MPRA
Munich Personal RePEc Archive

Introduction to Competition Economics

Merino Troncoso, Carlos

Universidad Nacional de Educación a Distancia

15 January 2023

Online at https://mpra.ub.uni-muenchen.de/115999/
MPRA Paper No. 115999, posted 14 Apr 2023 06:57 UTC

INTRODUCTION TO COMPETITION ECONOMICS

Carlos Merino Troncoso

CHAPTER 1.- DEMAND ESTIMATION

1. Introduction

In this chapter I will review the main methodologies used in economics for demand estimation, focusing on recent trends such as the structural approach and machine learning techniques. As one can imagine the literature review is extensive so due to space limitations I can only provide a summarized view of each theory. Nevertheless, the interested reader has a comprehensive bibliography at the end of the chapter for extensions and examples. There is also another barrier when explaining any concept in economics. Economics is widely based on Mathematics, Statistics and Econometrics so it is not possible to explain it without its usage. As it is not possible review econometrics and mathematics in this chapter I will refer to specific texts, and an appendix will give the reader a brief summary of the main concepts. Demand is usually the first step in the study of a market. Intuitively, suppliers only start production when they identify consumer interest in a particular good. All models reviewed try to solve the problems that traditionally have embarrassed demand estimation: identification, endogeneity and simultaneity. There is no perfect solution to them, each model has its advantages and limitations and are based on assumptions that are often irreal, so the model in itself is in all cases only an approximation of demand.

2. Traditional approach to consumer behavior

As I have mentioned above, my intention is to do a review on consumer demand covering the traditional approach developed by Marshall (1890), to more recent approaches used in Industrial Organization literature such as Nevo (2000) and a brief introduction of machine learning (2015). Demand is the most important component for empirical competitive analysis. It is not possible to quantify the change in a company's behavior if we do not have information about the potential response of consumers. Demand is based on consumer tastes and there are two empirical findings any consumer theory has to deal with: consumers tastes are heterogeneous and products are differentiated (2000) .

Basic concepts

We recall the basic elements of consumer demand. I only provide a quick overview of the main elements of traditional consumer theory. Davies and Garcés provide an application to antitrust analysis (2009)**.**

Demand function describe the amount of a good a consumer would buy based on variables that affect this decision such as the P_i vector of consumer's price or income y .Figure 1 shows a typical example of linear demand function: p=20-2q which is generally written as $p_i = D(q_i, y)$ and is the inverse demand Marshall curve[1] (1890).

The typical negatively sloped linear demand is based on strong assumptions. It is represented as the demanded quantity for each price level of the good when all other variables that affect demand such as income levels and prices of substitute or complementary products remain fixed. All variables remain fixed except price and quantity of the good when we move along the demand curve. The slope of the demand curve indicates at each point what the consumer would be willing to reduce (increase) the amount demanded if the price increases (falls) while income and any other variable remain fixed.

[1] *"There is then one general law of demand: - The greater the amount to be sold, the smaller must be the price at which it is offered in order to find purchasers; or, in other words, the amount demanded increases with a fall in price and diminishes with a rise in price."* Alfred Marshall, Principles of Economics (eighth edition), p. 99.

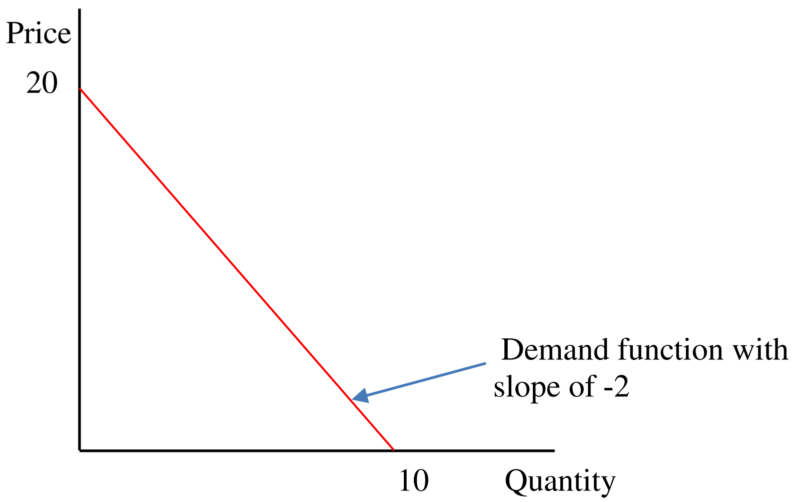


Figure 1: Inverse Demand

In the figure above (2009), if the price increases by 10 euros, demand will fall by 5 units. Consumers will not buy more units if the price is greater than 20 because at that level the price is greater than the value that the consumer assigns to the first unit of the good.

An interpretation of this curve is that it reflects the maximum price that the consumer wants to pay for units of the good q in question. Intuitively, consumers valuation will be lower as they own more units of the good. It is this declining marginal valuation that ensures that the demand curve typically has negative slope. Consumers will acquire a unit only if marginal valuation is greater than price so it can be said that the curve describes the consumers marginal valuation of the good.

Given this interpretation, the inverse demand curve describes the difference between the consumer valuation of each unit and the current price paid for it. This difference is called consumer surplus. At any price, we can add up the consumer surplus available in all units consumed (those with marginal valuation above price).

In a homogeneous product market, all goods are identical and perfect substitutes. This implies in theory that they all have the same price. In a market with differentiated product, they are not perfectly substitutes and prices will vary. In this market, demand for any product is determined by its price and that of potential substitutes. In practice, homogeneous markets end up being differentiated when viewed in detail. However, the assumption of homogeneity can be a reasonable approximation.

Demand functions

Demand functions are derived from consumer choice, assuming consumers maximize a utility function subject to a budget constraint. The existence of this utility function can be induced from some non-trivial assumptions. A detailed explanation of these assumptions is covered in Mas-Colell (1995). Maximizing utility is equivalent to choosing the most preferred product set given the budget constraint.

Mathematically, the problem can be represented as a consumer who chooses to maximize their welfare (utility function u) subject to disposable income y so that their total expenditure does not exceed their income:

$$Max\ u_i(q_1, q_2, \dots, q_n)$$

$$s.a. \sum_{i=1}^{n} p_i\, q_i = y$$

Where p_i and q_i are price and amounts of good, i, $u_i(q_1, q_2, \dots, q_n)$ is the individual's demand utility associated with i consuming that vector of amounts, y_i and is the individual's i disposable income.

The first order conditions of this problem are:

$$\frac{\delta u_i(q_1, q_2, \dots, q_n)}{\delta q_i} = \lambda p_i$$

$$\frac{\delta u_i(q_1, q_2, \dots, q_n)}{\delta p_i} = \lambda$$

Along with the budget constraint it gives *I+1* equations with *1+1* parameters: the quantities *I* and the value of the Lagrange multiplier, λ.

In the optimum, the first order conditions imply that the Lagrange multiplier is equal to the marginal income, which we assume constant. We assume, that consumer behavior is described by a utility function with an additive and separable q_i good, the price p_1 is standardized to *1*. The price of reference good q_1 is generally called money and its inclusion allows an interpretation of the first order conditions. In these circumstances a consumer that maximizes utility shall choose a basket of products so that the marginal profit of the last monetary unit disbursed on each product is equal to the marginal profit, i.e. 1.

In general, the solution to maximization describes the individual's problem as a function of the prices of all goods sold and consumer income. If we index the goods with i, we can write the individual demand as:

$$q_i = D(p_i, y)$$

A function of demanding a product i incorporates not only the effect of its own price on the amount demanded but also the effect of the disposable income and the price of other products whose supply may affect the quantity of the goods i purchased. In Figure 1 a change in price of product i represents a movement along the curve while a change in the income or price of other substitutes represent a change or rotation of the demand curve.

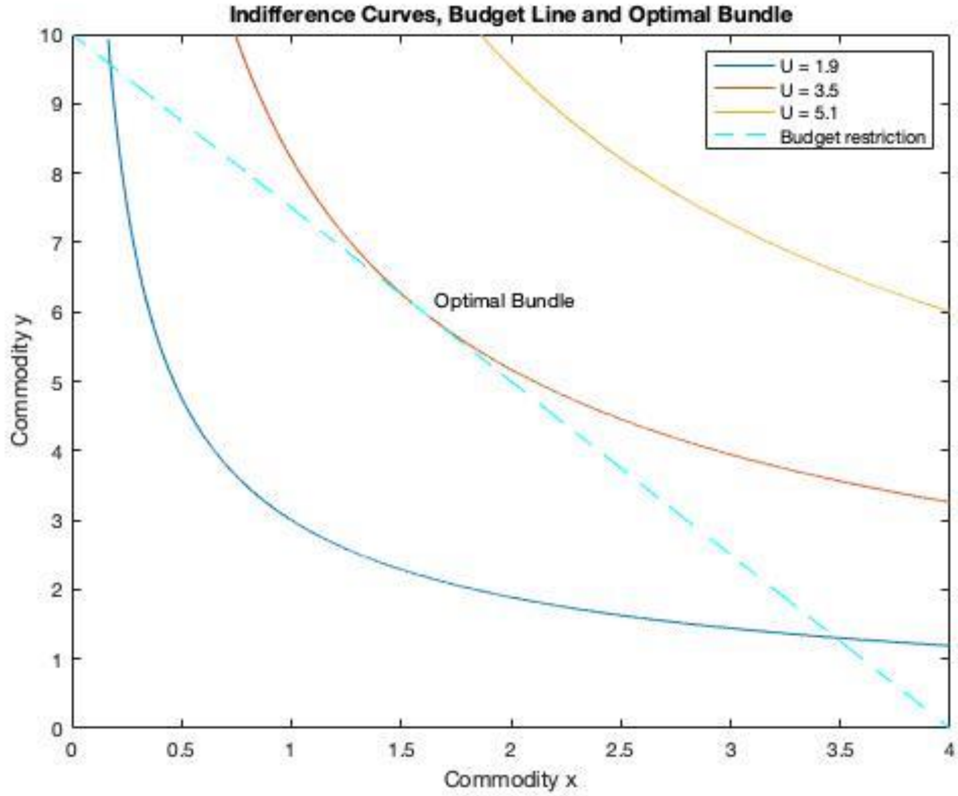
The utility generated by consumption is described by the direct utility u_i function, which relates the level of utility of the goods purchased and is not observable. We know that not all consumption levels are achievable because of the budgetary constraint and because the consumer will choose that set of goods that maximizes the utility function.

The indirect utility function, $V_i\big(q_1, q_2, \dots, q_{n,}, y\big) = u_i\big(q_1(p_1, \dots, p_n, y), q_2(p_1, \dots, p_n, y)\big)$ where , V_i describes the maximum utility a consumer can obtain at any price and revenue level. The following result will be important when developing demand functions that we estimate:

For any indirect function of V_i, there is a direct U_i function that represents the same preferences on goods provided that the indirect function of profit, is continuous in prices and income, not rising in prices, not decreasing in income, quasi-convex in *(p,y)* and homogeneous grade 0 in *(p,y)*.

This result although it seems purely theoretical can actually be very useful in practice. In particular, it will allow you to return to our demand $q_i = D(p_i, y)$ curve without explicitly having to solve the utility maximization problem.

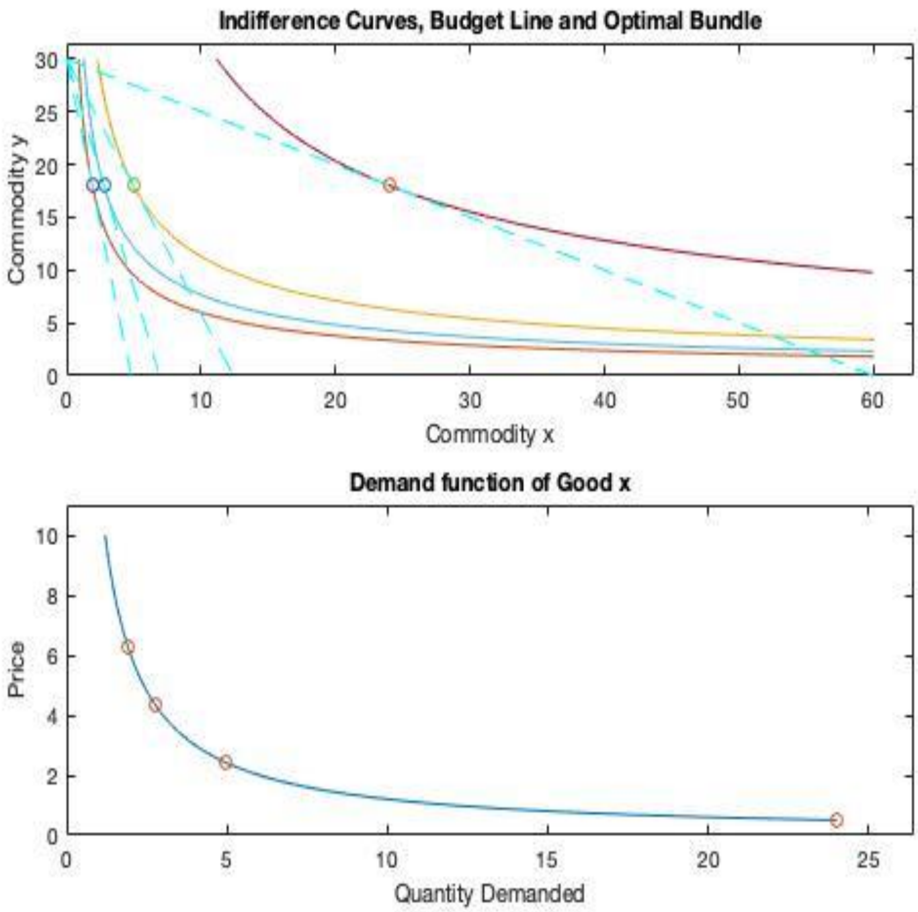
We can see the methodology graphically, starting with the consumer problem, maximizing a utility function (in this case a Cobb Douglas subject to budgetary restriction):



The graph shows the consumer problem. Starting with a Cobb Douglas Utility Function: $x^{0.4}y^{0.6}$ with prices of 2 goods x and y, Px = 5 , Py = 2 , and an income of 20, the highest utility achieved with that income is the optimal bundle x = 1.6 , y = 6 with an utility of 3.54.

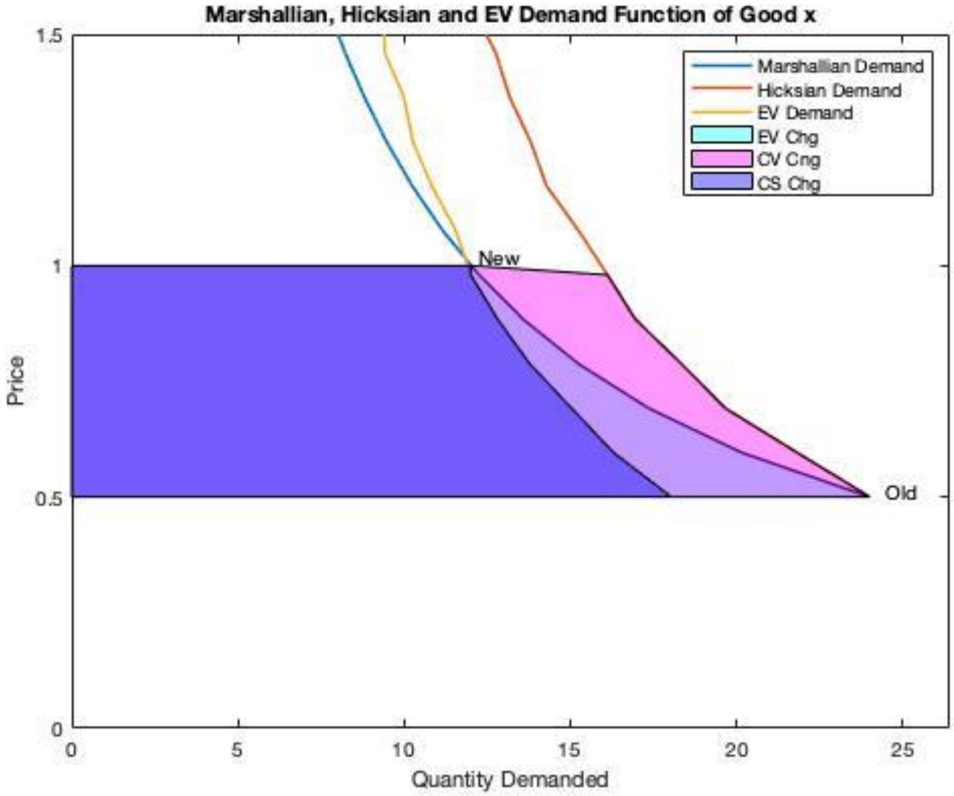
A price movement will change the quantity demanded, and so a demand function can be obtained, graphically[2]:

[2] Graphs built using Diagram Generator of Hang Qian (Iowa State University). Hang Qian (2020). Toolkit on Econometrics and Economics Teaching (https://www.mathworks.com/matlabcentral/fileexchange/32601-toolkit-on-econometrics-and-economics-teaching), MATLAB Central File Exchange. Retrieved December 28, 2020.



A demand curve is obtained as the consumer changes its optimal bundle of goods as prices changes. In this case, the starting position is Px = 0.5 , Py = 1 , Income = 30 with an initial bundle of x = 24 , y = 18 and an utility of 20.2. If the price of x increases to 1 with everything else constant Px = 1 , Py = 1 , Income = 30, the new bundle will be x = 12 , y = 18, with a lower utility of 15.3.

Finally the increase in prices reduces consumer surplus as shown in the graph below. Consumer surplus at initial bundle is 36 and at new bundle is 27.7 so the surplus reduction is 8.35. In the graph it is the area blue and magenta (trapezium with edges 0.5, 1, New, Old).



Demand Elasticity

In general, elasticity is a fundamental component of any competitive analysis as it provides a measurement of the consumer's response to a price increase. This parameter which represents demand sensibility to price changes is very relevant for firms when they set prices to maximize profits and play a central role in merger simulation.

The most useful measure of consumer sensitivity to changes in demand is demand-price elasticity[3], which measures demand sensitivity with respect to changes in prices of that good:

$$\eta_i = \frac{\%\Delta Q_j}{\%\Delta P_i} = \frac{\frac{\Delta Q_i}{Q_i}}{\frac{\Delta P_i}{P_i}}$$

Demand elasticity is the percentage change in quantity when prices increase in 1%. Marshall (1890) introduced the concept of elasticity and pointed out among his properties that he had no unit of measurement, unlike prices measured in a currency or the quantities to be measured in one unit of quantity per period.

For very small variations, demand elasticity can be expressed as the slope of the demand curve multiplied by the price-to-quantity ratio. Mathematically it can be written as the derivative of the price logarithm with respect to the logarithm of the demand curve:

[3] See USDA ERS database of demand elasticities-expenditure, income, own price, and cross price-for commodities and food products for over 100 countries in http://jaysonlusk.com/blog/2016/9/10/real-world-demand-curves.

$$\eta_i = \frac{P_i}{Q_i}\frac{\partial Q}{\partial P} = \frac{\partial Log Q_i}{\partial Log\, P_i}$$

Demand will be elastic at one point when elasticity is greater than one at absolute value, when the change in quantity will be greater than the price increase so revenue for a seller falls if the other parameters are maintained unchanged. A demand is inelastic if its elasticity is less than one at a certain price, and means that the seller can increase their revenue by increasing prices if all other parameters remain constant. Elasticity generally depends on the price level. Therefore, it makes no sense to talk about a product with elastic or inelastic demand but rather it would have to be said that it has an elastic or inelastic demand at a certain price or volume of sales. The elasticities calculated for aggregate demand are market elasticities for a given product.

Substitute and complementary goods

The cross price-elasticity of demand (2009) shows the effect of a change in the prices of another *good* k on the demand for good i. A higher price of good k can induce some consumers to substitute purchases of good k for good i. In this case, when the consumer increases their purchases of i when p_k increases, we will call i and k products as substitutes. For example two earphones from two different brands are substitutes if the demand for one falls with the fall in the price of the other because consumers replace the expensive one with the cheap one. Similarly, a fall in the price of air tickets will reduce the demand for train travel, keeping the price of train tickets constant.

In conclusion, we typically want to estimate the effect of prices on the quantity demanded. To do this, one has to build a demand model in order to estimate the impact of the change in price on the quantity demanded, as we described above. An important aspect of the demand function is its curvature and how it changes when we move along the curve. The curvature of demand determines the elasticity and therefore the impact of a change in price on the quantity demanded.

The simplest possible specification is a linear demand function such as $Q = \alpha - \beta P$ or its inverse: $P = \frac{\alpha}{\beta} - \frac{1}{\beta}Q$ where α and β are model parameters. This is the demand specification often appears in economics texts[4].

Taking into account disposable income we get the following functional form:

$$y_t = \alpha - \beta p_t + \gamma y_t$$

[4] The reader interested in the utility function that will result in a linear demand function will find it in http://www.its.caltech.edu/~kcborder/Notes/Demand4-Integrability.pdf

One of the earliest empirical demand estimation of the relation between price and consumer's demand is the paper written by Working on demand of agricultural products (1925), and the first to recognize its endogeneity (1927), meaning that when predicting the quantity demanded in equilibrium, the price is endogenous as producers change their price in response to demand and consumers change their demand in response to price. What this initial papers recognize is that price and quantity are related by both supply and demand curves which shift in response to non-price variables (right figure). In order to build a demand function it is necessary to assume constant a wide variety of non-price variables such as consumer preferences, incomes, prices of other goods, etc (left figure)[5].

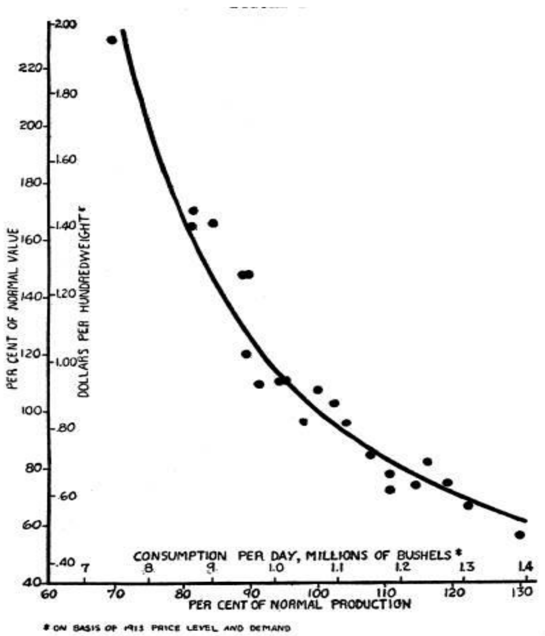
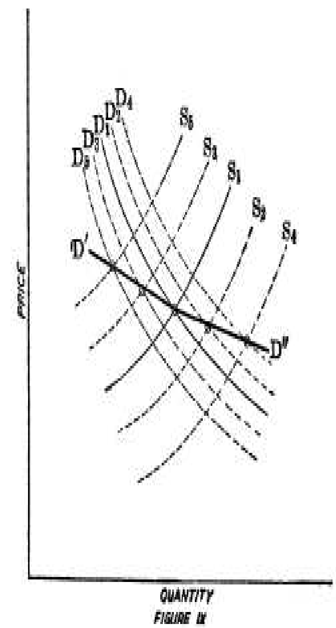


FIGURE 1. The Demand for Potatoes in the United States in Terms of Prices in St. Paul and Minneapolis



In the following section I introduce a structural model from the Industrial Organization literature which is a combination of a optimization agent problem with a statistical model. They solve some of the limitations of previous methodologies (endogeneity dimensionality, product

[5] See (2001): *"If the demand and supply curves shift over time, the observed data on quantities and prices reflect a set of equilibrium points on both curves. Consequently, an ordinary least squares regression of quantities on prices fails to identify—that is, trace out—either the supply or demand relationship. P.G. Wright (1928) confronted this issue in the seminal application of instrumental variables: estimating the elasticities of supply and demand for flaxseed, the source of linseed oil. Wright noted the difficulty of obtaining estimates of the elasticities of supply and demand from the relationship between price and quantity alone. He suggested (p. 312), however, that certain "curve shifters"—what we would now call instrumental variables—can be used to address the problem: "Such additional factors may be factors which (A) affect demand conditions without affecting cost conditions or which (B) affect cost conditions without affecting demand conditions." A variable he used for the demand curve shifter was the price of substitute goods, such as cottonseed, while a variable he used for the supply curve shifter was yield per acre, which can be thought of as primarily determined by the weather. ..."*

differentiation, etc.) at the cost of greater mathematical and statistical complexity. We need to study them as they have become standard in industrial organization and antitrust economics.

3. Structural Demand Estimation: Discrete Choice Random Coefficient Logit Model Differentiated Products.

As it is explained above linear demand models are easy to work and explain small dimension consumer demand problem. The traditional approach was based in specifying a demand function for each product as a function of its own price, prices of other goods and other variables, (for example, the Rotterdam model or Almost Ideal Demand Model)[6]. Models like the above ones will produce one demand equation for each product, but if we take into account heterogeneity and differentiated products, the system of equations and so the number of parameters to estimate can become easily intractable.

Differentiated products add some difficulties that can make demand estimation intractable as parameters are a square of the number of products in the estimated market. If we consider for example 50 demand equations for 50 products, it will imply 2.500 parameters to estimate (50 demand equations with 50 prices each). Economists have had to deal also with consumer heterogeneity which the traditional representative consumer demand theory based (1976) on was unable to explain.

Since 1980´s, industrial organization authors developed other methodologies capable of dealing with limitations of the traditional approximation, one example being the discrete choice model approach (2000). Products are seen as bundles of characteristics and consumers have preferences over those characteristics. Demand function maps a product in a space of characteristics such as quality, accessories, brand other than price.

The discrete choice random coefficient model developed in 1995 for automobile market (1995) solved many of the previous limitations, has been extensively upgraded and applied in marketing and industrial organization[7] and has become a standard IO model (Bresnahan, 1987). This model only needs as inputs aggregate market level price and quantity data for each product and allows endogeneity[8] of prices.

Berry, Levisnton and Pakes (BLP model) estimation of automobile market is based on observed characteristics of a car such as horsepower, miles per dollar, size and air conditioning. Power

[6] See (2019), for an application of Almost Ideal Demand System model of Deaton and Muelbauer (1980))
[7] See survey (2007).
[8] See Angrist and Krueger on instrumental variables to solve the endogeneity problem (2001) (prices correlated with econometric error term)

over weight and miles per dollar are measures of power and fuel efficiency while air conditioning would be a measure of luxury (at that time), size is also a measure of safety.
An application of BLP is provided by Nevo (2000) on cereal market using general data available for products such as average prices, aggregate quantities, product characteristics. Product characteristics used by Nevo are calories, sodium and fiber content. As it is shown below the model is more difficult to understand but is considerably more realistic.

The first equation of the BLP model is the consumer utility function. The (indirect) utility of consumer *i* when buying product *j* in month *t* is given by equation 1:

$$u_{ijt} = \alpha_i \left(y_i - p_{jt}\right) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}$$

Where,

i = 1, ..., m , j = 1, ..., n, t = 1, ..., r

Where y_i is consumer *i* income, p_{jt} is the observed price of product *j* in month t, x_{jt} vector containing 6 characteristics of *j* products in month *t,* ξ_{jt} unobserved characteristics of product j in month t, and ε_{ijt} unobserved stochastic disturbance with mean 0. ξ_{jt} attempts to capture unobserved or unquantifiable attributes or characteristics such as brand name or promotional activity which are essential to explain the data (2000). ε_{ijt} is a stochastic term added due to the inability to explain individual preferences in a complete and deterministic way.

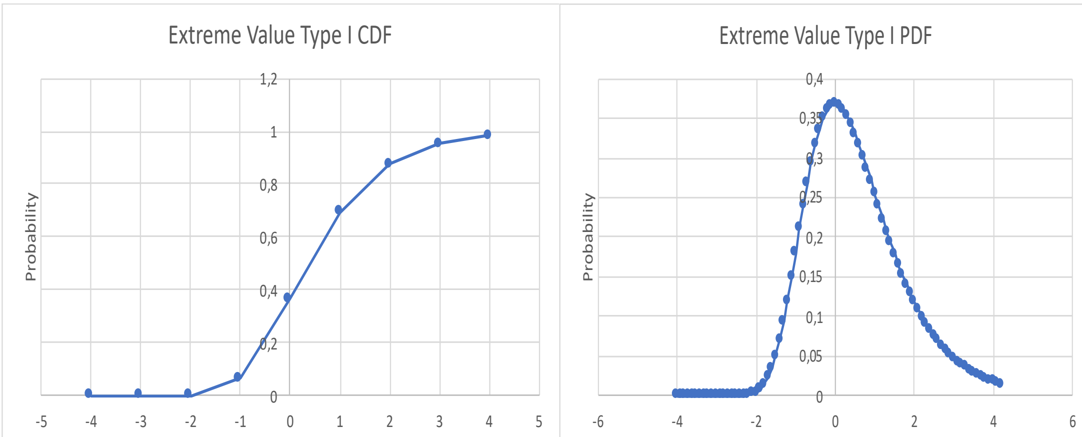
Parameters to estimate are now consumer i marginal utility to income, α_i, and marginal utility of each six characteristics, β_i.
Further it is needed to make assumptions on the stochastic disturbance. Each distributional assumption will generate different models. BLP assume that ε_{ijt} follows a Gumbel or (Generalized Extreme Value Distribution Type-I). For a mean zero and scale parameter one it has the density and cumulative distribution

$$F(x) = e^{-e^{(-x)}}$$

and probability density function

$$f(x) = e^{-x}\, e^{-e^{-x}}$$

This is the limiting distribution of the maximum value of a series of draws of independent identically distributed random variables. Next Figure 2 illustrates the density (PDF) which is standardly used in logit models because its cumulative distribution (CDF) is related to the probability of x being larger than any other of a number of draws, which is like the random utility from one choice being higher than that from a number of other choices (2006).



Consumer choice is to buy j product in month t if it yield him the highest utility. We do not have data at consumer *i* level, we only have market shares of *j* product and only a sample of consumer characteristics but not i´s marginal utility β_i.

There are two ways to estimate equation 1, one is to use the simple multinomial logit regression if we assume all consumer have the same tastes or marginal utility for characteristics. Another way is random-coefficient logit when we let fall this assumption.

Multinomial logit estimation

We can use a simple (multinomial because there are six choices) logit if we proceed assuming that consumers have identical preferences and aggregate them:

$$u_{jt} = \alpha\left(y - p_{jt}\right) + x_{jt}\beta + \xi_{jt} + \varepsilon_{jt}$$

$j = 1, \ldots, n,\ t = 1, \ldots, r$

We now assume that ε_{jt} follows a Gumbel or Type I extreme value distribution (with mean zero and scale parameter one), so we conclude that we are building a multinomial logit model. We can obtain the market share of product *j* through the probability of j having the greatest utility, which occurs if ε_{jt} is high enough relative to other disturbances[9]:

[9] See (2006) pg. 12-13 for an explanation of this equation.

$$s_{jt} = \frac{e^{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}}{1 + \sum_{k=1}^{n} e^{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}}$$

We can calculate the elasticities as the percentage change in the market share of product *j* when the price of product *k* goes up:

$$\eta_{jkt} = \frac{\%\Delta\, s_{jt}}{\%\Delta\, p_{kt}} = \frac{\delta s_{jt}}{\delta p_{kt}} \frac{p_{kt}}{s_{jt}} \begin{cases} \alpha p_{jt}(y_i - s_{jt}) \; if\; j = k \\ \alpha p_{kt} s_{kt} \;\; otherwise \end{cases}$$

As Rasmussen points out this model is unrealistic in at least two points. The first one is that for small market shares, $\alpha p_{jt}(y_i - s_{jt})$ is close to α so the elasticity is close to -αp_{jt} . This implies that the model delivers low elasticity for low prices, and this implies higher markups for products with low marginal cost when it is often the other way round (higher markups for higher marginal costs e.g. luxury cars)

The second point is that cross price elasticities $\alpha p_{kt} s_{kt}$ only depend on price and market share of product *k*. This means that if there is an increase of price of k, the model predicts that consumer will equally likely substitute it for the other substitute products, (see Nevo for an explanation of the blue bus/red bus example).

Random Coefficients Logit Model

An alternative to multinomial logit is random coefficients model where it is assumed that parameters (or marginal utilities of product characteristics) are different across consumers and are determined by consumer characteristics which are a function of fixed parameters that multiply observed characteristics and unobserved random characteristics:

$$u_{ijt} = \alpha_i y_i + \delta_{jt} + \epsilon_{ijt} + u_{ijt}$$

Where, δ_{jt} is the mean utility which is a component of utility from a consumer´s choice of product j that is the same across all consumers.

$\delta_{jt} = -\alpha p\text{-}p_{jt}+x_{jt}\beta + \xi_{jt}$

There is the heteroskedastic disturbance, u_{ijt}, and homoscedastic disturbance i.i.d., ϵ_{ijt}.

$$\binom{\alpha_i}{\beta_i} = \binom{\alpha}{\beta} + \Pi D_i + \sum v_i = \binom{\alpha}{\beta} + \binom{\Pi_\alpha}{\Pi_\beta} D_i + \binom{\sum \alpha}{\sum \beta} \sum(v_{i\alpha}|v_{i\beta})$$

$\boldsymbol{D_i}$ is a vector of consumer i observable characteristics, $\boldsymbol{v_i}$, is a vector of consumer unobservable characteristics; $\boldsymbol{\Pi}$ matrix of how parameters depend on consumer observables, Σ represent how those parameters depend on observables.

We have seen that the model deals with dimensionality, consumer heterogeneity and product differentiation. Furthermore, it takes into account price endogeneity or simultaneity when estimating demand and supply functions. For this reason an instrumental variable estimation is used in the BLP model. For every endogenous explanatory variable one has to find another variable correlated with that variable but uncorrelated with the error term.

An application of this methodology is provided by Nevo in the ready-to-eat cereal market. The data consists of quantity and prices for 24 brands of a differentiated product in 47 cities over 2 quarters. Product characteristics are sugar and mushy and some demographic variables are added such as log of income, income squared, age, child. The results of this estimation are shown below (2000). In the first column appear the marginal utilities (β) and show that the average consumer shows more preference for soggy cereal but it decreases with age and income, while the mean price coefficient is negative being less sensitive for children and wealthier consumers.

Table 1: Results full Model (see Nevo (2000)

			Dem. Vbles.			
	Mean	**Std Dev.**	**Income**	**Income^2**	**Age**	**Child**
Constant	*-1,87*	*0,38*	*3,09*	*0,00*	*11.859,00*	*0,00*
	(-0,2571)	*(0,1295)*	*(1,1962)*	----	*(1,0056)*	----
Price	*-32,43*	*1,85*	*1,66*	*-0,66*	*0,00*	*11,62*
	(7,748)	*(1,0811)*	*(172,9296)*	*(8,9871)*	----	*(5,1713)*
Sugar	*0,14*	*0,00*	*-0,19*	*0,00*	*0,03*	*0,00*
	(0,2571)	*(0,0123)*	*(0,0451)*	----	*(0,0371)*	----
Mushy	*0,83*	*0,08*	*1,47*	*0,00*	*-1,51*	*0,00*
	(0,0129)	*(0,2073)*	*(0,6957)*	----	*(1,0905)*	----

GMM objective: 14.9007

MD R-squared: 0.26471

MD weighted R-squared: 0.095502

run time (minutes): 0.67607

BLP method is more flexible and requires weaker assumptions than older models and has been widely used as it allows for many different types of firm and consumer behaviour at the cost of greater complexity. The model corrects endogeneity of prices with instrumental variables, reduces the number of parameters from a considerably high number of products to five or six characteristics and is grounded on consumer theory. Other methods used in the model such as GMM (Generalized Method of Moments) for estimation or the contraction mapping used for optimization fall out of the scope of this book.

4. Machine Learning for demand estimation

We have explained the traditional models that estimate consumer demand. Firms nowadays predict consumer demand using techniques based on computer science. I describe the main models described in Bajari et al. (2015). These techniques are possible now thanks to the availability of large data sets, greater processing speed and computer efficiency, easy access to cloud computing together with the development of computer science and statistics. Bajari et al. show that some of the computer science models predict more accurately consumer demand than traditional econometric models and are preferable to instrumental variable estimation as it is difficult to find plausible instrumental variables for all models. Varian (2014) considers that the focus of Machine Learning is finding some function that provides a good prediction of y as a function of x, most notably good out-of-sample predictions, penalizing models that are excessively complex, separating data into training, testing and validation data, and tuning parameters to produce the best out-of-sample predictions. Bajari compares machine learning model performance with standard linear regression. The models analyzed are stepwise regression, where the choice of predictive variables is automated adding the ones with highest correlation with th residual. Support Vector Machines (SVM) finds a function with a deviation no greater that ε for each data value and as flat as possible. LASSO (Least Absolute Shrinkage and Selection Operator) regression estimates parameters that minimize the sum of squared residuals plus a penalty term that penalizes models of bigger size (2014). Finally, regression trees were the data is split at several points for each independent variable, until the squared prediction error (error between the predicted values and actual values squared) falls under a threshold. Bajari demonstrates, concerning the former machine learning models, that they can produce superior predictive accuracy, or lower root mean squared error (RMSE), as compared to standard linear regression or logit models.

5. Conclusion.

This chapter provides a review of demand estimation methodologies, from the traditional estimation, to modern machine learning methodologies. As we mentioned before traditional methods were ill designed to deal with endogeneity, consumer heterogeneity and product differentiation. The demand characteristic approach of structural discrete choice methodology is focused on product differentiation and has become the standard methodology. Recently, easy access to massive data and computation capacity have smoothed the path for methods based on machine learning techniques. Some of these techniques have been outlined in this chapter. When compared to traditional econometric models ML techniques provide better prediction accuracy.

Bibliography

Ackerberg, D., Benkard, C. L., Berry , S., & Pakes, A. (2007). Econometric tools for analyzing market outcomes. In *Handbook of econometrics* (pp. 4171-4276). Elsevier.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives, 15*(4), 69-85.

Bajari, P., Nekipelov, D., & Ryan, S. P. (2015). Machine Learning Methods for Demand Estimation. *American Economic Review, 105*(5), 481-85.

Belloni, A., Chernozhukov, V., & Hansen , C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives, 28*(2), 1-23.

Berry, S., Levinston , J., & Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica, 63*(4), 841-890.

Bresnahan, T. F. (1987). Competition and Collusion in the American Automobile Industry: The 1955 Price War. *The Journal of Industrial Economics, 35*(4), 457-482.

Davis, P., & Garcés, E. (2009). *Quantitative Techniques for Competition and Antitrust Analysis.* Princeton University Press.

Deaton, A., & Muellbauer, J. (1980). An Almost Ideal Demand System. *American Economic Review, 70*(3), 312-326.

Marshall, A. (1890). *Principles Of Economics; an Introductory Volume.* London: Macmillan.

Mas-Collell, A., Whinston, M. D., & Green, J. (1995). *Microeconomic Theory.* Oxford University Press.

Merino Troncoso, C. (2019). *Market power and welfare loss* . Retrieved from mpra.ub.uni-muenchen.de.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of economics & management strategy, 9*(4), 513-548.

Rasmusen, & Rasmusen, E. (2006). *The BLP Method of Demand Curve Estimation in Industrial Organization.* Retrieved from http://www.rasmusen.org/papers/blp-rasmusen.pdf

Spence, A. (1976). Product Selection, Fixed Costs, and Monopolistic Competition. *Review of Economic Studies, 43*(2), 217-235.

Varian, H. (2014). Big data:New tricks of Econometrics. *journal of Economic Perspectives, 28*(2), 3-28.

Working, E. (1927). What Do Statistical "Demand Curves" Show? *The Quarterly Journal of Economics*, 212-235.

Working, H. (1925, August 1925). The Statistical Determination of Demand Curves. *The Quarterly Journal of Economics, 39*(4), 503-543.

CHAPTER 2: ESTIMATION OF A SUPPLY FUNCTION

In this chapter we provide a brief introduction of the production and supply function. The theoretical neoclassical model has limited capacity to explain data. Recent structural econometric techniques provide better estimation results and are able to explain entry, exit and investment decisions of firms.

1.- Introduction

In a similar way as in demand estimation, where we maximized utility functions, in this chapter we estimate a production function, productivity measures and cost functions. Technological efficiency is represented in the production function while economic efficiency is represented in the cost function. In this chapter we will replicate the main studies on production and cost function using standard software packages such as Excel or R.

Productivity estimates together with demand elasticity are key components of market structure studies. Traditionally productivity was estimated empirically using industry panels using a theoretical production function equation. These estimates were not consistent with data and did not explain market dynamics.

This chapter acknowledges how researchers have refined productivity estimation with the development of a broader dynamical structural approach in a similar way as in demand estimation. Firms decisions on investment strategies will depend on future expected profits than invest in research and development will have a higher probability of staying in the market while others with lower investment and lower productivity will have higher probability of exit. Firms will invest if they have expectations on future profits based on expectations and past achievements.

The chapter covers the dual approach of cost functions. Costs are a key component of the benefits and as such knowledge of the costs of a company or industry and the study of increasing returns of scale are fundamental to the analysis of competition. While the theoretical cost functions are generally known from the introductory economics courses, this chapter will also explain further research on increasing returns of scale and scope in industries where competition problems are common such as telecommunication and energy markets. The main concept here is economies of scale or the proportional increase in cost resulting from a small proportional increase in the level of output.

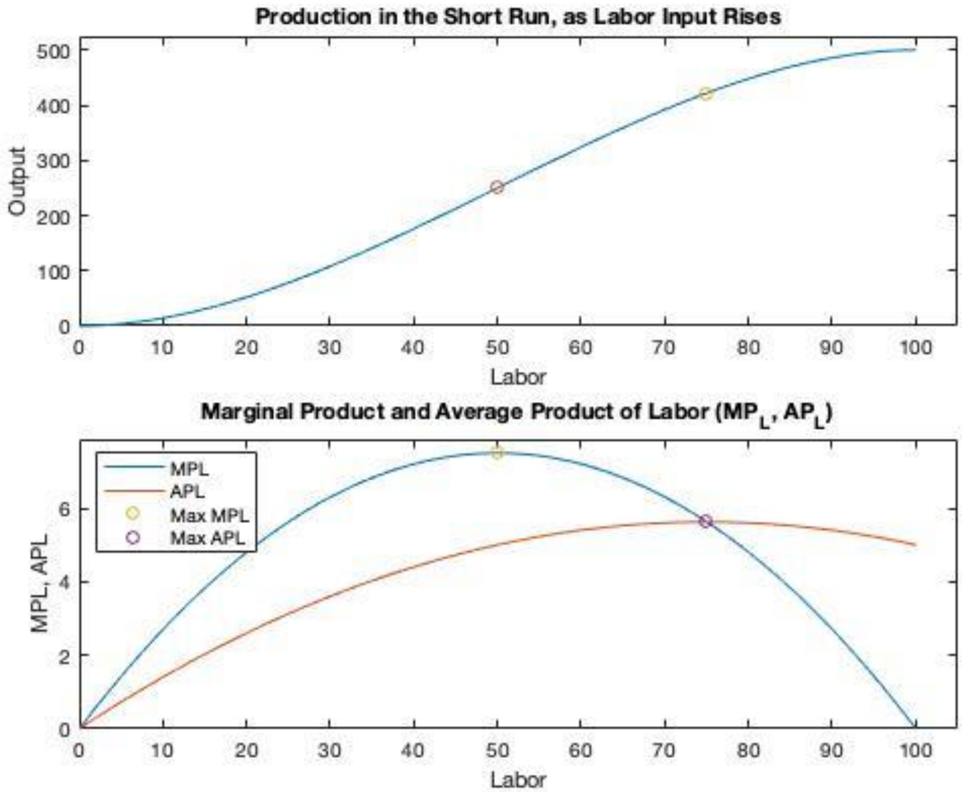
We begin with the traditional production function estimation, its endogeneity problems and further developments followed by cost function estimation developed by Viner (1932) and Sraffa (1926). Next scale economies and multiproduct scope economies are studied. Finally, the chapter explains other alternative methods that measure efficiency as a mere proportion of outputs and inputs while giving up production function determination. These methods are called frontier analysis approach and are divided into Stochastic Frontier Analysis (SFA) and Data Envelope Analysis (DEA).

2.- Production function estimation

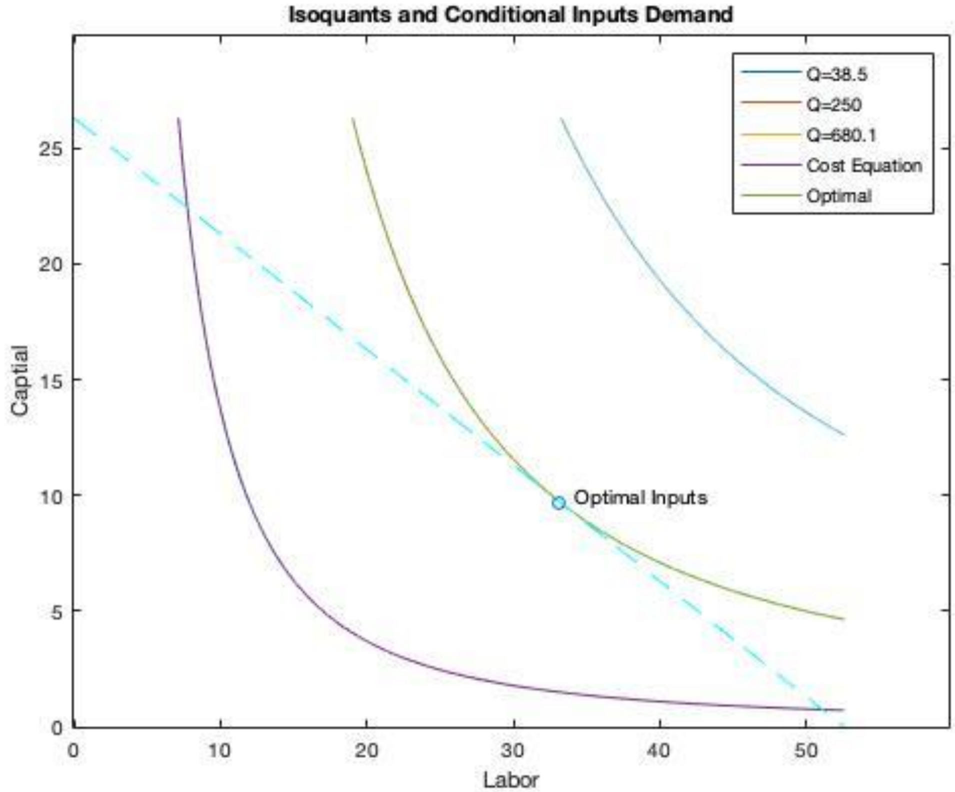
A production function is a mathematical function that relates the maximum quantity of output that can be obtained with inputs for example capital and labor (2007). Productivity is a measure of efficiency, the contribution of a particular input to the production. Syverson (2011) documented massive differences in total factor productivity (TFP) between industries in the US and stated that productivity is a matter of survival for business in a market. Duration, entry and exit in a particular market will most likely be related to a measure of efficiency or productivity of a firm. This is relevant when discussing market structure.

There are several economic properties of a production functions, namely nonnegativity, non-decreasing in inputs and concavity, that are generally expected and tested (for a complete description of assumptions of production functions in Microeconomics see Mas-Colell et al. (1995)). There are several mathematical functions that fit these assumptions, but Cobb-Douglas, Translog and Quadratic or Cubic production functions have been widely used in productivity analysis as we will see in this Chapter.

Classical graphs of the production function and average and marginal productivity are shown below assuming that only one input remains variable (e.g. Labor), while the rest are fixed. The graph represent a production function that only fulfills the general assumptions explained above in the segment of 50 and 100 of labor input. To the left of 50 labor units, the function is convex and to the right of 100, it is decreasing in labor, so the economically feasible region is only between 50 and 100. In the figure below average productivity AP (or output per unit of input) and marginal productivity MP (increase in output when input increases in one unit) are represented:



If we take into account two inputs the production function is generally represented by a family of isoquants which are combinations of two inputs (capital and labor) that are capable of producing an output level. There is one for each level of output and show higher level of output the farther they are from the origin. The expected shape of the production function implies convex isoquants with the below shape that can be easily minimized. The cost function will be the pairs of minimal cost for each unit of production. Unfortunately the theoretical model is too general and does not explain well specific production and cost data. Authors have dealt with different econometrical problems that surge when fitting easy functional forms to data.



As we mentioned, several mathematical functions that comply with minimal demanded assumptions have been fitted to production data and tested to comply with the abovementioned minimal assumptions.

A common function used in economics is Cobb-Douglas production function (see Ackerberg et al. (2007)):

$$Y = AL^{\beta_L}K^{\beta_K}$$

Where Y is output, K and L are capital and labor inputs, β are parameters and A is total factor productivity[10]. It is usually expressed as a linear equation in natural logarithms:

$$y = \beta_0 + \beta_k k + \beta_l l + \epsilon$$

β_0 is the mean efficiency level while ϵ are unobserved sources of differences such as managerial ability or technology differences between firms. Again, the problem if endogeneity appears when estimating this equation using simple OLS regression (see Econometric Appendix), as

[10] TFP - is the variation in output not explained by inputs depending on the functional approach, in this particular case the residuals or TFP would equal $\beta_0 + \epsilon$

acknowledged by Marschak et al. (1944) . The problem is that inputs K and L are not independent variables but are correlated with the unobserved error term ϵ. We face a problem of endogeneity if, for example, higher productivity companies, i.e. those with the highest unobserved productivity also demand a large number of inputs, see Davis and Garcés (2009), and a selected sample as only the most efficient firms would appear in the panel data, as the least efficient would have exited the market.

If we do not take into account the problem of endogeneity, our estimated coefficients on endogenous inputs would be overestimated. One possible solution is to use proxy instrumental variables, or variables correlated with the company's demand for input but not correlated with firm production.

Olley and Pakes (OP) (1996) suggested using investment as a proxy for productivity to control for endogeneity. We will see that this approach takes into account the decision process of a firm that will also be covered in chapter three, Market Structure and Dynamics. Furthermore, Levinsohn and Petrin (2003) refined the OP work.

OP designed an algorithm to simulate the decision process of an incumbent firm. Firms at the beginning of each period decide whether to exit or continue in the market. If they decide to exit they will receive a liquidation value of Φ, if they continue in the market they will choose inputs (labor, materials and energy) and a level of investment I_{it}. The sequence of decisions of a firm maximizing the expected discounted value of net future profits is shown in the following Bellman equation[11] :

$$V(k_{jt}, a_{jt}, w_{jt}, \Delta_t) = \max\{\phi(k_{jt}, a_{jt}, w_{jt}, \Delta_t), max\{\pi(k_{jt}, a_{jt}, w_{jt}, \Delta_t) - c(i_{jt}, \Delta_t) + \delta E[V(k_{jt+1}, a_{jt+1}, w_{jt+1}, \Delta_{t+1}) \vdots k_{jt}, a_{jt}, w_{jt}, \Delta_t, i_{jt}]\}$$

This equation describes the dynamic decision process of a firm. First, the firm decides to exit a market if its liquidation value, $\phi(\,)$, exceeds the expected discounted value of net future profits. Second, it decides on investing or not i_{jt}, which is the solution to the second maximization bracket, where π is the profit function and C is the cost of investment, δ is the discount function

[11] See appendix for an explanation of the Bellman equation.

and $E\ (\ )$ is the firm expectation conditional on information at t. Firms with positive productivity shocks in t will invest more in that period t. OP derive the unobserved productivity shock as:

$$\Omega_{it} = h(i_{it}, k_{it}, a_{it})$$

Firms will invest in the future if there is an increase in current productivity. OP derive the following Cobb-Douglas production function:

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \beta_a a_{it} + u_{it}$$

whereby,

$$u_{it} = \Omega_{it} + \eta_{it}$$

Substituting Ω_{it} and u_{it} in the production function, we obtain:

$$y_{it} = \beta_l l_{it} + \beta_m m_{it} + \beta_a e_{it} + (\beta_0 + \beta_k k_{it} + \beta_a a_{it} + h(i_{it}, k_{it}, a_{it})) + \eta_{it}$$

Where y is output, k and l are capital and labor inputs, β are parameters, a is age of the firm, all in logs, Ω_{it} is a productivity shock observed by the firm and η_{it} unobserved shocks. This specification takes into account the relation between inputs and Ω_{it} and controls for unobserved productivity, while traditional models based on OLS estimates will be biased upwards, this specification can be estimated with OLS without bias. With this development estimation of parameters is more accurate and a structural explanation of entry and exit of firms in a market.

The tables below show a dataset of 2544 Chilean firms consisting of value added, capital, labor, electricity, water, investment between 1986 and 1996 (2018). We will estimate parameters of the production function using the OP estimation technique on that dataset[12]:

Chilean firm-level production data 1986-1996

[12]See Rovigatti (2017).

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Y	2,544	13.003	1.463	8.770	11.965	13.884	18.580
Capital	2,544	11.668	2.157	2.708	10.498	13.049	18.217
Skilled Labor	2,544	1.770	1.358	0	0.7	2.7	7
Un. Labor	2,544	1.680	1.448	0	0	2.7	6
Elec	2,544	4.109	1.546	0.000	3.045	4.997	9.944
Water	2,544	11.137	2.333	1.072	9.855	12.631	18.088
Inv	2,544	11.137	2.333	1.072	9.855	12.631	18.088
idvar	2,544	15,770.490	7,355.940	10,007	11,497	16,337	40,475
timevar	2,544	2,001.042	3.223	1,996	1,998	2,004	2,006

The resulting parameters as expected are biased upward when using OLS without controlling for simultaneity and selection bias as shown in the table below (columns OLS and OP). Levinsohn and Petrin (2003) suggest an alternative approach to control for missing data (column LEVPET) Both Olley – Pakes and Levinson and Petrin show lower value of parameters. These parameter values show the marginal variation of output with an increase in one unit of input, e.g. an increase in one unit of capital increases output in 0.485 units. Recall that the negative value of water parameter does not comply with theoretical assumptions described above.

OLS, OP and LP methods: Chilean dataset

	OLS	OP	LEVPET
K	0.485^{***}	0.168	0.162^{***}
	(0.050)	(0.029)	(3.95)
Skilled	0.453^{***}	0.314	0.319^{***}
	(0.014)	(0.03)	(8.78)
Unskilled	0.362^{***}	0.256	0.258^{***}
	(0.013)	(0.027)	(9.79)
Water	-0.159^{***}	0.311	
	(0.046)	(0.208)	
_cons	7.625^{***}		
	(0.109)		
N	2544	2544	2544

SE in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Residual standard error: 0.7773 on 2539 degrees of freedom

Multiple R-squared: 0.7181, *Adjusted R-squared: 0.7177*

F-statistic: 1617 on 4 and 2539 DF, p-value: < 2.2e-16

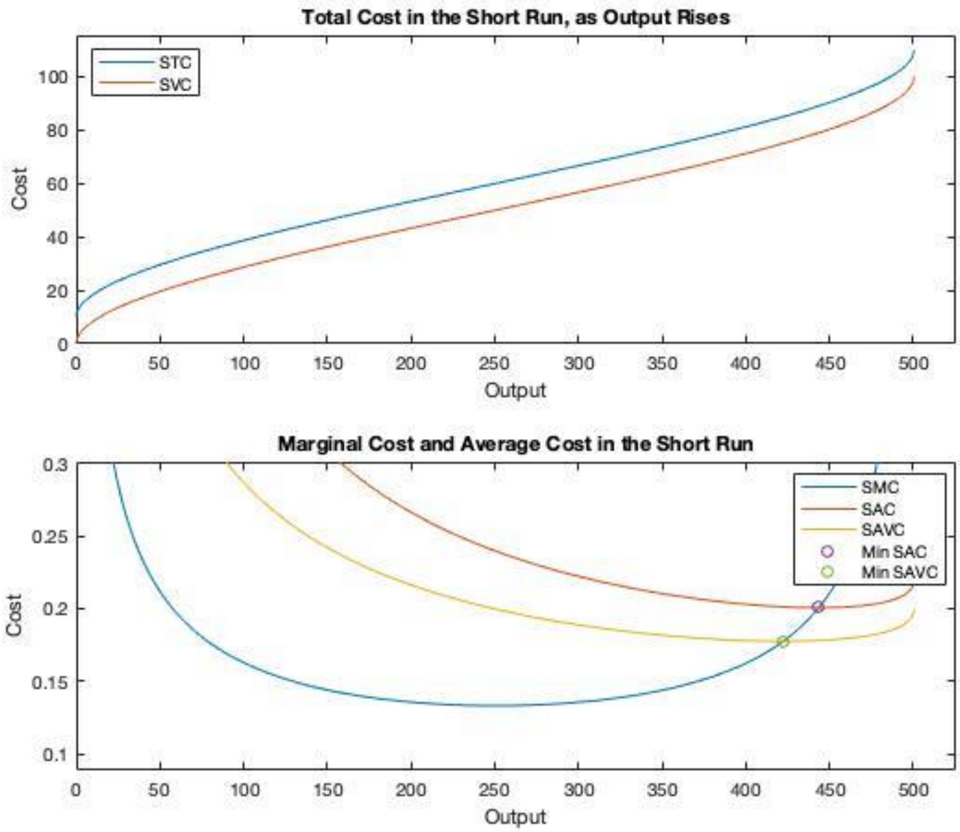
In conclusion, the naïve traditional method to estimate production functions was incorrect, several authors designed new methodologies that provide consistent estimations of firm productivity.

3.- Cost function estimation

If we take factor prices as given and multiplying, then by the quantities of production factors a cost equation is obtained. The cost function gives the minimal cost foreach level of production. The cost function has also general properties as the production function such as nonnegativity, nondecreasing in factor prices, non-decreasing in output, homogeneous and concave in input prices. Recall that through Shepard´s lemma (1953) deriving cost function with respect to price of factors one can obtain input demand functions and with them obtain the production function.

Viner developed a graphic representation of cost functions (1932). A cost function that complies with the minimum properties is represented below (STC are short term total cost and SVC short term variable cost). In the second graph below average and marginal cost are represented. Short term average cost function (SAC) (and equivalently variable cost curve SAVC) are unitary costs or total cost divided by total output and has a U-shape following production function curvature and fixed factor prices. Average cost decreases as output increases until the average cost is at the minimum. After this the law of diminishing returns makes average and marginal cost to rise. Marginal cost (SMC) is the derivative of the production function with respect to output or the infinitesimal change in cost when production increases in one unit. When prices are given in perfect competition, the part of marginal cost curve that is positively sloped (above min SAC) represents the short-term supply of the individual firm.

Short term means a period when some inputs remain fixed. Economies of scale surge when all factors are variable and an equivalent increase in all factors generates a more than proportionate increase in production. In the long term all factors are variable, technological change and entry and exit can affect cost functions. The minimum average cost is known as Minimum Efficient Scale (MES) is the optimal level of output to produce (Greer, 2010). We will see below a classic example of economies of scale known as natural monopoly, where average costs are always declining so efficient production is better if it is concentrated in a single firm.



Economies of scale

The classic work on scale economies is Nerlove (1961), that studies power generation in the US using data of 145 firms in 1955. The sample collected consisted of cost data, fuel prices, labor and capital prices and total production of electricity for each electricity generation plant. The summary statistics of Nerlove´s data is[13] shown below:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
cost	145	12.976	19.795	0.082	2.382	14.132	139.422
output	145	2,133.083	2,931.942	2	279	2,507	16,719
labor	145	1.972	0.237	1.450	1.760	2.190	2.320
laborshare	145	0.107	0.046	0.039	0.076	0.130	0.316
capital	145	174.497	18.209	138	162	183	233
capitalshare	145	0.427	0.115	0.090	0.351	0.486	0.845
fuel	145	26.177	7.876	10.300	21.300	32.200	42.800
fuelshare	145	0.467	0.123	0.070	0.394	0.544	0.759

Nerlove fitted the data to a Cobb-Douglas cost function derived from the following production function with three inputs: capital (K), labor (L) and fuel (F):

[13]Table built using Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

$$Y = AL^{\beta_L}K^{\beta_K}F^{\beta_F}u$$

From this production function he obtained a cost function in logs in terms of units of fuel as follows:

$$Ln\left(\frac{C}{P_F}\right) = \beta_0 + \beta_y LnY + \beta_L Ln\left(\frac{P_L}{P_F}\right) + \beta_K Ln\left(\frac{P_K}{P_F}\right) + \varepsilon$$

Where Y measures billion Kwatt hour of electricity produced, P_K price of capital, P_L price of labor (dollars per hour) and P_F is the fuel price in cents of dollar per million BTUs.

Nerlove estimated the model using traditional OLS regression using cost and input prices for 145 companies. Results are shown in the table below were Nerlove work is replicated using his original data:

Table: Regression results Nerlove (1963)

	Dependent variable:
	log(cost/fuel)
log(output)	0.721***
	(0.017)
log(labor/fuel)	0.594***
	(0.205)
log(capital/fuel)	−0.008
	(0.191)
Constant	−4.686***
	(0.885)
Observations	145
R^2	0.932
Adjusted R^2	0.930
Residual Std. Error	0.392 (df = 141)
F Statistic	640.096*** (df = 3; 141)
Note:	*p<0.1; **p<0.05; ***p<0.01

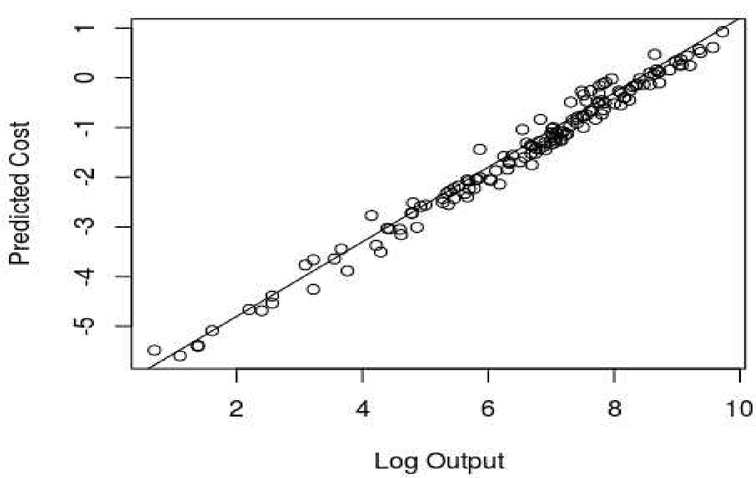
From the above theoretical curves it is expected that the parameter of output should be positive as well as input price parameters (an increase in output or input price parameters should always increase cost). Indeed, β_y, β_L parameters are positive and significant, the former means that an increase in 1% in output yields a 0.72% increase in total cost, considering all other variables to remain fixed (2010).

Nerlove rejects the hypothesis of constant returns of scale as the inverse of the log Y parameter[14] $(0.72)^{-1} = 1.39$ is greater than one. This proves that power generating plants have increasing returns of scale (positive economies of scale). The scale parameter is also the ratio of marginal to average cost:

$$\frac{\partial lnC}{\partial lnQ} = \frac{Q}{C}\frac{\partial C}{\partial Q} = \frac{MC}{AC}$$

This ratio is positive/negative for increasing/decreasing returns to scale. Nerlove also divided the data in 5 groups according to the size of the firms. In the five regressions returns to scale were lower as the size of the firm increased.

The following figure shows the Nerlove data (in log), the estimated costs based on the output level and below the estimated residuals as a difference between the estimated and the actual values. For the estimate to be consistent under OLS, the residuals need to have no dependence on explanatory variables. On the contrary, the graph shows that they depend on the output level which violates the requirement for consistent estimates. At low and high levels of production, residuals are positive so the true cost is higher than the estimated values. On the other hand, for output intermediate values, the true value of the costs is better than the estimate. The graph of the residuals reflects a U-shape, see (2009) and (2018).



[14] Because $r = \sum_i \beta_i$, see (2010)

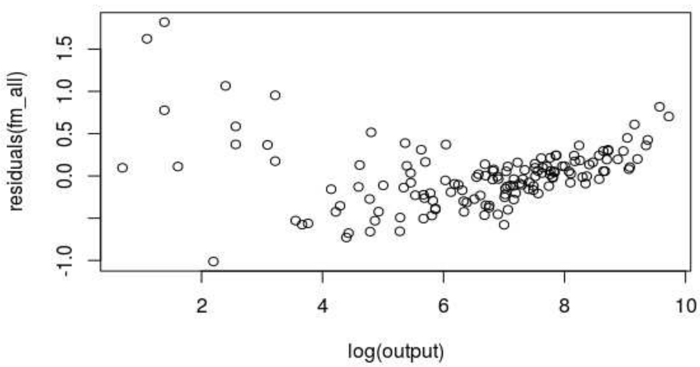


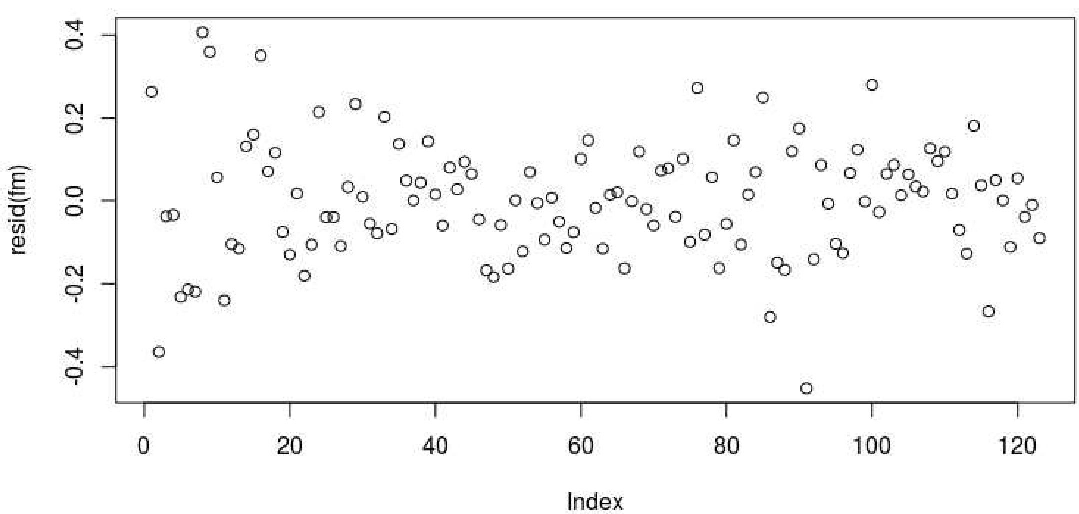
Figure: Nerlove graphs on fitted Log Cost and regression residuals

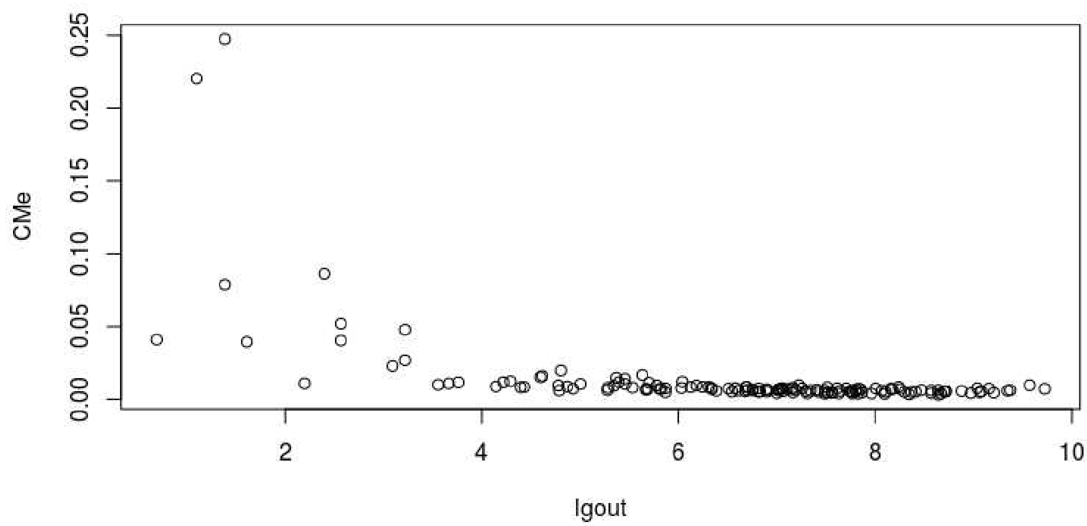
This diagnosis suggests that the shape of the cost function is incorrect, high residuals at low and high levels of output reveal an incorrect specification of the production function which should have a U type of cost function. The plot and data prove that there are increasing scale economies that are exhausted at a certain output level from which declining scale economies begin. Nerlove suggests that the specification can be corrected by expanding the function with a quadratic term. This generates a more flexible cost function that will allow cost to vary with the output level in a way that can generate economies of scale followed by diseconomies of scale as the level of production increases.

The table and figure below shows the Nerlove diagnostic test which consists of opposing the residues against the explanatory variable on the same chart. In this case using a Translog cost function, and unlike the previous specification, the graph shows that the expected value of the residuals in this regression is independent of the output level and remains around 0 as required by the MCO method. The table shows significant coefficients for output and labor.

Nerlove Translog Cost and regression residuals

	Dependent variable:
	log(cost/fuel)
log(output)	0.153**
	(0.062)
I(log(output)^2)	0.051***
	(0.005)
log(labor/fuel)	0.481***
	(0.161)
log(capital/fuel)	0.074
	(0.150)
Constant	−3.764***
	(0.702)
Observations	145
R^2	0.958
Adjusted R^2	0.957
Residual Std. Error	0.308 (df = 140)
F Statistic	800.667*** (df = 4; 140)
Note:	*p<0.1; **p<0.05; ***p<0.01





Christensen and Greene (1976) estimated the same cost function adjusted to the 1955 data by adding the 1970 data, using several models. Their model shows that the majority of firms were producing well beyond the efficient scale or with diseconomies of scale. These authors provided several models for data between 1955 and 1970 demonstrated the impact of technological progress over time moving the average cost function downwards and reducing de average cost of producing electricity.

Nerlove used a limited panel data and computing power available at that time and was strongly based on factor price taker assumption. Furthermore, other elements including long term contracts on fuel and trade unions regarding labor should be included in the study.

Multi-product firms and economies of scale

(1984) conducted an empirical estimate of AT&T's cost function and economies of scale and scope in local and long-distance calls. this study was relevant for the decision of the U.S. government in 1982 to break AT&T into several local firms while leaving AT&T in the long-distance call market (2009). The allegation of AT&T against it was the efficiencies gained from managing all telecommunications services in a single company would be lost if it was divided by region and activities. Evans and Heckman proved empirically that they were wrong. They used a "subaditivity" test for the cost function, a property that implies that the cost is lower when performed by a firm that by several firms. The failure of this test would mean that a single firm is more inefficient that several firms.

To do this they defined a cost equation for two products:

$C = C(q_L, q_T, r, m, w, t)$,

Where q_L the output level of local *L* calls is, and q_T is the output level of long-distance T calls. Cost functions depend on the price of production factors: *r* is the return rate of capital, w is the wage rate, and *m* is the price of the materials, *t* is a variable that captures technical change.

These authors used a translog function with two products and three production factors:

$$\log C = \alpha_0 + \sum_i \alpha_i \, logp_i + \sum_i \beta_i \, logq_i + \frac{1}{2}\sum_i \sum_j \gamma_{ij} logp_i logp_j +$$
$$\frac{1}{2}\sum_k \sum_j \delta_{kj} log\ q_k \log q_j + \sum_i \sum_k \rho_{ik} log\ q_k \log\ p_i + \sum_i \lambda_i \log t \log p_i + \sum_k \theta_k \log t +$$
$$\tau \log(t)^2 + \mu \log t$$

This cost function is much more general than that used at first by Nerlove based on a Cobb-Douglas function. It is more flexible and can approach any cost function. Let's define it for two inputs and two outputs.

The Translog cost function is presented in an unrestricted way but in the estimate a number of restrictions on the cost functions suggested by the theory are imposed. Authors impose homogeneity in input prices and symmetry on input prices. Using Shepard´s Lemma[15], one can obtain the three inputs (i=1,2,3) participation equations:

$$s_i = \alpha_i + \sum_j \gamma_{ij} logp_j + \sum_k \rho_{ik} log\ q_k \sum_i \lambda_i \log t$$

The estimate parameters of these equations are shown in the table below. A SURE estimator (seemingly unrelated regressions) was used (2015).

Selected estimated results Translog function with 31 observations

[15] The derivatives of the cost function with respect to input prices are the input demand functions. $l_j = \frac{\P C(q_1, q_2, p_1, p_2, p_3, t)}{\P p_j}$

So the participation of input j in total costs would then be:

$$s_j = \frac{p_j l_j}{C} = \frac{p_j}{C} \frac{\P \ln C}{\P \ln p_j} = \frac{\P \ln C}{\P \ln p_j}$$

By applying Shepard's lemma to the cost function we get the participation equations.

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*	
Constant	*9.0542*	*0.0044524*	*20.335.750*	*< 2.2e-16*	***
Capital	*0.654*	*0.1555707*	*42.075*	*0.0018071*	**
Labour	*0.354*	*0.1414821*	*51.771*	*0.0004148*	***
Local (output)	*0.226*	*0.2221301*	*10.197*	*0.3319212*	
Toll (Output)	*0.504*	*0.5275837*	*-0.4318*	*0.6750634*	
Technology	*-0.201*	*0.0780344*	*-0.1255*	*0.9025945*	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.9999439 Adjusted R-squared: 0.9998318

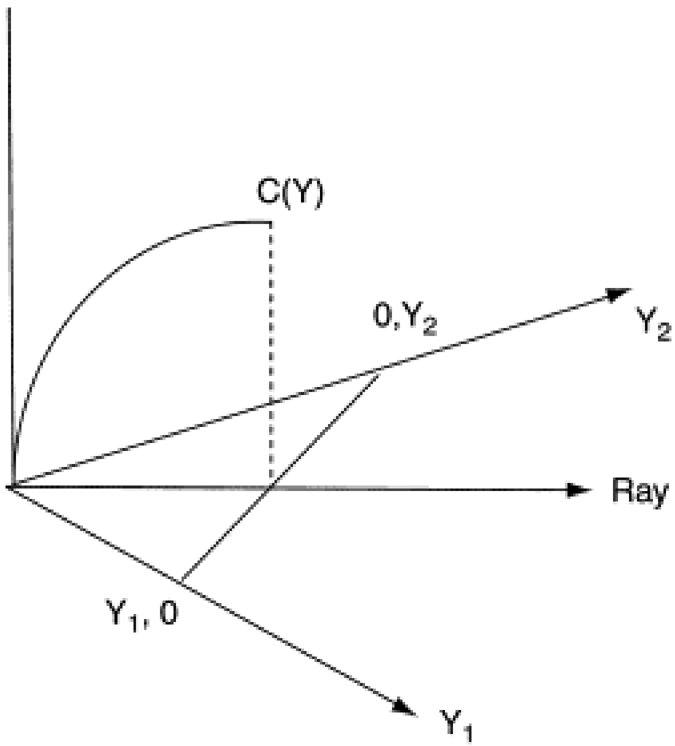
(Wales, 1987) proved that the function used was concave in input prices and monotonic but fails to comply non decreasing in output.

Subaditivity Test

Heckman and Evans provided a test to prove if Bell was a natural monopoly (1984). The traditional approach to natural monopoly was to evaluate if the company was at a level of output where average cost was decreasing. In case of multiproduct companies like a telephone firm (at that moment the two products were local and long call), Baumol (1977) considered that the necessary and sufficient condition in order to identify a industry as a natural monopoly was that the cost function must be sub additive. This test compares the cost of producing several products separately from the cost of producing them together. There is subbaditivity if it is cheaper to concentrate production in a firm than dividing it in several firms, so the relevant question is whether it costs more to produce the total output with multiple firms compared to the case where a single firm produces everything. If it costs more to distribute production among several companies, then we have that the firm (in this case AT&T) is a natural monopoly even though it produces several products (1977). The test is summarized by Baumol as:

$$C\left(\sum_i q_i\right) < \sum_i C(q_i)$$

Graphically, scale economies in multiproduct firms are represented graphically, when outputs increase proportionately as the gradient of the the perpendicular ray from the origin. The degree of scale economies may be interpreted as a measure of the percentage rate of decline or increase in ray average cost with respect to output (Baumol et al., 1982).



Evans and Heckman used a trans log production function with limited dataset to test for local subadditivity. Other authors such as Roller (1990b) contradict this Evans and Heckman and consider that Bell was a natural monopoly before splitting, and argued that the production function did not comply with general properties such as positive marginal productivity of production factors and was not suitable to explain technological progress.

Frontier Analysis

Another approach to production function approach is frontier analysis, which estimates a measure of efficiency and productivity of outcomes and has been extensively used to assess outcomes in utilities, banks, hospitals, etc. (2012). These methods consider the cost and production functions as "ideal" or "frontier" to be estimated.

Instead of using a mathematical function, this analysis only requires input and output data and delivers efficiency ratios for each firm in the data, so that one can compare branches, production units or firms. The most efficient units will form a frontier line below which less efficient units will appear. Efficiency is measured as the ratio of the sum of outputs produced by each unit divided by the sum of inputs spent in the production process.

Charnes et. al. (1978) estimated for the first time a production frontier technology analysis based on a previous paper of Farrel (1957). The data consists of seventy school sites with the following variables[16]:

Firm - school site number

Inputs

x1 - education level of the mother

x2- highest occupation of a family member

x3- parental visits to school

x4- time spent with children in school-related topics

x5 - the number of teachers at the site

Outputs

y1- reading score

y2- math score

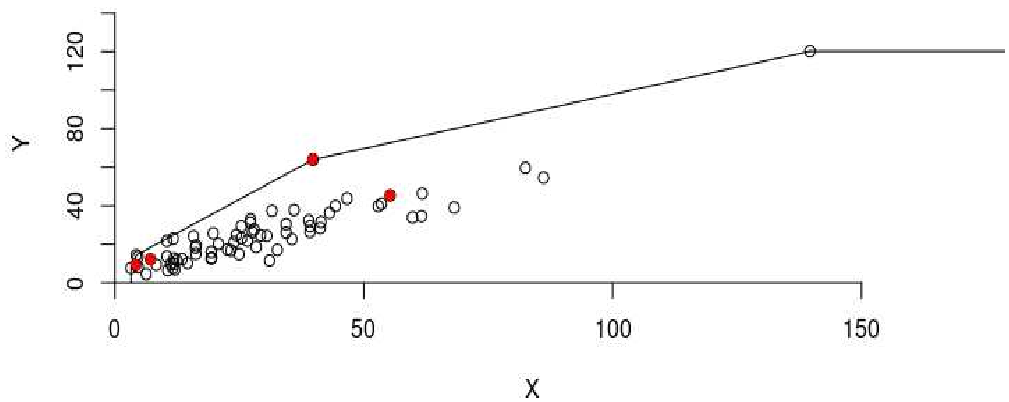
y3- self–esteem score

pft =1 if in program (program follow through) and =0 if not in program

name- Site name

The basic model of one input and one output is shown below and considers the maximum or border output that can be produced for each input level.

[16] See Benckmarking Package of Rstudio



Each point represents a school, and the border has been drawn to find a frontier that encompasses all the data. There are some technically efficient schools at the border, and others below the border that can improve their efficiency. Efficiency is measured through a ratio, OY/OX, so that a call center operating at the frontier had a technical efficiency of 100% while those within the border operate at a level below the efficiency level. A firm is defined as combinations of inputs and outputs weighted by their v_i weight. The optimization program then states that we must increase the output level of the firm k as much as possible subject to the requirement that we can find the smallest firm that could have currently produced that higher level of production given the current combinations of inputs and outputs observed in the data.

The advantage of the DEA is that no functional form of the cost or production frontier should be assumed, while its critics consider that it relies too much on data and its sensitivity to outliers can be troubling. On the other hand, this method avoids imposing a specific parametric function on the production and cost function[17].

Conclusion

This chapter has reviewed production function analysis and cost functions, efficiency and productivity. Authors searched for a functional form that explained production data and

[17] **Stochastic Frontier Models: SFA** are parametric models (1992) where econometric theory is used to estimate pre-specified functional form and inefficiency is modeled as an additional stochastic term.(see Methodology and Applications of Stochastic Frontier Analysis, Andrea Furková)

complied with the minimum economic assumptions such as monotonicity, homogeneity, and positive marginal productivity of factors. Newer techniques namely frontier analysis,

have opened possibilities for use in cases which have been resistant to other approaches due to complex (often unknown) nature of the relations between the multiple metrics labeled as inputs and outputs.

Bibliography

A. Charnes, W. C. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429-444.

Ackerberg, L. B. (2007). Econometric Tools for Analyzing Market Outcomes. In L. E. Heckman J, *The Handbook of Econometrics* (Vol. 6A Chapter 63, pp. 4171-4276). Amsterdam: North-Holland.

Battese, G. E. (1992). Frontier production functions and technical efficiency: a survey of empirical applications in agricultural economics. *Agricultural Economics, 7*(3-4), 185-208.

BAUMOL, W. J. (1977). Weak invisible hand theorem on the sustainability of multiproduct natural monopoly. *American Economic Review, 60*, 350 -365.

Christensen, L. R., & Greene, W. H. (1976, August). Economies of Scale in U.S. Electric Power Generation. *The Journal of Political Economy, 84*(4), 655-672.

Davis, P., & Garcés, E. (2009). *Quantitative Techniques for Competition and Antitrust Analysis.* Princeton University Press.

Evans, D. S. (1984). Test for Subadditivity of the Cost Function with an Application to the Bell System. *The American Economic Review, 74*(4), 615-623.

Farrell, M. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society, 120*(3), 253-290.

Greene, W. H. (2018). *Econometric Analysis, 8th Edition.* Stern School of Business, New York University.

Greer, M. (2010). *Electricity Cost Modelling Calculations.* Academic Press.

Henningsen, A. (2015). *Introduction to Econometric Production Analysis with R. Collection of Lecture Notes. Department of Food and Resource Economics, University of Copenhagen. Available at http://leanpub.com/ProdEconR/ . 2.* Copenhaguen: Department of Food and Resource Economics, University of Copenhagen. Available at http://leanpub.com/ProdEconR/ .

James Levinsohn, A. P. (2003, Abril). Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies, 70*(2), 317-341.

Marschak, J. &. (1944, Jul-Oct.). Random Simultaneous Equations and the Theory of Production. *Econometrica, 12*(3/4), 143-205.

Mas-Collell, A., Whinston, M. D., & Green, J. (1995). *Microeconomic Theory.* Oxford University Press.

Nerlove, M. (1961). *Returns to scale in electricity supply.* nstitute for mathematical studies in the social sciences.

Olley, G. S. (1996). he Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica, 64*(6), 1263-1297.

Röller, L.-H. (1990b). Proper Quadratic Cost Functions with an Application to the Bell System.". *The REview of Economics and Statistics, 72*, 202–210.

Rovigatti, G. (2017). Retrieved from prodest: Production Function Estimation in R. R package.: https://cran.r-project.org/web/packages/prodest/index.html

Rovigatti, G. a. (2018). Theory and practice of total-factor productivity estimation: The control function approach using Stata. *The Stata Journal, 18*(3), 618-662.

Sheppard, R. W. (1953). *Theory of Cost and Production Functions.* Princeton University Press.

Sraffa, P. (1926). The laws of returns under competitive conditions. *The Economic Journal* , 535-550.

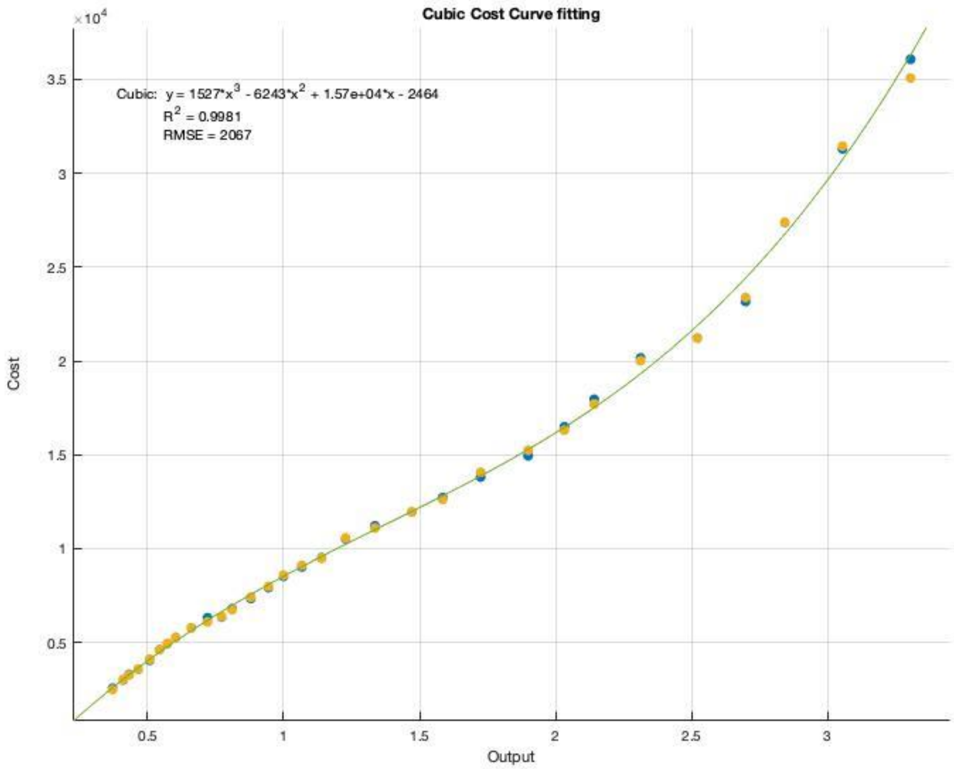
Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature, 49*(2), 326-365.

Viner, J. (1932). Cost curves and supply curves. *Zeitschrift für Nationalökonomie, 3*(1), 23-46.

Wales, W. E. (1987). Flexible Functional Forms and Global Curvature Conditions. *Econometrica, 55*, 43-68.

Worthington, A. C. (2012). *A review of frontier approaches to efficiency and productivity measurement.* Griffith University, Department of Accounting, Finance and Economics, Queensland.

Annex: Alternative Fitting of Evans Heckman data in a Cubic Production Function



CHAPTER 3: MARKET STRUCTURE

1.- Introduction

Market structure can be understood as the number of companies, size and products and role in determining competition in the market. Specifically, it studies how the number, organization, size of companies, potential competitors and the number of products affect business benefits and competition. This research has shaped the views of antitrust and regulatory authorities overseeing market structure and competition policy. For example, antitrust authorities seek to answer the following question: how many companies are necessary for competition to exist in this market? Or can investment in R&D, advertising or capacity slow down and reduce competition? Companies are also very interested in knowing how many companies a market can sustain.

There has been an evolution in the thinking of these economists in particular in the relationship between market structure and competition. In the 1950s and 1970s it was examined how variables such as business profits, advertising, R&D and prices differ in concentrated and low-concentrated markets. It was based on the basis that the market structure was exogenous.

Bain (1956) used it to explain whether a market was concentrated using various variables such as technological differences and differences in product markets. In the 1970s and 1980s, more influence was made of how the strategic behavior of companies could influence the market structure. The works started from considering a game on two levels, in the first the potential starters decided whether to enter or not, in the second the entrants competed in several dimensions. The increased availability of statistics allowed to know data at the enterprise level about business results and market structure.

Let's describe how economists have used game theory to build structural econometric models of entry, exit and market merger. Predictions about market structure depend on observable and non-observable variables: substitutability, entry costs and sunken costs, sensitivity of business profits to the entry and exit of competitors into the market, product replaceability, expectations of potential entrants to competition that they would once face in the market, prevalence and efficiency of potential entrants and endogeneity of fixed and sunk costs.

I begin by summarizing the econometric framework for analyzing cross-sectional data on the number and size of companies in different but related markets. This framework treats the number and identity of companies as endogenous results of a two-stage oligopolistic game. In the first stage they decide whether to operate and the characteristics of the product and in the second stage compete. This framework allows us to analyze the nature of competition in the market and the sources of business profits.

The construction of structural models allows estimating non-observable variables. We generally lack cost data so it is necessary to make inferences from prices and quantities and number of companies on demand, variable and fixed costs. Strong assumptions have to be made to do this.

2.- Bresnahan Reiss model

Bresnahan and Reiss (1991) proposed an empirical framework for measuring the effects of entry in concentrated markets. Building on models of entry in atomistic competitive markets, they concluded that the number of producers in an oligopolistic market varies with changes in demand and market competition. These analytical results structure their empirical analysis of competition in five retail and professional industries. Using data on geographically isolated monopolies, duopolies, and oligopolies, they study the relationship between the number of firms in a market, market size, and competition. Their empirical results suggest that competitive conduct changes quickly as the number of incumbents increases. In markets with five or fewer incumbents, almost all variation in competitive conduct occurs with the entry of the second or third firm. Surprisingly, once the market has between three and five firms, the next entrant has little effect on competitive conduct (see figure 4 below).

The paper shows how economists have used the number of companies in a market to retrieve information about the market demand and costs of each company. To do this, it is assumed that all potential constituents are equal.

2.1. Homogeneous simple model

Bresnahan and Reiss develop an empirical model of *N,* the number of homogeneous companies that choose to produce a homogeneous good. To do this, they develop a multiperiod oligopoly where *M* potential candidate to entry decide to enter and what to produce. When we develop an empirical model, we limit ourselves to the situation where we observe *N* but not the output of each q firm.

The empirical question of this model is what one can learn from the demand, cost, and competitive behavior of observing the *number of companies* N_1 , ..., N_T that have entered *T* markets. To do this N_i *(observable)* is related to the benefits of each company in the market *i.* Given N_i firms in the market *i*, each obtains the following profit:

$\pi(N_i) = V(N_i, x_i, \theta) - F_i.$

Where, *V(.)* represents variable benefits and *F* is a fixed cost. We consider under the assumption of homogeneity that all companies have the same variable function *of profits and fixed cost* F_i, vector x_i contains the market demand and cost variables *i* that affect business profits. θ It is a vector that contains the demand, cost and competition parameters that we try to estimate. To relate this profit function to the data on the number of companies we assume that in addition to *observing* N_i observamos *observe* x_i θ but not or fixed costs F_i. While in principle, $x_{i\ could}$ include endogenous variables such as prices or quantities, we simplify for the time being assuming that the only endogenous variable is the number of companies that entered the market. $I_{don't.}$.

Before beginning to estimate, the researcher must specify how the benefits depend *on N* and what is observable for the companies and for the researcher. These two decisions cannot be made independently as we will see below. The goal of introducing variables that the researcher does not observe is to rationalize why there is no exact relationship between xi and N_i.

We assume that the non-observable fixed costs are distributed independently in the markets according to the $\phi(F \mid x, \omega)$ distribution, which describes not only F_i but also the profit function of the company. $\pi(N_i)$.

Once the profit function is determined, the next task is to relate market entry decisions to N_i .

With symmetry between companies and perfect information and profits are a non-growing function N of, we only need two inequalities:

For N^* sign-ins:

$V(N^*, x, \theta) - F \geq 0,$

And for any other potential enterer

$V(N^* + 1, x, \theta) - F < 0.$

When we combine the two conditions, we set an upper and a lower limit on fixed costs

$V(N^*, x, \theta) - F \geq F \geq V(N^* + 1, x, \theta).$

These limits provide a basis for estimating the benefit and fixed costs parameters of θ, ω the information obtained from x_i y N_i. For example, we use the probability of observing N^* companies:

$$\begin{aligned}\mathrm{Prob}\big(V(N^*, x) \geqslant F \mid x\big) &- \mathrm{Prob}\big(V(N^*+1, x) > F \mid x\big)\\ &= \Phi\big(V(N^*, x, \theta) \mid x\big) - \Phi\big(V(N^*+1, x, \theta) \mid x\big)\end{aligned}$$

To construct the likelihood function for N^*. Under independent and identical sample assumptions this likelihood has a dependent variable

$$\mathcal{L}(\theta, \omega \mid \{x, N^*\}) = \sum_i \ln\left(\Phi\left(V(N_i^*, x_i)\right) - \Phi\left(V(N_i^* + 1, x_i)\right)\right)$$

It is essential to note that although the un noted profits of companies are statistically independent between markets, this likelihood function presumes that the profits of companies are economically independent between companies. These assumptions of independence will be much more realistic if we model a cross-section of several countries in several markets and not the same company over time or in several markets.

The simplicity of this function and its direct connection to the theory is especially important despite the extreme assumptions used. More importantly, we learn that we only have discrete data on the number of companies in different markets we will be forced to impose assumptions on the distribution of profits to know the parameters. For example, if we assume that the un not observed fixed costs have a normal iid distribution (independently and identically distributed) or logit then the above equation will have a probit or logit function.. But how do we know fixed costs are iid between markets? Economic science is not much help to know the distribution of fixed costs. Then without further information it will be difficult to recover observations only with N^*.

Given the arbitrariness of fixed cost assumptions, it seems necessary for researchers to explore the sensitivity of estimates to alternative distributions.

Therefore, we have developed an econometric model of the number of companies in a market based on two conditions of balancing the unfore observed profits of homogeneous companies. One is based on him N^* chosen to enter while the other assumption is that they chose not to $M - N^*$ enter. To know the number of companies researchers will have to make strong assumptions. Many of these assumptions can relax but come at a cost. We now discuss how we can make inferences about competition in V-specific models following Bresnahan and Reiss (1991).

2.2. V's relationship to the intensity of competition

We now study how to specify the profit function variable *V(N,x,* θ *)* and the fixed cost function *F(x,* ω *).*

There are two main approaches to specifying *as V(.)* depends on *N* and *x.* The first is to choose a simple parameterization of *V(.)* that meets that *V(.,x) is* non-growing in *N.* For example, the researcher can assume that x and 1/N *enter V linearly* with constant coefficients, and the coefficient over 1/N is limited to positive values. The advantage of this approach is that it results in a conventional probit model but as a disadvantage is the disrecognise of what it represents. θ.

A second approximation is to derive *V(.)* directly from specific assumptions about the cost and demand function and assumptions about the game after entry. This approach has as an advantage that clarifies which economic assumptions motivate the choice of V*(.)* and what the parameters represent. θ As a disadvantage is the complexity of the econometric calculation of the profit function.

Consider a market for a homogeneous good with *M* potential starters. Suppose that each incoming j has the variable cost function and that market $C_j(q_j)$ demand i has the form:

$$Q_i = S_i q(P_i),$$

Where *S* is the size of the market, *q* is the demand per capita and *P* is the price. In a standard Cournot model each entering company maximizes profits by choosing the output so that in the i *market,*

$$P_i = \frac{\eta_i}{\eta_i - s_j} MC_j \text{ for all} j\text{-}1,\dots\ j= ,N \leq M$$

Where MC_j is the marginal cost of inbound j, sj is market share j (equal to 1/N in the symmetrical case) and is equivalent to less market demand η_i elasticity i. In the above equation it is difficult to see how prices vary with the number of companies on the market. To explore N's *N* effect on prices it is useful to add the mark-up of companies to get the price equation: mark

$$P = \frac{N\eta}{N\eta - 1}\overline{MC},$$

Where $\overline{MC}$ is the average marginal costs of incoming *Ns.* This equation shows that the price of the industry depends not only on the number of companies entering the market but on the average marginal costs of companies. If we are interested in the size of the distribution of the starters, then we can add to get

$$P = \frac{\eta}{\eta - H}\overline{MC}^{w},$$

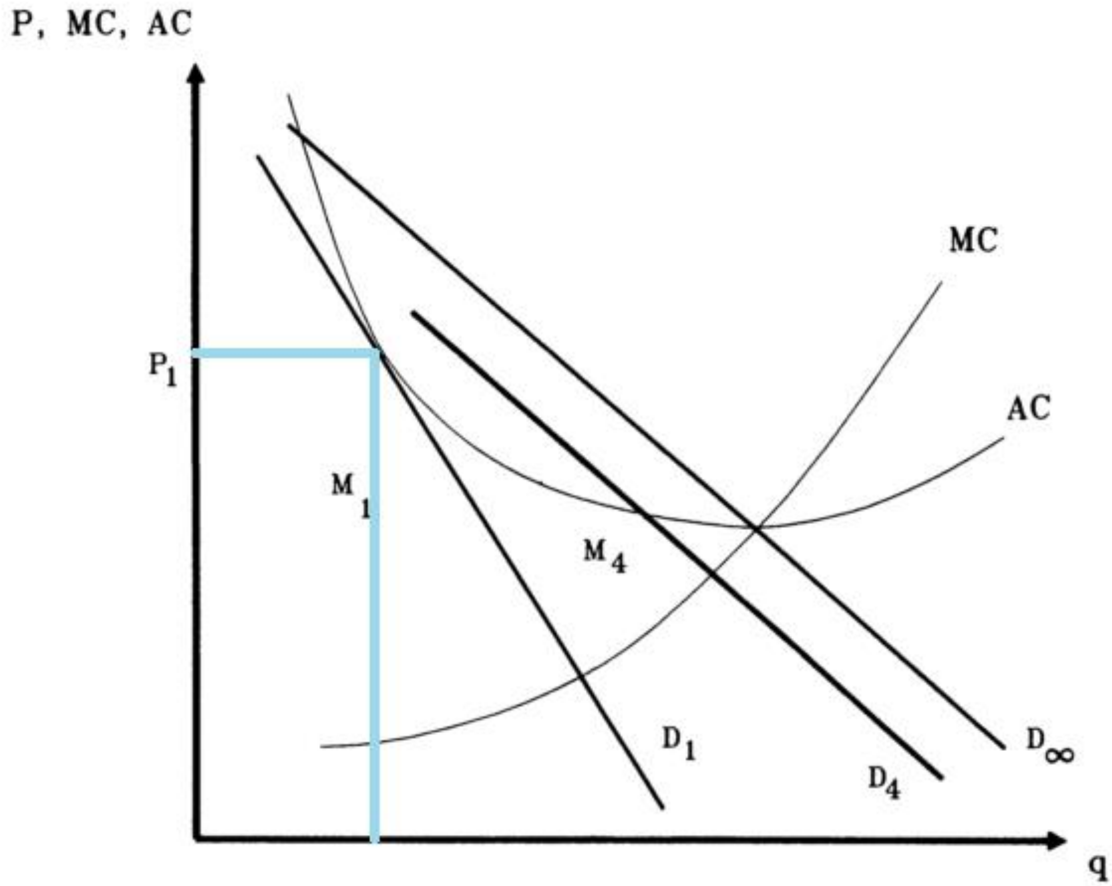


FIG. 1.—Breakeven firm demand and margins

(Bresnahan Reiss pg. 980)

The paper points out (see in figure 1), as the size of the market increases, market demand rotates outward. This increase increases the monopolist's profits. It also increases potential entrants' post entry profits. Most oligopoly theories predict that continued demand growth will encourage entry, thereby reducing incumbents' margins. Eventually, as market demand grows large relative to minimum efficient scale, firms' price-cost margins will tend to competitive levels. In figure 1, this occurs when the per firm demand curve D, passes through the bottom of average total cost.

The paper studies how margins change as entry changes. Since margins are unobservable, the authors use entry thresholds as a proxy of margins. Equivalently, the more efficient a monopolist is at surplus extraction, the greater this ratio. Bresnahan and Reiss studied ways a researcher can use price and quantity data to infer demand and costs, and competition between companies.

They calculate entry thresholds as the ratio of unobservable fixed costs F to variable profits per customer:

$$S_1 = \frac{F}{[P_1 - AVC(q_1, W)]d(Z, P_1)}.$$

The entry threshold ratio s_∞ / s_1 measures the fall in variable profits per customer between a monopoly and a competitive market. This scale-free measure of competition is bounded below by unity and increases with a steepening of the monopolist's demand curve.

Market structure models endogenize the number of companies in the market by modeling the decisions of potential enterers to enter the market or not. To do this they assume that companies expect non-negative profits, conditioned on the expectations or actions of competitors including those that did not enter. This is analogous to the revealed preference theory that is the foundation of discrete choice modelling models. By observing how business decisions economists can understand the underlying determinants of business profitability, the role of fixed costs, heterogeneity and the nature of competition in the market.

The data appears in table 3 and is explained in page 985:

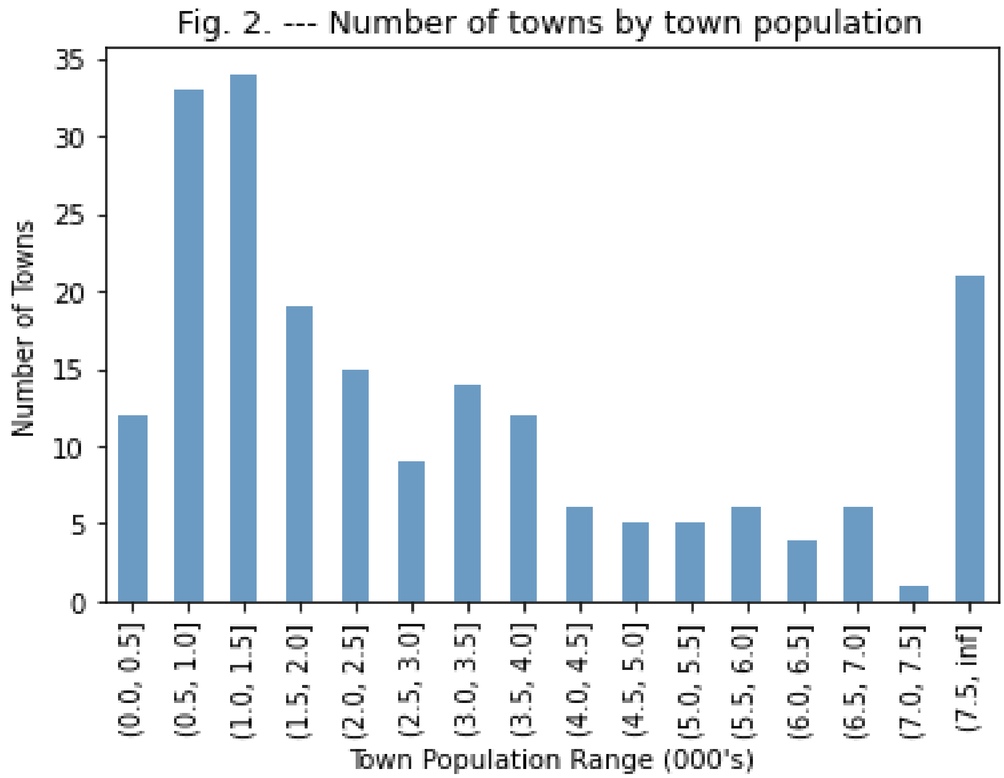
> *"Our sample contains 202 isolated local markets. A typical market in our sample is a county seat in the western United States. These county seats are separated from other towns in the county. Because most of the local population resides in or near the central town, its population provides a reasonable first approximation to S(Y).*
>
> *...*
>
> *We selected our sample of markets and industries using criteria developed in our earlier work (see Bresnahan and Reiss 1988, 1990). Briefly, we located towns or small cities in the continental United States that were at least 20 miles from the nearest town of 1,000 people or more. We elimi- nated towns that were near large metropolitan areas or were part of a cluster of towns. Our specific criteria exclude, for example, towns within 100 miles of a city of 100,000. We believe, a priori, that these selection criteria ensure that we can identify all relevant competitors.*
>
> *...*
>
> *We limited our study to industries or occupations in which we could identify all sellers of a narrowly defined product or service. We did not consider grocery and clothing stores, for example, because they sell a range of products."*

The variables included in the study are:

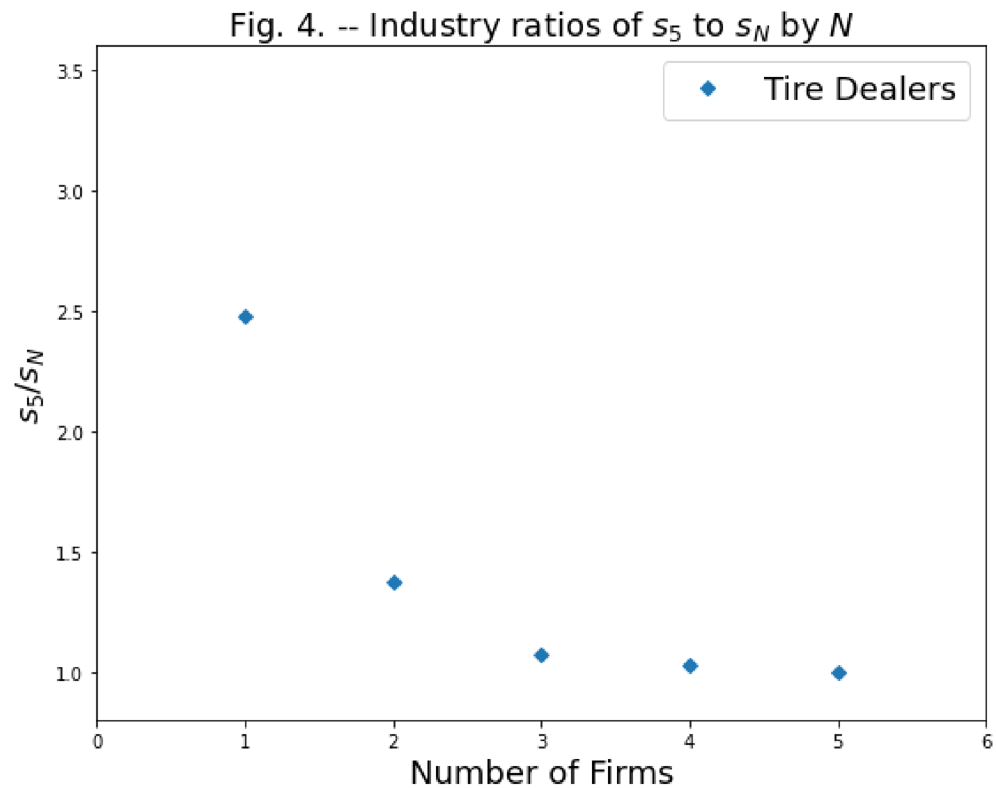
- ID: The unique ID associated with the represented market.
- TIRE: The number of incumbents "tire dealer" firms in the market
- TPOP: Town population.
- NGRW: Negative TPOP growth. This represents the negative growth in town population from 1970 to 1980.

- PGRW: Positive TPOP growth. Similar to NGRW. These growth terms capture entrants' asymmetric expectations about future market growth as well as lags in responses to past growth.
- OCTY: Commuters out of the city. There are commuters out of the county. It represents the Census Bureau's count of county residents who commute to work outside the county.
- OPOP: Nearby Population. This includes population within 10 miles of the town. (This would increase demand.)
- LANDV: Value per acre of farm-land and buildings ($1,000s)
- ELD: 65 years and older divided by the county population
- FFRAC: Fraction of land in farms
- PINC: Per capital income ($1,000's)
- LNHDD: Log of heating degree days

	count	mean	std	min	25%	50%	75%	max
TIRE	202.00	2.62	2.63	0.00	1.00	2.00	4.00	13.00
TPOP	202.00	3.74	5.35	0.12	1.05	2.10	4.34	45.09
NGRW	202.00	-0.06	0.14	-1.34	-0.04	0.00	0.00	0.00
PGRW	202.00	0.49	1.05	0.00	0.00	0.08	0.42	7.23
OCTY	202.00	0.32	0.69	0.00	0.08	0.16	0.32	8.39
OPOP	202.00	0.41	0.74	0.01	0.06	0.14	0.42	5.84
LANDV	202.00	0.29	0.23	0.07	0.14	0.24	0.35	1.64
ELD	202.00	0.13	0.05	0.03	0.10	0.13	0.16	0.30
FFRAC	202.00	0.67	0.35	0.00	0.32	0.83	0.94	1.27
PINC	202.00	5.91	1.13	3.16	5.17	5.87	6.68	10.50
LNHDD	202.00	8.59	0.47	6.83	8.45	8.70	8.93	9.20



The following figure replicates figure 4 of Bresnahan and Reiss only the case of Tire Dealers, (p. 996) It plots the ratio of the market size required to support five versus N firms, that is, s_5 / s_N,. This ratio equals one for N = 5. For N < 5, it can vary anywhere from one to infinity, depending on the entrants' estimated costs and variable profits (see eq. [l11). Figure 4 shows that these ratios are very near one once the market has more than two firms. In markets with two or fewer firms, however, they may be much greater than one.



In this paper I explain an algorithm designed by Pakes and Mc Guire [1] [2] to explain industry dynamics with differentiated products. The model has been widely used in industrial organization due to its generality and flexibility to explain industrial dynamics. The model could be used also as a tool to define a counterfactual structure for antitrust policy. Firms invest to improve product quality in a stochastic process were in some cases, investment produces positive outcomes that lead to higher profit while in other it produces negative outcomes that can eventually to exit the market. In the Pakes and McGuire [2] model, forward-looking oligopolistic firms compete in the product market and through their investment, entry, and exit decisions. By investing a firm aims to increase the quality of its product—and its profit from product market competition—over time. Investment, entry, and exit decisions are thus both dynamic and strategic. Similar models have been applied to advertising, auctions, collusion, consumer learning, environ- mental policy, international trade policy, learning-by-doing, limit order markets, mergers, network externalities, and other applied problems.

1 Introduction

Pakes and McGuire algorithm computes a perfect Nash Equilibrium for dynamic heterogeneus agent models that is able to reproduce the conduct of firms in a wide range of market structures. The model includes heterogeneity and uncertainty due to the random outcomes of the investment process. Firms invest to improve quality and production process which enable higher but uncertain profits. It could also be that the investment fails to produce higher returns leading to a possible market exit. Current profits depends on its own capacity, the capacity of other competitors in and outside the market. Actions such as investment, entry are intended to maximize the expected discounted value of future net cash flow conditional on current information set.

2 Pakes MacGuire Model

The model evolution is explained in [3]: The sequence of moves in the game is as follows. At the beginning of each period: (i) incumbents evaluate whether it is better to stay in the industry or exit, based on the exogenous scrap value they could obtain, and, if they stay, decide how much investment will maximize the present value of the sum of their expected future profits; (ii) potential entrants compare the exogenous costs of entry with the present value of the expected profits and decide whether or not to enter. These decisions are all simultaneous.

Entry and exit take one period to be completed: that is, firms leaving the industry still gain profits in the period they decide to exit, and entrants will only gain profits in the next period. After these decisions are made, spot market competition happens, idiosyncratic and common industry shocks to the firms' states are realized and the states of the firms evolve. PMG model consider time to be discrete and infinite. The profit function maps all combinations of states into the real numbers. Industry consists of N firms with potentially different product qualities (w

References

Bain, J. 1956. Barriers to New Competition. Cambridge, Mass.: Harvard University Press.

Berry, Steven, and Peter Reiss. "Empirical models of entry and market structure." *Handbook of industrial organization* 3 (2007): 1845-1886.

Bresnahan, Timothy F., and Peter C. Reiss. "Entry and Competition in Concentrated Markets." Journal of Political Economy, vol. 99, no. 5, 1991, pp. 977–1009. JSTOR, http://www.jstor.org/stable/2937655. Accessed 17 Oct. 2022.

[1] Chaim Fershtman and Ariel Pakes. A dynamic oligopoly with collusion and price wars. The RAND Journal of Economics, pages 207–236, 2000.

[2] Ariel Pakes and Paul McGuire. Computing markov-perfect nash equilibria: Numerical implications of a dynamic differentiated product model. The RAND Journal of Economics, 25(4):555–589, 1994.

[3] Paulo Henrique de Alcántara Ramos. Two essays on dynamic analysis of imperfectly competitive markets. 2017.

Chapter 5

Merger analysis

Introduction

The first step in determining the effects of a merger is an in-depth knowledge of the market where the transaction will take place, by defining the products that are consumer-substitutes. A market usually includes a group of competing products within a geographic area. The relevant product market and secondly the geographic market are described during the market definition analysis.

In some cases, the resulting entity will have no market power or that the merger will not increase this power. In any case it would not be necessary to establish which relevant market definition is correct. In many merger decisions the analysis does not go any further because under any market definition even the narrowest there is no risk to competition. If, for example, two companies have an identical market share in markets A and B, then there will be no difference between whether we consider one market for A and one for B or whether we consider a market for both products.

The definition of the market is important for two reasons: because it defines the scenario where the most immediate competition of the merged company is located and, secondly, we can only calculate market shares once the market is defined. This indicator is crucial as if the merger results in a reduced share (especially under all possible relevant market definitions) then it is possible to consider not going ahead with the analysis.

To define the market, it is usual to consider whether certain products can serve the same purpose and also with the same cost-efficiency. (e.g., a car helicopter to go to work is not substituted). The definition of the market is not always obvious: does a cola manufacturer compete with other cola-flavored beverages, non-alcoholic beverages, or other beverages? The market share will depend on this definition, and therefore the opinion of the authority on this merger.

It should also be noted that not all consumers are equal, it is necessary to distinguish between the average and the marginal consumer, and that the fact that many "captive" consumers will not switch to another product does not indicate that there are enough marginal consumers who are price sensitive and able to switch to another substitute product. The latter question is the most relevant for the purposes of market definition.

It should be mentioned that it is rarely possible to be able to define a market with mathematical precision and therefore the relevant market should be considered as an appropriate framework for the analysis of competitive effects. It is also important to recognize that competing products should not be seen within the market as 100% substitutes for any product within the market definition or off the market with 0% replaceability with products on the market. This false dichotomy is known as fallacy 0-1, one of the most serious problems of market definition especially in differentiated products. Despite this, market definition remains an important tool for analyzing the competition restrictions facing the merged company with its rivals.

In the case of a market with differentiated products, it can be difficult to define the market accurately, in this case it is necessary to sustain the market definition with a more rigorous analysis.

Use of Precedents

Investigations can be leveraged in previous cases when that market has already been analyzed. However, caution must be exercised as the conclusions would sometimes not apply to the merger in question:

- Market conditions change over time (the market is dynamic non-static), innovation can change the replaceability of products.
- The definition of a market applied to a geographical area may not apply to another area if there are differences in the purchasing power of consumers.
- The behavior of a company with market power can affect the market definition itself, may increase exchange costs for customers or close the market to other bidders.
- When the market includes differentiated products, by brand or quality, it can be difficult to define a market accurately.

Product market

The definition of the market focuses on the empirical question of the replaceability of products and services from the point of view of consumers. From a demand point of view, it is a question of analyzing whether consumers can replace between substitute products in response to a change in relative prices, quality, or availability. It is interesting to analyze from the point of view of demand, which products are seen by competitors as substitutes.

From a supply point of view, it examines the extent to which bidders may produce other products in the face of a change in relative prices, demand, or other market conditions. In some jurisdictions, only demand substitution is considered.

Analytical process to define the relevant market

The definition of the relevant market begins by considering the narrowest possible market where one or both parts of the merger operate.

The SSNIP or "small but significant non-transitory increase in Price" test is often used and is to analyze whether price increases (between 5 and 10%) the monopoly would be able to maintain the benefits. If consumers replace this product with the purchase of others, then the market is not properly defined, and other products must be added. The process is repeated to a point where the "hypothetical" monopoly reaches market power, i.e., it can maximize profits by keeping prices above market ones. This is defined as the relevant market. In practice, there is usually not enough data to perform an SSNIP test, so in those cases it is used as a conceptual framework.

Here are the data needed to find out if there is product replaceability by consumers:

- Features and use of consumer products and preferences. Although two products have similar characteristics, it is also necessary to consider brand loyalty and switching costs, white label products may have the same branded characteristics but not be in the same market. Related products physical characteristics do not have to be substitutes (trains at rush hour, in valley time), quite different products can be nearby substitutes (matches, lighters)
- Information can be extracted from the documents of the undertakings of the transaction on the products which they consider to be close substitutes to their own as well as the undertakings they consider to be their competitors.
- Consumer surveys to identify consumption patterns for example, how they have reacted to a price increase and how they would react to a hypothetical increase in the price of their product. The answers should be treated with caution, considering that the sample is representative. It is necessary to consider in the competition questions that may want to dynamite merger.
- High switching costs relative to the value of the product can make the replacement less likely. You can find out by asking consumers about past experiences when switching products.
- Similar product price developments: when they have nothing to do with costs or inflation developments, they can be proof that they are close substitutes. However, correlation data should be treated with caution.
- Indications that consumers buy a rival product as soon as there is a minor increase in the price of the product in question indicates that they are substitutes. If there is a divergence in time without substitution it would indicate that the products are in separate markets. This can be seen by looking at prices over time and seeing if product volumes diverge in response to a change in the prices of one of the products.

- Own elasticity or cross demand: Elasticity-price demand measures the rate at which the quantity of a product changes with a one percent price change. Cross elasticity measures the rate at which the amount of a product varies when the price of another product increases or decreases. It's hard to calculate. Companies may have an estimate.

- From the point of view of supply replaceability:

-Questionnaire to potential competitors as to whether substitution is technically possible, the cost of substitution and the time it would take to modify the manufacturing process. It is a question of asking whether it would be profitable to modify production to produce another product given a 5-10% increase in price.

- Ask potential bidders if they have surplus capacity or are interested in producing another product.

- It is even possible that new producers produce this product but consumers do not want to buy them so they should also be asked.

- The authorities should take into account that supply-side replacement must follow strict criteria by rejecting those that do not meet.

Geographic Market

The geographic market can be defined as the area where there is a substitutability between the products of the merged firms. The first approach would be to define the smallest area where a hypothetical monopolist would maximize his profits by imposing a reduced but significant and non-transient price increase. It can be local, regional, national, EU or global.

Identical to how it is studied in the case of the product market, it is a question of identifying substitutes that are close enough to prevent a hypothetical monopoly from sustaining a price increase of 5 or 10%. The analysis begins with the narrowest area to increase it to an area where that price increase could be sustained.

Imports

Imports can exercise a competitive barrier so that the market should be defined above national. A high volume of imports may indicate that the market is wider than domestic. The key question is whether importers can be restricted in imposing prices by domestic manufacturers so that they cannot raise prices.

It is also necessary to take into account:

- Product value / transport costs ratio: the greater the greater the replaceability with household products.

- Need to have a local presence.

- Tariffs, anti-dumping measures.
- Distance: reduced for local consumers, higher for wholesalers.
- Language barrier.
- authorization of local authorities (chemicals, pharmaceuticals)

B. Market structure and merger.

Once the relevant and geographical market has been defined, it is appropriate to analyse its structure and how it would change as a result of the merger.

Market shares, merger ratios and the Herfindal Hirschman Index (HHI) are used to analyse market structure and merger. It is essential to indicate that each of these measures can beused as an initial indicator of potential competitive problems in the market. But more in-depth researchis needed including an analysis of other market characteristics in addition to unilateral or coordinated effects.

Economic rationality: In the case of a merger between A and B, the market shares of A and B and their rivals may give an indication of whether the loss of competition between A and B is significant or whether the remaining competitors can limit monopolistic capacity so that this loss is not as significant. In general, at higher market shares of merged companies, the more likely the loss of competition is significant. In general, fewer companies on the market, the greater the likelihood that the merger will present a significant loss of competition for the other undertakings. However, merger indices are only an initial measure for knowing the effects of a merger.

Merger ratios

There are several merger rates on the market, which can serve as a threshold for identifying those mergers that are most likely to generate competition problems and therefore require further investigation. Information on market shares (from associations, consumers, suppliers or market research) must first be collected. This market information should give a precise indication of how competition works in a particular market. Sales volumes are the most commonly used figures because they allow you to know the amount of each product when they are differentiated. The number of bidders can be used in public procurement or tender markets. It can be difficult to obtain the necessary figures to carry out merger calculations on the relevant market, there is no market data, or it does not match the geographic market or other products must be added to reach the relevant market, etc. It may be advisable to ask the parties to estimate their market share and to explain how they have obtained this data and sources.

- Market shares

Odds indicate the percentage of total sales of the company's product and its rivals. It's the starting point of the analysis. First, the quotas of the companies being concentrated are added up and the increase in the share derived from the merger are the starting point for whether there are unilateral effects. It is also advisable to compare the combined share with that of other major market rivals. Clearly those mergers that create a high market share for the resulting company are the most worrying from the point of view of competition.

In some cases it is necessary to analyze the historical evolution of market share (when there is a lot of seasonality).

- Merger indices

HHI Index: It is about adding the square of the market shares of all companies active in the market. Absolute value is obtained or variation as a result of the merger can provide us with an indication as to whether the merger may be worrying or not. (ranges from 0 fully fragmented market to 10,000 where a company has 100% of the market share). The authorities should be more concerned about a highly concentrated market in the face of a very fragmented market for a possible merger, and they must also be more concerned about a merger that significantly increases the merger of the market compared to another that only marginally does. It is usually not possible to calculate the HHI because the quota of all companies is unknown, in that case the increment (Delta) could be used or the HHI used whenever the majority quotas are known.

There is no a priori merger figure from which we must be concerned.

price increases and the harm to competition from the merger (Pakes, 2010; Cheung, 2016).4

Unilateral effects

If alternatives were considered in the previous chapter to measure the impact of a merger on the level of merger on a particular market, this chapter analyses some anti-competitive effects of the transaction. It's about assessing how incentives to compete can change for participants and rivals after the operation. There are two theories: unilateral and coordinated effects.

In this section we study the unilateral or uncoordinated effects, which explain those effects arising from uncoordinated actions of the participants. It is a question of assuring whether, as a result of the transaction, shareholders can exercise their market power by increasing prices, reducing production or variety or quality as a result of the elimination of competition previously existing between the shareholders.

For example, the merger of two manufacturers of differentiated products A and B, before the merger if A raised the price lost customers in favor of B but after the merger internalizes in the new company A+B which can now increase the price of all its products without risk to lose customers. Rivals will also unilaterally raise the price. In addition to an increase in price there may be other negative effects on competition as a less incentive to innovate, diversify or improve quality.

Several factors allow us to assess the similarity of the products of the shareholder companies:

-internal documents: where their competitors are identified

- consumer opinion

-market research, analysts, etc.

- documents from other competition authorities

It is also necessary to take into account:

-Barriers to entry: if they are reduced the participants will not be able to take advantage of the market power they obtain.

- Buyer's market power: if you can search for other alternatives quickly.

- Participants are not close competitors: in this case pre-merger market shares are not adequate indicators of the level of rivalry between the shareholders.

- Response from competitors: They can alternatively increase production or reposition the.

- Other suppliers that consumers can buy from. What doesn't usually happen in the event of a preference for a brand or reputation.

- New incoming potential with new products.

- The new company may be able to limit the expansion of new competitors: e.g. Patents or trademarks, platforms.

In short, these are some characteristics that can cause the anti-competitive effects of a merger to be greater or lesser.

D. Coordinated effects

This section examines the situation in which the incentives of shareholders and rivals change after the merger from competing to coordinating, which ends up reducing social welfare. This behavior doesn't have to lead to collusion.

Coordinated effects arise when merger generates incentives for companies to coordinate their behavior rather than compete. It is a question of identifying what factors can lead companies to coordinate after the merger:

- Companies should be able to identify the terms of coordination (prices, e.g.)
- It should be costly to deviate from coordination, and it should be possible to detect and punish non-compliance.
- Little competition from outside the market (low probability of market enterers entering and de-stabilizing)

Let's look at each of them:

(a) Companies should be able to identify the terms of coordination. It is not necessary for companies to reach an explicit agreement to coordinate but to reach some kind of understanding, but for this three conditions are necessary: market transparency, homogeneous product, existence of "maveriks" (the market may be charged because it has lower costs or different policy) and Cross-participations: If there are cross-participations increases the likelihood of collusion.

b) It should be costly to deviate from coordination, and it should be possible to detect and punish non-compliance. For coordination to be sustainable, there needs to be sufficient transparency to be able to detect non-compliance and have credible means of punishment. It may be sufficient for companies to have an incentive not to deviate (to perhaps avoid a return to aggressive competition). The NAC must fully understand the market in question in order to deduce whether coordination is sustainable or not. For this condition to be met competitors must know the offers of the other companies, stable and predictable demand, regular and frequent orders, cross-participations reduce the incentive to breach the agreement, symmetrical companies in costs.

c) Little competition from outside the market: The market must be stable with limited competition that does not distort coordination. To do this, barriers to entry must be high and there must be not many small competitors. The evidence necessary to be able to analyze this condition are the market shares of all companies in the market, information about past entrans, data about buyers.

This information can give us guidance on whether market conditions can facilitate coordination or not but do not tell whether merger can make coordination easier or more likely.

E. Barriers to entry and expansion

A merger that increases market merger should not generate anti-competitive effects if new companies that curb the market power of merged companies can enter.

In this section we consider the entry of new companies or expansion of existing ones as a result of a merger. Market entry in this paragraph refers to the entry that would occur as a result of new market conditions generated by the merger.

In section B we analyze the restrictions on competition generated by companies already on the market, in this section we analyze those generated by companies that are off the market. In particular, NACs should study:

- If entry is possible and feasible. This requires cost-effective entry. The first thing is to see if there are barriers to entry that can be absolute (regulation, intellectual property rights, environmental restrictions, etc.), structural by demand cost or technology, economies of scale that can slow down entry, strategic advantages when incumbents have a privileged position over incoming ones.
- If the entry or expansion can prevent or reverse the anti-competitive effects of a merger.
- If the entry can take place in a reasonable period of time.

You can compare the costs of entering with the expected sales revenue and how long it takes to recover the entry costs to see if a potential sign-in is profitable. It is useful to ask customers if they want to switch to a new vendor as this can influence a signer's ability to affect the behavior of merged companies. It may also be interesting to have the history of market entries to be able to conclude that market entry is possible and can continue.

It should also be noted that it is not only a question of entering the market but that the entry is relevant enough to affect the merger. Therefore, it is necessary to analyze whether the new entry is of little relevant so that the merged ones can continue to increase prices.

F. EFFICIENCIES

Mergers can generate efficiencies, which are transferred to consumers in the form of lower prices or greater innovation. It is often defended that mergers generate efficiencies, although they are often not adequately justified by what THECs often dismiss them. Its quantification is usually the most speculative part of the merger.

Mergers often generate efficiencies by enabling better utilization of existing assets, allowing the merged company to achieve lower costs than they individually might have. Efficiencies can generate greater market rivalry so there would be no negative competitive effects as a result of a merger. Efficiencies mean lower costs, more capacity intensive use, economies of scale and scope, higher network, higher quality.

Efficiencies often have an impact on short-term pricing, if they reduce marginal or variable cost and maintain sufficient market competition, which is not the case for fixed costs. When the objective is to improve consumer well-being, only those efficiencies that are transferred to consumers will be taken into account. Other efficiencies such as those obtained in production are considered in jurisdictions where all social welfare is considered.

To incorporate efficiencies into a merger analysis, what is usually done is that they are part of the competitive analysis showing that the economic incentives to compete of the merged companies can be increased so that the merger does not affect consumer well-being. The valuation made by the participants is usually very optimistic, the tests have them which complicates the analysis. Some of the benefits from efficiencies are generally requested to be transferred to consumers in the form of lower prices and/or higher production.

Efficiencies should be the result of the merger and should not exist in the absence of the operation.

Critical Loss Analysis (CLA): was introduced in 1989 (Harris and Simons, 1989). The standard methodology for market definition in merger analysis begins with the *"hypothetical monopolist test": whether "a hypothetical profit-maximizing firm, not subject to price regulation, that was the only present and future seller of ... [a set of products] likely would impose at least a small but significant and non-transitory increase in price ('SSNIP') on at least one product in the market ..."* (U.S. Department of Justice and Federal Trade Commission, 2010).

Algebraically, we assume that industry profits are currently set as:

(1) $\Pi_0 = (P_0 - C)Q_0$ where fixed costs are assumed sunk and irrelevant in the short term and marginal costs C are assumed constant. The hypothetical monopolist of the industry would consider increasing price according to the following equation:

$$(2) \quad \Pi_1 = (P_0 + \Delta P - C)(Q_0 - \Delta Q)$$

and the critical issue is:

$$(3) \quad (P0 + \Delta P - C)(Q_0 - \Delta Q) \gtrless (P_0 - C)Q_0$$

Does the increase in profit from the price increase ΔP on the sales that remain outweigh the reduction in profit resulting from the loss of sales ΔQ? Simple algebra reveals the "critical loss" of output that would make the monopolist indifferent:

$$(4) \quad \frac{\Delta Q}{Q} = \frac{\Delta P / P}{M + \Delta P/_{P}}$$

where $M = (P_0 - C)/P_0$, the existing price-cost margin. If the loss of sales ΔQ resulting from the increase in price ΔP is higher than that which satisfies this equation, the price increase would not be profitable.

We now analyze whether the merger of two firms, say firms 1 and 2, would provide the merged firm with the unilateral incentive to increase its own price only. Before the merger, firm 1 sets prices according to the same profit-maximization principles as the hypothetical monopolist. Prices were increased until just the point at which the loss in sales – both to other firms in the market and to the rest of the economy – outweighed the benefits of higher prices. After the merger, however, some of the losses from a price increase for the product of firm 1 are newly internalized by the firm – they are "recaptured" by firm 2, now under the same control as firm 1. Thus we have a new factor in the calculations, the diversion ratio D_{12}, the share of the sales of 1 that are lost as the result of a price increase for 1 that are "recaptured" by firm 2 (Willig, 1991; Shapiro, 1996, 2010).

Thus the profit-maximization question facing the merged firm as it considers whether and by how much to raise the price of good 1 is now:

$(5)(P_1 + \Delta P_1 - C_1)(Q_1 - \Delta Q_1) + (P_2 - C_2)(Q_2 + \Delta Q_1 D_{12}) \gtreqless (P_1 - C_1)Q_1 + (P_2 - C_2)Q_2$ where $\Delta Q_1 D_{12} = \Delta Q_2$. As with equations (3) and (4), this translates into the critical loss of quantity W for the merged firm to raise price on good 1:

$$(6)\Delta Q_1 /Q_1 \gtreqless (\Delta P_1/P_1)/[(P_1 - C_1) + (\Delta P_1/P_1) - (P_2 - C_2)(P_2/P_1)D_{12}]$$

Comparing equation (4) – as interpreted for a single profit-maximizing firm premerger – and equation (6) – the calculation for the firm after merging with a competitor – shows that the difference is the last term in the denominator on the right-hand-side – and that because this is the subtraction of the product of three positive terms, it will tend to increase the size of the critical loss that would make a price increase unprofitable. The increase in the critical loss for good 1 following the merger – and thus the increase in the incentive of its producer to increase price – is a positive function of a) the margin earned on the second good 2, b) the ratio of the price of the second good 2 to the first good 1, and c) the diversion ratio of 1 to 2 – the percentage of sales of good 1 lost by a price increase that are recaptured by good 2, now owned by the same firm.

Upward Pricing Pressure (UPP) might be considered a more direct focus on the potential efficiencies of the proposed merger and whether they are likely to outweigh the loss of competition in the price-setting of the merged firm. In the previous section of the paper, we abstracted from the distinction between homogeneous and differentiated products, assuming that all firms charged the same price but that a single firm had the option to charge a different price. In this section we abandon this abstraction and embrace the distinction between these two types of goods that was such an important part of the revised Merger

Guidelines in 1992. As Shapiro (2010) discusses at length, while the 1982 and 1984 Guidelines focused on the danger that a merger among competitors would increase the likelihood of collusion – either explicit or tacit – in the more concentrated market, thus focusing implicitly on homogeneous products, the 1992 Guidelines added a second focus on the likelihood that a merger among competitors producing differentiated products would provide incentives for the merged firm to raise price unilaterally, regardless of the behavior of competitors in response. Thus a new emphasis was placed on the degree to which competing products were close or distant substitutes to each other – a concept implemented in the term 14 that we introduced in the previous section, the diversion ratio D_{12} between two firms **1** and **2**, the percentage of sales of good **1** lost in response to a price change for good **1** that are "diverted" to good **2**. A larger value for this diversion ratio D_{12} clearly suggested a merger that would be more troublesome from a competitive standpoint, ceteris paribus. However, the 1992 Guidelines, along with the subsequent literature, simultaneously added a new focus on another term in the denominator of equation (6): the price-cost margin earned in the production of good **2**. If this margin were "high", especially vis-à-vis the margin earned in the production of good **1**, the merged firm would be quite happy for sales of **1** to be diverted to sales of **2**; not so much if the margin earned on the production of **2** were "low". Thus increased attention came to be focused on the product D_{12} $(P_2 - C_2)$, the value to the merged firm of sales of **1** that were diverted by the price increase for **1** to sales of **2**, a value termed the "Gross Upward Pricing Pressure Index" (GUPPI) from the merger (Farrell and Shapiro, 2010a; Moresi, 2010). A corresponding term was calculated and considered for the merged firm's incentive to increase the price of good X.

GUPPI

But how are we to interpret and use GUPPI? Any non-zero diversion from W to X accompanied by any non-zero margin on X would yield a positive value for GUPPI. One option would be to proceed with the CLA for the analysis of unilateral merger effects, plugging an estimated value for GUPPI into equation (6). Again, for an evaluation of the impact of the merger one would also perform the exercise in reverse, analyzing the incentives created by the merger for a unilateral increase in the price of X, taking account of recapture in the sale of W through the corresponding diversion ratio D_{12}. A second option would be to use GUPPI directly to calculate the likely price impact of the merger directly, ignoring any possible efficiencies from the merger. Farrell and Shapiro (2010b) and Hausman, et al. (2011) provide formulas for doing so that rely on the assumptions of a linear demand curve and symmetric cross-price elasticities of demand along with estimated values for six parameters: the prices and margins for the two goods and the two diversion ratios. For good 1, the post-merger profit-maximizing price change equals the following:

$$(7)\ \Delta Q_1 / Q_1 = [D_{12}(P_2 - C_2) + D_{12}D_{21}(P_1 - C_1)]/[2(1 - D_{12}D_{21})P_1]$$

and correspondingly for good X. Again, what the formula makes most clear is one of the most important additions of the 1992 Guidelines to the 1982 and 1984 Guidelines: a significant incentive to increase price post-merger requires not only significant diversion ratios but also significant operating margins on the good or goods to which demand is diverted. Figure 1 shows makes this point graphically. A price increase for the first good moves out the demand curve for the second good. The "value of diverted sales" is the rectangle that represents the product of a) the volume of diversion from the first good to the second and b) the margin earned on the second. Only if both the base and the height of the rectangle are of non-trivial magnitudes is the area of the rectangle "large".

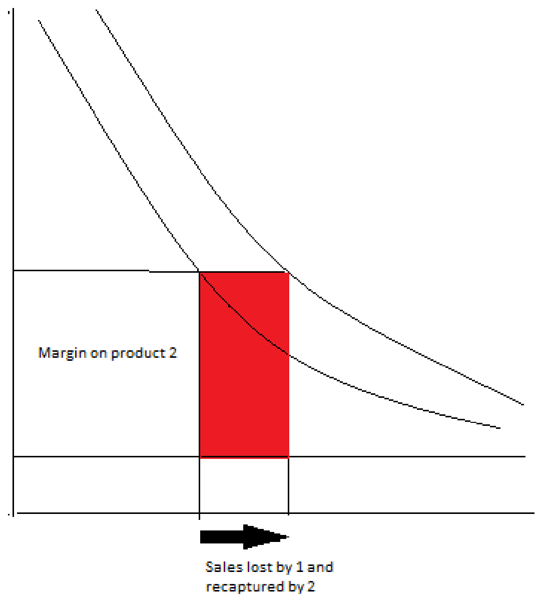


Figure 1. Value of diverted sales

Finally, information on diversion ratios and margins may be used to estimate the merger-specific marginal cost reductions ("efficiencies") that would be required to counterbalance the upward pricing pressure generated directly by the merger and so create a situation of unchanged pricing incentives for the merged firm.

UPP, Upward Pricing Pressure, adds estimates of post-merger marginal cost savings to the GUPPI calculations in order to calculate a measure of the "net" (efficiencies) incentives for the merged firm to increase prices. The formula for merger UPP for good 1 is as follows:

$$(8)\ \Delta P_1/P_1 = \{[2\boldsymbol{D_{12}}\,(\boldsymbol{P_2} - \boldsymbol{C_2}) - E_2\,(1 - (P_2 - C_2)(D_{21} - D_{12}))](\frac{P_2}{P_1}) + D_{21}(D_{21} + D_{12})(P_1 - C_1) - E_1[1 - (P_1 - C_1))(2 - D_{21}(\,D_{12} + D_{21})]\}/[4 - (D_{21} + D_{12})^2]$$

The Diversion Ratio

A requirement for making use of all these concepts is the estimation of diversion ratios. Unlike prices and (variable if not marginal) costs, these are not to be found in the account books kept by the companies. An approach, is to use the shares of competitors in a candidate or provisional market to estimate diversion ratios (Willig, 1991; Shapiro, 2010).

But there are often other good sources of information to guide the analyst in both evaluating the accuracy of market shares as indicators of diversion ratios and in judging how these estimates of diversion ratios might be adjusted to better reflect market realities. As Farrell and Shapiro (2010a) note: The diversion ratio might be estimated using evidence generated in the merging firms' normal course of business. Firms often track diversion ratios in the form of who they are losing business to, or who they can win business from. Consumer surveys can also illuminate diversion ratios, as can information about customer switching patterns. Interviews with, and documents supplied by, customers of the firms may yield subjective but informative information as to the particular qualities of differentiated products that make each a closer or more distant substitute for others as well as objective reports of past switching events and their rationales.

Applying Simulations to Competition cases

The analysis here is based on Epstein and Rubinfeld (2001) and Ivaldi and Verboven (2005). The advantage of this approach is that the model can be calibrated with relatively little information. We need to know:

the market shares of the firms in the relevant market, elasticity of market demand, own price elasticity for one firm and efficiency gains (claimed by firms). By applying a model to the data that we have, we can simulate the future effects of the merger.

The model makes the following assumptions:

Proportionality: if a firm raises its price, it loses market share; this lost market share is allocated to the other firms in the industry proportionally to these firms' market shares

Homogeneity: if all firms in the market raise their price by the same percentage, market shares are unaffected

Market shares of all firms (brands) in the market add up to 1.

The market share of firm i is defined as $si = \frac{piqi}{PQ}$ where pi denotes i's price, qi its quantity, Q total output on the market and P the aggregate industry price index.

Firm i´s demand in this model is written in terms of its market share:

$$si_i = a_i + b_{ii}ln(pi) + b_{ij}ln(p_j) + b_{ik}ln(p_k)$$

We only need to know:

- $\varepsilon = \frac{dln(Q)}{dln(P)}$ market demand elasticity
- $\varepsilon ii = \frac{dln(qi)}{dln(pi)}$ firm's own price elasticity
- $si \geq 0 \; for \; i = 1, \ldots, n \;\; with \sum ni = 1$

With this data and the above assumptions, we obtain all coefficients bij and demand elasticities εij. These can be obtained assuming that firms set their price (cost margin) based on $\mu i = -1/\varepsilon ii$, we know εii if we know the price cost margin. We assssume market shares of three firms to be 0,2, 0,3, 0,5 so we obtain the following coefficients and elasticities (see Epstein and Rubinfeld (2001)):

```
table of b coefficients:
```

b-coeff.	firm1	firm2	firm3
firm1	-0.400	0.150	0.250
firm2	0.150	-0.525	0.375
firm3	0.250	0.375	-0.625

```
table of elasticities:
```

elast.	firm1	firm2	firm3
firm1	-3.000	0.750	1.250
firm2	0.500	-2.750	1.250
firm3	0.500	0.750	-2.250

A firm chooses its price level p_i to maximize profits:

$$\pi_i = p_i q_i(p_i, p_{-i}) - c_i(q_i(p_i, p_{-i}))$$

First order condition is:

$$q_i + (p_i - c_q)\frac{\partial q_i}{\partial p_i} = 0$$

which become

$$\mu_i = \frac{p_i - c_q}{p_i}\frac{dp_i}{dq_i}\frac{q_i}{p_i} = \frac{-1}{e_{ii}}$$

The post merger price estimation is the following:

```
table of post merger outcomes:
```

firm	marketshare	margin	price increase
firm1	0.174	0.414	0.138
firm2	0.281	0.425	0.108
firm3	0.546	0.466	0.041

As one would expect, a merger between firms producing substitutes with no efficiency gains leads to higher prices for all firms. The market share of the firm that does not merge (firm 3) increases due to the merger as prices increase more for the merging firms.

Another merger simulation appears in the EC's decision in the Volvo-Scania merger as an example of a merger simulation using nested logit demand system (Case M1672 Volvo/Scania). The Commission noted in its decision that the results of the simulation confirmed the conclusions of the qualitative investigation. The merger between two truck manufacturers had 5 product markets where it appeared that a dominant company with a 50% share would be created in Sweden, Norway, Finland, Denmark and Ireland. The details of the simulation are contained in Ivaldi and Verboven (2005). The research focused on heavy trucks, of two types "rigid" and "tractor". The dataset consists of J × 1 vectors of the products' quantities q, prices p, and a J × K matrix of product characteristics x, including indicator variables for the products' subgroup and group and their firm affiliation. The dataset is for either one market or a panel of markets, for example, different years or different regions and countries. The panel is not necessarily balanced, because new products may be introduced over time, or old products may be eliminated, and not all products may be for sale in all regions (see Björnerstedt and Verboven, 2014).

In the common unit demand specification of the nested logit, consumers have inelastic conditional demands: they buy either a single unit of their most preferred product j = 1,...,J or the outside good j = 0. The potential market size is then the potential number of consumers I, for example, an assumed fraction γ of the observed population in the market, I = γL. Here the potential market size is the potential total budget B, for example, an assumed fraction γ of total gross domestic product in the market, B = γY .

Demand for heavy trucks was modeled as a sequence of consumer decisions (transportation companies) that chose the truck category and then the specific model within that category.

The model used is the "nested logit model" or nested logistic model by stimulating it using minimum squares in three stages. The aggregate two-level nested logit model gives rise to the following linear estimating equation for a cross section of products j = 1,...,J:

$$ln(sj/s0) \;=\; x_j\beta \;+\; \alpha\widehat{p_j} \;+\; \sigma_1\, ln(s_{j|h_g}) \;+\; \sigma_2\, ln(s_{h|g}) \;+\; \xi_j$$

Where $sj \;= \frac{pj\,qj}{B}$, $s_{j|h_g} \;= \frac{pj\,qj}{\Sigma_{Hhg}\,pj\,qj}$, $s\_(h|g) \;= \frac{\Sigma_{Hhg}\,pj\,qj}{\Sigma_{h=1}^{Hhg}\,\Sigma_{Hhg}\,pj\,qj}$

The price variable is $\hat{p}j \;=\; pj$ in the unit demand specification. The variable sj is the market share of product j in the potential market, $s_{j|h_g}$is the market share of product j in its subgroup h of group g, and $s_{h|g}$is the market share of subgroup h in group g. Furthermore, xj is a vector of observed product characteristics, and ξj is the error term, which captures the product's quality that is unobserved to the econometrician. Equation above has the following parameters to be estimated: a vector o mean valuations β for the observed product characteristics, a price parameter $\alpha < 0$, and two nesting parameters σ1 and σ2, which measure the consumers' preference correlation for products in the same subgroup and group. The model reduces to a one-level nested logit model with only subgroups as nests if σ2 = 0, to a one-level nested logit model with only groups as nests if σ1 = σ2, and to a simple logit model without nests if σ1 = σ2 = 0. The mean gross valuation for product j is defined as

$\delta j \equiv xj\beta + \xi j = ln(sj/s0) - \alpha pj - \sigma 1\, ln(sj|hg) - \sigma 2\, ln(sh|g)$, so it can be computed from the product's market share, price, and the parameters α, σ1, and σ2.

The results (see Björnerstedt and Verboven, 2014). show prices before and after the merger (in 1,000 Euro) and the percent-age price change averaged by firm. This information is provided standard, even without the detail option at the end. The merger simulations predict that GM will on average raise its prices by 7.6%, while VW will on average raise its prices by 3.6%. The rivals respond with only very small price increases (with the exception of Ford).

Pre-merger Market Conditions

Unweighted averages by firm

firm code |	price	Marginal costs	Pre-merger Lerner
BMW |	20.194	17.499	0.146
Fiat |	15.277	10.553	0.372
Ford |	14.557	11.923	0.207
Honda |	20.094	17.941	0.128
Hyundai |	12.915	10.849	0.179
Kia |	10.814	8.772	0.207
Mazda |	14.651	12.557	0.156
Mercedes |	25.598	21.569	0.162
Mitsubishi |	15.955	13.825	0.145

Nissan |	15.438	13.259	0.159
GM |	21.054	18.633	0.135
PSA |	16.243	13.533	0.194
Renault |	15.518	12.837	0.203
Suzuki |	9.289	7.226	0.234
Toyota |	14.560	12.430	0.172
VW |	18.990	16.388	0.181
Volvo |	23.167	20.912	0.099
Daewoo |	13.871	11.789	0.170

Prices

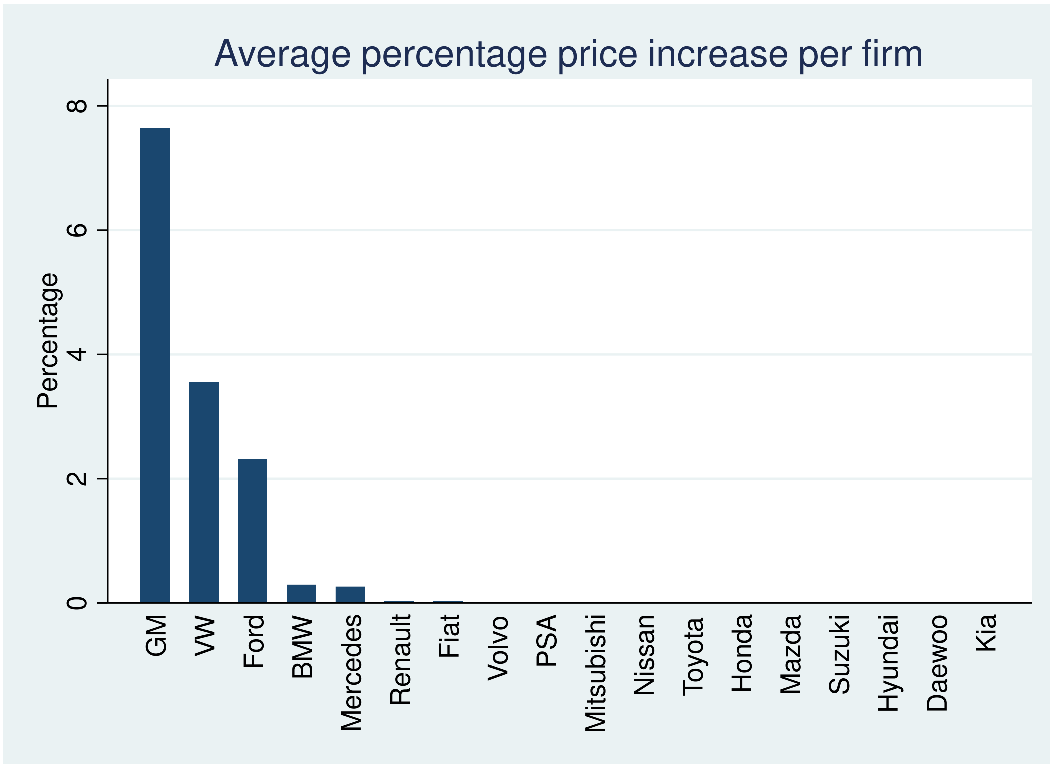
Unweighted averages by firm

firm code |	Pre-merger	Post-merger	Relative change
BMW |	17.946	18.002	0.003
Fiat |	15.338	15.341	0.000
Ford |	13.093	13.362	0.023
Honda |	15.778	15.780	0.000
Hyundai |	12.912	12.912	0.000
Kia |	11.276	11.276	0.000
Mazda |	14.229	14.231	0.000
Mercedes |	20.114	20.155	0.003
Mitsubishi |	15.832	15.834	0.000
Nissan |	15.101	15.103	0.000
GM |	19.921	21.054	0.076
PSA |	16.397	16.399	0.000
Renault |	15.292	15.295	0.000
Suzuki |	9.225	9.225	0.000
Toyota |	13.019	13.020	0.000
VW |	17.182	17.739	0.036
Volvo |	22.149	22.154	0.000
Daewoo |	13.483	13.484	0.000

Market shares by quantity			
Unweighted averages by firm			
firm code	Pre-merger	Post-merger	Difference
BMW	0.074	0.079	0.005
Fiat	0.043	0.045	0.003
Ford	0.095	0.132	0.037
Honda	0.012	0.012	0.001
Hyundai	0.006	0.006	0.000
Kia	0.003	0.003	0.000
Mazda	0.025	0.027	0.002
Mercedes	0.100	0.116	0.017
Mitsubishi	0.015	0.017	0.001
Nissan	0.025	0.027	0.002
GM	0.166	0.108	-0.058
PSA	0.034	0.037	0.003
Renault	0.051	0.054	0.003
Suzuki	0.006	0.006	0.000
Toyota	0.027	0.029	0.002
VW	0.300	0.280	-0.020
Volvo	0.012	0.013	0.001
Daewoo	0.006	0.007	0.001

	Pre-merger	Post-merger
HHS:	1501	1972
C4:	66.07	71.50
C8:	86.21	88.01

Change	
--	
Consumer surplus:	**-1,839,750**
Producer surplus:	**1,303,353**
--	



Bibliography

Amelio A, de la Mano M, Godinho de Mator M (2008) Ineos/Kerling merger: an example of quantitative analysis in support of a clearance decision. Comp. Pol. Newsletter 1:6569

Baker, JB (2011) Comcast/NBCU: The FCC Provides a Roadmap for Vertical Merger Analysis. Antitrust 25:36-41

Baker JB, Bykowsky M, DeGraba P, LaFontaine P, Ralph E, Sharkey W (2011) The Year in Economics at the FCC, 2010-11: Protecting Competition Online. Rev. Ind. Organ. 39:297-309

Baltzopoulos A, Kim J, Mandorff M (2015) UPP Analysis in Five Recent Merger Cases. Konkurrensverket Working Paper 2015:3.

Baumann MG, Godek PE (2009) Reconciling the Opposing Views of Critical Elasticity. GCP: Antitrust Chron. September

Carlton D, Perloff J (2005) Modern Industrial Organization, 4th ed. Pearson/AddisonWesley, Boston

Cheung L (2016) An Empirical Comparison Between the Upward Pricing Pressure Test and Merger Simulation in Differentiated Product Markets. J. Comp. Law & Econ. 12:701734.

CRA (2005) "Vertical arithmetic": The use of empirical evidence in vertical mergers. CRA Competition Memo, Charles River Associates, http://ecp.crai.com/publications/vertical_arithmetic.pdf

Farrell J, Shapiro C (2010a) Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition. B.E. J. of Theoretical Econ. 10:9

Farrell J, Shapiro C (2010b) Upward Pricing Pressure and Critical Loss Analysis: Response. CPI Antitrust J, February.

Fisher FM (1987) On the misuse of the profit-sales ratio to infer monopoly power. RAND J. of Econ. 18:384-396

Fisher FM, McGowan JJ (1983) On the misuse of accounting rates of return to infer monopoly profits. Amer. Econ. Rev. 73:82-97

Harris B, Simons J (1989) Focusing Market Definition: How Much Substitution Is Necessary? Research L. & Econ. 12:207-226 28

Hausman J, Moresi S, Rainey M (2011), Unilateral effects of mergers with general linear demand. Econ. Letters 111:119-121

Hüschelrath K (2009) Critical Loss Analysis in Market Definition and Merger Control. European Competition J. 5:757-794

Kaplow L, Shapiro C (2007) Antitrust. In: Polinsky AM, Shavell S (ed) Handbook of Law and Economics, v. 2, Elsevier

Katz ML, Shapiro C (2003) Critical Loss: Let's Tell the Whole Story. Antitrust spring 4956

Langenfeld J, Li W (2001) Critical loss analysis in evaluating mergers. Antitrust Bull. 299337

Moresi S (2010) The Use of Upward Price Pressure Indices in Merger Analysis. Antitrust Source, February

Moresi S, Salop SC (2013) v GUPPI: Scoring Unilateral Pricing Incentives in Vertical Mergers. Antitrust L. J. 79:185-214

Pakes A (2010) Upward Pricing Pressure Screens in the New Merger Guidelines: Some Pro's and Con's. Presented at DG Competition Authority, Brussels, http://scholar.harvard.edu/files/pakes/files/sdgcomp_0.pdf .

Pittman R (2009) Who Are You Calling Irrational? Marginal Costs, Variable Costs, and the Pricing Practices of Firms. https://www.justice.gov/atr/who-are-you-callingirrational-marginal-costs-variable-costs-and-pricing-practices-firms

Three Economist's Tools for Antitrust Analysis: A Non-Technical Introduction Pittman, Russell

Rybnicek J, Onken LC (2016) A Hedgehog in Fox's Clothing? The Misapplication os GUPPI Analysis. Geo. Mason L. Rev. 23:1187-1203

Shapiro C (1996) Mergers with Differentiated Products. Antitrust spring 23-30

Shapiro C (2010) The 2010 Horizontal Merger Guidelines: From Hedgehog to Fox in Forty Years. Antitrust L. J. 77:701-759

Sibley DS, Doane MJ (2002) Raising the Costs of Unintegrated Rivals: An Analysis of Barnes & Noble's Proposed Acquisition of Ingram Book Company. In Slottje, DJ, ed., Measuring Market Power (Elsevier)

Simons JJ, Coate MB (2014) United States v. H&R Block: An Illustration of the DOJ's New but Controversial Approach to Market Definition. J. Comp. L. & Econ. 10:543-580
29

Werden GJ (1996) A Robust Test for Consumer Welfare Enhancing Mergers among Sellers of Differentiated Products. J. Ind. Econ. 44:409-413

Werden GJ (2005) Beyond Critical Loss: Tailored Application of the Hypothetical Monopolist Test. Comp L. 69-78

Werden GJ (2008) Beyond Critical Loss: Properly Applying the Hypothetical Monopolist Test. GCP, https://www.competitionpolicyinternational.com/assets/0d358061e11f2708ad9d62634 c6c40ad/Werden,%20GCP%20Feb-08(2).pdf Willig R (2001) Merger Analysis, Industrial Organization Theory, and Merger Guidelines. Brookings Papers on Econ. Activity: Microeconomics 281-332

7. DAMAGE ESTIMATION

The estimation of damages within competition policy is one of the areas where quantitative analysis has been used most. Most of the work has been done in countries where courts impose fines or compensatory payments based on damages caused by infringing companies. The deterrent effect caused by fines as a result of criminal liability requires fines to be at least as high as the expected additional benefits of undertakings arising from the conduct to be deterred. The expected benefits are difficult to measure and in cases of cartels they tend to approximate the harm caused to consumers. This chapter examines the challenges facing in estimating the damage caused by the exercise of market power caused by a cartel.

7.1. Quantification of damage caused by a cartel

In competition policy it is often presumed that cartels are bad for consumers, by increasing the price and reducing the quantities available on the market. They are therefore illegal in most jurisdictions. However, since it can be very beneficial to participate in them, there is a temptation to collude when market conditions make it possible. Illegality per se does not have sufficient deterrent capacity when it is not accompanied by at least potential punishment that can eliminate the expected benefits of participation in the cartel. Cartels are punishable by ever higher fines and it is a criminal offence in some jurisdictions such as the US or the UK (in Spain art. 226 CP crime fraud tenders). In order for a fine to have sufficient deterrent capacity, its expected value must be linked to the benefits expected by the cartel. In the US it is common to sue participating companies for damages arising from the cartel and in Europe it is developing. (Damage directive [18]must be transposed by the end of 2016)

In these cases it is necessary to quantify the impact of a cartel and calculate the profits generated for companies and the harm caused to consumers. Below we analyze the effects of a cartel and subsequently the different techniques for quantifying damage, the pass on effect and finally determining the duration of a cartel.

7.1.1. Effects of a cartel

According to economic theory, cartels have two effects on well-being: first, they reduce the total well-being generated by the market and second redistribute income from consumers to businesses. The damage caused by a cartel is equal to the loss of consumer welfare due to the

[18] *Directive 2014/104/EU of 26 November 2014 on certain rules governing actions for damages under national law, for infringements of competition law of the Member States and the European Union.* "The purpose of the Directive is to ensure that any victim of a practice or conduct contrary to competition rules can claim from the infringer full re-respect for the damage that such practice may have caused him" (art. 1.1) See Judgment Courage/Crehan *C-453/99*

combination of those two factors. . . In fact, the damages are in practice defined restrictively asoverchargethatconsumers must pay for their purchases which are only part of the losses suffered by consumers.

7.1.1.1 Effects on the welfare of a Cartel

When several companies form a cartel, they coordinate to increase, even maximize, joint profits. If companies maximize joint profits, then the price of a cartel can be approximated by one that sets a monopoly where marginal income is equal to marginal cost. Compared to a competitive market where prices equal marginal cost this reduces quantity and increases prices. Since prices are higher in a cartel, companies can appropriate the consumer surplus that would go to consumers in a competitive market. In addition, the reduction in the aggregate amount produced causes total well-being to be reduced and results in a loss of well-being. The consequences of a cartel on what could be a ser un competitive market are reflected reflejadas in the following diagram (Figure 7.1.)

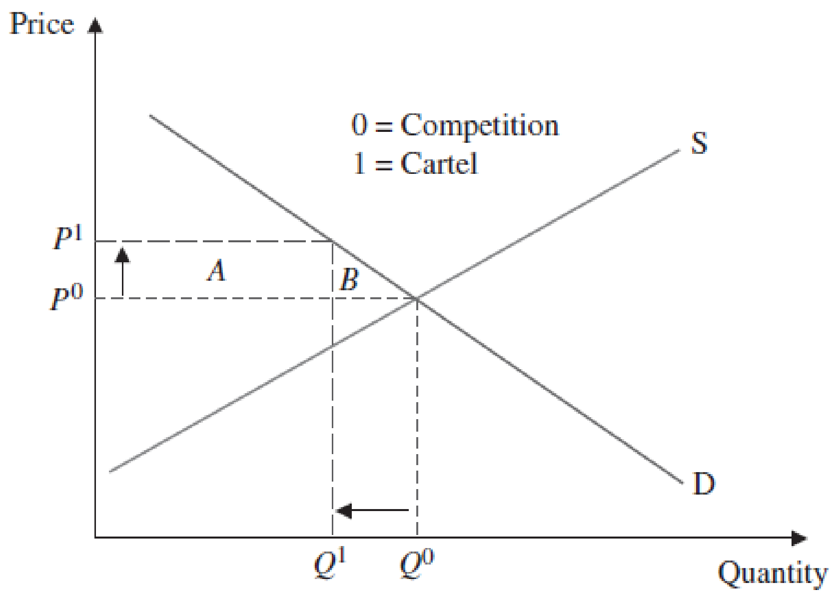


Figure 7.1. Welfare effect of a cartel.

Area A is the transfer of income from consumers to producers as consumers pay P1 instead of P0 and buy only Q1 instead of Q0. Area B represents the net loss of welfare, representing the loss of the consumer who has not been captured by the cartel. The total loss of well-being is therefore B due to the creation of a cartel. Total consumer damage is represented by A+B. The benefit of the cartel for a company is represented by A. If the total loss for the consumer is A+B, area B is usually ignored in the analysis. Although both regions should be estimated, in practice damages are often identified as the illegal appropriation of profits by companies represented by Area A. For practical reasons we assume that the illicit benefit and damage to the consumer is identical which is often referred to as *overcharge* or overprice. The overprice of a unit is the difference between P_1 and P_0. The total overprice is therefore, . $Q^1(P^1 - P^0)$ This approach is not unsatisfactory, in general, provided that the loss reflected in Triangle B is reduced in relation to the size of the transfer of consumers to companies associated with Area A.

7.1.1.2. Direct and indirect damage

Most cartels between undertakings are between undertakings that have vertical relations between them. To understand the consequences of this situation, consider the case of a downstream company being the consumer of a cartel industry,in which the price of the cartel is or affects the marginal cost of downstream companies. According to Van Dijk and Verboven (2007), we see that the damage to a downstream company can be broken down in three terms:

- A first element describes the reduction of downstream profits from the purchase of the inputs of a cartel. This is the direct oversteer of a posterized input..

- The second element describes the margin lost in units that can no longer be acquired under the cartel. Without the cartel we would have sold an extra $(q^0 - q^1)$ unit and gained a margin on $(p^0 - c^{comp})$ them. This effect on output is accounted for in damage calculations.

- A third element, is the increase in profits achieved by loading a higher downstream price and captures the pass-through effect of the increase in the cost per cartel to downstream customers. This is known as the pass-on effect that attenuates the damage suffered by a downstream company. It is also called the indirect effect on end consumers because it measures the overprice or injury suffered by those end consumers more than the current customer of the cartel, which is the downstream company. The treatment of indirect effects on both the calculation of damage to intermediate undertakings or the calculation of potential damage to the final consumer is determined by the legal framework.

Formally, the profits of a downstream company under a cartel can be expressed as follows:

$\pi^1 = (p^1 - c^{Cartel})q^1,$

Where Superscript 1 indicates the company's downstream prices, quantities, and profits under the cartel. In competition in the upstream market, the downstream company will have the following benefit,

$\pi_0 = (p^0 - c^{Comp})q^0,$

Where the superscript "0" indicates the prices, quantities, and profits of the intermediate company under competition. The difference between the two downstream benefits:

$\pi^0 - \pi^1 = (p^0 - c^{Comp})q^0 - (p^1 - c^{Cartel})q^1.$

By manipulating the equation, we get the expression of the difference in profit obtained with the three terms above:

$$\begin{aligned}
\Delta\Pi &\equiv \pi^0 - \pi^1 \\
&= (p^0 - c^{Comp})q^0 - (p^1 - c^{Cartel})q^1 + (q^1(c^{Comp} - c^{Comp}) + q^1(p^0 - p^0)) \\
&= -q^1(c^{Comp} - c^{Cartel}) + (q^0 - q^1)p^0 - (q^0 - q^1)c^{Comp} + q^1(p^0 - p^1) \\
&= -q^1\Delta c + (\Delta q)(p^0 - c^{Comp}) + q^1(\Delta p).
\end{aligned}$$

7.1.1.3 Empirical evidence

It is important to calculate the damages of a cartel in order to estimate the most appropriate compensation to give to victims or also to estimate the illicit profits of the cartelindustry, the proceeds of collusion, for the purpose of imposing an adequate fine. In any case, to quantify the damage there are several empirical and conceptual problems.

To get started, you have to define what we're going to quantify. In many cases, damages can be defined as the overprice to be paid by direct consumers of cartelized companies. This damage will be the lower level of the cartel's actual damage at any time from the reduction of the amount sold ignoring Triangle B.

Secondly, the calculation of damage can become very complicated if we add the potential dynamic effects. Dynamic effects can increase damage if competition had increased investment in R&D or quality. On the other hand, if the high profits had led to higher spending on quality or R&D then at least in principle the damage could be mitigated as incentives to innovate in a cartel could be taken into account. Generally, they are not usually taken into account but it is a possibility although very difficult to prove. In general, posters are often considered to cause long-term damage. It should be noted that these negative dynamic effects can occur and in those industries where they are important they could serve to aggravate the estimated damage caused by the cartel.

The hardest part of the analysis is the quantification of overprice. Calculating that element requires estimating the price in the case of an upstream competitive market. There are several techniques available to build the"but for Price" or prices that would have prevailed if the poster had not existed. Unfortunately, this is a counterfactual something that has not taken place. The duration of the cartel must also be defined since the damage must be calculated for the full duration of the cartel since consumers will be harmed and cartelised undertakings will benefit from the moment prices increase and as long as prices remain high. Since damage occurs over a long period of time, this must be transferred in real terms so that the fine is equivalent to the damage caused. In the next section we analyze the quantification of direct damage.-

7.1.2. Quantifying direct damage

To quantify the damage we need to estimate the price that could have occurred in the absence of a cartel during the cartel period. The price we need is not to observe why we should make assumptions and speculate since it is a prediction. Each method involves some assumptions that must be explained by the researcher. The reasonableness of the cases in particular in each case depends on the specific case, but when a cartel succeeds in increasing the price, the effects of the cartel must be analysed from the point of view of various methods, since this ensures that all the data available for the estimation of the price are used. Like any econometric model it is very important to test to see the robustness of the result in the face of changes in the specification, and there is no econometric exercise that is 100% robust.

The damage quantification exercise should be supported by an in-depth analysis of the particular industry where it is applied which helps to reinforce the justification of the methodology used. In addition, econometric results must be plausible in the context of industry.

7.1.2.1 Using a competency model

Given an econometric model that relates price to industry structure, it is possible to analytically derive the effect of moving from a competitive situation to a cartel. For example, under perfect competition without fixed costs the price will be equal to or almost equal to the marginal cost. The overprice of a cartel formed in that industry will be the difference between the price observed during a cartel and the marginal cost of the industry. The poster price is observable and the competitive price can theoretically be calculated if we have cost information. If the costs change during the cartel period the prices that would have prevailed during the competition period also change during the cartel period.

Specifically, we use the equations of perfect competition and poster. If we assume that the marginal cost is c_t and the linear demand reverse curve is , then $a_t - bQ_t = p_t$ maximizing the profit of a cartel means matching the marginal profit to the marginal cost:

$$\operatorname{Im} g_t(Q) = c_t \Leftrightarrow a_t - 2bQ_t = c_t \Leftrightarrow Q_t = \frac{a_t - c_t}{2b}.$$

If we replace this result in the demand equation we get the prices under the poster:

$$p_t = a_t - bQ_t = a_t - bQ_t = a_t - b(\frac{a_t - c_t}{2b}) = \frac{1}{2}a_t + \frac{1}{2}c_t.$$

In perfect competition the price would be $p_t = c_t$ and the amount of balance would be such that. $p_t = a_t - bQ_t$ The unit price envelope shall be the difference between the price and the marginal cost:

$$Sobreprecio = p^{Cartel} - p^{Comp} = \frac{1}{2}a_t + \frac{1}{2}c_t - c_t$$

In many cases, the mostrealistic "but for"scenario would be a Cournot oligopoly or other oligopoly. The analysis is complex because the amount produced by the cartel companies and the equilibrium price that would prevail in the absence of the cartel will be sensitive to changes in demand as companies take demand into account when setting prices in Cournot or under a cartel. Prices in oligopolistic markets may be less stable than in a perfect competitive situation.

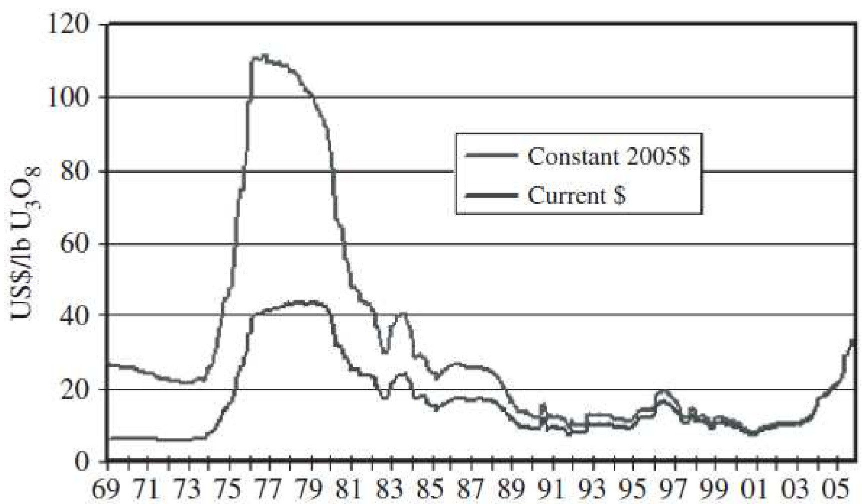


Figure 7.2. Price time series in suspected cartel. *Source*: uxc.com. The price of uranium 308. The reader may wish to speculate when the period of the cartel was.

7.1.2.2. Before and after

The before-and-after methodology uses historical series of prices of cartelized products as the main source of information. It analyzes prices before and after and compares them to prices that prevailed during the cartel. Damages are then calculated as a difference between cartel prices minus competitive prices multiplied by sales during the cartel:

$$Daños_5 = (P_t^{Cartel} - P_t^{Comp})Q_t^{Cartel}.$$

This is an extremely simple method but it can give us enough information in cases where the poster is stable and demand and supply does not change much. Figure 7.2. shows a time series of prices. The method is to join the price before and after the sign with a straight line. In cases where there is an underlying trend in the data, the trend can be taken into account in determining the prices that the market would hypothetically have in perfect competition. In the case of uranium there appears to be a declining price trend just before the cartel (constant 2005 dollars). When competition is re-established, prices return to a slightly lower level in real terms than during the cartel. In this case, a simple calculation of damages by this simple method would be the area above a straight that unites the competitive price of $21 in 1974 and the competitive price of $18 in 1989. Although the cartel lasted from 1972 to 1975, high prices remained much longer. It is therefore also necessary to examine whether these higher prices persisted because of coordination agreements have been extended at a stage of tacit collusion or perhaps costs increased greatly during that time for other reasons. Therefore, depending on the opinion of the courts on the final price of coluse prices,the calculation of damages has a very different appearance.(see Taylor and Yokell 1979)

Some price series are even harder to interpret. Figure 7.3 for example shows the prices of lysine, an additive between 1991 and 1996. The figure shows periods of high falls along with increases. From the chart we cannot infer the price in competition or the period of the cartel.

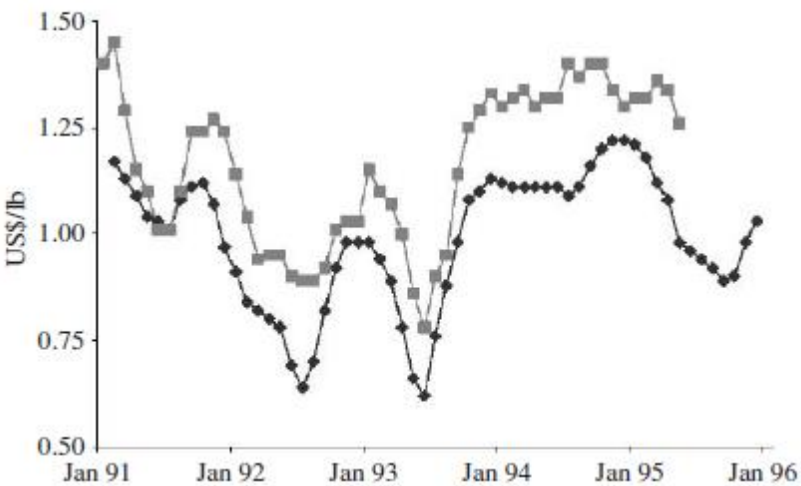


Figure 7.3. Lysine transaction prices in the U.S. and EU markets 1991–95.
Source: Connor (2008).

According to the facts of the case (see Connor 2004), it can be inferred that there is a first moment of price increase in 1992 followed by a collapse in 1993 of the cartel that restarted in 1993. The poster continues until 1995 where the investigation is made public. The applicants argued at the trial that the competitive price was May-June 1992 and April –July 1993 and the defendants considered aggressive competition not to be the most likely balance in such a concentrated and oligopolistic market..

The but but– for model requires assuming that market conditions do not vary, otherwise the methodology would be incorrect.

7.1.2.3.Ultivarian Mapproximation

These problems can be corrected from the above method taking into account changes in demand and supply conditions. If we run a price regression on demand and costs that affect prices but are not controlled by the cartel and also including a dummy variable by poster time. The dummy variable would capture the magnitude of the unexpected price increase that occurs during the poster period. Regression runs as follows:

$p_t = \alpha + \nu D_t + x_t \beta + \varepsilon_t$,

Where D_t it is the dummy variable that takes the value 1 if the poster is active during period t and 0 for the rest of the period and is the demand and cost vector that affect the price but are not x_t controlled by the poster. The coefficient ν gives us the over-price for the period.

Experts defending the respondent will want to include many variables to x to reduce the significance of the coefficient v and therefore the damage. It is necessary to avoid including irrelevant variables in the equation, particularly those that have spurious correlation with the dummy variable and should be robust to D_t small variations in the regression specification.

This method raises the question of whether the impact of a poster can be captured by a discrete increase in the price upwards during the poster. The coefficient ν in the dummy variable will measure the average price increase during the cartel period, regardless of the movements in market conditions that may have occurred. Changes in supply and demand may impact the cartel on prices, and a richer equation specification can capture a more complex effect. A Dummy variable assigns the same magnitude to the poster effect to each year and will give only an average for the entire period. Therefore, it is necessary to at least compare the results with other alternatives to check the robustness.

A second method is to predict the price in the absence of the poster. Using the pre-cartel and post-cartel data, you can estimate the price effects of demand and cost determinants. These parameter values can be used to but predict the but-for price during the poster. The difference between the current price and the price prediction gives us the estimated overprice. This method unlike simple but for allows for change in demand and supply conditions, although structural relationships between variables remain unchanged. In particular, it assumes that the conduct of companies and the way in which demand and costs affect prices remain stable. This assumption is violated if there is a technological change or a change in consumer preferences.

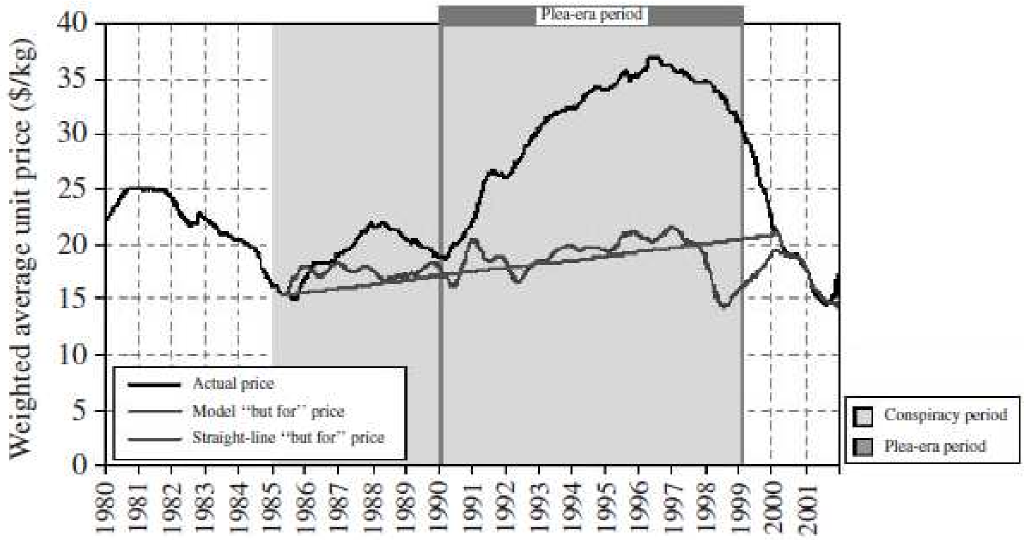


Figure 7.4. Vitamin E acetate oil USP price and "but for" price.
Source: Figure 14.2 of Bernheim (2002), also cited in Connor (2008).

This type of estimate was made in the case of the vitamin poster The estimated price in the absence of the poster was estimated with this regression:

$$P_t = \alpha P_{t-1} + \beta x_{t-1} + \varepsilon_t ,$$

Where P_t is the price of vitamins in the month t and are the supply and x_t demand variables. On the supply side, the exogenous determinants are the price of the raw materials needed to make the vitamins, the industry wage index, the interest rate, and the exchange rates of the currencies where the manufacturers are located. Also included are determinants of demand: population, per capita income, animals that consume these vitamins, amount of drugs produced, quantity of products that use vitamins and the price of substitutes. The delayed variable also appears, which introduces dynamicity to the equation. Bernheim (2002) estimated this equation using data from before the poster and 12 months later (see estimated prices in chart 7.4)

7.1.2.4 Yarsticks

When a cartel does not appear to be equally stable over the years or when demand and supply conditions have fluctuated significantly during the cartel, extrapolating the but for price from before or after prices will not produce the same result. An alternative method is to choose a price of a related product, which is not included in the cartel, and use it as a reference to build what would have happened at the price of a well-cartelized in case of competition. A price will be good reference if the product is closely related to the product under cartel. It should be similar in terms of demand, cost, market structure, in the same region, country. The market should behave similarly to how the cartelized market would behave if it is not cartelized..

The example of a cartel in the steel market(Sevy 2005), where a sign was detected and the following regression was used:

$$\begin{aligned}\text{Price}_{ijklt} &= \alpha + \beta\,\text{Costs}_{ijklt} + \gamma\,\text{Demand}_{ijklt} + \delta\,\text{Bargaining power}_{ijklt} \\ &\quad + \lambda\,\text{Discussions}_{ijklt} + \theta_k\,\text{Trend}_t + \rho_i + \eta_j + \mu_k + \tau_t + \varepsilon_{ijklt},\end{aligned}$$

MARKET POWER AND WELFARE LOSS

1.- Introduction

The economic literature has exhaustively studied the detrimental effect of monopoly power on resource allocation and consumer welfare, but there are hardly any studies on whether these effects can vary according to consumers' incomes. As exceptions, a first contribution on the effects of monopoly power on the distribution of wealth in the United States stands out (Urzúa 2008, Creedy, Dixon 1998, Comanor, Smiley 1975), while other authors point out the importance of the distributive effects of monopolies (McKenzie 1983). More interesting is the study on the distributive impact of monopoly power applied to the case of Australia (Creedy, Dixon 1998) that has recently been replicated for Mexico, using the same methodology (Urzúa 2008) and in which it not only analyzes the distributive impact of monopoly power but also the regional impact. [19] This last study divides the population into income strata according to their income and then estimates the loss of well-being caused in each stratum by the monopoly or oligopolistic powers. To do this, it is first estimated from a theoretical point of view, the loss of consumer welfare using the variation of the consumer surplus (although Creedy and Dixon use several methodologies in addition to the consumer surplus which guarantees the robustness of their results) estimating for this the price elasticity of the demand of every good subject to monopoly power. The study focuses on a set of assets for which it is presumed that the offering companies have a high market power using the data of amounts and expenditure per family consigned in a Survey of Family Budgets of Mexico (of Presupuestos Familiares, INEEncuesta). The next step is to estimate the elasticities of demand for the goods object of the study using Deaton's methodology that allows estimating the elasticities from the expenditure per family without the need to dispose of the prices of the goods. Finally, the distributive and spatial effects of the market power of some companies are presented. In this note we apply the same methodology developed by Urzúa to the case of Spain but using the data obtained from the Ine F-friendly P Survey.

2.- How to measure the loss of social welfare

As we have mentioned, first of all, we assume that the loss of social welfare is proportional to the loss of the consumer surplus and this proportionality factor is the same for all households. The estimation method based on the equivalent variation also used by Creedy and Dixon, requires the estimation of an indirect utility function that due to its complexity we will not estimate here.

[19] It has recently been applied to the case of South Africa (Njisane, Mann et al.) and Kenya (Silver, Begazo 2015)

We consider that the demand function is linear, so that the net loss of the consumer surplus, B, corresponds to half of the price difference multiplied by the reduction in the quantity demanded. So that:

(1) $B = \dfrac{(p^m - p^c)(q^c - q^m)}{2}$,

Prices and quantities refer to price and quantity monopoly (or any distortion of competition) and perfect competition. We assume that prices in perfect competition are equal to marginal cost.
The above equation can be approximated as:

(2) $\eta = \dfrac{(q^m - q^c)p^m}{(p^m - p^c)q^m}$ where η is the elasticity of demand. Reordering the above equation we obtain,

(3) $q_m - q_c = \eta\left(\dfrac{p_m - p_c}{p_m}\right)q_m$

Replacing in (1) we get

(3) $B = \left(\dfrac{p^m - p^c}{p^m}\right)^2 \dfrac{p^m q^m(-\eta)}{2}$.

To calculate (3) we need not only an estimate of the elasticity but also the quantity spent of each good (can be obtained from a Survey) and the estimated increase in relative prices due to market power (which depends on the industrial structure of each market). Following Creedy and Dixon, we add that in each market there are *K* identical firms with constant marginal costs, *c,* which behave using a Cournot conjecture.
Under these assumptions in the Cournot[20] equilibrium it is true that:

(4) $p^m\left(1 + \dfrac{1}{K\eta}\right) = c.$

So if p^c = c, then

(5) $\dfrac{p^m - p^c}{p^m} = -\dfrac{1}{K\eta}$.

It can now be substituted (5) in (3) to arrive at the measurement of the loss of well-being:

(6) $B = \dfrac{p^m q^m}{2K^2\eta}$,

In this way an estimate of the price elasticity, the expenditure of each good and the number of oligopolistic companies in that market is required for each good. This formula is applied for each

[20] see p. 219 (Tirole, 1988)

good and then the losses of the *n* goods are added to obtain the total loss of well-being relative to the total expenditure of each household (*and*) as:

$$(7)\ L = \sum_{i=1}^{n} \frac{B_i}{y},$$

It can be used (7) to calculate the distributional impact of the existence of market power throughout the country as follows: first, for each household the total relative loss is calculated and then the average losses relative to total expenditure are calculated among all households in each stratum. Subsequently, once the stratum with the lowest relative losses has been identified, all the losses are expressed again in a relative way, but now with respect to the least affected stratum.

3.- Distributive impact of market power in goods and services

In order to estimate the loss of well-being, we need estimates of elasticities of demand with respect to their prices. The INE's assumption of the INE reflects both household expenditure and the quantities consumed, so the unit value of the goods consumed by each household can be established indirectly. However, following Urzúa, it must be borne in mind that the unit value cannot necessarily be taken as the unit price of the good that all households would face and not just one in particular. Indeed, the good consumed by one household may have a different quality than that of another household, so it is not correct to equate unit values with prices. As an alternative, a standard methodology can be used (Deaton 1987) that estimates from household spending the elasticities indirectly and is used, for example, by Urzúa to arrive at the elasticities of the goods. Although Urzúa's study uses a set of goods previously indicated by the Competition Commission of Mexico, in our case, we concentrate on two goods where we assume the existence of monopoly power and possibly collusion. Instead of calculating the price elasticities of each good, in this case we use those estimated in several studies (Gálvez, Mariel et al. 2016, Santiago Álvarez-Garcí a,Desiderio Romero-Jordán, Marta Jorge-García 2016) that estimate both elasticities and that we use here, in the case of fuel it would be 0.5361 and in the case of electricity of 0.907 both with a negative sign. The other variable that needs to be specified is K, which is the number of companies in the market, and we assume for fuel that K =4 and in the case of electricity K = 5, according to the number of companies offering in the market.

Once the values of K have been established and the elasticities obtained in the previous studies have been used, the loss of the consumer's surplus can be calculated for each market for goods and services according to equation (6). In the case of an oligopoly of K firms with Cournot's conjecture, the existence of an optimum requires that the price elasticity of demand be negative and that:

(7) $\eta < -1/K$.

In case of monopoly ($\eta < -1$).

In the first place, households are classified by income strata, the average income is obtained by stratum as well as the corresponding share of the good in the total expenditure per household. The total relative loss due to the existence of market power of the bidders of each product is then calculated for each household using equation (7) and subsequently the average of the losses relative to the total expenditure has been calculated among all the households in each stratum. Finally, once the stratum with the lowest relative losses (the one with the highest income) has been identified, these have been expressed again in a relative way, but with respect to the welfare losses suffered by the least affected stratum. The results appear in the last column.

Intervalo de ingresos mensuales netos totales del hogar	*gasto medio por hogar total*	*gasto medio por hogar en gasolina*	*Gasto medio por hogar en electricidad*	*proporcion gasto gasolina sobre el total*	*proporción gasto en electricidad sobre el total*	*pérdida de bienestar absoluta gasolina (L)*	*pérdida de bienestar absoluta electricidad (L)*	*Perdida total bienestar*	*perdida de bienestar relativa (l)*
1 Menos de 500 €	11.038,39	658,20	358,47	0,0596	0,0017	0,0511416	0,0009391	0,0520807	1,300323828
2 De 500 a menos de 1000 €	12.136,26	620,60	510,11	0,0511	0,0014	0,0438565	0,0007586	0,0446151	1,113926611
3 De 1000 a menos de 1500 €	16.770,19	957,58	514,45	0,0571	0,0009	0,0489707	0,0005137	0,0494844	1,23550076
4 De 1500 a menos de 2000 €	21.547,41	1.239,02	606,99	0,0575	0,0009	0,0493157	0,0004958	0,0498115	1,243667623
5 De 2000 a menos de 2500 €	25.388,42	1.513,62	510,04	0,0596	0,0005	0,0511307	0,000267	0,0513977	1,283271039
6 De 2500 a menos de 3000 €	29.613,92	1.730,53	645,45	0,0584	0,0003	0,0501169	0,0001903	0,0503072	1,256044003
7 3000 a menos de 5000 €	36.624,96	2.084,45	556,33	0,0569	0,0003	0,0488109	0,0001775	0,0489884	1,22311689
8. 5000 a menos de 7000 €	50.738,82	2.469,64	789	0,0487	0,0002	0,0417441	0,0001165	0,0418606	1,045153687
9. De 7000 a menos de 9000 €	44.937,89	2.118,84	951,24	0,0472	0,0001	0,0404378	0,0000502	0,040488	1,010883324
10. 9000 o más €	63.705,66	2.975,10	0	0,0467	0	0,0400521	0	0,0400521	1

It is observed that the loss of well-being decreases as income increases. Therefore, the loss of well-being caused by market merger has a greater impact on those strata with lower incomes. The relative loss of the stratum with the fewest resources is about 30% greater than that ofthe richest stratum.

4.- Conclusions

In this note we have examined the distributional effects of imperfect competition in the case of Spain using a methodology developed by Creedy and Dixon and applied by Urzua. The methodology used starts from studying the loss of consumer welfare based on the difference between the hypothetical price in a competitive situation and in case of market power. Assuming that firms compete in quantities (Cournot model) and linear demand, the loss of well-being depends only on expenditure, price elasticity, and the number of firms in the market. If we divide by income strata, proportionally the loss of well-being will be greater for lower income levels, which accentuates social inequality.

5. Bibliography

ARGENT, J. and BEGAZO, T., 2015. Competition in Kenyan markets and its impact on income and poverty: a case study on sugar and maize.

COMANOR, W.S. and SMILEY, R.H., 1975. Monopoly and the Distribution of Wealth. *The Quarterly Journal of Economics,* , pp. 177-194.

CREEDY, J. and DIXON, R., 1998. The relative burden of monopoly on households with different incomes. *Economica,* pp. 285-293.

SURVEY OF FAMILY BUDGETS, INE, *INEBASE*.

DEATON, A., 1987. Estimation of own-and cross-price elasticities from household survey data. *Journal of Econometrics,* **36**(1-2), pp. 7-30.

GÁLVEZ, P., MARIEL, P. and HOYOS, D., 2016. Analysis of residential demand for basic services in Spain using a censored QUAIDS model. *Economic Studies,* **43**(1), pp. 5-28.

MCKENZIE, G.W., 1983. *Measuring economic welfare: new methods.* Cambridge University Press.

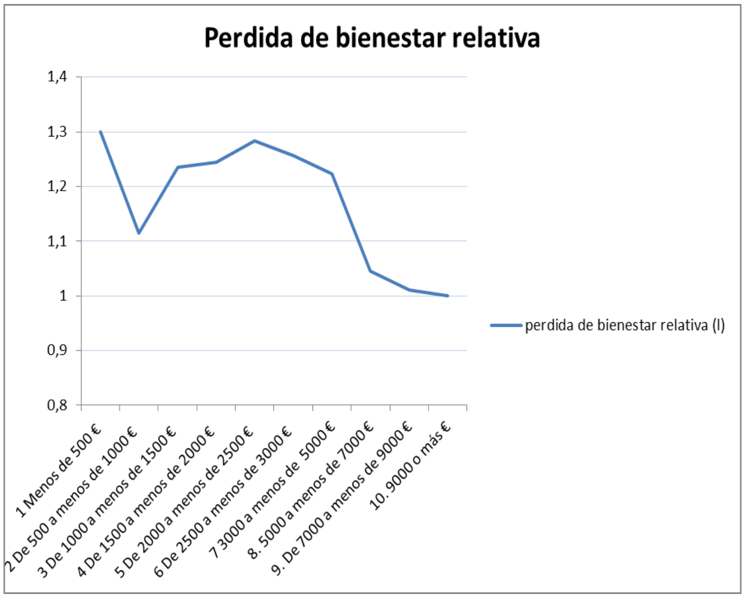
NJISANE, Y., MANN, R., MOJAPELO, K., MATHEBULA, P. and CLANCE, M., The burden of monopolies on South African households with different incomes: How Competition Policy can contribute towards a more equitable society* Preliminary Draft.

SANTIAGO ALVAREZ-GARCIA, DESIDERIO ROMERO-JORDÁN, MARTA JORGE-GARCÍA, January 2016. Exploring the fuel demand of Spanish households: a sensitivity analysis.

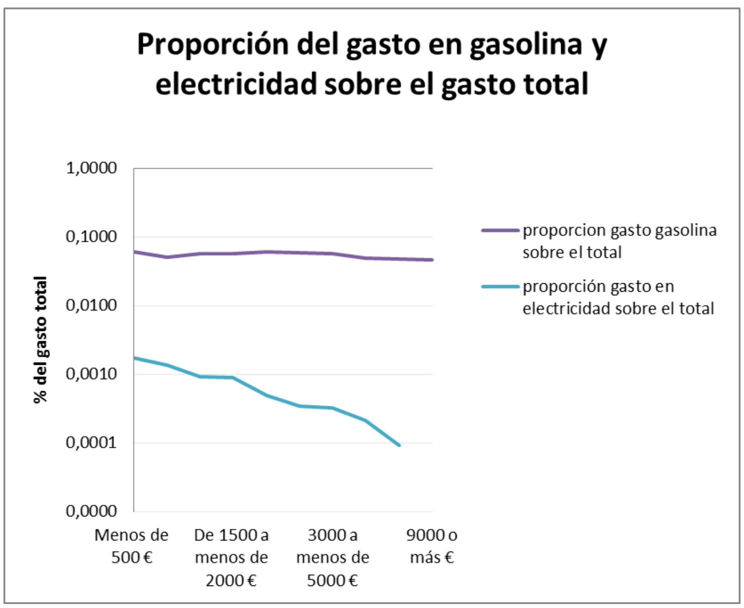
TIROLE, J., 1988. *The theory of industrial organization.* MIT press.

URZÚA, C.M., 2008. Evaluation of the distributional and spatial effects of companies with market power in Mexico. *Mexican Federal Competition Commission*.

Annex I.



Perdida de bienestar relativa
1,4
1,3
1,2
1,1
1
0,9
0,8
perdida de bienestar relativa (I)
1 Menos de 500 €
2 De 500 a menos de 1000 €
3 De 1000 a menos de 1500 €
4 De 1500 a menos de 2000 €
5 De 2000 a menos de 2500 €
6 De 2500 a menos de 3000 €
7 3000 a menos de 5000 €
8. 5000 a menos de 7000 €
9. De 7000 a menos de 9000 €
10. 9000 o más €



Proporción del gasto en gasolina y electricidad sobre el gasto total
% del gasto total
1,0000
0,1000
0,0100
0,0010
0,0001
0,0000
Menos de 500 €
De 1500 a menos de 2000 €
3000 a menos de 5000 €
9000 o más €
proporcion gasto gasolina sobre el total
proporción gasto en electricidad sobre el total

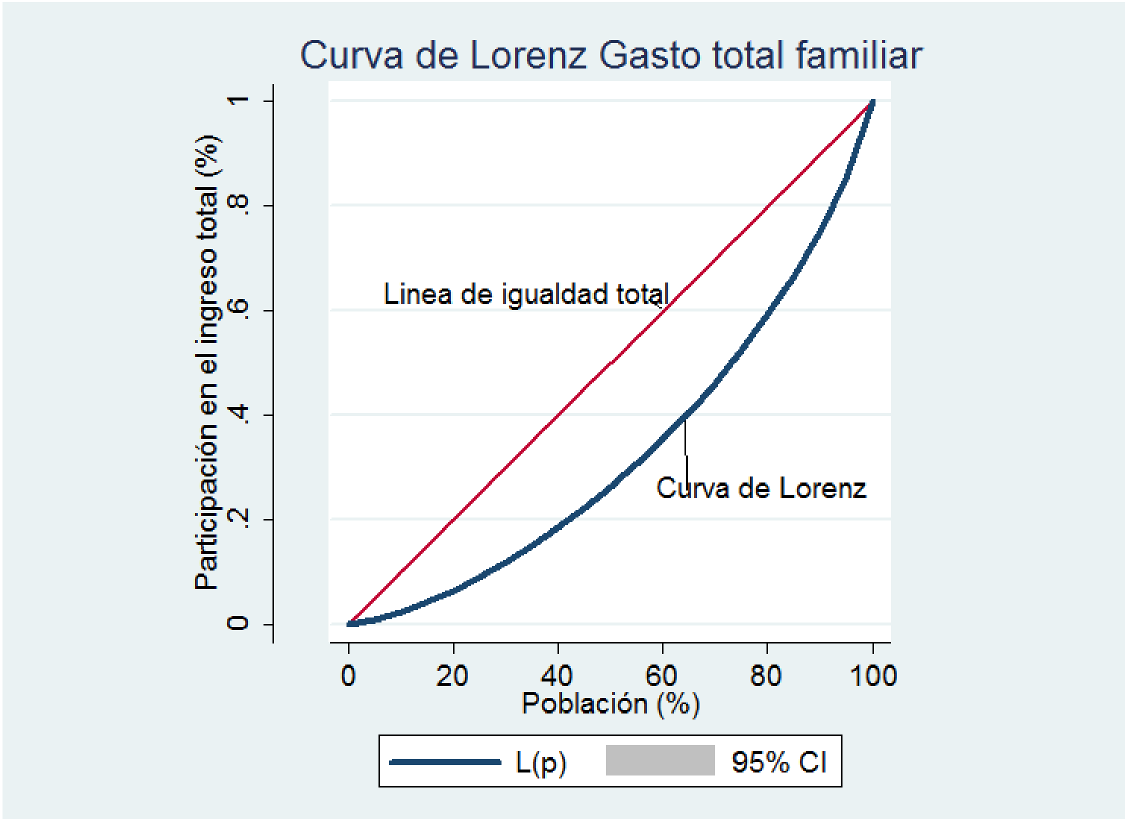
Annex II: Inequality indices

1: Lorenz curve: This is the most basic approach to evaluate the equality of an income distribution and consists of comparing the percentages of total income and the percentages of the population that correspond to those incomes. The idea is Lorenz's () and has become a key reference for the analysis of inequality since it incorporates a very intuitive graph, the Lorenz curve that also has its analytical definition.

The construction of the Lorenz curve is as follows: let be $y = (y_1, y_2, ..., y_n)$ an ordered distribution of rent, where $y_1 \leq ... \leq y_n$. The Lorenz curve is constructed as follows. We first order the cumulative population shares, from the poorest to the richest individual, on the horizontal axis and their corresponding percentages of income on the vertical axis. This produces a dimension 1 box containing a curve describing the distribution of income. The diagonal line represents perfect equality (each proportion of the population receives the same share of total income), so that the difference between this line and the actual distribution is a measure of inequality.

Formally for an ordered income vector from lowest to highest, the Lorenz curve is defined as a set of $[p_i, L(p_i)]$ points, where $p_i = \frac{i}{n}$ is the share of the population with income equal to or less than , y_i and can $L(p_i)$ be defined as:

$L(p_i) = \frac{\sum_{j=1}^{i} y_j}{n\mu}$, with $L(0) = 0, L(1) = 1$ (assuming,in the discrete version, that the intermediate points have been obtained through a linear interpolation). L_i is an abbreviation of $L(p_i)$.



Expenditure per Family	**Coef.**	**Desv. Typical**	**[95%**	**confidence interval]**
0	0			
5	.0094179	.0000131	.0093921	.0094436
10	.0245255	.000026	.0244746	.0245763
15	.0435237	.0000386	.043448	.0435994
20	.065763	.0000502	.0656645	.0658614
25	.0911084	.0000626	.0909857	.091231
30	.1193576	.0000744	.1192118	.1195033
35	.1506756	.0000868	.1505054	.1508457
40	.1851354	.0000987	.1849419	.1853289
45	.2226721	.0001099	.2224568	.2228874
50	.263311	.0001205	.2630748	.2635471
55	.3073042	.0001309	.3070476	.3075607
60	.3548825	.0001406	.354607	.355158
65	.4062984	.0001494	.4060056	.4065912
70	.4622915	.0001572	.4619833	.4625997
75	.523582	.0001637	.5232612	.5239028
80	.5909459	.0001679	.5906169	.5912749
85	.6657701	.0001686	.6654396	.6661006
90	.7509119	.0001639	.7505907	.7512331

95	.8520598	.000146	.8517737	.852346
100	1	.	.	.

The slope of the Lorenz curve between two consecutive points, , corresponds to $[p_{i-1}, L(p_{i-1})], [p_i, L(p_i)]$ the relative income, this $\frac{y_1 / n\mu}{y_i / n} = \frac{y_i}{\mu}$, implies that the slope increases as we move to the right (the richest part of the population). Income distributions can be compared from an ordinal point of view: if the Lorenz curve of one distribution is above another then the first is more egalitarian than the second, since it is closer to the line of total equality. However, such comparisons are difficult when the two Lorenz curves are cut and when there is no way to measure the inequality of one distribution with respect to the other. Therefore, a cardinal measurement is necessary and the most common is the Gini index that we analyze below. However, it should be noted that the Lorenz curve provides very interesting information, for example, for *L(0.5)=0.26* which means that the poorest 50% of the population obtains 26% of the total income. The lower the value of *L(p),* the more unequal the distribution will be.

2.- The Giniindex, proposed by Corrado Gini, at the beginning of the twentieth century () is the most used inequality index probably because it can be easily interpreted from the Lorenz curve. This is the ratio between the area between the Lorenz curve and the 45% line between the entire area below that line. Thus we obtain the Gini index as follows:

$G = \frac{A}{A+B}$, which can be expressed as follows since $A+B=0.5$:

$G=2A$.

$G = \frac{2}{n}\sum_{i=1}^{n}(p_i - L_i)$ indicates that the Gini index is twice the average difference between population and income share in the incomedistribution. For example, in the case of expenditure per household the Gini index would be 0,3428986.

3.- Palma's ratio

A much simpler approach to measuring inequality derives from the comparison between different quartiles of income distributions. The share of total income in the richest 1%, 5% or 10% of the population can be considered. It is usually formulated in ratios: the most frequent are 80-20, 10-90, 20-20 or the ratio between the median and the mean that indicates the asymmetry of the distribution. The advantage of these indicators is that they are very simple to calculate and intuitive, but they only give information on one or two population strata.

The next indicator is due to Gabriel Palma, a Chilean economist who in a 2011 article found a regular empirical relationship in income distributions between countries: middle-income groups defined as those between the decile 5 to 9, receive 50% of total income. The variability observed corresponds to the way in which the other 50% of the total income is distributed in complementary groups: 10% richer and 40% poorer (Palma ....). From these data, it can be deduced that most of the distributional conflict is concentrated in the battle between the rich and poor for half of the cake not enjoyed by the middle class. The palm indicator can be defined as the ratio between the aggregate income of the richest 10% of the population and the poorest 40%. This is a proposal that has received a lot of attention as it provides a simple measurement that is not arbitrary as it has been obtained from a confirmed empirical regularity ()

Annex III: Estimation of Price and Income Elasticities

There are different alternatives to model the demand for goods being the AIDS model, Almost Ideal Demand System, one of the most used (). The model essentially expresses the proportions of expenditure of each good, w_i as a function of prices, and real disposable p_i income:

(1) $w_{iht} = \alpha_{iht} + \sum_j \gamma_{ij} \log p_{jht} + \beta_i \log(X_{ht} / P_t)$ where i =1,... , n

Where X is the total expenditure on all goods and P is a quadratic price index defined as:

(2) $\log P = \alpha_0 + \sum_k \alpha_k \log p_k + \frac{1}{2} \sum_j \sum_k \gamma_{jk} \log p_k \log p_j$

Alternatively, Deaton and Muellbauer () developed a linear approximation (LAIDS) where the price index P is replaced by the Stone index () where:

(3) $\log P^* = \sum_{k=1}^{n} w_{kt} \log p_{kt}$

Substituting (3) in (1) we obtain the equations of the LAIDS model,

(4) $w_{iht} = \alpha_{iht} + \sum_j \gamma_{ij} \log p_{jht} + \beta_i \log(X_{ht} / P_t^*)$ where i=1,... ,n

The parameters $\alpha_{iht}, \gamma_{ij}$ and β_i, are estimated by imposing the following conditions:

i. Additivity: $\sum_{i=1}^{n} \alpha_i = 1$, $\sum_{i=1}^{n} \gamma_{ij} = 0$ $\sum_{i=1}^{n} \beta_i = 0.$

ii. Homogeneity of grade 0 in prices and expenses. $\sum_j \gamma_{ij} = 0.$

iii. Slutsky symmetry: $\gamma_{ij} = \gamma_{ji}$

Macroeconomic effects of Competition policy

October 3, 2021

Abstract

I estimate the macroeconomic impact of competition policy to deter collusion and merger control in the EU using a dynamic macroeconomic model as done by Dierx and Ilzkovit [2]. I then estimate the impact of competition policy using another macroeconomic model that includes Central Bank quantitative easing policies [4]. When these are included the effects on GDP growth are higher.

Introduction

There is a growing interest in the macroeconomic impact of competition and regulatory authorities, but very little research has been done in this subject. [2]. Previous papers estimate the direct impact of competition policy on consumers [2], leaving aside indirect deterrent effects[21]. This paper studies the effects of competition policy on production, employment and productivity in the EU. The direct effects are due to the interventions of the authorities against cartels or anticompetitive mergers, which end situations that would have reduced competition and increased prices. On the other hand, the indirect effects are divided between those that affect productivity, innovation and growth, and the dissuasive effects that are associated with the interventions of the authorities. For example, penalties on collusive agreements not only mean ending the infringement, but also discourage other companies from committing infringements. Because of the complexity of analysis indirect effects are usually excluded from the studies, although there is consensus as to the undoubted benefits of deterrent effects [2].

In this paper I follow [2], to estimate the impact of interventions for the EU between 2010 and 2018. I first explain the model, the mark-up shock simulation and the macroeconomic effects of the mark-up shock. I then explain an extension of the same model that includes Central Bank quantitative easing policies by Priftis and Vogel [4] to conclude with a comparison of results in both models [22].

[21] An exception is [3] who applied a model that includes not only static effects but also dynamic, and the differences between the effects in the short and long term. For this, they used a long-term general equilibrium model and calculated the positive effects that the Competition Agency policies had between 1998 and 2007 on the production, employment and labor productivity of the Netherlands.

[22] The comparative is done using a database of macroeconomic models [6] and Dynare/Matlab [1]. I would like to thank Adrien Dierx et al. and the authors of the Macroeconomic Model Database for their help.

Model

The effects of competition policy are transferred to a dynamic stochastic general equilibrium model as a permanent shock in the mark-up as a result of interventions made by the competition authority to increase the level of competition in the national market. We assume that companies operate in a monopolistic competition market and each company produces a variety of a national product that is an imperfect substitute for varieties produced by other companies.

Aggregate Demand

According to [2] there are two types of households: liquidity- and non-liquidity-constrained households. They possess the same utility function, non-separable in consumption and leisure with habit persistence in both consumption and leisure. Liquidity-constrained households do not optimize, they just consume their labor income. On the other side, non-liquidity-constrained households have access to domestic and foreign currency denominated assets, accumulate capital subject to investment adjustment costs and rent it to firms, earn profits from owning the firms and pay taxes. Income from foreign financial assets is subject to an external financial intermediation risk premium while real asset holdings are subject to an equity risk premium. Both types of households supply differentiated labor to a trade union which sets the wages by maximizing their joint utility (weighted by the share of each type). The wage setting process is subject to a wage mark-up and to slow adjustments in the real consumption wage. The wage mark-up arises because of wage adjustment costs and the fact that a part of workers index the growth rate of wages to past inflation.

Aggregate Supply

The final production of firm j in time t (Y_t^j) used capital (K_t^j) and labour (L^j_t) with Cobb-Douglas production function, with fixed cost (FC_Y^j):

$$Y_t^j = (L_t^j - FC_L^j)^\alpha (u_t^j K_t^j)^{1-\alpha} - FC_Y^j \tag{1}$$

The firm maximizes present value of profits (PR_t^j):

$$PR_{tj} = P_{tj} Y_{tj} - W_{tj} L_{jt} - i_{K,t} P_{I,t} K_{tj} \tag{2}$$

where P_t^j are firm prices, W_t salaries and $i_{K,t}$ cost of capital. In equilibrium where $P_t^j = P_t, \forall j$, firms use a mark up over marginal cost (MC):

$$P_t^j = (1+ \tau_t^j) MC_t^j \tag{3}$$

where τ_t^j is the mark-up over the price, that depends on elasticity of substitution between varieties 4^d, and the mark-up shock $\epsilon_{mpk,t}$:

$$\tau_t^j = 1/(\triangle^d - 1) + \epsilon_{mpk,t}. \tag{4}$$

I then simulate the impact of interventions of a competition authority as a reduction in mark-up through $\epsilon_{mpk,t}$ in the previous equation 4.

Foreign Sector

According to [5] the Foreign Sector: Demand behavior is considered the same for the home country and the rest of the world, therefore export demand and import demand are symmetric. Both equations are characterized by a lag structure in relative prices which captures delivery lags. Export firms buy domestic goods, transform them using a linear technology and sell them in the foreign market, charging a mark-up over the domestic prices. The same situation is faced by importer firms. Mark-up fluctuations arise because of price adjustment costs in both sectors. Markup equations are given as a function of past and future inflation and are also subject to random shocks.

Monetary Policy

Monetary policy is modelled using a Taylor rule, which allows for some smoothness of the interest rate response to the inflation and output gap

Mark-up shock simulation

We can obtain antitrust policy effects on savings by multiplying price reduction as a direct result of competition policy by market size.

Mark-up aggregate change ($4MUP_N$) due to N antitrust measures can be defined as:

$$\epsilon_{mpk,t} = \triangle MUP_N = \sum_k \left[\frac{\triangle P_k}{P_k} (1 + MUP_k) \right] \frac{GO_k}{GO} \tag{5}$$

where K_n are total sectors k where these interventions have reduced prices, equation $4P_k/P_k$ 5 shows that aggregate mark-up is weighted by the relative mark-up in the specific sector $(1+ MUP_k)$ and its share in total production k of the economy, $\frac{GO_k}{GO}$.

Direct effect of antitrust decisions

We can make a distinction between shocks that only have direct effects and other that have deterrent effects on other firms. In the first case, price changes in each sector k is computed as a weighted average of price changes in all affected markets n:

$$\frac{\triangle P_k}{P_k} = \sum_M \frac{\triangle P_n}{P_n} MS_{nk} + \sum_C \frac{\triangle P_n}{P_n} MS_{nk} \tag{6}$$

where $M_k y C_k$ are decisions on cartels and mergers that impact sector k. In each decision Competition Authority defines a relevant market. The weights MS_{nk} used to estimate price changes for each sector is defined as a share in the affected market n in sector $k(mkt_{nk})$ over the total value of production in that sector at a two digit level (GO_k):

$$MS_{nk} = \frac{mkt_{nk}}{GO_k} \tag{7}$$

We estimate merger decisions to reduce prices in 3 per cent and antitrust decisions 10 per cent against non intervention. Equation (6) is also:

$$\frac{\triangle P_k}{P_k} = -0,03 \sum_M MS_{nk} - 0,1 \sum_C MS_{nk} \tag{8}$$

Substituting equations (7) and (8) in (5), mark-up changes due to merger and antitrust decisions can be estimated in the following way:

$$\triangle MUP_N = -\frac{1}{GO} \sum_{k_N} \left[0,03 \sum_M mkt_{nk} + 0,1 \sum_C mkt_{nk} \right] \tag{9}$$

Deterrent effects of antitrust decisions

In general, only direct effects of decisions are estimated leaving deterrent effects due to the complexity of estimating unknown cartels that disappeared without being detected. As in [2] we assume that in each decision of a competition agency, price reduction impacts all the sector. An airline merger decision, for example, would have deterrent effects over total passenger air transport sector. Deterrent effects would impact all firms of the same sector.

When estimating mark-up shocks it is assumed that deterrent effects spill to the whole 4 digit NACE rev2 sector. The weights MS_{nk} are the proportion between the value added at four digit divided by the value added at 2 digit value and is used to calculate the price change in sector k due to n competition decisions.

$$MS_{nk} = \frac{VA4_{nk}}{VA2_k} \tag{10}$$

Size and duration of shock

One can obtain the direct impact of interventions by adding markup changes as a result of the merger and cartel decisions. Given that effects on prices have an impact in several years, consumers will benefit not only from the interventions in that year, but also from those carried out in previous years. [2] take into account decisions of 2015, that is, all the decisions made in 2014, in addition to the decisions made in previous years that still have an impact in 2015. The reduction in the markup MUP_N associated with these decisions is obtained from the equation 9 and added to reach a total effect in 2014 of 0.04 percentage points. This figure includes the effects of anticartel decisions.

However, the simulations presented consider not only the direct effect but also deterrent effects. Using the equation (9), the reduction of markup MUP_N derived from the decisions that still have an impact in 2015 can be calculated: the reduction of the markup is 0.57 percentage points in 2014, which corresponds to a reduction of 4.49 percentage points at the markup level. [2] consider that the magnitude of the shock and the simulation results come essentially from the deterrent effects of competition policy and not from the direct effects.

The competition authority is supposed to continue interventions at the same rate in the near future. This permanent shock can be applied to a baseline scenario where interventions are not performed. The assumption of a permanent shock reflects the idea that a single intervention by the authority will have little effect on firms. The deterrent effects of interventions of a competition authority mainly come from the expectation of firms that the antitrust authority will sanction if competition law is violated.

Macroeconomic effects of a mark-up shock using QUEST III

We apply the same permanent negative mark up shock as in [2] of 0,57 per cent to a perfecct foresight QUEST III model. The increase in consumption due to a fall in prices, will increase real salaries and employment due to competition. On the other hand, investment would be determined by production scale, while increasing factor remuneration. Regarding external demand, export and export increase similarly, due to respectively an increase in competitiveness of firms, and an internal demand increase. From the supply side, it is obtained an increase of employment above GDP increase in the four initial years after the measure is implemented, while on the long run it is slightly lower. This higher increase is due to the higher dynamism of an increase in competition. A more competitive environment will not only reduce prices but increase efficiency as they reach their optimal scale while searching for new resources and lower cost technologies In general, competition will stimulate innovation, technological progress and more efficient ways to provide services to society. The transmission mechanism is as follows: the interventions of the competition authority generate an increase in competition and a reduction in markups which leads to a decrease in prices (Equation 4). As companies think about the future when making decisions today, their demand for labor and capital is based on the future flow of benefits, taking into account the effect of markups on prices and demand. They take into account the direct effect of markups on future benefits, which will be negative due to lower markups, and at the same time they also take into account the increase in future demand for their products due to lower prices. To meet a greater demand, companies require more work and capital. However, the fall in the future profitability of the companies partially mitigates the increase in demand for the inputs since higher production costs and lower pre-payments can lead to lower profits generated by companies (Equation 2) .

Extension of QUEST III including Quantitative Easing (QE)

[4] extend the QUEST III to include assets and QE: QE is modelled as purchases of domestic long-term bonds in exchange for central bank liquidity. Next to physical capital and money, the model features long-term and short-term bonds. The assets are held by the household and long-term and short-term bond holdings are subject to portfolio adjustment costs. When the central bank intervenes by purchasing long-term bonds, private investors that aim at re-establishing the portfolio mix of short term and long-term assets can respond

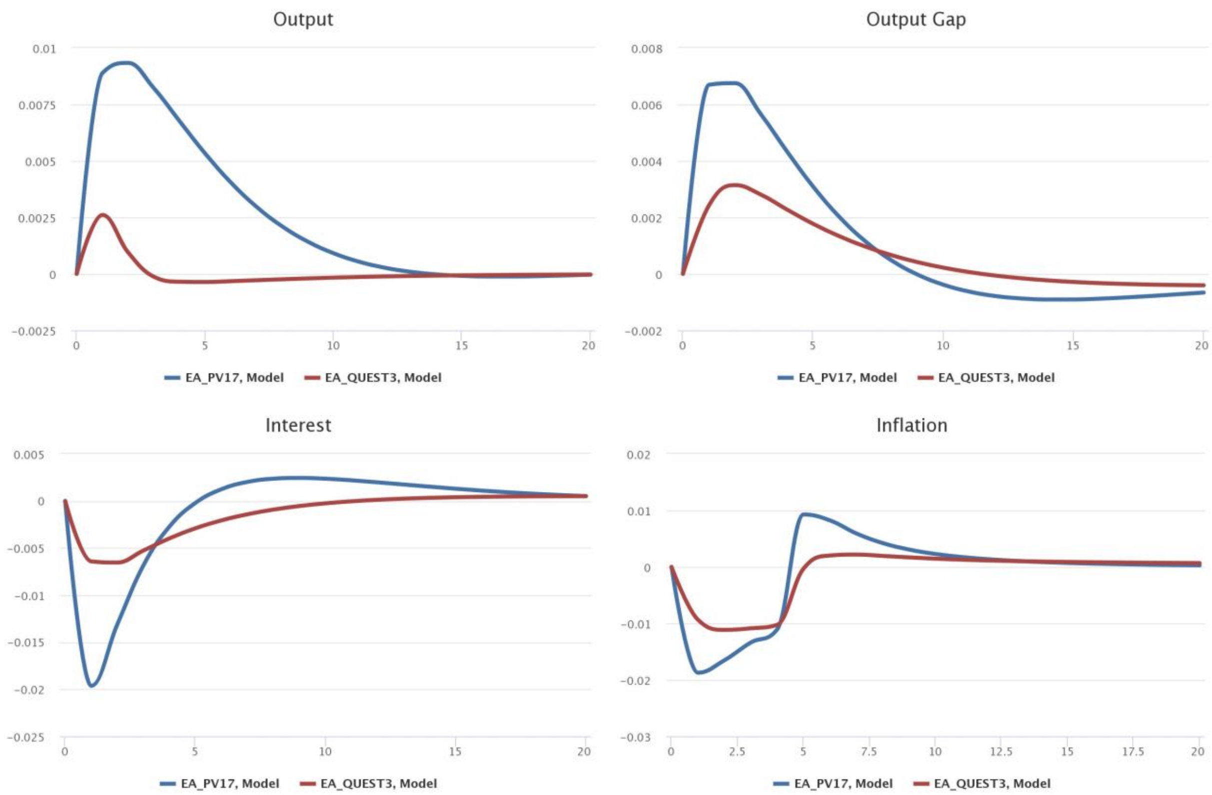


Figure 1: Macroeconomic Effects of the markup shock

by holding more corporate equity and foreign bonds, and by lowering their savings. The first response means portfolio reallocation towards equity and foreign-currency assets that increases the prices of corporate equity (rising stock market) and foreign currency (euro depreciation). Regarding transmission to the real economy, following [4] rising stock markets reduce the financing costs of corporations and lower the required return to capital, which translates into stronger investment and capital accumulation. Exchange rate depreciation strengthens net exports provided trade is sufficiently price elastic. The decline in savings associated with the general decline in returns on savings strengthens contemporaneous consumption demand. The figure and table below illustrate that competition policy interventions increase output, price reduction increase consumption. The table shows that the model that includes bond purchases (PV17) is more expansive than the benckmark model QUESTIII as GDP increases 0.3 percent after five years and 0.55 percent in 20 years. QUEST III estimates are slightly lower, 0.27 percent growth in five years and 0.47 percent in 20 years.

In conclusion, the expansive effect of a mark up shock due to competition policy is slightly greater if Quantitative Easing policies are taken into account than in standard tranditional macroeconomic model.

Conclusion

This essay aims to apply a general equilibrium dynamic model to show quantitative effects of competition policy interventions. Its implementation for impact assessment for antitrust policies is not common. Nevertheless, its dynamic and general equilibrium approach is better suited to explain and forecast the effects of interventions than partial equilibrium static models. Besides, the paper shows that the positive macroeconomic effects due to competition policy interventions are estimated to be higher when quantitative easing are included in the macroeconomic model.

Table 1: Macroeconomic Effects of the markup shock

	1 year		5 years		10 years		20 years	
	PV17	QUESTIII	PV17	QUESTIII	PV17	QUESTIII	PV17	QUESTIII
GDP	0.081	0.156	0.310	0.269	0.486	0.359	0.552	0.470
Inflation	-0.162	-0.007	-0.507	-0.03	-0.220	-0.02	-0.1373	-0.023
Consumption	0.041	0.161	0.184	0.239	0.300	0.327	0.296	0.431
Output Gap	0.060	0.003	0.218	0.0128	0.297	0.0195	0.237	0.019

References

[1] Stéphane Adjemian, Houtan Bastani, Michel Juillard, Fréderic Karamé, Junior Maih, Ferhat Mihoubi, George Perendia, Johannes Pfeifer, Marco Ratto, and Sébastien Villemot. Dynare: Reference manual version 4. Dynare Working Papers 1, CEPREMAP, 2011.

[2] Adriaan Dierx, Fabienne Ilzkovitz, Beatrice Pataracchia, Marco Ratto, Anna Thum-Thysen, and Janos Varga. DOES EU COMPETITION POLICY SUPPORT INCLUSIVE GROWTH? *Journal of Competition Law & Economics*, 13(2):225–260, jun 2017.

[3] Henk Don, Ron Kemp, and Jarig van Sinderen. Measuring the Economic Effects of Competition Law Enforcement. *De Economist*, 156(4):341–348, dec 2008.

[4] Romanos Priftis and Lukas Vogel. The macroeconomic effects of the ecb's evolving qe programme: a model-based analysis. *Open Economies Review*, 28(5):823–845, 2017.

[5] Marco Ratto, Werner Roeger, and Jan in't Veld. Quest iii: An estimated open-economy dsge model of the euro area with fiscal and monetary policy. *economic Modelling*, 26(1):222–233, 2009.

[6] Volker Wieland, Tobias Cwik, Gernot J. Müller, Sebastian Schmidt, and Maik Wolters. A new comparative approach to macroeconomic modeling and policy analysis. *Journal of Economic Behavior Organization*, 83(3):523 – 541, 2012. The Great Recession: motivation for re-thinking paradigms in macroeconomic modeling.

ABUSE OF DOMINANCE:

Analysis of Predatory Pricing

1. INTRODUCTION

Article 102 of the Treaty on the Functioning of the European Union prohibits abuses of a dominant position. According to the case-law, the fact that an undertaking is in a dominant position is not unlawful in itself and the dominant undertaking has the right to compete on the basis of its merits. However, the undertaking concerned has a special responsibility not to impede, by its conduct, the development of effective and undistorted competition in the common market. However, the list of conducts contained in Article102 TFEU does not constitute a closed catalogue but there are a number of abuses not mentioned in the Treaty which do not fall within any of the categories of paragraph 2 but which would fall within the generic concept of abuse contained in the first paragraph.

The notion of market closure strategies includes a wide variety of commercial practices carried out by a dominant undertaking the objective of which is, in general, to prevent the entry or profitable expansion of one or more competitors in the dominated market or in other markets related to the dominated one. The concept broadly encompasses the so-called "exclusion abuses" redefined by the European Commission and would be covered by Article 102 TFEU.

This work focuses mainly on a type of exclusionary abuse: predatory pricing, e.g. those practices of a dominant firm aimed at curbing entry, forcing out rival firms, or limiting the competitive capacity of their rivals. The identification of exclusion abuses is one of the most complex issues in the defense of competition as these practices are difficult to distinguish from competitive actions that seek to improve the welfare of consumers. In the specific case of exclusionary conduct based on predatory pricing, it seeks to analyze whether a reduction in the prices of the dominant undertaking is anti-competitive behavior: whether this behavior is intended to drive new competitors out of the market or is simply a competitive response which will lead to an improvement in the welfare of consumers. These behaviors are receiving increasing attention as liberalization, deregulation and privatization processes are completed in many countries which has led to markets typically with an incumbent company facing several competing rivalsand/or potential entrants, which creates incentives for exclusionary practices. It is added to this that there are more and more sectors (telecommunications, computing or internet) in advanced economies that exhibit" network effects" or network economies and "lock-in effects" or effects derived from the link with a provider. In this environment it can be very complex for an entrant to compete with the incumbent company, so it is necessary to pay special attention to possible abuses of exclusion.

2. PREDATORY PRICING

A predatory pricing situation can be defined as one in which an undertaking in a dominant position sets prices that involve sacrificing short-term profits in order to exclude or be likely to exclude competitors in order to maintain or strengthen its market power, to the detriment of consumers.

This definition contains two key elements: the existence of losses or giving up short-term profits and secondly the purpose of exclusivity or being likely to exclude one or more of its actual or potential competitors in order to strengthen or maintain its market power.

At first glance, one can deduce the difficulties involved in distinguishing lower prices derived from a predatory practice to those derived from the response to greater competition from rival companies.

We must be very cautious when analyzing predatory prices, also taking into account that incumbent companies could adopt a behavior based on high prices to avoid being accused of predatory prices, which can lead to a loss of allocation efficiency and entry of inefficient companies, adding also a loss of productive efficiency.

2.1. ARGUMENTS IN DEFENCE OF THE EXISTENCE OF PREDATORY PRICING

Although cases of predatory pricing have not been uncommon in the US and the EU, there has been no economic theory capable of explaining them until very recently, and thus there is still criticism of predatory pricing arguments. The generally accepted idea was that a large company faced with the entry of new competitors could start a price war that would end up generating losses for all companies, although small rival companies hardly have the resources to survive with losses for a long time. Sooner or later, they will leave the market which will allow the incumbent to increase prices and recover losses.

McGee (1958) already criticized the idea that an incumbent could drive out competitors using predatory pricing through four arguments:

- Due to its larger market size, a large company will end up having greater losses (the same unit loss by having more units will end up having a greater overall loss)

- Predation only makes sense if the incumbent firm raises prices once rivals exit the market. However, the assets and plants of the rivals do not disappear so that once the price has risen, the rivals can re-enter the market or can be acquired by other companies which would reduce the profits that the incumbent could obtain,

- The theory of predation assumes that the one who imposes predatory prices has ample resources and limited rivals. However, this should be explained in more depth rather than simply assumed.

- Rival companies could sustain themselves despite losses through bank financing (would not explain the situation to creditors)

- For predation to be rational it is necessary that it be feasible and more profitable than other alternatives such as merger with the rival company, thus avoiding losses.

The counterarguments to these points are as follows: in the first point, the incumbent company could discriminate prices and reduce them only in those markets where it competes with rivals which would give it a positive margin in almost all the units it sells which would reduce the cost of the predation strategy. The second point does not seem very feasible if we take into account the sunk costs involved in the entry and exit so it does not seem very feasible to leave and re-enter without incurring irrecoverable fixed costs, In addition, rival companies once they leave a market because of predatory behavior will be reluctant to re-enter the same market. As for the third point, it is refuted through the fact that predation can make financial restraint more rigid. In relation to the fourth point, the merger with the rival company may not be a good option because it could attract new competitors who would try to be sold to the incumbent obtaining profits. In addition, competition law could block this merger and finally it does not have to be exclusive of merger and predatory practices because these practices can lead to a purchase at a lower price.

From the arguments and counterarguments related to predatory prices we will then study how economic theory explains predatory prices and how they can take place.

2.2. RECENT THEORIES OF PREDATORY PRICING

In recent predatory pricing models, there is a common basis that is the existence of imperfect information or that situation in which companies have uncertainty. In all cases, the predator will try to use the imperfect information of the incoming one and will behave trying to make her believe that the incoming one will not make profits in the industry. As a result, the incoming company will exit, or its creditors will not wish to provide it with further financing.

For this to take place there must be uncertainty, since in a world where companies have complete information, predation would never be observed: it would be clear that the incumbent will have incentives to fight against entry so that potential rivals will not enter in the first place or will not apply predatory prices because she knows with certainty that they will not succeed.

Three sub-models of predatory pricing were identified:

- Reputation
- Signage
- Financial markets

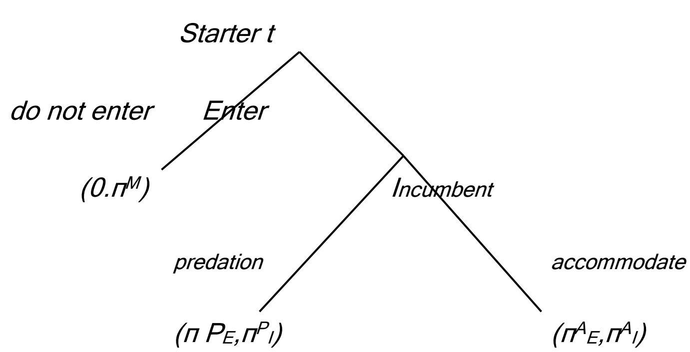
2.2.1. Reputation Models

Mc Gee argued that the conduct of an incumbent vis-à-vis potential competitors may affect future competitors as well. The incumbent company could initiate a price war to create a reputation as a strong and aggressive company to discourage entry.

Selten (1975) shows that a weak incumbent (equal cost as entrants) will allow entrants to settle in the market and will not fight them since it will not be profitable by generating losses for both incumbent and entrants. Therefore, the paradoxical result is that predation will never be observed. The main reason for this result is that the entrants know with total certainty that the incumbent has incentives to accommodate the rival if we take into account a single period. However, introducing uncertainty predation will occur,

Figure 1

Figure 1



Selten focuses on an example where he considers that a company that owns a chain of stores, in T different cities. In each market, the incumbent faces a competitor, for example, company 1 can enter market 1 in period 1, company 2 can enter market 2 in period 2, and so on.

In each period t, the game between the dominant firm and the incoming potential is the same and is illustrated in Figure 1. First, the company t has to decide whether or not to enter and then the incumbent has to decide whether it wants to fight against the entry (i.e., choose an aggressive action from the market), or accommodate it. If the incoming one is left out, then the gains for the incumbent and the incoming power are πM and 0 respectively. If there is entry, and the incumbent chooses to accommodate it, their respective payments are πAE,πAI. If the entry is followed by an aggressive reaction (predation), payments πPE,πPI.

Suppose the fight is costly for both players, in the sense that πAI> πPI and πAE >0> πPE. Suppose also that the incumbent obtains the greatest profits under monopoly: πM..> πAI

If the game is played only once (T = 1), it is evident that the threat of predation in case of entry is not credible, and the entry would take place and be accommodated in balance. Indeed, if the entry occurs, the incumbent prefers to accommodate him to fight against it, since he is πAI> πPI. The entrant correctly anticipates the choice of the dominant, and in the first period it is known that if he enters a weak market reaction will occur, so he will receive a reward πAE, while if he does not enter his profitability it is 0 < πAE. Therefore, he prefers to enter.

Selten's conclusion is to show that nothing changes in this result if the game is repeated several times, if this number is finite. Let us see what happens when the incumbent is faced with several starters.

Consider a game between the incumbent and the potential competitor in the last market T. Regardless of what has happened before (whether the entry occurred or not, whether it has been accommodated or not) the decisions to be taken in this market will have no more effect than on the current payments, since it is the last game. Therefore, companies behave as if they are playing the game the first and only time. The only equilibrium is the one above if T = 1: the incoming one foresees that the input would have a place, and therefore the input occurs.

Let us now consider what happens in the T-1 period, where the incumbent and the T-1 company play in the market in T-1. Again, companies know that regardless of what has happened in a period before and what they do today, in the next period the T entry is going to occur. Therefore, the period T - 1 again has no impact on the future game, and the incumbent must play as if it were the only (or the last) period of play. The T-1 company knows that the incumbent has no reason to fight the entry, and will enter accordingly, and of course the entrance will be accommodated.

The same reasoning will occur for the T-2 and in any previous period, so that the only Nash equilibrium of the game is one in which each entrant will enter its respective market, and each entry will be accommodated.

Contrary to what you would expect, there is no reputation effect on this game, and predation will not occur. Kreps and Wilson (1982) eliminating the assumption of perfect information demonstrate that a weak incumbent will fight against entry from the beginning to establish a reputation for being strong (an efficient company that can set prices below the costs of the incoming one) and thus discourage new entries. In general, the weak incumbent will fight against entry by reinforcing its reputation for being efficient, but it is a sacrifice of current benefits to stop new entrants from obtaining greater benefits in the future. At the beginning of the game, it is convenient for the incumbent to fight against the entry of new rivals, while as the game ends he is no longer interested in fighting but in allowing the entry of new rivals.

2.2.2. Signaling models

Also, these models like the previous one is based on imperfect information. Again the potential entrant has no idea whether the incumbent has low or high costs and the incumbent will try to exploit this uncertainty to avoid further entries.

The first model of signaling is due to Milgrom Roberts (1982b) that we will summarize below: In this model an inefficient or weak incumbent company will try to signal itself as a strong one since if the incoming ones detect that it is a weak incumbent then they will enter before the possibility of obtaining benefits. However, the strong incumbent has no interest in signaling herself as weak as this attracts new entries. Two balances arise in this game:

(a) Separate equilibrium: the incumbent sets a lower price than the normal monopoly in the first period so low that no weak incumbent could fix it because it would entail too high losses. Since then, it cannot be replicated to the strong incumbent then the inefficient or weak will set a monopoly price. This implies that the entrant knows whether the incumbent is strong or weak from the fixed prices: if they are low, it is an efficient incumbent, and there will be no entry, if they are high, it is an inefficient incumbent and then the entry will occur. In the latter case one could speak of predation since the efficient incumbent is sacrificing current benefits to stop the entry and thus earn more in the future. However, this is not the case since social welfare is not reduced by reducing the price in the first period below that of monopoly.

(b) Common equilibrium: In this equilibrium there would be no price at which the low-cost incumbent could sell and differentiate itself from the high-cost enterprise and will therefore set a monopoly price like the high-cost one.

This model is a continuation of the pre-game theory models called limit price theory of Bain (1949) and Sylos-Labini (1979) which refer to the possibility that the incumbent sets a price to avoid entry since it implies low profits.

However, in these signaling models they are not associated with the incumbent setting low prices. In the signaling model there may be a situation in which the incumbent sets high prices to avoid entry because this signals high costs that must also be assumed by the incoming potential which would prevent its effective entry (see Harrington, 1986).

Predatory prices in mergers: an extension of the previous signaling model allows to explain its use to reduce the purchase price of its rivals. Saloner (1987) modifies the Milgrom and Roberts model slightly, to allow companies to merge after the first period (also, in their model, entry into a market where there is an efficient incumbent might not result in losses, only lower profits). In this case, setting a lower price would be an optimal signal for the potential entrant if he must expect high profits after entry, but this time his goal is not to prevent entry but to improve the conditions in which the rival will accept to be acquired.

Other signal models:

There are other models where the incumbent could act strategically to make the incoming woman believe that she expects lower profitability once within the industry. Scharfstein (1984) analyzes a model called "test-market predation", where the entrant has a new product and is uncertain about the potential demand for that product. Given this uncertainty introduces the product in a test to know how it would be received by the market, the incumbent company could incur in various predatory practices (for example secret discounts on the price to certain customers) to make the potential entrant believe that the demand is low inducing it to leave the market. Fudenberg and Tirole (1986) also suggest that the incumbent could enter into a predation called "signal-jamming predation" that stops the entrant from improving their information. For example, in the "test market model" the entrant tries to hoard as much information as possible about the demand, and the predatory company derails this behavior by reducing prices, The entrant knows that her demand is artificially reduced due to the high prices imposed by the incumbent, but cannot access information about the demand under normal conditions. In the absence of information, the incoming company will prefer to exit the market,

2.2.3. Predatory prices in unfavorable markets

As we have seen above, one of the weaknesses of the predatory pricing fundamentals was that it does not explain why incoming or rival companies have limited access to financing, since with perfect capital markets a profitable project will always have financing available.

However, more modern theories of financial markets focus more on the imperfections of capital markets, which allows to better explain the difficulties of access to financing for rivals. With this modeling, access to financing will depend on whether the company is a victim of predatory actions since these actions affect the risk associated with a loan to this company.

The key point of this theory is the existence of imperfect information for lenders (they do not have accurate information about the industry that applies for the loan) which implies restrictions on credit which can cause a profitable project to not be financed. Translated this into the world of competition between a well-established company that has accumulated enough resources in the past and a potential entrant that does not have enough resources of its own and needs to borrow heavily. In such a situation, the application of predatory measures by the incumbent will reduce the possibility that the new company can obtain financing, since it reduces its profits which worsens its credit quality vis-à-vis the bank. Therefore, it is the aggressive behavior of the incumbent that reduces the funds available to the rival.

One objection to this argument is that the lender might have an interest in preventing predatory measures by announcing that it will finance the incoming company in any case. However, this also poses moral hazard problems for the bank.

In short, these models provide a compelling story of why predatory actions take place. Incumbent applies aggressive behaviors to modify the expectations of the profitability of the incoming rival. In this particular case, predatory actions affect the bank's assessment of the rival company and will result in a reduced ability to borrow and be forced out of the industry or downsize its operations.

2.3. PREDATORY PRICING IN PRACTICE

A clear lesson from the previous section is that incumbents may well use aggressive pricing strategies in order to curb entry and/or to expel or be likely to expel one or more competitors from the industry in order to bolster or maintain their market power, through different mechanisms: building a reputation for being aggressive, to scare off the company's potential competitors, setting low prices to signal potential entrants not to expect big profits, or eroding a rival's resources to make it harder for them to get financing. Undoubtedly, competition authorities must be prepared for the use by some operators of such predatory practices involving an abuse of the dominant position.

However, we must also try to get something more out of economic theory, not just the result that predatory pricing can occur. In all predatory pricing theories, there is a common mechanism: the predator sets low prices over a period, thereby sacrificing short-term profits, so that a rival (or its creditors) believes that it should not expect high profitability. When the rival reviews its plans (or its lenders cut off the financing) and leaves the market, or abandons the project of entering, or reduces the scale of its operations, the incumbent then increases its prices obtaining profits, which in the long term compensate for the losses of the beginning. Therefore, it makes sense to have a two-level predation test, as follows:

1. Industry analysis to determine the degree of market power of the allegedly dominant company. If the company is not dominant, dismiss the case,

2. If the company is dominant, proceed with the analysis of the relationship between price and cost:

- A price above the Average Total Cost should be considered lawful, without exceptions.
- A price below the Average Total Cost but higher than the Average Variable Cost must be presumed lawful, leaving the burden of proving otherwise to the Competition Authority.
- A price below the Average Variable Cost is considered unlawful, with the burden of proving otherwise falling on the defendant.

2.3.1 Ability to increase prices (Is there dominance?)

A necessary element for predation is the possibility that it can more than compensate the incumbent once the rival is expelled for the lower profits during the predation episode. Clearly, this requires the existence of market power on the side of the firm, and the greater the market power, the greater the probability of winning in the long run after the exit (or reduction in size) of the rival. Since market power is a question of degree, the question is where to draw the line above which a company can be accused of predatory practices: can a company in oligopoly, among many others in its sector that can exercise some market power, be accused of predation? Based on EU regulations, it is not a problem, as predatory practices fall into the category of abuse of a dominant position. Therefore, an oligopoly undertaking which does not have a dominant position shall not be guilty of abusive practices.

In the US, the issue is less clear, and courts have found companies guilty of predation, even low market shares. An example of this is Brooke Group, although the latest Supreme Court decision ultimately dismissed the predation claims, following a ruling finding her guilty of predatory practices. In that case, a cigarette producer, Liggett & Myers (of the Brooke Group) had filed a lawsuit against Brown & Williamson – which controlled 12% of the cigarette market – accusing it of having entered the generic and private brand segment of the market and of selling below cost to drive Liggett – the main company in this segment – out of the market.

However, it is necessary to establish high market power to avoid jeopardizing competition in the markets. It would be paradoxical for antitrust authorities to set up barriers to practices used by non-dominant firms to increase their market power.

Suppose, for example, that a company with a considerable market share (e.g., 20%) in an oligopolistic industry where there is a much stronger company (e.g., one with 60% of the market), tries to increase its share, decreasing its price. Suppose also that the lower price allows you to steal from customers both the leader and a smaller competitor, one that, for example, has 5% of the market.

The latter may accuse the company of applying predatory pricing and - if the price is below any specific cost measure, the Court will find evidence of predation. However, instead of behaving predatorily, the non-dominant predator might be trying to increase its market share through aggressive, but legal, behavior.

There are many reasons why a non-dominant company might set prices below costs as part of a normal competitive process.

Consider an industry characterized by switching costs: Most consumers would be trapped by the dominant company, and only significantly lower prices could convince many of them to go to another vendor instead. The same argument holds true for markets with network externalities: if they are significant enough, a substantial reduction in prices might be necessary for a company to gain enough customers and reach critical mass. Or industries characterized by steep learning curves or significant economies of scale: a company may want to drastically reduce prices and increase its output in order to go down the cost curve and increase its efficiency. Finally, a company may charge a price below cost if there is a complementarity with another market product. If the latter is more important to the company and is profitable in the short term it will decrease prices in the first market to increase demand in the complementary market.

These same arguments (with the exception of the defense of complementarity) would not be applicable to a dominant company, since it cannot be said that it needs to lower the price further to increase its sales, since it is already well established in the market: consumers are already blocked due to switching costs and network externalities, and presumably it has already reached the minimum efficient scale and benefited from the learning effects. Therefore, the market power test must take only the dominant firms. It is true that, in exceptional circumstances, there could be a non-dominant predator, who would go unpunished. This would be a small price to pay compared to a test that sets a lower fee for dominance. Finally, the existence of a dominant position must relate to the period in which the first allegedly predatory episode begins, not later.

2.3.2. Sacrifice of short-term benefits

In the predatory pricing models discussed above, the sacrifice of short-term profits implies that the predator-incumbent is not choosing the price, which he would optimally choose if he took as given the presence of the rival, but a lower price, say p', in which lower profits will be obtained. However, there is not necessarily a relationship between the price p' chosen to exclude the rival and any other measure of the incumbent's costs (e.g., p' may be above or below the marginal cost of the incumbent). In other words, the theory only says that the incumbent chooses a price that generates fewer profits that they might otherwise be able to make in the short run, but does not say whether such profit is negative or not:

If one wanted to apply the theory, one would find oneself on slippery slopes, where one of the necessary steps of predation research should be the calculation of the optimal price pa, and proof that the real short-term price p' is less than that. Clearly this could not be feasible in practice. Despite the knowledge of the company's managers, it is unlikely that they have any idea what the optimum price is: a posteriori, the competition authorities and the treasury will find it impossible to deduce that profits have been sacrificed in the sense that the price set by the incumbent is below that which he would have established if he had not tried to force a rival out of the market.

However, there is an alternative route to sacrificing current benefits. It is about reinterpreting the concept and defining it not as making lower profits that would otherwise be possible, but obtaining negative benefits, in the specific market where predation occurs. In other words, the sacrifice of short-term profits is achieved if the price predation p' assumed 'is less than (some adequate measure of) costs'. This approach introduces a clearer benchmark: during the predation episode, the predator's profits must be less than zero, or its price below cost. Unfortunately, however, this rule is far from easily applicable in practice. But before discussing the practical application of the standard, let's make some additional observations.

This rule is nothing other than what most Courts have done and what commentators have long suggested: a necessary (though not sufficient) condition for proving an accusation of predatory price dominance (or abuse of dominant position) is that the predator incurs losses during the predation period. . This rule makes a lot of sense. A company that makes a profit should be excluded from a predatory pricing complaint as it cannot be proven that it could have made higher profits if it had proceeded differently. A company that has negative profits, on the other hand, could be a predator, although there may be other reasons why a company wants to charge below cost, such as the sale of perishable products that would otherwise be unsellable (thus causing even greater losses), for example through promotional offers, stimulating sales of complementary products, and so on (see below).

This rule, however, is true that it leaves some cases of predation uncovered, as we have already seen that - in theory - a company can opt for strategic prices to exclude its rivals, even without going as far as selling below costs. Given the difficulty of proving that the price has been lower than a certain optimal price, there could still be another possibility, which is to find the documents on the premises of the alleged predator that show that the managers have been willing to sacrifice their profits in the short term to exclude their rivals. However, such tests should not replace objective evidence. Phrases like “we will reduce prices in order to give a signal that we are efficient”, or "... To get them out”, or even the existence of a business plan to reduce prices to make life difficult for competitors, maybe a complementary test, but it should not be taken as an independent and objective proof of the sacrifice of short-term profits, if at those prices the incumbent is making profits. If an inefficient rival enters the industry, an efficient incumbent might be entitled to reduce its prices in response to entry and might know that this is going to force the rival out. However, this is simply part of the competitive process, and the exit of an inefficient company does not harm well-being. Similarly, if an incumbent makes a rival come out by lowering prices, but at a level where it still makes a profit, this means that the rival is likely to be much less efficient than the incumbent. Therefore, it is unlikely that there is a great loss of well-being due to their departure. In short, a "price below cost" would not allow us to encompass all possible cases of predation. However, the cases not foreseen in this test are likely to be few or very special. The cost of making this kind of mistake seems small. Compare this with the mistake we would make if we allowed a predation finding with prices higher than costs. The absence of an objective rule based on observables ("optimal" prices are not observable) would introduce an element of legal uncertainty and arbitrariness. This could not only influence individual cases, but could also have dire consequences throughout the economy, as firms endowed with market power would hesitate to reduce their prices so that they would not be accused of forcing smaller competitors out of the industry or preventing the entry of new rivals.

Since low prices improve consumer surplus and welfare, it must be an objective of any competition policy to create favorable circumstances to achieve low prices, the risk of discouraging firms from setting low prices is simply too great.

What definition should be of costs in a price-below-cost test? Considering whether at a given price a company is making positive profits or not (i.e., it is selling above or below cost) is a difficult task.

First, one must decide which cost measure should be used in this evaluation exercise.

Areeda and Turner (1974), in an article that has influenced the defense of competition, argue that the best measure from the conceptual point of view should be the Marginal Cost since a company that sets the price below the marginal cost clearly will not maximize the benefits in the short term. However, they suggest that the Average Variable Cost - it is defined as the sum of all variable costs divided by production - be used in practice as a substitute for Marginal Cost, given "the difficulty of determining the marginal cost of a firm.

The incremental cost of preparing and selling the last unit cannot be easily deducted from the accounts of normal undertakings, since normally only the Observed Average Variable Cost is available. Consequently, it is quite possible that it will be necessary to use the latter as a Marginal Cost indicator. Therefore, Areeda and Turner (1974) suggest that:

(A) a price equal to or greater than the anticipated Average Variable Cost should be understood as a lawful situation.

(B) A price below the reasonably foreseen Average Variable Cost must be understood as an illicit situation.

There are probably not enough cases of predatory practices to say what the Courts of different jurisdictions currently consider predatory prices, however, according to Bolton, Rodley and Riordan (2000), undercurrent US law, a price above the Average Total Cost (ATC) is lawful, while a price below the Average Variable Cost (AVC) is at least suspicious. A price between AVC and ATC is sometimes considered illicit, depending on other factors.

In the case of the EU (Paragraphs 64 et seq. of Commission Communication 2009/C45/02) the following rule applies: "The Commission shall consider conduct to involve sacrifice (incurring losses or giving up short-term profits) if the dominant undertaking, by applying a lower price to all or a certain part of its production during the reference period or by increasing its production during the reference period, incurred or is incurring losses that could have been avoided. The Commission will take the AAC (Avoidable Average Cost which in most cases coincides with the Average Variable Cost although the AAC should be used when they differ from each other, for example when there is an increase in capacity to be able to apply predatory conduct) as an appropriate starting point for assessing whether the dominant undertaking incurs or incurred avoidable losses. If a dominant firm applies a price below the AAC to all or part of its production, it is not recovering the costs that could have been avoided if it did not manufacture that production: it is incurring a loss that could have been avoided. Thus, in most cases the Commission will consider prices below the AAC to be a clear indication of predation.

However, the concept of predation is not only about applying prices below the AAC. To demonstrate a predatory strategy, the Commission may also investigate whether the alleged predatory conduct produced in the short term a net income lower than would have been expected from reasonable alternative conduct, i.e. whether the dominant undertaking incurred a loss that it could have avoided. The Commission shall not compare actual conduct with hypothetical or theoretical alternatives that could have been more cost-effective. Only economically rational and practicable alternatives which, in view of market conditions and the commercial situation of the dominant undertaking, can realistically be expected to be more profitable will be considered. The Commission will analyze whether the competitor has an equal degree of efficiency, such as to determine whether the conduct is likely to be harmful to consumers. Normally only prices below the MLTIC (Medium Long-Term Incremental Cost) can exclude competitors with an equal degree of efficiency from the market. The Commission will investigate whether and how suspicious conduct will reduce the likelihood of rivals competing.

The Commission does not consider it necessary to prove that competitors have left the market in order to demonstrate that there has been an anti-competitive closure of the market. The possibility cannot be excluded that the dominant undertaking would prefer to prevent the rival from competing intensively and to force it to follow the prices of the dominant undertaking, rather than to eliminate it altogether from the market.

Finally, the Commission does not intervene only when it is likely that the dominant undertaking will be able to raise its prices above the level that existed on the market before the conduct. It is sufficient, for example, that the conduct is likely to prevent or delay a price decline that would have occurred without the conduct. The determination of the existence of harm to the consumer is not a mechanical calculation of profits and losses, and it is not necessary to prove that there were overall benefits. Probable harm to the consumer can be proved by assessing the likely market-closing effect of the conduct, combined with the examination of other factors, such as barriers to entry. In this context, the Commission will also examine the possibilities of re-entering the market.

Not only the Courts, but also some academics would rule out predatory pricing only if the price is above the Average Total Cost (i.e., if the company is able to recover its fixed costs as well), rather than above the average variable cost. Joskow and Klevoric (1979) suggest that it would be more appropriate to consider any price below average total costs as predatory (provided that the structural part of its test is met, i.e. if the alleged predator enjoys sufficient market power), since a situation in which a firm has losses cannot be equilibrium.

However, one problem with using Average Total Costs is that it would require a company to cover all sunken fixed costs, which is a very strict standard. Suppose, for example, that an incumbent has fixed expenses that she hopes to recoup through monopoly profits. Soon after, a new company unexpectedly enters the market, leading the competition to incumbents to reduce its price to a level where the sunken fixed investment is not recoverable. In this case, a price below the Average Total Cost would be predatory. This drawback is avoided by the concept of the Average Incremental Costs (AIC), defined by Bolton et al. (2000) as the cost per unit of production of additional production destined for predatory sale. Average Incremental Cost differs from Average Variable Cost in at least two ways:

- First, it is not measured on the entire output of the company, only on the increase in production used to supply the additional predatory sales.

- Second, incremental cost includes not only the variable cost, but the fixed costs incurred in new sales. Incremental cost is a better indicator than average variable costs or total costs, as it more accurately reflects the costs of making predatory sales.

Consequently, these authors presume infringement at a price below the Average Incremental Cost and a legal one above the Average Total Cost with a gray area in the middle. In summary, a number of cost rules have been proposed in the literature. In particular, both the mean variable cost (MVC) and the average incremental cost (AIC) are reasonable. Perhaps the CIM better matches the concept of predation, but it may not always be easy in practice to accurately identify the costs that are sustained for a given output, and/or to isolate the predatory output from total production. However, caution should be exercised in the search for abusive practices in cases where the price is above the average variable cost (or average incremental cost): the possibility that a company will be charged predatory prices if a price is set that allows it to recover the variable cost, but not the total fixed costs (i.e. a price above the average variable cost, but below the average total cost) seems too strict and could encourage companies to keep prices higher than they would otherwise have.

Consequently:

1. A price above the average total cost (ATC) should be considered lawful, with no exceptions.

2. A price below the average of the total costs but higher than the average variable costs is presumed lawful, with the burden of proving otherwise to the applicant or the competition authority.

3. A price below the average variable cost is presumed to be unlawful, with the burden of proving otherwise to the defendant.

For cases where the average incremental costs can be calculated, these shall be used instead of the average variable cost. To safeguard legal certainty, competition authorities shall express the chosen criterion with clear guidelines.

2.3.3. Predatory Pricing Test

This comments section deals with added elements that may arise during a predatory pricing case, intent, demonstration of recovery, proof of anti-competitive effects, and competitor pricing as a defense. It also deals with allegations of predatory pricing in high-tech markets, a topic that has recently been the subject of debate. Finally, he criticizes the recent laws and regulations adopted in some European countries, which prohibit companies from selling below costs.

a. INTENT (EXISTENCE OF A PREDATION PLAN)

In cases of predation, there is evidence through internal documents of the predator's alleged intention to exclude a new operator.

On the one hand, email, minutes and other internal documents where one or more administrators adopt a very strong language against competitors and establish that they want to expel them from the market, should not be given much importance, since they are part of the usual language of business (it would be much more suspicious of managers being kind to their competitors) On the other hand, it would be difficult to dismiss evidence that there is an articulated plan to try to exclude smaller rivals, at the cost of temporarily sacrificing profits. If there is evidence that a coherent business strategy has been put in place for exclusion purposes, and especially if the documents reveal the intention to produce at a loss to achieve this goal, then the burden of proving that there is no predation, after all, must be on the predator.

b. THERE IS NO NEED TO DEMONSTRATE EX-POST THE SUCCESS OF PREDATION.

The proposed test suggests looking at whether there is a chance of recovering losses in the long run. This assessment must in any case be made ex ante, rather than a posteriori, and the company cannot defend itself by showing that after all, it failed to recover losses. In other words, the predator should not benefit from the fact that he has miscalculated the opportunity to recover losses, or that the rival has become tougher than he initially thought, or that the action of the competition authority has led to the conclusion of the predatory episode before it could reach its objective. Similarly, if the market power and price above the cost levels of the test are satisfied, the fact that predation was not effective, and exclusion did not happen should not be admitted as a defense.

c. NO NEED TO PROVE HARM TO CONSUMERS

Competition policy aims to protect competition, not competitors. Therefore, it might seem strange that there is no explicit requirement of harm to consumers to prove predation. However, predatory pricing will translate into a lower consumer surplus in most cases. It is possible to conceive models in which predation occurs without harming the consumer surplus (for example, in signaling models). However, the fact that the predator expects to recover losses by raising prices after the exclusion of some leads to a sufficient time horizon for the high future prices to have a higher weight than the current low prices. Consequently, a negative effect on the net consumer surplus can be expected, leading to the presumption of anti-competitive effects. In particular, it should not be accepted as a defense that consumers have a posteriori benefit from the predation episode: the fact that low prices do not follow other high prices could be due to: predatory miscalculations, anti-trust action, or increased resistance of the rival company. Conversely, the alleged predator may possibly have an efficient defense of having below-cost prices (for example, when active in complementary goods markets). If you can prove this and prices are below the Average Variable Costs, then this would exclude anti-competitive effects.

d. EQUALIZING COMPETITOR PRICES AS A DEFENSE

It is part of the normal competitive process that the entry of a competitor triggers a response from the incumbent that leads to a decrease in prices. Therefore, the observation of a price drop by a dominant firm after an entry is not evidence of strategic behavior. A rule that dictates otherwise and forcing a dominant firm not to react to entry would prejudice the competition process and lead to market distortions (any inefficient firm would be invited to enter). However, a rival's price equalization should not be accepted if it leads the incumbent to prices below the average variable cost.

As we have seen above, there are many reasons why an entrant may wish to set prices below cost over a period in order to win new customers and overcome the competitive disadvantage created for example, by switching costs, network economies, and scale. In such situations, an overly aggressive price may be the only instrument available to a new company to win customers. While initial losses are perfectly justified for the incoming, the same is not true for the incumbent with a dominant position. Consequently, a competitor's price equalization should not be an acceptable defense if it involves prices lower than the average variable costs for the incumbent. Conversely, a price above the ATC is legal, even if it means harming a smaller competitor or a new company.

e. ABUSIVE PRICING IN HIGH-TECH MARKETS (see Network Externality Annex)

In recent years, there have been Numerous Cases where they have been involved high-tech companies, such as IBM, Intel, and Microsoft. Microsoft, in particular, has been the subject of several investigations in the US. and the EU. Without discussing the merit of these allegations here, some commentators have argued that traditional antitrust analysis is not well suited to high-tech markets such as software characterized for very high fixed costs and very low marginal cost (and average variable costs). This means that we should expect very intense competition in the early stages of the market, until the market leans in favor of a company, which will become dominant and use its market power to recover the losses suffered in the initial periods, until a new technology replaces the previous one, ending the previous leadership. However, these markets do not seem to differ from others of less technological advancement: there are multiple examples of industries with strong sunk fixed costs.: by example advertising expenses in consumer products as the chemical industrial, pharmaceuticals, and engineering characterized by high investment in R&D.

Network Economies are also seen in traditional products such as toys, shoes, and the design for example, fashion and of course network externalities play a role in electrical appliances, fixed telephony, railways and records. A company that is charging prices below ATC it is not necessarily guilty, even if it has a dominant position. And, certainly, if two or more companies are competing for the market in each sector and neither of them has a dominant position, the fact that set a price below cost No would imply a breach of the antitrust rules provided that the first condition is not met (absence of dominance). In other words, if there is no dominance ex ante, the case of predation it must be dismissed. The existence of network effects and significant R&D expenditure should not be an excuse for a dominant company Try to support its market power through anti-competitive means. The rules should apply to high-tech markets, a lot how in the traditional ones, and if a dominant firm tries to support its predatory pricing position, it must be punished for it.

Another interesting case is one in which a dominant company tries to enter a new market for a complementary or independent product, which is often the case in high-tech markets (think, for example, of Microsoft, which has progressively focused on Internet browsers, audio and video software, instant messaging services, software for handheld devices, and so on). If the products are complementary, remember that it is more efficient for a monopolist to produce two products than for them to be produced by two different monopolists.

f. PRICE BELOW THE AVERAGE VARIABLE COST

It is not a general rule. In most EU countries, there are laws and regulations that apply to specific sectors or the entire economy, and prohibit below-cost prices, promotional sales, gifts, offers and retail discounts above a certain threshold. In general, these rules are the result of negotiating with small traders who see their businesses threatened by aggressive price cuts by hypermarkets because they are abusive and unfair. However, the consequent obligation to set a price higher than cost applies to any undertaking, irrespective of the market power it possesses and the reasons for the price cut. There is no justification for a regulatory approach, which protects competitors rather than promoting competition. To repeat, there are many reasons (switching costs, network externalities, promotions, pricing and complementary products) why companies want to sell at a loss temporarily. If it wants to safeguard small businesses, this must be the subject of other economic policy measures (tax breaks for example).

3. CONCLUSION

Predatory pricing refers to short-term price reduction in an effort to exclude rivals in order to gain or protect a incumbent's market power. Despite the interest they have aroused, they are a complex form of anti-competitive behavior. There has been considerable debate about the frequency with which predation occurs as well as the proper form of control. Unfortunately, the exact frequency remains unknown, as there are no reliable statistics available, but they seem to appear very infrequently and are often simply the result of price competition. A promising course of action against predatory pricing is the "two-tier" rule, a criterion according to which competition authorities first determine whether the market in question is susceptible to predation and in the second stage analyze prices in relation to costs in the light of the economic context. The possibility for companies to hide predatory activities increases the likelihood that predation will succeed (due to information asymmetry) and go unpunished (because prices appear to be higher than costs). Other factors must also be analyzed such as intentionality through actions such as litigation against the rival, intervention before regulatory agencies, interference with relations between the rival and its suppliers or customers and changes in contracts or exclusivity agreements.

Finally, it is worth highlighting other tactics apart from predatory pricing whose effect would be to increase the costs of rivals. These practices can also be used as an instrument of protectionism by a domestic company that seeks to limit international competition, through the imposition of litigation on other importers or by request of government agencies to effectively block imports. Note, however, that the notion of price predation is not straightforward and it is possible that genuinely pro-competitive activities, such as product innovation, may be misinterpreted as a predatory strategy. Therefore, it is a theory that must be applied carefully.

4. BIBLIOGRAPHY

Areeda, I., and D. Turner. 1975. Predatory Pricing and Related Practices under Section 2 of the Sherman Act Harvard LawReview 85

Bain, 1949. A Note an Pricing in Monopoly and Oligopoly. American Economic Review 39: 448-464.

Bernheim, B. D. and M. D. Whinston. 1998. \Exclusive Dealing." The Journal of Political Economy. 106(1): 64-103.

Bolton, P. and D. Scharfstein. 1990. \A Theory of Predation Based on Agency Problems in Financial Contracting." American Economic Review. 80: 93-106.

Bolton, P., J. Brodley, and M. Riordan. 2000. \Predatory Pricing: Strategic Theory and Legal Policy." Georgetown Law Journal. 89: 2495-2529.

Carlton, D.W. and M. Waldman. 2002. \The Strategic Use of Tying to Preserve and Create Market Power in Evolving Industries." Rand Journal of Economics. 33: 194-220.

Communication from the Commission — Guidelines on the Commission's control priorities in its application of Article 82 of the EC Treaty to the abusive exclusionary conduct of dominant undertakings (Text with EEA relevance) (2009/C 45/02)

Elzinga, K.G. and D.E. Mills. 2001. \Predatory Pricing and Strategic Theory". Georgetown Law Journal. 89: 2475-2494.

Fudenberg, D. and J. Tirole. 1986. \A Signal-Jamming Theory of Predation." Rand Journal of Economics. 17: 173-190.

Fumagalli, C. and M. Motta 2009 "A simple Theory of Predation", CEPR Discussion Paper No. 7372.

Innes, R. and R. J. Sexton. 1994. \Strategic Buyers and Exclusionary Contracts". American Economic Review, 84(3): 566 - 584.

Joskow, P., and A. Klevorick. 1979. A Framework for Analyzing Predatory Pricing Policy. Yale Low Journal 89: 213-270.

Karlinger, L. and M. Motta. 2007. "Exclusionary Pricing and Rebates when Scale Matters." CEPR D.P. No. 6258.

Kreps, D. and R. Wilson. 1982. \Reputation and Imperfect Information." Journal of Economic Theory. 27: 253-279.

Predatory Price Cutting: The Standard Oil (N.J.) Case. John S. McGee. Journal of Law and Economics, Vol. 1. (Oct., 1958), pp. 137-169.

Milgrom, P. and J. Roberts. 1982. \Predation, Reputation and Entry Deterrence." Journal of Economic Theory. 27: 280-312.

Motta, M. (2004). Competition Policy: Theory and Practice. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511804038

Rasmusen, E.B., J.M. Ramseyer and J.S. Wiley Jr. 1991. Naked Exclusion." American Economic Review. 81: 1137-1145.

Saloner, G. 1987. Predation, Mergers, and Incomplete Information". Rand Journal of Economics. 18: 156-186.

Scharfstein, D.S. 1984. \A Policy to Prevent Rational Test-Marketing Predation". Rand Journal of Economics. 2: 229-243.

Segal, I.R. and M.D. Whinston. 2000a. \Naked Exclusion: Comment." American Economic Review. 90: 296-309.

Selten, R. (1975): A Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games, International Journal of Game Theory, 4(1), 25—55

Tirole, J. (1988): The Theory of Industrial Organization, MIT Press.

The digital economy.

Network efficiencies and technology **adoption processes**

Objective

The objective of this article is to theoretically construct a demand curve characterized by network externalities and to study whether empirical evidence supports the theoretical form of it. Secondly, several possible equilibria are analyzed assuming different behaviors of companies in the presence of these externalities.

1.- Introduction.

The study is applicable to a multitude of markets with network externalities, especially all those associated with the Internet (including telephone, e-mail, Internet, hardware, software, video, credit cards, etc.).

Markets with network externalities can be defined as those where the utility derived from consumption or subscription to a network depends on the number of consumers who consume that product (or are subscribed to a network). (see Liebobitz and Margolis,1995a)[23] Thus, no one would subscribe to a telephone, fax, email or internet network for example if there is no user registered in that service.

In these markets, producers tend to operate with economies of scale (thus the production of software entails huge initial versions although the cost of producing a CD ROM or distributing it over the Internet is almost zero), and consumers can be trapped in a specific technology (an operating system for example, Windows, Unix, Apple, etc.) given the high costs involved in switching to another technology (costs known as switch costs).

The presence of externalities and increasing returns at scale implies that in these markets the first fundamental theorem of welfare economics is not fulfilled (a competitive general equilibrium is optimal).

2.- Contributions of economic theory to markets with network externalities.

The existence of interdependencies between consumers has long been analyzed in economics (Veblen, 1899 and Leibenstein, 1950), as traditional economics is unable to explain certain consumer behaviors, specifically that some consumers took into account when consuming a product whether or not other consumers also consumed that product.

With the development of telecommunications and computing arises the study of the externalities that cause a consumer to other consumers to adhere to a network (direct network externality).

[23] These authors consider the above definition as network effect, while network externality is a particular case where there is also a market failure consisting in that market participants are not able to internalize the profit generated by a new actor. In markets where these benefits are internalized, there would be no externalities, only network effects.

Rohlfs,1974, was the first to model a demand curve in the presence of network externalities. Katz and Shapiro, 1985 applied it to analyze a Cournot oligopoly with network externalities, and defined an equilibrium point known as FECE (Cournot equilibrium in which all agents meet their expectations).

The literature on network externalities has distinguished direct network externalities, such as those that take place in any telecommunications network (e.g. fax, telephone, internet networks, are useful and more demanded the more users are connected) from indirect ones. when they do not directly affect users but do incentivize investment and the development of new applications (e.g. if more users use a given hardware more software applications will be developed for this hardware), see Katz and Shapiro (1985). There are numerous markets in which the existence of these externalities is observed (TV, telecommunications, social networks, operating systems, etc.)

In them, compatibility is essential to take advantage of these externalities and many of the anti-competitive behaviors observed in the market concentrate on limiting this compatibility through, for example, restrictive licenses that allow companies to enjoy monopolistic situations (the case of end-user licenses of software manufacturers).

From a dynamic point of view, competition is aggressive in the initial moments, but once adopted a type of technology it is difficult to transfer users to other technologies (for example, the QWERTY keyboard once adopted by a large user base makes it almost impossible to replace it with a more efficient one such as DVORAK) Therefore, bidders initially compete to achieve an extensive base of users using their technology (it can be an operating system, for example)once achieved it is very difficult for users to switch to another technology because of the learning costs involved.

In this context, once a critical user base has been achieved, the incentive of the network owner is that their technology is incompatible (the case of Microsoft software, for example) thus avoiding competition from other potential bidders.

Another effect that manifests itself in these markets are the switch costs or costs of switching or from one technology to another, see Farrel & Klemperer (2007). Users once accustomed to a software package, for example, will be reluctant to switch to another technology because of the learning costs involved (consumers are held captive or "locked-in"). In this context, the owner of this technology will set higher prices and obtain greater benefits in the face of the scarce existing competition.

The presence of these network externalities, and lock-in effects lead to the resulting markets not being competitive but *imperfect* or *oligopolistic.*

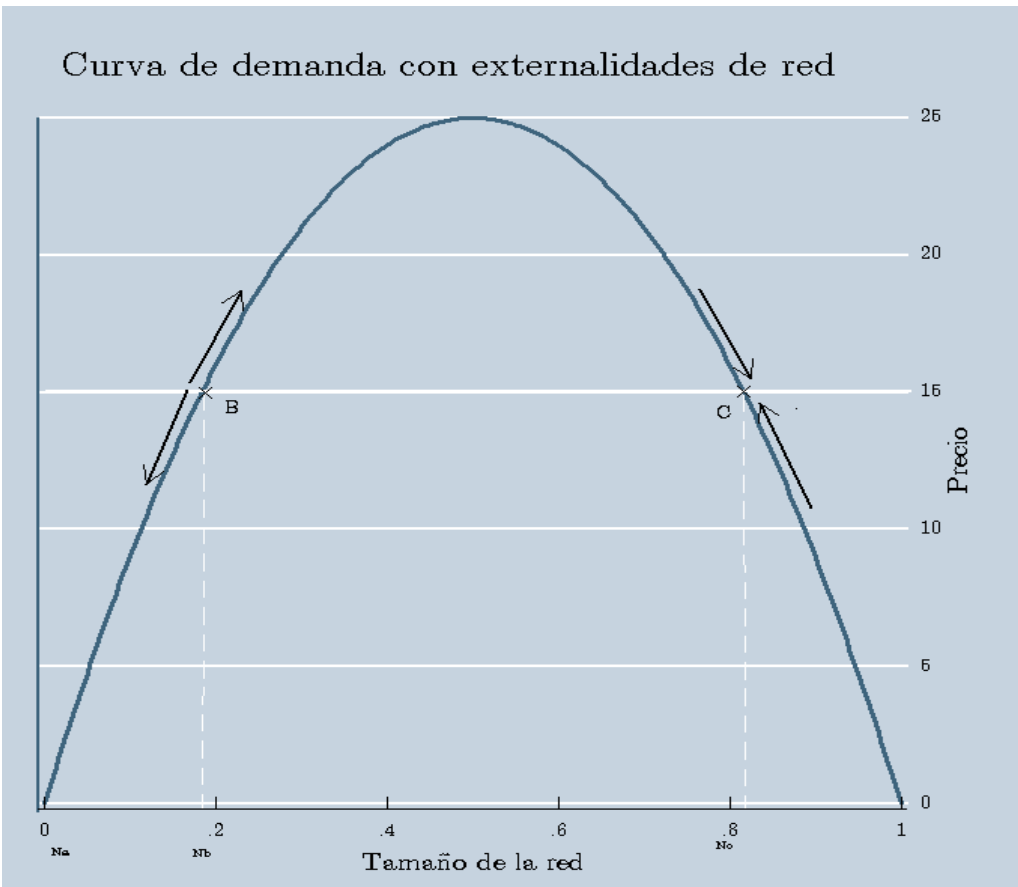
3.- Demand curve in the presence **of network externalities (see** Varian,1998 and Oz Shy).

M. Katz and C. Shapiro develop a static method with 2 companies that compete in quantities and produce identical goods with a constant and identical marginal cost. The good generates network externalities for the plaintiffs. In summary we can explain this model:

The demand curve with network externalities (initially obtained by Rolfs, 1974) differs from a usual demand curve in that it includes a new variable which is the consumer's expectation of the size of the network (n) . Including this variable, the price-quantity ratio of the traditional demand curve breaks down, it will no longer be decreasing throughout the curve, but will have increasing segment (users expect an increase in the size of the network) and decreasing (from a certain network size users do not expect growth in the network).

Let's look at an example to build this demand with 100 consumers.

The price that a consumer would be willing to pay for a good with network externalities would be $p = vn$, where $n = 100 - v$ [24] being v the consumer's valuation of the good. One can represent the demand for a good in the presence of externalities as follows:



If we consider the offer with constant returns of scale we would have several equilibria:

- ✓ N_a: Stable equilibrium: Intersection of the supply line on the OY axis. It can be interpreted as a point where there is no consumer in the network the valuation of the good will be null and void and no one will acquire it.
- ✓ N_b: Unstable equilibrium: To the left of the equilibrium the demand is below the supply, the valuation of the consumer is lower than the cost so it will not be incorporated into the

[24] We consider that there are 100 consumers and that each one has a valuation of the good v = 1,...,100, the number of people willing to buy andl would be equal to 100 – v, if p=Vn in the presence of network externalities, then p=v(100-v)

network. On the right, the valuation is higher than the cost, so it will always be incorporated into the network.

This point is also the minimum size of the network or critical mass or that minimum network size so that potential users are compensated to join it. On the left of it the technology will end up failing and on the right a process of expansion would begin until reaching the size of stable equilibrium.

- ✓ N_c: Stable equilibrium: When demand is above supply the benefit for the consumer of entering the network is above cost so that more inputs will be produced into the network up to a "stable" equilibrium point. On the right, the cost outweighs the benefit of adhering, so consumers will stop consuming the good or will not join the network.

In general, if there are n consumers connected to a network, each user can potentially communicate with n(n-1) users (see Metcalfe, 1973), an additional consumer would add2n potential communications in the system. In general, the demand function would take the following form in the presence of network externalities:

$$P = (1 - n)n^k$$

3.1. Equilibrium with Monopoly

Let us now study the equilibrium in the case of a monopolist whose aim is to maximize his profit in the face of a demand with network externalities (we assume that the marginal costs are constant and equal to 15). The first order conditions in equilibrium from the demand function with network externalities with the previous form would be:

$$\frac{\partial \pi}{\partial n} = P + n\frac{\partial P}{\partial n} - \mathrm{C_n} = \mathrm{n^k(1 - n)(k + 1) - n^{k+1} - C_n} = 0$$

In our example, (k=1) would $C_n = 15$ equal to $100(2n - 2n^2) = 15$, which clearing equals n =0,9625, with a price equal to 3.609.

3.2. Equilibrium with Oligopoly

If we now assume two companies that take the production of the other as given (Cournot oligopoly) the demand would have the following form:

$P = (1 - n)n^k$where $n = n_1 +\ n_2$ and maximizing profit, we would reach the following conditions of the first order:

$$\frac{\partial \pi}{\partial n} = P + n[kn^{k-1} - (k + 1)n^k] - C^{'}$$

In a symmetrical equilibrium where $n_1 = n_2$ this ecuation is summarized in:

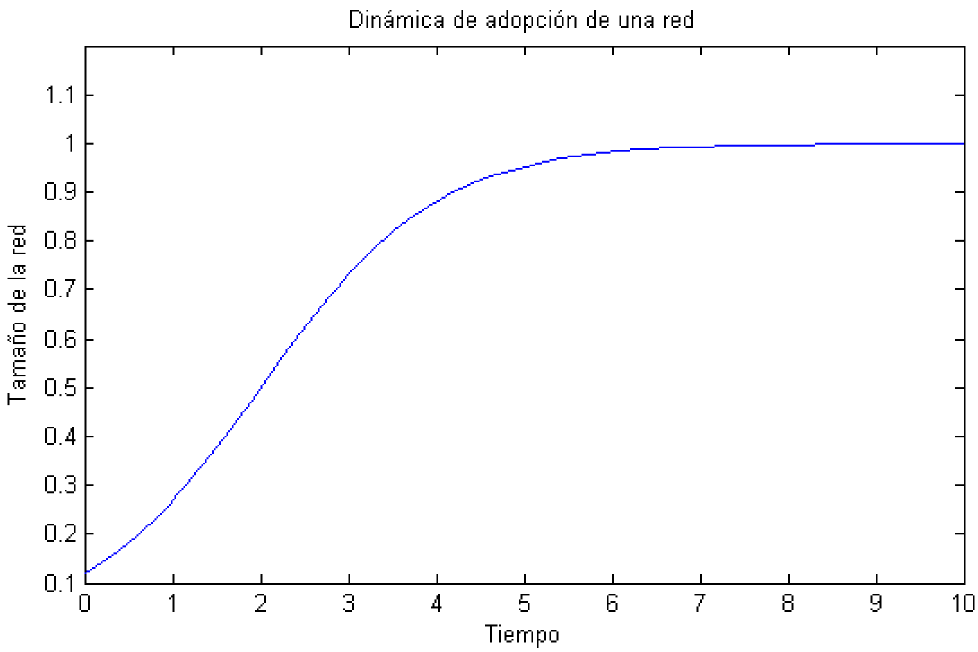
$$(1 + k)n^k - (1 - k)n^{k+1} - C^{'} = 0$$

In our example, with k=1, 2 companies and C=15 would be equivalent to:

$$2n - n^2 - 15 = 0$$

3.3. Dynamic Equilibrium

From a dynamic point of view from the starting point where there is no user, technological advancement will reduce costs by shifting the cost curve downwards, until users join that little by little and as the cost of joining the network there will be a massive input that will increase the network to a new equilibrium point this time stable $N_{c.}$ Graphically, it can be approximated according to a graph similar to this:(Economides, 2006)



We can use internet adoption data (number of internet hosts worldwide) between 2000 and 2013 and we see that they behave in a similar way to what is stated in the previous section.

Many authors consider (Varian Shapiro, 1999a, Economides,2003) that the dynamic behavior of a technology subject to network externalities follows a temporal evolution in a sigmoidal form.

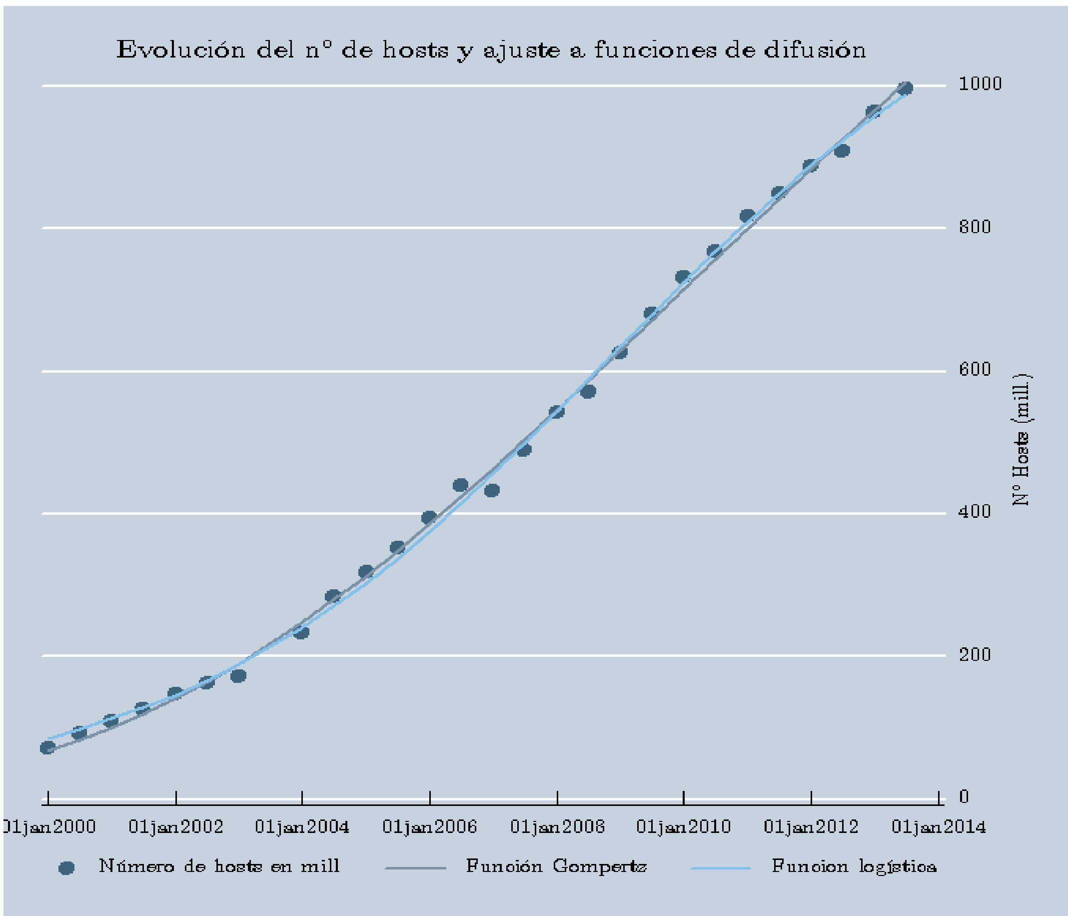
Among the numerous theoretical models that represent this evolution, we choose two of the most common (Logistic or Gompertz model) and study if the empirical data we have fit these models. Using nonlinear regression with STATA 11.1 *(nl*command) we see that the real data conform to both theoretical models: a logistic function and a Gompertz function, [25]concluding that its dynamic behavior is that of a technology subject to network externalities.

[25] The equation of Gompertz $y = e^{-e^{\beta t + k}}$and the logistic equation $Ln\left(\frac{x(t)}{1-x(t)}\right) = ba + bt$ are two of the most used models to explain dissemination processes Technological (See Barrigüete, 2007).

Tabla de resultados de regresiones no lineales (Logística y Gompertz)

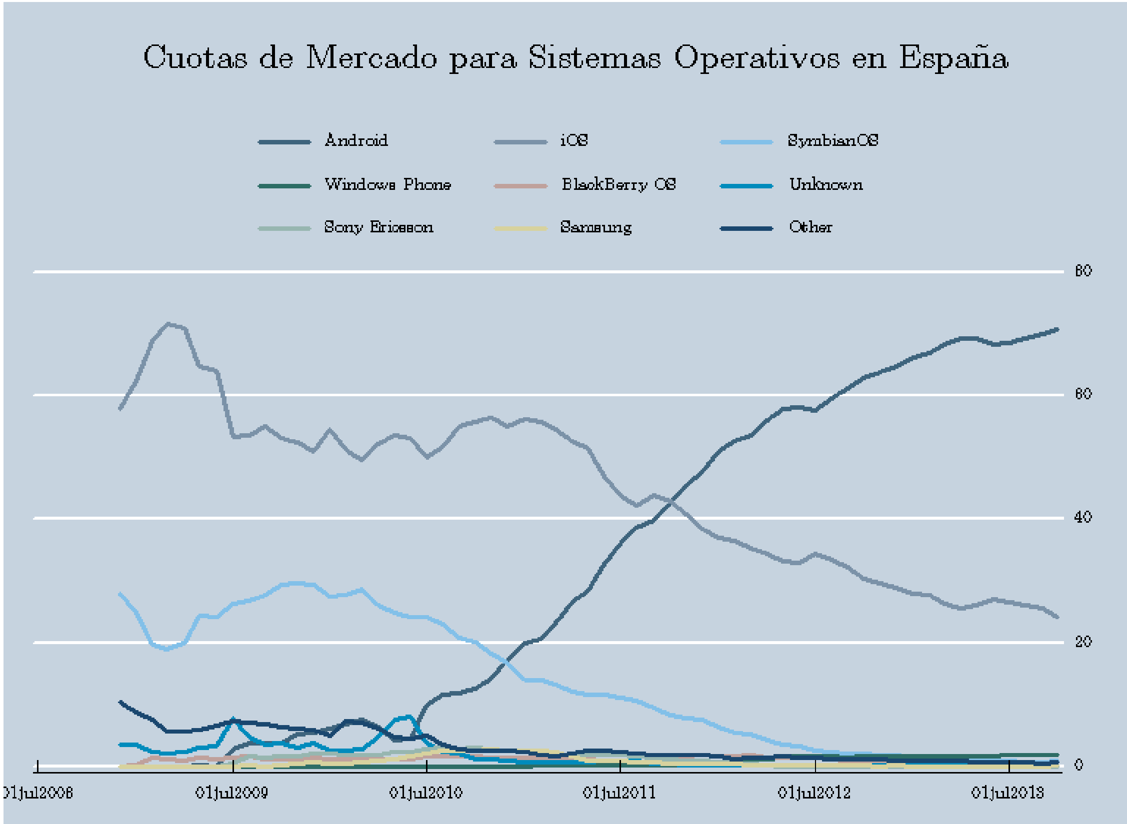
	(1) hosts (mill)		(2) hosts (mill)	
b1				
Constant	1239.5***	(40.59)	1905.3***	(16.89)
b2				
Constant	0.000805***	(37.57)	0.000334***	(20.08)
b3				
Constant	17837.3***	(244.30)	18202.0***	(97.61)
Observations	27		27	
Adjusted R-squared	1.000		1.000	

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

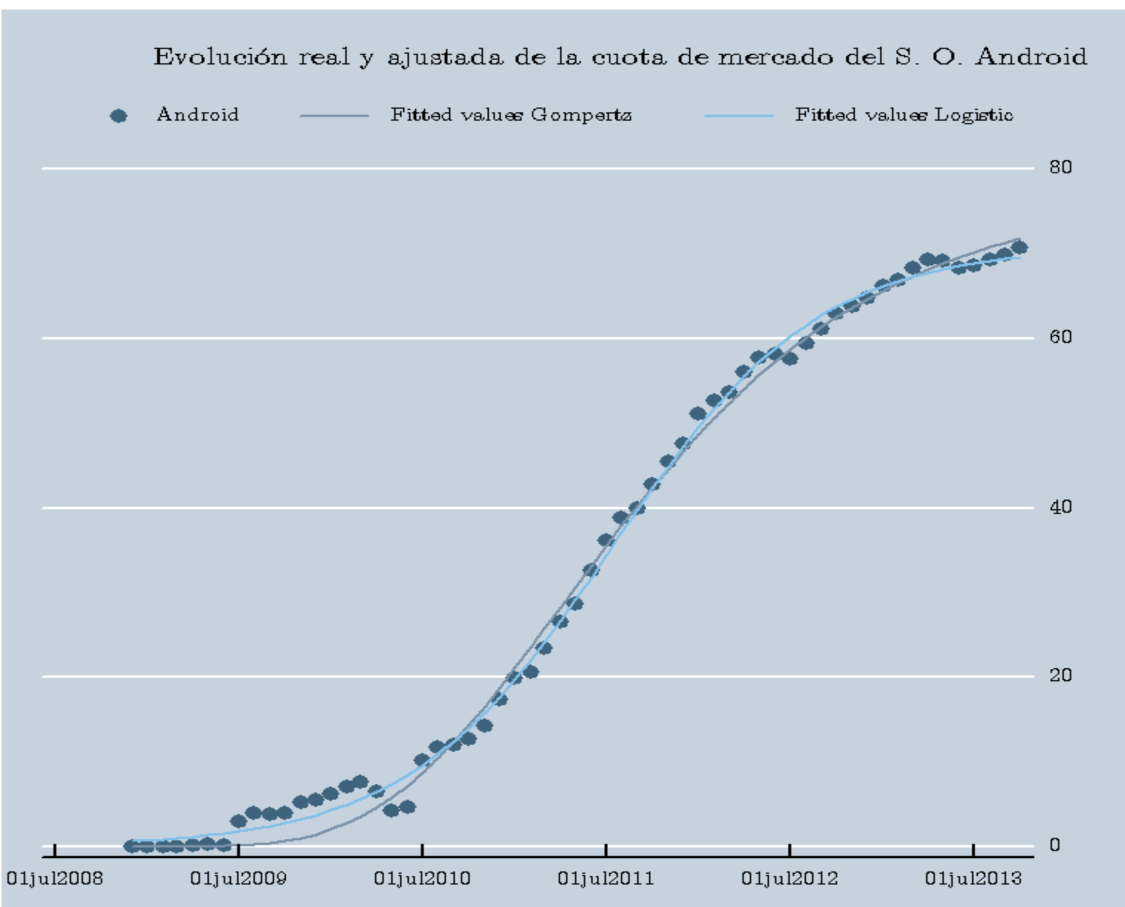


Source: Internet World Stats

Another example confirms what has been said above. From the market share data of the main mobile operating systems in Spain in the last four years, we observe that the evolution of the Android operating system in recent years seems to follow the same dynamic behavior in which network externalities would be present. We readjust the empirical data we have with theoretical mathematical models that have a sigmoidal or S-shaped behavior and again we observe an almost total adjustment of the empirical data with the theoretical models which comes to prove the presence of network externalities in that market.



Source: http://gs.statcounter.com/

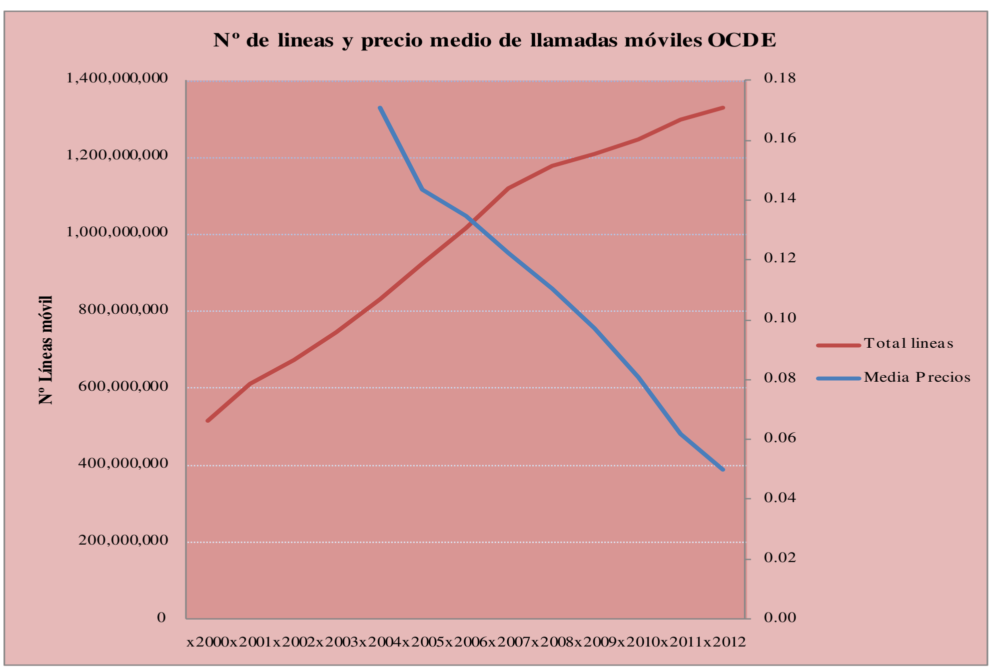


4. Empirical evidence for the presence of network externalities

Multiple authors have analyzed the presence of network externalities in markets: the most cited are Greenstein (1993), Gandal (1994, 1995) and Saloner and Shepard (1995) who focus on compatibility and network externalities to demonstrate empirically that the value of hardware depends on the variety of existing compatible software. The first studies the network externalities and compatibility of IBM's first servers, the second in the spreadsheet market, and the third focuses on the compatibility of ATM networks. Rysman (2004) develops a structural model to examine the importance of network effects in the yellow pages market. In another area, differentiated goods and from a dynamic point of view, Clements and Ohashi (2005), use a logit model to check if there are indirect network externalities in the US video game market.

In general, the main problem that network externality research encounters is the availability of data. Another problem is to distinguish whether the increase in the number of users of a network is the result of network effects or are derived from a reduction in prices and costs. To distinguish these two effects Gandal (2000) has data on CD titles published at all times, also know the fixed costs of the CD industry and assume that the market for CD players behaves as competitive so prices are exogenous. Park (2004) develops a dynamic model of competition in network industries, for readers-video recorders where pricing is analyzed over time where companies initially lower prices to subsidize the first users and thus build a network that will allow them to charge higher prices later.

We will apply the theoretical model of Economides and Himmelberg,1995 to the mobile internet market using a data panel of mobile line subscriptions and prices (average of 1 minute of a mobile call) in the last 12 years in 30 OECD countries.



Source: OECD Communications Outlook 2013

We start from a utility function as follows: $U(y, n^e) = y + h(n^*)$ where and es the maximum amount you would pay for 1 unit of that good in a network of size, $n^e$and is the function $h(n^*)$of externalities that captures the influence of the expected size of the network in the willingness to pay for that good by a consumer where that $h^{'} > 0$ y $h^{''} < 0$ is, the utility of the individual grows with the expected size of the network but in a decreasing way. In equilibrium: $y + h(n^{*}) = p$ and since $n = 1 - y$(consumers are evenly distributed between 0 and 1 and indexed in and, see Katz and Shapiro,1985) we arrive at:

$P = (1 - n) + \mathrm{h}(n)$ where we hypothesize that the number of future lines is an approximate function of the lines in the t-1 period i.e., consumers use the information they have of the size of the network in t-1 to predict the future size of the network. From this we assume that the function has the following mathematical form: $h(n) = k + b_1 n_{t-1} + b_2 \log(n_{t-1})$ substituting:

$P_t = 1 - n_t + k + b_1 n_{t-1} + b_2 \log(n_{t-1})$

In equilibrium,$n_t = n_{t-1}$, so we can estimate p using the following function:

$P_{i,t} = b_0 + b_1 n_{i,t} + b_2 \log n_{i,t} + b_3 X_{i,t} + e_{i,t}$

(For the function to have an inverted U shape the parameters must be $\beta_1 < 0$ and $\beta_2 > 0$)

where $P_{i,t}$ is the dependent variable representing the average price of a 3-minute national mobile call for country i in year t; $n_{i,t}$ is the number of mobile lines for country i in year t, is the $\log n_t$ logarithm of the number of lines and contains a set of control variables (GDP per capita, population, n of calls) $X_{i,t}$:

Regression Results Table

	(1) Price	(2) Price	(3) Price
Price1		0.00205	-0.00320
		(0.46)	(-1.02)
Lines	-4.94e-08***	-6.13e-08***	-4.98e-08***
	(-58.44)	(-55.43)	(-51.74)
Log Lines	1,642***	2,042***	1,657***
	(50.77)	(51.94)	(46.42)
GDP	0.000000222**	0.000000684***	-7.01e-09
	(3.13)	(6.22)	(-0.48)
Population	1.60e-10	1.31e-10	3.54e-13
	(1.99)	(1.01)	(0.08)
Log calls	-0.00167	-0.00386*	-0.000116
	(-1.55)	(-2.44)	(-0.58)
L.pm		-0.297***	
		(-14.22)	

Constant	-26.64***	-33.10***	-26.92***
	(-49.81)	(-51.89)	(-45.88)
Observations	150	125	150
r2	0.990		0.990

t statistics in parentheses
Source: OECD
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We performed several regressions with panel data to estimate the average price of a three-minute call in 30 OECD countries using as independent variables:

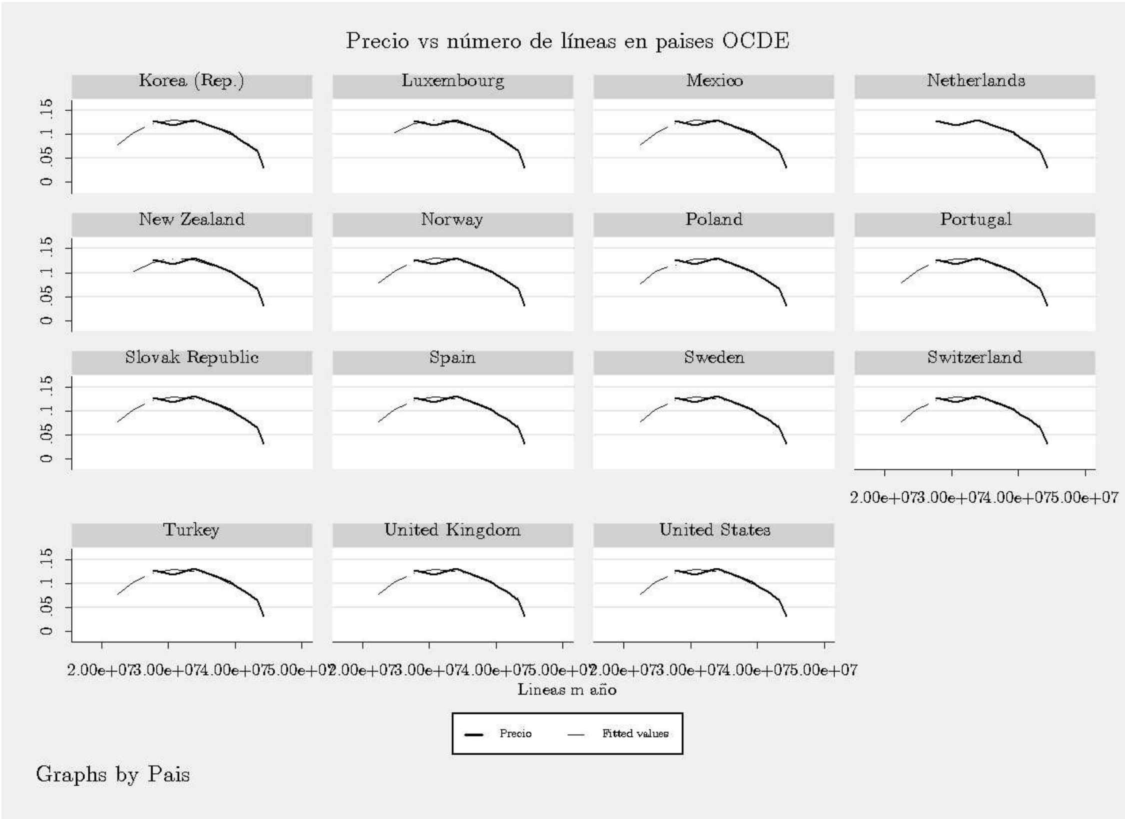
1. Price1: the dependent variable with a delay,
2. Lines m year: the number of lines (annual average),
3. Loglineas: the logarithm of this variable,
4. PIBP: GDP per capita,
5. Population: Population,
6. Logllamadas: Logarithm of the number of calls.

In the first column (1), we have estimated the price using a fixed effects regression whose results for the variables lines (negative sign) and loglineas (positive sign) are expected and significant so we can say that the hypothesis is fulfilled that the maximum amount that consumers are willing to pay for mobile calls is increasing, it reaches a maximum and then decreases.

Both the variable lines and log lines are significant and have the expected sign which comes to corroborate that demand or maximum amount that the user is willing to pay is increasing, reaches a maximum and then is decreasing. The coefficient of the logarithm of calls is not significant but the negative sign is correct, so an increase of 1% in the number of calls means a fall in the price of 0.0017. The coefficient of GDP per capita is also significant and with the correct positive sign, while that of population also has the correct sign but is not significant.

The presence of residue autocorrelation forces us to change the specification of the equation to a dynamic using a price delay as an exogenous variable and using the Arellano-Bond estimate (Column 2). Column 3 shows the results of an ordinary least squares regression.

In the graph below, the graphical representation of the price and the estimated price appears using regression (2) based on the number of lines, in fifteen OECD countries which allows us to conclude that the estimate is very close to the way that the theory predicts for the demand for a good subject to network externalities:



5.- Conclusions.

In this chapter we have constructed a demand function in the presence of network externalities applied to the case of the mobile phone network. We have verified the presence of network externalities in the mobile network and plotted the demand function (or valuation performed by the users of it) in the form of an inverted U and we have also proven that it has that form in practice.

We have used several estimation models from a panel of data obtained from the 30 OECD countries in the last 12 years to demonstrate the importance of network externalities in demand which should be taken into account when predicting the future development of one or another network.

6.- Bibliography

Clements, M., and H. Ohashi 2005. Indirect Network Effects and the Product Cycle: Video Games in the U.S., 1994-2002. Forthcoming, Journal of Industrial Economics.

Edward Elgar, Competition Policy in Network Industries: An Introduction, in Dennis Jansen (ed.) *The New Economy and Beyond: Past, Present and Future* (2006)

Ellison, Glenn and Sara Fisher Ellison, *Journal of Economic Perspectives,* 19(2), Spring 2005, pp. 139-158.

Farrell and Klemperer, Coordination and Lock-In: Competition with Switching Costs and Network Effects, in eds. Armstrong, M. and Porter, R. *Handbook of Industrial Organization,* Vol 3, North Holland, Amsterdam, The Netherlands. 2007

Gandal, Neil; Hedonic price indexes for spreadsheets and an empirical test for network externalities, the RAND Journal of Economics,160-170, 1994, JSTOR

Gandal, N., M. Kende, and R. Rob 2000. The Dynamics of Technological Adoption in Hardware/Software Systems: The Case of Compact Disc Players. RAND Journal of Economics, 31: 43-61.

Greenstein, Shane, Did Computers Diffuse Quickly? Best versus Average Practice in Mainframe Computers, 1968-1983. *Working Paper* No. 4647, National Bureau of Economic Research, Cambridge, MA. February 1994.

Jeffrey Rohlfs, A Theory of Interdependent Demand for a Communications Service *The Bell Journal of Economics and Management Science* Vol. 5, No. 1 (Spring, 1974), pp. 16-37Published by: RAND CorporationArticle Stable URL:http://www.jstor.org/stable/3003090

Leibenstein, H. (1950),"Bandwagon, Snob and Veblen Effects in the Theory of Consumers' Demand," *Quarterly Journal of Economics,* 64(2): 183-207.

López Sánchez, J. I. & Arroyo Barrigüete, J. L. (2006). "Network Externalities in the Digital Economy: A Theoretical Review". *Industrial Economics,* 361, 21-32.

Metcalfe, R. "Packet Communication", MIT Project MAC Technical Report MAC TR-114, (dec., 1973)

Michael L. Katz; Carl Shapiro Network Externalities, Competition and Compatibility", The American *Economic Review,* Vol. 75, No. 3. (Jun., 1985), pp. 424-440.

OzShy, The Economics of NetworkIndustries, *Graduate Lecture Notes.*

Rysman, M. (2004). Competition Between Networks: A Study of the Market for Yellow Pages. Review of Economic Studies, 71, 483-512

Saloner, G. and A. Shepard 1995. Adoption of Technologies with Network Externalities: An Empirical Examination of the Adoption of Automated Teller Machines. *RAND Journal of Economics* 26, 479-501.

S. J. Liebowitz & Stephen E. Margolis, 1994. "Network **Externality: An Uncommon Tragedy,"** Journal of Economic Perspectives, American Economic Association, vol. 8(2), pages 133-150, Spring.

CARTELS AND PROBABILITY OF DETECTION

Abstract

The article analyzes the probability of cartel detection in Spain using a sample of cartels discovered between 2011 and 2016. These types of studies have been questioned for using a sample considered biased *from* all *cartels. This* chapter *uses two alternative methodologies to conclude,*as well as (Harrington and Wei 2017), *that there is no such bias,* and that the estimation *of the probability detection is around 15%* using the *traditional Bryant-Eckard method. it is acceptable, although it should be interpreted as a probability of disappearance* (*or death)* of *the cartels, which only coincides with the probability of detection when we assume that all cartels disappear by detection. We must therefore consider that the result is an* upper limit below which the *real* probability *of detection* should be *placed.*

1.- Introduction

Cartels represent a significant damage to society in terms of higher prices and lower supply, higher public spending and economic inefficiency, reasons among others that justify their consideration as illegal and their persecution in most countries.

In order to calculate the effect that cartels have on the economy and to design dissuasive sanctions against these infringements, it is necessary to have an estimate of several parameters: the magnitude of the illicit benefit, the overcharge and the probability of detection.

The objective of this article is to estimate the third parameter: the probability of detection, and the application of the same or to the particular case of Spain. Surprisingly, there are no estimates of the probability of detection applied to the case of Spain. This article tries to fill this gap.

Indeed, in order to establish dissuasive penalties against cartels, that is to say, those capable of preventing the formation or continuation of a cartel, the unlawful benefit derived from the agreement must be less than the expected fine multiplied by the probability of detection.

According to (García-Verdugo, Merino Troncoso and Gómez Cruz 2016) the optimal sanction is *"one that discourages a company from participating in a cartel, and this is achieved when there is no expected net gain from participating in prohibited conduct, that is, when the expected illicit gain from entering the cartel is less than the expected loss. Therefore, in order to* *obtain a reference value for sanctions, it is necessary to begin in any case with the estimation of the benefit provided by cartel membership."*

The expression of the illicit benefit, or difference between the profit obtained by the members of the cartel and the profit obtained under conditions of competition, according to (Buccirossi and Spagnolo 2005) would be:

$$\Delta\pi = f(S; m, k, \varepsilon) = k\frac{(1+m)(1-\varepsilon k) - \varepsilon m}{(1+m)(1+k)(1-\varepsilon k)}S$$

It shows that the illicit profit is a proportion of the total volume of sales in the affected market S and that percentage depends on three parameters: m is the profit margin in the competitive *mark-up* scenario, k is the percentage of price increase as a result of cartelization ("cartel *overcharge*"), and ε is the price elasticity of demand (in absolute value). [26]

In (García Verdugo, Merino Troncoso and Gómez Cruz2016), the formation and stability of a cartel in a dynamic environment in which the annual probability of detection is λ; if the cartel is detected it will be dissolved and each company will pay a fine F [27] . The benefit of not complying with the cartel agreements is at least slightly higher than the benefit of complying with them, since the company that breaches them would set a price below that set by the cartel, thus increasing its sales at the expense of its cartelized competitors; in turn, they assume that the benefit of belonging to the cartel is greater than the benefit in a competitive situation.

The expression of the optimal deterrent fine from a dynamic perspective, i.e. the sufficient condition for the fine to have a deterrent effect is:

$$F \geq \frac{\Delta\pi}{\lambda}$$

A fine equal to or greater than would be sufficient for at least one of the companies in the sector to breach the collusion agreement, while the others would respond by choosing the competitive strategy and the cartel would be dismantled. According to the above expression, a fine would be dissuasive if it were at least equal to the annual illicit profit multiplied by a factor equivalent to the inverse of the annual probability of detection. $\frac{\Delta\pi}{\lambda}$

[26] If the value of the elasticity with a negative sign is used, the expression of the illicit benefit would be Analog but with a positive sign in the terms where elasticity appears.

[27] Both the probability of detection and the sanction are assumed to be constant over time.

In short, it is established that the probability of detection of the infringement must necessarily play a fundamental role in the determination of the fine in order for it to have a deterrent effect. An OECD report on the fight against cartels clearly underlines:[28]

"Effective sanctions against cartels should take into account not only the amount of gain realized by the cartel but also the probability that any cartel will be detected and prosecuted. Because not all cartels are detected, the financial sanction against one that is detected should exceed the gain realized by the cartel".

As we have said, the probability of detection is difficult to estimate, on the one hand, because the available sample of cartels is affected by a selection bias (Heckman 1979),since data are only available on the number of cartels that have been detected, but it is unknown how many are still operating in the market.

On the other hand, in general, we can only estimate a lower limit of the duration of a cartel because we are not sure when the cartel was exactly born, we can only infer it from the evidence in the file while we do know exactly the date of the discovery and death of the cartel. The durations of cartels discovered are therefore underestimates of the actual duration of the cartels which in turn are overestimates of the probability of detection.

In this paper, we first conducted a review of the existing literature on probabilities of cartel detection and then analyzed the available data of cartels in Spain, on which we applied, first, the methodology of Bryant and Eckard to the data of cartels in Spain to obtain a result of probability of detection . Secondly, we apply to the same data from Spain the most recent Harrington methodology that is based on Bayes' theorem, to finish with the main conclusions and possible future lines of research.

2.- Review of the literature on the probability of detection of cartels

In general, previous studies try to explain the probability of detection exogenously using theoretical mathematical models.

(Bryant et al. 1991) were the first to estimate the annual probability of detection of price agreements in the US and concluded that they ranged between 13% and 17%, from a database of 184 cartels detected in the US between 1961 and 1988. In their article they consider that the cartels are born and disappear when detected following a homogeneous stochastic process of birth and death, a methodology frequently used in biology and ecology, to describe population dynamics. These authors use the following intuition: if the durations of infringements are in most

[28] Development. Competition Committee. *Fighting hard-core posters: Harm, effective sanctions, and leniency programs.* OECD PUBLISHING, (2002), Report on the nature and impact of hard-core cartels and sanctions against cartels under national competition laws.

cases short-lived versus a few long-term, then the probability of detection should be high, and the total number ofactive cartels should be reduced. . Other authors have followed this methodology to estimate the number of active cartels and the probability of detection in different regions and countries.

(Combe, Monnier, and Legal 2008) they follow the same line as Bryant and Eckard's cartel detection model, using in this case EU data on the durations of 86 cartels between 1969 and 2008. They arrive at similar results: the annual probability of detection is between 12.9 and 13.3%.

Other authors use these percentages directly in their articles to estimate the optimal sanction against cartels, such as, for example, (Katsoulacos and Ulph 2013) or (Harrington 2014). As we have seen, in this line of research, the calculation of the probability of detection does not take into account the behavior of companies or the competition authority, but the creation and destruction of cartels follows a mathematical process of creation and destruction.

Although it is usual not to have the data on undiscovered cartels, the exception is (Hyytinen et al. 2010) that they have had the complete sample of cartels during the period in which they were legal in Finland and have thus been able to study the process of creation and destruction of all cartels in that country until their prohibition, unaffected by selection bias.

The usual, however, is to have a sample similar to that of the first authors to study the probability of detection, (Bryant et al. 1991) who recognized that it can only be concluded that the estimated probability (the probability of detection would be the inverse of the duration of the cartel) of 13 and 17%, it constitutes an upper limit on the probability of the disappearance of a cartel, since if it is not detected it will open up to have a longer duration.[29]

And it is that one of the restrictive assumptions of this model is that the cartels disappear only by detections, but they do not enter to analyze the possible disappearance of the cartel due to their natural instability.

(Ormosi et al. 2011) uses another capture-recapture model, often used in biology to estimate population data (size, strengths, survival rate). The authors conclude that in the EU the probability of detection is less than a fifth between 1985 and 2005.

(Miller et al. n.a.) it hasalso used a stochastic model but in this case to show the change in the number of detections before and after the introduction of the leniency program in the US in 1993, concluding that the program increased the detection and deterrence capacity of the US competition authorities.

[29] Other authors argue that the bias goes in the opposite direction since only long-term cartels are detected, short-lived ones go unnoticed by the competition authorities.

In the line of Miller's research, other authors have constructed a more complex model of an endogenous nature that explains the behavior of companies and authority (Harrington and Chang 2009a),especially the creation and dissolution of cartels. However, as far as we know its practical application is limited to the case already mentioned in Finland (Hyytinen et al. 2010), obtaining a result substantially lower than the previous studies (the probability obtained is only 4%).

The study conducted by (Harrington and Wei 2017) on cartels detected in the US reaches surprising conclusions. The bias derived from calculating the probability of detection from the sample of detected cartels is non-existent in a model where all cartels follow the same process of birth and death, while in a model that admits heterogeneity of the cartels, this error does appear, although they consider that it is smaller. of what was previously thought.

One can object to this study that only takes into account the detection of cartels during the period between 1960 and 1985 before the implementation of the leniency program in the United States. The authors, unlike the previous models, consider that a poster can disappear not only because it is detected but also because of internal instability.

Finally, (Park, Lee, and Ahn 2018) use a Bayesian methodology to calculate the probability of detection in the US, with results consistent with the previous authors, between 11.4% and 17.4%. However, this probability is calculated from a sample of investigated posters of which it differentiates between sanctioned and unpunished.

In this article we are going to analyze, first, the available data of cartels in Spain, data that are similar to those that any jurisdiction has and that are those used in the main studies. Secondly, we analyzed Bryant and Eckard's method to estimate the probability of detection and applied it to the data from Spain, and finally, we do the same with the Harrington and Wei model, reaching similar conclusions.

3. – Sample dataset

We have data on the duration of 114 cartels detected in Spain between 2011 and 2016. The data analyzed are those of cartel durations, although these data must be taken with some caution. We assume that the cartel is terminated at the time of the opening of the file while the date of constitution of the cartel can be deduced from the information contained in the file. There may be some ambiguity about this start date, since sometimes it is only known approximately as the minimum date of commencement of the cartel. It is possible that (Bryant et al. 1991) the start date refers to the oldest existing evidence. In any case, we have excluded from the sample those files of which we lack information on the duration of the infringement.

Bryant et al. use two methods to calculate the durations of infringements, the first being the maximum duration calculated with the available information, i.e. the farthest possible start date and the most recent possible to estimate the longest possible duration of the cartel. The second method is the minimum possible duration calculated with the available data (most recent evidence of cartel initiation and oldest date of completion of the cartel).

The following table shows the statistical data of durations of the sample of cartels in Spain using the two previous methods:

Duration 1	Difference between the final maximum date and the initial minimum date of the poster
Duration 2	Difference between the minimum end date and the initial maximum date

Table 1: Frequencies of cartel durations

	Frequencies	
Interval – Years of duration	Duration 1	Duration 2
0 to 2	20	24
2 to 4	11	15
4 to 6	14	12
6 to 8	14	20
8 to 10	12	15
10 to 12	7	12
12 to 14	5	5
14 to 16	10	3
16 to 18	5	4
18 to 20	9	1
> 20 years	8	4
Sample size	115	115
Stocking	7,65	5,55
Median	5,83	4,92
Min	0,04	0,04

Max	34,01	29,26

Figure 1: Histogram durations cartels discovered Spain

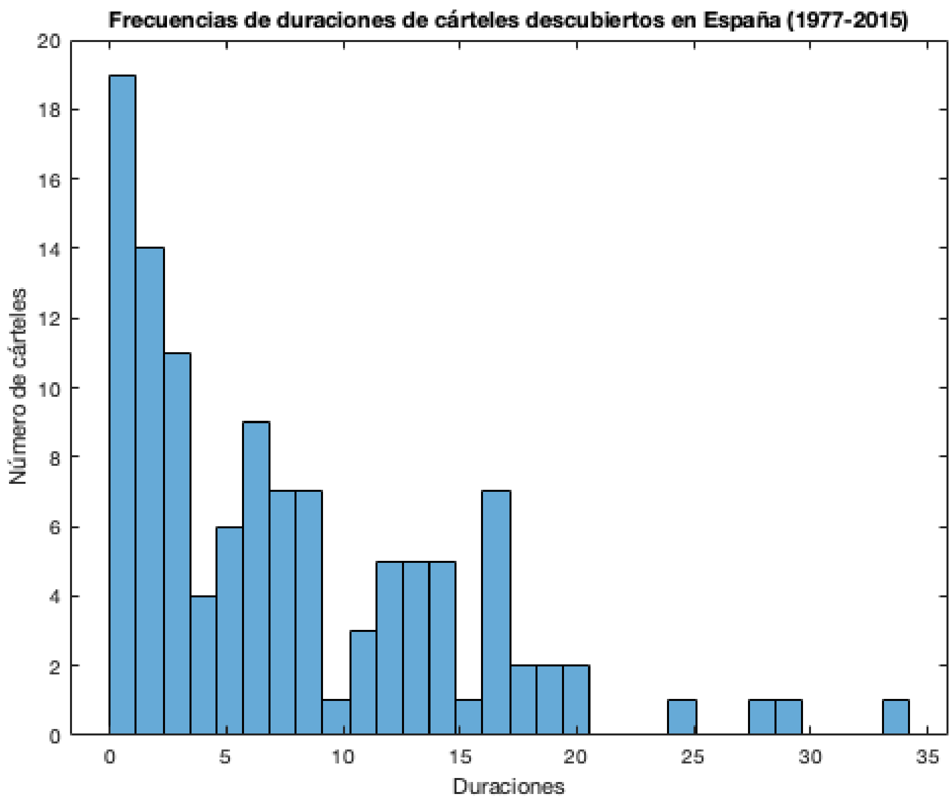
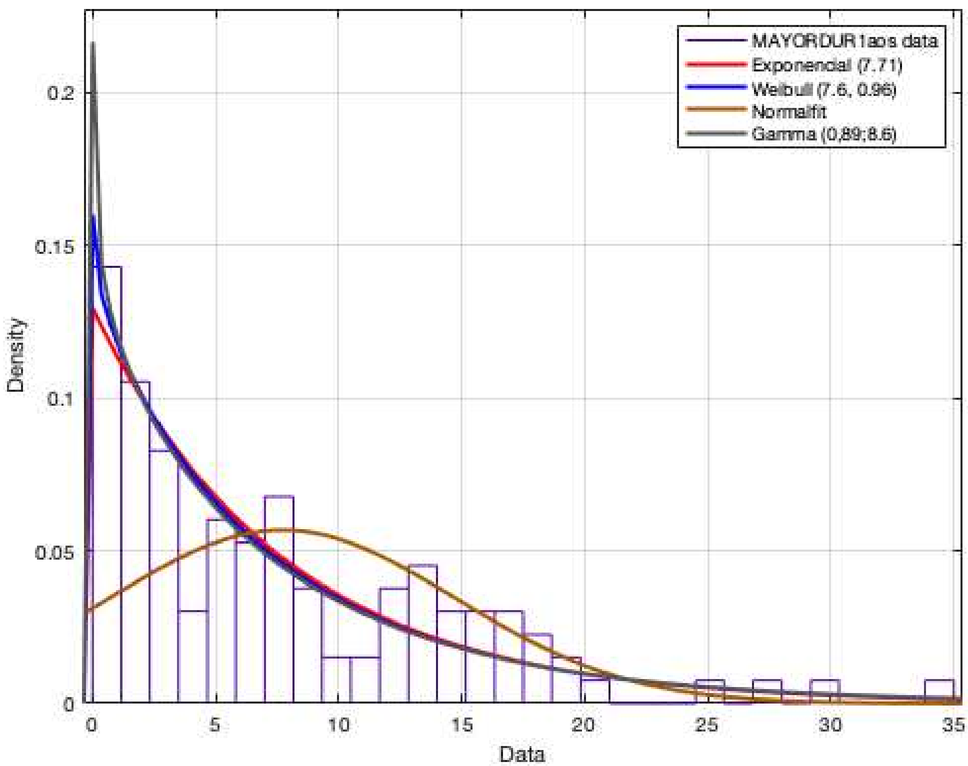


Figure 1 shows the duration frequencies of 114 cartels discovered. For the duration calculated using the Duration 1 method the average is about 7.6 years with a minimum of 0.04 years and a maximum of 34 years. The median (5.83 years) is lower than the average (7.65 years) which reflects the bias of the distribution to the right, so there are more durations below the average than above it. There is a lot of variability in the data, which is reflected in a standard deviation greater than the mean. In the case of Duration 2, the mean is 5.55 and the median is 4.92 and the data vary between a minimum of 0.04 and 29.26 years.

These results are similar to those of other studies, for example , (Levenstein and Suslow 2006) with an average duration of 5 years and a distribution of a bimodal nature or two maximums, that is, numerous cartels with a reduced duration (less than one year) numerous with durations between 5 and 10 and, to a lesser extent, decades of decades.

In Figure2, the same histogram of the durations in years is shown, but now with several curves representing the best fit of the data to distributions of typical probability such as Normal, Gamma, Exponential or Weibull. As seen in the graph, the last two are the ones that best fit the data. Conversely, a good fit is not obtained using the normal or Gaussian distribution:

Figure 2: Fitting standard probability distributions to the frequency histogram:



Most studies on detection probabilities[30] have considered that the data of cartel durations follow exponential distributions, since they are the ones that best fit strictly positive and *memoryless* stochastic processes. , where the result at a time only depends on the immediately preceding moment.

Although we only have data from a sample of discovered cartels we can try to adjust them to several theoretical distributions using the technique of maximum probability or likelihood (MLE):

Table 2: Results of adjustment to different probability distributions

[30] Bryant-Eckard they used an exponential function. (Hellwig and Hüschelrath 2018) Used one distribution Weibull for the case of cartels in the EU.

Type of Distribution	Normal	Gamma	Weibull[31]	Exponential
MLE	-383.619	-346,49	-346.84	-346,9
Stocking	7.71723	7,72	7.72885	7,72
Variance	49.4546	66,43	64.2586	59.56
Parameter A (Shape)	7.71723 (Mu)	0.896525	7.60567	0.522418 (Mu)
Parameter B (Scale)	7.0324 (Sigma)	8.60794	0.964349	

The best result is obtained with any of the distributions of the exponential family versus the normal distribution, for example, whose result in terms of likelihood is lower. Bryant and Eckard also concluded that the data obeyed an exponential function.

Next, we leave aside the analysis of these particular data and their conclusions, to go on to build a general theoretical model that explains the process through which a cartel goes through until its disappearance and with which we can explain these data of particular durations of cartels detected.

3.- Estimation of the probability of detection

When estimating the probability of detection from the data we have, the durations of the cartels detected, we need a theoretical model that explains the process of creation and destruction of cartels during a period of time from which to draw conclusions.

[31] The Gamma, Weibull and Exponential distribution are related since the Gamma distribution with a parameter equal to 1 turns out to be equivalent to an Exponential of parameter 1/Beta, while a Weibull distribution with a parameter equal to 1 turns out to be an Exponential of average equal to Beta.

We analyze, first, the model of the first authors, Bryant and Eckard, who used a continuous mathematical model of birth and death, a model known as the Poisson process, although the model is limited to the extent that its results do not take into account the unstable nature of cartels that can disappear spontaneously, undetected.

Secondly, we analyze the model developed by Harrington and Wei, which avoids analyzing the process of creating cartels and concentrates on the process of disappearance taking into account not only the death by detection of the cartel but also by its internal viability. It seems, therefore, that the process of disappearance or death of the cartels would be based on this model, but a theoretical analysis of the process of birth of the cartels remains to complete the theoretical study. The behavior of the agents when deciding between creating a cartel or competing should be introduced into the model (see, for example, the study carried out by Motta and Polo). This remains pending for future study.

Next, we analyze the two methods applied to the data obtained for the case of Spain:

3.1.- Bryant and Eckard Model

We begin by analyzing the method used by Bryant and Eckard, who used a continuous random process of birth – death to explain the evolution of cartels over time.

The authors start from restrictive assumptions: cartels are created and destroyed according to a random and independent process, where at the end of the period all cartels disappear by detection, there are no other options. On the other hand, it is assumed that a long period of time has already passed since the process of creating cartels began so we would be in a steady state, that is, the rates of creation and destruction of cartels are constant.

In this context, considering these assumptions the authors construct a function of plausibility or probability of observing *n* cartels of durations in a time interval: $L_1, \ldots, L_n, T_2 - T_1$

$$V(L_1, \ldots, L_n) = \theta^n \lambda^n e^{-\theta(T_2 - T_1)]} e^{-\lambda \sum L_i}$$

Where the parameters θ and λ are the probability of birth and probability of detection, respectively. Taking logarithms and deriving one arrives at the maximum plausible estimators (MLE) of the parameters of interest:

$\frac{1}{\theta}$= $(T_2 - T_1)/n$ *and* $=\Sigma(Li)/n\frac{1}{\lambda}$

Therefore, at steady state, the parameters, $1/\theta$ *and* $1/\lambda$ are simply the average of the times between cartels and the average of cartel durations. (Keiding 1975, p. 1) comes to the same conclusion. (an explanation of the process is given in Annex 1).

The estimated parameters λ and θaccording to the methodology of these authors applied to the data of durations in Spain would be:

Table 3:Bryant-Eckard estimate results applied to cartel data in Spain

	Measurement Duration 1[32]	*Measurement Duration 2*
Probability of Detection ()λ	*0,1295*	*0,18018*
Average duration	*7,71*	*5,5*

The results of the estimate are shown in Table 3. The average duration to detection is 7 years, using the first method of calculating durations and 5.5 for the second method. The annual probability of detection is between 13% and 18%. These results are similar to those obtained by Bryant and Eckard, who stated that the average duration of a cartel in the US was 5 to 7 years on average while the estimated annual probability was between 12.8% and 17.4%.

To conclude, the authors elaborate a function to explain the cartels that survive this detection process in each period, a decreasing function as it approaches the end of the period which agrees with the initial assumptions they established. According to the authors, since we observe the process only betweenT_1 and T_2 we expect the live cartels in *t* to disappear as we approach T_2,counting the posters alive in t, and still alive in T_2, they reach that the surviving cartels in each period t are:

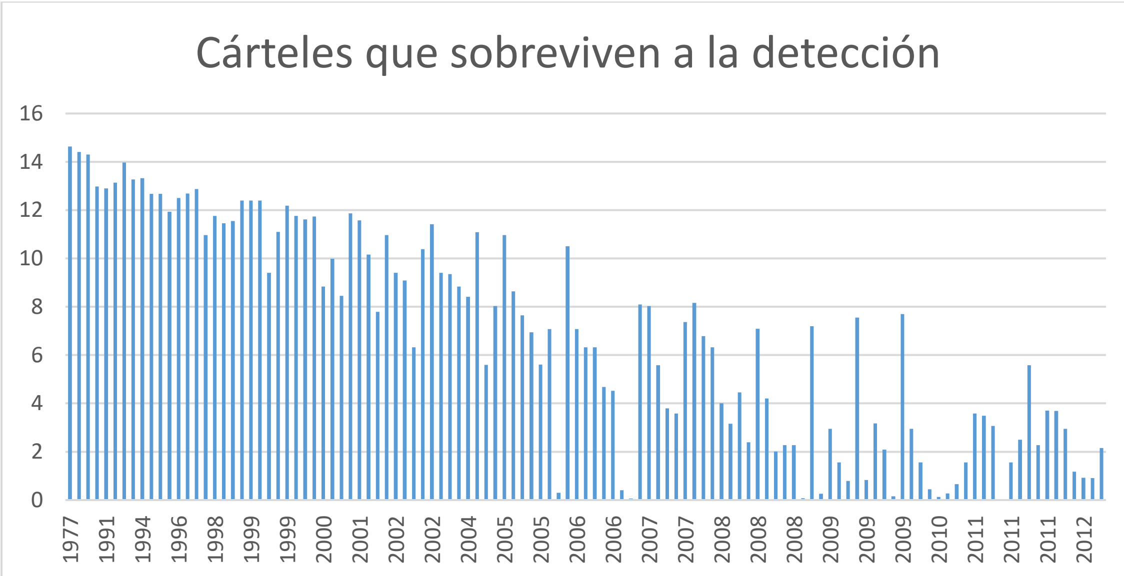
$$(\frac{\theta}{\lambda})(1-e^{\{\lambda(T_2-t)\}}) ,$$

If we use the same formula in our sample of cartels detected in Spain we also obtain that, as t approaches T_2,*N(t)* decreases. We have built the same graph with data from Spain to get the following graph:

Figure 3: Posters that survive detection in Spain using the Bryant-Eckard methodology.

32

DUR1	*Difference between the maximum date of the final and the minimum date of the start of the poster*
DUR2	*Difference between the average of the end date and the average of the start date*

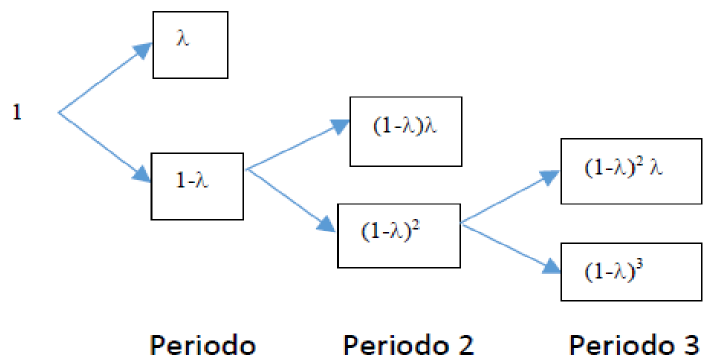


3.2.- Methodology based on a **Discrete Model. Homogeneous posters.**

The above analysis suffers from several limitations, the fundamental one is that in this process of creation and destruction of cartels does not take into account the unstable nature of the same and therefore it should be taken into account that a part of the cartels disappear for internal reasons and not for detection. It also seems more realistic to consider the process of birth and death as a discrete process, rather than a continuous one as the first authors proposed.

To make a theoretical model on this process we must explain the factors that give rise to the creation and destruction of cartels. As we have said, to include the creation of cartels in this model we should have a model that explains this process that we do not have at the moment. However, if we can explain how from a population of cartels that we assume is stable over time, it is decreasing because of two factors: detection and destabilization. Starting from an initial number of cartels standardized to 1 (Harrington and Wei 2017), in each period a proportion is discovered λwhile 1- λ are not discovered (we assume that it is the same in all λperiods):

Figure 4: Proportion of cartels in each period (detection only)



In n periods, the cartels discovered would be $\lambda(1\text{-})\lambda^{n\text{-}1}$ while those that are not discovered and, therefore, survive would be as a sum $\sum_{n=1}^{\infty}(1-\lambda)^{n-1}$of columns 3 and 2, respectively:

Table 4: Proportion of cartels in each period (detection only)

Time period (n)	Live posters	Posters detected
1	*1*	λ
2	$(1\text{-})\lambda$	$(1\text{-})\lambda\lambda$
3	$(1\text{-})\ \lambda^2$	$(1\text{-})\ \lambda^2\lambda$
4	$(1\text{-})\ \lambda^3$	$(1\text{-})\ \lambda^3\lambda$
…	…	….
Σ	$\sum_{n=1}^{\infty}(1-\lambda)^{n-1}$	$\lambda\sum_{n=1}^{\infty}(1-\lambda)^{n-1}$

The sum of the column of cartels detected would be equal to:

$$\lambda\sum_{n=1}^{\infty}(1-\lambda)^{n-1} = \frac{(1-\lambda)\lambda}{1-(1-\lambda)\lambda} = \frac{(1-\lambda)\lambda}{1-\lambda+\lambda^2} \qquad (1)$$

This would be a discreet result equivalent to the model of Bryant and Eckard, although in this case we only take into account the disappearance of the cartel by detection, as Harrington does.

Now we introduce greater realism to this sequence if we take into account not only the detection but also the internal dissolution of the cartel to explain the disappearance of the cartels.

In this case, the equation consists of starting from the initial number of cartels normalized to 1, but now in each period a proportion is discovered while 1- are not discovered,λλ(we assume that it is the same in all periods). Now of these λundiscovered cartels there is a proportion q that disappear due to internal destabilization. The sequence would then be:

Figure 5: Proportion of cartels in each period (detection and destabilization)

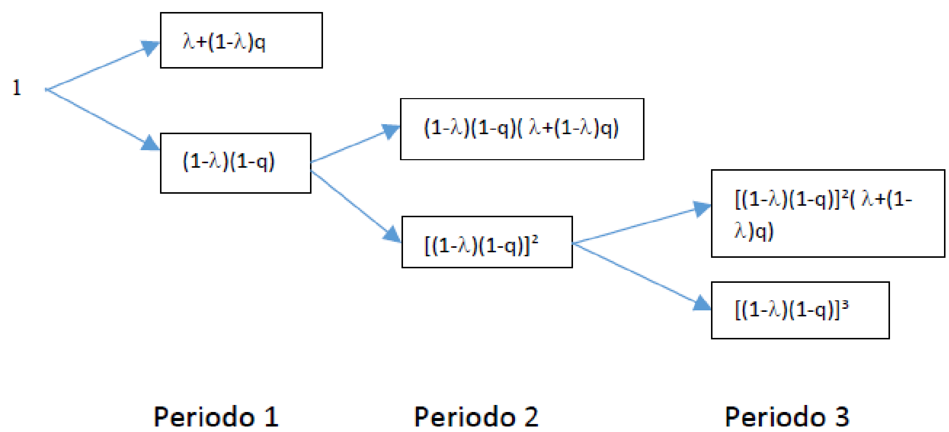


Table 5: Proportion of cartels in each period (detection and destabilization)

Time period (n)	Number of live posters in each period t	Number of dead posters (detection or internal dissolution) in each period t
1	1	$\lambda+(1-)q\lambda$
2	$(1-)\,(1-q)\lambda$	$[(1-)(1-q)]\,(\,+(1-)q)\lambda\lambda\lambda$
3	$[(1-)(1-q)]\,\lambda^2$	$[(1-)(1-q)]\,\lambda^2(\,+(1-)q)\lambda\lambda$

4	[(1-)(1-q)] λ^3	[(1-)(1-q)] λ^3(+(1-)q)$\lambda\lambda$
...	...	....
Σ	$\sum_{n=1}^{\infty}[(1-\lambda)(1-q)]^{n-1}$	$[(\lambda+(1-\lambda)q)]\sum_{n=1}^{\infty}[(1-\lambda)(1-q)]^{n-1}$

The sum of the column of missing cartels taking into account detection and destabilization, weighted by the duration of each cartel is equal to the average duration of the cartels (see Harrington):

$$[(1-\lambda)(1-q)]\sum_{n=1}^{\infty} n[(1-\lambda)(1-q)]^{n-1} = [(\lambda+(1-\lambda)q)]\frac{1}{1-(1-\lambda)(1-q)}$$

$$= \frac{1}{q+\lambda-\lambda q} \qquad (2)$$

This would be the result in the event that we consider that λ and $q$they are the same for all cartels, that is, they are homogeneous. However, it is more realistic to consider that they are heterogeneous, so they have different parameters since their stability can be affected by different characteristics depending on the sector in which they are located. Also the detection parameter varies depending on the ease or difficulty to be detected which also depends on the sector and the number of companies that can facilitate detection.

In order to obtain empirical results, it is necessary to make a hypothesis about what values the parameters adopt and the relationship they have between them. λqWe see thisin thenext section.

3. 2. 1. Discrete and heterogeneous model. Estimation of the **probability of detection.**

In this case we consider (Harrington and Wei 2017) the most realistic hypothesis: the posters are not homogeneous and, therefore, each poster has different parameters (λ, p). Before continuing, we t know that to make an assumption about how p behaves and λ, in this sense we consider that each parameter follows a probability distribution of type Beta defined so that the joint distribution of the two parameters is defined by five parameters $\theta=\{,c\mu_{\lambda p},,c\mu_{\lambda p},\}$, $\psi$where μ is the mean and c the coefficient of variation for each parameter, while it would be the correlation coefficient between bothψ parameters. The different combinations of the parameters are collected in the joint density function. $k(\lambda, p)$

Now it is a question of selecting values of the parameters with which we obtain the highest values of likelihood using the following formula based on the theoretical model of the previous section, with which we obtain the probability that the card has the duration *t* conditioned to be discovered:

$$\iint \frac{p[(1-\lambda)(1-p)]^{t-1}}{\iint (\frac{p}{\lambda+p-\lambda p})k(\lambda,p)d\lambda dp} k(\lambda,p)d\lambda dp(3)$$

To estimate the parameters, we use a Bayesian approximation that requires, first of all, to specify some a priori values of the parameters, and that given the ignorance we have about the parameters we assume that they follow a uniform distribution, which assigns the same probability to each value. The next step is to calculate the values a posteriori that are obtained by multiplying the values a priori by their plausibility calculated according to (3).

We will only use those values whose plausibility (or probability) is greater than what would be obtained if we simply use values obtained from a uniform distribution, rejecting the values with less plausibility. We repeat this process with 200,000 values of the parameters, accepting only 1% of the values.

That set of parameter values results in inserts (3) which is derived from the theoretical model and which reflects the unconditional probability of cartel death:

$$\iint [1-(1-\lambda)(1-\mathrm{q})]k(\lambda,\mathrm{q};\theta)d\lambda d\mathrm{q} \qquad (4)$$

3.3. 1. Empirical result

As we can see in the following table we have obtained some values for the parameters *p* (probability that the cartel disappears due to destabilization) and (probability that the cartel disappears by detection), λ once the model is applied, lower and with a lower standard deviation than the apriori values, which as we remember come from a di totally random attribution.

The mean values of $\lambda$and a $q$posteriori are 0.0854 and 0.0831, lower than the a priori values of 0.110, in thepresentcases, and a reduction also in the values of the coefficients of variation from (0.449;0.446) to (0.4028;0.3581),which means that cartels are more likely to have low values of probability of detection and destabilization, and that their variability measured through the coefficients is smaller.

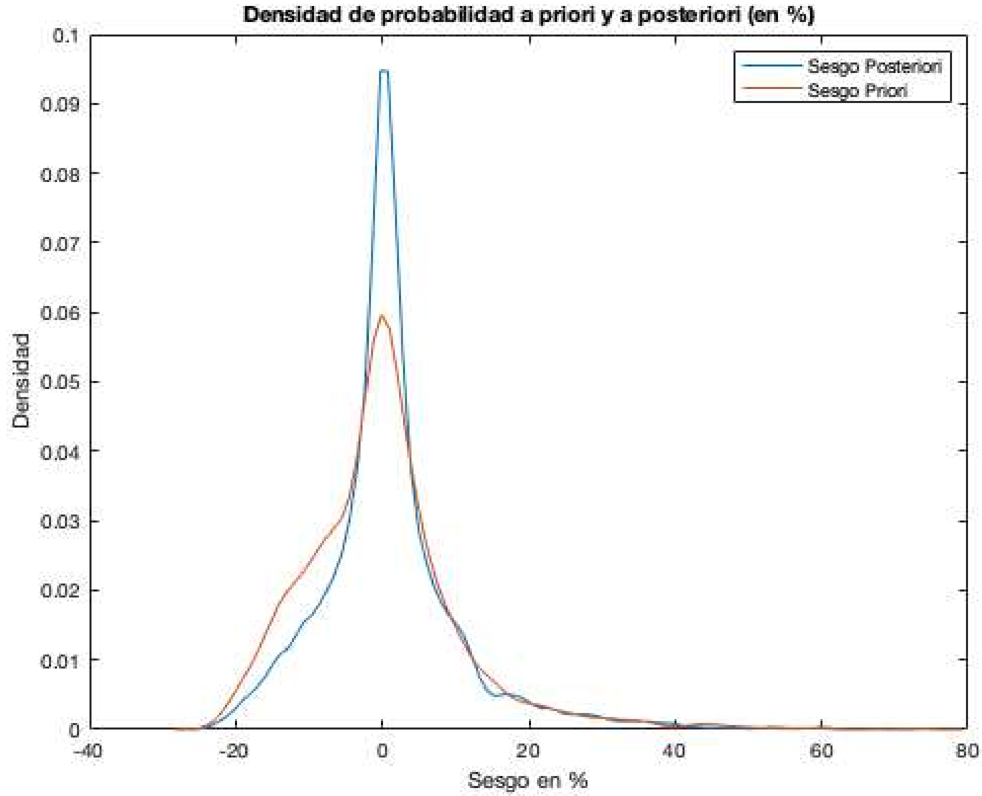
Table 6: Statistical results

	Stocking Prior	**Desv. Est.**	**Stocking Posterior**	**Desv. Est.**

Average Lambda value (λ)	0,109	0,0506	0,0854	0,0384
Mean Value Rho (q)	0,110	0,0505	0,0831	0,035
Coef. Lambda	0,449	0,261	0,4028	0,2594
Coef. Rho	0,446	0,2601	0,3581	0,2515
Correlation	5,46	2,5964	5,4233	2,609
Prob. Cartel Death	0,2066	0,063	0,162	0,0256

The bias or error derived from using a sample of detected cartels versus the alternative of using the entire cartel population is close to 0 as seen in the graph below, with a standard deviation a posteriori of 9.5% lower than that obtained a priori of almost 11%. Therefore, the a posteriori distribution is less dispersed than the a priori and more concentrated in values around an error of 0.

Figure 6: Priority and posterior probability density



With two typical deviations above and below the mean in the a priori distribution, the bias would be between 20% and -18%, while in the case of the a posteriori distribution it is between 19% and -18%. This means that the average duration of cartels discovered obtained in our sample of cartels discovered of 7.6 years, could be at most overestimatedby19%, or at least underestimated by 18% the duration of the cartels, that is, the actual duration of the cartels would be in an interval between 6.2 and 9.1 years. If we use the a priori distribution the dispersion would be greater, and the interval would be between 5.96 and 9.21 years.

With this, we can conclude that the error of using a sample of discovered cartels as a proxy for all cartels is small.

Second, we can find the probability of a cartel's death using (4).

Figure 7 shows that the highest density is around 0.16, which is the mean. In a range of two standard deviations from the mean, between 0.11 and 0.21, the results of previous studies are included. For example, the result using the Bryant-Eckard model was 13-17%. Figure 8 shows how the probability density is concentrated in that interval.

Therefore, we agree with the conclusions of Harrington's paper, the error in using a sample of discovered cartels instead of the entire cartel population is small, so the results using a restricted sample are acceptable.

In any case, the results are always about probabilities of death of the cartel, so the correct thing would be to talk about probabilities of disappearance or death both in the case of Bryant Eckard and this last result.

Figure 7: Density unconditional probability of death

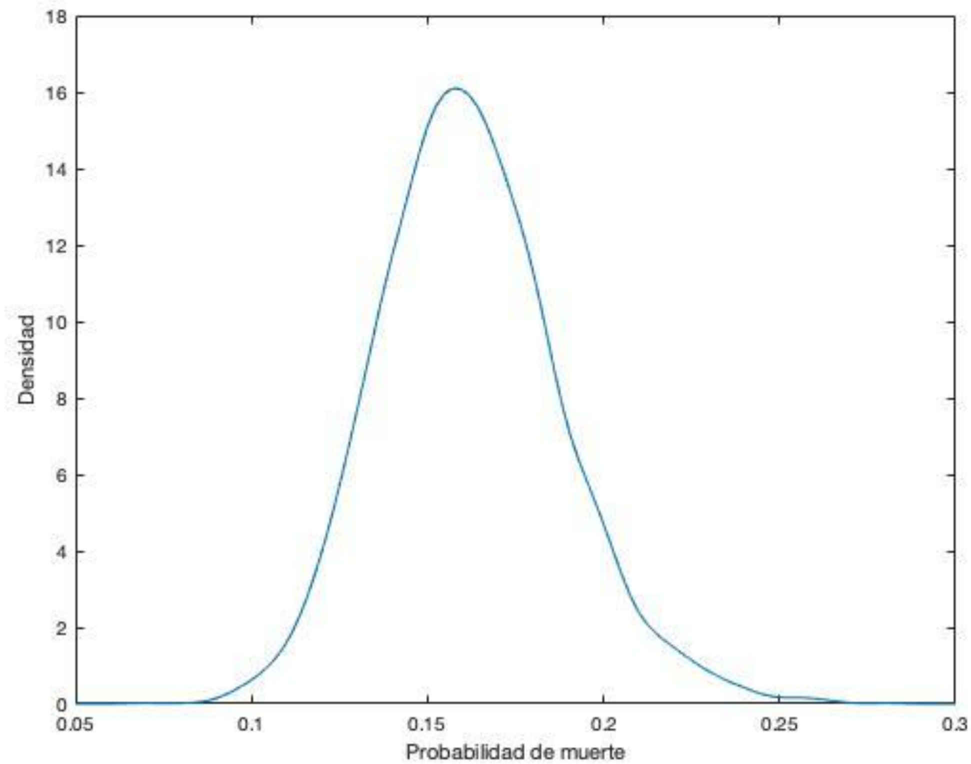
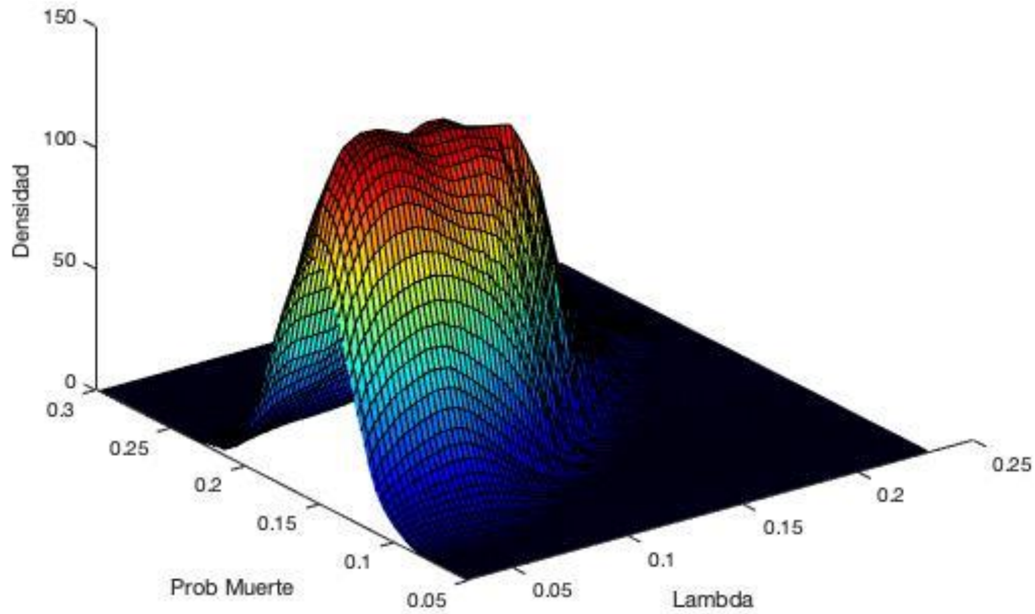


Figure 8: Density probability of death versus parameter λ



6. Conclusions

Bryant and Eckard were the first to estimate the probability of cartel detection using a sample of cartels in the U.S. during the years 1961 and 1988. The probability of detection obtained was between 13% and 17% and the average duration of seven years. Using the same procedure, Combe and Monnier came up with an annual probability of 12.9% and 13.2%, for a sample of posters in the EU between 1969 and 2007.

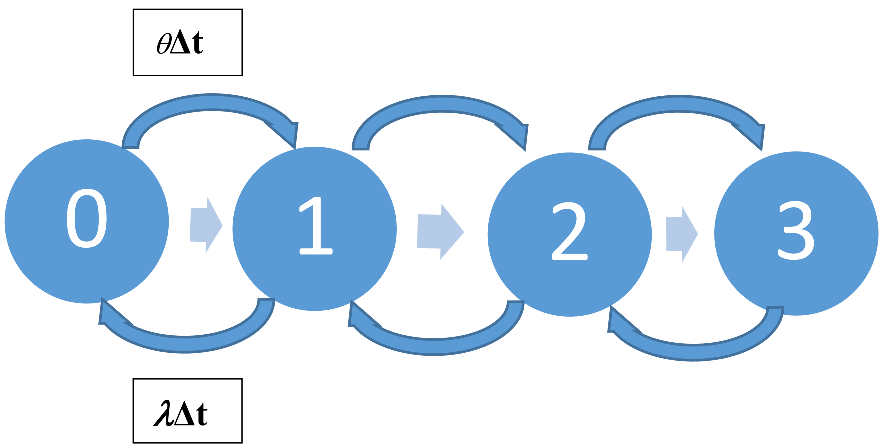
In this article the same methodology is applied to calculate the duration until detection and the probability of detection with data from Spain, and a probability of between 13% and 18% is obtained. The average number of cartels detected according to Bryant et al. at any given time is between 36 and 50. In our model the average number of cartels would be between 16 and 22.

These authors wonder whether these results can be extended to those undetected cartels, although it is apparently not possible to estimate their number, they consider that if the duration of a discovered cartel is not greater than that of the undiscovered, then they deduce that the duration estimates are also applicable to the undiscovered (as a lower limit), and the estimated probabilities are an upper limit of the probability that a cartel will disappear not by detection but by internal disagreements.

In the second part, the article applies a model developed by Harrington that combines both the sample of detected cartels, the theoretical model of the birth of cartels and our a priori belief about the behavior of the cartels, obtaining a probability of disappearance of the cartels of 16.2%, within the interval obtained using the Bryant-Eckard model, although now as we have seen we have taken into account the entire possible population of cartels and not only those discovered.

Annex 1: Bryant-Eckard:

Bryant-Eckard approximation can be explained by considering that the creation of cartels follows a Poisson distribution of parameter, θwhile the duration of the cartel follows a negative exponential distribution. $\lambda e^{-\lambda t}$ The following diagram explains the process of creating and discovering cartels (or death because we believe they are only destroyed with detection) (Ross, n.d.) , the numbers reflect the number of cartels and the arrows the transitions from one state to another, i.e. from one number of cartels to another. There are only two options, either the number of cartels is increased by one or it is reduced by one:



In state 0herself may be because a poster has been discovered, so it goes from 1 to 0 with probability λ, or because no cartel is born for a short time interval △t. This translates into a transition equation:

$$\mathbf{p_0 = p_0\,(1 - \Delta t) + p\theta_1\,\lambda\,\Delta t}$$

Successively we would have:

$$\mathbf{p_1 = p_0\,\theta\,\Delta t + p_1\,(1 - \Delta t - \Delta t) + p\theta\lambda_2\,\lambda\,\Delta t}$$

$$\mathbf{p_2 = p_1\,\theta\,\Delta t + p_2\,(1 - \Delta t - \Delta t) + p\theta\lambda_3\,\lambda\,\Delta t}$$

$$\mathbf{p_3 = p_2\,\theta\,\Delta t + p_3\,(1 - \Delta t - \Delta t) + p\theta\lambda_4\,\lambda\,\Delta t}$$

...

Simplifying we get:

$$\mathbf{p_0\ \theta\,\Delta t = p_1\ \lambda\ \Delta t}$$

$$\mathbf{p_1(\Delta t + \Delta t) = p\theta\lambda_0\ \theta\,\Delta t + p_2\ \lambda\ \Delta t}$$

$$\mathbf{p_2(\Delta t + \Delta t) = p\theta\lambda_1\ \theta\,\Delta t + p_3\ \lambda\ \Delta t}$$

$$\mathbf{p_3(\Delta t + \Delta t) = p\theta\lambda_2\ \theta\,\Delta t + p_4\ \lambda\ \Delta t}$$

If we eliminate Δt and simplify we get:

$$\mathbf{p_0\ \theta = p_1\lambda}$$

$$\mathbf{p_1\ \theta = p_2\lambda}$$

$$\mathbf{p_2\ \theta + p_2 = \lambda\ p_1 + \theta\ p_3\lambda}$$

$$\mathbf{p_3\ \theta + p_3 = \lambda\ p_2 + \theta\ p_4\lambda}$$

...

Which allows us to reach:

$$\mathbf{p_0\ \theta = p_1\lambda}$$

$$\mathbf{p_1\ \theta = p_2\lambda}$$

$$\mathbf{p_2\ \theta = p_3\lambda}$$

$$\mathbf{p_3\ \theta = p_4\lambda}$$

...

Or what is the same:

$$\mathbf{p_1 = /\ p\theta\lambda_0}$$

$$\mathbf{p_2 = /\ p\theta\lambda_1}$$

p_3 = / p$\theta\lambda_2$

p_4 = / p$\theta\lambda_3$

...

So you get to:

p_1 = / p$\theta\lambda_0$

p_2 = (/)$\theta\lambda^2$ **p_0**

p_3 = (/) $\theta\lambda^3$ **p_0**

p_4 = (/) $\theta\lambda^4$ **p_0**

...

Since **$p_0 + p_1 + p_2 + p_3 + ... = 1$** and replacing the previous p_i in this equation we obtain the following geometric sequence:

$p_0 + (\theta / \lambda)^1 p_0 + (\theta / \lambda)^2 p_0 + (\theta / \lambda)^3 p_0 + ... = 1$

$p_0 (1+(\theta / \lambda)^1 + (\theta / \lambda)^2+(\theta / \lambda)^3+.... = 1$

Which can be simplified if we subtract this sequence by it multiplied by / a:θλ

$$P_0 = 1 - (\frac{\theta}{\lambda})$$

The steady state is obtained as follows:

p_0 = (/)$\boldsymbol{\theta\lambda^0}$ **(1-(/))**$\boldsymbol{\theta\lambda}$

p_1 = (/)$\boldsymbol{\theta\lambda^1}$ **(1-(/))**$\boldsymbol{\theta\lambda}$

p_2 = (/)$\boldsymbol{\theta\lambda^2}$ **(1-(/))**$\boldsymbol{\theta\lambda}$

p_3 = (/)$\boldsymbol{\theta\lambda^3}$ **(1-(/))**$\boldsymbol{\theta\lambda}$

....

Or what is the same:

$P[n] = (\theta/\lambda)^n(1 - \theta/\lambda)$

If we limit the time period between t_1 and t_2 ($t_2>t_1$) (Clarke n.d.) , assuming that θ /<1, reaches: λ

$$\frac{1}{\theta} = (t_2 - t_1)/n \text{ and } = \Sigma(Li)/n\frac{1}{\lambda}$$

Annex: Estimation of the number of uncovered cartels

Although this calculation was not performed by Bryant-Eckard or Harrington, we tried to approximate the number of existing but undiscovered cartels from the sample of cartels discovered using also a survival model.

We maintain the same assumptions of independence from the durations of the different cartels and that these are created through a Poisson process. There are, therefore, a number of cartels discovered during the period we know N_1 and a number of cartels created but not discovered during the N_2 process. All the cartels created end up being discovered.

Assuming that we know of the estimate in the previous section, we can calculate the number of cartels created but not discovered during the process λ(Ross, pg. 316) using the following equation:

$$N_2(t) = \frac{n\lambda(1 - e^{-\frac{T2-T1}{\lambda}})}{(T_2 - T_1) - \lambda(1 - e^{-\frac{T2-T1}{\lambda}})}$$

Now we give values to the parameters obtained from the estimate of the previous section: t = 14,609 days, average duration from creation to detection = 2817 days, and number of cartels detected during the period n = 115, we would obtain 27 existing but undetected cartels:

$$n = \frac{(115x2.817)x(1 - \exp(-\frac{14.609}{2.817}))}{(14.609 - 2.817)x(1 - \exp(-\frac{14.609}{2.817}))} \approx 27$$

This estimate has the limitation that the rate of birth of cartels is constant during the period.λ

Bibliography

Bryant, Peter G, Edwin Woodrow Eckard, Peter G Bryant, and Edwin Eckard. 1991. "Price Fixing: The Probability of Getting Caught." *The Review of Economics and Statistics* 73 (3): 531–36. https://econpapers.repec.org/article/tprrestat/v_3a73_3ay_3a1991_3ai_3a3_3ap_3a531-36.htm.

Buccirossi, Paolo, and Giancarlo Spagnolo. 2005. "Optimal Fines in the Era of Whistleblowers-Should Price Fixers Still Go to Prison?"

Clarke, A. Bruce. n.a. "Maximum Likelihood Estimates in a Simple Queue." *The Annals of Mathematical Statistics*. Institute of Mathematical Statistics. Accessed December 2, 2018. https://doi.org/10.2307/2237068.

Combe, Emmanuel, Constance Monnier, and Renaud Legal. 2008. "Posters: The Probability of Getting Caught in the European Union." *Bruges European Economic Research Papers*. https://ideas.repec.org/p/coe/wpbeer/12.html.

García Verdugo, Javier, Carlos Merino, and Lorena Gómez Cruz. 2016. "Economic Valuation of the Sanctions of the Spanish Competition Authority (2011-2015)." *Competition Yearbook, No. 1, 2016, pp. 345–379,*no. 1:345–79. https://dialnet.unirioja.es/servlet/articulo?codigo=6044346.

Harrington, Joseph E. 2014. "Penalties and the Deterrence of Unlawful Collusion." *Economics Letters* 124 (1): 33–36. https://doi.org/10.1016/J.ECONLET.2014.04.010.

Harrington, Joseph E., and Myong-Hun Chang. 2009. "Modeling the Birth and Death of Cartels with an Application to Evaluating Competition Policy." *Journal of the European Economic Association* 7 (6): 1400–1435. https://doi.org/10.1162/JEEA.2009.7.6.1400.

Harrington, Joseph E., and Yanhao Wei. 2017. "What Can the Duration of Discovered Cartels Tell Us About the Duration of All Cartels?" *The Economic Journal* 127 (604): 1977–2005. https://doi.org/10.1111/ecoj.12359.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153. https://doi.org/10.2307/1912352.

Hellwig, Michael, and Kai Hüschelrath. 2018. "WHEN DO FIRMS LEAVE POSTERS? DETERMINANTS AND THE IMPACT ON CARTEL SURVIVAL." *International Review of Law and Economics* 54 (June): 68–84. https://doi.org/10.1016/J.IRLE.2017.11.001.

Hyytinen, Ari, Frode Steen NHH Bergen, Otto Toivanen HECER, Frode Steen, Otto Toivanen, Susanna Fellman, Joe Harrington, Vesa Kanniainen, and Howard Smith. 2010. "Uncovered Posters*." https://helda.helsinki.fi/bitstream/handle/10138/16777/cartelsu.pdf?sequence=1.

Katsoulacos, Yannis, and David Ulph. 2013. "Antitrust Penalties and the Implications of Empirical Evidence on Cartel Overcharges." *The Economic Journal* 123 (572).

Keiding, Niels. 1975. "Maximum Likelihood Estimation in the Birth-and-Death Process." *The Annals of Statistics* 3 (2): 363–72. https://doi.org/10.1214/aos/1176343062.

Levenstein, Margaret C, and Valerie Y Suslow. 2006. "What Determines Cartel Success?" *Journal of Economic Literature* 44 (1): 43–95. https://doi.org/10.1257/002205106776162681.

Miller, Nathan H, Severin Borenstein, Joseph Farrell, Russell Pittman, Carl Shapiro, John Sutton, Sofia Villas-Boas, and Gregory Werden. n.a. "Strategic Leniency and Cartel Enforcement." Accessed September 15, 2018. https://doi.org/10.1257/aer.99.3.750.

Ormosi, Peter L, Stephen Davies, Joseph Harrington, Morten Hviid, Bruce Lyons, Franco Mariuzzo, Matthew Olczak, Greg Shaffer, Iwan Bos, and Greg Sha¤er. 2011. "A Tip of the Iceberg? The Probability of Catching Cartels A Tip of the Iceberg? The Probability of Catching Cartels I Am Grateful for the Useful Comments To." http://competitionpolicy.ac.uk/documents/8158338/8253131/CCP+Working+Paper+11-6.pdf/9d9c5c1d-35fe-4139-baa0-06838f7d4fe9.

Park, Jihyun, Juhyun Lee, and Suneung Ahn. 2018. "Bayesian Approach for Estimating the Probability of Cartel Penalization under the Leniency Program." *Sustainability* 10 (6): 1938. https://doi.org/10.3390/su10061938.

Ross, Sheldon M. n.d. *Introduction to Probability Models*.

Ross, Sheldon M. n.d. "Introduction to Probability Models, Tenth Edition." Accessed November 1, 2018. https://pdfs.semanticscholar.org/13a3/db2415fb38ec77c165a5dfea1913a22655be.pdf.

Appendix: An estimation of demand in telecommunications market in Spain

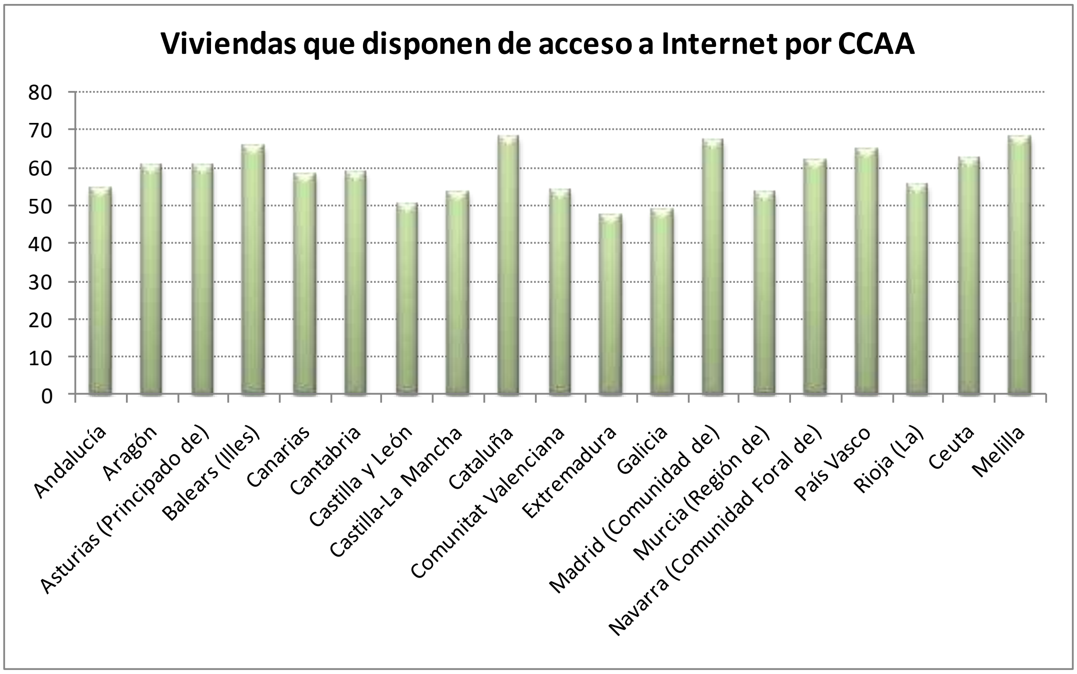
In this annex, the demand for the Internet in Spain is analyzed. To do this, we use the Survey on equipment and use of information and communication technologies in homes and using econometric techniques it is provided an estimate internet demand *functions, incorporating the sociodemographic characteristics of individuals. We have two econometric models, one for broadband access at home and* anot*her* for the use of *the internet. We found that 59.1% of the Spanish population has internet access at home, and 57.8% use broadband connections.*

Study of the demand for internet access and use in Spain

The objective of this section is to analyze the demand for the Internet in Spain. To do this, we used the *Survey on equipment and use of information and communication technologies in* multi-year households as well as econometric techniques to estimate the demand for the internet, incorporating the sociodemographic characteristics of individuals. We start with a graphical analysis of the data that allows us to search for relationships between the distinct characteristics. Then we will analyze two econometric models, one for broadband access in homes and another for the different uses of the internet. We found that 59.1% of the Spanish population has internet access at home, and 57.8% use broadband connections. This demand is positively related to income and other technological attributes and negatively to socio-demographic attributes such as habitat and age.

The demand for the internet in Spain

In Spain, the main study of the demand for internet use and access stands out as the National Institute of Statistics (INE) called "Survey on equipment and use of information technologies at home (TIC-H)" which has information since 2001. The data in this section were taken from the MULTI-year INE survey (2004, 2008, 2009 and 2010) on equipment and use of information technologies in the home (TIC-H). The first thing that is observed is the high heterogeneity in the access and use of the internet in Spain. The general figures contained in Figure 1 show that, while internet access is increasingly common in Spanish households (59.1% of homes in 2010 compared to 54% in 2009), the distribution of this access is not homogeneous.



SOURCE: NATIONAL INSTITUTE OF STATISTICS

The figure allows us to deduce that internet access is more widespread in Catalonia (68.3%), the Basque Country (64.3%), Madrid (66.9%) and Melilla (68.1%), compared to Extremadura (47.4%), Castilla-La Mancha (56.3%) and Galicia (48.9%).
As a starting point to analyze the demand for the internet is the conventional telephone service, but unfortunately, we must point out notable differences, (always distinguishing between access and use as different concepts). The first is that differences in end-use are much more important in the demand for the Internet than in that of conventional telephony. Although there are many different uses in conventional telephony, in none of these is the speed (or bandwidth) of real importance.

In the demand for the Internet, end-use is important, since productivity may be, for example, of little importance for conventional e-mail transmission, but crucial for the exchange of large data files, graphics, videos or photographs. In conventional telephony, traffic per actual minute is the appropriate output measure; However, for many Internet end uses, bits per second (i.e., speed) is the crucial magnitude. Among other things, this means that the end uses of the Internet are important when analyzing Internet demand. This brings us to the second complication: the determination of the ability to access. In traditional telephony, greater need for voice capacity can be met by adding more lines. As such, speed is not a problem. However, in most end uses of the Internet, however, this is not the case. Speed is very important, and can only be obtained through increasing the capacity of incoming access, not simply multiplying the number of lines.

The implications for modelling the demand for Internet access and use are as follows:
1. Demand should be measured by usage (e.g. online minutes or sites visited) and in Mbits per second, rather than simply in terms of minutes as in regular telephony demand.
2. Consumer surplus shall, at least in principle, be measured over all end-uses.
3. The demand for access should be addressed based on different types of access (i.e. dial-up, ADSL or cable).
In short, in this section of the work, we estimate a model of access and use to the internet and compare the types of broadband connections using the characteristics of the consumer contained in a survey conducted by the National Institute of Statistics (INE) from two points of view: On the one hand, we will take into consideration the individual conditioned by education, experience

and income which, in turn, is conditioned by the size of the family and on the other, the family group, equipment and household income.

Data: Survey on equipment and use of information technologies in the home (TIC-H)

The objective of this section is to provide empirical evidence on the use and access to the internet in Spain. For them we use the information contained in the "Survey on equipment and use of information technologies at home (TIC-H)" carried out by the INE throughout Spain. The adoption of a defined type of access, the type of connection and the use given will be endogenous variables in the model proposed in this section.

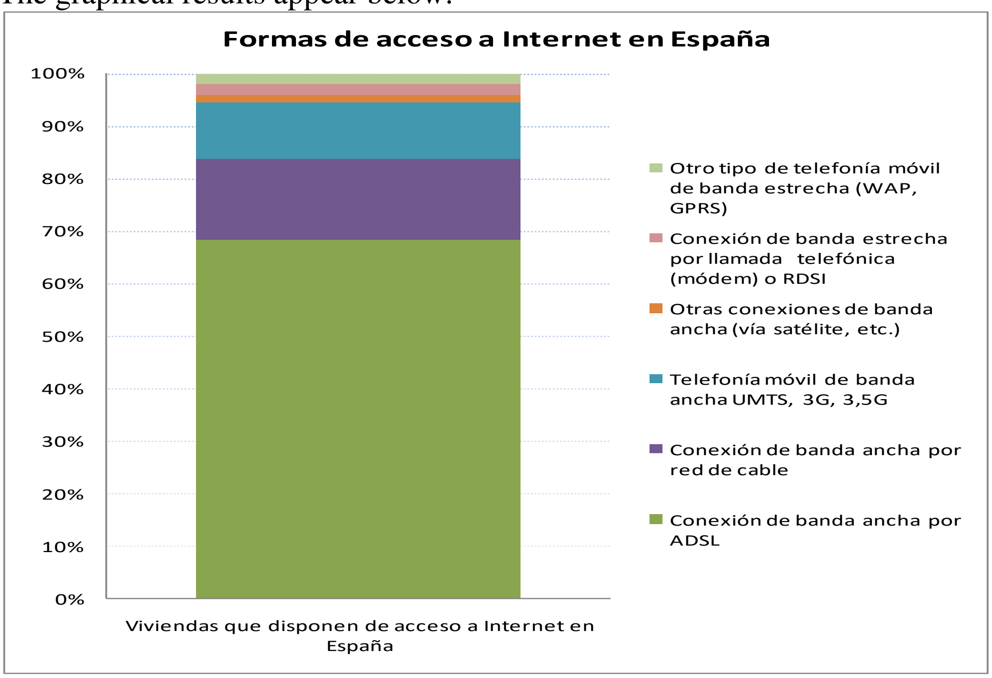
The survey aims to provide information on the various information and communication technology products of Spanish households, as well as the uses made by Spaniards of these products, the Internet and electronic commerce.

The variables studied are among others the equipment in information technologies (television, computer, telephone, video,...), availability of access and form of internet connection per home. Computer, internet and email uses by household size and by household members according to sex, age, level of education and employment status.

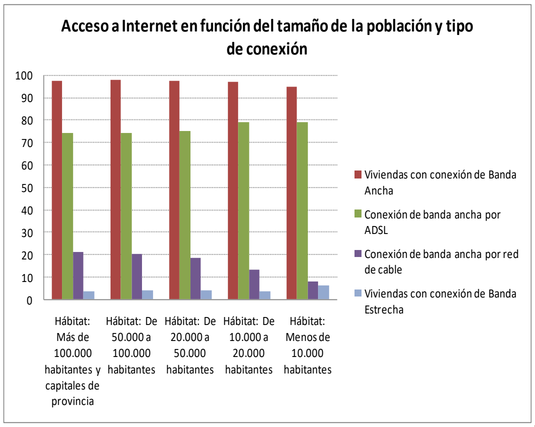
Internet access from home

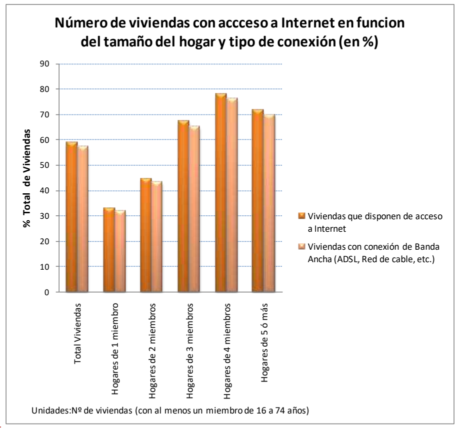
We start from a sample of answers to the question " You have vd. Any kind of internet access from home? "

The graphical results appear below:



SOURCE: NATIONAL INSTITUTE OF STATISTICS

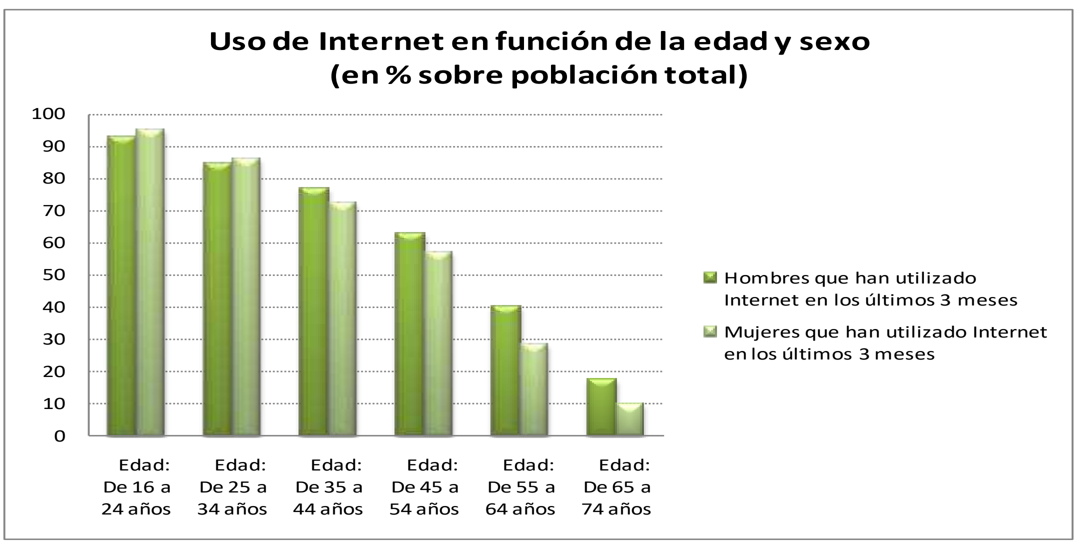




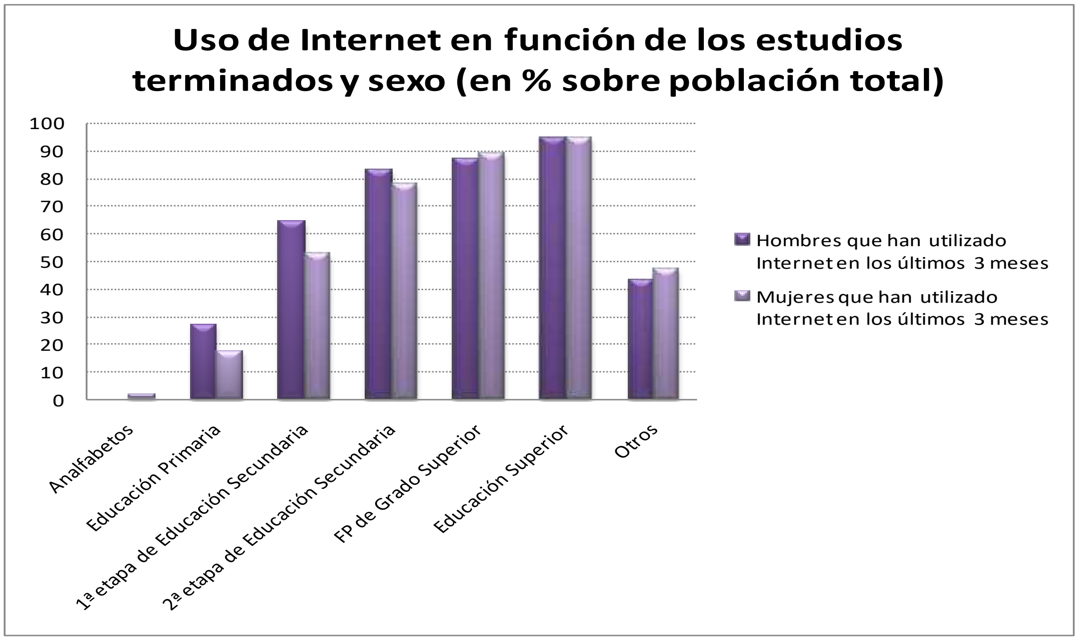
SOURCE: NATIONAL INSTITUTE OF STATISTICS

The figures above refer to the type of internet access based on the size of the population and the size of the household. Thus, of all households with internet access (59.1%), most already have broadband access (97.1%) while only 4% have narrowband access. The last table shows considerable differences between the percentages of access through one channel or another depending on the size of the household.

Internet Use

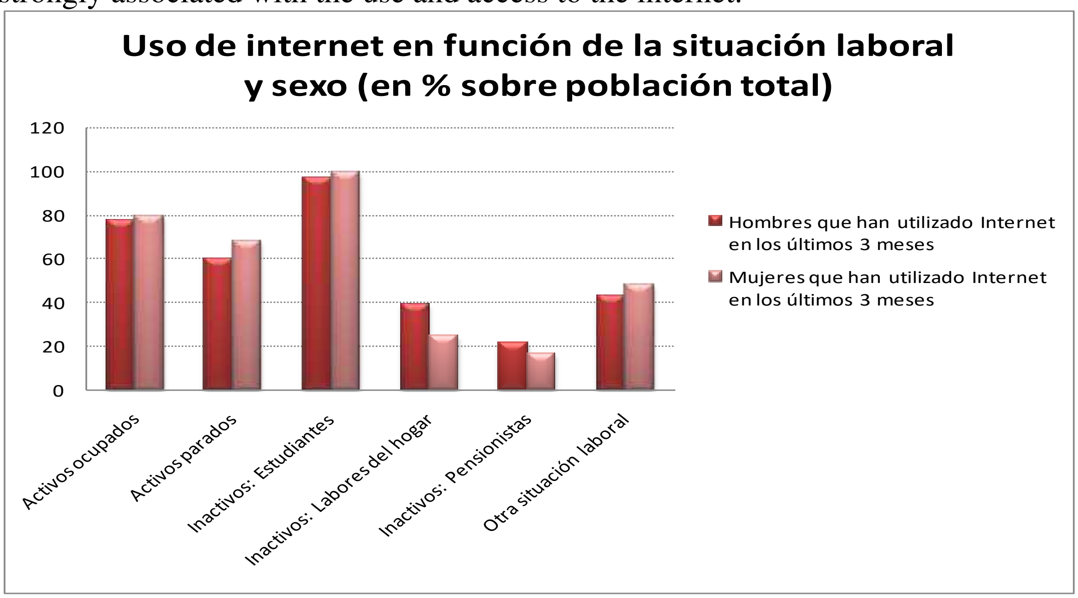


SOURCE: NATIONAL INSTITUTE OF STATISTICS



SOURCE: NATIONAL INSTITUTE OF STATISTICS

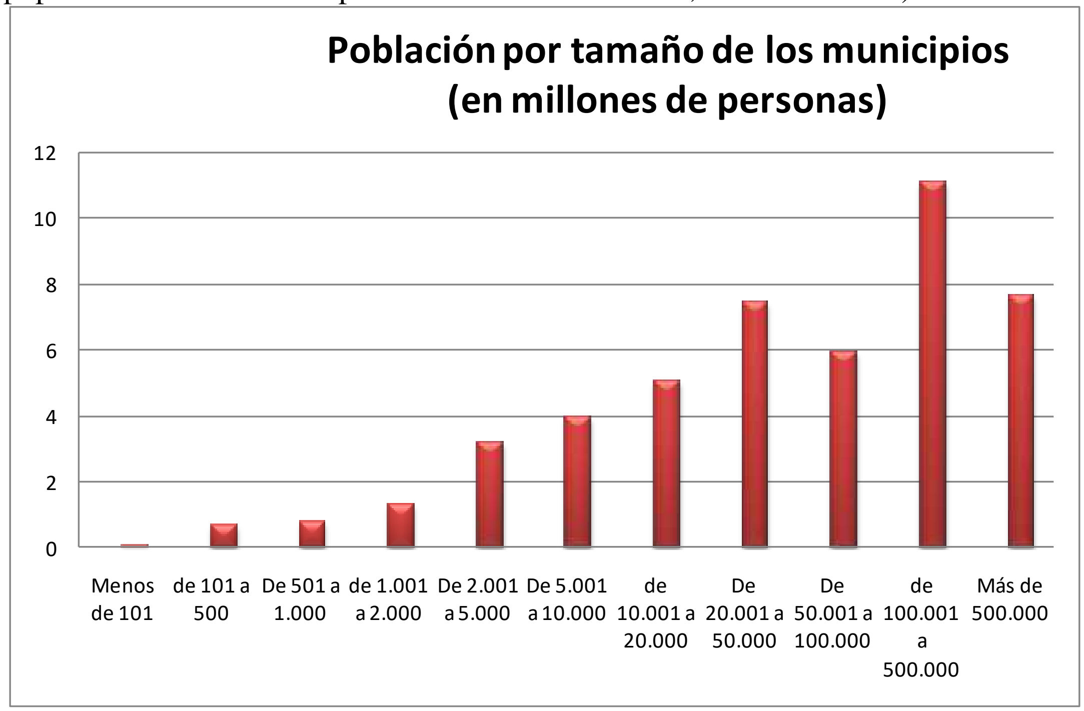
As for the level of qualification, we see in the figure above that the use of the internet is higher in people with higher education completed (29%);they are closely followed by those people who have completed the second stage of secondary education(27,5%)and then by 21,6 % of people who have completed the first stage of secondary education and 13.9% of people with higher vocational training. As expected, people without any training represent only 0.023%of respondents and are barely noticeable in the graph. It is evident that the level of education is strongly associated with the use and access to the internet.



SOURCE: NATIONAL INSTITUTE OF STATISTICS

Taking into account the professional situation of the respondents, the figure above shows, as expected, that the people who use the internet the most are the employed workers who connect to the internet from home and workplace although the latter in a smaller proportion (64. 7% in total); they are followed by the unemployed who connect mainly from housing (13.5%) and students (12.7%) who connect mainly from housing and study center. The remaining 9.2%is divided between people who are not part of the working population (housewives, pensioners and unemployed).

In general, more than half of the people who use the internet are people whose habitat is in large cities (with more than 100,000 inhabitants) and provincial capitals (if we look at the distribution of the population by size of municipalities in the graph below, about 40% of the Spanish population lives in municipalities with more than 100,000 inhabitants.) .



SOURCE: NATIONAL INSTITUTE OF STATISTICS

In the case of household size, those who use the internet the most come from households of three members, followed by those belonging to households with four or more members.

After this basic description of the data, we will analyze and measure the dependencies between the variables. We start by using a demand model designed to explain the relationships between the variables of interest.

Modelling Internet demand in Spain

The first thing to distinguish is access to and use of the internet. It is evident that the use of a service is only possible if you have access to the internet, so the use is conditioned to access. But also access is conditioned to use, because depending on the use the individual will decide whether or not to join the network.

We used the Survey on equipment and use of information and communication *technologies in households in* Spain to analyze the characteristics of internet users in Spain and study the relationships between sociodemographic characteristics with habitat, sex, age, studies. Costumes the demand for the internet taking into account the following elements:

Internet access= f(economic,technologies,socio-demographic)

Internet usage= g(economic, technol.,socio-demographic acceso)

Internet access will depend on economic, technological and sociodemographic factors. The use of the internet will depend on all the above factors and is conditioned to have access (second formula).

We approximate broadband internet access (= y_{2i} availability of broadband connection at home) through the following equation:

$$y_{2i} = \beta_0 + \beta_1 renta + \beta_2 pc_i + \beta_3 móvil_i +$$

$$\beta_4 frecuencia_i + \beta_5 nivelest_i + \beta_6 totmen_i + \beta_7 habitat_i + \beta_8 sexo_i + \beta_9 edad_i$$

$$+ \beta_{10} telev + \beta_{11} edad^2{}_i + \varepsilon_i$$

I=1,...,19,644

Where the exogenous variables are summarized in the following table:

Exogenous variables

Variable name	Definition
income	*family income*
PC	*1=has ord. at home, 0 = no*
mobile	*1=have mobile. at home, 0 = no*
frequency	*internet use last three months*
levelest	*respondent's level of education*
totmen	*number of children between 10 and 15 years old*
Habitat	*population* size
sex	*1 = male, 0=female*
age	*age of respondent*
telev	*Television at home*
age^2	*age square*

The endogenous y_{2i} variable is a binary variable that refers to whether the individual has broadband at home (1 if he has, 0 does not have).

We describe below the mean, standard deviation, and the maximum and minimum values of the endogenous variable (broadband internet access) depending on the values adopted by the exogenous variables (age, computer availability, levels of disposable income, level of education and sex of the respondent):

Disponibilidad de banda ancha y frecuencia de uso de internet (últimos 3 meses)

Frecuencia	Media	Desv. T.	Min	Max
1	.737914	.4398144	0	1
2	.7196602	.4493016	0	1
3	.6945813	.4611528	0	1
4	.6290323	.4850235	0	1
Total	.7292229	.4443928	0	1

Disponibilidad de banda ancha y disponibilidad de ordenador

pc	Media	Desv. T.	Min	Max
0	.2429907	.4309078	0	1
1	.7215979	.4482374	0	1
Total	**.7159666**	**.4509776**	**0**	**1**

Disponibilidad de banda ancha y renta disponible de los hogares				
Renta	Media	Desv. T.	Min	Max
1	.6335404	.4820867	0	1
2	.6968366	.4597263	0	1
3	.7022638	.4573755	0	1
4	.7719403	.4197066	0	1
5	.7426573	.4372717	0	1
Total	**.7159666**	**.4509776**	**0**	**1**

Disponibilidad de banda ancha y nivel de estudio				
n. de estudio	Media	Desv. T.	Min	Max
1	.7391304	.4439611	0	1
2	.6818505	.4659237	0	1
3	.7057393	.4558206	0	1
4	.7307153	.4436919	0	1
5	.700291	.4583528	0	1
6	.7386976	.4394355	0	1
Total	**.7159666**	**.4509776**	**0**	**1**

Disponibilidad de banda ancha y disponibilidad de televisión en el hogar				
telev	Media	Desv. T.	Min	Max
0	.75	.4423259	0	1
1	.7158765	.4510208	0	1
Total	**.7159666**	**.4509776**	**0**	**1**

[33]

Disponibilidad de banda ancha y disponibilidad de teléfono fijo en los hogares				
Telef1	Media	Desv. T.	Min	Max
0	.3317073	.4712097	0	1
1	.7438377	.4365382	0	1
Total	**.7159666**	**.4509776**	**0**	**1**

33

Stratum 0 : Provincial capitals with more than 500,000 inhabitants.

Stratum 1 : Rest of provincial capitals.

Stratum 2 : Municipalities (not provincial capitals) with more than 100,000 inhabitants.

Stratum 3 : Municipalities (not provincial capitals) with more than 50,000 and less than 100,000 inhabitants.

Stratum 4 : Municipalities with more than 20,000 and less than 50,000 inhabitants.

Stratum 5 : Municipalities with more than 10,000 and less than 20,000 inhabitants.

Stratum 6 : Municipalities with less than 10,000 inhabitants.

Disponibilidad de banda ancha y edad

edad	Media	Desv. T.	Min	Max
16	.7346939	.4430055	0	1
17	.75	.4345241	0	1
18	.781457	.414633	0	1
19	.7092199	.4557413	0	1
20	.7459016	.4371484	0	1
21	.7894737	.4094824	0	1
22	.7627119	.4272345	0	1
23	.7361111	.4422781	0	1
24	.7380952	.4414263	0	1
25	.7435897	.438529	0	1
26	.6666667	.4732424	0	1
27	.6992481	.4603188	0	1
28	.7372263	.4417557	0	1
29	.8	.4013865	0	1
30	.7114094	.4546353	0	1
31	.6772152	.4690278	0	1
32	.7040816	.4576234	0	1
33	.6939891	.4620986	0	1
34	.7014925	.4587459	0	1
35	.6866359	.4649335	0	1
36	.6302083	.4840103	0	1
37	.6940639	.4618584	0	1
38	.6593886	.474953	0	1
39	.693617	.4619747	0	1
40	.7165354	.4515697	0	1
41	.6875	.4644811	0	1
42	.6550388	.4762796	0	1
43	.7261905	.4467998	0	1
44	.7639485	.4255684	0	1
45	.7111111	.4542568	0	1
46	.7293578	.4453141	0	1
47	.6995708	.4594316	0	1
48	.7330317	.4433799	0	1
49	.7511737	.4333514	0	1
50	.7101449	.4547948	0	1
51	.7386935	.4404547	0	1
52	.7888199	.4094194	0	1
53	.654321	.4770638	0	1
54	.6923077	.4631607	0	1
55	.7432432	.4383274	0	1
56	.7293233	.4459892	0	1
57	.7217391	.4501038	0	1
58	.7285714	.4462934	0	1
59	.7461538	.4368942	0	1
60	.7163121	.4523943	0	1
61	.6326531	.4845607	0	1
62	.6699029	.4725473	0	1
63	.81	.3942772	0	1
64	.6966292	.4623186	0	1
65	.7105263	.4565315	0	1
66	.7619048	.4293388	0	1
67	.7333333	.4459485	0	1
68	.7647059	.4273363	0	1
69	.65	.4830459	0	1
70	.75	.4380188	0	1
71	.7948718	.4090739	0	1
72	.7948718	.4090739	0	1
73	.7391304	.4439611	0	1
74	.6111111	.4944132	0	1
75	.6285714	.4902409	0	1
76	.7666667	.4301831	0	1
77	.5925926	.5007117	0	1
78	.68	.4760952	0	1
79	.6875	.4787136	0	1
80	.7407407	.4465761	0	1
81	.7222222	.4608886	0	1
82	.7222222	.4608886	0	1
83	.8421053	.3746343	0	1
84	.5555556	.51131	0	1
85	.7692308	.438529	0	1
86	.5833333	.5149287	0	1
87	.7333333	.4577377	0	1
88	.6	.5163978	0	1
89	.8	.4472136	0	1
90	.8571429	.3779645	0	1
91	.7142857	.48795	0	1
92	.6666667	.5163978	0	1
93	1	0	1	1
94	.5	.7071068	0	1
95	.6666667	.5773503	0	1
96	1	.	1	1
97	.5	.7071068	0	1
98	.	.	.	.
99	.	.	.	.
100	.	.	.	.
102	1	0	1	1
Total	**.7159666**	**.4509776**	**0**	**1**

SOURCE: INE SURVEY ON EQUIPAMIENTO AND USE OF TECHNOLOGIES INHOUSEHOLDS

Model estimation results

The bottom tab shows the results of the estimation using a probit model correcting the selection bias following Heckman's methodology:

I. Equation to estimate:

$$y_{2i} = \beta_0 + \beta_1 renta + \beta_2 sexo_i + \beta_3 nivelest_i + \beta_4 edad_i + \beta_5 telef1 + \beta_6 edad^2{}_i + \varepsilon_i$$

II. Selection equation:

The selection equation studies the variables that determine the decision to have or not to have a computer at home which in turn affects the decision to have broadband at home:

$$pc_i = \beta_0 + \alpha_1 renta + \alpha_2 sexo_i + \alpha_3 nivelest_i + \alpha_4 edad_i + \alpha_5 telef1 + \alpha_6 edad^2{}_i + \alpha_7 pc_i + \mu_i$$

Therefore, the decision to have broadband depends on several variables (income, level of education, sex, age, and availability of telephone at home) but also considers the decision to acquire or not a computer. Both decisions are correlated (we can see this correlation through the correlation between the random term of the probit estimate without taking into account the selection bias and the one that includes the selection bias (it is defined by the variable rho: If rho is not significant, we cannot reject the null hypothesis of absence of selection bias. In this case Rho is significant, so it can be said that there is selection bias.

```
Probit model with sample selection              Number of obs      =     19644
                                                Censored obs       =     10550
                                                Uncensored obs     =      9094

                                                Wald chi2(6)       =    106.02
Log likelihood = -10668.74                      Prob > chi2        =    0.0000
```

forint1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
forint1						
sexo	-.0325947	.0279057	-1.17	0.243	-.0872888	.0220995
edad	-.0077103	.0042173	-1.83	0.068	-.0159761	.0005555
nivelest	-.0050798	.010487	-0.48	0.628	-.0256339	.0154743
renta	.0302899	.0108998	2.78	0.005	.0089266	.0516532
telef1	.6747373	.0742124	9.09	0.000	.5292836	.820191
age2	.0000799	.0000446	1.79	0.073	-7.57e-06	.0001673
_cons	.1865553	.1335731	1.40	0.163	-.0752431	.4483537
select						
sexo	.0575152	.0268234	2.14	0.032	.0049424	.1100881
edad	-.0003816	.0039495	-0.10	0.923	-.0081225	.0073593
nivelest	.1015332	.0099415	10.21	0.000	.0820483	.1210181
renta	.0770227	.0096427	7.99	0.000	.0581233	.0959222
pc	2.744552	.0427907	64.14	0.000	2.660684	2.82842
telef1	1.321077	.0360034	36.69	0.000	1.250512	1.391642
age2	-.0000998	.0000396	-2.52	0.012	-.0001775	-.0000221
_cons	-3.42171	.1088865	-31.42	0.000	-3.635123	-3.208296
/athrho	-.4906683	.0679734	-7.22	0.000	-.6238937	-.357443
rho	-.4547467	.0539168			-.5538332	-.3429598

```
LR test of indep. eqns. (rho = 0):   chi2(1) =    56.26   Prob > chi2 = 0.0000
```

The following table shows not only the results of the regression with correction of selection bias but also a probit estimates without correcting the selection bias:

RESULTADOS DE LAS REGRESIONES

	Regresión con corrección del sesgo de selección (Heckman)	Regresión Probit sin corrección del sesgo de selección
EcuaciónI		
sexo	-0.0326	-0.0229
	(-1.17)	(-0.80)
edad	-0.00771	-0.00695
	(-1.83)	(-1.61)
nivelest	-0.00508	0.0140
	(-0.48)	(1.33)
renta	0.0303**	0.0477***
	(2.78)	(4.36)
telef1	0.675***	1.077***
	(9.09)	(19.49)
age2	0.0000799	0.0000504
	(1.79)	(1.10)
_cons	0.187	-0.425***
	(1.40)	(-3.70)
Ecuación II		
sexo	0.0575*	
	(2.14)	
edad	-0.000382	
	(-0.10)	
nivelest	0.102***	
	(10.21)	
renta	0.0770***	
	(7.99)	
pc	2.745***	
	(64.14)	
telef1	1.321***	
	(36.69)	
age2	-0.0000998*	
	(-2.52)	
_cons	-3.422***	
	(-31.42)	
athrho		
_cons	-0.491***	
	(-7.22)	
N	**19644**	**9094**

t statistics in parentheses
* p<0.05, ** <0.01, *** p<0.001

The following conclusions can be inferred from the results of the estimate:

- The higher the rent, the greater the demand for broadband access at home. This fact is explained by a positive sign of the equity coefficient.
- The variables sex, age or level of education are not significant. On the contrary, the variable that reflects the availability of telephone in the home (telef1) is significant.
- Uncorrected probit regression overestimates coefficients versus regression with correction selection bias.

Using the internet to make online purchases

In this section we try to contribute to a better understanding of the behavior of consumers when they purchase products online considering the fact that not all products or services are equally suitable for online sales. For various products, the Internet shows varying degrees of suitability as a means of purchase.

We describe below the mean, standard deviation and the maximum and minimum values of the endogenous variable (internet purchases) depending on the values adopted by the exogenous variables (availability of landline, computer, levels of disposable income, level of education and sex of the respondent):

Compras por internet en función de la disponibilidad de teléfono fijo en la vivienda (1=dispone de telefono fijo, 0=no dispone)

telef1	media	desv. est.	min	max
0	0,29	0,45	0	1
1	0,43	0,50	0	1
Total	**0,41**	**0,49**	**0**	**1**

Compras por internet en función de la disponibilidad de ordenador en la vivienda (1=dispone de ordenador, 0=no dispone)

pc	media	desv. est.	min	max
0	0,18	0,38	0	1
1	0,44	0,50	0	1
Total	**0,41**	**0,49**	**0**	**1**

Source: INE Survey on equipamiento and use of technologies inhouseholds

Compras por internet en función de categorías de renta disponible de la familia

renta	media	desv. est.	min	max
1	0,24	0,43	0	1
2	0,34	0,47	0	1
3	0,47	0,50	0	1
4	0,61	0,49	0	1
5	0,36	0,48	0	1
Total	**0,41**	**0,49**	**0**	**1**

Disponibilidad de banda ancha y nivel de estudio

N. de estudio	Media	Desv. T.	Min	Max
1	.7391304	.4439611	0	1
2	.6818505	.4659237	0	1
3	.7057393	.4558206	0	1
4	.7307153	.4436919	0	1
5	.700291	.4583528	0	1
6	.7386976	.4394355	0	1
Total	**.7159666**	**.4509776**	**0**	**1**

Source: INE Survey on equipamiento and use of technologies inhouseholds

Compras por internet en función del sexo de la persona informante (0=mujer, 1=hombre)

Sexo	media	desv. est.	min	max
0	0,36	0,48	0	1
1	0,45	0,50	0	1
Total	**0,41**	**0,49**	**0**	**1**

Source: INE Survey on equipamiento and use of technologies inhouseholds

Disponibilidad de banda ancha y nivel de estudio				
edad	Media	Desv. T.	Min	Max
16	.7346939	.4430055	0	1
17	.75	.4345241	0	1
18	.781457	.414633	0	1
19	.7092199	.4557413	0	1
20	.7459016	.4371484	0	1
21	.7894737	.4094824	0	1
22	.7627119	.4272345	0	1
23	.7361111	.4422781	0	1
24	.7380952	.4414263	0	1
25	.7435897	.438529	0	1
26	.6666667	.4732424	0	1
27	.6992481	.4603188	0	1
28	.7372263	.4417557	0	1
29	.8	.4013865	0	1
30	.7114094	.4546353	0	1
31	.6772152	.4690278	0	1
32	.7040816	.4576234	0	1
33	.6939891	.4620986	0	1
34	.7014925	.4587459	0	1
35	.6866359	.4649335	0	1
36	.6302083	.4840103	0	1
37	.6940639	.4618584	0	1
38	.6593886	.474953	0	1
39	.693617	.4619747	0	1
40	.7165354	.4515697	0	1
41	.6875	.4644811	0	1
42	.6550388	.4762796	0	1
43	.7261905	.4467998	0	1
44	.7639485	.4255684	0	1
45	.7111111	.4542568	0	1
46	.7293578	.4453141	0	1
47	.6995708	.4594316	0	1
48	.7330317	.4433799	0	1
49	.7511737	.4333514	0	1
50	.7101449	.4547948	0	1
51	.7386935	.4404547	0	1
52	.7888199	.4094194	0	1
53	.654321	.4770638	0	1
54	.6923077	.4631607	0	1
55	.7432432	.4383274	0	1
56	.7293233	.4459892	0	1
57	.7217391	.4501038	0	1
58	.7285714	.4462934	0	1
59	.7461538	.4368942	0	1
60	.7163121	.4523943	0	1
61	.6326531	.4845607	0	1
62	.6699029	.4725473	0	1
63	.81	.3942772	0	1
64	.6966292	.4623186	0	1
65	.7105263	.4565315	0	1
66	.7619048	.4293388	0	1
67	.7333333	.4459485	0	1
68	.7647059	.4273363	0	1
69	.65	.4830459	0	1
70	.75	.4380188	0	1
71	.7948718	.4090739	0	1
72	.7948718	.4090739	0	1
73	.7391304	.4439611	0	1
74	.6111111	.4944132	0	1
75	.6285714	.4902409	0	1
76	.7666667	.4301831	0	1
77	.5925926	.5007117	0	1
78	.68	.4760952	0	1
79	.6875	.4787136	0	1
80	.7407407	.4465761	0	1
81	.7222222	.4608886	0	1
82	.7222222	.4608886	0	1
83	.8421053	.3746343	0	1
84	.5555556	.51131	0	1
85	.7692308	.438529	0	1
86	.5833333	.5149287	0	1
87	.7333333	.4577377	0	1
88	.6	.5163978	0	1
89	.8	.4472136	0	1
90	.8571429	.3779645	0	1
91	.7142857	.48795	0	1
92	.6666667	.5163978	0	1
93	1	0	1	1
94	.5	.7071068	0	1
95	.6666667	.5773503	0	1
96	1	.	1	1
97	.5	.7071068	0	1
98	.	.	.	.
99	.	.	.	.
100	.	.	.	.
102	1	0	1	1
Total	**.7159666**	**.4509776**	**0**	**1**

Taking advantage again of the INE survey on Equipment and Use of Information and Communication Technologies in Homes we can study the relationship between online purchases and sociodemographic variables (sex, age, level of education) and technological:

I. Equation to estimate:

$$Compras \text{ por internet} = \beta_1 + \beta_2 sexo + \beta_3 edad + \beta_4 nivelest + \beta_5 renta$$
$$+\beta_6 pc + \beta_7 telef1 + \beta_8 edad^2{}_i + \varepsilon_i$$

As in the previous section we relate online purchases with the availability of broadband internet access which in turn is conditioned by the availability of ordered:

II. Selection equation:

The selection equation studies the variables that affect the decision to have a computer at home or not:

$$pc_i = \beta_0 + \alpha_1 renta + \alpha_2 sexo_i + \alpha_3 nivelest_i + \alpha_4 edad_i + \alpha_5 telef1 + \alpha_6 edad^2{}_i + \alpha_7 pc_i + \mu_i$$

Therefore, the decision to buy online depends on several variables (income, level of education, sex, age and availability of telephone at home) but also on having or not having access to the internet that depends in turn on the decision to have or not to have a computer. The decision to buy online and to have a computer are correlated (The rho comes to represent the correlation between errors between the probit estimate and the one that includes the selection bias: In this case Rho is also significant, so it can be said that there is selection bias.

The following table shows the results of the regression with correction of selection bias and in the following in two columns the results of the first compared with a probit regression without correction of selection bias:

```
Probit model with sample selection              Number of obs      =     19644
                                                Censored obs       =      9985
                                                Uncensored obs     =      9659

                                                Wald chi2(6)       =    686.59
Log likelihood = -12202.63                      Prob > chi2        =    0.0000
```

compras	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
compras						
sexo	.3526348	.0264431	13.34	0.000	.3008072	.4044624
edad	.0269745	.0052043	5.18	0.000	.0167742	.0371747
nivelest	.3566749	.0196075	18.19	0.000	.3182448	.3951049
renta	.058257	.010308	5.65	0.000	.0380537	.0784602
telef1	.4259306	.0394703	10.79	0.000	.3485702	.5032911
age2	-.0005832	.0000624	-9.35	0.000	-.0007055	-.000461
_cons	-2.72244	.1296413	-21.00	0.000	-2.976532	-2.468347
select						
sexo	.1798374	.0252342	7.13	0.000	.1303793	.2292954
edad	-.0492238	.0046721	-10.54	0.000	-.058381	-.0400667
nivelest	.5350594	.0098767	54.17	0.000	.5157014	.5544174
renta	.043415	.0088665	4.90	0.000	.026037	.060793
telef1	.5875906	.0324518	18.11	0.000	.5239863	.6511949
age2	-.0000777	.0000463	-1.68	0.093	-.0001684	.000013
_cons	.1777841	.1163613	1.53	0.127	-.0502799	.4058481
/athrho	.5249208	.1647578	3.19	0.001	.2020015	.8478401
rho	.481489	.1265617			.1992981	.6899394

LR test of indep. eqns. (rho = 0): chi2(1) = 16.85 Prob > chi2 = 0.0000

RESULTADOS DE LAS REGRESIONES		
	Regresión con corrección del sesgo de selección (Heckman)	Regresión Probit sin corrección del sesgo de selección
Ecuación I		
sexo	0.353***	0.332***
	(13.34)	(12.27)
edad	0.0270***	0.0296***
	(5.18)	(5.54)
nivelest	0.357***	0.280***
	(18.19)	(25.49)
renta	0.0583***	0.0516***
	(5.65)	(4.90)
telef1	0.426***	0.351***
	(10.79)	(9.33)
age2	-0.000583***	-0.000509***
	(-9.35)	(-7.96)
_cons	-2.722***	-2.382***
	(-21.00)	(-20.37)
ecuación II		
sexo	0.180***	
	(7.13)	
edad	-0.0492***	
	(-10.54)	
nivelest	0.535***	
	(54.17)	
renta	0.0434***	
	(4.90)	
telef1	0.588***	
	(18.11)	
age2	-0.0000777	
	(-1.68)	
_cons	0.178	
	(1.53)	
athrho		
_cons	0.525**	
	(3.19)	
N	**19644**	**9659**
t statistics in parentheses		
*** p<0.05, ** p<0.01, *** p<0.001**		

We can summarize the main results:

- In the case of the relationship between purchases by internet and the sex of the person surveyed, we observed, unlike the usual purchases in store, the greater probability that it is the man who makes purchases online.

```
Marginal effects after heckprob
      y  = Pr(compras=1) (predict)
         =  .27071208
```

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
sexo*	.1004751	.00892	11.26	0.000	.082989	.117962	.445734
edad	.0104528	.0017	6.14	0.000	.007114	.013791	51.5336
nivelest	.0724388	.00373	19.44	0.000	.065137	.079741	3.38816
renta	.0157549	.00343	4.59	0.000	.009026	.022484	2.6769
telef1*	.0890148	.01084	8.21	0.000	.067767	.110263	.808797
age2	-.0001509	.00002	-7.80	0.000	-.000189	-.000113	3020.86
pc*	0	0	.	.	0	0	.58033

(*) dy/dx is for discrete change of dummy variable from 0 to 1

- Disaggregated by categories of products purchased online, the probability of acquiring, for example, food or travel is higher for women, while computers and financial services are more likely to be purchased by men (see regression table for each product purchased). A possible explanation is the greater experience of one or the other in the purchase of a certain type of article in the case of men would be computer and electronic products and in the case of women in food, event tickets and travel.

RESULTADOS DE LAS REGRESIONES

	Alimentación	Eq. informático	Viajes	Ser. Financieros
sexo	-0.171**	0.614***	0.0221	0.357***
	(-2.69)	(10.22)	(0.49)	(5.02)
edad	0.0319*	0.000641	0.0438***	0.0685***
	(2.11)	(0.05)	(4.22)	(3.70)
nivelest	0.0462	0.0580*	0.214***	0.114***
	(1.68)	(2.37)	(11.04)	(3.74)
renta	0.00765	-0.0242	0.0423*	0.0392
	(0.27)	(-0.99)	(2.19)	(1.27)
telef1	0.0739	0.0736	-0.0899	-0.0110
	(0.68)	(0.78)	(-1.22)	(-0.10)
age2	-0.000248	-0.0000913	-0.000466***	-0.000764***
	(-1.41)	(-0.58)	(-3.73)	(-3.44)
_cons	-2.556***	-1.942***	-2.284***	-4.139***
	(-7.13)	(-6.32)	(-9.43)	(-8.99)
N	**3297**	**3297**	**3297**	**3297**

*** p<0.05, ** p<0.01, *** p<0.001**

t statistics in parentheses

- The higher the income, the greater the purchases online. This fact is explained by a positive sign of the equity coefficient.

- Age: We observe that the probability of buying online increases with age which can be explained because as age increases, the knowledge of the products to be purchased increases, which allows them to buy without physically seeing or touching the product, or because time saving becomes a vital variable as age increases.
- Level of studies: We observe that education is an important determinant of online purchases so there is a positive and significant relationship between the level of education and the probability of buying online. By product category we observe the most significant coefficients in financial services, travel and computer services.

Conclusions

In this paper we have analyzed the telecommunications sector from a descriptive perspective and then focus on analyzing the demand for the internet in Spain. To do this we have first studied the characteristics that can affect access and then use. An extension of this analysis would be to carry out an analysis of panel data, that is, the behavior of a group of families over a period of time, so that we could evaluate from a dynamic point of view the demand for the internet in Spain.

In the last section of the work, we have analyzed factors that make certain internet users acquire products online in addition to considering that the importance of certain variables will be greater or lesser depending on the type of product to be purchased. We can say that the future evolution of online retail sales in the case of Spain will be conditioned by the penetration of interest, availability and familiarity of consumers in computers and the internet as well as internet security. Therefore, policies should be aimed at improving these three variables and should also be considered in the marketing strategies of companies.

Bibliography

A.L. (Ed.), Down to the wire: Studies in the diffusion and regulation of telecommunications technologies, Nova Science Publishers, Haupaugge, NY, p. 119-138.

Armstrong, M. (1998), "Network interconnection in telecommunications." Economic Journal 108, 545-564.

Armstrong, M. (2002). The Theory of Access Pricing and Interconnection. In: Cave,M.E., Majumdar, S.K., Vogelsang,I. (Eds.), Handbook of Telecommunications Economics. North-Holland, Amsterdam.

Aron, D.J. and D.E. Burnstein (2003), "Broadband adoption in the United States: An empirical analysis", in Shampine,

Baranès, E. and M. Bourreau (2005), "An economist's guide to local loop unbundling",Communications &Strategies, Vol. 57, p. 13-31.

Calzada, J. y Valletti, T. (2007), "Network Competition and Entry Deterrence.", forthcoming in Economic Journal.

Cave, M. (2003), "The economics of wholesale broadband access", MMR Beilage 10/2003, p. 15-19.

Cave, M., and Vogelsang, I. (2003), "How access pricing and entry interact",Telecommunications Policy 27,717-727.

Cave, M. (2006), "Encouraging infrastructure competition via the ladder ofinvestment", Telecommunications Policy 30, 223-237

CRA International, 2008, "Average pay TV revenues per subscriber across Europe", Report prepared for BSkyB in the framework of the 2nd research on pay television carried out by OfCom. Available in http://www.ofcom.org.uk/tv/paytv/

Crandall, R., A. Ingraham and H. Singer (2004), "Do unbundling policies discourage CLEC facilities-basedinvestment?", Topics in Economic Analysis & Policy, Vol. 4.

Davidson R, MacKinnon J (1993) Econometric Theory and Methods. Oxford University Press

DeGraba, P. (2002), "Bill and keep as the efficient interconnection regime?: A reply." Review of Network Economics 1, 61-65.

Denni, M. and H. Gruber (2006), "The diffusion of broadband telecommunications: the role of competition",Departmental Working Papers of Economics – University 'Roma Tre' nº 60.

Distaso, W., P. Lupi and F. Manenti (2006) "Platform competition and broadband uptake: Theory and empirical evidence from the European Union", Information Economics and Policy, Vol. 18, p. 87-106.
Gual,J., 2003, "Market Definition in the Telecoms Industry", CEPR Discussion Papers 3988, C.E.P.R. Discussion Papers.
Heckman, James J, 1979. "Sample Selection Bias as a Specification Error," Econometrica, Econometric Society, vol. 47(1), pages 153-61, January
Hoernig, S. (2007), "On-net and Off-net Pricing on Asymmetric Telecommunications networks." Information Economics and Policy, 19:171-188.
Howel, B. (2002), "Broadband uptake and infrastructure regulation: Evidence from the OECD countries", ISCR Working Paper BH02/01.
IDATE, 2009, "FTTH European Panorama. December 2008", presentation prepared for the FTTH Council Europe Conference, held in Copenhagen on February 11, 2009. Available in http://www.ftthcouncil.eu/documents/studies/Market_Data-December_2008.pdf
Jullien, B., Rey, P. (2008), "Notes on the economics of termination charges." IDEI Report #6 Telecommunications.
Laffont, J.-J., Rey, P., Tirole, J. (1998a), "Network competition: I. Overview and nondiscriminatory pricing." RAND Journal of Economics 29, 1-37.
Liu, H., P. K. Chintagunta and T. Zhu (2008), "Complementarities and the Demand for Home Broadband Internet Services," Chicago Booth School of Business Research Paper No. 08-30.
López, A.L., y Rey, P. (2008), "Foreclosing Competition through Access Charges and Price Discrimination." Mimeo, available in: www.angelluislopez.net, www.idei.fr.
López, A. L., and Vives, X. (2008), "Inversión en broadband: Competencia en infraestructuras y competencia en servicios." Economists "Spain 2007. A balance." Vol. 116, 258-265.
López, A. L. (2009), "Deployment of telecommunications networks and broadband diffusion." Economists "Spain 2008. A balance." Vol. 119, 283-288.
Maldoom, D., Marsden, R., Sidak, J.G., and Singer, H.J. (2003), "Competition in Broadband Provision and its
Implications for Regulatory Policy". Report prepared for the Brussels Round Table, London. Available in SSRN: http://ssrn.com/abstract=463041.
Montero, J.J., "Telecommunications Law." Tirant Lo Blanch, 2007.
OECD, (2007), "Convergence and Next Generation Networks", Ministerial Backgroud Report, Directorate for Science, Technology and Industry, Committee for Information, Computer and Communications Policy, DSTI/ICCP/CISP(2007)2/FINAL.
OECD, (2008), "Developments in Fibre technologies and investment", OECD Digital Economy Papers 142,OECD, Directorate for Science, Technology and Industry.
Pérez Amaral T, Alvarez F, Moreno B (1995) Business Telephone Traffic Demand in Spain 1980–1991: an Econometric Approach. Information Economics and Policy 7: 115–134
Rappoport P, Taylor L, Kridel D (2002) The Demand of Broadband: Access, Content, and the Value of Time. In: Crandall RW, Alleman JH (eds) Broadband: Should We Regulate High-Speed Internet Access? AEI-Brookings Joint Centre for Regulatory Studies, Washington, D.C.
Reisinger, M. (2004), "The Effects of Product Bundling in Duopoly", Discussion Papers in Economics 477, University of Munich, Department of Economics.
Sangwon L. (2006), "Broadband deployment in the United States: Examining the Impacts of Platform Competition", The International Journal on Media Management, Vol. 8, p. 173-181.
Taylor LD (2000) Towards a framework for analyzing internet demand. Manuscript, University of Arizona U.S. Department of Commerce, and National Telecommunications & Information Administration(2002) A Nation Online: How Americans Are Expanding Their Use of the Internet
Ulset, S. (2002), "Mobile virtual network operators: a strategic transaction cost analysis of preliminary experiences", Telecommunications Policy, 26, pp. 537-549.

Valletti, T.M., C. Cambini (2005), "Investments and network competition. " RAND Journal of Economics 36,446-468.

Varian H (2002) The Demand for Bandwidth: Evidence from the INDEX Project. Mimeo, University of California, Berkeley

Vogelsang, I. (2003), "Price regulation of access to telecommunications Networks." Journal of Economic Literature 41, 830-862.

Wright, J. (2002), "Bill and keep as the efficient interconnection regime?." Review of Network Economics 1, 54-60.

Varian H (2002) The Demand for Bandwidth: Evidence from the INDEX Project. Mimeo, University of California, Berkeley

Review of Math

Descriptive Statistics

Given n numbers x_i: $i = 1, \ldots, n$, their average or mean is its summation divided by n:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The summation operator is a shorthand for sum of a sequence of numbers:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

When the x_i are a sample of data on a particular variable, we often call this the sample average (or sample mean) to emphasize that it is computed from a particular set of data. The sample average is an example of descriptive statistics; in this case, the statistic describes the central tendency of the set of x_i.

Some basic properties about averages are:

First, the sum of these deviations (d) between the observations x_i and the average: $d_i = x_i - \bar{x}$ is always zero:

hence:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

In regression analysis, we need to know some algebraic facts involving deviations from sample averages. An important one is that the sum of squared deviations is the sum of the squared x_i minus n times the square of x:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=0}^{n} (x_i{}^2 - \overline{nx})$$

Given a data set on two variables, $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$, it can also be shown that

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i y_i - \mathrm{n}\overline{xy})$$

The average is the measure of central tendency. However, it is sometimes informative to use the median (or sample median) to describe the central value. To obtain the median of the n numbers $\{x_1, \ldots, x_n\}$, we first order the values of the x_i from smallest to largest. Then, if n is odd, the sample median is the middle number of the ordered observations. For example, given the numbers {−4, 8, 2, 0, 21, −10, 18}, the median value is 2 (because the ordered sequence is {−10, −4, 0, 2, 8, 18, 21}). If we change the largest number in this list, 21, to twice its value, 42, the median is still 2. By contrast, the sample average would increase from 5 to 8, a sizable change. Generally, the median is less sensitive than the average to changes in the extreme values (large or small) in a list of numbers. This is why "median incomes" or "median housing values" are often reported, rather than averages, when summarizing income or housing values in a city or county. If n is even, there is no unique way to define the median because there are two numbers at the center. Usually, the median is defined to be the average of the two middle values (again, after ordering the numbers from smallest to largest). Using this rule, the median for the set of numbers {4, 12, 2, 6} would be (4 + 6)/2 = 5.

Linear Functions

Linear functions play an important role in econometrics because they are simple to interpret and manipulate. If x and y are two variables related by

$$y = \beta_0 + \beta_1 x,$$

then we say that y is a linear function of x, and β_0 and β_1 are two parameters (numbers) describing this relationship. The intercept is β_0, and the slope is β_1.
The defining feature of a linear function is that the change in y is always β_1 times the change in x:

$$\Delta y = \beta_1 \Delta x$$

where Δ is variation. In other words, the marginal effect of x on y is constant and equal to β_1. Linear functions can be defined for more than two variables. Suppose that y is related to two variables, x_1 and x_2, in the general form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

β_0 is still the intercept (the value of y when $x_1 = 0$ and $x_2 = 0$), and β_1 and β_2 measure slopes. From the above equation, the change in y, for given changes in x_1 and x_2, is

$$\Delta y = \beta_1 \Delta x_1 + \beta_2 \Delta x_2$$

If x_2 does not change, that is, $\Delta x_2 = 0$, then $\Delta y = \beta_1 \Delta x_1$, so $\beta_1 = \Delta y / \Delta x_1$ measures how y changes with x_1, holding x_2 fixed, β_1 is often called the partial effect of x_1 on y. Because the partial effect involves holding other factors fixed, it is usually named *ceteris paribus*. The parameter β_2 has a similar interpretation: $\beta_2 = \Delta y/\Delta x_2$ if $\Delta x_1 = 0$, so that β_2 is the partial effect of x_2 on y.

Proportions and Percentages

Proportions and percentages play such an important role in applied economics: many quantities reported in the popular press are in the form of percentages. A percentage is easily obtained by multiplying a proportion by 100. For example, if the proportion of adults in a county with a high school degree is .90, then we say that 90% (90 percent) of adults have a high school degree.

When using percentages, we often need to convert them to decimal form. For example, if a state sales tax is 6% and $100 is spent on a taxable item, then the sales tax paid is 100(.06) = $6. If the annual return on a certificate of deposit (CD) is 10% and we invest $3,000 in such a CD at the beginning of the year, then our interest income is 3, 000(.1) = $300.

We are often interested in measuring the changes in various quantities. Let x denote some variable. Let x_0 and x_1 denote two values for x: x_0 is the initial value, and x_1 is the subsequent value. For example, x_0 could be the annual income of an individual in 1994 and x_1 the income of the same individual in 1995. The proportionate change in x in moving from x_0 to x_1, sometimes called the relative change, is simply

$$x1 - x_0 \,/\, x_0 = \Delta x / x_0$$

In other words, to get the proportionate change, we simply divide the change in x by its initial value. This is a way of standardizing the change so that it is free of units. For example, if an individual's income goes from $40,000 per year to $50,000 per year, then the proportionate change is 10.000/40.000 = .25.

It is more common to state changes in terms of percentages. The percentage change in x in going from x_0 to x_1 is simply 100 times the proportionate change:

$$\%\Delta x = 100(\frac{\Delta x}{x0})$$

Where "%Δx" is "the percentage change in x." For example, when income goes from $30,000 to $50,000, income has increased by 60%; to get this, we simply multiply the proportionate change, .6, by 100.

Again, we must be on guard for proportionate changes that are reported as percentage changes. In the previous example, for instance, reporting the percentage change in income as .6 is incorrect and could lead to confusion.

It is important to distinguish percentage point change and a change in percentage which is common in social sciences. To illustrate, let x denote the percentage of adults in a particular city having a college education. Suppose the initial value is $x_0 = 30$ (30% have a college education), and the new value is $x_1 = 32$. We can compute two quantities to describe how the percentage of college-educated people has changed. The first is the change in x, Δx. In this case, $\Delta x = x_1 - x_0$: the percentage of people with a college education has increased by 2 percentage points. On the other hand, we can compute the percentage change in x: %Δx = 100[(32 − 30)/30] = 6.667. In this example, the percentage point change and the percentage change are very different. The percentage point change is just the change in the percentages. The percentage change is the change relative to the initial value.

Some Special Functions and Their Properties

In Section 1.2, we reviewed the basic properties of linear functions. We already indicated one important feature of functions like $y = \beta_0 + \beta_1 x$: a one-unit change in x results in the same change in y, regardless of the initial value of x. As we noted earlier, this is the same as saying the marginal effect of x on y is constant, something that is not realistic for many economic relationships. For example, the notion of diminishing marginal returns is not consistent with a linear relationship.

Quadratic Functions

One simple way to capture diminishing returns is to add a quadratic term to a linear relationship. Consider the equation

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \quad (1)$$

where β_0, β_1, and β_2 are parameters. When $\beta_1 > 0$ and $\beta_2 < 0$, the relationship between y and x has the parabolic shape. The fact that equation (16) implies a diminishing marginal effect of x on y is easily seen from its graph.

When $\beta_1 > 0$ and $\beta_2 < 0$, the maximum of the function occurs at the point

$$x^* = \beta_1/(-2\beta_2)$$

For example, if $y = 6 + 8x - 2x^2 2$ (so $\beta_1 = 8$ and $\beta_2 = -2$), then the largest value of y occurs at $x^* = 8/4 = 2$, and this value is $6 + 8(2) - 2(2)^2 = 14$.

The statement that x has a diminishing marginal effect on y is the same as saying that the slope of the function in above equation decreases as x increases. An application of calculus gives the approximate slope of the quadratic function as

$$slope = \Delta y/\Delta x \approx \beta_1 + 2\beta 2x$$

for "small" changes in x. The right-hand side of the above equation is the derivative of the function in equation (1) with respect to x. Another way to write this is

$$\Delta y \approx (\beta_1 + 2\beta 2 x)\Delta x \text{ for "small" } \Delta x. \quad (2)$$

For many applications, equation (2) can be used to compute the approximate marginal effect of x on y for any initial value of x and small changes. And we can always compute the exact change if necessary.

The Natural Logarithm

The nonlinear function that plays the most important role in econometric analysis is the natural logarithm. In this text, we denote the natural logarithm, which we often refer to simply as the log function, as

$$y = \log(x) \quad (3)$$

For our purposes, only the natural logarithm is important, and so log(x) denotes the natural logarithm. The function $y = \log(x)$ is defined only for $x > 0$. One important difference between the log and the quadratic function is that when $y = \log(x)$, the effect of x on y never becomes negative: the slope of the function gets closer and closer to zero as x gets large, but the slope never quite reaches zero and certainly never becomes negative.

The following are useful facts for natural log:

1. *$\log(x) < 0$ for $0 < x < 1$*

2. *$\log(1) = 0$*

3. $\log(x) > 0$ for $x > 1$

4. $\log(x_1 x_2) = \log(x_1) + \log(x_2)$

5. $\log(x_1/x_2) = \log(x_1) - \log(x_2)$

6. $\log(x^c) = c\log(x)$, c is any number.

The logarithm can be used for various approximations that arise in econometric applications. First, $\log(1 + x) \approx x$ for $x \approx 0$. Let x_0 and x_1 be positive values. Then, it can be shown (using calculus) that

$$log(x1) - log(x_0) \approx (x1 - x_0)/x_0 = \Delta x/x_0 \qquad (3)$$

for small changes in x. If we multiply equation (3) by 100 and write $\Delta \log(x) = \log(x_1) - \log(x_0)$, then

$$100\Delta \log(x) \approx \%\Delta x$$

One can write the elasticity of y with respect to x as

$$\frac{\Delta y}{\Delta x}\frac{x}{y} = \frac{\%\Delta y}{\%\Delta x}$$

In other words, the elasticity of y with respect to x is the percentage change in y when x increases by 1%. If y is a linear function of x, $y = \beta_0 + \beta_1 x$, then the elasticity is

$$\frac{\Delta y}{\Delta x}\frac{x}{y} = \beta_1 \frac{x}{y} = \beta_1 \frac{x}{\beta_0 + \beta_1 x}$$

which clearly depends on the value of x. (This is a generalization of the well-known result from basic demand theory: the elasticity is not constant along a straight-line demand curve.)
Elasticities are of critical importance in many areas of applied economics, not just in demand theory. It is convenient in many situations to have constant elasticity models, and the log function allows us to specify such models. If we use the approximation in equation (3) for both x and y, then the elasticity is approximately equal to $\Delta \log(y)/\Delta \log(x)$. Thus, a constant elasticity model is approximated by the equation

$$log(y) = \beta_0 + \beta_1\, log(x), \qquad (4)$$

and β_1 is the elasticity of y with respect to x (for $x, y > 0$).
Equation (4) defines a constant elasticity model. Such models play a large role in empirical economics. There are many other models that use log function.

The Exponential Function

The exponential function is the inverse to the log function: $\log[\exp(x)] = x$ for all x, and $\exp[\log(x)] = x$ for $x > 0$. Two important values of the exponential function are $\exp(0) = 1$ and $\exp(1) = 2.7183$ (to four decimal places). $\text{Log}(y) = \beta_0 + \beta_1 x$ is equivalent to

$$y = exp(\beta_0 + \beta_1 x).$$

If $\beta_1 > 0$, then x has an increasing marginal effect on y. Properties:

$$exp(x_0 + x_1) = exp(x_1)exp(x_2) \; and \; exp[c.log(x)] = x^c.$$

Differential Calculus

Let y = f (x) for some function *f.* Then, for small changes in x,

$$\Delta y = \frac{df}{dx}\Delta x \tag{3}$$

where *df /dx* is the derivative of the function *f*, evaluated at the initial point x_0 or dy/dx.

In applying econometrics, it helps to recall the derivatives of a handful of functions because we use the derivative to define the slope of a function at a given point. We can then use equation (7) to find the approximate change in y for small changes in x. In the linear case, the derivative is simply the slope of the line, as we would hope: if $y = \beta_0 + \beta_1 x$, then $dy/dx = \beta_1$.

If $y = x^c$, then $dy/dx = cx^{c-1}$. The derivative of a sum of two functions is the sum of the derivatives: $d[f(x) + g(x)]/dx = df(x)/dx + dg(x)/dx$. The derivative of a constant times any function is that same constant times the derivative of the function: $d[cf(x)]/dx = c[df(x)/dx]$.

Some functions that are often used in economics, along with their derivatives, are

1. $y = \beta_0 + \beta_1 x + \beta_2 x;\ dy/dx = \beta_1 + 2\beta_2 x$
2. $y = \beta_0 + \beta_1 / x;\ dy/dx = -\beta_1/(x^2)$
3. $y = \beta_0 + \beta_1 \log(x);\ dy/dx = \beta_1/x$
4. $y = exp(\beta_0 + \beta_1 x);\ dy/dx = \beta_1\, exp(\beta_0 + \beta_1 x)$

When *y* is a function of multiple variables, then partial derivatives are the norm. For:

$$y = f(x_1, x_2), \tag{8}$$

There are two partial derivatives, one with respect to x_1 and one with respect to x_2. The partial derivative of y with respect to x_1, denoted here by $\partial y/\partial x_1$, is just the usual derivative of equation (8) with respect to x_1, where x_2 is treated as a constant. Similarly, $\partial y/\partial x_2$ is just the derivative of equation (8) with respect to x_2, holding x_1 fixed.

Review of Probability

Random Variables

A random variable is one that takes on numerical values and has an outcome that is determined by an experiment. For example, the number of heads appearing in 10 flips of a coin is an example of a random variable. Before we flip the coin 10 times, we do not know how many times the coin will come up heads. Once we flip the coin 10 times and count the number of heads, we obtain the outcome of the random variable for this particular trial of the experiment. Another trial can produce a different outcome.

To analyze data collected in business and the social sciences, it is important to have a basic understanding of random variables and their properties. Following the usual conventions in probability and statistics, we denote random variables by uppercase letters, usually *W, X, Y ,* and *Z*; particular outcomes of random variables are denoted by the corresponding lowercase letters, *w, x, y,* and *z.*

We indicate large collections of random variables by using subscripts. For example, $X_1, X_2, \dots, X_n$; and the outcomes would be denoted $x_1, x_2, \dots, x_n$. Random variables are always defined to take on numerical values, even when they describe qualitative events. For example, consider tossing a single coin, where the two outcomes are heads and tails. We can define a random variable as follows: *X = 1* if the coin turns up heads, and *X = 0* if the coin turns up tails. A random variable that can only take on the values zero and one is called a binary or Bernoulli random variable.

Discrete Random Variables

A discrete random variable is one that takes on only a finite or countably infinite number of values. A Bernoulli random variable is the simplest example of a discrete random variable. The only thing we need to completely describe the behavior of a Bernoulli random variable is the probability that it takes on the value one. In the coin-flipping example, if the coin is "fair," then P (X = 1) = 1/2 (read as "the probability that X equals one is one-half"). Because probabilities must sum to one, P (X = 0) = 1/2, also.

One can define a Bernoulli random variable as X = 1 if event happens, and X = 0 if not.

The probability of the event happening is θ, so that $P\,(X = 1) = \theta\; P\,(X = 0) = 1 - \theta$ For example, if θ = 0.75, then there is a 75% chance that the event happens and a 25% chance that it doesn't occur.

More generally, any discrete random variable is completely described by listing its possible values and the associated probability that it takes on each value. If *X* takes on the *k* possible values $\{x_1, x_2, \dots, x_k\}$, then the probabilities $p_1, p_2, \dots,\ p_k$ are defined by

$$p_j = P\,(X = x_j\,), j = 1, \dots, k$$

where $0 \le p_j \le 1$, for $j = 1, \dots, k$ and

$$p_1 + \cdots + p_k = 1$$

The *probability density function (pdf)* of *X* summarizes the information concerning the possible outcomes of *X* and the corresponding probabilities:

$$f\,(x_j\,) = p_j\,, j = 1, \dots, k$$

f (x) is the probability that the random variable X takes on the value of x.

Continuous Random Variables

A variable X is a continuous random variable if it takes on any real value with zero probability. A probability density function for continuous random variables, provides information on the likely outcomes of the random variable. We use the pdf of a continuous random variable only to compute events involving a range of values. For example, if a and b are constants where a , b, the probability that X lies between the numbers a and b, P (a ≤ X ≤ b), is the area under the pdf between points a and b. The entire area under the pdf must always equal one.

When computing probabilities for continuous random variables, it is easiest to work with the cumulative distribution function (cdf) :

$$F\,(x)\ \equiv\ P\,(X\ \leq\ x) \qquad (9)$$

For a continuous random variable, F (x) is the area under the pdf, f , to the left of the point x. Because F (x) is simply a probability, it is always between 0 and 1. Further, $if\ x_1 < x_2, then\ P\,(X\ \leq\ x_1)\ \leq\ P\,(X\ \leq\ x_1), that\ is, F\,(x_1)\ \leq\ F\,(x_2)$. This means that a cdf is an increasing (or at least a nondecreasing) function of x.

Two important properties of cdfs that are useful for computing probabilities are the following:

$$For\ any\ number\ c, P\,(X\ >\ c)\ =\ 1\ -\ F\,(c)$$

$$For\ any\ numbers\ a\ <\ b, P\,(a\ <\ X\ \leq\ b)\ =\ F\,(b)\ -\ F\,(a)$$

Probability Distributions that involve more than one Random Variable

In econometrics, we are usually interested in how one random variable, call it Y, is related to one or more other variables. For now, suppose that there is only one variable whose effects we are interested in, call it X. The most we can know about how X affects Y is contained in the conditional distribution of Y given X. This information is summarized by the conditional probability density function, defined by

$$f_{Y\,|X}\,(y|x)\ =\ f_{X,Y}\,(x,y)/f_X\,(x) \qquad (11)$$

When X and Y are discrete. Then,

$$f_{Y\,|X}\,(y|x)\ =\ P\,(Y\ =\ y|X\ =\ x),$$

where the right-hand side is read as ”the probability that Y = y given that X = x.”

Characteristics of Probability Distributions

Measure of Central Tendency: Expectation (Mean)

If X is a random variable, the expected value (or expectation) of X, denoted E(X) or μ, is a weighted average of all possible values of X. The weights are determined by the probability density function. Sometimes, the expected value is called the population mean, especially when we want to emphasize that X represents some variable in a population.

For discrete random variables,

$$E(X)\ =\ x_1\,P\,r(X\ =\ x_1\,)\ + x_2\,P\,r(X\ = x_2\,)\ + \cdots\ + x_k P\,r(X\ =\ x_k)$$

For continuous random variables,

$$E(X)\ = \int_{-\infty}^{+\infty} x\,f(x)dx$$

which we assume is well defined. This can still be interpreted as a weighted average. For the most common continuous distributions, $E(X)$ is a number that is a possible outcome of X.

Properties of expectation:

1. For any constant $c, E(c) = c$.

2. For any constants a and $b, E(aX + b) = aE(X) + b$.

3. If $\{a_1, a_2, ..., a_n\}$ are constants and $\{X_1, X_2, ..., X_n\}$ are random variables, then $E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$.

Measure of Variability:

Variance

Another important feature of a random variable is variance, which is used to measure the spreads of distributions. For a random variable X, let $\mu = E(X)$. There are various ways to measure how far X is from its expected value, but the simplest one to work with algebraically is the squared difference, $(X - \mu)^2$. One such number is the variance, which tells us the expected distance from X to its mean:

$Var(X) \equiv E(X - \mu)^2$ From above equation, it follows that the variance is always nonnegative.

Properties of variance:

1. For any constant $c, Var(c) = 0$

2. For any constants a and $b, Var(aX + b) = a^2Var(X)$.

3. $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$, where $Cov(X,Y)$ denotes the covariance between X and Y.

Another measure that is related to variance is standard deviation, denoted as $sd(X)$, which is the positive square root of the variance $sd(X) = \left|\sqrt{Var(X)}\right|$.

Covariance

The covariance between two random variables X and Y, sometimes called the population covariance to emphasize that it concerns the relationship between two variables describing a population, is defined as the expected value of the product $(X - \mu X)(Y - \mu Y)$:

$$Cov(X,Y) = E[(X - \mu X)(Y - \mu Y)]$$

which is sometimes denoted σ_{XY}. If $\sigma_{XY} > 0$, then, on average, when X is above its mean, Y is also above its mean. If $\sigma_{XY} < 0$, then, on average, when X is above its mean, Y is below its mean.

Covariance measures the amount of linear dependence between two random variables. A positive covariance indicates that two random variables move in the same direction, while a negative covariance indicates they move in opposite directions.

Properties of covariance:

1. If X and Y are independent, then $Cov(X,Y) = 0$

2. For any constants $a_1, a_2, b_1,$ and b_2,

$$Cov(a_1X + b_1, a_2Y + b_2) = a_1a_2Cov(X,Y)$$

An important implication of the second property is that the covariance between two random variables can be altered simply by multiplying one or both of the random variables by a constant.

Correlation

Suppose we want to know the relationship between amount of education and annual earnings in the working population. We could let X denote education and Y denote earnings and then compute their covariance. But the answer we get will depend on how we choose to measure education and earnings. The second property of covariance implies that the covariance between education and earnings depends on whether earnings are measured in dollars or thousands of dollars, or whether education is measured in months or years. It is pretty clear that how we measure these variables has no bearing on how strongly they are related. But the covariance between them does depend on the units of measurement.

The fact that the covariance depends on units of measurement is a deficiency that is overcome by the correlation coefficient between X and Y:

$$Corr(X,Y) = \frac{Cov(X,Y)}{sd(X)sd(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Properties of correlation:

1. $-1 \leq Corr(X,Y) \leq 1$

2. For constants a_1, b_1, a_2, b_2, with $a_1 \, a_2 > 0$,

$$Corr(a_1 X + b_1, a_2 Y + b_2) = Corr(X,Y)$$

If $a_1 \, a_2 < 0$,

$$Corr(a_1 X + b_1, a_2 Y + b_2) = -Corr(X,Y)$$

Review of Mathematical Statistics

Populations, Parameters, and Random Sampling

Statistical inference involves learning something about a population given the availability of a sample from that population. By population, we mean any well-defined group of subjects, which could be individuals, firms, cities, or many other possibilities. By "learning," we can mean several things, which are broadly divided into the categories of estimation and hypothesis testing.

The first step in statistical inference is to identify the population of interest. This may seem obvious, but it is important to be very specific. Once we have identified the population, we can specify a model for the population relationship of interest. Such models involve probability distributions or features of probability distributions, and these depend on unknown parameters. Parameters are simply constants that determine the directions and strengths of relationships among variables.

For reviewing statistical inference, we focus on the simplest possible setting.

Let Y be a random variable representing a population with a probability density function f (y; μ), which depends on the single parameter μ. The probability density function (pdf) of Y is assumed to be known except for the value of μ; different values of μ imply different population distributions, and therefore we are interested in the value of μ. If we can obtain certain kinds of samples from the population, then we can learn something about u. The easiest sampling scheme to deal with is random sampling.

Random Sampling: If $Y_1, Y_2, \ldots, Y_n$ are independent random variables with a common probability density function $f\ (y;\ \mu)$, then $\{Y_1, Y_2, \ldots, Y_n\}$ is said to be a random sample from $f\ (y;\ \mu)$ [or a random sample from the population represented by $f\ (y;\ \mu)$].

The random nature of $Y_1, Y_2, \ldots, Y_n$ in the definition of random sampling reflects the fact that many different outcomes are possible before the sampling is actually carried out. For example, if family income is obtained for a sample of n = 100 families in the United States, the incomes we observe will usually differ for each different sample of 100 families. Once a sample is obtained, we have a set of numbers, say, $\{Y_1, Y_2, \ldots, Y_n\}$, which constitute the data that we work with. Whether or not it is appropriate to assume the sample came from a random sampling scheme requires knowledge about the actual sampling process.

Finite Sample Properties of Estimators

In this subsection, we study what are called finite sample properties of estimators. The term "finite sample" comes from the fact that the properties hold for a sample of any size, no matter how large or small. Sometimes, these are called small sample properties.

Estimators and Estimates

To study properties of estimators, we must define what we mean by an estimator. Given a random sample $\{Y_1, Y_2, \ldots, Y_n\}$, drawn from a population distribution that depends on an unknown parameter θ, an estimator of θ is a rule that assigns each possible outcome of the sample a value of θ. The rule is specified before any sampling is carried out; in particular, the rule is the same regardless of the data obtained. As an example of an estimator, let $\{Y_1, Y_2, \ldots, Y_n\}$ be a random sample from a population with mean μ. A natural estimator of μ is the average of the random sample:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$\bar{Y}$ is called the sample average. Given any outcome of the random variables $Y_1, Y_2, \ldots, Y_n$, we use the same rule to estimate μ : we simply average them. For actual data outcomes $y_1, \ldots,\ y_n$, the estimate is just the average in the sample

$\bar{Y}$ = (y1 + y2 + · · · + y/n)/n.

More generally, an estimator W of a parameter u can be expressed as an abstract mathematical formula:

$$W\ =\ h(Y_1, Y_2, \ldots, Y_n)$$

for some known function h of the random variables $Y_1, Y_2, \ldots, Y_n$.

Unbiasedness

An estimator, W of θ, is an unbiased estimator if $E(W\ )\ =\ \theta$, for all possible values of θ.

It means that if we could indefinitely draw random samples on Y from the population, compute an estimate each time, and then average these estimates over all random samples, we would obtain u.

The unbiasedness of an estimator and the size of any possible bias depend on the distribution of Y and on the function h.

We now show that the sample average Y is an unbiased estimator of the population mean μ, regardless of the underlying population distribution:

$$\overline{E(Y)} = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n = \mu$$

For hypothesis testing, we will need to estimate the variance s2 from a population with mean μ. Letting $Y_1, Y_2, \ldots, Y_n$ denote the random sample from the population with $E(Y) = \mu$ and $Var(Y) = \sigma^2$, define the estimator as:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

which is usually called the sample variance. It can be shown that S^2 is unbiased for σ^2 E(S) = σ^2 .

Simple Linear Regression

In economics we are mostly interested in" explaining y in terms of x" or "studying how y varies with changes in x". For example,

Three issues in writing down a model that will explain y in terms of x:

1. Since there is never an exact relationship between two variables in economics. How do we "allow" x to affect y?

2. What is the functional relationship between y and x?

3. How can we be sure we are capturing a ceteris paribus relationship between y and x?

One way to resolve the above three issues is to write down a linear equation relating y to x with an error term u.

$$y = \beta_0 + \beta_1 x + u \qquad (12)$$

We typically refer y as the Dependent Variable and x as the Independent Variable. We also typically refer β_0 and β_1 as intercept and slope coefficient, respectively. And we refer u as error terms.

The model in equation (12) solves the first two issues. Adding the additional error term, allow x and y to have some, but not exact, relationship. Moreover, by writing down the linear function like in equation (12), we assume x can only affect y linearly. The linearity of equation (12) implies that a one-unit change in x has the same effect on y, regardless of the initial value of x. This is unrealistic for many economic applications.

Concerning the third issue we are only able to get reliable estimators of β_0 and β_1 from a random sample of data when we make an assumption restricting how the unobservable u is related to the explanatory variable x. Without such a restriction, we will not be able to estimate the ceteris paribus effect, β_1. Because u and x are random variables, we need some knowledge based on probability.

We now turn to the crucial assumption regarding how u and x are related.

We assume that the average value of u does not depend on the value of x, and it is equal to zero. We can write this assumption as

$$E(u|x) = E(u) = 0, \quad (13)$$

Equation (13) says that the average value of the unobservables is the same across all slices of the population determined by the value of x and that the common average is necessarily equal to the average of u over the entire population. When assumption (13) holds, we say that u is mean independent of x. With this assumption, we can show that

$$E(y|x) = E(\beta_0 + \beta_1 x + u|x) = \beta_0 + \beta_1 x + E(u|x) = \beta_0 + \beta_1 x$$

This result indicates that the our simple regression model is associated with the conditional expectation of y given x. The following figure shows the regression line,

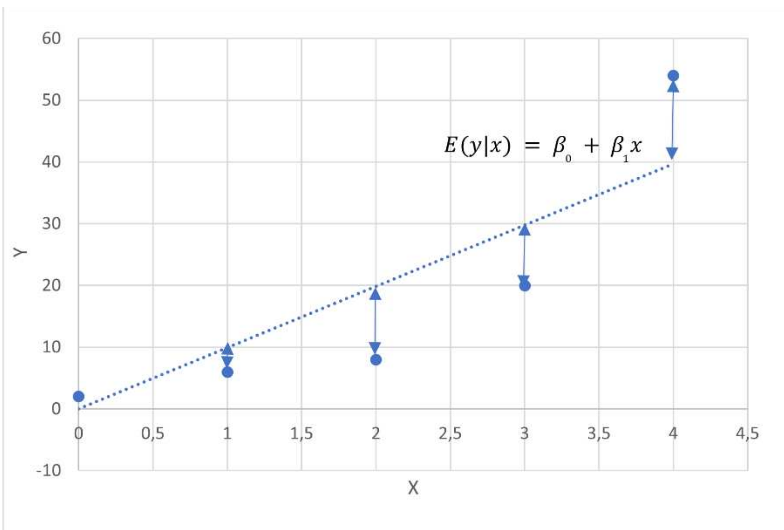


Figure 1: Population Regression Line

Now we know what we are looking for in a simple regression model in population. But the question is, how can we estimate β_0 and β_1 with a finite sample observations. We have two methods to get our estimators.

Least Square Method

We want to use sample information to reveal the population parameters that we are interested. In simple regression model, we are interested in estimating β_0 and β_1, and we want to use a random sample from the population to reveal them.

Based on equation (12), we can write the following model for each observation

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

The goal is to minimize the sum of square of error terms by choosing the values of β_0 and β_1. The problem is to solve:

$$\min \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x)^2$$

If we derivate this expression and rearrange and solve we obtain:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{X})(Y_i - \bar{Y}\,)}{\sum_{i=1}^{n} (x_i - \bar{X})(x_i - \bar{X}\,)}$$

The numerator is an estimator of covariance between x and y, and the denominator estimator of variance of x. This implies:

1. The sign of $\hat{\beta}_1$ depends on the sign of the covariance between x and y. When x and y are positively (negatively) correlated, $\hat{\beta}_1$ is positive (negative).

2. We need to have variability in X to have an meaningful estimator. If the denominator (variance of x) is very small or 0 $\hat{\beta}_1$ would become very big or infinity.

The above least square estimators are usually called OLS estimators. Intuitively, OLS is fitting a line through the sample points such that the sum of squared residuals is as small as possible, hence the term least squares. The residual, $\hat{u}_1$, is an estimate of the error term, u, and is the difference between the fitted line (sample regression function) and the sample point (dotted lines between regression line and observations):

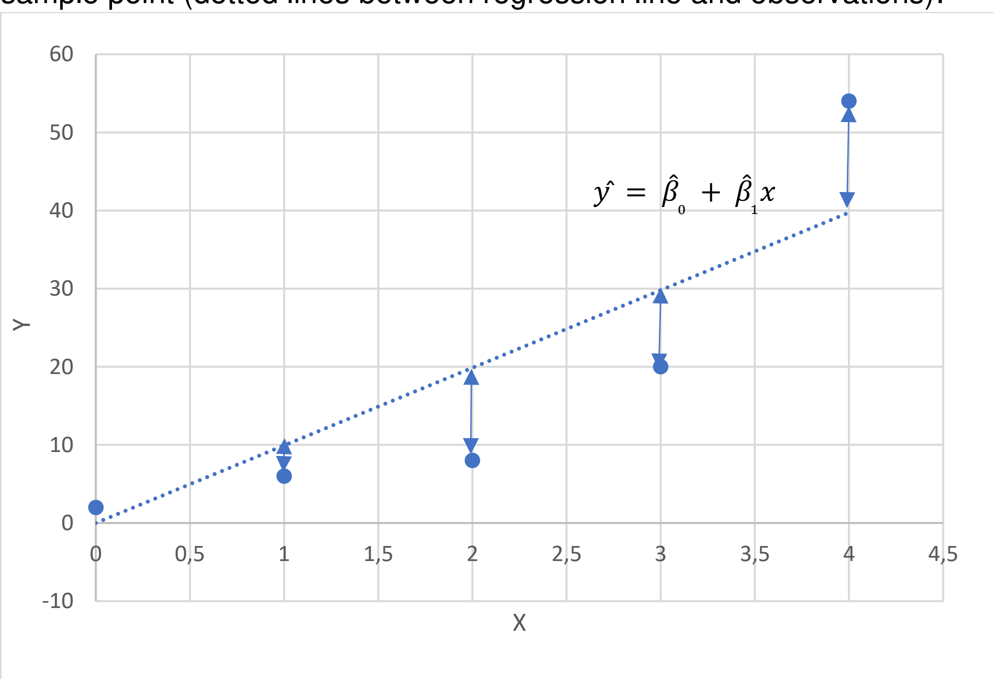


Figure 2: Sample Regression Line

In figure 1, we have conditional expectation of y given x as the population regression line, and the difference between the actual y and the population regression line is u, the error term. In figure 2, we have sample regression line, $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x$, and the

difference between the observed y and the sample regression line is u, the residuals. In fact, the sample regression line will only be equal to the population regression line" in average". Finally, $\hat{y}_i = y_i + \hat{u}$. For $i = 1, \dots, n$. So for every observation y_i is equal to the fitted value $\hat{y}_i$ where we can get to SST=SSE+SSR which means that the "sum of square variation" (SST) is equal to the "sum off square of explained variation" (SSE) and the "sum of square of residuals" (SSR). This is the basis of R-squared which be explained as "the percentage of the total variation of the data that can
be explained by the regression model". A larger R-square means that the model can explain more total variation of the data, therefore is in general preferred in simple regression model:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

Unbiasedness of the OLS estimators

Under four conditions an OLS estimator is unbiased $E(\widehat{\beta_0}) = \beta_0$ and $E(\widehat{\beta_1}) = \beta_1$

1. Population model is linear in parameters as $y = \beta_0 + \beta_1 x + u$

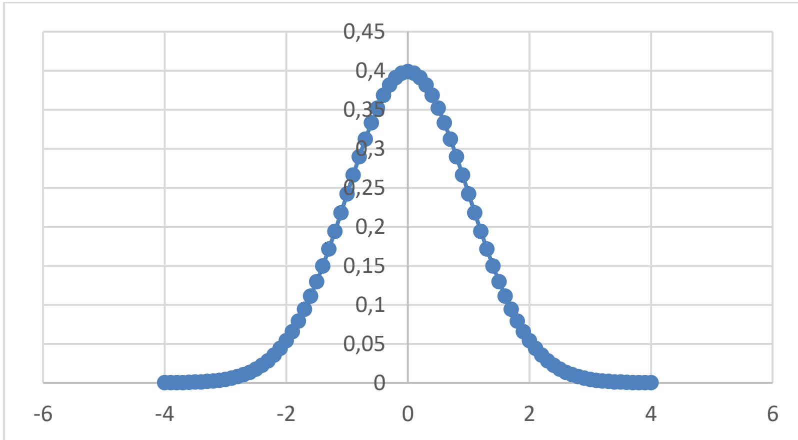
2. One can random sample of size n, $\{(x_i, y_i) : i = 1, 2, \dots, n\}$, from the population model. Thus, we can write the sample model $y_i = \beta_0 + \beta_1 x_i + u_i$.

3. $E(u|x) = 0$ and thus $E(u_i|x_i) = 0$.

4. There is variation in the x_i

Variance of the OLS estimators

In social sciences there is usually only one sample so a low variance is very important as one random draw of the random variable can reveal a lot of information about the expectation. The following are two figures of the density functions from two different random variables.



2 Figure 3: Normal Distribution with Expectation = 0 and Variance = 1

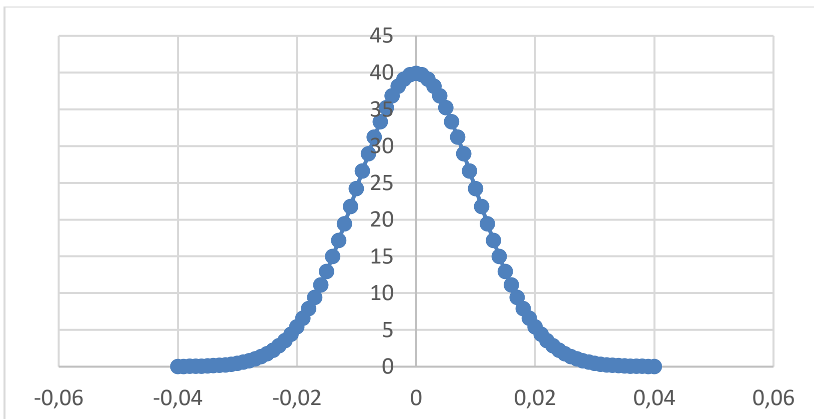


Figure 4: Normal Distribution with Expectation = 0 and Variance = 0.01

In the first case, when variance = 1, one random draw can be any number between -4 and 4, which can be far away from the expectation. When variance = 0.01, one random draw can be any number between -0.04 and 0.04, which are all close to the expectation. If the variance of the estimator is small, one random draw of β_1 based on our one sample will not be too different from the expectation. In this case, one random draw of β1 is good enough for us to learn about β_1.

5. Homoskedasticity: $Var(u_i|X) = \sigma^2$

We assume that the conditional variance of error terms is a constant for all the different values of x.

The variance of β_1 is obtained from the definition of $\widehat{\beta_1}$:

$$Var\ (\beta_1|X) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{X})^2}$$

The main implication from this equation is that for large sample sizes and/or larger sample variation the denominator becomes large and so the variance becomes smaller.

Now the only thing left in the above estimator is that we don't know σ^2. In econometrics, we usually employ the following estimator for σ^2 in simple regression model,

$$\sigma^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)$$

And finally using the above estimator we obtain the standard error (se) used for measuring the accuracy of the estimator:

$$se\left(\widehat{\beta}_1\right) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}}$$

If all the 5 assumptions in the previous section, then simple regression model can be a good choice. But this is usually not the case. In reality, we usually need to control more than one factor in our analysis. In this case, we will need multiple regression model, the definition of multiple regression model is as following:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, E(u|x_1, x_2, \ldots, x_k) = 0 \quad (31)$$

In the above multiple regression model, we have k different factors that we want to control. If there are more than one factor that need to be controlled, but we somehow only control one of them as it is the case of simple regression, we will get biased estimators.

Bibliography

Apostol, Tom M., Mathematical analysis (2nd edition), Addison-Wesley, Reading, Massachusetts, 1974

A. C. Chiang and K. Wainwright, Fundamental methods of mathematical economics, 4th ed. New York: McGraw-Hill, 2005.

AISTEDA Econometrics study guide, Econometrics and Mathematical Economics

K. Sydsæter, P. J. Hammond, and A. Strøm, Essential mathematics for economic analysis, 4th ed. Harlow: Pearson, 2012.