Lecture Notes on Industrial Organization

Xavier Martinez-Giralt
CODE and Department of Economics
Universitat Autònoma de Barcelona

Lecture Notes on Industrial Organization

Contents

List of Figures

List of Tables

Preface

Chapter 1

Introduction

Industrial Organization[1], Industrial Economics, Oligopoly, Imperfect Competition, ... All these are well known labels to address one of the oldest problems in economics, namely how prices arise in the market when there are few competitors.

We will start with a review of the ideas of the founding fathers of the oligopoly problem, Cournot, Bertrand, Edgeworth, Stackelberg, Chamberlin, Robinson, and Hotelling. Next we will present the contrast between the so-called classical Industrial economics with the modern industrial economics. Finally, we will address the issue of the adequacy of the partial equilibrium framework where most of the contributions in this area are developed.

1.1 A historical appraisal.

A complete account of the early ideas in what today is known as Industrial Organization can be found in Schumpeter (1958). For our purpose, Cournot (1838) was the first in proposing a solution concept to determine market prices under oligopolistic interaction. By means of an example of two producers of mineral water deciding production levels and competing *independently*, Cournot proposes that the price arising in the market will be determined by the interplay of aggregate supply and demand. Also, such a price will be an equilibrium price when every producer's production decision maximizes its profits conditional on the expectation over the production of the rival. It is worth noting that this equilibrium involves a price above the marginal cost of production. This concept of equilibrium is precisely what Nash (1950) proposed as solution of a non-cooperative game when we consider quantities as strategic variables. Next, Cournot tackles the case of complementary products. Interestingly enough he assumed in this case

[1]This chapter is based on Martin (2002, ch. 1), Tirole (1988, Introduction), and Vives (1999, ch. 1)

that producers would choose prices and applied the same solution concept, namely a Nash equilibrium with prices as strategic variables. In this case, the equilibrium price is larger than the monopoly price.

Cournot's contribution was either ignored or unknown for 45 years until Bertrand (1883) published his critical review where he claims the the obvious choice for oligopolists competing in a homogeneous product market such as the proposed by Cournot would be to collude, given that the relevant strategic variables must be prices rather than quantities. In particular, in Cournot's example, the equilibrium price will equal marginal cost, i.e. the competitive solution.

The criticism of the Cournot model continued with Marshall (1920) and Edgeworth (1897). Marshall thought that under increasing returns, monopoly was the only solution; Edgeworth's main idea was that in Cournot's set up the equilibrium is indeterminate regardless of products being substitutes or complements. For substitute goods with capacity constraints (Edgeworth (1897)) or with quadratic cost (Edgeworth (1922)) he concludes that prices would oscillate cycling indefinitely. For complementary products the indetermination of the equilibrium is "at least very probable" (Vives (1999 p. 3)).

This demolition of Cournot's analysis was called to an end by Chamberlin (1929) and Hotelling (1929) after the observation that neither assumption of quantities or prices as strategic variables are correct in an absolute sense:

> Equilibrium in the Bertrand model with a standardized product is quite different from equilibrium in the Cournot model. The Cournot model emphasizes the number of firms as the critical element in determining market performance. Bertrand's model predicts the same performance as in long-run equilibrium of a perfectly competitive market if as few as two producers supply a standardized product.
>
> The qualitative nature of the predictions of the Cournot model are robust to the introduction of product differentiation. The same cannot be said of the Bertrand model. (Martin (2002, p.60)).

From that point Cournot's model served as a departure point to other analysis. Hotelling (1929), Chamberlin (1933), and Robinson (1933) introduced product differentiation. Hotelling's segment model introduces different preferences in consumers and provides the foundation for location theory by assuming consumers buying at most one unit of one commodity; Chamberlin and Robinson considered a large number of competitors producing slightly different versions of the same commodity (thus allowing them to retain some monopoly power on the market) and assumed that consumers had convex preferences over the set of varieties. Stackelberg (1934) considered a sequential timing in the firms' decisions, thus incorporating the idea of commitment.

Some years later, von Neumann and Morgenstern (1944) and Nash (1950, 1951) pioneered the development of game theory, a toolbox that provided the most flourishing period of analysis in oligopoly theory along the 1970's. Refinements of the Nash equilibrium solution like Selten's subgame perfect equilibrium (1965) and perfect equilibrium (1975), Harsanyi's Bayesian Nash equilibrium (1967-68), or Kreps and Wilson's sequential equilibrium (1982) have proved essential to the modern analysis of the indeterminacy of prices under oligopoly.

Also, the study of mechanisms allowing to sustain (non-cooperative) collusion was possible with the development of the theory of repeated games lead by Friedman (1971), Aumann and Shapley (1976), Rubinstein (1979), and Green and Porter (1984). Figure 1.1 summarizes this discussion.

1.2 Oligopoly theory vs. the SCP paradigm.

Industrial Economics, as we have already mentioned, deals with the study of the behavior of firms in the market. The field as a separate area within microeconomics appears after the so-called monopolistic competition revolution, linked to the names of Mason (1939) and Bain (1949, 1956) ("Harvard tradition"). Barriers to entry was the central concept giving rise to market power. The approach is essentially motivated by stylized facts arising from an empirical tradition seeking how the structural characteristics of an industry determine the behavior of its producers that, in turn, yields market performance. This framework of analysis is known as the Structure-Conduct-Performance paradigm. Martin (2002), p.6 reproduces the figure 1.2 from Scherer (1970) showing the SCP paradigm. Schmalensee (1989) provides a very nice survey of this approach.

This paradigm dominated the evolution of the field for three decades. During these years research was mainly discursive and informal and independent of the formal microeconomic analysis of imperfect markets. Basically, the SCP provided a general framework allowing the implementation of public policies from empirical regularities observed in many industries. The early seventies witnessed a major revolution in the analysis, leading to the so-called "new industrial economics". Following Martin (2002), p.8, three factors are behind this evolution. (i) the conclusions of the formal microeconomic models are not qualitatively different from those of the SCP paradigm.; (ii) empirical economists held that market structure should be treated as endogenous rather than exogenous with respect to conduct and performance. This raised the need for a theoretical foundation of the econometric models (to be found in the microeconomic models of oligopoly); (iii) last but not least, the application of game theory to the modeling of oligopolistic interaction provided the definite element to replace the SCP paradigm and place Oligopoly Theory (understood as the analysis of strategic interactions being cen-

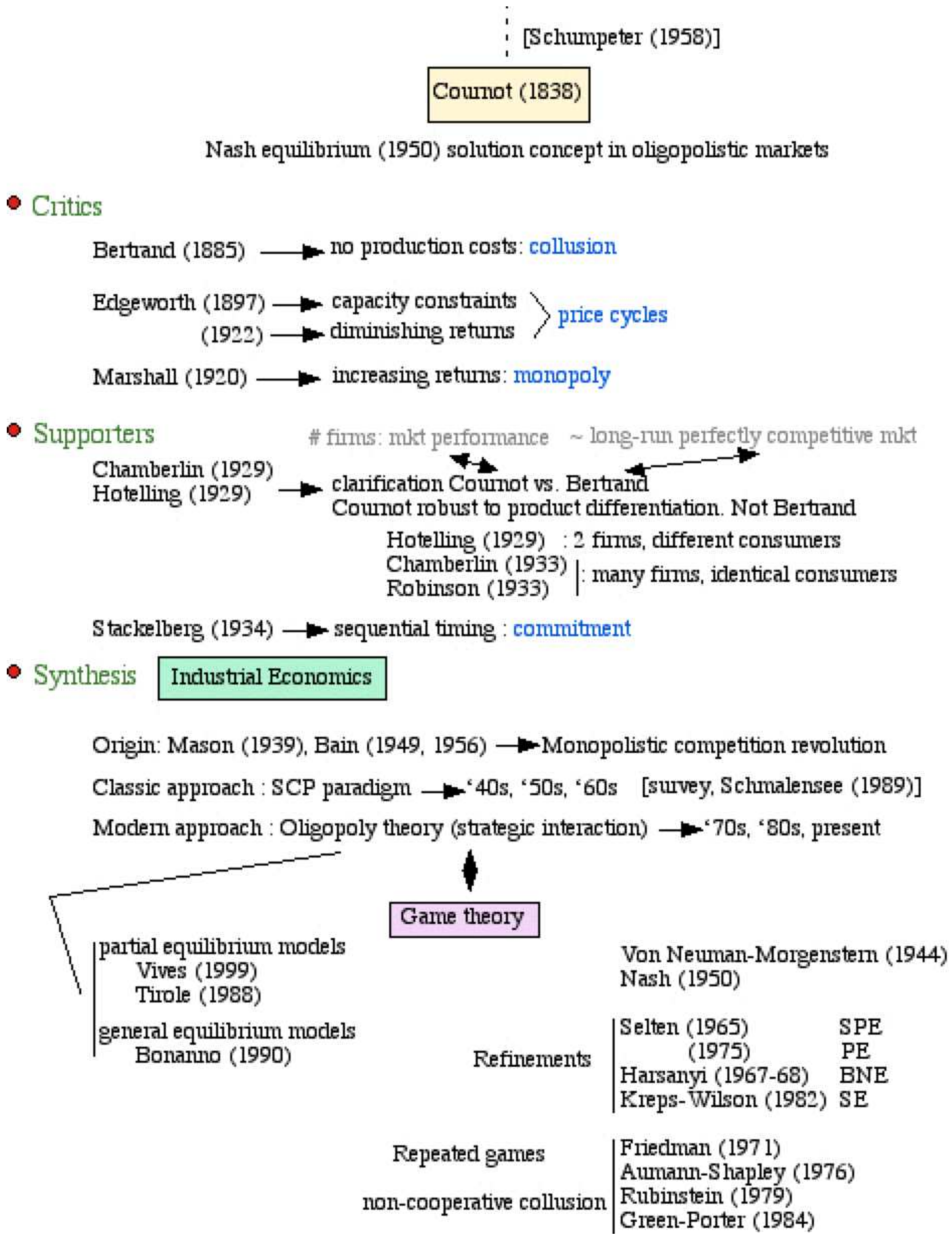

Figure 1.1: History of IO.

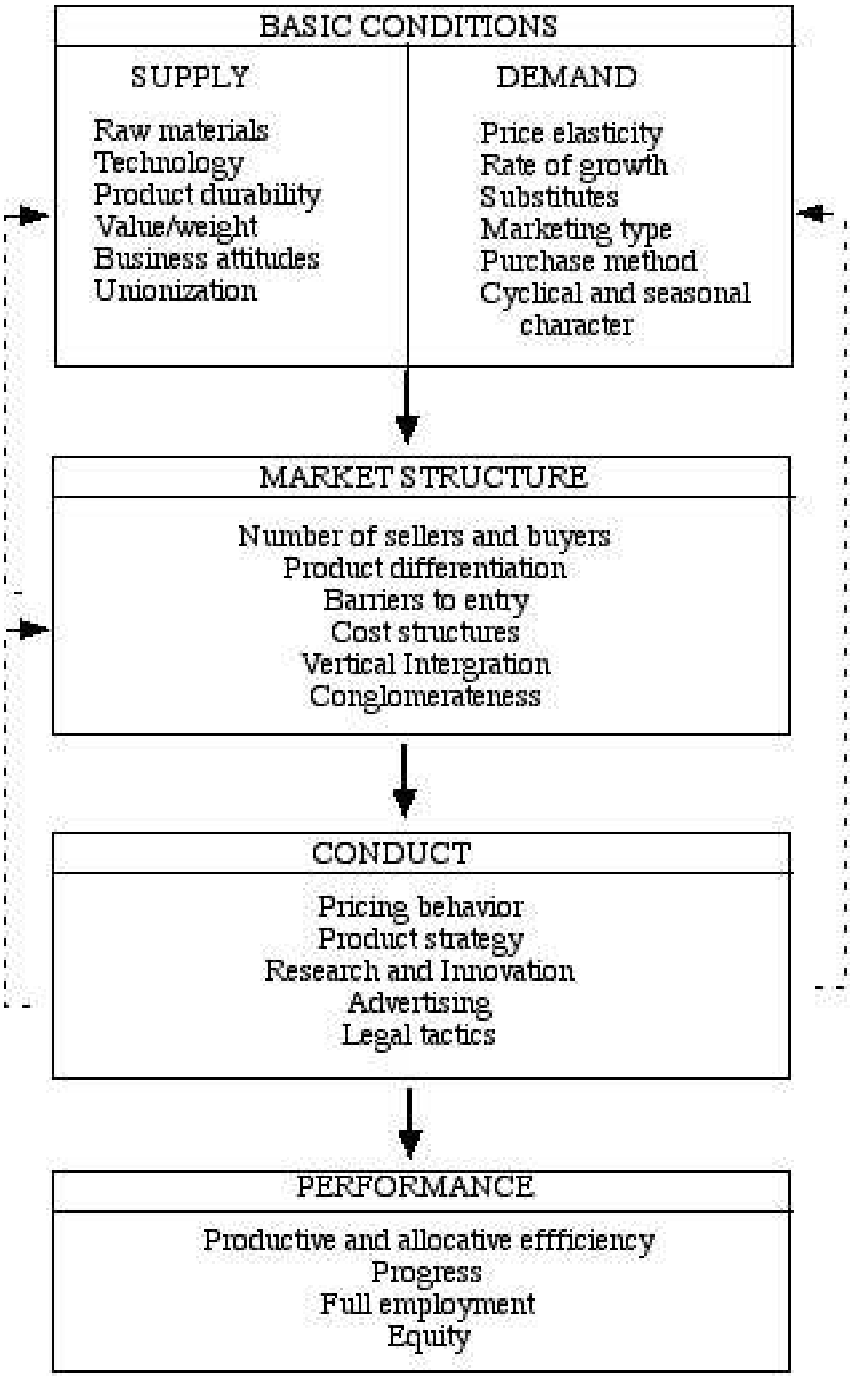

Figure 1.2: The Structure-Conduct-Performance paradigm .

tral to the determination of market performance) and the standing methodology.

The two decades from the early 1970's until the late 1980's has been the most flourishing period of theoretical development in industrial organization. The main methodological difference with respect to the SCP paradigm is that game-theoretical models are rather specific and their predictions about equilibrium behavior often not robust to minor changes in the set of underlying assumptions.

Most of the literature on oligopoly theory has been developed using models of partial equilibrium[2]. That is the model focuses in an industry and the interactions with the rest of the economy are neglected. This approach goes back to Marshall (1920). His idea is that the partial equilibrium model only makes sense when the income effects are small. In this case the share of consumer expenses in the industry under analysis will be small and small changes in the industry should not give rise to variations in the other markets of the economy. Vives (1999, section III.2) presents a rigorous foundation for these ideas. Following Cournot, it is generally assumed that firms face a downward sloping demand curve (except in the models of spatial competition). Also, it is assumed that welfare changes are adequately measured by variations in consumer surplus, a concept introduced by Dupuit (1844). We will study this last concept and will see that the consumer surplus is a precise measure of the change in consumer welfare only when preferences are quasilinear. Then we will verify that such assumption is meaningful only when the income effects are small, that is when the share of consumer expenses in the industry under analysis is small.

1.3 Variations of prices and welfare.

Variations in the economic environment (price changes, taxes, etc) give rise to variations in the consumers' welfare. Thus, it is reasonable to try to obtain quantitative estimations of those changes in prices and welfare with clear economic interpretations.

The classic and most used measure of welfare variation is the *consumer surplus*. The problem with this measure is that it is precise only in the special case of quasilinear preferences. En general, consumer surplus only gives an approximation of the impact on welfare of a variation of some basic magnitude of the economy. Therefore, before focussing the attention in the consumer surplus, we

[2]There is however a whole line of general equilibrium models of oligopoly started by Negishi (1961) and continued by Gabszewicz and Vial (1972), Shitovitz (1973), Novshek and Sonnenschein (1978), Mas-Colell (1982), and Codognato and Gabszewicz (1991). See also a survey paper by Bonanno (1990).

will examine some more general methods. These are the *compensating variation* and the *equivalent variation*.

1.3.1 Price indices.

To start at the beginning, let us suppose that the basic economic magnitude suffering a variation are prices. Thus, we have to construct some measures of that variation that will prove useful in the study of the impact on welfare. These are the *price indices*[3].

Definition 1.1 (Price index)**.** *A price index measures the impact on the welfare level following a price variation.*

Let us consider two price vectors $p^0, p^1 \in \mathbb{R}^l_+$, where p^0 represents the initial situation and p^1 the new price level after the variation. How can we measure the impact of this price variation on the cost of living?

A first approach consists in considering a reference consumption bundle, x_i^R and evaluate it at both systems of prices. We obtain an index of the following type:

$$PI(p^0, p^1, x_i^R) = \frac{p^1 x_i^R}{p^0 x_i^R}. \tag{1.1}$$

This index measures the cost of the bundle x_i^R at prices p^1 with respect to the cost of this very bundle at prices p^0. The relevance of the index thus obtained depends on how representative is the bundle x_i^R in the economy.

Two indices built in this fashion are linked to the names of Laspeyres and Paasche. The difference between them is the reference consumption bundle. The former uses the bundle in the initial period x_i^0, the latter the resulting bundle after the price variation, x_i^1:

$$PI_L(p^0, p^1, x_i^0) = \frac{p^1 x_i^0}{p^0 x_i^0},$$
$$PI_P(p^0, p^1, x_i^1) = \frac{p^1 x_i^1}{p^0 x_i^1}.$$

The drawback of this family of indices is that given that the consumption bundle is fixed, they cannot capture the substitution effects associated with the price change.

An alternative family of indices overcoming this problem should estimate the impact of the price change on the utility level. The natural way of constructing such an index should use the expenditure function, that measures the cost of

[3]see Villar (1996, pp.72-73)

reaching a certain utility level at a given prices. Therefore, from a utility level of reference u_i^R we can construct the so-called *true price index* as

$$TPI(p^0, p^1, u_i^R) = \frac{e_i(p^1, u_i^R)}{e_i(p^0, u_i^R)}.$$

Naturally, the usefulness of this index depends on how representative is the utility level u_i^R. With the same logic behind the Laspeyres and Paasche indices, we can obtain the corresponding true Laspeyres price index and true Paasche price index.

$$TPI_L(p^0, p^1, u_i^0) = \frac{e_i(p^1, u_i^0)}{e_i(p^0, u_i^0)},$$
$$TPI_P(p^0, p^1, u_i^1) = \frac{e_i(p^1, u_i^1)}{e_i(p^0, u_i^1)}.$$

It is easy to prove (this is left to the reader) that the true Laspeyres price index is a lower bound of the Laspeyres index, and that the true Paasche price index is an upper bound of the Paasche price index, that is,

$$TPI_L \leq PI_L,$$
$$TPI_P \geq PI_P.$$

Finally, note that when preferences are homothetic, the true Laspeyres and Paasche price indices coincide (again, the proof of this statement is left to the reader).

1.3.2 Welfare variations

We will examine the effect of a price change (due, for instance to a variation in taxes) on the welfare level of an individual. Let us consider, as before, two price vectors $p^0, p^1 \in \mathbb{R}_+^l$, where p^0 represents the initial situation and p^1 the new situation. Let us also assume that wealth remains constant in both scenarios. This is a simplifying assumption. To see the effect of the price change on consumers we only have to compare the utility levels in both situations evaluated at the corresponding consumption bundles, $u_i(x_i^0)$ and $u_i(x_i^1)$. Similarly, we can compare the indirect utility levels $v_i(p^0, w_i)$ and $v_i(p^1, w_i)$. These comparisons are *ordinal*, that is they only tell us whether the consumer is better off or worse off after the price variation, but do not tell us anything about how much better off or worse off the consumer is.

To overcome this limitation of the analysis, we can consider the expenditure function as representation of the indirect utility function. Thus, let us consider a reference price vector p^R together with the price vectors p^0 y p^1. Next, let us compute $e_i[p^R, v_i(p^j, w_i)]$, $j = 0, 1$. These functions tell us the amount of money, given prices p^R, necessary to achieve the utility level $u_i^j = v_i(p^j, w_i)$. Given that the expenditure function is strictly increasing in u_i, we can think of the expenditure function as a monotonic increasing transformation of v_i and therefore an alternative representation of the individual utility. This argument allows us to express the indirect utility in monetary units (Euros), and thus obtain a quantitative measure of welfare variation. In other words, the difference,

$$e_i[p^R, v_i(p^1, w_i)] - e_i[p^R, v_i(p^0, w_i)]$$

tells us how much does our welfare level vary when prices change from p^0 to p^1 in Euros relative to the price vector p^R. Of course, the selection of p^R is crucial to obtain a meaningful interpretation. The obvious candidates are the prices corresponding to the initial situation or to the final situation. As in the case of the price indices, this gives rise to two different measures of welfare variation. Before introducing these measures, let us recall that we are assuming that wealth remains constant form one situation to the other, namely $e_i[p^0, v_i(p^0, w_i)] = e_i[p^1, v_i(p^1, w_i)] = w$.

Definition 1.2 (Equivalent variation)**.** *The equivalent variation is the change in consumer wealth equivalent, in terms of welfare, to the price change:*

$$EV_i(p^0, p^1, w) = e_i[p^0, v_i(p^1, w_i)] - e_i[p^0, v_i(p^0, w_i)] = e_i[p^0, v_i(p^1, w_i)] - w.$$

The equivalent variation tells us that the price change from p^0 to p^1 has the same impact on welfare as an income change from w_i to $(w_i + EV_i)$. Therefore, EV_i will be negative when the price change will worsen the situation of the consumer and positive otherwise.

Definition 1.3 (Compensating variation)**.** *The compensating variation is the change in consumer wealth necessary to maintain that consumer in the initial welfare level after a price change has occurred:*

$$CV_i(p^0, p^1, w) = e_i[p^1, v_i(p^1, w_i)] - e_i[p^1, v_i(p^0, w_i)] = w - e_i[p^1, v_i(p^0, w_i)].$$

The compensating variation is a modification in the wealth of the individual to maintain him (her) in his (her) initial utility level. Hence, that modification will be negative (an increase in income) when the change of prices will worsen the situation of the consumer and positive otherwise.

We can thus conclude that when we compare two scenarios, both measures go in the same direction, although not in the same magnitude given that the price vectors at which both scenarios are evaluated are different. This statement does not hold true when we compare more than two situations. In that case the EV_i turns out to be a better measure than the CV_i (see Villar (1996, p.75)).

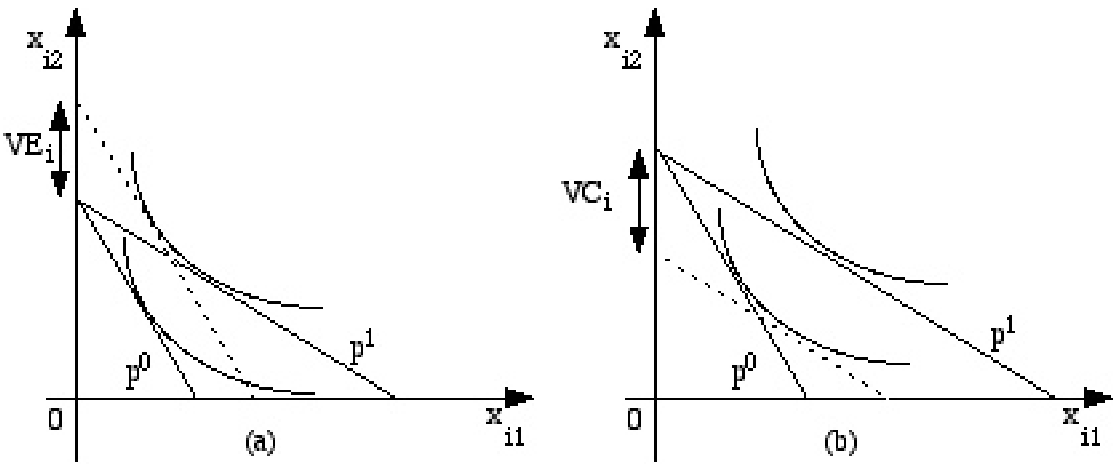

Figure 1.3: Equivalent variation and compensating variation

Figure 1.3 illustrates the argument when the price of good 1 decreases from p^0 to p^1. Section (a) in the figure represents the equivalent variation of income, i.e. how much additional money is needed at the price vector p^0 to maintain the consumer at the same welfare level as with prices p^1. Part (b) represents the compensating variation of income, that is how much money do we have to subtract from the consumer to maintain him (her) at the same welfare level as with prices p^0.

1.3.3 Consumer surplus

The concept of *consumer surplus* gives an approximation of the impact of a change of prices on consumer welfare. In contrast with the equivalent variation and the compensating variation, it is easier to compute because it uses the demand function. However, it only allows to obtain an approximation to the true value (except in one particular case that we will examine below).

To illustrate the idea of consumer surplus, let us consider a market of a good where a monopolist knows the demand curve of the consumer. This monopolist by setting a price p^0 would sell x^0 units, so that its revenue would be p^0x^0. Let us assume that the monopolist would like to sell precisely these x^0 units to the consumer. In an effort to maximize profits, and given that the monopolist knows the demand function, it can sell every unit separately until it reaches the quantity x^0. According to demand function, it can sell the first unit at a much higher price than the competitive price, the second unit to a slightly lower price, and so on until reaching the x^0th unit that sells at the price p^0. The difference in revenues obtained by the monopolist using this discriminatory mechanism and the uniform

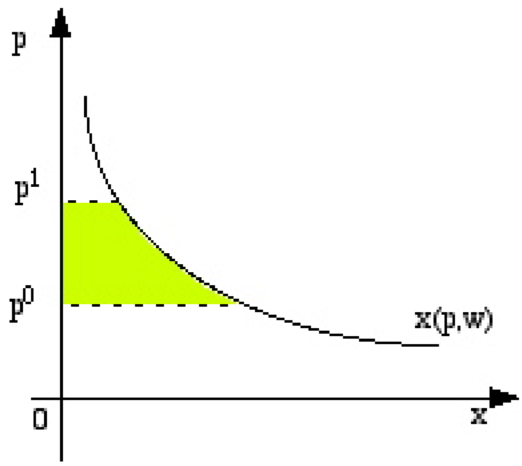

Figure 1.4: Consumer surplus

price is called consumer surplus. This surplus captures the rents the the consumer saves because the firm cannot set a price for every unit that the consumer buys.

In a similar way, we can compute the consumer surplus when the price of the good varies. Figure 1.4 illustrates the argument. Consider an initial situation where the price of the good is p^0. At this price, the consumer, given his (her) demand function, buys x^0 units. Assume that for some reason, the price increases to p^1 so that the consumer, whose income remains constant, reduces its demand to x^1. The colored area illustrates the variation of the surplus of our consumer and offers an idea of the impact of the price change of his (her) welfare.

Formally, the consumer surplus in the case of figure 1.4 is given by,

$$CS_i = \int_{p^0}^{p^1} x(t)dt$$

The consumer surplus coincides with the equivalent variation and the compensating variation when preferences are quasilinear. For any other preferences, the consumer surplus only offers an approximate measure bounded by the equivalent variation and the compensating variation. We will analyze both situations in turn.

Quasilinear utility.

Assume our consumer lives in a world of two goods with prices $p_1 = 1$ and p_2. Also, assume his (her) income is w_i and the utility function can be represented as

$$U_i(x_{i1}, x_{i2}) = x_{i1} + u_i(x_{i2}),$$

so that the utility function is linear in one good. Assume the function $u_i(x_{i2})$ is strictly concave.

The problem of the consumer is

$$\max_{x_{i1},x_{i2}} x_{i1} + u_i(x_{i2})$$
$$\text{s.a } x_{i1} + p_2 x_{i2} = w_i$$
$$x_{i1} \geq 0.$$

This problem may have two types of solution according to the consumption of good x_{i1} be positive or zero.

Consider first $x_{i1} > 0$. We can reformulate the problem as,

$$\max_{x_{i1},x_{i2}} x_{i1} + u_i(x_{i2})$$
$$\text{s.a } x_{i1} + p_2 x_{i2} = w_i$$

or also,

$$\max_{x_{i2}} w_i - p_2 x_{i2} + u_i(x_{i2})$$

The first order condition, $u_i'(x_{i2}) = p^2$, tells us that demand of good 2 only depends on its own price and is independent of income. In other words, we can write its demand as $x_{i2}(p_2)$. We obtain the demand of good 1 from the budget constraint $x_{i1} = w_i - p_2 x_{i2}(p_2)$.

When $x_{i1} = 0$, demand of good 2 is simply $x_{i2} = \dfrac{w_i}{p_2}$.

How does the consumer decide his (her) consumption plan? Given that the utility (sub)function on good 2 is strictly concave, the consumer will start consuming good 2 until the marginal utility of an additional euro spent in good 2 will be equal to $p_1 = 1$. From that point on, the increases in income will be devoted to consumption of good 1. For the sake of the argument, let us assume that initially our consumer has zero income and we increase it marginally. The increase in utility is $\dfrac{u_i^{'}(w_i/p_2)}{p_2}$. If this increase in utility is larger than 1 (the price of good 1), the consumer obtains more utility consuming good 2. This behavior will remain the same until the marginal increase in income make the marginal utility of that income spent in good 2 equal to the price of good 1. Then our consumer will be indifferent between consuming either good. From that point on, further increases in income will be devoted to increase the consumption of good 1.

The level of utility (welfare) obtained by the consumer is simply the sum of the utility derived from the consumption of every good, i.e.,

$$U_i(x_{i1}, x_{i2}) = w_i - p_2 x_{i2}(p_2) + u_i(x_{i2}(p_2)).$$

To illustrate this welfare level on the demand curve of good 2, we only need to integrate,

$$w_i - p^2 x_{i2}(p_2) + u_i(x_{i2}(p_2)) = w_i - p_2 x_{i2}(p_2) + \int_0^{x_{i2}} p(t)dt.$$

Leaving aside the constant w_i, the expression on the right hand side of this equation is the area under the demand curve of good 2 and above price p_2.

The general case

When the utility function representing the preferences of our consumer is not quasilinear, the consumer surplus can only offer an approximation to the welfare variation associated to a price change.

Recall that the equivalent variation and the compensating variation of the consumer when the price of a good varies from p^0 to p^1 (given the prices of the other goods and consumer income) are:

$$\begin{aligned} EV_i(p^0, p^1, w) =& e_i[p^0, v_i(p^1, w_i)] - e_i[p^0, v_i(p^0, w_i)] \\ CV_i(p^0, p^1, w) =& e_i[p^1, v_i(p^1, w_i)] - e_i[p^1, v_i(p^0, w_i)]. \end{aligned}$$

Also, recall that the compensated demand function is the derivative of the expenditure function, $h_i(p, u_i) \equiv \frac{\partial e_i}{\partial p}$, so that we can rewrite the equivalent variation and the compensating variation as,

$$\begin{aligned} EV_i(p^0, p^1, w) =& e_i[p^0, u_i^1] - e_i[p^0, u_i^0] = \int_{p^1}^{p^0} h_i(p, u^1)dp, \\ CV_i(p^0, p^1, w) =& e_i[p^1, u_0^i] - e_i[p^1, u_i^0] = \int_{p^1}^{p^0} h_i(p, u^0)dp, \end{aligned}$$

that is, the compensating variation is the integral of the compensated demand curve associated at the initial utility level while the equivalent variation is the corresponding integral associated at the final utility level.

The correct measure of welfare is thus an area given by a compensated demand function. The problem, as we already know, is that such demand is unobservable. This is the reason why the consumer surplus, obtained on the (observable) marshallian demand is often used as an approximation. The question that remains is how good this approximation is. To answer this question we start by recalling the Slutsky equation,

$$\frac{\partial x_{ij}}{\partial p_k} = \frac{\partial h_{ij}(p, u)}{\partial p_k} - \frac{\partial x_{ij}}{\partial w_i} x_{ik}(p, w_i).$$

When the good is not inferior, i.e. $\dfrac{\partial x_{ij}}{\partial w_i} > 0$, we obtain

$$\frac{\partial x_{ij}}{\partial p_k} < \frac{\partial h_{ij}(p, u)}{\partial p_k}.$$

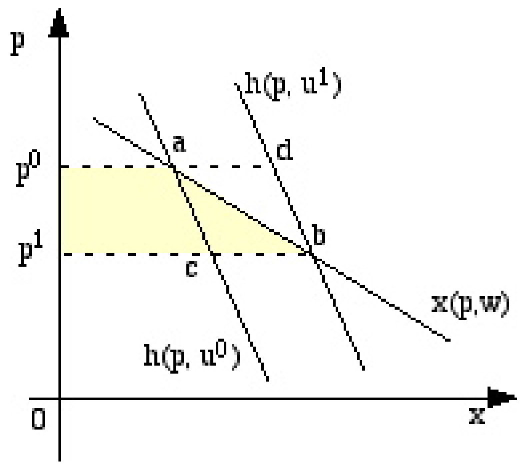

Figure 1.5: Consumer surplus as an approximation

that is, the slope of the compensated demand is larger that the slope of the marshallian demand. Figure 1.5 shows the relationship between the equivalent variation, the compensating variation, and the consumer surplus.

The initial situation is given by a price p^0, so that consumer is at a on the marshallian demand curve. The final scenario appears after a decrease in price to p^1, so that consumer is at point b on the marshallian demand curve.

The compensating variation is computed from the initial utility level, u^0, and is given by the area below the compensated (hicksian) demand at the point a, that is the area p^0acp^1.

The equivalent variation is computed from the final utility level, u^1, and is given by the area below the compensated (hicksian) demand at the point b, that is the area p^0dbp^1.

The consumer surplus is the area below the marshallian demand curve between points a and b, that is the area p^0abp^1.

Comparing these areas we realize that $CV \leq CS \leq EV$. In particular, if there are no income effects $\frac{\partial x_{ij}}{\partial w_i} = 0$, (this is the case of quasilinear utility) the three areas will coincide. This means that for small income effects, the consumer surplus represents a good approximation to the equivalent variation and to the compensating variation.

1.4 Producer surplus and deadweight welfare loss.

The producer surplus is the profit of the firm in the industry (net of fix costs). The next two figures show the marginal cost curve (i.e. the supply curve under perfect

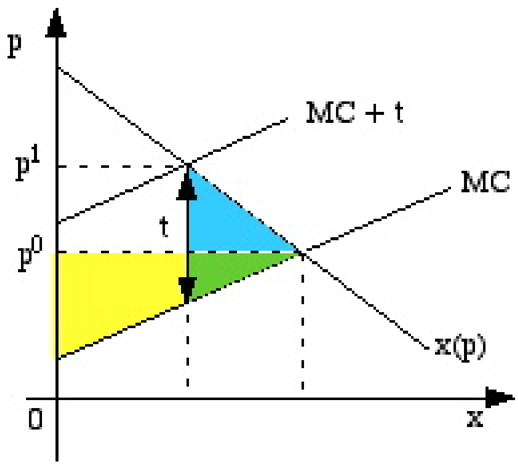

Figure 1.6: Taxation and deadweight welfare loss

competition). Profit is the difference between revenues (p^0x^0) and cost, where total cost is the integral of marginal cost (recall we are assuming away fixed/sunk costs). Accordingly, profit is the area between the marginal cost curve and the horizontal line at price p^0.

The aggregate welfare of the economy is the sum of the consumer surplus and the producer surplus. Looking at figures 1.6 and 1.7 it is easy to understand that total surplus is maximized when price equals marginal cost. Any deviation of the price away from the marginal cost represents a welfare loss. A monetary measure of this loss of welfare is the so-called “dead-weight loss”. Figure 1.6 shows the dead-weight loss associated with the introduction of a tax on a commodity. Assume an initial state where the economy is perfectly competitive, so that the equilibrium price is p^0. Then a unit tax $t$is imposed on each unit sold. This raises the equilibrium price to p^1 and lowers the equilibrium consumption to x^1. The welfare loss is thus the difference in total surplus between both situations. This is the area of the blue/green triangle. It is given by $t(x^0 - x^1)/2$.

The second example shows the dead-weight loss associated with the transition from a competitive economy to a monopolized economy. As before the initial situation is a perfectly competitive economy with an equilibrium price p^0. A monopoly would set a price equation marginal revenue and marginal cost, that is p^1. The dead-weight loss is the area of the blue/yellow triangle.

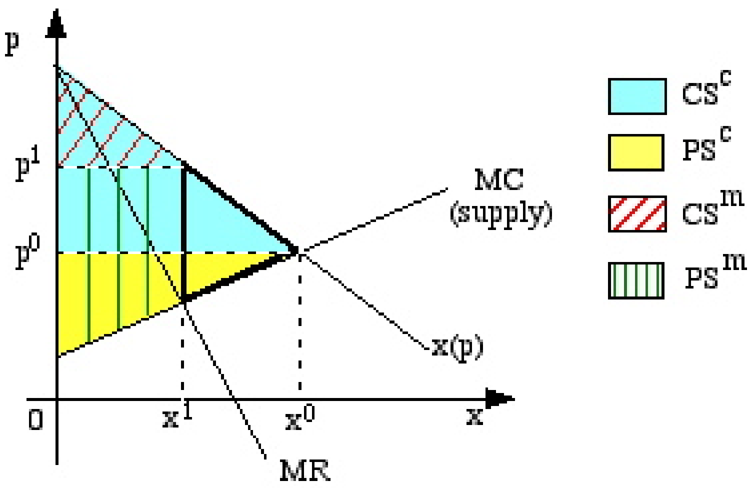

Figure 1.7: Monopoly and deadweight welfare loss

1.5 Market and market power.

1.5.1 Definition of the market.

Defining a market is not an easy task. It is obvious that we do not want to restrict ourselves to the case of a homogeneous good. A first approximation could be the principle that two goods belong to the same market only if they are perfect substitutes. That would be too a restrictive definition because it would have as consequence that there would be only one firm in each market. But very few firms have an absolute monopoly power. A common feature in real markets is that consumers after the increase of the price of one good react (partially) substituting that good by purchasing other alternative goods. The definition cannot be too general either. Considering any two substitute goods as belonging to the same market would lead to an economy with a single market since any good can be directly or indirectly a potential substitute of any other good. Such a definition would not allow partial equilibrium analysis.

At this point thus, we realize that the “correct” definition of a market has to be contingent to the problem we want to tackle. For instance, let us consider the case of coal. If the problem we face is the design of energy policy, the relevant market is the energy market (including coal, petrol, nuclear, . . .); to evaluate the effects of a merger between two providers, we would need a narrower definition of the relevant market.

An ideal scenario to define a product market consists in having a set of goods with very high cross price elasticities (in absolute value) among them and very low with respect other commodities not in the set. This definition refers to demand elasticities but also in a subsidiary way to supply elasticities. Let us consider some examples:

(i) cross elasticity between lead free 95 and 98 octanes gasoline is very high. They are two close substitutes that clearly should belong to the same market;

(ii) cross elasticity between consumption of gasoline and mineral water is very low. They are two independent products. They should not belong to the same market;

(iii) cross elasticity between shoes for the right foot and the left foot is (in absolute value) very high: they are two perfect complements that must belong to the same market (actually, in this extreme case the correct definition of the market should be that of pairs of shoes).

This rule involving elasticities even though contains clear ambiguities (it is not clear what does sufficiently high cross elasticity means) is not always easy to apply. First, product differentiation is a gradual phenomenon and therefore determining the critical value of the cross elasticity is not easily determined. Second, we should be aware that two products may be substitutes indirectly through a chain of substitution as it appears often in the pharmaceutical products: a certain drug is effective in treatments A and B; another one is effective in treatments A and C. Both the strict definition of market based on therapeutical effectiveness (A, B, C) as the more general definition $(A + B + C)$ are inconsistent. It is also worth mentioning that a market definition using a geographical criterion raises similar problems. For example, what is the proper definition of the market for wine: the world, Europe, Spain, Catalonia, Barcelona?

Despite all these difficulties, for statistical purposes there exist classification systems of the economic activities in every country. In Spain we have the so called “Clasificación Nacional de Actividades Económicas”, CNAE among others (see appendix). We can also mention the classification system NACE proposed by the Eurostat. These classifications are divided in sections from one digit, until four digits. Often these classification systems are used as proxies for market definitions. This is a problem because usually the grouping in these systems are done using criteria from the supply side of the market, while the definition of the market emphasizes the demand side of the market. For example, production of wine and production of cava belong to different groups (because the technology is different) but from the demand point of view should be considered as belonging to the same market.

The supply side aspects in the definition of a market offers some advantages from the industrial organization point of view. A well known example is provided by McKie (1985) (see Cabral (1994), p.21): In 1964 the US Air force opened a contract for the provision of a certain type of radar. The contract was assigned to *Bendix*, enterprise that maintained a monopoly status for some years. This situation led a second firm, *Wilcox*, to sue *Bendix* for abuse of monopoly position. The Federal Trade Commission voted in favor of *Bendix*. The reason was that from the demand point of view, looking at the elasticity of demand, *Bendix* can be considered a monopoly. Nevertheless, according to the classification of industrial activities we find a certain number of firms with similar technological capacity to *Bendix*. Therefore, any of them could win the next contract when it would become public. Actually, this is what happened in 1969 when *Honeywell* obtained the new contract.

1.5.2 Concentration measures.

Once we have defined the market (the industry) both from the cross elasticities and from the supply side perspectives, we also need a measure of the relative importance of every firm in the market, and a statistical method to compute an index giving information on the degree of concentration of the market. Even though there are obvious difficulties to agree upon a criterion to measure the relative size of a firm (see e.g. Hay and Morris (1996) or Eraso Goicoechea and Garcia Olaverri (1990)), one possibility is to use the market shares.

Consider an industry with n firms producing a homogeneous good. The distribution of these firms according to their market share is given by a vector m ordered from biggest to smallest:

$$m = (m_1, m_2, \cdots, m_n), \quad m_i = \frac{q_i}{\sum_{i=1}^n q_i} \geq 0, \quad \sum_{i=1}^n m_i = 1.$$

That is, we are working in a unit simplex S_{n-1} in $\mathbb{R}^n$. For every number n a concentration measure is a real application C_n defined on S_{n-1}.

Following Encaoua and Jacquemin (1980), any proper concentration measure has to satisfy some requirements and properties.

Requirements

- Unidimensionality.
- The concentration measure must take values in $[0, 1]$.
- Independence of the market size.
- Symmetry: the concentration measure has to be invariant to permutations of the market shares of the firms, that is,

$$C_n(m_1, m_2, \cdots, m_n) = C_n(m_{\pi(1)}, m_{\pi(2)}, \cdots, m_{\pi(n)})$$

for any permutation π de $\{1, 2, \cdots, n\}$

Properties
n fix

- P1 (Transfer principle): Transferring part of the production from a smaller firm to a bigger one cannot decrease the concentration measure, i.e.

  $$C_n(m_1, \ldots, m_j, \cdots, m_k, \cdots, m_n) \leq C_n(m_1, \ldots, \widetilde{m}_j, \cdots, \widetilde{m}_k, \cdots, m_n),$$

  where $m_j > \widetilde{m}_j, \widetilde{m}_k > m_k$, and $m_j - \widetilde{m}_j = \widetilde{m}_k - m_k$.

- P2 (Homogeneity principle): Given the number of firms in the industry, the concentration measure takes its minimum value when all firms have the same market share, i.e. $\min C_n(m_1, m_2, \cdots, m_n) = C_n(m, m, \overset{\text{n times}}{\cdots\cdots}, m)$.

- P3 (Lorenz criterion): consider two industries with the same number of firms. Let the aggregate production of the k, $(k = 1, 2, \cdots n)$ bigger firms in the first industry be larger than or equal to the corresponding aggregate production of the k bigger firms in the second industry. Then, this inequality must also hold between the concentration measures of the two industries, Formally, $C_n(m_1^1, m_2^1, \cdots, m_k^1) \geq C_n(m_1^2, m_2^2, \cdots, m_k^2)$ when $\sum_{i=1}^k m_i^1 \geq \sum_{i=1}^k m_i^2$.

n variable

- P4: If two or more firms merge, the concentration measure cannot diminish. Formally, $C_n(m_1, \cdots, m_j, \cdots, m_k, \cdots, m_n) \leq C_{n-l}(m_1, \ldots, m_{jl}, \cdots, m_n)$

- P5: If in an industry all firms' market shares are equal ($m_i = m_j, \quad i \neq j, \quad i, j = 1, 2, \cdots, n$), the concentration measure cannot be increasing in the number of firms, i.e. $C_n(m, \cdots, m) \leq C_{n+l}(m, \cdots, m)$

These properties are not independent. Encaoua and Jacquemin show the following implications:

- If P1 holds, then P2 i P3 also hold ($P1 \Rightarrow P2, P1 \Rightarrow P3$)

- If P2 and P4 hold, then P5 also holds ($P2 \wedge P4 \Rightarrow P5$)

Definition 1.4 (Concentration measure)**.** *Let $h(m_i)$ be a function defined in $[0, 1]$ assigning a weight to the relative production of a firm i, i.e. $m_i \to m_i h(m_i)$. A concentration measure is defined as,*

$$C_n(m_1, m_2, \cdots, m_n) = \sum_{i=1}^n m_i h(m_i)$$

We can recall now the most usual concentration indices.

- Concentration ratio:
$$C_k = \sum_{i=1}^{k} m_i.$$
This is a measure that considers $h(m_i) = 1$. This index measures the relative production of the k largest firms in the market over the total production of the industry. The index takes values in the interval $C_k \in [\frac{k}{n}, 1]$. The problem of this index is the arbitrariness of k.

- Herfindhal index:
$$C_H = \sum_{i=1}^{n} m_i^2.$$
This index defines $h(m_i) = m_i$. Therefore, it is a measure that overestimates the relative importance of the large firms against the smaller ones. The index takes values in the range $C_H \in [\frac{1}{n}, 1]$. Also, the number $\frac{1}{C_H}$ is called "Adelman's equal number" associated to the Herfindhal index. It represents the number of equal sized firms whose distribution results in the same concentration measure as the one given by C_H. Note that the Herfindhal index is defined over all firms in the industry.

- Entropy index:
$$C_E = \sum_{i=1}^{n} m_i \log_a m_i, \quad a > 1.$$
This index defines $h(m_i) = \log_a m_i$. Therefore, it is a measure that underestimates the relative importance of the large firms against the smaller ones. The index takes values in the range $C_E \in [\log_a \frac{1}{n}, 0]$. Also, the number $\frac{1}{a^{C_E}}$ is called the "equal number" associated to the entropy index. It represents the number of equal sized firms whose distribution results in the same concentration measure as the one given by C_E.

1.5.3 Degree of Monopoly.

The most popular measure of monopoly power of a firm was proposed by Lerner (1934) and thus called Lerner's index:

$$L_i = \frac{P - MC_i}{P} = 1 - \frac{MC}{P}$$

The index takes values in the range $L_i \in [0, 1)$. When a firm behaves competitively, its price equates its marginal cost, $L_i = 0$. As the firm increases its ability to set a price above marginal cost the index increases. In the limit, when the margin of the price over the marginal cost is infinitely large, $L_i \to 1$.

From the individual Lerner indices in an industry, we can obtain an **aggregate index of monopoly power**. Let $\Im_n(L_1, L_2, \cdots, L_n)$ be the aggregate index of monopoly power in an industry with n firms. This index has to satisfy three properties:

1.- The value of $\Im_n(L_1, L_2, \cdots, L_n)$ must lay in the range defined by the extreme values of the distribution of individual Lerner indices $(L_1, L_2, \cdots, L_n)$ i.e. $\max\{L_1, L_2, \cdots, L_n\} \geq \Im_n(L_1, L_2, \cdots, L_n) \geq \min\{L_1, L_2, \cdots, L_n\}$. If $L_1 = L_2 = \cdots = L_n = L$, then $\Im_n(L_1, L_2, \cdots, L_n) = L$. This result has two interpretations. Either the industry is perfectly competitive or it is a perfect cartel. In the latter case, the optimal assignment of market shares is such that the marginal costs of the different firms coincide; even if market shares are not equal, the individual Lerner indices coincide. Hence, it is important to make the distinction between the distribution of market shares and the distribution of monopoly power in an industry.

2.- In an industry, some of its members may be "price takers" so that their monopoly power are nil, i.e. $L_i = 0$. Even though, these components must also be included in $\Im_n(L_1, L_2, \cdots, L_n)$, because each member of the industry has its weight in the computation of the aggregate index. In other words, if $L_n = 0$, $\Im_n(L_1, L_2, \cdots, 0) \neq \Im_{n-1}(L_1, L_2, \cdots, L_{n-1})$.

3.- If two or more firms merge, the aggregate index of monopoly power must not decrease, i.e. $\Im_n(L_1, L_2, \cdots, L_n) \leq \Im_{n-1}(L_{1,2}, L_3, \cdots, L_n)$.

Some indices satisfying these properties are:

- Aggregate concentration ratio:

$$\Im_k = \frac{1}{n}\sum_{i=1}^{k} L_i$$

- Aggregate Lerner index

$$\Im_a = \sum_{i=1}^{n} m_i L_i$$

  This is an arithmetic average so that large firms get more weight.

- Aggregate entropy index

$$\Im_g = \prod_{i=1}^{n} (L_i)^{m_i}, \quad L_i \neq 0$$

  This is a geometric average so that large firms get less weight.

1.5.4 Concentration indices and degree of monopoly.

Consider a homogeneous industry with n firms. Market (inverse) demand is given by $p = f(q)$ where $q = \sum_{i=1}^{n} q_i$. Every firm has a technology described by $C_i(q_i)$ and a production capacity limit v_i. Firm i's profits are,

$$\pi_i(q_1, q_2, \cdots, q_n) \ = \ q_i f(q) - C_i(q_i),$$

where $q_i \in [0, v_i]$. Assuming the proper conditions on demand and cost functions, a noncooperative Cournot equilibrium[4] will be interior and unique, i.e. $q_i^0 \in (0, v_i), \ \forall i$, so that the system of first order conditions of the profit maximization problem evaluated at the equilibrium satisfies,

$$f(q^0) + q_i^0 f'(q^0) - C_i'(q_i^0) \ \equiv \ 0 \quad (i = 1, 2, \cdots, n),$$

where $q^0 = \sum_{i=1}^{n} q_i^0$. Rearranging terms, we can write,

$$f(q^0) - C_i'(q_i^0) \ = \ -q_i^0 f'(q^0) \quad (i = 1, 2, \cdots, n).$$

Dividing both sides by $f(q^0)$ and multiplying and dividing the right hand side q^0, we obtain,

$$\frac{f(q^0) - C_i'(q_i^0)}{f(q^0)} \ = \ -\frac{f'(q^0)}{f(q^0)} q^0 \frac{q_i^0}{q^0} \quad (i = 1, 2, \cdots, n).$$

The left hand side of this expression is firm i's Lerner index evaluated at the equilibrium output vector. The right hand side contains two elements:

- $\frac{q_i^0}{q^0} \ \equiv \ m_i^0$ i.e. firm i's equilibrium market share, and
- $$-\frac{f'(q^0)}{f(q^0)} q^0 \ = \ -f'(q^0) \frac{q^0}{f(q^0)} \ = \ -\frac{\partial f(q^0)}{\partial q^0} \frac{q^0}{f(q^0)} \ = \ \frac{1}{\varepsilon}.$$

Therefore, we can write the first order condition evaluated at the equilibrium production plan as

$$L_i^0 \ = \ \frac{1}{\varepsilon} m_i^0. \tag{1.2}$$

Now we can show the (direct) relationship between a concentration index and an adequate aggregate index of monopoly power:

[4]see chapter 3 on the concept of noncooperative equilibrium.

- Concentration ratio

$$\Im_k = \frac{1}{n}\sum_{i=1}^{k} L_i = \frac{1}{n}\frac{1}{\varepsilon}\sum_{i=1}^{k} m_i = \frac{1}{n}\frac{1}{\varepsilon}C_k;$$

- Herfindhal index

$$\Im_a = \sum_{i=1}^{n} m_i L_i = \frac{1}{\varepsilon}\sum_{i=1}^{n} m_i^2 = \frac{1}{\varepsilon}C_H;$$

- Entropy index [5]

$$\Im_g = \prod_{i=1}^{n} L_i^{m_i} = \prod_{i=1}^{n} \frac{1}{\varepsilon} m_i^{m_i} = a^{C_E}\frac{1}{\varepsilon}.$$

All these indices are reasonable in the sense that they satisfy the axioms proposed by Encauoa and Jacquemin (1980), but nothing is said on their usefulness. Are they a useful instrument for policy design? To answer we can think of relating these indices with the productivity of the industry. In particular, we can examine the relationship between concentration and industry profits.

Start assuming all firms in the industry with the same market shares (symmetry). The only reasonable concentration measures are equivalent to the number of firms in the industry: $C_k = \frac{k}{n}$, $C_H = \frac{1}{n}$, $C_E = \ln\frac{1}{n}$. Bertrand's model tells us that market price and industry profits are independent of the number on firms in the industry. Thus, concentration and profitability are not related. Cournot's model (with fixed number of firms) shows a negative correlation between concentration and profitability.

If firms are *asymmetric*, because they may have different costs for instance, the concentration measure is not ambiguous any more. Assume, to illustrate, that firms have constant marginal costs, $C_i(q_i) = c_i q_i$, and compete in quantities. The aggregate industry profits are:

$$\Pi = \sum_{i=1}^{n}\Pi_i = \sum_{i=1}^{n}(f(q)-c_i)q_i = \sum_{i=1}^{n}\frac{f(q)m_i q_i}{\varepsilon} = \frac{f(q)q}{\varepsilon}\Big(\sum_{i=1}^{n} m_i^2\Big) = \frac{f(q)q}{\varepsilon}C_H,$$

5

$$C_E = \sum_{i=1}^{n} m_i \log_a m_i = \sum_{i=1}^{n}\log_a m_i^{m_i} = \log_a \prod_{i=1}^{n} m_i^{m_i};$$

taking antilogs

$$a^{C_E} = \prod_{i=1}^{n} m_i^{m_i}$$

where we have used (1.2). Assume also that consumers spend a fix proportion of their incomes in this market, that is, $pq = k$ where k is a positive constant. Then, demand elasticity $\varepsilon = 1$, and the above expression reduces to $\Pi = kC_H$. In this particular case, the Herfindhal index gives an exact measure (up to a proportionality constant) of industry profitability.

The asymmetries between firms tend to generate high correlation between concentration indices and industry profitability. This is so because asymmetries in costs generate asymmetries in production levels and thus increase the concentration indices. Also, the most efficient firms in the industry obtain rents increasing the global industry profit. For instance, under Bertrand competition with constant marginal cost the firm with the lowest marginal cost gets all the market (so that concentration index will be highest) and obtains positive profits. When firms are more symmetric concentrations indices usually are not so high and firms obtain small profits. This phenomenon also appears under Cournot competition.

Summarizing, concentration indices are useful because they are easy to compute and give an economic interpretation of how competitive an industry is. Unfortunately, there is no systematic relation between these indices and the relevant economic variables to evaluate changes in technology, demand, or economic policies. Even if it would be possible, we should be aware that the indices are endogenous measures, so that correlations could not be interpreted in a causal sense.

1.5.5 Volatility measures

A limitation of the concentration measures is their static character. The introduction of dynamic considerations in the analysis of market structure leads to the so-called volatility measures.

The degree of competitiveness in a market is related not only with the distribution of market shares, but also with the relative position of the firms as time goes by. Assume that a certain market has at any point in time a dominant firm, but that firm varies in the different periods. It is quite likely that this market approaches more a competitive behavior than another market with less concentration but with more stable position of the firms in the market.

One of the most popular measures of dynamic competitiveness (or stability of market shares) is the *instability index*. It is defined as,

$$I \equiv \frac{1}{2}\sum_{i=1}^{n} |m_{i2} - m_{i1}| \in [0, 1),$$

where m_{i2} and m_{i1} represent firm i's market shares in periods 1 and 2 respectively, and n denotes the number of firms in any period. It is easy to verify that I varies between 0 (minimum instability) and 1 (maximum instability). When market shares remain constant along time we obtain $I = 0$. The situation where $I = 1$ corresponds to a scenario where all firms present in the market in the initial period have zero market share in the next period (i.e. none remains in the market). Naturally, this instability index also presents some problems of measurement and interpretation.

Exercises

1. Show

   a) $TPI_L \leq PI_L$ and $TPI_P \leq PI_P$.

   b) when preferences are homothetic, then $TPI_L = PI_L$ and $TPI_P = PI_P$.

2. Consider a quasilinear utility function $U_i(x_{i1}, x_{i2}) = x_{i1} + u_i(x_{i2})$ where $u_i(x_{i2})$ is strictly concave. Show that $CV = CS = EV$.

3. Show that $C_H = \dfrac{1}{n} + nV(m_i)$ where C_H denotes the Herfidahl index, n the number of firms and $V(m_i)$ the variance of market shares. From the equation above, provide an interpretation to the Adelman number defined as C_H^{-1}.

4. Compute the extreme values of the (i) concentration ratio, (b) Herfindhal index, and (iii) entropy index.

5. Consider a n-firm Cournot oligopoly where each firm has a constant marginal cost c_i. Market demand is described by a well-behaved function $p = f(Q)$. Show that the ratio between industry profits and industry revenues equals the ratio between the Herfindhal index and the elasticity of demand.

6. Show that under the conditions of problem 2, the "average Lerner index" $(\sum_i m_i L_i)$ equals the Herfindhal index divided by the demand elasticity.

7. Consider a market with linear demand $Q = 1 - P$. Two firms operate with constant marginal costs, c_1 and c_2 such that $c_1 + c_2 = 2c$ where c is a constant. Show that when firms become more asymmetric (i.e. c_i moves away from c) Cournot competiton yields a higher concentration index and a higher industry profit.

8. Consider a market of a certain (homogeneous) product described by the list of firms with market shares above 2% as shown in the following table:

Firm	Share	Firm	Share
1	14.19	8	2.60
2	12.71	9	2.54
3	11.02	10	2.50
4	10.56	11	2.14
5	9.50	12	2.10
6	7.92	13	2.03
7	3.00		

These 13 firms together cover 82.81% of the market. Compute the upper and lower bounds of the Herfindhal index for this market.