from complementary goods. For instance a laser printer allows to sell toner cartridges and a razor allows to sell blades.

The price for regular rooms is fairly large, and some rooms are left unused, whereas capacity is binding for meeting rooms. The overall revenue is 40,000.

Now, what is wrong with this model?

The price of the regular rooms is fairly large, and it is a value such that demand for meeting rooms is exactly the availability.

A common practice in revenue management is inventory rationing, i.e., the restriction of sales. This is the core tool in quantity-based revenue management, as we shall see.

Indeed, we may decouple sales from demand in our problem, by solving the following alternative model formulation:

$$\begin{aligned} \max \quad & p_1 q_1 + p_2 q_2 \\ \text{s.t.} \quad & q_1 \leq 500 - p_1 - 5p_2, \\ & q_2 \leq 10 - 0.01p_1 - 0.05p_2 \\ & p_1, p_2, q_1, q_2 \geq 0 \\ & q_1 \leq 250, \ q_2 \leq 6. \end{aligned}$$

Here, the linear demand function gives an *upper bound* on the quantity offered.

Now, the optimal solution is

$$p_1^* = 250, \ p_2^* = 0, \ q_1^* = 250, \ q_2^* = 6,$$

with revenue 62,500. The problem with the first model formulation is that this solution would yield a demand 7.5 for the second good, which is not feasible if it is tightly linked with sales. Decoupling by rationing provides us with additional degrees of freedom.

Part 2

Game theoretic models

References

- C. d'Aspremont, J. Jaskold Gabszewicz, J.-F. Thisse, *On Hotelling's "Stability in Competition"*, Econometrica 1979, Vol. 47, No. 5, pp. 1145–1150
- O. Shy. *Industrial Organization: Theory and Applications*. MIT Press, 1995.
- J. Tirole. *The Theory of Industrial Organization.* MIT Press, 2003.

Sometimes, pricing decisions cannot be taken while ignoring the overall context.

Strategic consumers are a concern in a B2C setting. Strategic interactions are even more relevant in a B2B setting, where manufacturers may be connected by a supply chain on which intermediate goods are traded and transformed, or where producers and retailers in charge of distribute end items interact.

In such a context, prices may play an important role as a tool to coordinate actions and share risks.

Last but not least, what about competition among firms? We should distinguish between competition in terms of prices or quantities, and identical vs. differentiated goods/services.

These problems may be (at least partially) addressed by game theoretic models.

For the sake of simplicity, we shall only consider rather stylized games:

- There are only two decision makers (players); each player has an objective (payoff) that she wants to maximize and there is no form of cooperation.

- Only one decision has to be made by each player; hence, we do not consider sequential games in which multiple decisions are made over time.

- We assume complete information and common knowledge. Formalizing these concepts precisely is not that trivial, but loosely speaking they mean that there is no uncertainty about the data of the problem, nor about the mechanisms that map decisions into payoffs. The two players agree on their view of the world, the rules of the game, and know the incentives of the other party; furthermore, each player knows that the other one has all of the relevant information.

In order to get closer to a formalization of the problem, let us consider the decision problem

$$\begin{aligned} \text{max} \quad & \pi_1(x_1, x_2) + \pi_2(x_1, x_2) \\ \text{s.t.} \quad & x_1 \in S_1, \ x_2 \in S_2 \end{aligned} \tag{4}$$

The objective function (4) can be interpreted in terms of a profit depending on two decision variables, x_1 and x_2, which must stay within feasible sets S_1 and S_2, respectively.

Note that, even though the constraints on x_1 and x_2 are separable, we cannot decompose the overall problem, since the two decisions interact through the two profit functions $\pi_1(x_1, x_2)$ and $\pi_2(x_1, x_2)$.

Nevertheless, by solving the problem, we may find optimal decisions, x_1^* and x_2^*, yielding the optimal total profit

$$\pi_{1+2}^* = \pi_1(x_1^*, x_2^*) + \pi_2(x_1^*, x_2^*)$$

In doing so, we assume that there is either a single stakeholder in charge of making both decisions, or a pair of cooperative decision makers, in charge of choosing x_1 and x_2, respectively, but sharing a common desire to maximize the overall sum of profits.

But how about the quite realistic case of two *noncooperative* decision makers, associated with profit functions $\pi_1(x_1, x_2)$ and $\pi_2(x_1, x_2)$, respectively?

Decision maker 1 wishes to solve the problem

$$\begin{aligned} \text{max} \quad & \pi_1(x_1, x_2) \\ \text{s.t.} \quad & x_1 \in S_1, \end{aligned} \tag{5}$$

whereas decision maker 2 wishes to solve the problem

$$\begin{aligned} \text{max} \quad & \pi_2(x_1, x_2) \\ \text{s.t.} \quad & x_2 \in S_2. \end{aligned} \tag{6}$$

Unfortunately, these two problems, stated as such, make no sense. Which value of x_2 should we consider in problem (5)? Which value of x_1 should we consider in problem (6)? We must clarify how the two decision makers make their moves.

1. One possibility is that the two decision makers act sequentially. For instance, decision maker 1 might select $x_1 \in S_1$ before decision maker 2 selects $x_2 \in S_2$. In this case, we may say that decision maker 1 is the *leader*, and decision maker 2 is the *follower*. In making her choice, decision maker 1 could try to anticipate the reaction of decision maker 2 to each possible value of x_1.

2. Another possibility is that the two decisions are made simultaneously. In such a case, we need conceptual tools to understand which kind of decisions we might expect.

Game theory aims at finding a sensible prediction of an *equilibrium solution* $(x_1^{\text{e}}, x_2^{\text{e}})$, which depends on the precise assumptions that we make about the structure of the game. Whatever equilibrium solution we obtain, it cannot yield an overall profit larger than π^*_{1+2}, as the following inequality necessarily holds:

$$\pi^{\text{e}}_{1+2} = \pi_1(x_1^{\text{e}}, x_2^{\text{e}}) + \pi_2(x_1^{\text{e}}, x_2^{\text{e}}) \leq \pi_1(x_1^*, x_2^*) + \pi_2(x_1^*, x_2^*) = \pi^*_{1+2}$$

If this inequality were violated, (x_1^*, x_2^*) would not be the optimal solution of problem (4). This means that, if decentralize decisions, the overall system is likely to fail in achieving the overall optimal performance.

In the following, we consider Nash equilibrium for the case in which a continuum of infinite actions is available to each player.

First, we analyze the behavior of two firms competing with each other in terms of quantities. Both firms would like to maximize their profit, but they influence each other since their choices of produced quantities have an impact on the price at which the product is sold on the market. This price is common to both firms, as we assume that they produce a perfectly identical product.

This kind of competition is called *Cournot competition*; the case in which firms compete on prices is called *Bertrand competition*.

In the case of competition on prices, we should distinguish homogeneous or differentiated products.

After dealing with static games with simultaneous moves, we will discuss a simple example of sequential moves.

A game with simultaneous moves, where actions are quantities, leads to the *Cournot–Nash equilibrium*. This may be relevant in markets, like energy, where firms selects quantities and prices are settled by a possibly complicated auction mechanism.

To clarify the concept, it is useful to tackle a very simple model, in which we assume that each firm has a cost structure involving only a variable cost:

$$\mathsf{TC}_i(q_i) = c_i q_i, \qquad i = 1, 2$$

TC_i denotes total cost for firm i, c_i is the variable cost, and q_i is the amount produced by firm $i = 1, 2$.

The total amount available on the market is $Q = q_1 + q_2$, and it is going to influence price according . to a linear inverse demand function:

$$P(Q) = a - bQ, \qquad a, b > 0;\ a \geq c_i.$$

Incidentally, this stylized model assumes implicitly that all produced items are sold on the market. Then, the profit for firm i is

$$\pi_i(q_1, q_2) = P(q_1 + q_2)q_i - \mathsf{TC}_i(q_i) = \big[a - b(q_1 + q_2)\big]q_i - c_i q_i, \qquad i = 1, 2.$$

Assuming that the two firms make their decisions simultaneously, it is natural to wonder what the Nash equilibrium will be (we assume complete information and common knowledge).

We can find the equilibrium by finding the best response function $R_i(q_j)$. The stationarity condition for the profit of firm 1 yields

$$\frac{\partial \pi_1(q_1, q_2)}{\partial q_1} = a - 2bq_1 - bq_2 - c_1 = 0 \quad \Rightarrow \quad R_1(q_2) = \frac{a - c_1}{2b} - \frac{1}{2}q_2 \tag{7}$$

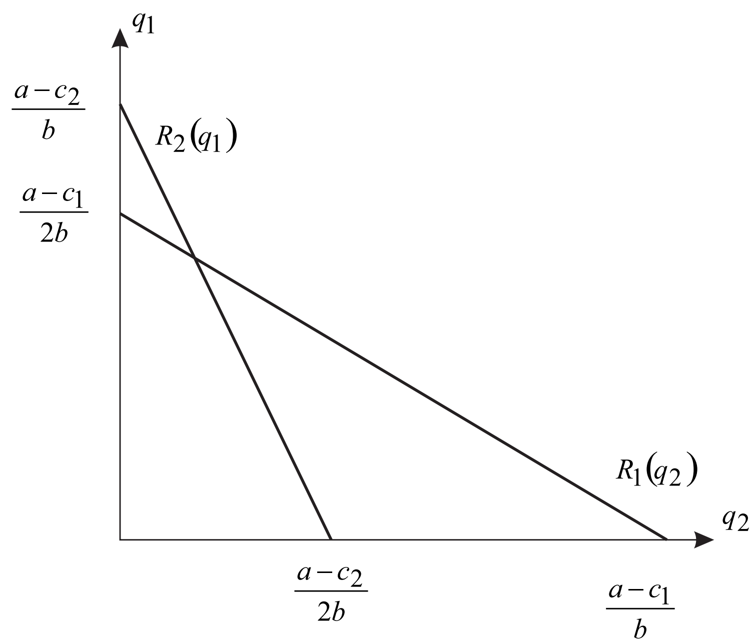
By the same token, for firm 2 we obtain

$$R_2(q_1) = \frac{a - c_2}{2b} - \frac{1}{2}q_1 \tag{8}$$

To solve the problem, we should find where the two response functions intersect; in other words, we should solve the system of equations

$$\begin{cases} q_1^{\mathrm{c}} = R_1(q_2^{\mathrm{c}}) \\ q_2^{\mathrm{c}} = R_2(q_1^{\mathrm{c}}) \end{cases}$$

where we use the superscript "c" to denote Cournot equilibrium. Here, response functions are downward-sloping lines.



Hence, to find the Nash equilibrium we simply solve the system of linear equations

$$\begin{cases} q_1^{\mathsf{c}} = \dfrac{a - c_1}{2b} - \dfrac{1}{2} q_2^{\mathsf{c}} \\ q_2^{\mathsf{c}} = \dfrac{a - c_2}{2b} - \dfrac{1}{2} q_1^{\mathsf{c}} \end{cases}$$

which yields

$$q_1^{\mathsf{c}} = \frac{a - 2c_1 + c_2}{3b}, \quad q_2^{\mathsf{c}} = \frac{a - 2c_2 + c_1}{3b} \tag{9}$$

The resulting equilibrium price turns out to be

$$p^{\mathsf{c}} = \frac{a + c_1 + c_2}{3} \tag{10}$$

and the profit of each firm is

$$\begin{aligned} \pi_i^{\mathsf{c}} &= (p^{\mathsf{c}} - c_i) q_i = \left(\frac{a + c_i + c_j}{3} - c_i \right) \left(\frac{a - 2c_i + c_j}{3b} \right) \\ &= \frac{(a - 2c_i + c_j)^2}{9b} = b (q_i^{\mathsf{c}})^2 \end{aligned} \tag{11}$$

If a firm manages to reduce its cost, it will increase its produced quantity and profit as well. If the firms have the same production technology (i.e., $c_1 = c_2$), then we have a symmetric solution $q_1^{\mathsf{c}} = q_2^{\mathsf{c}}$, as expected.

Let us consider two firms that compete on prices for a homogeneous good, under the simple assumption of linear demand and cost functions.

To analyze the problem, we have to clearly specify how consumers react to prices. Since we are assuming a homogeneous product, it is sensible to assume that consumers will buy the cheaper one; if prices are the same, let us assume that market is equally split between the two firms. Here, we do not consider capacity constraints.

Then, the quantity sold by firm i (where $i = 1, 2$, and $j = 2, 1$ refers to the competitor) is

$$q_i = \begin{cases} 0 & \text{if } p_i \geq \alpha/\beta, \\ 0 & \text{if } p_i > p_j, \\ \dfrac{\alpha - \beta p}{2} & \text{if } p_i = p_j \equiv p < \alpha/\beta, \\ \alpha - \beta p_i & \text{if } p_i < \min\{p_j, \alpha/\beta\}. \end{cases} \tag{12}$$

The idea is that there is no demand when a firm price exceeds the limit price a/b or when the price is strictly larger than the competitor's price. Demand is split equally when prices are the same (and below the limit). Otherwise, the cheaper firm captures all demand.

A Bertrand–Nash equilibrium is a quadruple of prices and quantities, $\left(p_1^b,\ p_2^b,\ q_1^b,\ q_2^b\right)$, such that:

1. p_1^b solves the problem $\max_{p_1}\ \pi_1\left(p_1, p_2^b\right) = \left(p_1 - c_1\right)q_1$, for $p_2 = p_2^b$.

2. p_2^b solves the problem $\max_{p_2}\ \pi_2\left(p_1^b, p_2\right) = \left(p_2 - c_2\right)q_2$, for $p_1 = p_1^b$.

3. The quantities q_1 and q_2 are given by Eq. (12).

Clearly, prices will not be set below marginal costs c_1 and c_2, respectively.

The key concept behind Bertrand–Nash equilibrium is **undercutting**. If $c_1 < c_2$, firm 1 can undercut firm 2 by applying a price strictly less than p_2. Strictly speaking, there is no equilibrium, since there is no maximum (but only a supremum). In fact, there is a sort of discontinuity, unlike Cournot competition, since an infinitesimal decrease of price may induce a jump demand for a firm.

However, it may be shown that there is an equilibrium when the marginal cost is the same c for both firms:

$$p_1 = p_2 = c, \quad q_1 = q_2 = \frac{\alpha - \beta c}{2}.$$

It is fairly easy to prove the result by contradiction; for any other setting of prices, a firm will have an incentive to deviate.

The net result, is that there may be no profit for either firm.

If the marginal costs are different, finding a maximum (and an equilibrium), requires that prices can only change by a given amount ϵ. This could be, for instance, a cent of Euro.

If $c_2 - c_1 > \epsilon$, then firm 1 may undercut by setting $p_1 = p_2 - \epsilon$, and the resulting equilibrium is

$$p_2 = c_2, \quad p_1 = c_2 - \epsilon, \quad q_1 = \alpha - \beta(c_2 - \epsilon), \quad q_2 = 0.$$

Leaving technicalities aside, this prediction is a bit at odds with empirical findings.

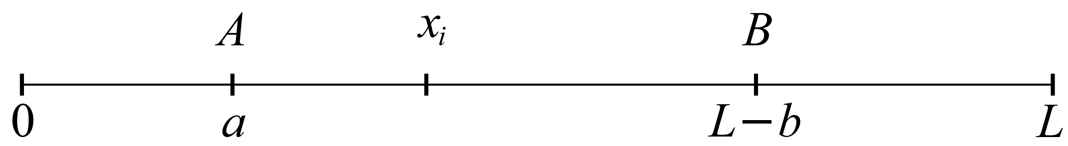
In practice, the cost structure is not simply linear, there are capacity constraints, and goods are not actually homogeneous. Other complications may arise, so that there need not be a single firm on the market.

Seeing things the other way around, the above result suggests that firms should differentiate their offerings, and take advantage of dishomogeneity among consumers.

To deal with nonhomogeneous (differentiated) items, we need a consumer choice model.

Let us consider a one-dimensional model of consumers' preferences. The items produced by two firms, A and B, are characterized by a feature, measured on a scale from 0 to L. Let us disregard production costs, for the sake of simplicity.

Firm A produces an item with feature at level a, and firm B produces an item with feature at level $L - b$, with distance b from the maximum level.



This geometrical representation is known as the Hotelling's linear street model.

Each consumer has a preferred level for the feature, which is represented by her position on the street. Let us assume that consumers are uniformly distributed on the street, from level 0 to level L.

Each consumer has a preferred brand, the one closer to her taste. If the two prices, p_A and p_B, are the same, each consumer would just buy the closer brand. However, there is a tradeoff between price and satisfaction.

Let us imagine that we may express the (dis)utility of each consumer by measuring a "transportation" cost τ. Hence, for a consumer at position x on the scale, utility is

$$u(x) = \begin{cases} -p_A - \tau \cdot |x - a| & \text{if she chooses brand } A, \\ -p_B - \tau \cdot |x - (L - b)| & \text{if she chooses brand } B. \end{cases}$$

Then, we may look for a critical customer, located at x_i, such that $a < x_i < L - b$, who is indifferent between the two brands:

$$-p_A - \tau \cdot (x_i - a) = -p_B - \tau \cdot (L - b - x_i),$$

which yields

$$x_i = \frac{p_B - p_A}{2\tau} + \frac{L - b + a}{2}. \tag{13}$$

As a reality check, note that when prices are the same, x_i is the midpoint between the two brand locations. Firm A will capture all demand in the interval $[0, x_i)$, so its demand function is just given by x_i, whereas the demand function for brand B is

$$L - x_i = \frac{p_A - p_B}{2\tau} + \frac{L + b - a}{2}.$$

We may wonder whether there is a Bertrand–Nash equilibrium in prices, for fixed locations.

Firm A, for a given p_B solves

$$\max_{p_A} \pi_A = \frac{p_B p_A - p_A^2}{2\tau} + \frac{L-b+a}{2} \cdot p_A,$$

whereas firm B, for a given p_A, solves

$$\max_{p_B} \pi_B = \frac{p_A p_B - p_B^2}{2\tau} + \frac{L+b-a}{2} \cdot p_B.$$

The first order stationarity conditions are

$$\begin{aligned} \frac{\partial \pi_A}{\partial p_A} &= \frac{p_B - 2p_A}{2\tau} + \frac{L-b+a}{2} = 0, \\ \frac{\partial \pi_B}{\partial p_B} &= \frac{p_A - 2p_B}{2\tau} + \frac{L+b-a}{2} = 0, \end{aligned}$$

which yield the equilibrium prices

$$p_A^e = \frac{\tau(3L-b+a)}{3}, \qquad p_B^e = \frac{\tau(3L+b-a)}{3}.$$

The profit for firm A is

$$\pi_A^e = x_i^e p_A^e = \frac{\tau(3L-b+a)^2}{18}.$$

In the symmetric case $a = b$, which does *not* imply that the locations are the same, the prices are the same and profit boils down to

$$\frac{\tau L^2}{2}$$

for both firms. We notice that the larger the transportation cost, the larger the prices. This also applies in situations with switching costs.

To be precise, we should not take for granted that there is a pair of strictly positive equilibrium prices. The following may be shown:

1. If the brands are homogeneous (located at the same point), $p_A^e = p_b^e = 0$, because of undercutting.

2. If the brands are not too close, then the above equilibria apply; otherwise, there is no equilibrium (see next discussion).

3. There is no equilibrium in the game where both locations and prices are decision variables.

Let us consider the symmetric case, where $a = b$. Based on the previous discussion, we should have

$$p_A^e = p_B^e = \tau L, \qquad \pi_A^e = \pi_B^e = \frac{\tau L^2}{2}.$$

Actually, we may prove that this holds only for $a \leq L/4$.*

Let us assume that $p_B = \tau L$, the equilibrium price, and let us change p_A. Actually, firm A captures the whole market if even the consumer located at $L - b$ chooses A. This happens if the price p_A plus the transportation cost is smaller than $p_B = \tau L$. Observing that the distance between A and B is $L - 2a$, A captures the market if

$$p_A < \tau L - \tau(L - 2a) = 2a\tau = p_{A,\text{lim}}.$$

In this case, the profit for A is an increasing linear function of p_A, and its limit value is

$$\pi_A^\circ = p_{A,\text{lim}} L = 2a\tau L.$$

For the limit price, there is an ambiguity, as consumers to the right of B are indifferent. We might assume that that market segment is split in equal parts. Anyway, there is a discontinuity in the profit of A as a function of its price.

*See d'Aspremont et al. for a thorough discussion. The present discussion is based on Shy.

Above $p_{A,\text{lim}}$, the market is split, and from Eq. (13) we obtain

$$\pi_A = \left[\frac{p_B - p_A}{2\tau} + \frac{L - b + a}{2}\right] p_A = \left[\frac{\tau L - p_A}{2\tau} + \frac{L}{2}\right] p_A = p_A L - \frac{p_A^2}{2\tau}.$$

At equilibrium, when $p_A = \tau L$, this boils down to $\pi_A^e = \tau L^2/2$. Then, when p_A gets large enough, firm B captures the whole market, and π_A drops to 0.

We have an equilibrium (i.e., there is a max and not only a sup), if $\pi_A^e \geq \pi_A^\circ$, which requires

$$\frac{\tau L^2}{2} \geq 2a\tau L \qquad \Rightarrow \qquad a \leq \frac{L}{4},$$

which means that the firms should be far enough from each other to ensure existence of an equilibrium.

So far, we have assumed that the two competing firms play simultaneously and have to make similar decisions. It is sometimes more natural to assume that one of the two players moves first.

For instance, a manufacturer may have to decide the wholesale price, and a retailer the purchased amount (assuming that the selling price is fixed).

Hence, we may also wonder what happens in the quantity game, if we assume that firm 1, the leader, sets its quantity q_1 before firm 2, the follower. Unlike the simultaneous game, firm 2 knows the decision of firm 1 before making its decision; thus, firm 2 has perfect information.

The analysis of the resulting sequential game leads to *von Stackelberg equilibrium*.

Firm 1 makes its decision knowing the best response function for firm 2, as given in Eq. (8). Hence, the leader's problem is

$$\max_{q_1} \pi_1^{\mathsf{s}} = P\Big(q_1 + R_2(q_1)\Big)q_1 - c_1 q_1 = \left[a - b\left(q_1 + \frac{a - c_2}{2b} - \frac{q_1}{2}\right)\right] q_1 - c_1 q_1$$

where the superscript "s" refers to von Stackelberg competition. Applying the stationarity condition yields

$$q_1^{\mathsf{s}} = \frac{a - 2c_1 + c_2}{2b} = \frac{3}{2} q_1^{\mathsf{c}} \qquad (14)$$

Firm 1 produces more in this sequential game than in the Cournot game.

If we plug this value into the best response function $R_2(q_1)$, we obtain

$$\begin{aligned} q_2^{\mathsf{s}} &= \frac{a-c_2}{2b} - \frac{a-2c_2+c_1}{4b} \\ &= \frac{a-3c_2+2c_1}{4b} = \frac{a-2c_2+c_1+(c_1-c_2)}{4b} \\ &= \frac{3}{4}q_2^{\mathsf{c}} + \frac{c_1-c_2}{4b}. \end{aligned} \tag{15}$$

The output of firm 2 is a fraction of that of the Cournot game, plus a term that is positive if firm 1 is less efficient than firm 2.

Now it would be interesting to compare the profits for the two firms under this kind of game. This is easy to do when marginal production costs are the same; we illustrate the idea with a toy numerical example.

Two firms have the same marginal production cost, $c_1 = c_2 = 5$, and the market is characterized by the price/quantity function

$$P(Q) = 120 - Q$$

In this example we compare three cases:

1. The two firms collude and work together as a cartel. We may also consider the two firms as two branches of a monopolist firm.
2. The firms do not cooperate and move simultaneously (Cournot game).
3. The firms do not cooperate and move sequentially (von Stackelberg game).

In the first case, we just need to work with the aggregate output Q. The monopolist solves the problem

$$\max \pi^{\mathrm{m}} = (120 - Q)Q - 5Q.$$

We solve the problem by applying the stationarity condition

$$120 - 2Q - 5 = 0 \qquad \Rightarrow \qquad Q^m = 57.50$$

which yields the following market price and profit:

$$p^{\mathrm{m}} = 120 - 57.5 = 62.50, \qquad \pi^{\mathrm{m}}_{1+2} = (62.50 - 5) \times 57.50 = 3306.25$$

In the second case, the solution given by (9) is symmetric:

$$q_1^{\mathrm{c}} = q_2^{\mathrm{c}} = \frac{120 - 10 + 5}{3} = 38.33$$

The overall output and price are

$$Q^{\mathrm{c}} = 2 \times 38.33 = 76.77, \qquad p^{\mathrm{c}} = 120 - 76.77 = 43.33$$

respectively. The profit for each firm is

$$\pi_1^{\mathrm{c}} = \pi_2^{\mathrm{c}} = (q_1^{\mathrm{c}})^2 = 1469.19$$

Note that the total overall profit is

$$\pi_{1+2}^{\mathrm{c}} = 2 \times 1469.19 = 2938.89 < 3306.25 = \pi_{1+2}^{\mathrm{m}}$$

In fact, the monopolist would restrict output to increase price, resulting in a larger overall profit than with the Cournot competition. So, collusion results in a larger profit than competition, which is no surprise.

Let us consider now the von Stackelberg sequential game. Using (14) and (15), we see that

$$q_1^{\mathrm{s}} = \frac{120 - 10 + 5}{2} = 57.5, \qquad q_2^{\mathrm{s}} = \frac{120 - 10 + 5}{4} = 28.75$$

from which we see that, with respect to the simultaneous game, the output of firm 1 is increased whereas the output of firm 2 is decreased.

The total output and price are

$$Q^{\mathsf{s}} = 57.5 + 28.75 = 86.25, \qquad p^{\mathsf{s}} = 120 - 86.25 = 33.75$$

respectively. The price is lower than in both previous cases, and the distribution of profit is now quite asymmetric:

$$\begin{aligned}
\pi_1^{\mathsf{s}} &= (33.75 - 5) \times 57.5 = 1653.13 \\
\pi_2^{\mathsf{s}} &= (33.75 - 5) \times 28.75 = 826.56 \\
\pi_{1+2}^{\mathsf{s}} &= 1653.13 + 826.56 = 2479.69
\end{aligned}$$

The overall profit for the sequential game is lower than for the simultaneous one; however, the leader has a definite advantage and its profit is larger in the sequential game.

The toy example above shows that the privilege of moving first may yield an advantage to the leader.

Given the structure of the game, it is easy to see that the leader of the sequential game cannot do worse than in the simultaneous game; in fact, she could produce the same amount as in the Cournot game, anyway.

However, this need not apply in general. In particular, when there are asymmetries in information or things are random, the choice of the leader, or its outcome when there is uncertainty, could provide the follower with useful information. The following example shows that being the first to move is not always desirable.

Let us consider the battle of the sexes, where we assume that Juliet has the privilege of moving first.

	Romeo	
Juliet	Horror	Shopping
Horror	(1, 3)	(0, 0)
Shopping	(0, 0)	(3, 1)

whatever her choice, Romeo will play the move that allows him to enjoy her company. Hence, she will play *shopping* for sure and is certainly happy to move first.

The situation is quite different for the payoffs below:

	Romeo	
Morticia	Cinema	Restaurant
Cinema	(5, -100)	(0, 1)
Restaurant	(0, 1)	(5, -100)

In this case, Romeo is indifferent between going to cinema or restaurant. What he really dreads is an evening with Morticia. It is easy to see that this game has no Nash equilibrium, as one of the two players has always an incentive to deviate.

An equilibrium can be found if we admit mixed strategies, in which players select an action according to a probability distribution, related to the uncertainty about the move of the competitor. We do not consider mixed strategies here, but the important point in this case is that no player would like to move first.

We noted that the first version of the battle of the sexes is a stylized coordination game for two firms that should adopt a common standard; in this second version, one firm wants to adopt the same standard as the competitor, whereas the other firm would like to select a different one.

As an interesting application to pricing, let us consider a B2B setting, whereby a producer and a retailer interact by prices (See Tirole, 2003, Chapter 4).

It is interesting to compare pricing decisions and overall supply chain profits in two settings:

1. The vertically integrated firm, in charge of both production and distribution, where one decision maker is in charge of every decision.

2. The decentralized scheme, in which the producer decides the wholesale price and the retailer decides the market price.

In both cases, we consider a market with a simple linear demand function

$$d(p) = 1 - p,$$

where p is the market price, and assume a linear cost structure, with marginal cost $c < 1$.

The overall problem for a vertically integrated firm is

$$\max_p \, (p - c) \cdot (1 - p).$$

Solving the problem yields optimal price, market demand, and profit given by

$$p^*_{\text{vi}} = \frac{1+c}{2}, \qquad d^*_{\text{vi}} = \frac{1-c}{2}, \qquad \pi^*_{\text{vi}} = \frac{(1-c)^2}{4},$$

respectively.

If the chain is not integrated, each player will set a price. Let us assume that the producer is the leader in a sequential game, where she fixes a wholesale price p_w. We need the response function of the retailer, who sets the market price p_m by solving, for a given wholesale price p_w,

$$\max_{p_m} \; (p_m - p_w) \cdot (1 - p_m).$$

This yields the best response function

$$R_m(p_w) = \frac{1+p_w}{2}.$$

Then, the producer's problem becomes

$$\max_{p_w} \; (p_w - c) \cdot \left[1 - \frac{1+p_w}{2}\right],$$

which gives the equilibrium price

$$p^*_{w,\text{dec}} = \frac{1+c}{2}.$$

Then, the retail price is larger than in the vertically integrated case (remember that we assume $c < 1$),

$$p^*_{m,\mathsf{dec}} = \frac{3+c}{4},$$

and market demand is smaller,

$$d^*_{\mathsf{dec}} = \frac{1-c}{4}.$$

The overall profit of the decentralized control mechanism is the sum of two profits:

$$\pi^*_{\mathsf{dec}} = (p^*_{w,\mathsf{dec}} - c) \cdot d^*_{\mathsf{dec}} + (p^*_{m,\mathsf{dec}} - p^*_{w,\mathsf{dec}}) \cdot d^*_{\mathsf{dec}} = \frac{(1-c)^2}{8} + \frac{(1-c)^2}{16} = \frac{3(1-c)^2}{16} < \pi^*_{\mathsf{vi}}.$$

This kind of issue is known as double marginalization, and it is due to the fact that both players must apply a markup to the marginal cost they see, which is increasing along the chain.

In a more realistic setting, we should also consider demand uncertainty, which introduces issues in risk sharing.

In that case, pricing policies (two-part tariffs and buyback contracts) may be used to overcome the issue and coordinate the supply chain.

Part 3

Discrete choice models

- V.R. Rao. *Applied Conjoint Analysis*. Springer, 2014.
- K.E. Train. *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge University Press.
- G. van Ryzin. *Future of Demand Management: Models of Demand*. Journal of Revenue and Pricing Management, vol. 4, no. 2, 2005, pp. 204–210/

Link: `https://sawtoothsoftware.com/resources/technical-papers`

Conjoint analysis is a widely used method in marketing (and other fields) to design product and services, to assess price impact, and to build market simulators.

The underlying idea is that consumers react to alternatives represented by combination of attribute values.

There are a few variations of conjoint analysis, but the basic ones are:

- ratings-based conjoint analysis (the traditional one);
- choice-based conjoint analysis (the most common one).

The preference rating Y_i for alternative i could be modeled as

$$Y_i = \sum_{k=1}^{m} U_k(x_{ik}) + \epsilon_i,$$

where each function U_k captures the *partworth* for attribute k, and ϵ_i is an error term.

Note that asking consumers to prioritize attributes is typically useless (everything is important!). Asking for a rating Y_i may be a better option, and it lends itself to classical regression analysis (possibly using dummy variables).

Classical conjoint analysis infers partworths on the basis of consumer ratings of a set of alternatives (stimuli, profiles, etc.).

The additive character of the model is a clear limitation, but there are ways to introduce interactions among attributes, which may be quite relevant for specific values (e.g., "red" and "Ferrari").

However, assigning rating is not the way consumer make their choices. To build a better preference model, we must collect data from actual choices among a set of alternatives (including the *no purchase* one).

We may have:

- *revealed preference* data, where we observe actual market choices;

- *stated preference* data, where we analyze the reaction of a panel of consumers to a set of carefully crafted stimuli (hypothetical alternatives).

Choice-based conjoint analysis relies on stated preferences, as this offers extreme flexibility in stimuli generation and data collection (at the cost of some potential loss in realism).

Choice-based Conjoint Questions

Choice Model: Preview Questionnaire

Version: 1

Q1. Which of the following would you choose?

	AT&T	verizon	Sprint	T-Mobile
Price	$70 per month per line	$30 per month per line	$80 per month per line	$50 per month per line
Hotspot data	50GB	10GB	20GB	Unlimited
Hotspot speed	4G	3G	5G	5G
Streaming	HD Streaming (1080p quality)	Ultra HD Streaming (4K quality)	Ultra HD Streaming (4K quality)	HD Streaming (1080p quality)
Channels	Live TV: 35+ channels + Netflix + Hulu + Prime + HBO + Showtime	Live TV: 35+ channels + Netflix + Hulu + Prime	Live TV: 35+ channels + Netflix	Live TV: 35+ channels
Music	Tidal	Apple Music	Amazon Music	Spotify

Q2. Which of the following would you choose?

AT&T	verizon	Sprint	T-Mobile

DISPLAYR

RESEARCHSOFTWARE

A key role is played by careful design of experiments (orthogonal designs, etc.) in order to maximize the information collected, without exerting an excessive cognitive burden on the user (too many alternatives per choice, and too many choices).

In addition to classical issues in design of experiments, it is important:

- To choose (possibly combine) attributes and their levels.
- To consider prohibitions (nonsensical combinations).
- To avoid inconsistent profiles (better performance for a lower price).
- To take care of ordinal variables.

A typical model that can be estimated by CBC is a multinomial logit (MNL) model.

Sophisticated CBC version exist, based on Hierarchical Bayesian models, to account for consumer heterogeneity.

By assessing partworth utilities, we may predict the market share of items that are not currently offered, as well as the impact of pricing decisions.

Let us assume that the consumer chooses one among n goods, according to the utility of each alternative. The utility of good is $i \in [n]$ is $V_i + \epsilon_i$, i.e., the sum of a deterministic and a random component. The deterministic component is often referred to as **representative utility**.

Alternative i is selected if

$$V_i + \epsilon_i > V_j + \epsilon_j, \qquad \forall j \neq i.$$

Hence, the probability of selection for alternative i is

$$\pi_i = \mathsf{P}\{V_i + \epsilon_i > V_j + \epsilon_j,\ \forall j \neq i\} = \mathsf{P}\{\epsilon_j < \epsilon_i + V_i - V_j,\ \forall j \neq i\}.$$

Assuming independence among the random components, conditional on ϵ_i, we have

$$\pi_i \mid \epsilon_i = \prod_{j \neq i} F_j(\epsilon_i + V_i - V_j),$$

where F_j is the DF of ϵ_j. Hence,

$$\pi_i = \int_{-\infty}^{+\infty} \prod_{j \neq i} F_j(\epsilon_i + V_i - V_j) f_i(s)\, ds,$$

where f_i is the PDF of ϵ_i.

Now let us assume that all random components are i.i.d. and follow a Gumbel distribution, whose DF and PDF are, respectively,

$$F(x) = \exp\left(-e^{-x}\right), \quad f(x) = e^{-x}\exp\left(-e^{-x}\right).$$

Then

$$\pi_i = \int_{-\infty}^{+\infty} \prod_{j\neq i} \exp\left(-e^{-(s+V_i-V_j)}\right) \cdot e^{-s}\exp\left(-e^{-s}\right) ds.$$

To find the integral, we may observe that $s + V_i - V_j = s$ for $j = i$, so that we may include the factor $\exp\left(-e^{-s}\right)$ inside the product and then collect terms not depending on j:

$$\begin{aligned}\pi_i &= \int_{-\infty}^{+\infty} \prod_{j=1}^{n} \exp\left(-e^{-(s+V_i-V_j)}\right) \cdot e^{-s}\, ds = \int_{-\infty}^{+\infty} \exp\left(-\sum_{j=1}^{n} e^{-(s+V_i-V_j)}\right) \cdot e^{-s}\, ds \\ &= \int_{-\infty}^{+\infty} \exp\left(-e^{-s}\sum_{j=1}^{n} e^{-(V_i-V_j)}\right) \cdot e^{-s}\, ds.\end{aligned}$$

Now we apply a change of variable, $t = e^{-s}$, so that $-e^{-s}\, ds = dt$ and the lower and upper integration limits are set to $+\infty$ and 0, respectively.

Then, adjusting for the change in sign,

$$\pi_i = \int_0^\infty \exp\Big(-t\sum_{j=1}^n e^{-(V_i-V_j)}\Big)\,dt = \frac{\exp\big(-t\sum_j e^{-(V_i-V_j)}\big)}{-\sum_j e^{-(V_i-V_j)}}\Bigg|_0^\infty$$
$$= \frac{1}{\sum_j e^{-(V_i-V_j)}} = \frac{e^{V_i}}{\sum_j e^{V_j}},$$

which corresponds to the MNL model.

The MNL model is quite common in machine learning and classification, and it is relatively simple to deal with in optimization models (e.g., optimal assortment decisions).

Nevertheless, it has definite limitations, most notably the IIA (Independence from Irrelevant Alternatives) assumption. Let us consider the relative odds of choosing items i and k:

$$\frac{\pi_i}{\pi_k} = \frac{e^{V_i}\big/\sum_j e^{V_j}}{e^{V_k}\big/\sum_j e^{V_j}} = e^{V_i-V_k},$$

which does *not* depend on any other alternatives. Hence, the relative odds do not change when additional alternatives are included in the choice set.

Example (the red–blue bus problem). Consider a travel mode decision between car and bus (say, a blue bus), where we assume that the representative utilities are equal, so that $\pi_c = \pi_{bb} = 1/2$. Now let us introduce a red bus, for which it is reasonable to assume $\pi_{rb} = \pi_{bb}$.

With a MNL model, we have $\pi_c = \pi_{rb} = \pi_{bb} = 1/3$, whereas it would be more reasonable to assume a split in the probability of choosing a bus, so that $\pi_c = 1/2$ and $\pi_{rb} = \pi_{bb} = 1/4$.

Nested logit models are a possible approach to circumvent IIA, when needed.

If we assume that the random component of utility is normally distributed, we obtain a probit model.

In Extreme Value Theory (EVT), among other things, we investigate the distribution of block maxima.

Consider an i.i.d. sample of size n from a distribution with distribution function $F_X(x)$. What is the distribution of $M_n = \max\{X_1, \ldots, X_n\}$? What happens when $n \to \infty$?

Clearly, given independence,

$$F_{\mathsf{max}}(x) = \mathsf{P}\{M_n \le x\} = \mathsf{P}\{X_1 \le x, \ldots, X_n \le x\} = \mathsf{P}\{X_1 \le x\} \cdots \mathsf{P}\{X_n \le x\} = F_X(x)^n.$$

For a uniform random variable $U \sim \mathsf{U}(0,1)$,

$$F_U(x) = x \quad \Rightarrow \quad F_{\mathsf{max}}(x) = x^n, \qquad x \in [0,1].$$

When $n \to \infty$, the distribution collapses into a degenerate one, with a probability mass concentrated on $x = 1$. Note that, in this case, $\mathsf{E}[M_n] = \frac{n}{n+1}$.

A degenerate case like this, due to the bounded support, is not quite interesting. If we consider a random variable with an upward unbounded support, things should be more fun, but the distribution will shift towards $+\infty$. So what?

In the case of a sum $S_n = \sum_{k\in[n]} X_k$, which is relevant to the Central Limit Theorem, we essentially normalize the variable, by defining normalizing sequences a_n and b_n:

$$\lim_{n\to\infty} \mathsf{P}\left\{\frac{S_n - a_n}{b_n} \le x\right\} = \Phi(x), \qquad a_n = n\mathsf{E}(X_1),\ b_n = \sqrt{n\mathsf{Var}(X_1)}.$$

To obtain a limit distribution for block maxima, we can introduce normalizing sequences d_n and c_n and check whether

$$\lim_{n\to\infty} \mathsf{P}\left\{\frac{M_n - d_n}{c_n} \le x\right\} = \lim_{n\to\infty} F_X^n(c_n x + d_n) = H(x), \tag{16}$$

for some non-degenerate distribution function $H(x)$

Example. Consider an exponential distribution with rate $\beta > 0$, for which $F(x) = 1 - e^{-\beta x}$, $x \ge 0$. If we choose $c_n = 1/\beta$ and $d_n = (\log n)/\beta$, we find:

$$F^n(c_n x + d_n) = \left(1 - \frac{1}{n}e^{-x}\right)^n, \qquad x \ge -\log n,$$

$$\lim_{n\to\infty} F^n(c_n x + d_n) = \exp\left(-e^{-x}\right), \qquad x \in \mathbb{R}.$$

Definition: generalized extreme value (GEV) distribution. The distribution function of the standard GEV distribution is given by:

$$H_\xi(x) = \begin{cases} \text{exp}\left[-(1+\xi x)^{-1/\xi}\right], & \xi \neq 0, \\ \text{exp}\left(-e^{-x}\right), & \xi = 0, \end{cases}$$

where $1 + \xi x > 0$. If we introduce a location parameter $\mu \in \mathbb{R}$ and a scale parameter $\sigma > 0$, we obtain a parameterized family of distributions

$$H_{\xi,\mu,\sigma}(x) \doteq H_\xi\left(\frac{x-\mu}{\sigma}\right).$$

The parameter ξ defines the shape of a family of similar (i.e., of the same type) distributions.

We recall that when two random variables V and W differ in distribution by a scale parameter $a > 0$ and a location parameter $b \in \mathbb{R}$, i.e., $V \stackrel{d}{=} aW + b$, they have the same type (are similar).

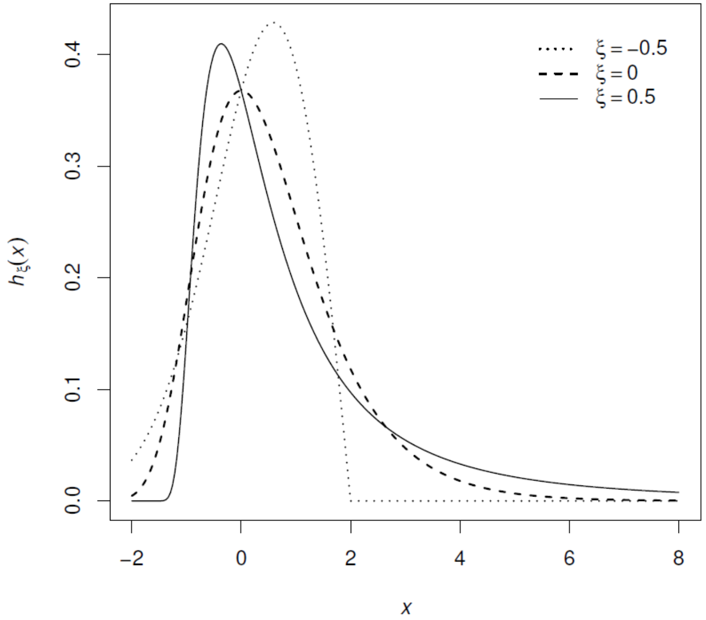
The figure shows densities for some values of the shape parameter (assuming $\mu = 0$, $\sigma = 1$).

- If $\xi > 0$, we have a Fréchet distribution.
- If $\xi = 0$, we have a Gumbel distribution.
- If $\xi < 0$, we have a Weibull distribution.

The Weibull distribution has a finite right endpoint.

The Fréchet distribution features a slower decay than Gumbel.

Note: The Gumbel distribution, among other things, plays a key role in discrete choice models.



If condition (16) holds for some non-degenerate H, then we say that the distribution function F is in the **maximum domain of attraction** of H: $F \in \text{MDA}(H)$.

In the case of an exponential distribution, we have $F \in \text{MDA}(H_0)$.

Theorem (Fisher–Tippett–Gnedenko). If $F \in \text{MDA}(H)$, for a non-degenerate H, then H must be a GEV distribution of type H_ξ for some value of ξ.