



Dispensa BIE - Prof. Luca Grilli

Business & industrial economics (Politecnico di Milano)



Scansiona per aprire su Studocu

BUSINESS & INDUSTRIAL ECONOMICS

PROF. LUCA GRILLI

POLITECNICO DI MILANO

A.A. 2018/2019

Business and industrial economics

EFFICIENCY, COORDINATION, ECONOMIC ORGANIZATION

The **economic organizations** are entities created by individuals and evaluated on the base of their capacity to reach goals. These **goals** are not only economic: they could be related to the capacity of satisfying the economic needs of participants such as the consumption of goods. The individuals have to know their **preferences**, which mean needs and priorities. We can describe their satisfaction through a **utility function** and the goal is to maximize it.

We live in a context of **scarcity of resources**: if we choose to consume more one good, we will have to reduce the consumption of another good. Moreover, the satisfaction of the utility function of an individual is at the expense of the others.

Economy as a whole is the highest-level organization; it contains the lower-levels such as **markets and firms**. There are other formal economic organizations such as labor unions, government agencies, universities and so on. They are independent legal entities: they can enter binding contracts and providing enforcement of those contracts in their own name.

The place where individuals and firms interact are:

- **Product market**: households express a demand of goods and producers supply these goods.
- **Market for inputs**: producers demand labor and components.

These markets are clearly interrelated: a decision in the product market has an impact on the market for inputs.

Pareto efficiency

An allocation of resources A is inefficient if there is some other available allocation B that everyone concerned likes at least as A and that one person strictly prefers. In such a case A is Pareto dominated by B (B is Pareto superior to A) and it is clearly wasteful from a society point of view. Otherwise A is Pareto efficient (or Pareto optimal).

This is a way to **evaluate different choices** leading to different allocation of resources and products. The Pareto efficient depends on the **group of people** and **set of available resources** considered. If there are more people and resources, the Pareto efficient changes.

Giving all resources to a single insatiable and completely selfish individual would be Pareto efficient, even if there were **not ethical considerations**.

Given a set of resources, there are many efficient allocations. Therefore, the Pareto efficiency has a **normative power**: it has the positive power to predict the outcomes of allocations but with negative implications from an ethical point of view.

The efficiency principle: If people are able to bargain together effectively and can effectively implement and enforce their decision, then the outcomes of economic activity will tend to be

efficient (at least for the parties of the bargain). If people recognize an inefficient allocation, **they can migrate to a Pareto efficient** position through the coordination without hurting anybody.

Indeed, since **efficient choices and allocations are less vulnerable**, we should expect inefficient arrangements being supplanted over time, while efficient ones survive.

The Edgeworth box: if we consider two consumers, A and B, and their endowment of goods 1 and 2, which are: $\omega^A = (\omega_1^A; \omega_2^A)$ and $\omega^B = (\omega_1^B; \omega_2^B)$.

Example: $\omega^A = (6, 4)$ and $\omega^B = (2, 2)$

The total quantities available are:

$$\omega_1^A + \omega_1^B = 6 + 2 = 8$$

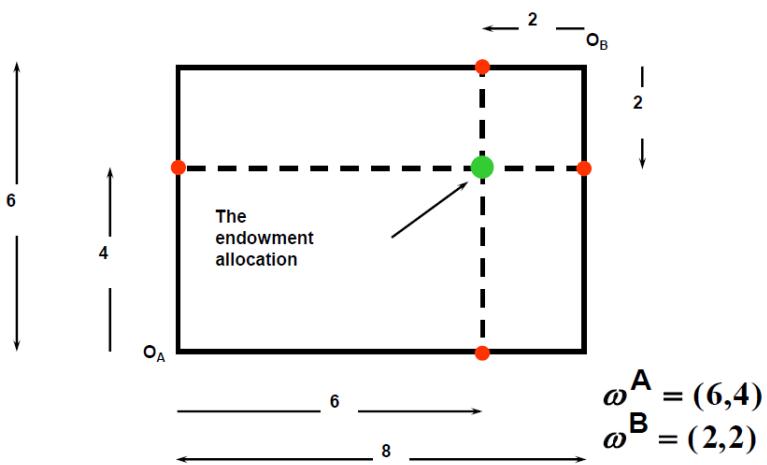
$$\omega_2^A + \omega_2^B = 4 + 2 = 6$$

$$\omega_2^A + \omega_2^B$$

The dimensions of the box are the quantities available of goods.

$$\omega_1^A + \omega_1^B$$

The box includes **all the feasible allocations of goods** between the two consumers.



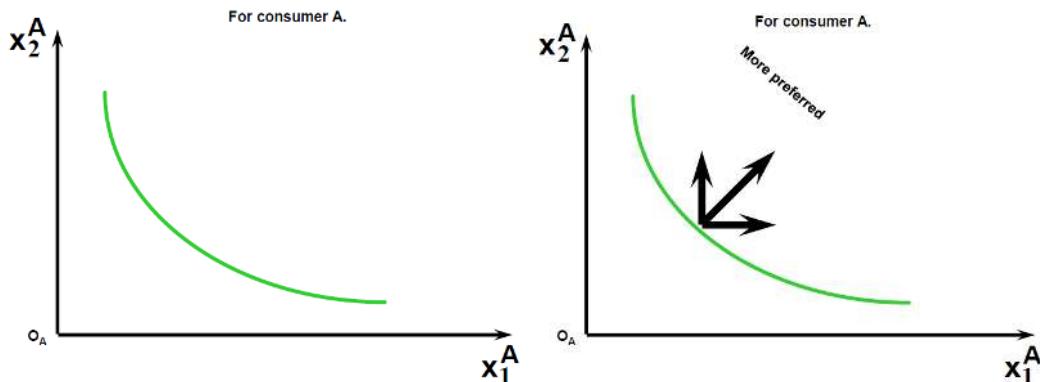
The allocations to consumers A and B are $(X_1^A; X_2^A)$ and $(X_1^B; X_2^B)$. An allocation is feasible only if:

$$X_1^A + X_1^B \leq \omega_1^A + \omega_1^B$$

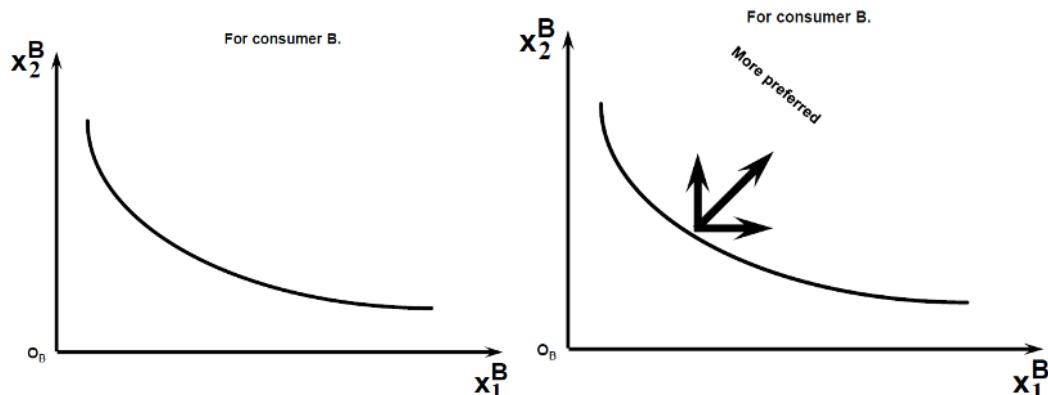
$$X_2^A + X_2^B \leq \omega_2^A + \omega_2^B$$

There is a number of feasible allocations. One or both **consumers will block an allocation**. The consumers prefer to move to an allocation rather than another on the base of their **preferences**.

We can add the preferences to the Edgeworth box, drawing the **indifference curve**, which represents the allocations of goods that give the same level of satisfaction to the consumer. For consumer A, the indifference curve will be the following:

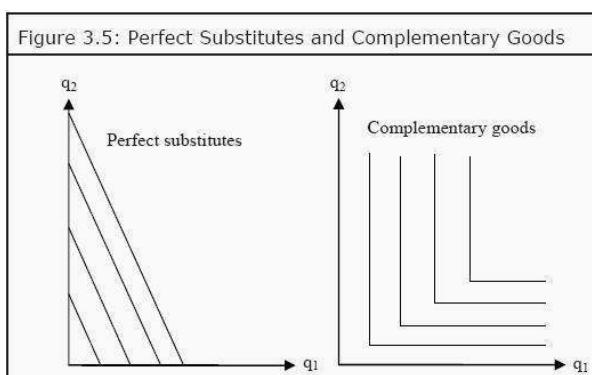


For consumer B, it will be:



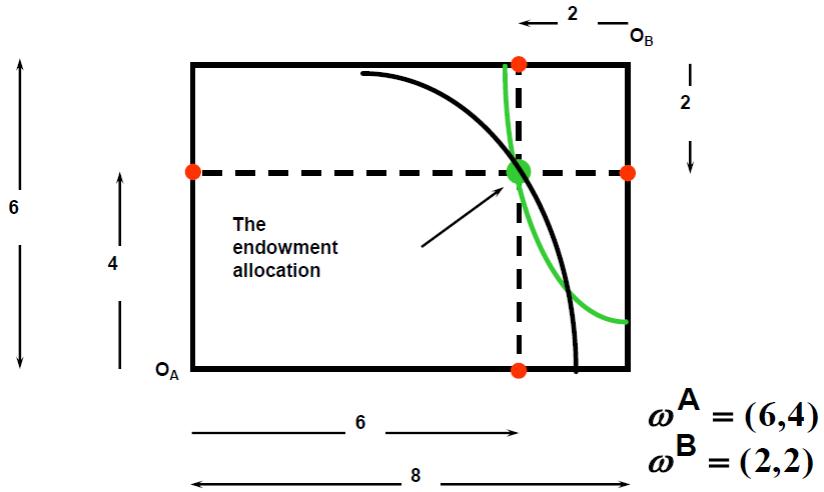
The indifference curves have a **negative slope** because if I want to maintain the same level of utility increasing the consumption of one good, I have to reduce the consumption of the other. The indifference curves of a consumer never cross each other.

Notes: we can observe the indifference curves of the perfect substitutes and complementary goods. In the case of perfect substitutes, the extreme quantities of the goods (on the axes) give the same level of utility to the consumer. In the case of complementary goods, if the consumer buys one of the two goods, he must buy the other (for example shoes).



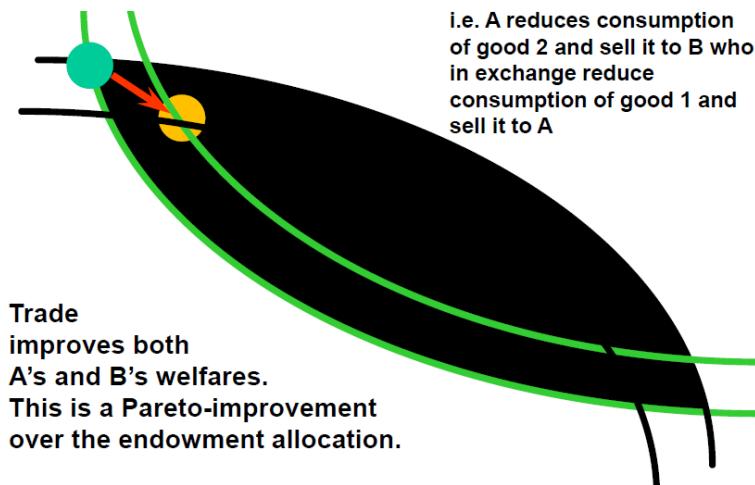
We always assume **well-behaved indifference curves** with monotonic preferences. This means that the more consumption provide a better utility. Furthermore, we assume **convex preferences**, which means consumers prefer to have a balanced mix of goods rather than the extreme quantities.

Mixing the indifference curves of the two consumers and adding them to the Edgeworth box, we find the following graph:

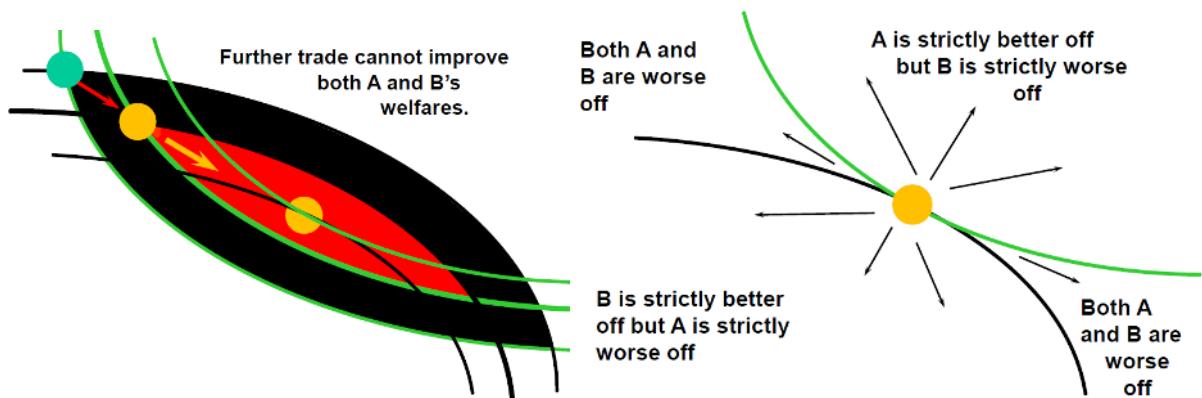


Within the area between the two curves, both consumers may find allocations that give them a better level of utilities. Therefore, an allocation that improves the welfare of a consumer without reducing the welfare of the other is a **Pareto-improving allocation**.

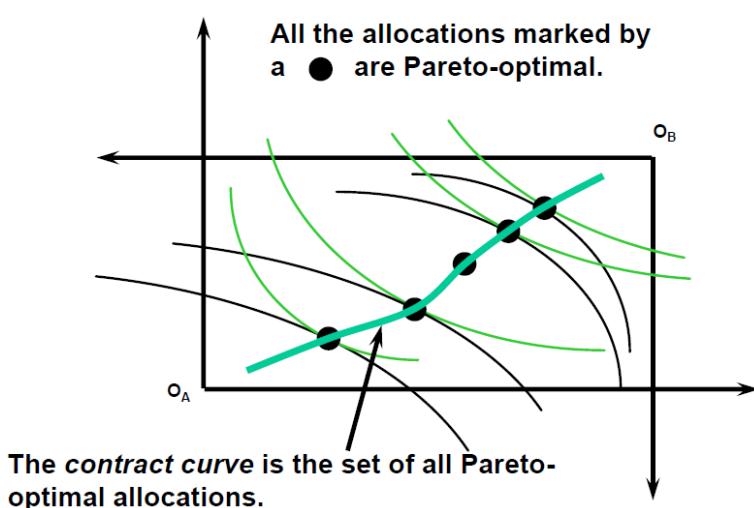
Trades between the two customers allow them to move to a better allocation.



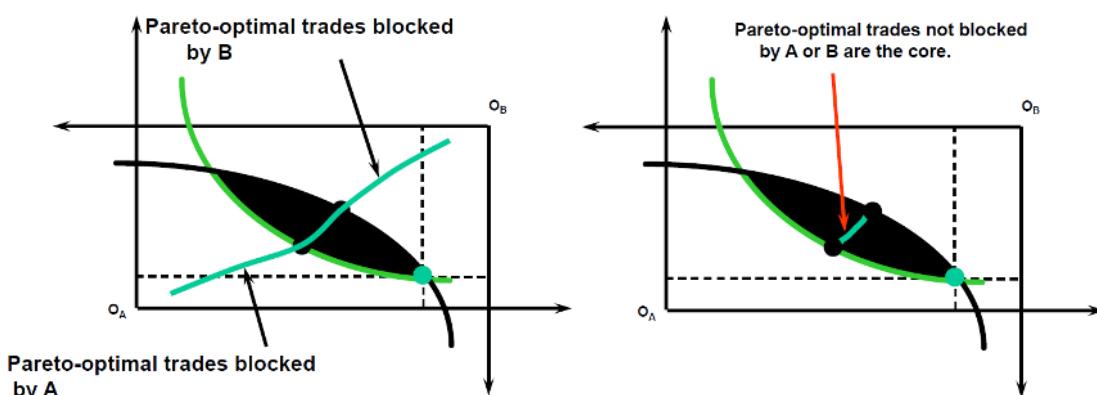
These trades end when another trade cannot improve the utility of one or both the consumers. This happens when the curves of the two consumers are tangent.



This allocation is **Pareto optimal** because the only way one consumer's welfare can increase is to decrease the welfare of the other.



There are many points in the box where the curve of one consumer is tangent to the curve of the other: all these points form the **contract curve**. If there is an initial allocation of resources, there is **limited set of allocations reachable** in the contract curve. Mixing the contract curve and the indifference curve within the box, we can find the **Pareto efficient allocations not blocked by a consumer**. These Pareto optimal allocations, called the **Core**, are welfare improving for both consumers relatively to their own endowments.



Rational trades should achieve a core allocation. The ultimate trade should be into the core. Therefore, trades can increase the **welfare of a community**. Moreover, they allow **specialization** with positive impacts on productivity and production. Through specialization and coordination,

people produce more and transact in order to acquire goods or services they desire. There are more resources available. There are ***four steps in trading***, one step requiring the following:

- 1) ***Productivity*** that means producing more outputs with the same inputs.
- 2) ***Specialization***: focusing on one task, we become a master of this task increasing productivity.
- 3) ***Coordination***: every task requires a complementary task on which someone else is specialized. People can produce more if they specialize and cooperate.
- 4) ***Information***: producers have to gather relevant information for instance about the complementary tasks. There are two ways to collect information:
 - ***Centralized planning***: it can be applied to the highest level of economic organization with a centralized decider (communism).
 - ***Autonomous decentralized decisions***: each individual takes choices to maximize its utility. If coordinated, these choices lead to the optimal allocation (capitalism).

Competitive markets

In ***competitive markets***, it is easier to achieve an efficient allocation because ***price*** is a vehicle of information about scarcity of resources and may involve different behaviors of economic agents. For example, if price rises, it means the offer is much lower than demand. If every relevant good is traded in a market at publicly known prices (i.e. if there is a complete set of markets), and if households and firms act perfectly competitive (i.e. as price takers), then the market outcome is ***Pareto optimal***. Therefore, when markets are complete, any ***competitive equilibrium is necessarily Pareto optimal***. ***Competitive markets are efficient***: there are no market failures and it is possible to reach Pareto efficiency. There are some assumptions:

- ***No market power***: firms are price takers.
- ***Complete information***: consumers know in advance the quality of goods.
- ***No externalities***

The ***competitive market mechanism*** always achieves an efficient allocation but it is not the only mechanism able to achieve it. Nevertheless, this mechanism is a ***very simple*** one: every individual simply maximizes its own utility while only knowing its own preferences and the market prices. Other ***allocation mechanisms*** require much more information, especially in a large economy with several markets and agents.

Competitive markets are an example of ***perfect competition*** in every market: every consumer will consume a quantity of any good n produced in the economy until the marginal utility he/she gets from the consumption equals the price he/she pays for obtaining it ($MU_n = P_n$). Furthermore, any good n produced is traded at a price equal to its marginal cost ($MC_n = P_n$).

Each individual is maximizing his/her utility by acquiring/selling products/services at the minimum price possible given costs (***allocative efficiency***), and products/services are produced at the minimum cost possible given the price of the factors (***productive and technical efficiency***). No Pareto improvements are possible in the economic system by moving resources from one production to another one.

$$\frac{MU_1}{MU_2} = \frac{P_1}{P_2} = \frac{MC_1}{MC_2}$$

The perfect competition is the ***best social solution*** and a point of reference in order to maximize the social welfare. Nowadays there are ***hybrid solutions*** between the two extreme: the first extreme refers to transactions between separate individuals and the second to the elimination of the price system with a regime of central planning within a single organization. These hybrid solutions involve the ***interaction among firms*** through the markets but within which activities are explicitly coordinated by plans and hierarchical structures. They are an effective mechanism for achieving coordination.

Competitive markets are the ultimate goal but there are three main ***market imperfections*** not allowing the perfect competition:

- Market power
- Externalities
- Asymmetric information

THEORY OF PRODUCTION

In economics, resources are limited. Companies and economic agents have to face a trade-off between:

- What they aspire to do: profit maximization.
- What they can do: technology constraints.

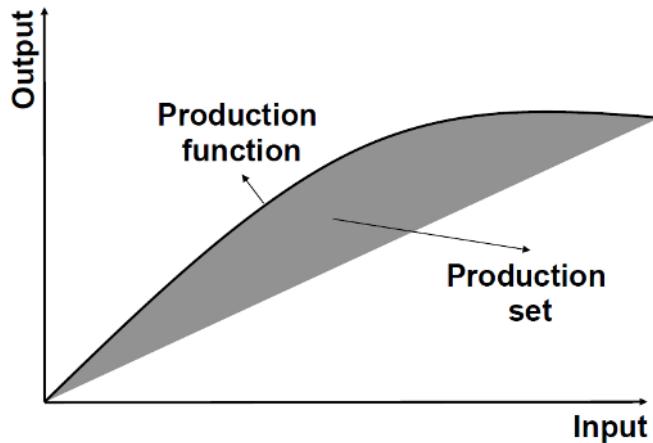
The ***technology*** is the process that receives some inputs (labor and materials) and returns some outputs. The technology constraints refer to the feasibility of certain combinations of inputs and outputs. Given a certain input, it is possible to achieve only a certain output level. This is valid also for customers having income constraints.

The ***production set*** is a combination of inputs and outputs achievable through a certain technology. We can represent it on the Cartesian plane: if we suppose the existence of one single input x and one single output y , the production set is given by (x,y) . Given x input units, the technology allows to produce only y output units.

The production function $f(x)$ represents the ***maximum output vector achievable given a certain input vector***.

$$f : R^N \rightarrow R^N \quad MAX \{Y\} = f(X)$$

It represents the ***efficiency frontier of the production set***. Given one single input and one single output, the representation of the production function is the following:



Analyzing the production function, we can consider two types of input: labor (K) and capital (C). The production function will be:

$$y=f(L, K)$$

We can define the ***labor marginal productivity*** as the output variation given a variation of one labor unit (maintaining the same amount of capital).

$$MP_L = \frac{f(L + \Delta L, K) - f(L, K)}{\Delta L}$$

Given infinitesimal variations of the labor:

$$MP_L = \lim_{\Delta L \rightarrow 0} \frac{f(L + \Delta L, K) - f(L, K)}{\Delta L} = \frac{\partial f(L, K)}{\partial L}$$

The derivative of the production function for L attains the labor marginal productivity.

Furthermore, we can define the ***capital marginal productivity*** as the output variation given a variation of one capital unit (maintaining the same amount of labor). We can follow the same steps to calculate it:

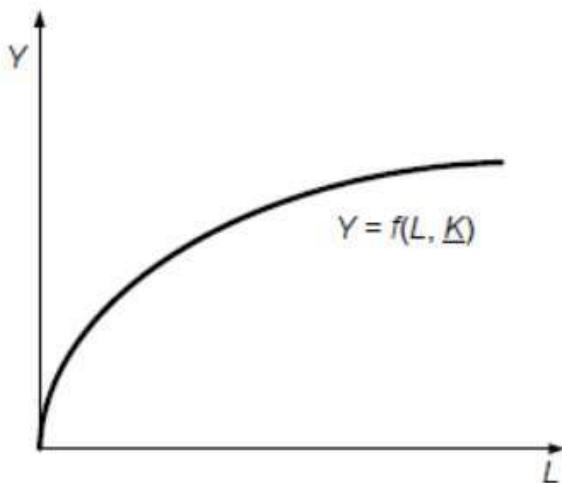
$$MP_K = \lim_{\Delta K \rightarrow 0} \frac{f(L, K + \Delta K) - f(L, K)}{\Delta K} = \frac{\partial f(L, K)}{\partial K}$$

If the production increases, it is not necessary that productivity increases.

The ***returns to scale*** characterize the ***correlation between output and input variation***:

- ***Constant returns to scale***: the output changes proportionally with the input.
- ***Increasing returns to scale***: the output changes more than proportionally with the input.
- ***Decreasing returns to scale***: the output changes less than proportionally with the input. It means that keeping constant all the others production factors, beyond a certain production level, additional units of an input cause a decreasing marginal productivity (for example too many people working on the same task).

Over the **short run**, it is possible to change the quantity of **labor only** (for example the recruitment of new employees). The following graph represents the production function in the short run keeping the capital constant: the production increases but at a lower rate.



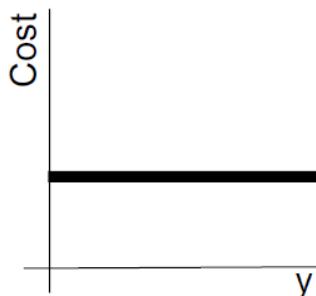
Over the **long run**, it is possible to change the quantities of **both the inputs**, labor and capital (for example the new production plants).

We can define the **cost function** that represents the minimum cost of production to produce the outcome y . It shows the total cost of inputs the firm needs to pay to produce output y .

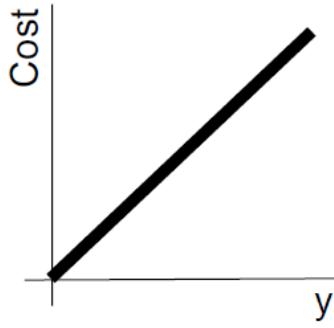
$$C(y) = f(y)$$

We can analyze the cost taxonomy; there are several types of cost:

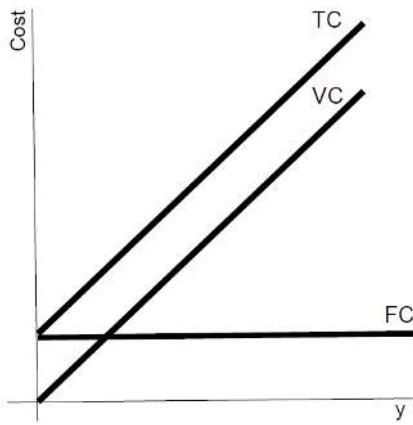
- The **opportunity costs** are the value of the best alternative forgone, in a situation of limited resources. They have to be taken into account when making economic decisions.
- The **fixed costs** (FC) do not vary with the output and are relevant over the long run. They are wages, rents, plant and so on.



- The **variable costs** (VC) vary depending on the output and are relevant over the short run. They are raw materials, energy and so on.



- The **quasi-fixed costs** do not depend on the output, but the company faces them in case it produces a certain threshold (for example recruitment costs).
- The **total costs** (TC) are the sum of fixed and variable costs.



- The **average cost** (AC), or unitary cost, is the cost for every unit produced:

$$AC(y) = \frac{C(y)}{y} = \frac{VC(y) + FC}{y} = \frac{VC(y)}{y} + \frac{FC}{y}$$

- The **average fixed cost** (AFC) is:

$$AFC(y) = \frac{FC}{y} \quad \begin{matrix} \text{If } & y \rightarrow 0 \Rightarrow AFC(y) \rightarrow \infty \\ \text{If } & y \rightarrow \infty \Rightarrow AFC(y) \rightarrow 0 \end{matrix}$$

- The **average variable cost** (AVC) is:

$$AVC(y) = \frac{VC(y)}{y}$$

AC=AFC+AVC AVC=AC-AFC

The average variable cost is lower than the average cost. **The AVC curve is below the AC curve.** The distance between the AVC and AC curves reduces as y increases:

$$AC(y) = AVC(y) + AFC(y) \quad \text{and} \quad AFC(y) \rightarrow 0 \quad \text{if} \quad y \rightarrow \infty$$

- The **marginal cost** (MC) is the total cost of producing $y+1$ units of output minus the total cost of producing y units of output. In other words, it is the cost of producing one additional unit of a good. It is given by the **first derivative of the total cost** or the variable cost:

$$MC(y) = \frac{\Delta C(y)}{\Delta y} = \frac{C(y + \Delta y) - C(y)}{\Delta y}$$

$$MC(y) = \lim_{\Delta y \rightarrow 0} \frac{C(y + \Delta y) - C(y)}{\Delta y} = \frac{\partial C(y)}{\partial y}$$

$$MC(y) = \frac{\partial C(y)}{\partial y} = D[VC(y) + FC]dy = \frac{\partial VC(y)}{\partial y}$$

In order to understand the relationship between AVC and MC, we compute the derivative of AVC:

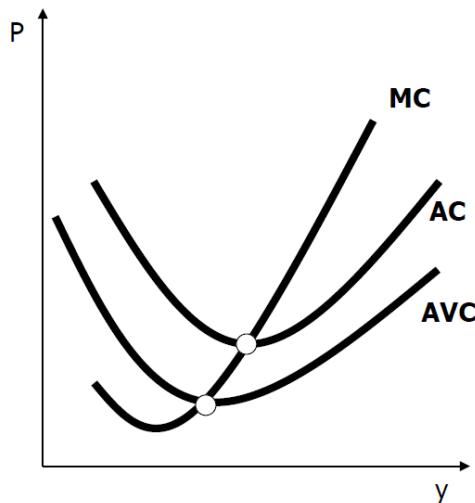
$$\frac{\partial AVC(y)}{\partial y} = \frac{\partial \left[\frac{VC(y)}{y} \right]}{\partial y} = \frac{VC'(y) \cdot y - VC(y)}{y^2} = \frac{1}{y} \cdot \left[VC'(y) - \frac{VC(y)}{y} \right] = \frac{1}{y} \cdot [MC(y) - AVC(y)]$$

$$\frac{\partial AVC(y)}{\partial y} > 0 \quad \text{if} \quad [MC(y) - AVC(y)] > 0 \Rightarrow MC(y) > AVC(y)$$

$$\frac{\partial AVC(y)}{\partial y} < 0 \quad \text{if} \quad [MC(y) - AVC(y)] < 0 \Rightarrow MC(y) < AVC(y)$$

In the increasing part of the AVC curve, the MC curve is above the AVC. Instead, in the decreasing part of the AVC curve, the MC curve is below the AVC curve.

The two curves intersect each other at AC minimum point.



Marginal cost is the appropriate cost concept to decide **how much to produce**. Indeed, if the MC decreases, the production should be increased.

Average cost is the appropriate cost concept to decide **whether to produce**. If the AC is lower than the price, the production can start.

- The **avoidable and not avoidable cost**: the difference between these two types of cost refers to the decision taken, for example increasing or decreasing production, make or buy decisions and so on. The avoidable costs depends on the decision taken. Not avoidable costs incur anyway, regardless the decision taken.
- The **sunk costs** depend on the specificity of the asset: if the asset is specific, it cannot be used for alternative productions. It becomes an exit/entry barrier. The opportunity cost is null.

Profit maximization and cost minimization

In economics, the firm is traditionally considered a black box whose major goal is to maximize its profit. Symmetrically, another major goal of the firm is to minimize costs.

In reality, the concepts of ***profit maximization*** and ***cost minimization*** do not always coincide. Paradoxically, a firm can minimize its costs dismissing its production activity (for example before selling a brand). Furthermore, the concept of profit maximization implies not only the possibility of cost reduction, but also of revenue increase.

We can define the ***marginal revenue*** (MR) as the total revenue change due to the production of one additional unit. It is the derivative of revenue with respect to output:

$$MR = \frac{\Delta TR(q)}{\Delta q} = \frac{TR(q + \Delta q) - TR(q)}{\Delta q}$$

$$MR = \lim_{\Delta q \rightarrow 0} \frac{TR(q + \Delta q) - TR(q)}{\Delta q} = \frac{\partial TR(q)}{\partial q}$$

$$TR = p \cdot q \quad \epsilon_p = \frac{\partial q}{\partial p} \cdot \frac{p}{q}$$

$$MR = \frac{\partial p}{\partial q} q + p = -\frac{1}{\varepsilon} \cdot p + p = p \left(1 - \frac{1}{\varepsilon}\right)$$

$$\text{Profit maximization: } \max_q \pi(q) = p(q) \cdot q - TC(q)$$

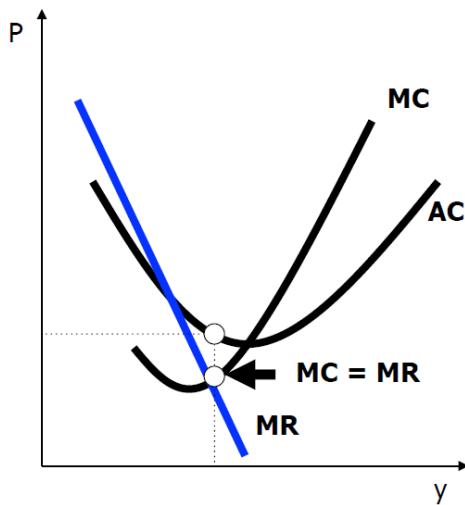
$$\pi(q) = TR(q) - TC(q)$$

$$\frac{\partial \pi(q)}{\partial q} = \frac{\partial TR(q)}{\partial q} - \frac{\partial TC(q)}{\partial q} = 0$$

$$MR = \frac{\partial p(q)}{\partial q} q + p(q) = -\frac{1}{\varepsilon} \cdot p(q) + p(q) = p \left(1 - \frac{1}{\varepsilon}\right)$$

$$MR = p \left(1 - \frac{1}{\varepsilon}\right) = MC$$

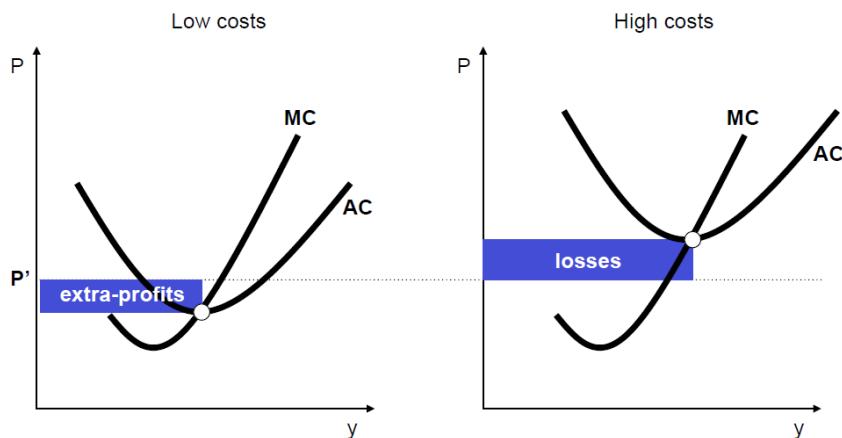
The optimal production is in the point where $MC=MR$. Once found the equilibrium, it is possible to find the average cost. The price must be higher than the average cost.



If $MR > MC$, the firm is maximizing profit on each unit but not maximizing the overall profit. An output increase implies higher profits.

If $MR < MC$, the firm is making losses on each unit produced. An output decrease implies higher profits.

In the point $MC=MR$, the profit is maximized and we can define the quantity produced and supplied.



In the long run, companies facing losses will go out of the market.

The goal of profit maximization is an assumption of classical economics; however, companies might be unable or unwilling to maximize their profits. There are some other types of goals:

- **Profit satisficing:** In organizations, only shareholders want to maximize their profit. Workers have usually little incentives to maximize profits. They make enough profits to keep their job, but then they pursue other objectives.
- **Increasing market share:** often firms seem to be most concerned with increasing their market share in order to obtain economies of scale, prestige and monopoly power in the future.
- **Non-economics motives:** a goal should be the corporate social responsibility towards stakeholders (employees, society, and environment).

The supply

The supply is the **amount a company is willing to produce at a given price**. The **produced quantity** is found at the point where $MR=MC$. The **market supply** is given by the **horizontal sum** of the supplies of each individual company.

The supply is positively inclined as long as companies tend to increase their production as prices increases. The **shape of the supply schedule** depends on:

- **Technology**: higher technology allows lower MC (downshift of the MC curve) and higher supplied quantity for the same price.
- **Prices of the factors of production**: lower input costs allow lower MC and higher supplied quantity for the same price.
- **Regulation**: regulations as the production subsidies allow lower MC and higher supplied quantity for the same price.

Productive efficiency

The productive efficiency represents how close the actual production cost is to the lowest cost achievable. Considering the possible evolution of the efficiency over time, there are two options:

- **Static efficiency**: it refers to the efficiency at a particular point in time.
- **Dynamic efficiency**: it refers to the improvement over time of products and production techniques.

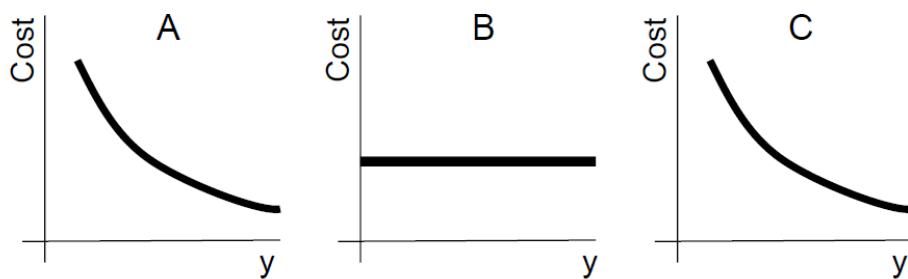
Economies of scale

Economies of scale are realized when the **average cost decreases** as the output and the productivity increases and the **total costs are spread over a greater range of output** (even if it increases).

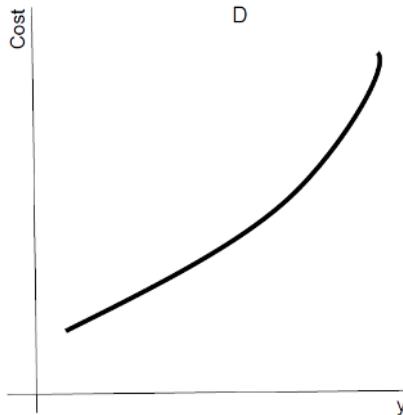
Diseconomies of scale are realized when the average cost increases as the output increases.

Increasing economies of scale impacts:

- The average fixed cost decreases as the output increases (A).
- The variable average cost does not change as the output increases (B).
- The total average cost decreases as the output increases (C).



Decreasing economies of scale (diseconomies of scale) impact on the average cost:



We can define the **scale elasticity** as the ratio between a cost variation and an output variation. It is the percentage of cost increase after an increase of 1% in output.

$$\varepsilon_S = \frac{\frac{\Delta C}{C}}{\frac{\Delta y}{y}} = \frac{\Delta C}{\Delta y} \cdot \frac{y}{C}$$

The **index of economies of scale** is:

$$S = \frac{1}{\varepsilon_S}$$

- If $S=1$, constant economies of scale
- If $S>1$, increasing economies of scale
- If $S<1$, decreasing economies of scale

The economies of scale can be:

- Internal: advantages that arise due to the firm growth.
- External: advantages firms can gain due to the industry growth (normally associated with a particular area).

There are different types of **internal economies** of scale:

- **Technical economies of scale:** they are achievable through:
- Specialization: large organizations can employ specialized labor.
 - Indivisibility of plant: machines cannot be broken down to do smaller jobs.
 - Principle of multiples: some firms need more than one machine of different capacities to be more efficient.

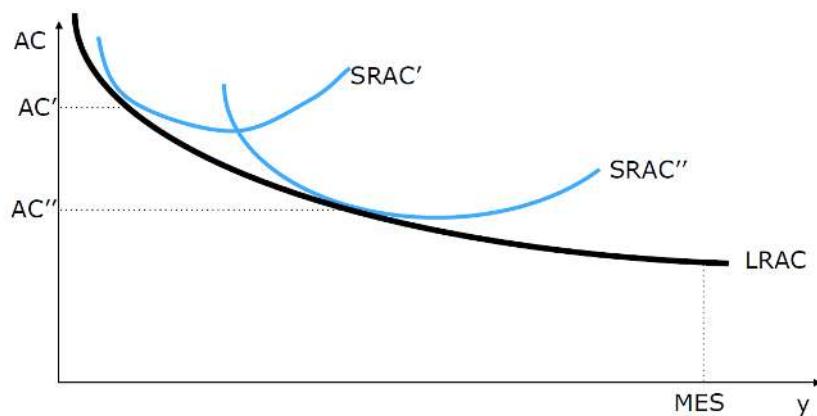
- The principle of massed reserves: large companies can split the risk related with reserves (spare parts, finished goods, maintenance staff and so on) more effectively than small companies can.
 - Increased dimensions: bigger transportation containers can reduce average cost. The two-thirds rule states that the cost of equipment increases roughly as the surface area, while the capacity increases roughly as the volume.
- **Commercial economies of scale:** large firms can negotiate favorable prices because of buying in bulk. They have high negotiating power. Due to their market power, they may have advantages in keeping prices higher.
- **Financial economies of scale:** large firms can negotiate cheaper financial deals, can be more flexible about finance (share options, rights issues and so on), and can utilize skills of merchant banks to arrange finance.
- **Managerial economies of scale:** use of specialist (accountants, marketing, lawyers, production, human resources and so on).
- **Risk-bearing economies of scale:** diversification, markets across regions/countries, product ranges, R&D.

There are different sources of **external economies of scale**:

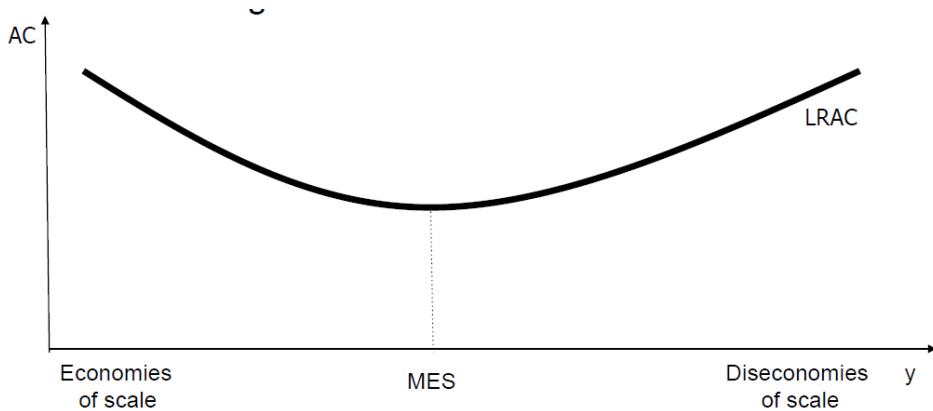
- Supply of skilled labour
- Reputation
- Local knowledge and skills
- Infrastructure
- Training facilities

In the **short run**, only variable inputs and variable costs can vary. In the **long run**, also the fixed inputs and the fixed costs can vary. Over the long run, there is not a distinction between fixed costs and variable costs as long as all the inputs can vary.

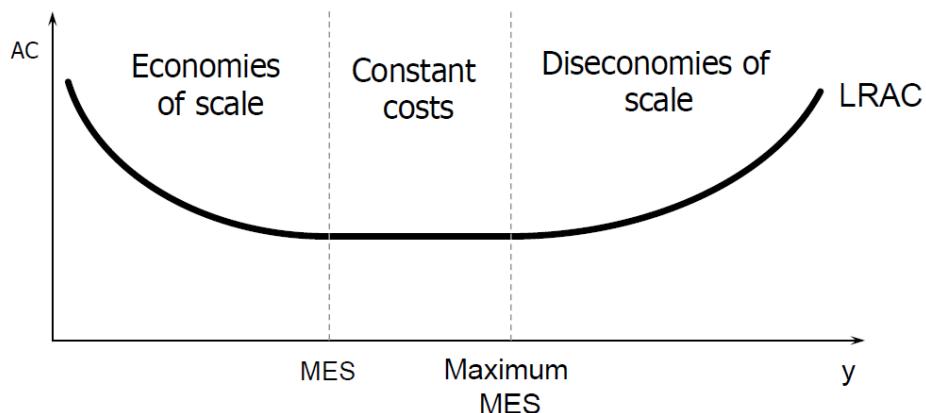
Therefore, we have **short run average cost (SRAC)** and **long run average cost (LRAC)**. Firms can minimize costs from both perspectives. LRAC is an envelope curve of SRAC because over the long run the firm varies also the capital in order to minimize costs.



The LRAC is traditionally considered U shaped due to increasing and decreasing economies of scale. The **minimum efficient scale (MES or efficient scale of production)** is the point where the firm minimizes its LRAC.



It is possible to identify also ***the maximum MES*** beyond which the LRAC increases.



The LRAC is decreasing until the MES and not decreasing beyond the MES.

The firm is in ***equilibrium*** producing at the MES:

- $Y < MES$: increasing y , firm can reduce AC (economies of scale).
- $Y > MES$: increasing y , AC increases (diseconomies of scale).
- $Y = MES$: no incentives to change the production levels.

The MES can be very high because of ***high fixed costs*** that need to be distributed on an increasing production volume. It is the case of a ***natural monopoly***: one single firm can work satisfying the MES.

The MES can be ***at industry level*** or ***at a plant level*** being the ***minimum efficient plant size (MEPS)***, which means the outcome level where increasing the scale of production of an individual plant provides no significant unit cost benefits. Beyond this threshold, the cost stop decreasing.

The ***diseconomies of scale*** refer to a large scale production that can lead to increasing average costs due to:

- Management problems
- Maintaining effective communication
- Coordinating activities across the globe
- Demotivation and alienation of staff
- Divorce of ownership and control

- Market saturation

In economics, we assume that firms are ***mono-product*** and ***mono-plant***, even if in reality the situation is different. Therefore, economies of scale depend on the level of analysis:

- Single product (intermediate or final)
- Plant/business unit
- Company
- Industry/group of companies

If we consider a ***specific product***, regardless the overall production of the plant/business unit, we need to ***define the product*** and its homogeneous category. We have to decide ***how to assess the production*** in terms of number of items produced in one unit of time, production lot and planned level of production. Finally, we have to ***define the time horizon***: the introduction of specialized and productive machineries can be beneficial in the long term for large quantities, while in the short term for little quantities the switching costs can be too high.

Considering the economies of scale related to a ***specific plant/business unit***, the AC is reduced for large plants due to general expenses decrease, staff activities and technical infrastructure.

At a ***company level***, the AC decreases for large companies typically due to commercial, managerial, financial and risk-bearing economies of scale. The diseconomies of scale lead to a risk of saturating the market: as the dimension increases, the company looks for new/different market opportunities.

Economies of scope

The average cost of producing two or more products (product mix) together is lower than the cost of producing them separately.

Products: $Y = (y_1, y_2, \dots, y_M)$

Production function: $Y = f(X)$

Cost function: $C(Y) = C(y_1, y_2, y_3, \dots, y_M, p, T)$

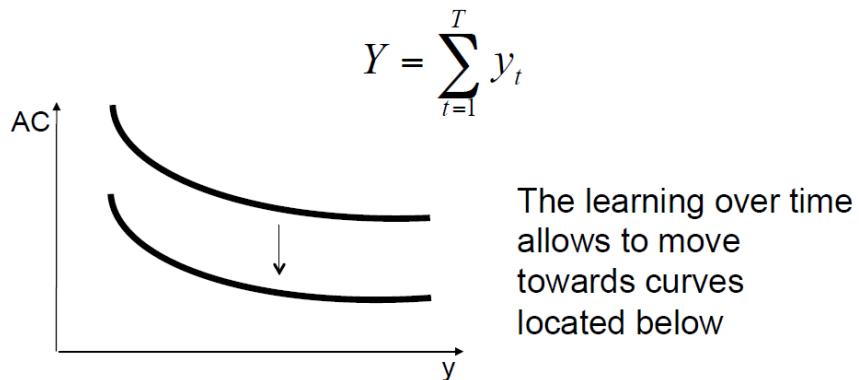
Economies of scope: $C(Y) < \sum_{i=1}^p C(\underline{y}^i)$

Sources of economies of scope:

- The fact that the production processes are not saturated
- Shared use of indivisible inputs
- Joint production

The AC decreases as the company gains experience because it learns how to become more efficient. A proxy of the underlying learning is the cumulative production where:

- Y is the cumulative production
- y_t is the production per year
- T is the number of years since the company is operating



The cost function at time t depends on:

- Outcome
- Prices
- Cumulative production as a proxy of experience

$$C_t = C(y_t, p_t, T, Y)$$

Over the long run, cost decreases due to:

- Greater productivity of the production factors
- Better managerial procedures
- Less wastes

Measurement techniques of economies of scale

We have two types of methods:

1. Methods relying on the cost function

a. Engineering approach

The steps are the following:

- Development of a cost function (constituted by several dots)
- Identification of the AC for several production level
- Identification of the MES where the curve changes its slope

At an industry level, we can use this method considering a sample of companies with different production levels but similar features.

There are some limitations of this method related to the need of identify a representative sample of companies, the need to process a relevant amount of information and the risk of referring the productivity changes only to the production level and not to other issues such as the organizational efficiency.

b. Econometric estimates

We have to identify the cost function analytically and the coefficient from empirical data. We have to include an error considering the mismatch between empirical data and real observations. The most adopted cost function is the Cobb-Douglas.

The major limitation refers to the fact that the results depend on the cost function adopted and the sample representativeness.

2. Methods relying on the MES

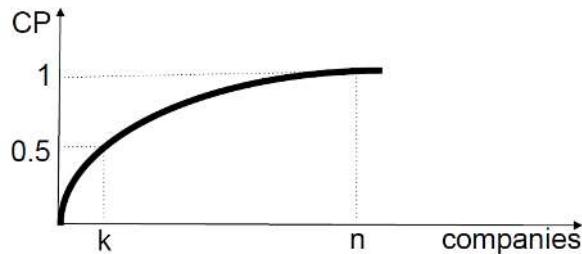
a. Statistical proxies

The proxy indirectly captures a phenomenon that cannot be directly measured. The methods consists of a comparison of the MES in several industries in order to assess the height of the barriers to entry. We have to calculate the cumulative production of the bigger companies compared to the entire population:

$$CP_k = \frac{\sum_{i=1}^k y_i}{\sum_{i=1}^n y_i}$$

For example, the first ten companies produce 70% of the industry output.

We can identify a concentration curve that represents the cumulative production as a function of companies ordered from the biggest to the smallest.



The highest the slope of the concentration curve, the highest is the concentration of the industry. This means that few companies capture high proportions of the industry total production.

We can identify two proxies of the MES:

- Average size of the first k companies: $MES = \frac{\sum_{i=1}^k y_i}{k}$
- Size of the k -th company: $MES = y_k$

The major limitations refer to the data availability at a firm level and to the fact that the solution depends on the adoption of size classes.

b. Stigler's surviving technique

It is based on the idea that the most efficient companies have higher probability to survive (Darwinian natural selection). Therefore, companies with low costs (operating at the MES) are more likely to survive.

Implementation:

- Clustering companies according to their size (number of employees, turnover and so on)

- Computing the level of production for each cluster over several periods: if the level of production decreases, the company is not efficient and AC is not minimized. If it increases, the size is efficient.

The major limitations refer to the need of considering long span of time in order to guarantee the selection. Moreover, it is possible to observe a market share growth in two contiguous classes not understanding which of the two is operating at the MES. Another problem is that it is based on not realistic assumptions:

- Prices and technology are constant
- Companies have the same goal (increasing market share)
- Companies operate in similar business context
- No barriers to entry
- No collusive agreements

Limitations of the economies of scale model

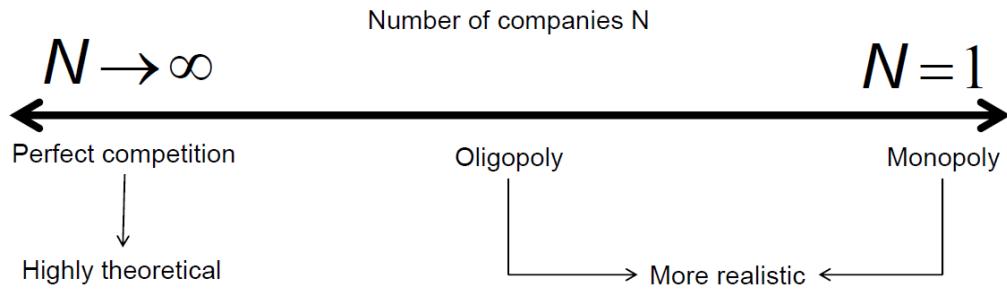
The ES model is based on some critical assumptions:

- In reality, it is not always possible to discretionary change the input level and some inputs are fixed over the short run.
- Companies are not always able to minimize their costs: the information is not perfect and the sunk costs do not allow choosing always the best combination of inputs.
- Companies do not always want to minimize their costs: they want to preserve or increase their market share and they want to penetrate new markets.
- Companies face switching costs and difficulties related to the adaptation of production processes: they need to re-bargain existing contracts with suppliers, train the employees, adopt new technologies and so on.
- Companies need time to improve (dynamic ES).

BASICS OF COMPETITIVE STRUCTURES

There are different market forms: we can compare them looking at the **market power**. The market power is the capability of a firm to set a price above the cost. Therefore, the profit achieved by a company depends on the market structure, which means barriers, number of firms and so on (it is not an internal characteristic). The monopoly (or firm with a dominant position) is characterized by the highest level of market power.

We can classify the market structures considering the number of firms:



- The perfect competition consists of infinite companies producing homogeneous products.
- The oligopoly consists of limited number of companies producing homogeneous products or differentiated products. In order to study the oligopoly, it is common to use the game theory.
- The monopoly consists of one single company producing a product without substitutes.

There is not always a correlation between the market power and the number of firms: for instance, a dominant firm has the market power even if there are other firms.

Perfect competition

The perfect competition is the market structure with the **highest production efficiency** because costs are minimized and the **highest allocative efficiency**, because the price is the lowest possible.

Since the **maximum social welfare** is the sum of the consumers' surplus and the producers' surplus, this is an ideal situation for the policy makers.

Firms do not gain any extra-profits. Therefore, this is the worst possible situation for firms.

A perfectly competitive market is a market where firms are **price-takers**, i.e. they do not determine their selling price since it is settled by the market, i.e. by the interaction of demand and supply.

The perfect competition refers to five central assumptions:

- **Atomicity:** there are several small firms (small compared with the demand of customers). The production capacity is limited.
- **Homogeneous products** not only physically but also from a customers' perception point of view due to brand reputation, advertising and so on.
- **Perfect information:** every agent, firm and consumer knows the price charged by every firm.
- **Technological symmetry:** every firm has the access at the same conditions to the available production technologies.
- **No entry and exit barriers (costs).**

Analyzing the **dynamics of the perfect competition**, companies are price takers because their products are homogeneous:

- If the company raises the price, consumers will buy the products from competitors and the demand for the company is null (perfect information). Therefore, no (rational) firm will raise the price.
- Firms cannot collude, given their high number (atomicity).

- If the company reduces the price, consumers will try to buy all the products from that company. However, the company is not able to serve the entire market due to its limited production capacity (atomicity).

Furthermore, companies do not gain extra-profits because they are a signal for new potential entrants:

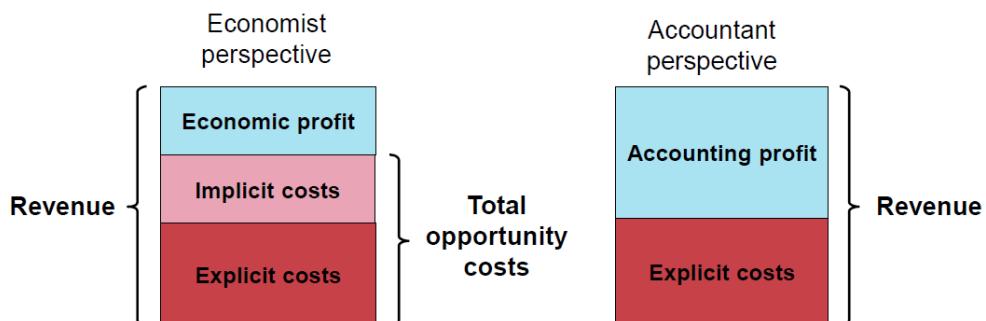
- If a firm makes extra-profit, this extra-profit will attract other firms to acquire the technology required (technological symmetry) and enter into the market (no barriers).
- This induced competition will erode in the long-run any possibility of extra-profits for the firms with the adoption of the best technologies possible by all firms.

The perfect competition equilibrium is:

$$p = MC = \min AC$$

This is the maximum productive efficiency since firms produce the maximum quantity of the good at the lowest possible cost. Furthermore, this is the maximum allocative efficiency, since the firms sell the maximum quantity of the good at the lowest price.

From an economic perspective, we consider explicit and implicit costs that are the total opportunity costs. Therefore, in perfect competition, the firms realize “normal” profits, but they should not realize extra-profits over the long run.



We can understand better why $p = MC$:

$$\max \pi = p \cdot q - TC(q)$$

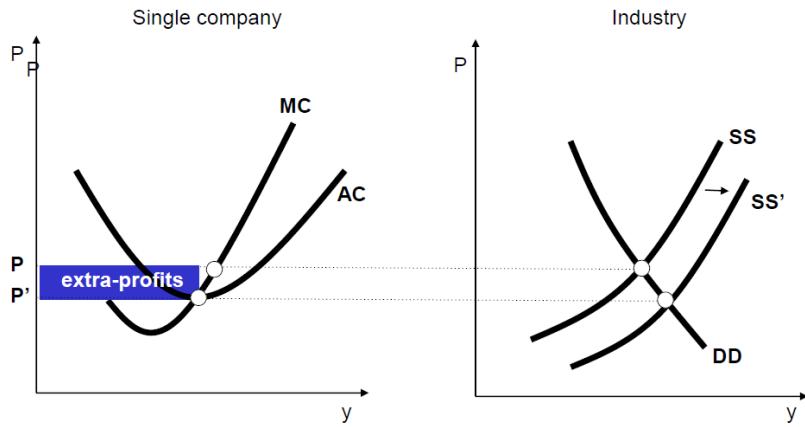
$$\pi = TR - TC$$

$$\frac{\partial \pi}{\partial q} = \frac{\partial TR}{\partial q} - \frac{\partial TC}{\partial q} = 0$$

$$MR = MC$$

$$p = MC$$

In order to understand the dynamics, we look at the **Demand (DD) and Supply (SS) curves**.



At price P , the profits realized attract new firms in the industry over the long run. Entrants will enter, start operations with easy access to technology and compete. The SS shifts to the right, quantity goes up, and the equilibrium price decreases. Over the long run $p = \min AC$ is the equilibrium.

In the opposite situation, if the firms incur in a loss, they want to move inputs to an alternative use. Some firms will exit the market and the SS will shift to the left.

There are two ***shutdown rules***:

- ***Production stop***: In the short run, firms incur in the fixed costs anyway. The company will stop production if keeping production alive will not help in recovering the variable cost. In other words:

$$TR < VC$$

$$\frac{p \cdot q}{q} = \frac{VC}{q}$$

$$p < AVC$$

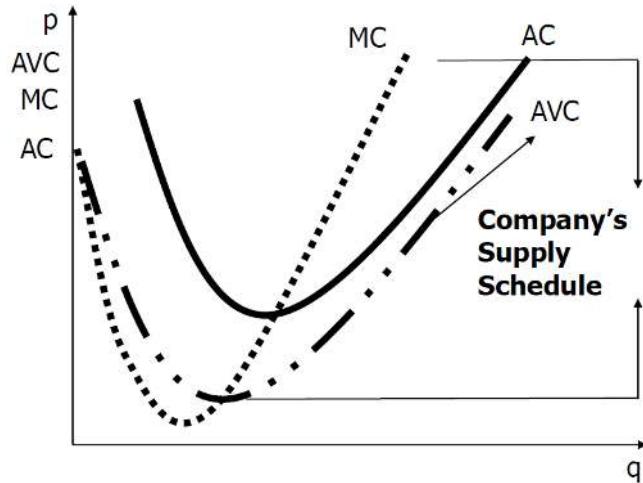
If this is true, firms will stop production.

- ***Market exit***: over the long run, firms do not incur in fixed costs anyway. A firm will get out of the market if:

$$TR < TC$$

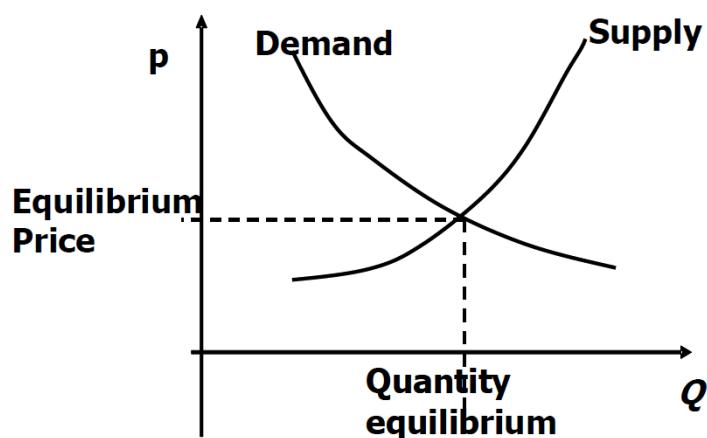
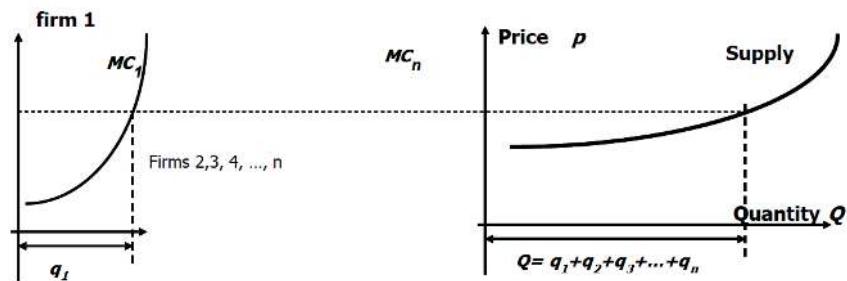
$$\frac{p \cdot q}{q} = \frac{TC}{q}$$

$$p < AC$$

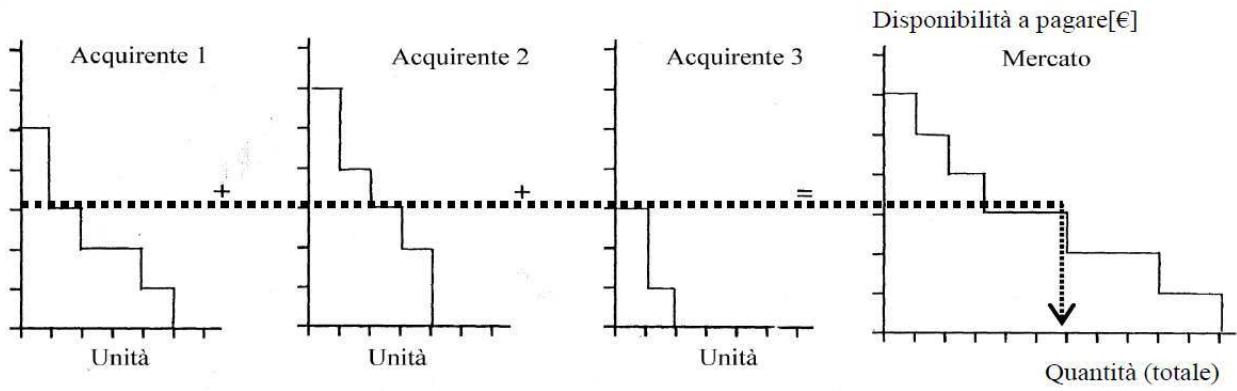


In perfect competition, the short run SS coincides with the increasing part of the marginal cost curve lying above the **average variable cost**. The long run SS coincides with the increasing part of the marginal cost curve lying above the **average cost curve**.

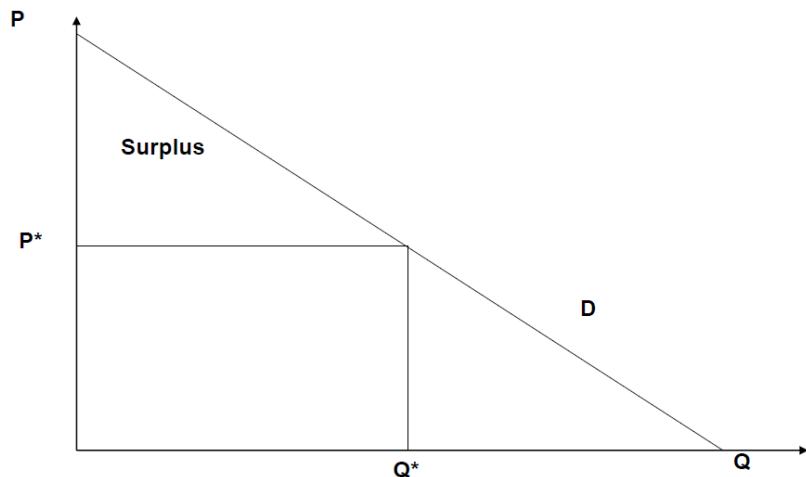
In perfect competition, the market supply curve is the sum of individual supply curves (willingness to sell) and is equal to their marginal cost curves.



The demand curve has a negative slope because consumers get decreasing marginal utility consuming goods. Summing horizontally the demand curve of each consumer, we obtain the aggregate demand.



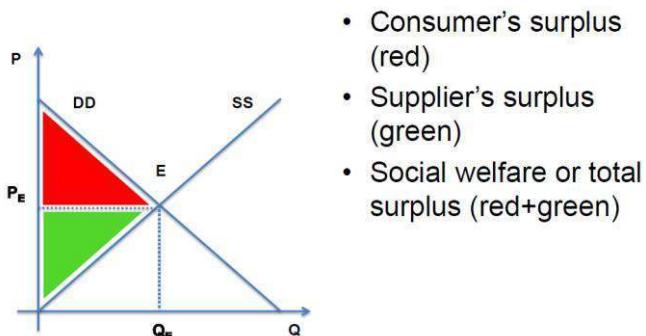
Consumer's surplus is the difference between the prices an individual is willing to pay to have a certain good or service and the market price for the same good or service. It measures the consumers' welfare. The total consumers' surplus from the aggregate demand curve is the following.



Producer's surplus is the difference between the prices of a certain good or service paid to the producer and the price the producer is willing to accept for selling the same good or service. It measures producers' welfare.

The **social welfare** is the sum of consumer's surplus and producer's surplus

$$W = S^c + \sum v\pi$$



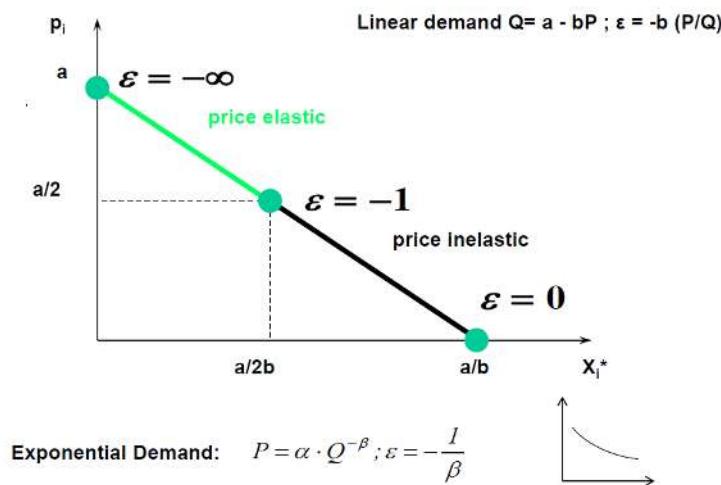
In perfect competition, we have the maximum area of surplus and welfare.

We can introduce the concept of demand elasticity, which is the variation of demand consequently to price changes.

$$\varepsilon_p = \frac{\frac{\Delta q}{q}}{\frac{\Delta p}{p}} = \frac{\Delta q}{\Delta p} \cdot \frac{p}{q}$$

Where q is the demanded quantity and p the price. For infinitesimal changes:

$$\varepsilon_p = \frac{\partial q}{\partial p} \cdot \frac{p}{q}$$



In the case of exponential demand, ε_p is constant along the demand curve. We can calculate the elasticity starting from the demand curve or the inverse demand curve.

Potential competition (contestable markets)

A **contestable market** is a market where firms from other markets/sectors can perform a **hit and run competition** with no costs of entry and exit. Hit and run competition means take the money and go away very fast. It looks similar to the perfect competition in terms of results.

We have some pre-requisites:

- No requisites on the number of firms in the market
- No entry and exit barriers (no sunk costs and non-re-deployable investments).
- Perfect information for consumers (they are able to react immediately to price differentials between companies)
- Time requested for the incumbent to retaliate to the entry of the new firms (by lowering price) is superior to the time needed for the entrant to make all the investment necessary to operate in the focal market. It means that there is no possibility to enter the market.

The number of firms is not always a good predictor of the market power that is difficult to observe in the real world.

It leads to the following results:

- Incumbent firms are forced to settle a **price near to the average cost** in order not to “turn on” the signal of extra-profits.
- Actually, every extra-profits will be captured and exploited by new entrants with a hit and run competition.
- If the market is contestable, the **number of firm is a poor predictor** of the market power, and even a market with only one firm may behave more similarly to perfect competition rather than monopoly. The number of firms is not always a good predictor of the market power.

The assumptions of perfect competition and contestable markets are difficult to observe in the real world. They are not so realistic but they are useful as **benchmarking models** for comparing results (in terms of efficiency and social welfare) of more realistic market structures, where firms have market power (oligopoly, monopoly).

Monopoly

The **highest level possible of market power** is the monopoly. There are no possible entrants and one single firm faces the whole demand. It is completely free to choose the price, being a price setter.

First of all, we have to note that **choosing price is equivalent to choosing the quantity** to produce and deliver on the market. The quantity becomes a decision variable.

There is **a surplus transferred from consumers to producers** with respect to the perfect competition. Monopoly is **inefficient in terms of social welfare** and avoided by policy makers.

$$\begin{aligned} \text{Max } \pi &= TR(q) - TC(q), p(q) \cdot q - TC(q) \quad MR = MC \\ \text{F.O.C} \quad p + q \cdot \frac{\partial p}{\partial q} &= MC \quad \text{or also:} \quad p \left[1 + \frac{q}{p} \frac{\partial p}{\partial q} \right] = MC \\ \text{Optimum} \quad p \left[1 - \frac{1}{|\varepsilon|} \right] &= MC \end{aligned}$$

Example with a linear demand curve:

$$p = a - bq$$

$$TR = pq = (a - bq)q = aq - bq^2$$

$$MR = a - 2bq$$

$$TC = F + cq^2$$

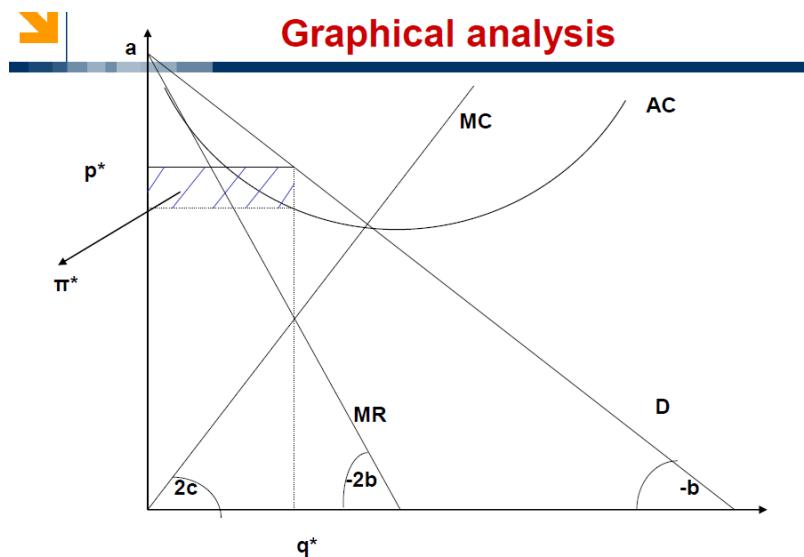
$$MC = 2cq$$

$$\text{Max } \pi \longrightarrow RM = CM, a - 2bq = 2cq \longrightarrow q^* = a/[2(b+c)],$$

$$\text{therefore } p^* = a - (ab)/[2(b+c)]$$

$$\pi^* = p^*q^* - F - cq^{*2}$$

Seeing it graphically:



The **demand elasticity** influences the price definition: the monopolist will settle a higher price as long as it faces an inelastic demand, while the more the demand curve is elastic, the more it will fix a price close to MC , and similar to the one that would emerge under perfect competition. In the case of rigid demand, if the firm rises the price, it will lose a relatively lower number of customers. Therefore, prices are much higher.

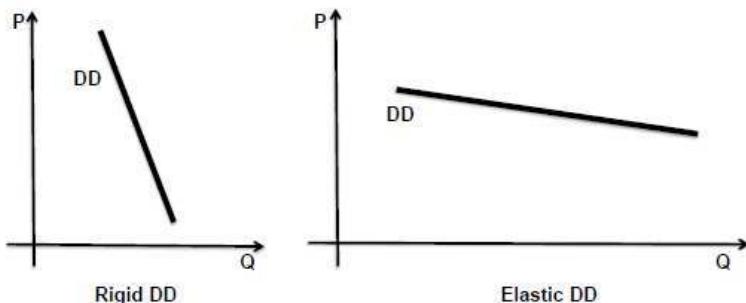
Starting from the optimum:

$$P \left[1 - \frac{1}{|\varepsilon|} \right] = MC$$

The mark-up will be:

$$\frac{P - MC}{P} = \frac{1}{|\varepsilon|}$$

We can identify graphically demand curve for the firm:



Comparing the demand curves, in perfect competition the firm DD is horizontal and elasticity tends to infinity ($p=MC$).

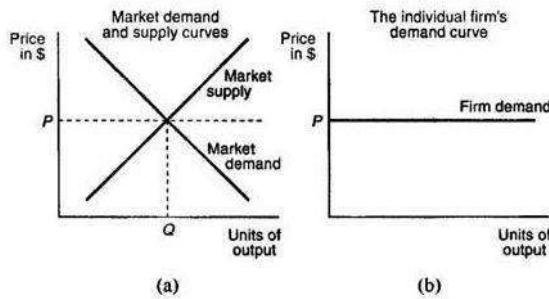
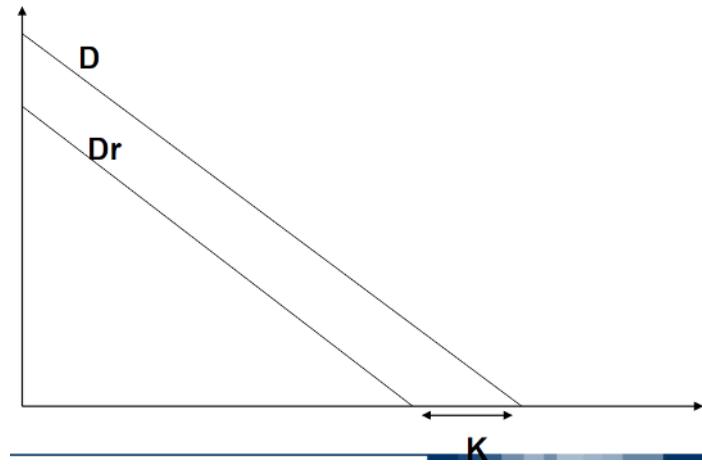


Figure 1
Market and firm demand curves in a perfectly competitive market

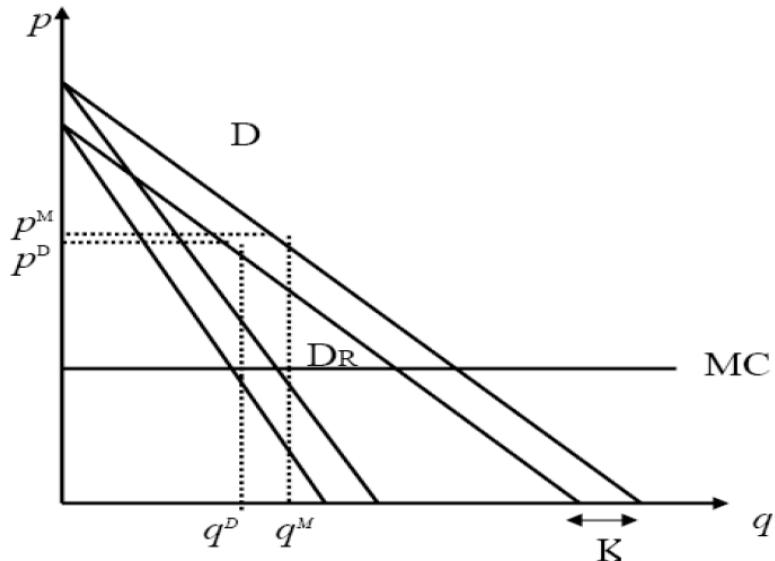
Dominant firm

The analysis on pure monopoly we conducted so far can also be applied to the market context in which there is only **one very large firm** and a **set of very small firms** with a limited production capacity.

If K is the total production capacity of the set of small firms, these latter typically fix a price only marginally inferior to the one fixed by the large firm and produce a quantity such as their capacity is saturated:



The dominant firm may act as a monopolist for the residual demand not covered by the small companies. We can identify graphically the optimum for a dominant firm:



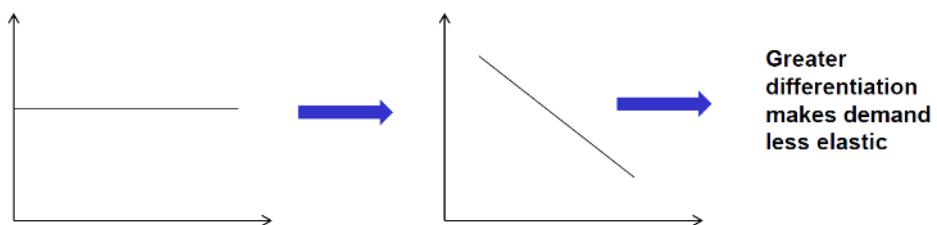
Monopolistic competition

The perfect competition assumption of homogeneous products is not realistic. In reality, all products are differentiated, at least slightly, and over a wide range of economic activities differentiation is of considerable importance. **Firms compete on differentiation.**

The central assumption are the same of the perfect competition, except for the homogeneity:

- Atomicity
- Perfect information
- Technological symmetry
- No entry and exit barriers

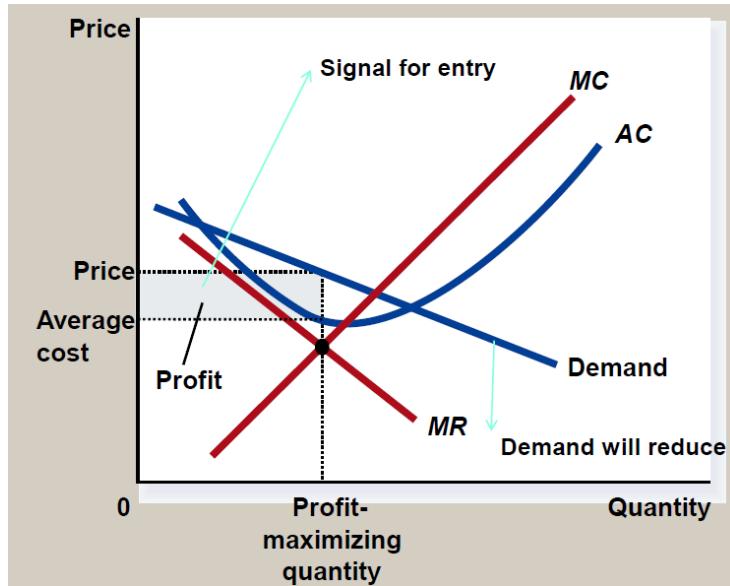
The huge change with respect to the perfect competition is that the firms are **price makers**. Each firm faces a **downward-sloping demand curve**.



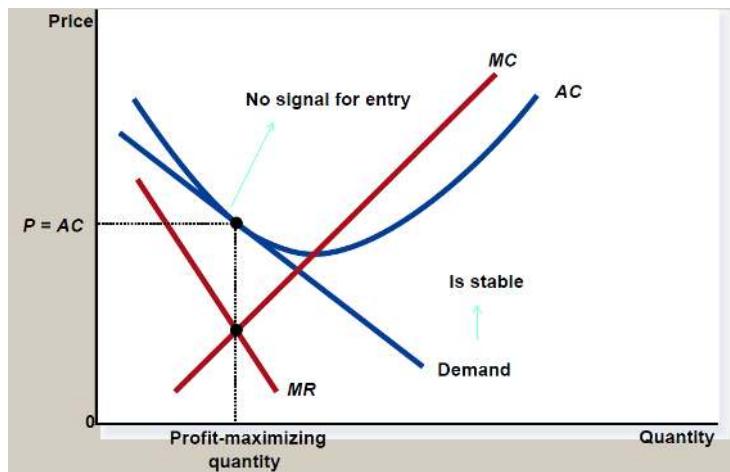
Demand has to be the more inelastic possible and the goal is to build **strong preferences**.

There is **not mutual dependence** among firms: the behavior of a single firm cannot influence the others. This is for the atomicity assumption: firms cannot collude agreeing a price and if they collude, there are no entry barriers: this is a signal for new entrants.

In the short run, the firm gains extra-profits, which are a signal for new entrants (the same in perfect competition). The demand curve shifts down until it is tangent to AC and there are no more extra-profits.



This situation leads to a stable demand, being the equilibrium over the long run.



In this equilibrium, there are **no extra-profits**: it is a situation of allocative and productive inefficiency: price is major than MC (consumers will go out of the market) and AC are not the minimum maximizing quantity.

This type of competition has elements of the perfect competition and the monopoly (freedom to choose the price). There is a debate on if this competition is good or bad. There are several brands: this means more options for consumers but more costs in advertising for firms. This is a waste of resources.

Oligopoly

The relevant outcomes measures (prices, quantities, profits, consumers' surplus, social welfare) are in the middle of those that emerge in perfect competition and monopoly. In both cases, firms do not care about other firms' reaction: there is no strategic interaction. In oligopoly, there are few firms and each decision taken by each single firm can influence other firms' profits. There is a **strategic interaction** studied through the game theory. Thus, oligopoly is more realistic.

The number of companies in the market is limited, higher than one but not tending to infinite. They are **price-maker**.

Use of the game theory

Strategic interaction means the production and price decisions of each company consider other players' decisions. The game theory is an analytical apparatus needed to study oligopolistic models. A game is characterized by:

- **N rational players** belonging to the strategic context;
- **A strategy** (set of possible actions) that can be:
 - Pure strategy, if it is deterministically chosen;
 - Mixed strategy, if a probability is associated to the decision;
- **Payoff**: a utility/profit functions associated to each strategy.

The simplifying assumptions adopted in the oligopolistic models is that there are **two companies** producing **homogeneous products**.

Rational players mean that:

- They can assess the **consequences of their strategy**;
- They order the decisions on the basis of the **preferences** and the expected **outcome** achieved by each strategy;
- They can understand which is the **best strategy** for them;
- Each **player knows** that the other player is rational;
- **The information** regarding the strategies of the game and the payoffs is **perfect**.

Models can be:

- **One shot**: decisions are simultaneous;
- **Sequential**: one of the two companies takes its decisions before the other; therefore, the other company benefit from an **information advantage**.

The strategic variables in these models are **price** and **quantity**.

We can introduce the concept of **collusion**, which means companies jointly agree on prices and/or quantities in order to **maximize the industry profit**. They jointly behave as **monopolist**. The collusion is modeled in the game theory through **cooperative games**. It is possible to demonstrate that:

$$\pi_{Monopoly} > \sum_{i=1}^N \pi_{Oligopoly}$$

Analyzing the game theory, we can identify a **dominant strategy**, which is the best strategy regardless the strategy of the other player. This dominant strategy eliminates the **dominated strategies**.

		Supplier 2	
		Cooperate	Not cooperate
Supplier 1	Cooperate	20, 20	0, 30
	Not cooperate	30, 0	10, 10

The solution can be:

- A **Nash equilibrium**: a set of strategies constitutes a Nash equilibrium if no player can unilaterally change its strategy in a way that improves its payoff after considering an opponent's choice.
- **Pareto-efficient** if there is not a situation where simultaneously a player can improve its payoff and the payoff of the other does not get worst.

In the example, the solution is a Nash equilibrium but is not Pareto-efficient.

We have three models to describe the oligopoly:

- Bertrand
- Cournot
- Stackelberg

	Decision on quantities	Decisions on prices	Type of model/game
Sequential decisions	Quantity leadership (Stackelberg)	Price leadership	Sequential and competitive
Simultaneous decisions	Quantity choice (Cournot)	Price choice (Bertrand)	One-shot and competitive
Collusion	Quantity joint decision	Price joint decision	Cooperative

Cournot's model

In the **Cournot's model**, the strategic variable is the **quantity**. It considers two firms with identical MC, a linear demand curve and no fixed costs.

It is based on very restrictive **assumptions**:

- Only two firms (**duopoly**);
- **No potential entrants** (closed markets);
- **Homogeneous products** in terms of quality and technology;

- **No variations** in **demand**;
- No variations in the **cost function**;
- Only one strategic variable: quantity;
- **Simultaneous decisions** about the production level (quantity) before deciding the price;
- **Price determined by the market** at a level where the demand equals the joint production of the two firms.

Firms choose how much they want to produce and the **price is given by the aggregated demand** (under the assumption of standard goods, the DD has a negative slope):

$$p_1 = p_2 = P(Q) = P(q_1 + q_2)$$

There are some critiques to the model:

- **The competition on quantity is not realistic**: firms compete on several things and usually on prices. The demand is determined according to the demand diagram.
- However, we have to interpret the quantity as **the production capacity**. Firms decide their production capacity in advance and subsequently they compete on prices.

The firms' profit functions are formalized in terms of quantity:

We assume a **linear cost function**:

$$TC = c \cdot q$$

We have:

$$\pi_1 = q_1 \cdot [p(q_1 + q_2) - c]$$

$$\pi_2 = q_2 \cdot [p(q_1 + q_2) - c]$$

If we assume a **linear cost function**, the **fixed costs are zero** by definition. Therefore, the average cost and the marginal cost coincide.

$$MC = \frac{\partial TC}{\partial q} = c \quad AC = \frac{TC}{q} = c$$

Thus, the profit will be:

$$\pi = TR - TC \Rightarrow \pi = p \cdot q - FC - AC \cdot q$$

$$\Rightarrow \pi = (p - AC) \cdot q \Rightarrow \pi = (p - c) \cdot q$$

The firms strategically interact through the quantity. We can find the **equilibrium** where, given the competitor's choice, the firms choose the strategy enabling to maximize their profit.

Assuming that:

- Firms can choose the quantity they prefer in the interval [zero; +infinity);
- Both the profit functions can be differentiated in quantity;

We can write the **first order conditions**:

$$\begin{cases} \frac{\partial \pi_1}{\partial q_1} = f(q_1, q_2, c_1) = 0 \\ \frac{\partial \pi_2}{\partial q_2} = g(q_1, q_2, c_1) = 0 \end{cases}$$

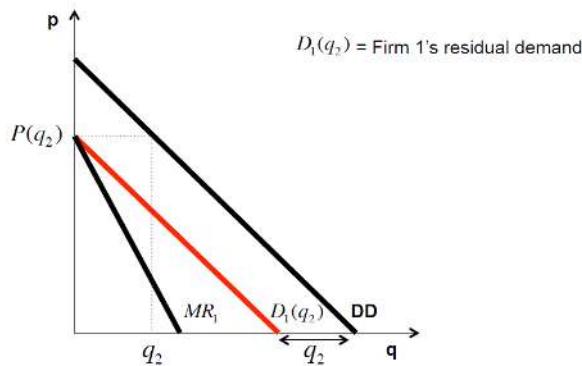
This system leads to the identification of the quantity maximizing the profits of the two firms (necessary condition). The sufficient condition is that the stationary point must be a maximum.

In order to identify the equilibrium, we need two steps:

1. We have to identify the ***optimal choice*** of each firm given the conjectures regarding the competitor's behavior. This choice is given by the ***reaction function***, which describes the quantity produced by each player in order to maximize its profit given the competitor's production quantity (locus of points of maximum profits).
2. We have to ***put the two reaction curves together*** in order to find the combination of decisions mutually compatible.

Given the quantity produced by firm 2 (q_2), the firm 1 residual demand is found $D_1(q_2)$. The residual demand is the relationship between the quantity produced by firm 1 and the price, assumed a certain quantity produced by firm 2. On the residual demand, firm 1 is a monopolist because $MR=MC$. We can find the optimal quantity for firm 1 given the assumption on the quantity produced by firm 2 using firm 1 reaction function. The process ends up in the equilibrium.

Graphically:



Note: Marginal Revenue is a curve with twice the slope of the Demand Diagram and with the same vertical intercept. This results from our assumption that the demand is linear. In general, the Marginal Revenue curve has the same intercept as the demand curve and a higher slope (in absolute value), not necessarily twice the slope of the demand curve

For example:

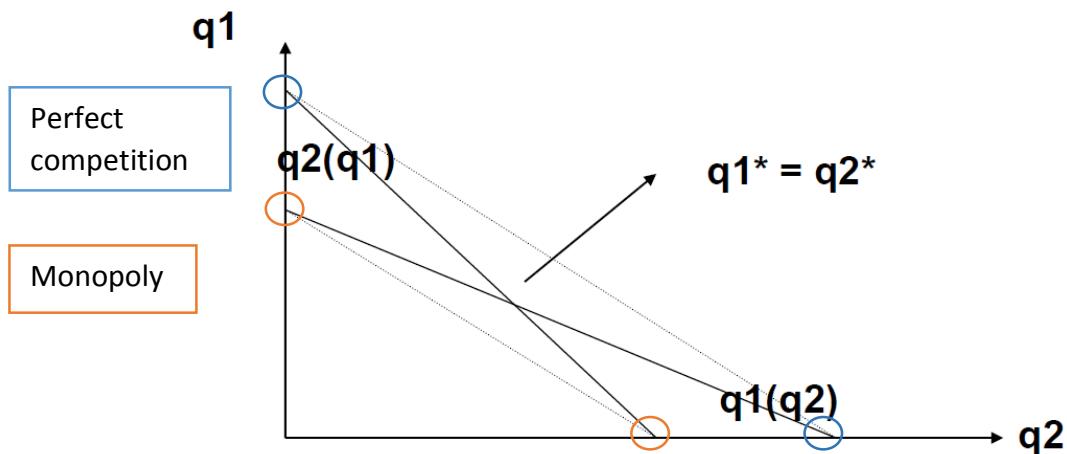
$$p = a - b(q_1 + q_2)$$

$$\Pi_1 = pq_1 - cq_1 = [a - b(q_1+q_2)]q_1 - cq_1$$

$$\Pi_2 = pq_2 - cq_2 = [a - b(q_1+q_2)]q_2 - cq_2$$

$$d\Pi_1/dq_1 = a - 2bq_1 - bq_2 - c = 0 \longrightarrow q_1^* = [(a - c)/2b] - (1/2)q_2^*$$

$$d\Pi_2/dq_2 = a - 2bq_2 - bq_1 - c = 0 \longrightarrow q_2^* = [(a - c)/2b] - (1/2)q_1^*$$



The two curves are the ***reaction functions*** of one firm with respect to the quantity delivered by the other. The intersection between the reaction functions is the best response of one firm given the quantity of the other.

It is a ***Nash equilibrium***: in an operational definition, it is a bi-univocal correspondence between the two responses of the players.

This equilibrium depends on the ***firm's belief about what the other firm is doing***. It is a point where a firm chooses its optimal quantity given the quantity it assumes for the other firm. Those conjectures are correct.

We can find a ***dynamic*** of the oligopoly: where the responses of the two firms are equal to zero, the competitive structure is the perfect competition. The other two points on the axes refer to the monopoly: the quantity in these two points is the best response of one firm if the other decides not to produce.

If one firm decides for a quantity, we can derive the best response of the other from its reaction function. Going on in this way, the process ends up in the equilibrium.

The quantities produced in perfect competition and monopoly are the two lines between the points: we can notice that the quantity produced in oligopoly is in the middle of the quantities produced in monopoly and perfect competition.

In ***the real world***, firms' strategic behaviors depend on several variables: price, quantity, R&D investments, product features, commercialization modes and so on. All these decisions implies strategic interaction and interdependence. A firm can take these decisions ***over the short-run*** (very quickly: price), ***over the middle-run*** (production scales, incremental innovations, features and distribution of the products) ***and over the long-run*** (radical innovation, entry decisions).

Stackelberg's model

It is a ***sequential competition*** articulated in two steps, in each step firms take strategic decisions. The simultaneity of decisions assumption is not valid but ***the two steps are dependent***: there is a strategic interaction between them.

The **competitive environment changes** according to the decisions taken in the first step. The second step inherits these conditions about price and/or quantity. The constraints and asymmetries in the first step can turn into positive **extra-profits** in the second step that means price is higher than the marginal cost. The results can be different if compared with the simultaneous competition.

The two-steps model solution is a **backward induction**: firstly, we solve the second step given the decision taken in the first step. Therefore, we solve the first step in order to maximize the payoff.

In the sequential model, one of the two firm precedes the other in choosing the strategy: it is called **first-mover advantage**. The first mover is the **leader** and the other is the **follower**, reacting to the leader's decision.

The backward induction is based on the assumptions that the leader:

- Knows the game rules (perfect information);
- Is rational;
- Knows that the follower is rational;
- Knows that the follower in the second step will try to maximize its profit given the leader's decision.

In the Stackelberg's model, the leader decides the quantity at time zero and the follower reacts in time one. We can talk about **quantity leadership**. In the **real world**, there are industries with one dominant firm or **natural leader**. In these cases, the other firms observe the leader's decisions regarding products to place in the market before doing it themselves.

We have to note that the Stackelberg's model is different from the case where there is one dominant firm on the market and the remaining firms satisfying the very little residual demand. In this model, **the leader and the follower equally share the market**.

The model:

- Firm 1 is the leader: in t=0 it chooses the quantity q_1
- Firm 2 is the follower: in t=1 reacts with the quantity q_2

The market price is given by the inverse demand function

$$p(Q) = p(q_1 + q_2)$$

The leader's profit:

- $\pi_1 = p(q_1 + q_2) \cdot q_1 - c(q_1)$
- Is function of q_2 , i.e. of the (next) follower's choice

The leader takes its decision considering the future choice of the follower: it considers the follower's profit maximization issue (game theory).

Imposing the first order condition:

$$\frac{\partial \pi_1}{\partial q_2} = 0 \Rightarrow p(q_1 + q_2) + \frac{\partial p(q_1 + q_2)}{\partial q_2} \cdot q_2 - MC(q_2) = 0$$

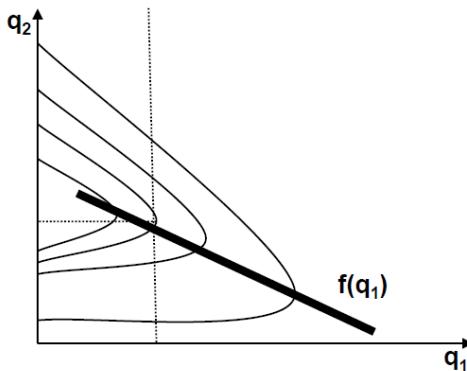
$$p(q_1 + q_2) + \frac{\partial p(q_1 + q_2)}{\partial q_2} \cdot q_2 = MC(q_2)$$

The **follower reaction function** is:

$$q_2 = f_2(q_1)$$

It describes the output level maximizing the follower's profit, given the level of production chosen by the leader.

We can introduce the isoprofit curves: points where profit is the same, moving from the curve it decreases/increases.



In the optimal points on the isoprofit curves, their tangent lines must have a vertical slope. Applying the first order condition for a maximum, we can find:

$$\frac{\partial \pi_2}{\partial q_2} = 0 \Rightarrow MR_2 = MC_2$$

If $q_1=0$, we have the highest possible profit for firm 2, which is a monopolist.

Considering the leader's choice, it considers the future follower's choice choosing its output level. Therefore, the profit maximization depends on it:

$$\max_{q_1} \pi_1 = p[q_1 + f_2(q_1)] \cdot q_1 - c(q_1)$$

Deriving the leader's profit on the quantity, we can find its reaction function.

Putting together the two reaction functions, we can find the **equilibrium**. The total output will be the sum of the two firms' outputs and it will be greater than the output under monopoly conditions.

Bertrand model

Assumptions:

- **Two firms** producing **homogeneous products**
- **No potential entrants** (closed markets);
- **No variations in demand**;

- No variations in the **cost function**;
- Only one strategic variable: **price**;
- **Simultaneous decisions** about price;
- **Consumers demand the good to the company with the lowest price** (no differences of products, no versioning, no switching costs)

Assuming a linear cost function $TC = c \cdot y$

If i sets a price

- Lower than j, it captures the entire market demand
- Equal to j, it shares the market demand with j
- Greater than j, it has a null market demand (consumers demand the good to j)

$$D_i(p_i, p_j) = \begin{cases} D(p_i) & \text{if } p_i < p_j \\ \frac{1}{2}D(p_i) & \text{if } p_i = p_j \\ 0 & \text{if } p_i > p_j \end{cases}$$

The two firms choose a price in order to maximize profits:

$$\pi_i(p_i, p_j) = p_i \cdot D_i(p_i, p_j) - c \cdot D_i(p_i, p_j) = (p_i - c) \cdot D_i(p_i, p_j)$$

$$\pi_j(p_i, p_j) = p_j \cdot D_j(p_i, p_j) - c \cdot D_j(p_i, p_j) = (p_j - c) \cdot D_j(p_i, p_j)$$

The game is not cooperative: each firm maximizes its own profit.

The Bertrand's model solution is a **Nash equilibrium**: no one can unilaterally change its position and improve its situation. Each company's price maximizes the profit, given the other's choice.

$$(p_i^*, p_j^*)$$

$$\begin{cases} \pi_i(p_i^*, p_j^*) > \pi_i(p_i, p_j^*) & \forall p_i \neq p_i^* \\ \pi_j(p_i^*, p_j^*) > \pi_j(p_i, p_j^*) & \forall p_j \neq p_j^* \end{cases}$$

It is possible to demonstrate that in the Nash equilibrium each firm chooses a price equal to its marginal costs.

$$p_i = p_j = c$$

Firms cannot set different prices between each other and greater than c. Price cannot be lower than c, otherwise it would make losses instead of profits. Given the market conditions, firms are identical and the game is symmetric. The reasoning developed for one player is perfectly applicable to the other.

We can analyze three cases:

1. $p_i > p_j > c$

In this case, firms set different prices and higher than the marginal cost. Firm i has null demand, while j captures the entire market demand and makes extra-profits. This solution is not a Nash equilibrium, i finds convenient to reduce the price in order to make extra-profits. This situation goes on until the equilibrium.

2. $p_i > p_j = c$

In this case, firms set a different price and equal to the marginal cost. For both the firms the profit is null and they have an incentive to deviate from their choices. The firm i has null demand that is an incentive to reduce the price. The firm j has a null extra-profit, thus it has an incentive to raise the price until it is only a bit lower than the competitor's price in order to capture the entire market.

3. $p_i = p_j > c$

In this case, they both gain extra-profits but we assume that they are competing and not colluding. Therefore, they will set a lower price in order to get the entire market instead of sharing.

Each company has always the incentive to revise its price decision, unless the price for both the companies is equal to the marginal cost. The equilibrium is the following:

$$(p_i^*, p_j^*) = (c, c)$$

$$\begin{cases} \pi_i(p_i^*, p_j^*) > \pi_i(p_i, p_j^*) \quad \forall p_i \neq p_i^* \\ \pi_j(p_i^*, p_j^*) > \pi_j(p_i, p_j^*) \quad \forall p_j \neq p_j^* \end{cases}$$

This is a Nash equilibrium: none of the two companies has an incentive to change its choice, given the other's choice. If price is higher than the marginal cost, the company loses the entire demand. If price is lower than the marginal cost, the firm makes losses instead of profits.

This solution depends on the assumptions of the model:

- If products are **homogeneous**, the demand depends only on price. With differentiation and information asymmetries, price can be higher without decreases in demand.
- If each company has the **same technology**, the marginal cost is equal for both of them. With different technologies, one company can be more competitive than the other is.
- If each company is able to satisfy the entire market demand, there are **no capacity constraints in production**. With capacity constraints, a different equilibrium is feasible.
- It is a **one-shot (static) game** where competition is for a period only. In sequential (dynamic) games, where we consider the time dimension, it is possible to obtain the equilibrium with price higher than the marginal cost. There are three possible scenarios:
 - **Price war**: the firm that has decided price at the beginning does not get the entire market demand as long as the other firm further reduces price. Price war should be avoided because some companies go out of the market but the others incur great losses.
 - The **simple threat** stemming from the second firm might be enough to keep the price low.

- **Trigger strategy:** the players decide to cooperate. One player punishes the others if they do not cooperate. In particular, there are two options: the first is a **tit-for-tat strategy**, which means the punishment goes on until the other player starts to cooperate again. The second option is a **grim trigger strategy**, which means the punishment goes on indefinitely after the other player defects just once and the cooperation shut down.

Introducing **capacity constraints** in the Bertrand model, if one firm produces a maximum quantity of k_i and beyond k_i the costs go to infinity (extreme case of decreasing returns to scale), the results differ from the standard Bertrand model: it is possible to achieve a price higher than the marginal cost. It is a solution of the Bertrand paradox.

There are some **additional assumptions** to the ones of the standard model:

- The two companies face capacity constraints k_1 and k_2 .
- Each company cannot sell more than k_i .
- After this threshold, consumers demand the good to the other.

The demand function of one company depends on:

- The price level set by the two companies (strategic interdependence);
- The capacity constraints;
- The rule of distribution R.

If company 2 sets a price higher than company 1, its demand would be null without capacity constraints for company 1 but positive with capacity constraints for company 1.

$$\text{If } D_1(p_1) = D(p_1) > k_1 \rightarrow D_2(p_2) = D(p_2) - k_1$$

It means that if capacity constraint is lower than demand, companies can set different prices: if a firm satisfies its demand, the other can set a higher price to its **residual demand**. With capacity constraints, even if a company set a lower price, it is not able to satisfy the entire demand: a residual demand remains for the other.

We find a **Nash equilibrium**:

$$p_1 = p_2 = p(k_1 + k_2)$$

In the equilibrium, each company sets a price in order to match the market demand with the aggregated production capacity.

$$D = k_1 + k_2$$

In the equilibrium, there is a **saturation of the production capacity**. The game can also be not symmetric.

We can demonstrate that in the equilibrium companies have not an incentive to move from their position. Analyzing alternatives for firm 2:

1) $p_2 < p(k_1 + k_2)$

Firm 2 catches the entire market demand but it cannot satisfy it. It just sells the same quantity at the lower price. It has no incentive to move from its position as long as its profits decrease.

2) $p_2 > p(k_1 + k_2)$

The firm 1 catches the entire demand, but it cannot satisfy it. Firm 2 has a positive demand equal to $D(p_2) - k_1$. The marginal revenue is greater than the marginal cost for each unit produced lower than k_2 . The revenue loss is greater than the saving gained anyway, thus $p_2 > p(k_1 + k_2)$ implies a lower production for firm 2. Profits are not maximized. Firms do not have an incentive to move from the equilibrium setting a higher price.

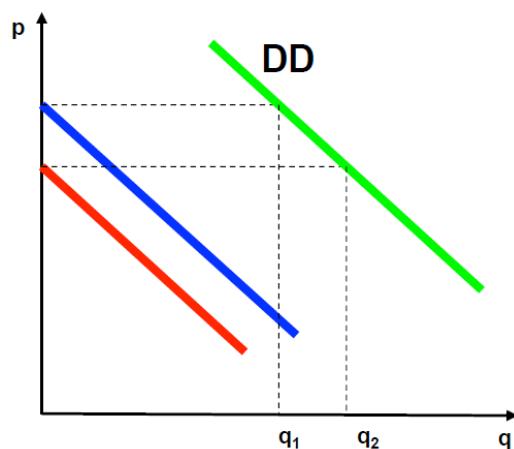
The equilibrium depends on the size of **capacity constraints compared with the market demand**. High capacity constraints are able to satisfy the entire market demand; therefore, if both firms are able to satisfy the entire demand, we can apply the standard model. Limited capacity constraints impose equilibrium prices higher than the marginal cost.

Entry barriers

An entry means the entrance in a market or industry of a new firm producing a good that is a **perfect substitute** of the goods already produced in that market or industry.

The degree of substitutability depends on the **consumers' preferences** as long as the goods can have different features. It is important how consumers perceive the features of a product. A new entry does not imply the creation of a new firm; it could be a matter of **diversification**.

The decision to enter a new market depends on the **expected profits**, which are a function of the **production costs** and the **demand conditions after the entry**. Indeed, the entry has an impact on the market quantity and price. Therefore, new entrants have to forecast the reaction of incumbents.



The blue curve represents the demand for the entrant if incumbents produce q_1 . The red curve represents the demand for the entrant if incumbents produce q_2 . The green curve is the market demand at an industry level. The new entrant will face a potential demand given by:

$$PD = \text{Market Demand} - q_{\text{incumbents}}$$

Where $q_{\text{incumbents}}$ is the quantity that incumbents decide to produce after the entrance. It means the new entrant gets the **residual demand**.

The entry barriers are obstacles preventing new firms from:

- Starting to produce goods of a certain type
- Entering a market/industry and compete against incumbents

In a market **without entry barriers**, every firm can enter and compete with the incumbents, and over the long run, **the price is equal to the average cost**. We can demonstrate that:

1) $p > AC$

Firms make extra-profits. Without entry barriers, new firms enter the market.

$$\pi = p(q) \cdot q - AC(q) \cdot q = [p(q) - AC(q)] \cdot q > 0$$

Due to the increase in the supply, the price decreases. New firms will go on entering until $p=AC$.

2) $p < AC$

Firms make losses. Without entry barriers, the most inefficient firms leave the market. The price increases due to the reduction of the supply. Firms will go on leaving the market until $p=AC$.

Entry barriers allow the incumbents **to keep the price higher than the average cost** exploiting a **market power**. However, the threat of new potential entries affects the price set by incumbents. After new entries, incumbents risk to make losses. Therefore, the relation **$p=AC$ might be valid also in the short-run** as long as the incumbents tend to keep lower prices in order to prevent new entries.

The entry barriers are **additional costs sustained by new entrants** with respect to costs sustained by incumbents. Analyzing the entry barriers taxonomy:

1. **Institutional/legal barriers**: there are different frameworks:

- An administrative authorization in order to perform a business activity
- Institutional monopoly through the grant of patent

2. **Structural technological conditions**

- Absolute cost advantage for the incumbents
- Economies of scale create a cost advantage and a price decrease

3. Incumbents' **greater capability to negotiate contracts** due to contractual power and trust

4. **Demand conditions such as consumers' preferences** for existing products: consumers tend to buy products closer to their preferences and they do not want to bear switching costs. New entrants have to incur in advertisement costs with uncertain outcomes in order to change the customers' behavior.
5. **Firms' strategic behaviors:** the incumbent has the incentive to prevent new entries in order to preserve its extra-profits.

If there is more than one typology of entry barriers in the same industry, the maximum difference between the price and the average cost without incurring in new entries is given by the height of the highest entry barrier.

There are different definitions of entry barrier:

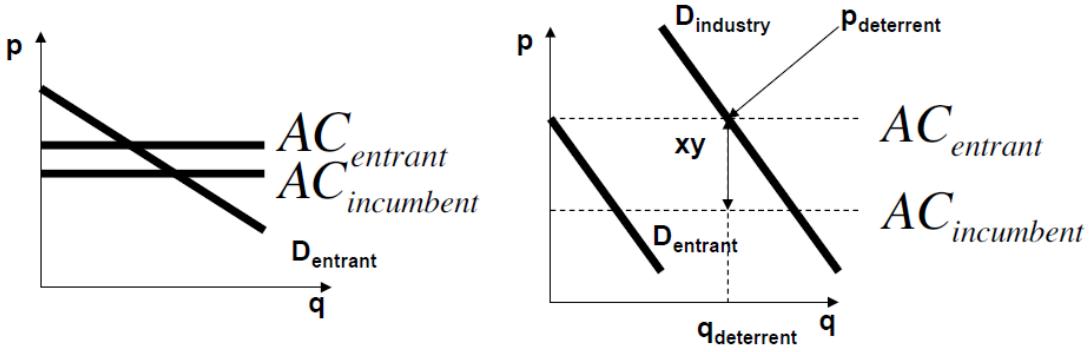
- **Bain's definition:** some rents are granted to incumbents, even though the conditions for competition are guaranteed. Non-efficient firms do not enter the market and there are no social surplus losses.
- **Stigler's definition:** entry barriers are given by the cost that a firm entering new markets has to bear, but that the incumbents do not have to bear. Therefore, there is an asymmetry and in particular an advantage for incumbents if compared with new potential entrants.
- **Weizsäcker's definition:** it adds to the Stigler's definition a ***distortion in resource allocation***, so that there are inefficiencies and loss of social surplus if compared with the perfect competition. This inefficiency is related to a very little protection of incumbents, for example if new entrants can copy the incumbents, they do not have incentive to spend in innovation, or to a very high protection of incumbents, that implies too much protection.

Strategic interdependence

Once described the entry barriers, we can introduce the **Sylos Labini's postulate**: new potential entrants behave as if they were able to forecast that the **incumbents will keep their production at the same level** as before the new entrance. Therefore, they assess whether entering or not comparing the **actual Demand Diagram** and **their own cost function** (including opportunity costs achievable in other industries).

This is a strong assumption in particular for small markets where it is difficult that companies maintain a high level of production after an entrance. Nevertheless, we assume that the new entrants' conjectures are correct.

Therefore, the ***potential demand*** for the new entrants is the difference between the market demand and the quantity already offered by the incumbents.



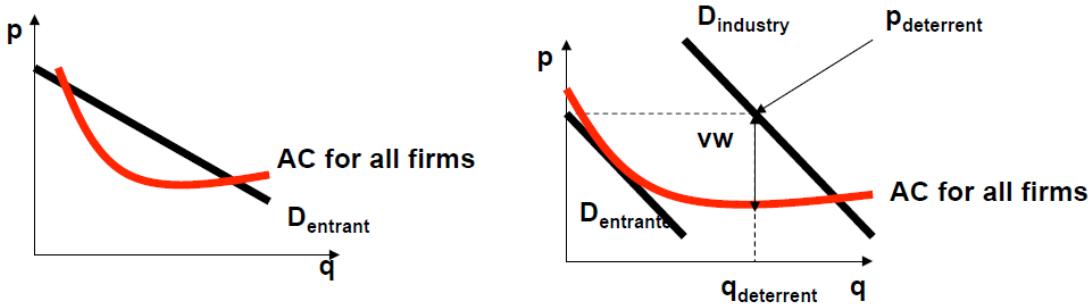
If $p_{before\ the\ entrance} - AC_{incumbents} > AC_{entrants} - AC_{incumbents}$, the demand of the new entrant is above the cost function and new firms enter the market. The **deterrent price and quantity** are those values preventing the new entrance.

The greater the difference between incumbents' costs and new entrants' costs, the greater can be the difference between price and costs of incumbents without incurring in new entries.

We assume that new entrants have to face additional cost anyway, having a higher AC with respect to incumbents. It is not always true; it could be possible that they have a new efficient technology and a lower AC.

The **maximum value of the difference between price and average cost** without incurring in new entries depends on:

- The breadth of the market
- The demand elasticity
- The production size
- The speed of the average cost decline



We can introduce the **Bain-Sylos Labini-Modigliani model (B-SL-M)**. There are some assumptions:

- Complete information
- Sylos Labini's postulate
- 2-steps competition: in $t=1$, the incumbent is the monopolist and it decides both price and quantity. It is the leader. In $t=2$, a new potential entrant decides whether entering in the market. It is the follower.
- Linear costs (constant average cost)
- Incumbent's **absolute cost advantage**: incumbent's costs are always lower than new potential entrants' costs. Several situations can be included in this model:

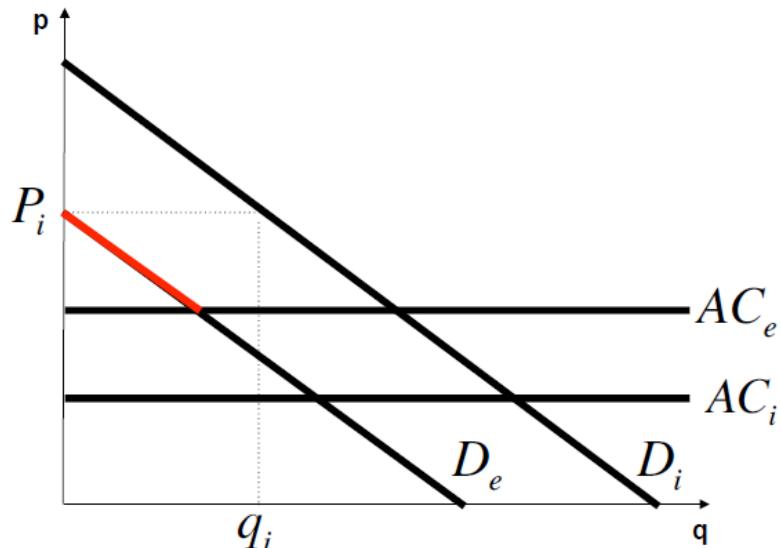
- **Product differentiation** that leads to higher advertisement costs in order to convince consumers to switch from one product to another.
- **Institutional barriers**, for example the payment of the royalties related with a certain patent.

The existence of absolute cost advantage is not a sufficient condition to have an entry barrier. There are entry barriers because the incumbents are also exploiting their market power.

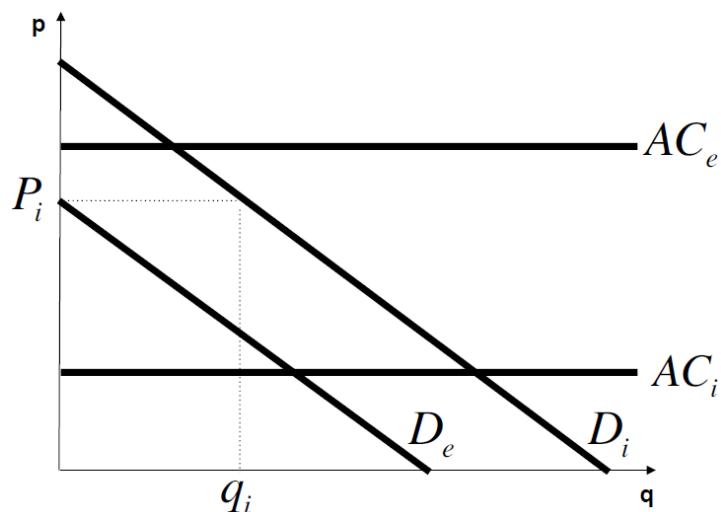
The entrant takes the entry decision comparing the potential demand and the cost function. It will actually enter the market if it can obtain positive profits:

$$p_{\text{before the entrance}} - AC_{\text{incumbents}} > AC_{\text{entrants}} - AC_{\text{incumbents}}$$

The entrant's residual demand diagram will have a part above the average cost curve assuring positive profits because $p > AC_e$. Therefore, they will enter the market.



In the following example, there is no entrance because the entrant cannot make positive profits.



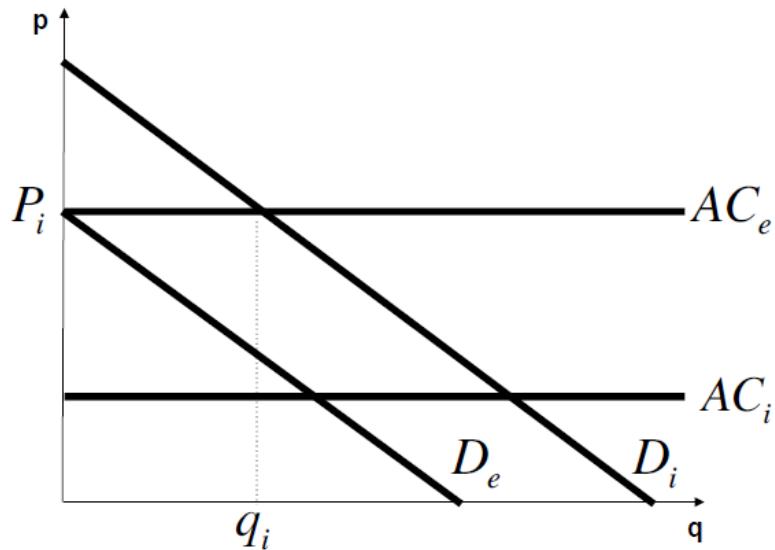
The difference between price and average cost of the incumbent is **a proxy of the height of the entry barrier**.

The **limit price** is where:

$$p_{\text{before the entrance}} - AC_{\text{incumbents}} > AC_{\text{entrants}} - AC_{\text{incumbents}}$$

$$p_{\text{before the entrance}} = AC_{\text{entrants}}$$

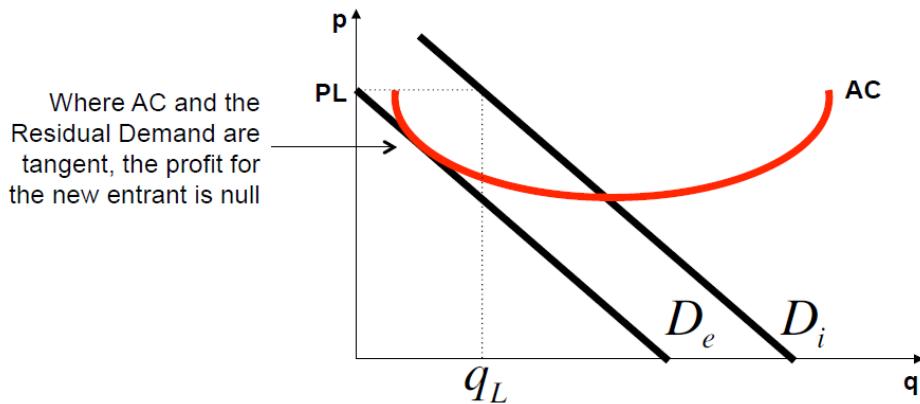
It is the highest price that can be charged without incurring in new entries. For higher prices, the entrant enters the market. For lower prices, the entrant does not enter the market.



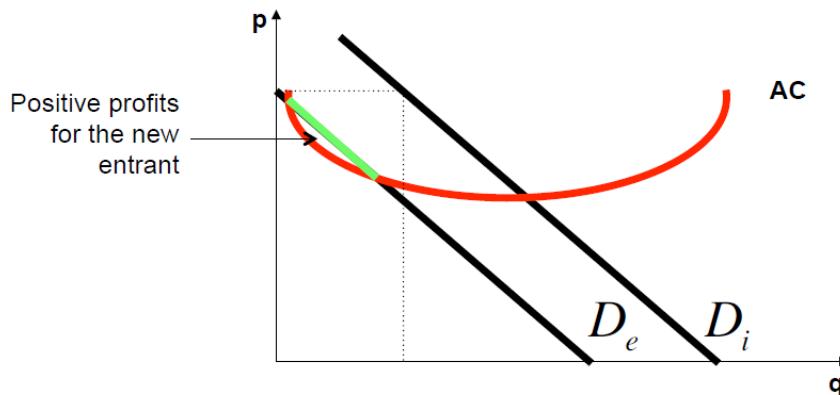
Introducing in this model the economies of scale, we have decreasing average cost before the MES. The Cournot's paradigm states that the higher the MES, the farther the industry equilibrium is from the perfect competition. In the B-SL-M model, **the higher the economies of scale, the higher is the price limit**.

Adding the assumption of same cost function for the two firms, that means **no absolute cost advantage**, we can solve the problem with the **backward induction**:

- 1) SL postulate: the new entrant knows that incumbents produce the same quantity as in t=1.
It decides considering the residual demand.
- 2) The incumbent sets the quantity so that the residual demand will allow only null profits for the new entrant. It determines the residual demand for the new entrant considering the average cost curve. The new entry turns to be not convenient.



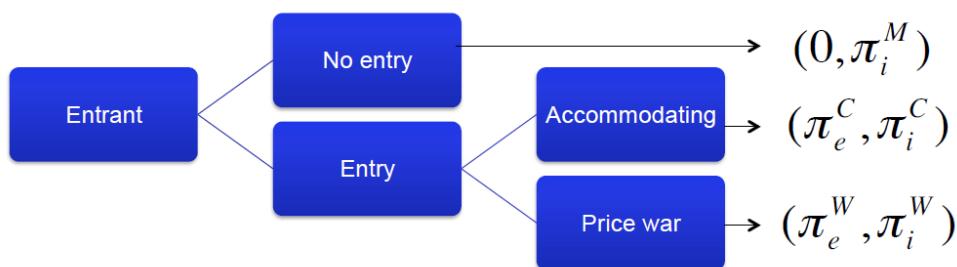
If AC is higher than the residual demand, the price is lower than the AC. Therefore, the entrant would realize losses. If the price is higher than the limit price, there is a part of the residual demand above the average cost. Therefore, new entrant gains positive profits.



A **limitation of the SL postulate** is related to the **irrational conjectures** of the potential entrant regarding the leader's behavior. After the entry, the incumbent might not find convenient anymore to produce the limit quantity at the limit price. It might choose:

- An **accommodating strategy** in order to avoid oversupply that would reduce the price below the average cost. It means to produce less in order to maintain extra-profits.
- A **price war**, keeping the production to the limit quantity and leading to an oversupply in the market. The consequences are a price below AC and losses for both the players.

The Dixit Model removes the SL postulate, changing the problem:



In the first step, the new entrant decides whether entering the new market. In the second step, the incumbent decides whether producing the limit quantity or adopting a different strategy.

If the potential entrant decides not to enter, the incumbent will make a ***monopolistic profit***. This is a ***Nash equilibrium*** because the incumbent prevents itself from a price war preferring not to enter and the incumbent makes monopoly profits. It is not a perfect equilibrium because if the entrant decides to enter, the incumbent will prefer the Cournot profit instead of the losses. ***The threat to engage in a price war is not credible*** because the monopolist does not have incentive to implement this threat. ***The entry will not incur if and only if the Cournot profit is negative***, but the potential entrant would not enter anyway for negative profits.

The Dixit model suggests that ***barriers can be only structural***: the incumbent always adopts an accommodating strategy and the entry process continues until the Cournot profit is negative.

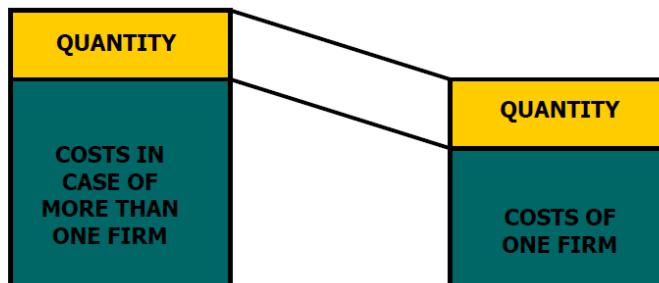
Structural	Strategic
Profits stemming from accommodating strategies are negative anyway	On purpose firms' behavior (as in the Price limit model)

If the potential entrant decides to enter, the incumbent might adopt an accommodating strategy, leading to Cournot duopoly profits, or a price war, leading to losses for both the parties.

There is not a right model to describe reality; it depends on the assumptions compared with the real situation in order to understand the possible players' behaviors.

Natural monopoly

The assumption is the ***subadditivity of the cost function***, which means an industry where multi-firms production is more costly than production by a single firm.



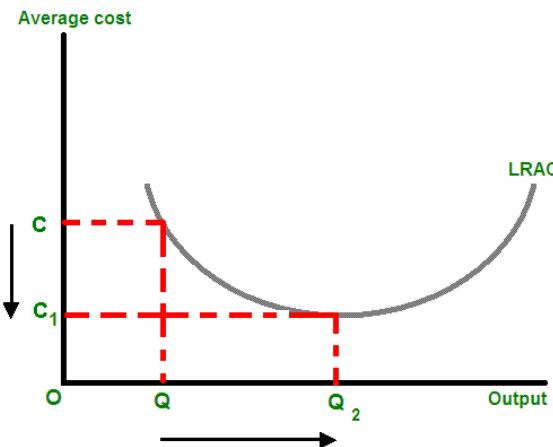
$$TC(Q) < \sum_{i=1}^n TC(q_i) \quad \sum_{i=1}^n q_i = Q, \forall n \geq 2$$

There are two conditions:

- ***Productive efficiency***: monopoly leads to a higher level of productive efficiency.
- ***Dynamics efficiency***: the capacity to produce technological change and innovation is higher in a monopoly.

In this case, the **economies of scale** are a sufficient but not necessary condition to prove subadditivity. In a natural monopoly, if one firm produces all the quantity, it is more convenient from a productive point of view. Nevertheless, we can be in a natural monopoly even if there are not economies of scale.

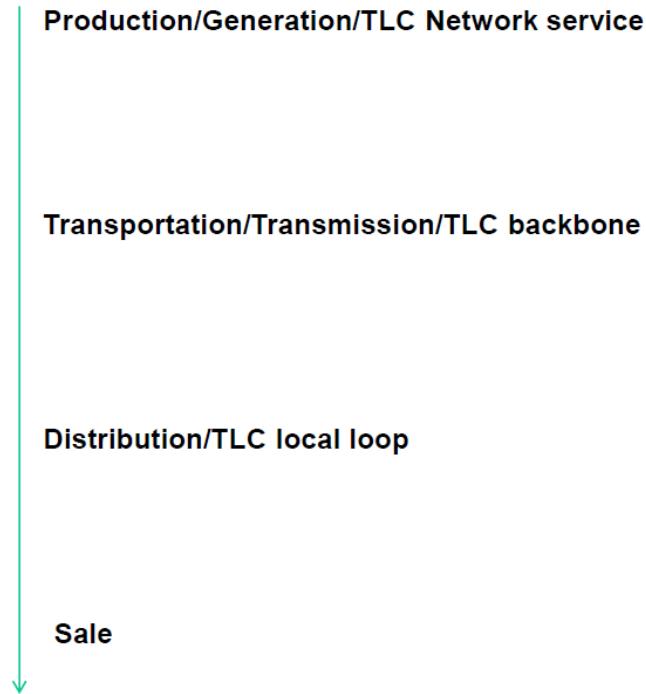
We can introduce the concept of **minimum efficient scale (MES)**: it is the minimum cost of production if one firm produces all the quantity. The TC area of one firm is less than the sum of the other firms TC areas.



The closer we are to the MES, the more likely is to be in a natural monopoly because a second firm has to cover the cost gap.

The natural monopoly is a crucial economic concept behind the **public service utilities**. These utilities are natural gas, electricity, water, telecommunications, and railroads. These businesses deliver an essential good or service through a wide network infrastructure.

There are different phases to deliver these services:



The subadditivity of cost influences all the phases. In the past, this led to a natural monopoly. After that, the European liberalization process opened the ***competition in production and sale***, while the other phases are still a natural monopoly.

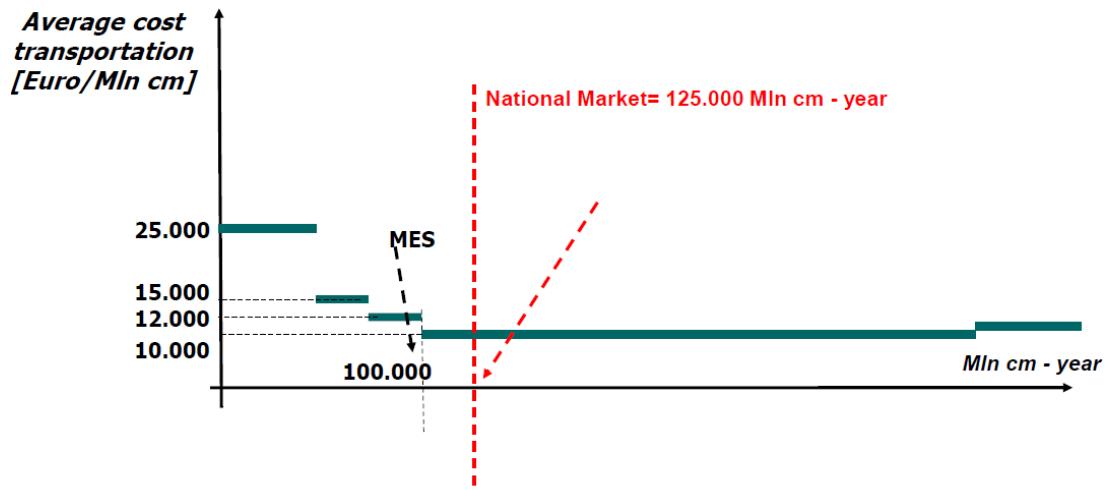
In order to understand, if we are in a situation of natural monopoly we have to find the MES:

$$MES = \frac{\partial AC}{\partial q} = 0$$

Once found the MES, we have to look at the ***demand curve***. If the demand curve intersect the AC curve far from the MES, it is not a natural monopoly.

Demand is the reason why policy makers decided to open competition in some phases of the utility services.

For example, considering the ***transportation of natural gas***, the size of the market is close or slightly higher than the MES. The total cost suffered by one firm is lower than the cost split among more companies.



Nº of firms	Average cost transportation	Total transportation costs
1	10.000 Euro/Mln cm-year	1.250 Mln Euro
2	15.000 Euro/Mln cm-year	1.875 Mln Euro
3	25.000 Euro/Mln cm-year	3.125 Mln Euro

In this industry, there are strong economies of scale related to:

- Great amount of **indivisible resources**
- **Specialization of workforce**: there are specialized network planners and geographically specialized shippers.

There are other industry-specific reasons related to the **volumetric returns to scale of the pipelines** (consequently to a diameter increase of the pipe) and the **administrative costs** for acquiring the right to transit bargained with foreign countries which are relatively insensitive to the volume of gas transported.

From a **social welfare** point of view, the productive efficiency forces the policy makers to rely on a single firm. The monopoly outcome leads to a **loss of allocative efficiency**. The policy makers can re-balance this situation using **regulation**. There are two types of regulations:

- **Ex-post regulation (Antitrust)**: it is collection of laws that regulate the conduct and organization of business corporations to promote fair competition for the benefit of consumer. There two areas of great importance:
 - Anticompetitive practices (e.g. cartels)
 - Abuse of dominant position

Policy punishes anti-competitive behaviors.

- **Ex-ante regulation:** specific regulatory commissions that right from the beginning set invasive and pervasive “rules of the game” (e.g. including prices of final or intermediate services/products).

For utilities, firm is not private (profit goal) but **state-owned** with the objective of maximizing the social welfare. These monopolies are possible with an implicit regulation. Nevertheless, the privatization process of state-owned companies changes their goals introducing the need of an ex-ante regulation stated by a commission (in Italy, AGCOM and AEEGSI).

The regulations impose to the monopolist a price that maximizes the social welfare being more similar to costs. If $p^M=MC$, the firm can recover only variable costs losing the fixed costs. From a social point of view, this is the best solution. Often the regulation imposes $p^M=AC$: therefore, firm does not incur in a loss without having an extra-profitability. From a social point of view, this is the second best solution.

There are two elements reinforcing the need of regulation:

- **Sunk costs:** if market is contestable, even without a regulator, the firm may choose a price near to MC in order to prevent new entrants. Whereas, in monopoly, the sunk costs originate exit and entry barriers. In the case of utilities, it is difficult to recover the non-redeployable investments selling the assets because their future value is uncertain. The reasons of this uncertainty are:
 - **Long-life plants:** it is difficult to predict what will be the market condition at the end of the asset life. Returns from the investment embrace a long period and it is possible that the demand and the preferences change during this time.
 - **Second-hand markets are imperfect:** investments are specific for a given geographical and institutional context. Only the sellers know very well the quality of the asset (information asymmetry). Therefore, the potential acquirers are few and willing to pay a lower price.

Implications:

- **Under-investment:** only state could afford these types of investments.
- **No competition:**
 - Exit barriers: the incumbent company is scared of losing all the investments made in case of exit. Therefore, it is better to remain in the market even if the profitability is low.
 - Entry barriers: the new entrants risk to be locked-in in case of entry and are conscious of the resilience of the incumbent.
- **Inelastic demand:** utilities are necessary for the population without having perfect substitutes. The monopolist optimum price is:

$$\frac{p - MC}{p} = \frac{1}{|\varepsilon|}$$

There are two reasons of competition in the production and sale phases:

- **Demand-side explanation:** we are asking more and more these services causing a movement of demand (shift to the right).

- **Supply-side explanation:** the demand is very far from the MES. In this case, there is not a natural monopoly but competition.

Firms strategy

Companies should try to make their ***DD as rigid as possible***, getting closer to a monopoly:

- Promoting and **advertising** the brand, the product or the service;
- **Differentiating** the offering (for example through innovation);
- Choosing the **proper portfolio of products**, including complementary products in order to widen the product range.

For instance, especially in the High-Tech and ICT industry, some companies prevent the interoperability between different systems.

PRICE DISCRIMINATION

Price discrimination is the ***capacity of firms to price consumers differently*** considering what they want to purchase.

The increasing ***use of Internet*** as a market place has led to think that we would enter to an area of perfectly competitive market. The two main reasons lie in the demand side, where we can observe information enhancements, and in the supply side, with information improvements. There are ***five central assumptions in a perfectly competitive environment***:

1. Atomicity;
2. Product homogeneity;
3. Perfect information: every agent (firms, consumers) knows the price charged by every firm within the market;
4. Technological symmetry: every firm has access to the available production technologies;
5. No entry and exit barriers.

Moreover, in Internet ***there are not the so called “menu costs”***: changing the price of a good costs just a click and does not require printing a menu. Thanks to ***low geographic barriers*** and to the massive use of Internet, markets are becoming more and more global. Has perfect competition become more common? The answer is not really. This is because ***the great variability of prices of goods sold in Internet***. Why for many homogeneous goods sold in the web, we do not observe a unique price and perfect competitive outcomes? There are five reasons:

1. In internet ***brand reputation matters and differentiates firms***;
2. ***Service-premium strategies*** enable firms to differentiate homogeneous products. Some examples are delivery time capability, refund policies, ease of ordering through websites, and so far so forth;
3. This increasing information can enable ***more collusive behaviors*** between firms;

4. **Increased recognizability of consumers** through their web shopping: there are more possibilities of tailoring offers to customers and propose different versions of products allowing price discrimination strategies.

Now we can analyze in details all these five reasons.

Brand reputation

We have to take into account that all transactions in internet occur with a physical and temporal distance among buyer, seller and product, so **trust between buyer and seller** is extremely important. Trust can be enhanced by the good reputation of a firm.

Services premium strategies

In Internet, we can find several portals that compare prices of the same goods sold on different websites. For instance, if we are looking for a certain book and we find out that this book is available in different websites at different prices, why should we buy it at the highest one? Moreover, why does the company price it higher? The reason is that the **same product can be differentiated by the use of service-premium strategies** (customers support, ability to deliver on time...). The more a firm invests in these services, the more the product will be differentiated, causing a higher price.

Collusion

From a research about the monitoring of prices of the same book charged by two firms (Amazon and Barnes and Noble), we can observe how these two companies have raised prices simultaneously. This study has shown how prices were low until the book was between the best sellers, but when the book has begun to be sold less, the price has been raised. This is a typical temporal dynamic of a market, in which big players are colluding. Why? The reason is rooted in a famous article of Rotemberg and Saloner, 1986. The main intuition is that firms collude on price in periods of:

- **High demand:** high incentive to set a lower price than the cartel (higher profits at the expense of the other firms). In order to make the cartel sustainable, the price has to be reduced. In order to reduce it, the firms have to deviate from the cartel;
- **Low demand:** low incentive to deviate from the cartel (lower profits at the expense of the other firms). Collusive prices can be raised since incentive to deviate is low and the cartel is still sustainable.

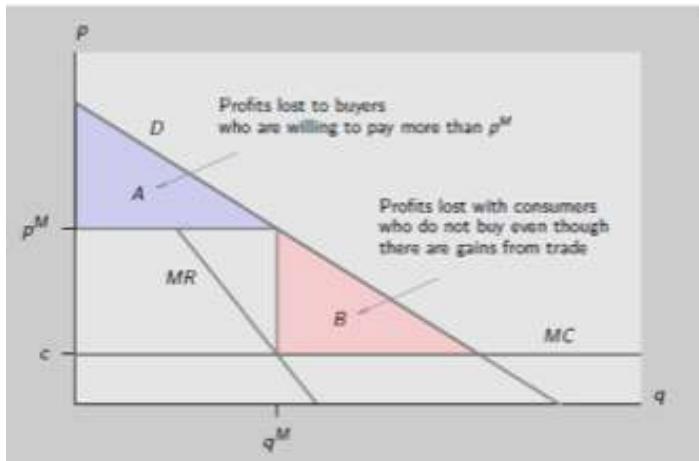
Increased recognizability of consumers and possibilities for personalized offers

All this leads to implement price discrimination strategies.

Why do firms want to discriminate, if they have the capability to do so?

Because **if they set a unique price that maximizes profits**, they incur in two types of losses:

- **Losses from buyers** who are willing to pay more than what they actually pay (triangolo lilla);
- **Losses from consumers** who do not buy the good because the price is too high for them (triangolo rosa).



Perfect price discrimination – 1st grade

It consists in pricing each consumer with a different price for each unit sold, **price that exactly matches his/her willingness to pay** (under the hypothesis that each consumer buys only 1 unit).

This way, the firm is able to recover both losses: the producer's surplus is maximized, the social welfare the same (so this is an efficient solution). The problem is that **there is not the consumer's surplus**: there are no differences between price and willingness to pay. Since all the pie goes to the firm, this type of discrimination is considered a **sort of utopia**, a highly theoretical possibility but hard to be developed in reality. This is because a firm is unable to know in advance the willingness to pay of each consumer. Secondly, even if this becomes possible, it is impossible to prevent bad behaviors of consumers (for instance, they can buy at a low price and then resell the good at a higher one).

Internet has made this discrimination less utopian, because through the web it is easier to know the willingness to pay.

Price discrimination – 3rd grade

It consists in setting **different prices for different groups of customers** (group pricing).

Within the same group, prices are homogeneous.

Sometimes it is convenient to subdivide the market considering some exogenous indicators:

- Age;
- Occupation;
- Geography.

Firms aim at **maximizing their profits in each submarket**. Since each submarket must have a **different elasticity**, this means different prices. Companies have to set **higher prices where the demand is inelastic**.

The trick is to apply the elasticity rule to each market segment:

- In sub-market i : max $\pi_i = RT(q_i) - CT(q_i)$ o $p_i(q_i) \cdot q_i - CT(q_i)$

$$\text{first order condition: } p_i + q_i \cdot \frac{\partial p_i}{\partial q_i} = MC \longrightarrow p_i \left[1 + \frac{q_i}{p_i} \frac{\partial p_i}{\partial q_i} \right] = MC$$

$$p_i \left[1 - \frac{1}{|\varepsilon_i|} \right] = MC \longrightarrow \frac{p_i - MC}{p_i} = \frac{1}{|\varepsilon_i|}$$

- In sub-market j repeat the same and obtain: $\frac{p_j - MC}{p_j} = \frac{1}{|\varepsilon_j|}$

Exercise:

Suppose we have two sub-markets, one with an inelastic demand, and the other with elastic one. We call UPTOWN the richer sub-market and DOWNTOWN the poorer one.

P_U	Q_U	TR_U	MR_U	P_D	Q_D	TR_D	MR_D
8	0	0	-	8	0	0	-
7	1	7	7	7	0	0	-
6	2	12	5	6	0	0	-
5	3	15	3*	5	0	0	-
4	4	16	1	4	1	4	4
3	5	15	-1	3	2	6	2
2	6	12	-3	2	3	6	0
1	7	7	-5	1	4	4	-2

P_U : price for the Uptown sub-market;

Q_U : quantity bought in the Uptown sub-market with a certain price;

TR_U : total revenues in the Uptown market;

MR_U : marginal revenues in the Uptown market.

*: $MR = TR_{i+1} - TR_i$ (MR_U al prezzo di 5 sono pari a 15-12)

Suppose that there are no fixed costs and that marginal costs are constant for each market and equal to 2.

Scenario 1: price discrimination (3rd) is possible

In each sub-market we want to maximize the profit. So $MR = MC$.

The firm will produce until $MR > MC$. Then, when $MR < MC$ that unit will not be produced anymore. The optimal choice will be setting a price of 5 in the U market (with a MR of 3, then MR will be of 1, so it is not good), and a price of 3 in the D market (with a MR of 2).

The profits will be:

$$TR_U = 15$$

$$TC_U = 3 \times 2 = 6$$

$$\Pi_u = 15 - 6 = 9$$

$$TR_D = 6$$

$$TC_D = 2 \times 2 = 4$$

$$\Pi_D = 6 - 4 = 2$$

$$\text{Total profits achievable} = \Pi_{UD} = 9 + 2 = 11$$

Scenario B: price discrimination is NOT possible

P _T	Q _T	TR _T	MR _T
8	0	0	-
7	1	7	7
6	2	12	5
5	3	15	3
4	5	20	5
3	7	21	1
2	9	18	-3
1	11	11	-7

MC=2 and constant

The optimal price is 4 (after, MR=1, so it would be a loss).

$$TR_T = 20$$

$$TC_T = 2 \times 5 = 10$$

$$\Pi_T = 20 - 10 = 10 (< \Pi_{UD} = 11)$$

We can notice that the 3rd type of price discrimination when possible brings higher profits. It is important that price in markets with inelastic demand should be higher than price in markets with elastic demand.

Perfect price discrimination – 2st grade

It is characterized by **self-selection by consumers**: seller cannot directly identify consumer type, but can still induce consumers to distinguish themselves. This selection may be based on the willingness of consumers to consume:

- Different quantities (so price paid by consumers depends on the quantity of the good consumed: **non-linear pricing**);
- Different versions of the same product (**Versioning**).

A typical non-linear pricing technique is the two-part tariff:

- With identical consumers (same demand curve), the firm may obtain the maximum surplus possible (same as perfect discrimination)
- With heterogeneous consumers, the firm will opt for multiple two-part tariffs. If there are 2 typologies of consumers with $CS_2(p) > CS_1(p)$, we will have: $A_1 < A_2$ e $p_1 > p_2$

Versioning

The aim is to sell ***more than one version of the same product at different prices*** targeting different segments of consumers.

This applies also when the low quality version has the same cost of production of the high quality one (or even higher):

- Software: basic version obtained by degrading the premium through the disablement (bearing some costs) of some functions;
- Information services about share prices: the delayed version is produced with some additional costs with respect to the immediate version.

Companies leave that consumers choose the version they prefer (self-selection).

In order to make a profitable versioning, there is a key constraint (***incentive***): you ***cannot make the inexpensive version too attractive to those willing to pay more***. Therefore, companies need to lower the price of the premium version and lower the quality of the basic one.

One additional constraint (***participation***) refers to the fact that the cheap version must be sufficiently cheap that low types are willing to purchase.

The second additional constraint (***good design***): it is impossible for consumers to transform the basic into the premium version.

What about consumer welfare? Is versioning good or bad for consumers?

It is difficult to say a priori. The answer crucially depends on whether ***versioning is able to enlarge the customer base or not:***

- ✓ If the number of consumers increase, the consumer's surplus may not be inferior with versioning;
- ✓ If the number of consumers does not increase, then versioning enables the producer to gain more at the expense of consumers. In this case, versioning leads to a lower consumers' surplus.

There are not general rules to ***decide how many versions*** a company should implement. More versions lead to:

- ❖ More possibilities to capture all the value from consumers;
- ❖ More personalization costs (which presumably are convex in the number of versions);
- ❖ Risk of cluttering effect for potential consumers.

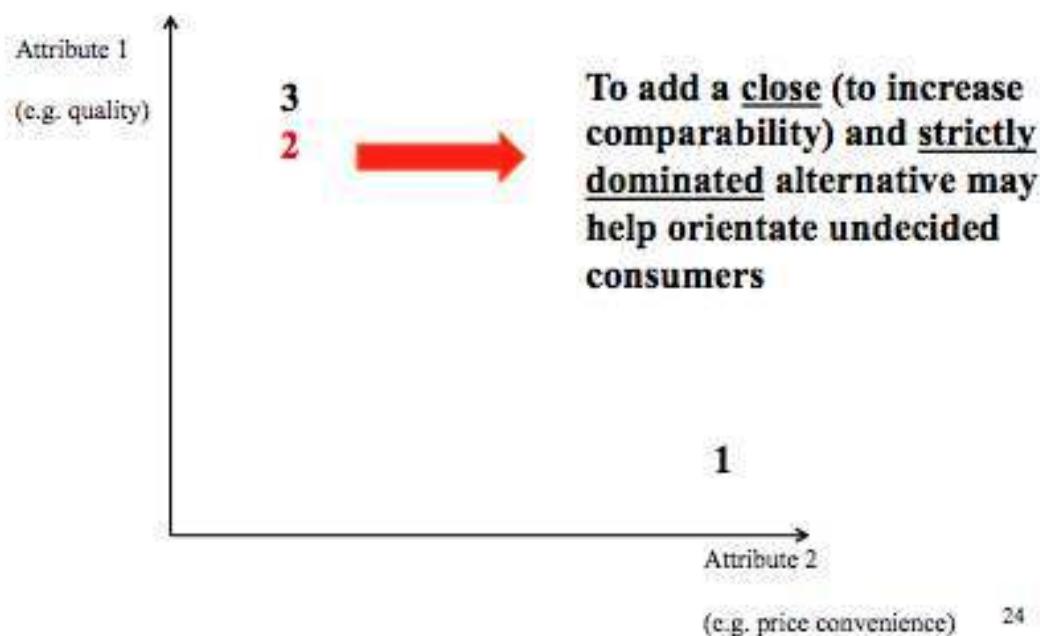
Shapiro and Varian said that "if you cannot decide how many versions to have, choose three." 3 can be better than 2 for the phenomenon of ***extremeness aversion***: that is the risk that 2 versions (heavy

and light) are felt by potential consumers as too big or too small, with the risk that a high percentage will opt for the light version, thus generating less revenues for the company.

Adding another category (super gold premium) and making the previous premium version as the medium one can produce some advantages.

If switching from 2 to 3 is not possible for some reasons, note that a firm ***might always artificially increase the number of versions*** and sell the preferred version exploiting the ***cognitive inclination of consumers***.

We can introduce the ***Decoy effect***:



Bundling

From microeconomic analysis, we know that ***pricing policies of complementary goods should be coordinated*** since the price of an item influences the demand for the complementary item and vice versa.

It is optimal for a firm ***pricing an item below its marginal costs for stimulating the demand of the complementary good***. This way the firm active in both markets is able to internalize the demand spillovers that a firm only active in one market is unable to do.

Bundling means ***offering two or more distinct products as a package at a single price***. Two kinds of bundling:

1. **Pure bundling**: sale of the package but not of single components;
2. **Mixed bundling**: sale of the package and of single components (the package price is lower than the sum of prices of single components).

Bundling is ***a sort of 2nd price discrimination and versioning***.

The logics underlying the two types of bundling are similar. We are going to focus on the pure one to make general considerations.

We can say that a profitable bundling ***reduces the willingness to pay dispersion***. When bundling is especially profitable? When the heterogeneous and negatively correlated willingness to pay of consumers for single components (someone prefers more one component, someone else another one) but overall, willingness to pay of consumers for the two products should be similar.

The dispersion of the ***sum of the willingness to pay single components has to be greater than the dispersion of the willingness to pay the bundle***.

Bundling is not always profitable. When a small group of potential consumers shows ***relatively a very low willingness to pay (the bundle and single components)***, trying to target these consumers and sell them the bundle may not be convenient.

EXTERNALITIES

An externality (spillover) is a cost or a benefit imposed upon someone by actions taken by others (with no compensation)

- An externally imposing benefit is ***a positive externality*** (for example scientific advance, a well-maintained property next door that raises the market value of your property and a network externality).
- An externally imposing cost is ***a negative externality*** (for example air and water pollution, traffic congestion, loud parties next door, second-hand cigarettes smoke).

There are two types of externality:

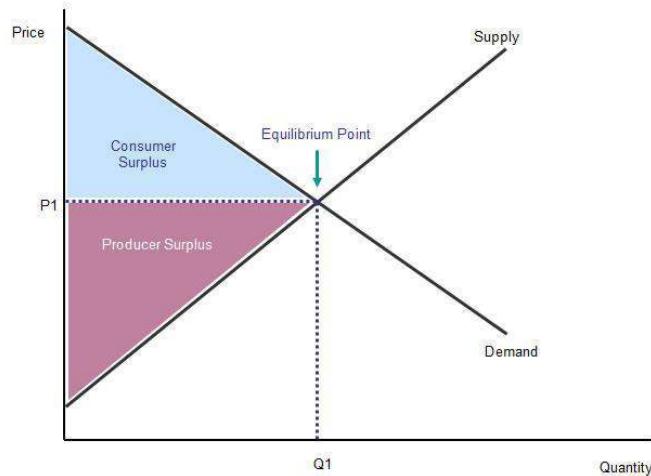
- ***Consumption externalities***: the consumption of a good by agent A has a direct impact on agent B's utility (for instance smoking, loud music, tidy garden).
- ***Production externalities***: production actions by agent A have a direct impact on agent B's utility (for instance bee-keeper and apple orchard, polluting firm and fisherman).

Externalities causes ***Pareto inefficiency***:

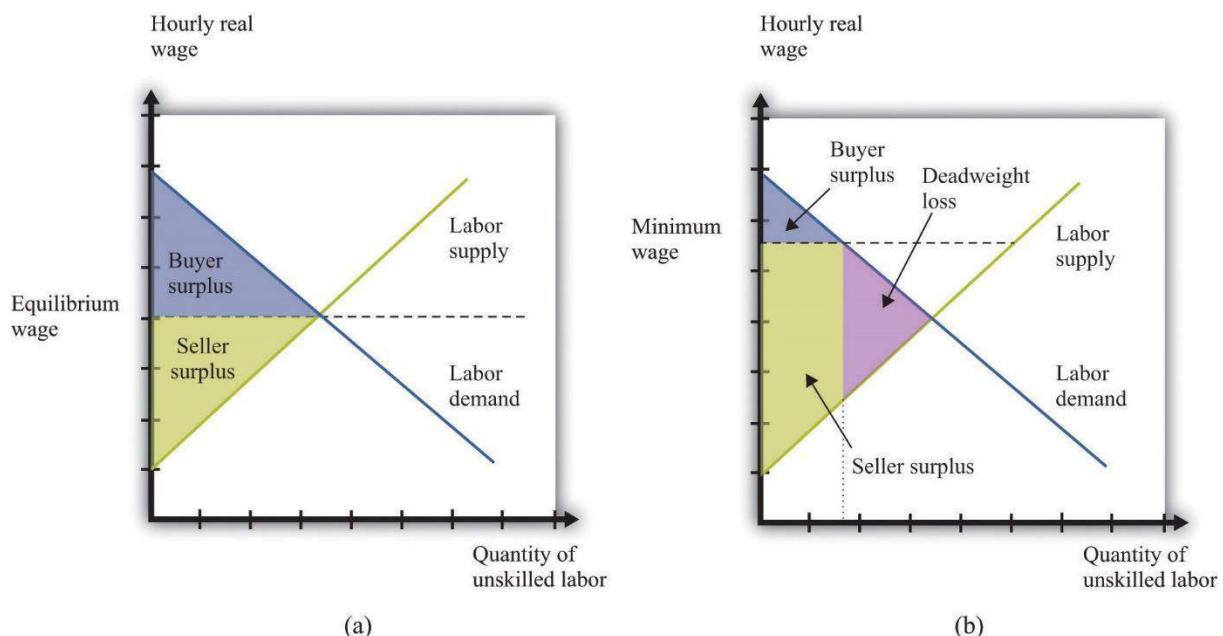
- Too much scarce resource is allocated to an activity that causes a negative externality
- Too little resource is allocated to an activity that causes a positive externality.

The problem is that the externalities are not internalized in the economic system.

In **perfect competition**, we have the maximum social welfare. This means that the **supply curve** reflects the **private marginal cost (PMC)**, which is equal to the **social marginal cost (SMC)**. Moreover, the **demand curve** reflects the **private marginal benefit (PMB)**, which is equal to the **social marginal benefit (SMC)**.



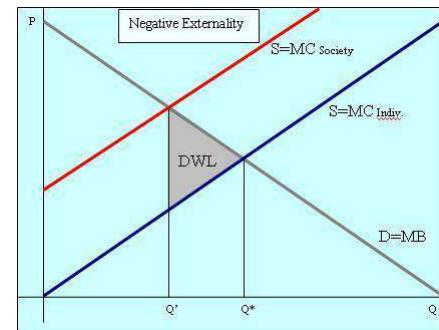
The **market power** is the **first cause of inefficiency** because the consumer surplus decreases and there is a **deadweight (welfare) loss**, which means something not produced.



Returning to perfect competition, the producer surplus is the same but the consumer surplus is more because the **total welfare area is much bigger**.

The second source of market imperfection is the externality.

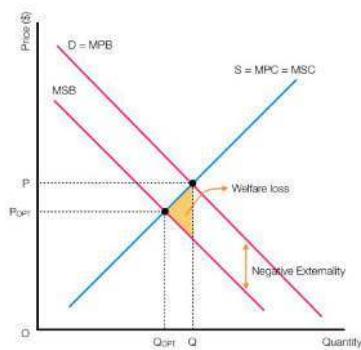
- **Negative production externality:** a good produced exert an additional cost to the community. This means that the supply curve is not still equal to the social marginal cost but becomes the sum of the private marginal cost and a public marginal cost caused by the externality. The supply curve shift to the left and the equilibrium is not a social optimum.



- **Positive production externality:** a good produced exert a benefit to the community, which is not considered in the private marginal cost. This leads to a shift of the supply curve to the right.

- **Positive consumption externality:** a good consumed exert a benefit to the community causing a shift of the demand curve to the right.

- **Negative consumption externality:** we have a welfare loss for consuming a good too much. It is related to a loss of social marginal benefit and a shift of the demand curve to the left. For instance, the consumption of alcohol causes violence and vandalism that is negative for the entire community.



Generally, there is ***no match between the private marginal cost/benefit and the social marginal cost/benefit***. Private decisions have negative/positive impacts on the society as a whole without costs or compensation.

Coase's theorem:

- (Weak version: efficiency proposition) When parties can **bargain without cost** (no transaction costs), ***the resulting outcome will be efficient***, regardless of ***how the property rights are specified*** (e.g. Polinsky, 1974, pp 1665, 'Economic Analysis as a Potentially Defective Product: A Buyer's Guide to Posner's Economic Analysis of Law', 87 Harvard Law Review).
- (Strong version: efficiency + invariance propositions): When parties can bargain without cost, the resulting outcome will be efficient and ***the level of the externality generated the same***, regardless of how the property rights are specified.

Ronald Coase's intuition was that most externality problems are due to an ***inadequate specification of property rights*** and, consequently, an ***absence of markets*** in which ***trade can be used to internalize external costs or benefits***.

The strong version of the theorem is hard to find in reality. There is only one case where we can identify the strong version, which is the case of quasi-linear preferences. It means a particular shape of indifference curves:

$$U(x_1, x_2) = x_1 + v(x_2)$$

One good enters the utility function in a non-linear way that causes changes in the demand curves, but it is a very uncommon condition.

The weak version is more possible. The key message is that **Pareto improvements are possible** when there are ways that allow the economic system to **internalize the externality through market mechanism**. There are two ways:

- 1) **Bargaining of consumers** (Edgeworth box): if there are two roommates, one is a smoker and the other is a non-smoker, their endowment is composed by money and smoke/clean air. They can bargain: the smoker will pay for a unit of smoke; the non-smoker will pay for a unit of smoke reduction.
- 2) **Merging of firms** (steel mill and fishery example)

This theorem refers to two assumptions:

- No bargaining costs
- Small numbers involved

If the numbers involved are larger (externalities affect society at a large scale) and bargaining costs are high, **public policy has a more invasive role** than just assigning property rights. It has an active role to make the market move. There are two possible approaches:

- **Command and control**: the government imposes a limit to a negative production externality. For example, it imposes a limit to the pollution produced by a firm.
- **Pigouvian taxes** (or subsidy): the government imposes taxes on the units of negative externality produced. For example, it imposes a tax on the units of pollution. This tax should be equal to the cost of each unit of pollution, not considered before (additional social marginal cost). Producing more pollution, the firm incurs in a tax. In this way, the private marginal cost returns equal to the social marginal cost. The problem is the difficulty to estimate the cost produced to the society and therefore the right value of the tax. On the contrary, if a firm produces a benefit for the society, the government has to incentive this production assigning a subsidy to the firm (for example production of clean technology).
- **Tradable permits**: it means to create an artificial market in the spirit of the Coase's theorem. For example, the government puts a limit to the pollution of firms (command and control) and gives permits to pollute. In this way, it creates a market of permits. The firms have a willingness to pay and sell the permits. If the willingness to pay of one firm is greater than the willingness to sell of the other, both parties have an interest to bargain in order to exchange permits. Therefore, it is a market. Following the dynamic of this market, costs of reducing pollution goes to company more efficient in doing that.

Public goods

The public good is a positive externality stretched at the maximum. It is characterized by:

- **Indivisibility**
- **Non-excludability** from consumption: once a consumer has bought a public good, he cannot exclude other consumers from the consumption.
- **Non-rivalry**: if a consumer consumes a unit of a public good, he does not hurt the consumption of others.

A public good produces a **great amount of positive externalities**: if someone buys a public good, everybody can enjoy it. In order to understand if a good is a public good, we have to analyze if we have **the technical capability to exclude non-payers** from the good consumption. If it is technically possible, we have to understand **if it is economically feasible**. Two examples are the defense and the basic science.

The problem is why should I bear a cost for a public good? If someone else buys it, I can benefit from it without paying. There is a **reservation price**, which is the price a consumer is **willing to pay in order to buy the public good**. The consumer will buy the public good, only if this willingness to pay is greater than the cost of the public good (**Pareto dominant solution**).

In the real world, the reservation price of each person depends on his personal wealth. Whether or not to provide a public good will depend on reservation prices and distribution on **wealth among the members of a community**. If someone has a willingness to pay higher than the cost of the public good and has all the wealth, the acquisition of the public good would be a Pareto improvement. On the contrary, if someone has a willingness to pay lower than the cost of the public good and has all the wealth, the Pareto efficient solution is not to provide the public good.

This is the **free-riding problem**: a free rider is a person who receives the benefit of a good but avoids paying it. Using the game theory, we can demonstrate that the free-riding problem leads to no provision of the public good, even if it is Pareto efficient. **Only government can remedy** to this problem:

- Providing the public good and paying it with **tax revenues**
- Enforcing the **excludability mechanisms** (transforming the public goods into club goods)

Asymmetric information

In purely competitive markets, all agents have full information about traded commodities and other aspects of the market. There is no perfect information for all products: for example, a doctor knows more about medical services than the buyer does. Imperfectly informed markets with one side better informed than the other are characterized by asymmetric information.

Asymmetric information affects the dynamic of a market in different ways:

- 1) **Adverse selection**: it is a problem of hidden information. In a market, there can be good or bad products. The problem is that there is an **externality** between the sellers of good products and bad products; when an individual decides to try to **sell a bad product**, it **affects the purchasers' perceptions of the quality of the average product on the market**. This lowers the price that they are willing to pay for the average product, and thus hurts the

people who are trying to sell good products. This type of externality leads to a ***market failure***. For example, in the market of innovation finance, there are good or bad innovation projects. Banks or other providers of finance for the project do not know the quality of the project. Due to this risk, in the equilibrium, banks pose unfavorable conditions for lending, such as high interest rates. The capable innovators may prefer to give up searching for external debt finance (discouraged borrowers) or search for other financing sources.

- 2) **Moral hazard:** in a transaction, it occurs when ***the party with more information*** about its actions or intentions has a tendency or incentive to ***behave inappropriately*** from the perspective of the party with less information. Moral hazard is a problem of ***hidden actions***: if the actions of the party with more information were observable, there would be no problem. It is a market failure because the moral hazard leads to ***less trade than the optimum***. Following the previous example, a bank cannot control strategies and choices of the innovators. Therefore, the innovators may put in place actions not in the interest of investors. The solution can be the imposition of milestones to control how the project is going on, or a co-investment in order to guarantee the high effort of the innovators.

Adverse selection is an outcome of an informational deficiency. High-quality sellers can improve information through signals of high quality, such as warranties, references from previous customers and professional credentials. Companies sustain the cost of signaling in order to improve information, for example patents for innovative companies. Signaling allows to reach a ***pooling equilibrium***, which means the total output of the company does not change but resources are wasted in order to signal. For example, if a company has to distinguish between good workers and bad workers, a signaling mechanism is the education. Assuming that education does not change the output produced by bad/good workers, it is a pooling equilibrium because the firm wasted resources in education in order to signal without being able to separate bad and good workers. It means that the firm is not able to reach a ***separating equilibrium***. It hurts the market efficiency. An alternative solution is to put incentives, which allow the company to go from a pooling equilibrium to a separating equilibrium. The incentives have the purpose to make the workers' type reveal itself.

CONCENTRATION INDICES

The concentration indices are useful to understand **whether a market is concentrated**. Understanding the degree of concentration is necessary to **policy makers' decisions**.

Concentration refers to the number and size (or size distribution) of the firms belonging to one industry.

- **Concentrated industry:** very few firms or many firms with few of very large size;
- **Dispersed industry:** many firms without a dominant position.

Concentration has a deep effect on firms' behavior and offers indications on competition mechanisms.

The **size of a firm** can be computed according to several dimensions, such as:

- **Number of employees:** it is not sufficient because the firm could be inefficient needing more workforce.
- **Output:** it is difficult to compare different outputs, understanding whether they are substitutes.
- **Turnover:** it is different among different industries, for instance in the cosmetic industry there are high margins with low turnovers.
- **Added value**

The **market share** of a company is the ratio between the turnover/output of the firm i and the total turnover/output of the industry j. If there are N firms operating in the market:

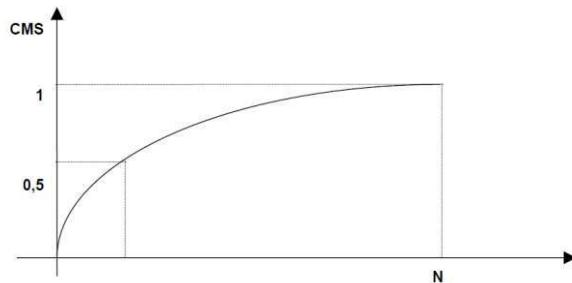
$$s_{ij} = \frac{q_{ij}}{\sum_{k=1}^N q_{kj}} \quad 0 \leq s_{ij} \leq 1 \quad \text{and} \quad \sum_{k=1}^N s_{kj} = 1$$

The **concentration vector** has the market shares as composing elements ordered decreasingly.

$$S = (s_1, s_2, \dots, s_N)$$

The **concentration curve** is represented in the Cartesian plan (for a certain industry at time t)

- Horizontal axis: firms in decreasing order of size (from the largest to the smallest)
- Vertical axis: Cumulated of Market Shares (CMS)



Concentration indices synthetize in one index the entire information contained in the concentration vector, facilitating the comparison between the degree of concentration in different periods and industries. They can be:

- **Absolute indices:** weighted sum of market shares. Using different weights, we identify different concentration indices.

$$I = \sum_{i=1}^N h_i(s_i) s_i \quad \text{with } h_i(s_i) = \text{weight}$$

- **Relative indices:** they refer to the degree of inequality between firm sizes within a certain industry.

Analyzing the **absolute concentration indices**:

- **Concentration ratio:** it is the sum of the market share of the first k firms in the industry in increasing order.

$$C_k = \sum_{i=1}^k s_i$$

The weight $h_i(s_i)$ is equal to 1 for the first k firms

The weight $h_i(s_i)$ is equal to 0 for the remaining N-k firms

As rule of thumb k is usually equal to 3, 4, 8, 20

The higher the concentration ration, the higher the concentration in the industry.

$$0 < C_k \leq 1$$

- C_k is equal to 1 when the first k firms serve the entire market
- C_k is equal to 0 in perfect competition

There are two problems related to this ratio:

- The arbitrary choice of k;
- The provision of the same result for industries with different concentration. Considering a certain number of firms can lead to lose important information.

- **Herfindal index:** it is the weighted sum of the market shares of all the firms in one industry. The share itself gives its weight.

$$HI = \sum_{i=1}^N s_i^2$$

In the sum of the square of the market shares, smaller firms contribute very little to the final value of the index.

$$\frac{1}{N} \leq HI \leq 1$$

- if $HI = \frac{1}{N}$: there are N firms with the same size

The market share of each firm is 1/N of the market

In perfect competition $HI=0$ ($N \rightarrow \infty$)

- If $HI=1$: there is one monopolistic firm

- **Entropy index:** it is a weighted sum of the market shares of all firms. The weight of each share is equal to the logarithm of the inverse of the market share considered.

$$EI = \sum_{i=1}^N s_i \cdot \ln\left(\frac{1}{s_i}\right)$$

Smaller firms provide a greater amount contribution to the total amount of the index.

$$EI = \sum_{i=1}^N 1 \cdot \ln\left(\frac{1}{1}\right) = \sum_{i=1}^N 1 \cdot 0 = 0$$

$$EI = \sum_{i=1}^N \frac{1}{N} \cdot \ln\left(\frac{1}{\frac{1}{n}}\right) = \frac{1}{N} \cdot N \ln N = \ln N$$

The entropy index provides **a measure of the disorder**: in industries where there is only one firm, the entropy is minimal. In industries where there are several identical firms, the entropy increases together with the increase of N.

Concentration indices are used to **compare the degree of concentration**:

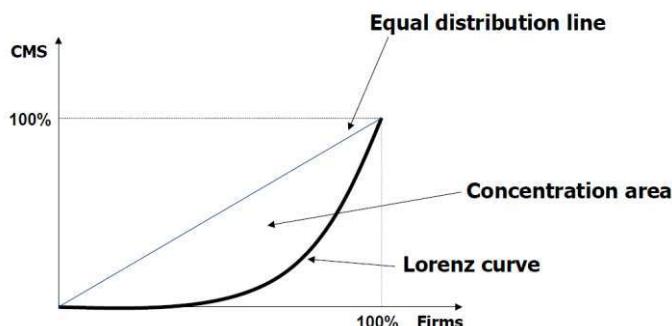
- Between **different industries at the same time**;
- In **the same industry across time**.

The **degree of disaggregation** needs to be the same in order to compare the results. It means to understand if companies sell the same product to consider them in the same industry. The concentration depends on the degree of disaggregation of the industry: if disaggregation increases, concentration increases.

The degree of disaggregation is defined relying on **the classification of production activities**. There are several national and international classifications of production activities allowing to define the industries adopting a common criterion (for example ATECO in Italy, NACE in Europe, SIC in the US).

Analyzing the **relative concentration indices**, we can use two methods, adopted in labor economics in order to measure income inequality and income distribution:

- **Lorenz curve:**



On the horizontal axis, we have the firms ordered by increasing size. On the vertical axis, we have the cumulated market share.

The Lorenz curve lies between the **absolute equality line**, where firms size and output coincide (the first 20% of firms produces the 20% of output), and the **maximum inequality**, where there is one firm producing the entire industry. The higher the concentration area, the higher the degree of concentration of the industry.

- **Gini index:** looking at the Lorenz curve, it is the ratio between the concentration area and the area underlying the diagonal.

The Gini index varies between 0 and 1:

- All the firms have the same size:
 - The equal distribution line and the Lorenz curve coincide
 - The concentration area is null
 - The Gini index is equal to 0
- One firm produces for the entire industry
 - The concentration area coincides with area underlying the diagonal
 - The Gini index is equal to 1

There is the ***Encava and Jacquemin qualitative classification*** that synthetizes the industry concentration in a finite number of categories:

1. **Monopoly:** there is one firm with a market share higher than 80%;
2. **Dominating firm:** there is one firm with a market share between 50% and 80%, and the others are much smaller;
3. **Duopoly:** 2 firms of similar size control 80% of the market;
4. **Asymmetric oligopoly:** 3 or 4 firms control 80% of the market, the highest share is around 40%;
5. **Symmetric oligopoly:** 3 or 4 firms equally control 80% of the market;
6. **Asymmetric competition:** the largest firm holds a market share between 20% and 50%;
7. **Symmetric competition:** the largest firm controls 20% of the market maximum;

This classification is only a point of reference: the situation could be different, for example, one firm with a market share of 75% could be a monopolist.

TRANSACTION COSTS

The transaction costs are a market imperfection that explains the firms' growth strategy. In the neoclassical vision of markets, the exchange of goods between firms and consumers is ***without frictions or costs in the market***. If markets are perfectly competitive, the economic system may reach Pareto optimal allocations. The ***Transaction Cost Economics (TCE)*** doubts the existence of these "perfect" markets: ***transactions are not instantaneous*** and the ***use of markets is costly***: time & money to search for sellers & buyers, negotiate exchange terms, write contracts, inspect results, and enforce deals. In many circumstances, ***alternative governance structures***, such as the firm, could be more efficient than the market.

The three pillars of the TCE are the reasons why markets are costly:

1. Agents are characterized by bounded rationality;
2. Opportunism, not of all agents but of a large part;
3. Relationship-specific investments.

Without all these three conditions the market works efficiently, otherwise the use of market implies substantial costs.

1. **Bounded rationality:** Bounded rational agents **experience limits** in **formulating and solving complex problems and in processing** (receiving, storing, retrieving, transmitting) **information** (Herbert Simon). Does bounded rationality mean that people (and therefore their actions) are irrational? Not at all. People making choices are **intendedly rational**. They want to make rational decisions, but they cannot always do so (Bryan D. Jones).

In decision-making, rationality of individuals is limited by:

- The **information** they have: the economic agent does not know perfectly the present state of the nature and he is not able to forecast all the possible states of nature on which he will finds himself after a decision (**substantial limit** caused by environmental uncertainty).
- The **cognitive limitations** of their minds: even if it is possible to know the present and forecast all the possible future states of nature, there is not an algorithm enabling to find optimal solutions in a reasonable time. There is a “cognitive” and “time” problem (**procedural limit**). Agents rely on **heuristics**: if a decision unsatisfied them, they will not take it again. They base decisions on past experience and approximations for their behavior.

In the TCE, the **bounded rationality** implies two important facts. If two or more parties want to **regulate a transaction by a contract**:

- It is very difficult to **include in the contract all possible details** of interest for the object of the contract: they cannot include every likely (and less likely) circumstance of interest in the present and in the relevant future.
- Even if this could be possible, the parties **are not able (or it would be too costly) to negotiate** every one of these single details (procedural limit).

Therefore, **contracts are incomplete**: the parties of a contract have and can exert a certain **degree of discretion** over the fulfillment of contractual clauses. They have discretionary power of interpretation.

2. **Opportunism:** some economic agents pursue their own utility and interest even if this is detrimental to the utility of someone else and are selfish. Not all economic agents are opportunists, but it is enough that some of them behave opportunistically in order to incur in **additional costs of transaction**.

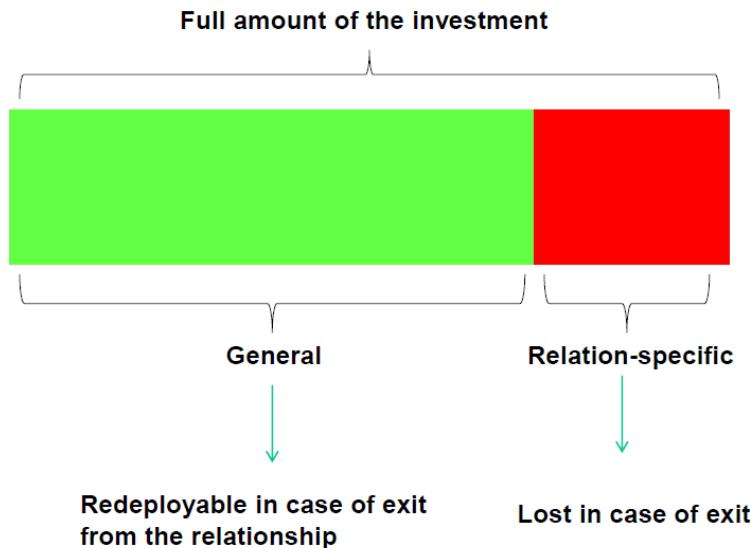
Ex-ante it is not possible to distinguish between opportunistic and not opportunistic economic agents. Opportunistic behaviors can appear during the writing of the contract or also after.

- Ex ante opportunism: **hidden information**;
- Ex post opportunism: **hidden actions**.

3. **Relationship-specific investments:** investments that are specific to the relationship between the parties and to the nature of the counterparts. There are three more important types of relationship-specific investments:

- **Dedicated (often physical) asset specificity:** it refers to the supplier-buyer relationship along the supply chain. There are **assets** whose physical and engineering properties are **specifically tailored to a particular transaction**. For example, supplier produces something only for one buyer making specific investments.
- **Site specificity:** it refers to assets that are located **side-by-side to economize on transportation or inventory costs or to take advantage of processing efficiencies**. For instance, the producer decides to locate its facilities near to one buyer in order to facilitate just-in-time processes, but there are no other firms in the surroundings to whom selling the goods.
- **Human asset specificity:** Some employees of the firms engaged in the transaction may have to acquire **relationship-specific skills, know-how and information**. For example, a company buys a software, trains its employees for its usage, but only a specific software house sell, and teach this software.

Considering the full amount of investment, there could be a general part and a relationship-specific part.



In a transaction cost perspective, most of investments should have at least a relationship-specific part. The relation-specific part of the investment (red area) is **not recognized outside the relationship**. The more the red area is large, the more a firm would be reluctant to exit from the relationship (once entered). The more asymmetric the red areas are between parties, the more the firm facing low relationship-specific investment may **"exploit" the firm facing high relationship-specific investment**. If one company has only the green area, it does not care a lot about the relationship having a stronger position.

A **Rent** is the **profit a company expect to get from a contractual relationship** if everything goes as planned. For example, the supplier agrees with a buyer for a certain quantity to produce: the supplier's alternative is to sell the same quantity on the market. The supplier will accept these terms as long as **extra-profits are non-negative**: in particular if the price agreed with the buyer is greater than the price on the market. Nevertheless, **the buyer could be interested in exploiting the red area** after the contract, **acting opportunistically**. Considering that, we could calculate the **quasi-rent**, which is the extra profit that one gets if the deal goes ahead as planned, versus the profit one would

get if he had to turn to his next-best alternative (in our case Market). Once agreed a price higher than the one on the market, the buyer could act opportunistically ***renegotiating the contract clauses in order to obtain a lower price***. The buyer will try to lower the price until the supplier is willing to stay in the relationship, which means until its profits are greater than the profits gained on the market, even if negative. This is the ***hold-up problem***.

Contracts are not able to prevent companies from the hold-up problem because they are ***incomplete***. For example, the buyer could pretend that the quality of supplied products is not enough, justifying the request of a lower price. The supplier could fight the buyer ***in court for breach of contract***. This is itself a potentially ***expensive move***. Contractual incompleteness leads by definition to ambiguity in contractual terms so the outcome of the trial (even if you are in principle right) cannot be taken for granted. Actually, it can be ***highly uncertain***.

Most of the time the supplier ***is better off accepting buyer's revised offer than not accepting it***. This of course also depends on the extent to which the buyer wants to appropriate the quasi-rents.

If there is not bounded rationality, there are complete contracts. If opportunism does not exist, there is no need of complete contracts. If there are not relationship-specific investments, there is no need to defend from opportunism, because there is no loss in switching from a transaction to others. Thus, if only one of these pillars is not met, the market will still work efficiently.

Would the supplier ever be willing to enter into a market relationship with one buyer? If none is willing to do that, there is one possibility: the buyer produces the good itself.

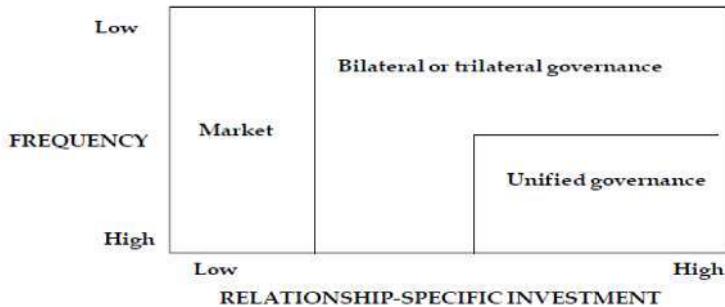
Given that markets are not always perfect and their use may entail substantial costs, the main aim of the TC perspective is to ***individuate for each transaction the best governance structure possible***. There are three alternatives:

- ***Market***, which may present costs (related to the use of contracts);
- ***Hierarchy*** (transactions are internalized and carried out inside the firm, costs arise from the administrative bureaucracy that internalize and monitor the exchanges);
- ***Hybrid forms***;

The most efficient solution (the market vs. planned solutions) is identified through a ***Darwinian selection process*** where the best option finally survive. Note that from this perspective, ***firm is an efficient response to a market failure***.

Williamson classified transactions considering ***two determinants***:

- a. Relation-specific investment needed for carrying out the transactions
- b. Frequency of transactions



The hybrid forms are bilateral or trilateral governance, such as joint ventures and alliances. An increase in uncertainty of transactions (which may affect the degree of incompleteness of contracts) is likely to lead to less hybrid forms.

Example: Fisher body and General motors. During '800, car bodies were made by wood. General motors wanted to create the first metal body. Wooden body were made through short term contracts because there were not specific investments (made by artisans). Metal bodies required specific machines tailored for specific types of car. They required long-term relationships. In 1919, GM and FB agreed for a 10-year contract based on output, price and quality. Due to the explosion of the automobile market, GM decided to re-discuss the contract. GM wanted FB located much closer by in order to control the quality of products because of its dissatisfaction. FB was very reluctant because it would have been a hostage of GM due to the high specific investments. In 1926, GM acquired FB. It is a case of external growth.

Therefore, the options could be an ***internal growth or an external growth through mergers and acquisitions.***

THEORY OF THE FIRM

The ***neoclassical firm*** is a sort of machine: inputs enter a black box and are worked out with an output. The ultimate goal of the firm is to maximize profit. It is a sort of robot. There are two problems: firms are not a black box, there are individuals and therefore different objectives.

Literature tries to address the following problems:

- Why do firms exist remain unresolved;
- The nature, role, characteristics of individuals within the firm and their mutual relationships do not matter for explaining firm's existence and its behavior;
- Within market or industry, firm-specific heterogeneity is not acknowledged;
- Profits remain the sole ultimate goal;

The theory of the firm consists of a number of ***economic and managerial theories*** that investigate the nature of the firm addressing one or more of the problems above: its existence, behaviour and goals.

Firms consist of different kinds of agents: the owners and the providers of inputs (managers, workers). There are two possible approaches investigating what is a firm:

1. **Contractual approach**: there is a bundle of contracts linking owner and providers of inputs. The parties of these contracts have different objectives and the firm is not necessarily an independent subject with its own objectives.
2. **Holistic approach**: the firm goes behind the linkages between owners and providers of inputs. It is an **anthropomorphic vision**: firm is a social unit that overcomes the sum of its parts and it has its own objectives.

Contractual approach

There are two families of theory:

1. Principal-agent theory (agency problem): it consists of the managerial theory of the firm (Baumol) and the team production.
2. Incomplete contracts or property rights theory: it refers to the development of the TCE (Grossman and Hart).

Principal-agent theory

It refers to the **agency problem**: an agent agrees to work for the principal to pursue principal objectives but he has an incentive to pursue his own objectives, not necessarily aligned with those of the principal. The principal incurs **in some costs to control and monitor the agents' behavior**.

According to the Cabral's definition, a principal wants an agent to act in the principal's interest but possesses less information than the agent possesses. It is a problem of **asymmetric information between parties**.

Managerial theories of the firm

In many firms, **management is separated from ownership**:

- Managers want to pursue **their own goals** rather than owners' goals
- Managers ensure an **acceptable levels of profit to satisfy owners**, but tend to pursue goals which enhance their own utility

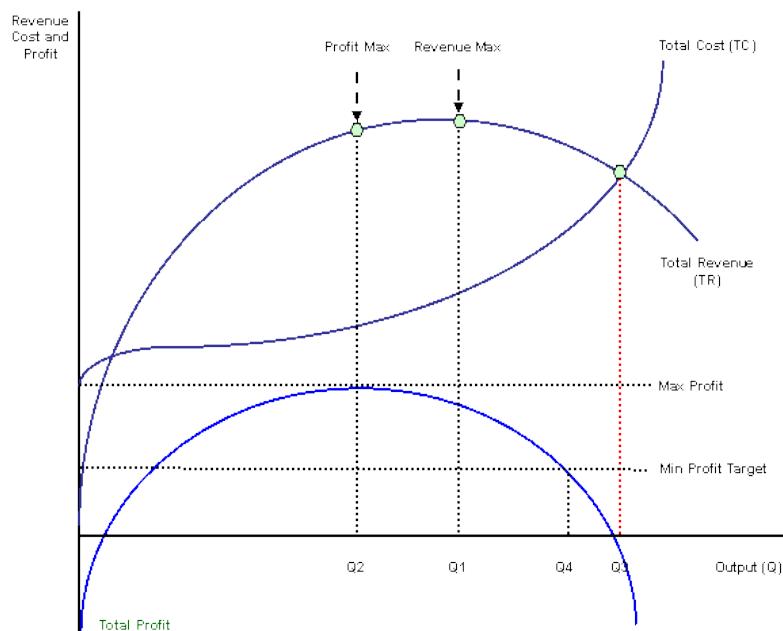
According to that, we can introduce the **Baumol's model of sales revenue maximization** as an **alternative goal** to profit maximization.

Why do **managers** prefer to **maximize sales revenues** (and have a satisfactory level of profits) rather than **profits**?

1. **Compensations of top executives** (Chief Executive Officers, CEOs) rises with firm size (i.e. **sales revenues**);
2. **Large (and growing) sales revenues** enhance a **CEO's prestige** and his/her visibility (they are more likely to be hired by another firm);

The Baumol's model states that:

- Firm's decision-making is limited to a **single period**.
- During this period, the **CEO attempts to maximize revenues**.
- Sales revenue maximization is subject to **provision of a minimum required profit** to ensure a fair dividend to shareholders. Thus assuring the stability of the CEO's job.
- Conventional cost and revenue functions are **total cost increasing** and **downward sloping demand curve**.



The quantity maximizing sales (Q_1) is different from the quantity maximizing profit (Q_2).

If profit for Q_1 is above the minimum profit target, managers choose to produce Q_1 . If this condition does not hold, managers must reduce production.

There are some extensions of the Baumol's model: Marris's model of managerial enterprise (based on growth rather than size) and Williamson's theory of managerial discretion are based on similar intuitions and reach comparable results.

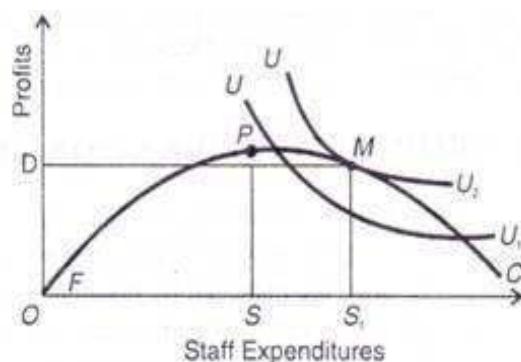


Fig. 3

This leads to two types of firm:

- **Owner-controlled firm:** ownership coincide with control;
- **Managerial firm:** ownership and control happen to be in the hands of two separate sets of agents: shareholders and managers. This is the principal-agent problem.

In a perfect informed world, the outcomes of these two models should be the same. In the real world, there is **asymmetric information between parties**: owners do not know what managers are maximizing because they are not managing the company: they are external. Managers are better informed. Moreover, the ownership fragmentation reduces the shareholders incentive to overcome the information asymmetry.

Summing up, the empirical evidence on the managerial theory is that:

- **Profit rates should be higher in owner-controlled firms** than in managerial firms (public firms).
- **Monitoring by shareholders should be higher when the ownership is more concentrated:** they have a bigger stake in the company.

The empirical evidence is mixed because shareholders of public companies have important **governance mechanisms** at their disposal:

- Board of directors (insider and outsider members);
- Mixed-payment schemes for managers, balancing risks and incentives.

Team production

The **firm** enables resource owners to **increase productivity** through cooperation and specialization. This enhances the output level, which is the joint result of the **combined efforts of all the inputs** working at the same time. The problem is that **individual contribution** of each member of the team to the final output **cannot be isolated** and measured.

The risk is the **free-riding tendency**, which means each member tries to do the less possible. Firm is a way to avoid this tendency being a **monitoring device**.

Without monitoring, each person will shrink and hope for a free ride on the effort of other people. It is necessary an **insider monitor** of the teamwork: the team requires a monitor that puts effort in order to observe the individual members and to check that their efforts is satisfactory.

This monitor needs **some incentives** to bother. If he were simply a member of the team whose job is to check the commitment of the others, the monitor would have the same incentive to shrink as anyone else.

He needs to be a "**residual claimant**". Team members receive a contractual reward in the form of wage, and the monitor receive whatever remains after these payments have been made. The more effectively the team operates, the bigger the residual will be, and hence the monitor will have an interest in promoting the efficiency of the team.

A possible critique to this model is that it considers the entrepreneur as a monitor, a coordinator of factors of production, but it is much more than that. This model is not a solution to the problem of firms' existence: it considers only how firms are created.

Property right theory

It is an evolution of the TCE (Coase and Williamson), according to which firm arise as a result of efficiency coordination, in particular the costs of internal production are lower than the sum of the market price and the transaction costs.

The basic idea of the property right approach (Grossman and Hart) is that contracts are largely incomplete. The incompleteness of contracts means that there are ***non-contractible elements*** due to difficulties in contemplating in advance all possible future contingencies and measuring performance under each contingency.

Firm (i.e. Ownership of assets) conveys the "***residual rights of control***", which are all the decision rights not specified in a contract, and this organization is the only way to rapidly capture business opportunities.

Therefore, firms vertically integrate because it is the only way to exploit opportunities.

Holistic approach

Resource-based view of the firm (RBV)

Profits are often heterogeneous and sticky with-in markets and industries. Firms want to achieve competitive advantage, which is sustainable through time. The source of this competitive advantage is not easily imitable for other firms.

RBV does not challenge the neoclassical profit maximization assumption, but it is rather interested in understanding what a ***sustainable competitive advantage*** is (that originate persistent high profit levels) and how do firms can achieve it.

In the neoclassical firm, gaining profits depends on the characteristics of the market. In RBV, company is a bundle of resources and profits depends on these resources.

The competitive advantage comes from the possession of one resource (core competence, distinctive capability) that other firms do not possess and the way the firm combines the resources in order to reach the best strategic fit with the external environment. These resources could be:

- **Physical:** plant equipment, location, access to raw materials;
- **Human:** training, experience, judgment, decision-making skills, intelligence, relationships, knowledge;
- **Organizational:** culture, formal reporting structures, control systems, coordinating systems, informal relationships.

It is important to consider the ***external environment***, which sets the rules of the game: some resources are key for a wide range of environments some others not. **A resource is not valuable everywhere.** Sometimes it is a matter of ***combining resources*** rather than collecting.

For instance, if a company steals some scientists and creative people from another, it has to integrate these people in their department. These people have to establish a new way of work and make relationships with new peers and departments. Moreover, they have to deal with a completely new corporate culture. Because of these reasons, it is not sufficient to move a resource to another environment in order to have better performances. The company has to find the right balance between its resources.

In order to generate a sustainable competitive advantage, resources should be:

1. ***Rare***: not available for every company;
2. ***Imperfectly mobile***;
3. ***Imperfectly imitable***;
4. With ***no strategically equivalent substitutes***.

Companies need some isolating mechanisms preventing other firms from appropriating the resource:

- ***Path-dependency*** of production processes and resources accumulation (e.g. Brand reputation)
- ***Intellectual property rights (IPR) regime*** (e.g. patents)
- ***Causal ambiguity***: firms do not know what to imitate, for example managerial processes. It is powerful when company with competitive advantage does not know the sources of its advantage.

The evolutionary view of the firm (Nelson and Winter)

Profits (might) remain the goal of the firm, but "***bounded rationality***" is present: firms are not capable to maximize. They adopt a satisfaction criterion.

Firms put in place strategies and practices that lead them to achieve a ***satisfactory level of performance***. As time rolls by, these strategies and practices become ***routines*** (firm is a bundle of routines).

Until this set of routines lead to satisfactorily performances, firm will not change them, while as soon as performances deteriorate and are no more satisfactory, firms will try to change them.

This search for ***new practices and strategies*** is constrained by the existing set of routines (i.e. a sort of firm's "genetic heritage") on which the firm relies upon. If the search is successful, i.e. leads to the identification of new strategies and practices (that will later become new routines and will modify the firm's "genetic heritage"), ***the firm will survive***; otherwise, it will be pushed out of the market.

The neoclassical approach is not so bad; it has a predictive capability of market behavior even if without the descriptiveness of firm behavior.

For **Hall and Hitch** (1939), firms may not consciously apply the profit maximization rule ($MC=MR$), but this does not mean that they have not internalized the rule with the practice.

Generally, firms set prices considering the full cost and adding a **reasonable mark-up**, even if they do not know the marginal cost. This mark-up is lower when the demand is elastic and higher when it is inelastic.

$$\frac{p - MC}{MC} = \frac{1}{\varepsilon}$$

In managerial firms there are several arguments (in addition to those already exposed) that could **discipline the behavior of managers** (and realigning their conduct to profit max):

- **Market labor discipline** (e.g. reputation);
- **Product market discipline** (e.g. competition): they cannot diverge too much because some potentials of the firm would not be exploited;
- **Capital market discipline** (e.g. takeovers risk).

DETERMINANTS AND INSIGHTS INTO FIRM BOUNDARIES

Horizontal and vertical integration

Quantity, production and make-or-buy decisions are taken over both the short and the long run, as long as profits are maximized and costs are minimized during both the short run and the long run.

Over the short run, we talk about “tactic” or “planning”, referring to 1 fiscal year. Decisions regards changes in the production levels (for example, accepting a new non-strategic customer), optimization and variation of the production mix, make-or-buy choices for contingent reasons (for example, demand's peak). Characteristics:

- We have to maximize the difference between revenue and avoidable cost (direct cost);
- We do not consider the fixed cost;
- We assume sales and cash flow coinciding;
- In order to take the optimal decision, we have to assess different scenarios.

Over the long run, we talk about “strategy”, referring to more than 1 fiscal year. Decisions regards the organizational structure, the adopted technology, the size of the plant, the purchase of a new machinery, the investments in new products or processes, the entry in or exit from the industry, the diversification, the mergers and acquisitions, and the delocalization of some activities. Characteristics:

- We consider the total cost;
- Revenue and cost flows must be discounted;
- We consider the hidden costs.

The **horizontal** boundaries of the firm refer to the **quantity** and the **varieties of products** a firm produces. They are related with the concepts of:

- Economies of scale
- Economies of scope

The **vertical** boundaries of the firm refer to the choice regarding which activities are performed **inside** the firm and which are **out-sourced** and left to the market. Thus, it is a **make-or-buy** decision. Organizing the horizontal and vertical boundaries is an important part of the **business strategy**.

As already said, the size is important in firms that remain focused on a single activity (economies of scale) or related activities (economies of scope). However, some companies are **conglomerates** involved in **unrelated diversification**.

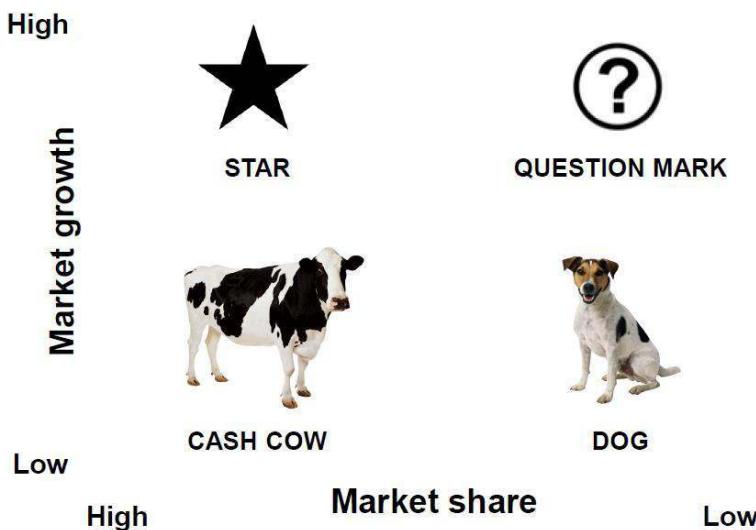
Costs of diversification:

- Costs of financing: money needed to acquire a company or to set-up an entrepreneurial venture.
- Costs of bureaucracy: incentives, information flows, and cooperation can suffer in large organizations.

Benefits of diversification:

- ✓ Higher efficiency thanks to the spreading of firm's underutilized organizational resources.
- ✓ Managers' preferences thanks to the spreading of managerial talent across businesses without enjoying the economies of scope.

Firms should use their internal capital market (equity market) instead of the external capital market (debt), according to the **BCG growth/share paradigm**.



According to the product lifecycle, firms use profits from established *cash cow* products to fund increased production of early stage *question mark* and rising *star* products. Profitable divisions effectively cross-subsidize the money losers:

- **Reduction** of the conglomerate's **value**
- **Lower incentives** to perform better for the **managers** of the money-losing divisions

Analyzing the BCG matrix:

- **Question Mark**: the market is strongly growing, but the firm does not have a relevant share in that market. Further investments are needed in order to increase the market share and let the product become a star.
- **Star**: the market is strongly growing and the firm has a relevant share in the market. Further investments are needed to support the growth leading to a cash cow.
- **Cash Cow**: market with little growth where the firm has a relevant market share. Cash Flows are high, the investments in new technologies are little and the profit is high. They are successful products that can finance other products.
- **Dog**: market with little growth where the firm has a little market share. The profits hardly cover the costs or there are losses.

The external capital market is more sophisticated and efficient. However, firms should use their **internal capital market**, because:

- Firms may find difficult to find external providers willing to fund new ventures due to **asymmetric information** (insiders vs. outsiders)
- Outsiders are reluctant to provide capital to firms that have **existing debt** (existing bondholders have to be paid first)
- External finance consumes **monitoring** resources (in order to guarantee bondholders' interests)
- Large diversified firms may have greater opportunities to identify **skilled workers** and assign them to appropriate and challenging jobs
- **Family firms** prefer diversifying the risk

The **family firms** are characterized by ownership in the hand of a family (at least partially) and constitute a very high share of the firms in a market (especially in Italy). In these firms, a great amount of the family wealth is invested within the firm.

- The risk of **bankruptcy** of the firm put at risk the wealth of the family.
- There is a **socio-emotional wealth issue**: they need to maximize their social prominence and prestige.
- Family firms tend to **diversify** their activities.

Managers get some benefits from diversification:

- Improve their **social prominence, public prestige** and **political power** (empire building)
- Limit managerial risk: by diversifying their firm, managers limit the **risk of extremely poor overall profitability** (however, financial markets are more efficient to spread shareholders' risk)

The mechanism through which shareholders control corporations and managers is the **corporate governance**. The mechanisms are:

- 1) **Board of directors:**
 - Monitor the management to ensure that actions are taken to increase shareholder's value.
 - However, CEOs may exert considerable control over the selection of new directors.
- 2) **Market for corporate control:** it is a deterrent to inappropriate management actions. If a bad manager is running a company, if a manager overpays for a diversifying acquisition, or if the stock market expects the firm to overpay for additional acquisitions in the future:
 - The **market price** of the firm's share **falls**.
 - It is an opportunity for another entity to try a **hostile takeover** and appoint its own managers. For example, another entity buys the company at a low price and with the potential to increase profitability with a better management.

This is the so-called "**capital market discipline**" and it is not free from criticism. If a raider can change management in a profit-increasing way, why cannot current shareholders do the same? One possible answer is that the raider possesses information that shareholders do not: for example, the raider may be a firm from the same or related industry.

However, if that were the case, why would an individual shareholder sell his or her shares to the raider? Surely, if the raider is to change management and increase firm value, then the optimal strategy is not to sell the shares.

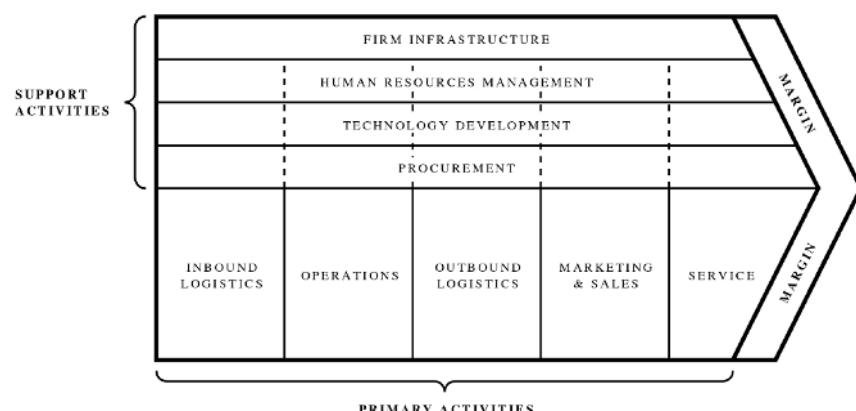
- 3) **Managers' labor market:** it constitutes an indirect incentive to pursue the organizational goals rather than the personal goals. Indeed, **managers' reputation** circulate in the labor market. **Potential employers** know whether managers pursue personal or organizational goals.
- 4) **Incentives to managers:** in the period 1980-1990, the market experienced a wave of leveraged buyout (LBO) transactions, where cash flows of the target firms were used to undertake unprofitable acquisitions instead of paying shareholders' dividends. This led to change the incentives to managers:
 - **Employee Stock Options (ESO):** some shares are given to the manager, and he/she has the opportunity to receive a gain in case he/she is able to increase the value of the share (however risk of short termism...).
 - **Incentives on performance indicators** accounting for the cost of capital (e.g. Economic Value Added: important to have high returns and low cost of capital and it is not important the amount of the cash flow).

Diversification allows greater growth of the business; however, the **growth** could be **not profitable**. Some **theoretical explanations** support the need of diversification but some others do not. Empirical studies show that the performance of **diversified** firms is **lower** if compared with **more focused** firms; however, correlation does not imply causality!

When it comes to the question of whether firms should diversify, the answer is "**it depends**, only when the economics make sense".

Considering the **vertical boundaries**, we have to take a make-or-buy analysis in order to decide whether implementing some business activities in-house or through an external supplier.

Introducing the Porter's value chain, the early steps in the production process are upstream, the later steps are downstream and the support activities are provided all along the chain.



There are different types of outsourcing:

Type of outsourcing	Definition	Example
Full outsourcing	contracting out of an entire business process	eCommerce
Functional outsourcing	contracting out of business function	Legal services
Selective outsourcing	fine-slicing of the value chain and outsourcing of very specific tasks	Testing a new software
Co-sourcing	the combining of services from within and outside the business	Milk from your farm and another
Multi-sourcing	the adoption of more than one supplier	Milk from 2 farms
Offshoring Outsourcing	the supplier is outside the national boundaries	Li&Fung: https://www.youtube.com/watch?v=h1f7hSCs8QI
In-sourcing or back-sourcing	firms deciding to perform in-house activities formerly performed in outsourcing	Belfe (clothing Italian firm) in 2004

The outsourcing relationships are characterized by:

- Market firms: they are outside specialists that can performs vertical chain tasks. They are usually recognized leaders in their field.
- Merchant coordinators: they are independent firms that specialize in linking suppliers, manufacturers, and retailers.

Outsourcing advantages:

- ✓ Market firms can achieve ***economies of scale*** (lower production costs thanks to the possibility to aggregate the demand of several customers).
- ✓ The ***value*** of the service is regulated by the ***market discipline***.
- ✓ Market firms can benefit from ***learning economies*** (e.g. consultancy).
- ✓ Market firms may have ***patents*** or proprietary information.

Outsourcing disadvantages:

- The customer has to incur in ***transaction costs*** (search, negotiating and monitoring costs).
- ***Know-how*** regarding product design and production processes may be lost in favor of outside firms.
- The customer has to incur in higher vulnerability to ***intellectual property losses*** and ***technology leakage:***
 - Firms do not want to compromise the ***source of their competitive advantage***
 - ***Well defined patents*** can help but may not provide full protection
 - ***Contracts with non-compete clauses*** can be used to protect against leakage of information (however hard to enforce in practice).

Wholly owned solutions advantages: It is usually easier to ***coordinate*** the business flows (e.g. the launch of a Heineken marketing campaign must coincide with increased production and distribution by its bottlers)

- Firms make decisions that depend in part on the decisions made by other firms along the vertical chain

- A good fit have to be accomplished in all dimensions of production (e.g.: timing, size, color and sequence)
- Without good coordination, bottlenecks arise in the vertical chain
- Coordination is especially important when design attributes are present; design attributes are attributes that need to relate to each other in a precise fashion (e.g. Fit of auto sunroof glass to opening, Boeing case)
- Timely delivery is critical
- Small errors can be extremely costly

Wholly owned solutions disadvantages:

- **Agency costs:** slacking by employees and the administrative effort to deter slacking
- **Influence costs:** managers try to influence the allocation of scarce resources in order to guarantee higher resources to their department: not efficient decisions are taken and a lot of effort is wasted just to allocate resources within the same organization

The **Make-or-Buy choice** should rely on the comparison between:

- The **implementation costs** of performing the activity in-house (the cost of equity, agency and influence costs, learning curve, capacity to achieve economies of scale...)
- The **costs of outsourcing** the activity
 - The **price** and **quality** assured by the supplier
 - “**Hidden costs**” (coordination costs, transaction costs, risk of intellectual property losses, cultural distance...)

There are some **make-or-buy fallacies**. In particular, it is not true that:

- 1) Firms should *make* an asset, rather than *buy* it, if that asset is a source of competitive advantage for that firm: an asset that is easily obtained from the market cannot be a source of advantage.
- 2) Firms should *make*, rather than *buy*, to avoid paying a profit margin to independent firms: always consider that the equity has a cost.
- 3) Firms should *make*, rather than *buy*, to tie up a distribution channel (vertical foreclosure) denying access to the competitors. Competitors will find alternative channels; the costs of acquiring a monopolist is very high (e.g. full value of the monopoly power) and it constitutes a possible violation of anti-trust laws.
- 4) Firms should *make*, rather than *buy*, because a vertically integrated producer will be able to avoid paying high market prices for the input during periods of peak demand or scarce supply. Alternatives to reduce input price risk are:
 - Long term contracts
 - Forward or future contracts
 - The capital tied up in vertical integration could be used as a contingency fund to deal with price fluctuations

Reality is more complex than the traditional make-or-buy dilemma:



According to the Williamson's classifications, the two determinants of the make-or-buy choice are the relationship-specific investments and the frequency of transaction. An increase in uncertainty of transactions, affecting the incompleteness of contracts, is likely to lead to less hybrid forms.

The traditional TCE approach explains the dilemma between the costs to have complete contracts (buy) and the costs of internalizing the activities (make), but it fails in including intermediate forms, such as the joint venture, strategic alliances, franchising and so on. There are three possible explanations to this situation:

1. **Hostage**: it is an asset with a value higher than the gain stemming from the other party opportunistic behavior. The threat is credible only if the value of the hostage is asymmetric for the two parties. The hostage is a mechanism able to avoid the opportunism. Both the parties could have their own hostage.
2. **Win-win**: both parties can gain from the relationship. However, it is not clear what happens if one of the two parties receive higher benefits.
3. **Cooperation**

According to the **TCE**, integration allows complete contracts. On the opposite side, the **property rights theory** (Grossman-Hart) states that even within organizations contracts are incomplete. Integration does not eliminate the possibility to have opportunistic behaviors. Therefore, **two incomplete solutions** are assessed and it will be chosen the **proprietary structure minimizing the surplus loss** (in comparison with the situation where the parties cooperate on the base of complete contracts).

In particular, the **property rights theory (PRT)** states that integration determines:

- Who controls the resources
- Who makes the decisions
- To whom profits are allocated

When the wrong firm has ownership rights, efficiency suffers. The **residual right of control** is the right to decide on all situations that are not included in the contract. This right is usually attributed to the owner of the firm.

Ownership is defined as the person or group who has the **residual rights of control** over certain assets. **Vertical integration** transfers the residual rights of control to the firm.

The **costs** of leaving some residual rights of control must be compared with its **returns**. Leaving the residual rights of control to one party instead of another one is **always a second-best decision**

Residual rights of control should be given to the **party that is providing the highest contribution to the final surplus** (the best solution among the second-best decisions). Usually the identification of the best solution among the second-best decisions require **practical experience**.

There are some alternative forms of organizing transactions:

1. Non-integration: the two firms are independent.
2. Forward integration: firm 1 (upstream) owns the assets of firm 2. This means firm 1 forward integrates into the function performed by firm 2 by purchasing control over firm 2 assets.
3. Backward integration: firm 2 (downstream) owns the assets of firm 1. This means firm 2 backward integrates into the function performed by firm 1 by purchasing control over firm 1's assets.

We have to notice that asset specificity often requires performing activities far removed from the core competence of the firm. Moreover, the form of integration affects the incentives of both managers to invest in relationship-specific investments.

Non-integration is the best arrangement when the investments in relationship-specific assets of both firms are of comparable importance (**no unbalances between the parties**).

Vertical integration is desirable when one firm's investment in relationship-specific assets has a significantly greater impact on the value created in the vertical chain than does the other firm's investment (**unbalances between the parties**).

Employees could be **selfish**. In order to align managers and firm interests, we need **rewards and penalties**; otherwise, there would not be benefit from integration.

According to PRT, central **decision-making rights for an activity** should be delegated to those managers whose decisions will have the greatest impact on the performance of that activity. We have also to consider the **path dependence**, which means the decisions taken in the past affect today's set of alternatives.

For example, firms manufacturing for **internal use** do not typically sell excess output to other firms because they **do not have the skills to sell to multiple buyers**. If a **supplier is acquired**, the firm have the **skills to sell** to others besides itself, and selling outside would be neither a distraction nor an activity for which the firm lacked the requisite skills.

Integration decision

For each step in the vertical chain, the firm has to decide between **market exchange** and **vertical integration**. The degree of vertical integration differs:

- Across **industries** (e.g. firms in the aluminum industry are more vertically integrated than firms in the tin industry).
- Across **firms** within an industry (e.g. Hyundai is more vertically integrated than Honda).
- Across **transactions** within firms (e.g. U.S. firms tend to outsource transportation services to a much greater degree than warehousing or inventory management).

Firms should economize, which means balancing technical and agency efficiency:

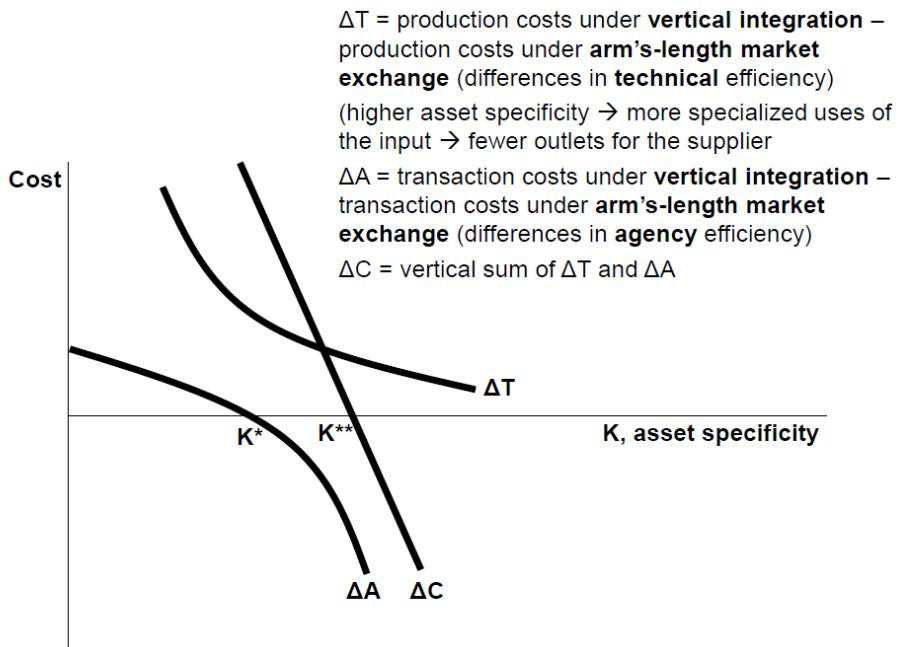
- **Technical efficiency:** the firm is using the least-cost production process.
- **Agency efficiency:** the extent to which the exchange of goods and services in the vertical chain has been organized to minimize the coordination, agency and transaction costs.

The balance is necessary to minimize the sum of technical and agency inefficiencies. Using the **market** leads to **higher technical efficiency** compared to vertical integration (power of market discipline). The difference of technical efficiency of market over vertical integration (ΔT) depends on the nature of the assets involved in production. ΔT is positive for any level of asset specificity because outside suppliers can aggregate demands from other buyers and thus can take advantage of economies of scale and scope. As the assets become more specific, the market firm's advantage becomes weaker. Therefore, **ΔT declines with greater assets specificity**.

At **high levels of asset specificity, differential agency efficiency (ΔA)** of market over vertical integration is **negative**. When specific assets are involved, potential for a holdup is high and the transaction costs are higher. At low levels of asset specificity, ΔA is likely to be positive. Without the holdup problem, market exchange could be more "agency efficient" than in-house production.

The **combined differential efficiency** of market over vertical integration (ΔC) will be **negatively related to asset specificity**. This is the **efficiency trade-off**.

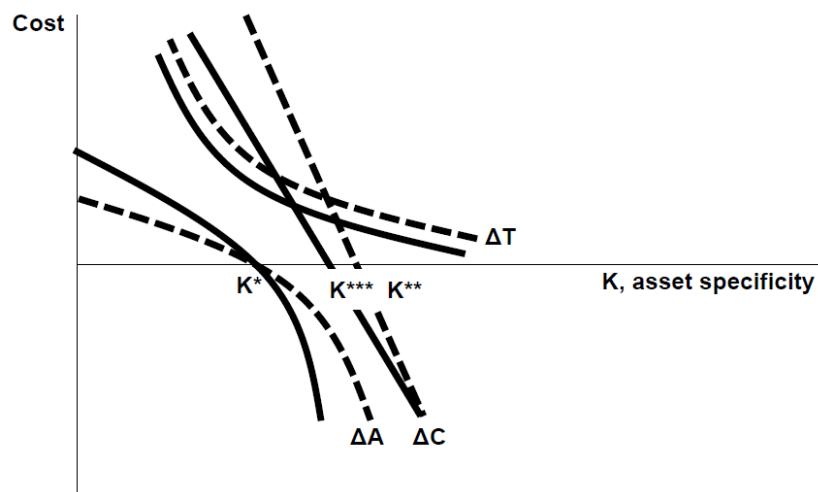
- At high levels of asset specificity, vertical integration is more efficient.
- At low levels of asset specificity, market firms have an edge.



Introducing the **scale of transaction**, if it increases:

- The vertically integrated firm enjoys better economies of scale: therefore, the differential technical efficiency decreases for every level of asset specificity.
- The differential agency efficiency becomes more sensitive to asset specificity: it increases with scale for low asset specificity and decreases with scale for high asset specificity.

The combined differential efficiency (ΔC) **sharply declines** for low asset specificity. The **degree of asset specificity** at which **market** is just **competitive** with **vertical integration** declines K^{***} . **Vertical integration** is preferred to market exchange over a **larger range of asset specificity**.



If the firm is considering whether to make or buy an input requiring significant upfront setup costs (source of **economies of scale**), and there is a **large market outside the firm for the input**, then the firm should **buy** the input from outside market specialists

Firms with large market shares are likely to be **vertically integrated** as long as in-house input production can take as much advantage of economies of scale as an outside market specialist.

Considering the **real world evidence** (Anderson and Schmittlein), it refers to the fact that in the electronic components industry, firms rely on their own sales force:

- When there is a greater asset specificity
- When they are larger manufacturers
- When performance measurement is more difficult

If firms have the market power (they set a price above the marginal cost), this leads to the problem of double marginalization.

When a firm with market power (e.g. an input supplier) vertically integrate with another firm with market power (e.g. a manufacturer), both the upstream supplier and the downstream buyer exploit their market power and apply a **double mark-up**.

Thus, the price of the finished good will **exceed the price that maximizes the joint profits of the supplier and buyer**. **Integration is justified** in cases where the two firms with market can **jointly** maximize their **profits** and **improve** their condition if compared with a situation where profits are maximized independently between each other.

The alternatives to vertical integration are:

- 1) **Tapered integration**: it is **mixture** of vertical integration and market exchange. A firm may **produce part of its input** on its own and **purchase the rest**. A firm may **sell part of its output** through **in-house sales efforts** and sell the rest through independent distributors.

Advantages	Disadvantages
Additional input/output channels without massive capital investments	Possible loss of economies of scale
Information about costs and profitability from internal operations can help in negotiating with market firms	More difficult coordination between the two production units (product specifications, delivery times...)
Threat of outsourcing can impose discipline on internal channels	Monitoring problem (e.g. the firm may mistakenly establish the performance of an inefficient internal supplier as the standard to be met by external suppliers)
Threat of self manufacture can impose discipline on external suppliers	Managers may maintain inefficient internal capacity because these facilities had formerly been critical to the firm
Internal supply capabilities will protect against potential holdups	

2) **Franchising:** this means the **franchiser** gives partial ownership rights to **franchises**. The franchises put up the capital to build and operate their stores and pay a fee for the right to use franchiser's name and business model. Franchisers may require franchises to purchase from designated suppliers, offer specific products, and conform to architectural and design guidelines the point of sale. The franchiser performs tasks involving substantial scale economies due to purchasing and branding. **Free-riding problems** are limited through tight quality controls and frequent surprise inspections.

3) **Strategic alliances and joint ventures:** they are in the middle of pure market exchange and full vertical integration. They are a way to organize complex business transactions collectively without sacrificing autonomy:

- **Strategic alliances** involve cooperation, coordination and information sharing for a joint project by the participating firms
- A **joint venture** is a particular type of strategic alliance where a new independent organization is created and jointly owned by the promoting firms

The parties of the alliance remain independent. They can be horizontal, between firms of the same industry, or vertical, between firms at different stages of the value chain, or involving firms that neither are in the same industry nor related through the vertical chain.

Alliances rely on greater **trust, reciprocity, cooperation, coordination and information sharing** than would occur in arm's length contracts. **Disputes** are rarely litigated but resolved through **negotiation**.

Features leading to strategic alliances:

- The transaction involves **impediments** to comprehensive **contracting**
- The transaction is **complex**, not routine
- The transaction involves the creation of **relationship-specific assets by both** parties in the relationship
- It is excessively **costly for one party to develop all the necessary expertise** to implement the activity (learning curves, indivisibilities...)

- The **market opportunity is transitory** thus independent parties do not find convenient to merge
- The market opportunity occurs in an **institutional context with unique features** requiring a local partner with relationship in that environment

The risk of leaking of private information and losing control over proprietary information is more severe in alliances than in traditional market transactions because:

- The relationship is more complex and requiring high degree of knowledge sharing.
- The contract is not very well specified.

Other problems are:

- Coordination issues
- No formal mechanisms for making decisions
- No formal mechanisms for resolving disputes
- Agency and influence costs

4) **Collaborative relationships:** traditionally, **Japanese and Korean** industrial firms have been **less vertically integrated** compared to their western counterparts and have organized the vertical chain using **long-term relationships** rather than arm's length transactions.

- **Keiretsu** in Japan: Japanese manufacturers maintain close, informal, long term relationship with their network of subcontractors:
 - High asset specificity in the relationship
 - Members have strong institutional linkages
 - Links are further strengthened by social affiliation and personal relationship among executives
 - Easy coordination and no holdups when vertical chain activities are performed by keiretsu members
 - However, recent research indicates that members have extensive business dealings also outside their keiretsu

The members of a keiretsu work with each other through implicit contracts, which are unstated understanding between firms in a business relationship.

Long-term relationship can make firms behave cooperatively towards each other without any formal contract. The **threat of losing future business and steam of profits** is enough to deter opportunistic behavior in any one period. The desire to protect one's **reputation** in the market place can be another mechanism that makes implicit contracts viable.

- **Chaebol** in Korea

Determinants of internationalization and the multinational corporations

Multinational corporations (MNCs) expanded a lot in the last 20-30 years, and they are still growing (average growth 2-3 times higher than the growth of the GDP).

The internationalization process implies the growth of the firm beyond the home country boundaries through:

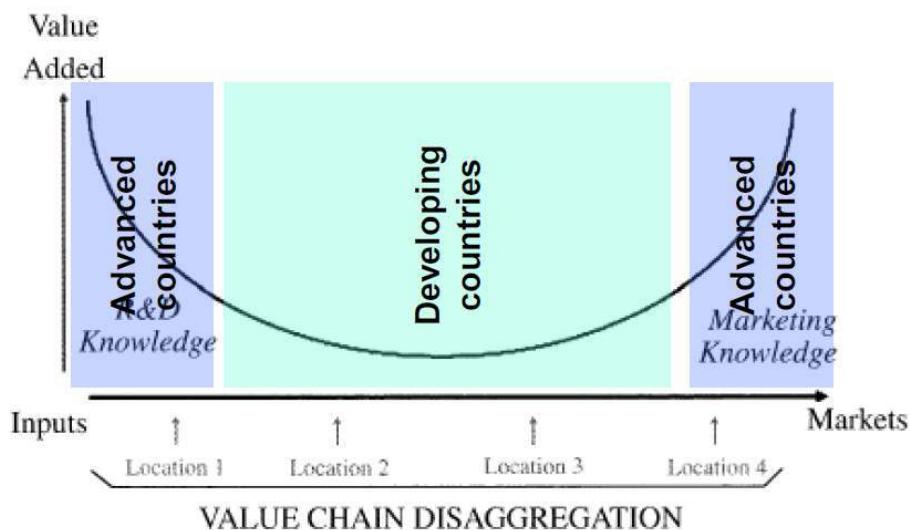
- **Export**
- **Foreign direct investments (FDI)**
- **Intermediate forms**: joint ventures, strategic alliances, licensing, partnership and so on

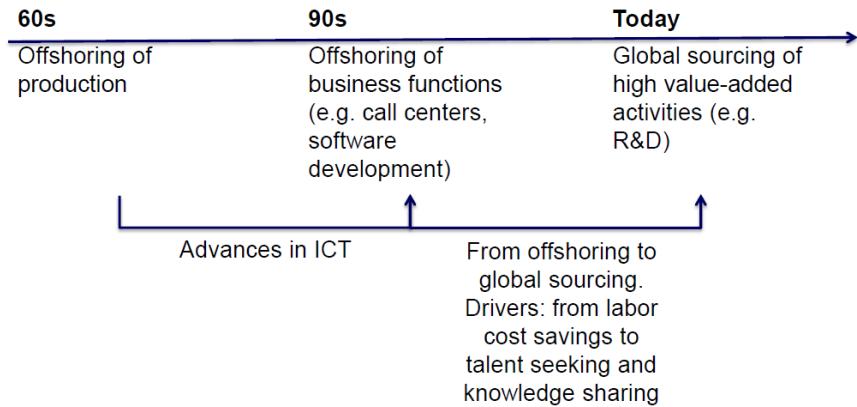
MNCs consist of one or more FDI, considering that the definition of FDI does not include portfolio investments such as the purchase of shares.

The internationalization process characteristics have changed over time:

60s	Today
Manual and labour intensive	Technology and capital intensive (the impact of ICT)
Orientation towards production	Orientation towards customers
One product	Product diversification
National business context	International business context
Long product life cycle	Short product life cycle
Loyal customers	Unstable and demanding customers' needs
Mere profit maximization	Greater attention towards social, environmental and ethical issues

Nowadays, **the phenomenon of offshoring** is spreading, consisting of disintegration, relocation and reintegration of a company from one country to another. This leads to the **value chain disaggregation in multinational enterprises**, exploiting more than one location.





Today also high value activities, such as R&D and marketing, are increasingly offshored.

Offshoring raises concerns of **high unemployment rates** and **loss of core competencies** in advanced countries. Therefore, US and European policy makers are considering to bring back manufacturing activities (re-shoring) and keep the existing manufacturing at home.

Moreover, companies having offshored their activities are **unsatisfied by the performances**. The **business context has changed** (for example with a labor cost inflation) and they want to benefit from **geographic, cultural and linguistic proximity**. There are three solutions:

- Back re-shore at home
- Near re-shore in a nearby country
- Keep the production at home

Economic theories of Multinational Corporations

Before Hymer's theory, several economists investigated international trades. These theories regarded trades and capital movements across borders (not FDI):

- The main determinant of movements of goods (trade) across borders is the **difference in factor endowments**
- The main determinant of movements of funds (capital) across frontiers is the **difference in interest rates**

These theories were assumed to extend and apply to all types of investment. However, FDI were not considered as an autonomous category needing explanations. There was no theory studying the FDI by large companies for production and direct investment purposes.

Hymer's theory

Hymer is the first scholar developing a fully coherent theory and his approach has led to many theoretical developments. He distinguished between:

- Portfolio of investments (purely financial)

- FDI providing the firm control over the business activities abroad

The question is: why is the MNC that takes up investment opportunities in the host country rather than the local firm?

In the **Hymer's seminal work**, FDIs allow the growth of the firm benefitting from increasing economies of scale and providing the solution to market imperfections.

If there are **market imperfections**, the process of internationalization depends on the coordination of production via **internal hierarchy** (vertical integration, FDI), that is more advantageous than imperfect market transactions (export).

For FDIs to prosper there must be market imperfections that create both **advantages and conflicts**:

- Imperfections in the good market
- Imperfections in the factor market
- Internal and external economies of scale
- Governments' interference with production or trade

FDIs allow exploiting the advantages and reducing conflicts due to market imperfections, in particular:

- ✓ They reduce the competition
- ✓ They increase the market power
- ✓ They increase extra-profits

FDIs involve extra-costs and risks:

- Costs of **communication and information acquisition** (in a different cultural, linguistic, legal, economic and political context)
- Costs due to the **less favorable treatment** given by host countries governments
- Costs and risks of **exchange rate fluctuation**

Therefore, the MNCs **competitive advantage** must allow to overcome the local producers advantage, otherwise there is no need of FDI.

Concerning the determinants of FDI, they are **not motivated by searching for low costs**; otherwise, local firms should successfully compete with foreign ones. The determinants are:

- ✓ Exploitation of **specific advantages** thanks to specific knowledge or technological resources (the ownership of a specific technology, products, brands)
- ✓ **Removal of conflicts** through collusion with rivals
- ✓ **Diversification of products**, or market locations, or production locations

Summing up, the advantages stem from the possibility to solve market imperfections, the exploitation of economies of scale, and the adoption of a resource that is known only by a specific firm and not to local firms.

Later Hymer moved towards a **Marxist approach** stressing the **conflicts and contradictions of the internationalization of production**:

- Conflicts between different parts of the same firm

- Conflicts between the firm and its labor force
- Conflicts between the firm and governments
- Conflicts between developed and developing countries

The MNC activities have effects on labor, politics, economic, policy, division of labor and so on. At a ***micro level***, MNC is a strong progressive force enabling:

- The planning and organization of production on a worldwide scale
- Increase in productivity
- Spread of new technology and new products

At a ***macro level***, the spread of activities into many sectors and countries leads to conflicts and contradictions as long as developments and opportunities spread unevenly. There is a trade-off between ***micro planning and macro anarchy***.

Critiques to Hymer:

- The determinants of FDIs are determinants of investments in general, thus underplaying the specificity of internationalization.
- Hymer overemphasizes the risks and costs of foreign operations as long as they have declined considerably since the 60s.
- Hymer underplays the advantages of internationalization and multinationality per se, including the advantages of spreading the production over many countries.

There are some other studies before the Vernon's theory:

- **Kutznets**(1953): the growth of demand tends to be slow in the innovation phase of a product, then to accelerate and, finally, to slow down again
- **Posner** (1961) analyses how initial product innovation in one country leads to cumulative technological and trade advantages (dynamic economies of scale, learning-by-doing and a tendency towards accumulation of inventions)
- **Hirsch** (1965; 1967) analyses the phases of product's life in relation to the technology and scale production, to the type of labor skills needed, and to the countries' competitive advantages
- The origin of **Vernon's work** (1966) is in the technological gap theory and the literature on the life of the product putting particular emphasis on international production and trade

Vernon's theory

Introducing the ***Vernon' theory***, all the countries have the same access to scientific knowledge; however, there is a large gap between the knowledge of a scientific principle and the application of such knowledge in a marketable product

US market characteristics are:

- Consumers have high average income per capita;
- Very large market;

- High unit labor costs and a large supply of capital (market abundant in capital and scarce in labor).

New products (whose demand requires high income per capita) have a **large market in the US** and a fertile ground for the designed products to save labor in the production phase.

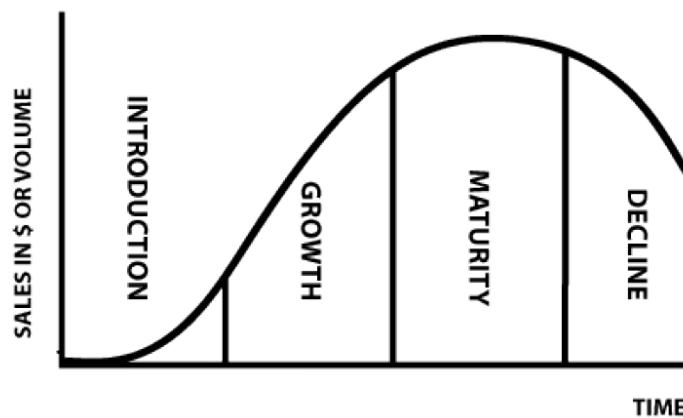
In early stages, new products will be located in the US, in spite of high production costs:

- Producers are interested in **flexibility** in adapting the product changing inputs as necessary. There is no need of standardization and efficiency.
- **Proximity of production to the market** is essential to ensure effective communication between producers and consumers.
- The product enjoys a **high degree of differentiation** and a monopolistic position (no real competition). Therefore, there is low price elasticity and it is easy to translate high costs into high prices.
- Target consumers have **high income** and are therefore prepared to pay high prices.

As demand for the product grows and the product reaches **maturity**:

- The need for flexibility declines and the product becomes **more standardized**. Therefore, production processes require high capital intensity and unskilled labor.
- **Imitation** becomes easier. The technology is widespread and imitable.
- The **competition increases** affecting the price that the firm can charge and cost cutting becomes necessary.
- The **need for proximity to customers declines** and the demand spreads to other high-income countries (European countries).
- As the product matures, costs become more important and **production is delocalized**.

Analyzing the product lifecycle:



- 1) **Introduction:** there is high **uncertainty** related to:

- The potential success of the product
- Its required features
- Production technologies

The firm is innovating and benefitting from a ***monopolistic situation***. It is not convenient to go abroad as long as the ***sunk costs would be too high*** in a phase where the entrepreneur is not fully aware regarding the potential of the new product.

- 2) **Growth:** once the product becomes successful, it is more ***stable in features***, the production technology can be ***standardized***, and the firm benefits from ***monopolistic market power***. Increasing the production the firm can benefit from increasing ***economies of scale*** and ***experience***. Therefore, profitability can improve through ***international expansion***, firstly exporting and, after a certain threshold, producing directly where the final product is sold.

$$\text{Production Cost}_{\text{abroad}} \leq \text{Production Cost}_{\text{home}} + \text{Transportation cost}$$

- 3) **Maturity:** as the product gains success, new entrants are attracted (imitation or replication of technologies). The ***monopolistic advantage is gradually eroded*** and the firm needs to become ***more efficient***. The market must be present in foreign countries in order to avoid ***protectionist behavior*** by foreign governments. The ***rent*** of the monopolistic firm is gradually ***eroded***. The firm will try to settle in countries where the ***costs for production are lower*** in order to become more efficient and competitive.
- 4) **Decline:** after a certain threshold, costs cannot be reduced anymore. The competitive pressure become higher over time. A new product needs to be introduced before the previous one will see its decline.

The MNC maximizes its profit benefitting from ***temporary monopolies***.

FDI is preferred to ***export*** if:

- ✓ Threat of rivals beginning imitations in foreign countries
- ✓ Production costs are lower in European countries
- ✓ Threat of import controls by European governments

Gruber et al. added (1967):

- If production costs in Europe are low enough to outweigh transportation costs, there are possible exports from Europe to the US market (market sourcing strategy)
- Direct production abroad may have effects on the pattern of international trade in Europe (e.g. if the US firm has invested in the UK, the production facilities in this country can be used to export to Italy or other European countries)

Critiques:

- In the 1980s the differences in per capita income, cost of labor, size of markets and consumer tastes between Europe and the US have narrowed considerably. Therefore, product life-cycle theory less applicable.
- Deterministic: there is a hierarchy of countries in terms of innovation potential, stages of development and income per capita. The technology transfer is from the most to the less developed countries. Cantwell (1989; 1995) shows that, even if MNCs are leaders in

innovation, this does not necessarily lead to their home country being the leader (innovative activity has its origin in many countries).

Knickerbocker's theory

Knickerbocker was a student of Vernon at Harvard Business School. His source of the data is the "Harvard Multinational Enterprise Study". Thus, it is a US-centric study

He gave a definition of FDI: capital flow resulting from investment by an enterprise "in assets outside its home country in order to control, partially or fully, the operation of these assets".

He stated that firms expanding internationally belong to oligopolistic industries characterized by:

- Few sellers
- Products that are close substitutes
- Substantial market interdependence among the competitive policies of these firms" (oligopoly is defined in terms of both the market structure and the behavior of firms)

The interdependence of firms leads to a pattern of action and reaction, move and counter-move (as in a game of chess). Each firm combines aggressive and defensive moves:

- **Aggressive** moves are made to improve the firm's position (e.g. new technologies, new sources of raw materials...).
- **Defensive** moves are made to offset aggressive policies by the firm's opponents; the advantage that the aggressor may gain could, in the long run, be highly detrimental to its rivals, who therefore have to react in order to minimize the risks.

Firms are aware that, in a situation of roughly equal strengths, aggressive moves are likely to lead to defensive ones and thus to the risk of **mutually destructive competition**. There is no price war, but market-enhancing competition **via advertising**.

The **oligopolistic equilibrium** is defined as a state of affairs among sellers such that "all rivals having roughly the same competitive capabilities, there is little reason for any one rival to expect that it can, with impunity, improve its market position at the expense of the others".

Graham's theory

At first glance, it seems that the firm must compare the monopolistic profits in its own market with the two oligopolistic profits in the home and in the host market. However, the firm must also consider the behavior of the other firm (strategic interdependence) in order to assess if it is convenient to penetrate the new market.

It is a prisoner's dilemma: the strategy of entering the new market is dominant for both the players (even when the monopolistic profit is higher than the sum of the two oligopolistic Cournot profits); rationally the two players discard the most efficient solutions as long as they cannot collude.

Internationalization and the transnational corporation

The background are Coase and Williamson: we have to compare the costs of carrying out transactions through the market and of organizing the allocation of resources internally.

Markets are characterized by:

- Structural imperfections: some firms have monopolistic or oligopolistic market power (no perfect competition)
- Transactional imperfections: information asymmetry between buyers and sellers.

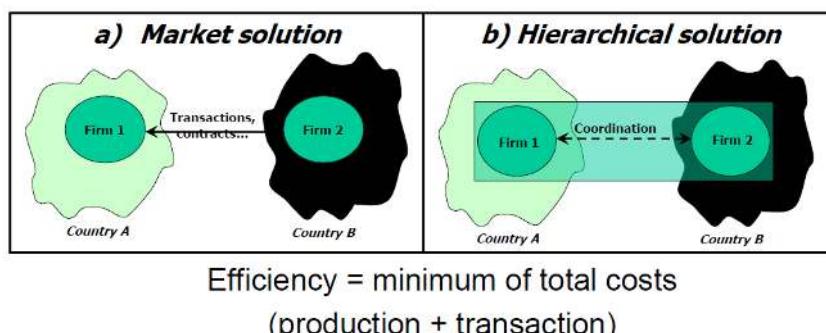
The extension of the transaction cost theory from the firm in general to the international firm is due to:

- McManus
- Buckley and Casson
- Teece, Rugman, Caves and Hennart

The MNC is seen as the mechanism that efficiently transfer tangible and intangible assets when the markets are not perfect and its price mechanisms do not work efficiently.

The **McManus** theory refers to the fact that if the control is internationally centralized, productive activities in different countries will earn a higher total income. The control is distributed between different locations if the sum of the values of resources in two or more interdependent productive activities is greater than it would be if the activities were conducted autonomously.

Hierarchy is more efficient than the market because transactions and contracts are substituted by coordination, minimizing the total cost (production and transaction).



Buckley and Casson studied that:

- Firms maximize profit in a world of imperfect markets
- When markets in intermediate products have transactional imperfections, there is an incentive to bypass them by creating internal markets.
- Internalization of markets across national boundaries generates MNEs

The two most important areas for internationalization are:

- 1) **Markets for intermediate products:** especially before the Second World War, the major factor that contributed to the emergence of MNEs was demand for primary goods, leading to vertical integration across frontiers and to internalization of intermediate markets.
- 2) **Markets for knowledge:** knowledge is a public good within the firm; its costs of transmission are low and it can be easily internationalized .

Transaction costs are high in trans-border activities. Therefore, there is a greater incentive to internalize such activities by direct production abroad rather than via market transactions or licensing.

Caves (1982) used a model of multi-plant production to explain the MNE and its presence in more than one country. There are three types of MNEs:

1. **Horizontally integrated MNE:** same line of goods from its plants in each geographic market. This is because some tangible and intangible (e.g. research, knowledge, brand names) proprietary assets that used in the firm are more productive than in outside environment and thus give a higher return.
2. **Vertically integrated MNE:** produces outputs in some of its plants that serve as inputs to its other activities. This is because there is a transaction cost reduction.
3. **Diversification of production:** plants' output are neither vertically integrated nor horizontally related to one another. This is because of the principle of risk aversion.

Dunning's Eclectic Framework

Dunning synthetizes previous approaches and he introduces a comprehensive framework (not properly a theory) studying **why firms become multinationals**.

The main assumption is that the degree and the structure of foreign activities depend on the existence of three types of advantages:

1. **Ownership**
2. **Locational**
3. **Internalization**

→ this is the so-called OLI eclectic paradigm.

The **ownership advantages** are the following:

- Competitive advantages are specific to a particular enterprise;
- They enable the company to take advantage of investment opportunities wherever they arise;

We can distinguish **three types** of ownership advantages:

1. **First type:**
 - Standard advantages which any firm can have over another producing in the same location;
 - Independent of whether the firm is multinational or not;
 - They include benefits of:
 - Special access to input or/and markets;
 - Established market position;

- Superior technical and/or organizational knowledge;
 - Size;
 - Monopoly position.
2. **Second** type:
 - Advantages which a branch of a national enterprise may have over a new enterprise;
 - Related to all those economies and benefits deriving from belonging to a larger pre-existing organization;
 - They include benefits of:
 - Access to cheaper inputs;
 - Knowledge of markets and local production conditions;
 - Access to innovation and technology available at zero or low marginal cost (if already developed by other branches of the company).
 3. **Third** type:
 - Advantages deriving from the multinationality of the enterprise;
 - Firms with a history of international operations are in a better position than firms with experience at the national level only;
 - Greater capability to take advantage of different factor endowment and market situations.

The ***locational advantages*** are the following:

- They are specific to a country, which are likely to make it attractive for foreign investors;
- Advantages are linked to the geographical and political space (e.g. quality of transportation and communication, legal and commercial infrastructures, government policies, quality and price of inputs).

Finally, the ***internalization advantages*** are the following:

- Benefits deriving from producing internally to the firm (→hierarchies) rather than markets;
- They allow to bypass external markets and the transaction costs associated to them;
- If internalization advantages exist, the firm will implement FDIs, otherwise it will turn to the market.

The conditions under which **Foreign Direct Investment (FDI)** takes place are:

1. The enterprise must possess net ownership advantages against firms of other nationalities in serving particular markets;
2. The enterprise must have benefits from internalizing the use of resources in which it has an advantage rather than selling them on external markets;
3. The country where the FDI takes place must offer special locational advantages to be used in conjunction with those deriving from ownership and internalization → locational advantages.

There are four major Dunning's later developments:

- 1) **Operationalization:** the original eclectic approach is far too wide to be useful in practice. There are too many variables, so the explanatory and predictive power is low. The key operationalization is the contextualization of the variables linked to the four sets of advantages.

This Dunning's development distinguishes FDI according to whether is:

- ✓ **Resource** seeking: firms go abroad in order to get the cheaper or more productive resources;
- ✓ **Market** seeking: firms go abroad in order to penetrate new markets; the physical presence is needed to develop distribution channels and to know the market (follow-the-customer);
- ✓ **Efficiency** seeking: it includes the two previous categories in order to improve the overall efficiency of the firm: production where inputs are cheaper or more productive + establish distribution channels in important markets;
- ✓ **Assets** seeking: firms go abroad in order to acquire some strategic assets that they are unable to develop internally (R&D for example). They go abroad not to exploit a competitive advantage but in order to develop it.

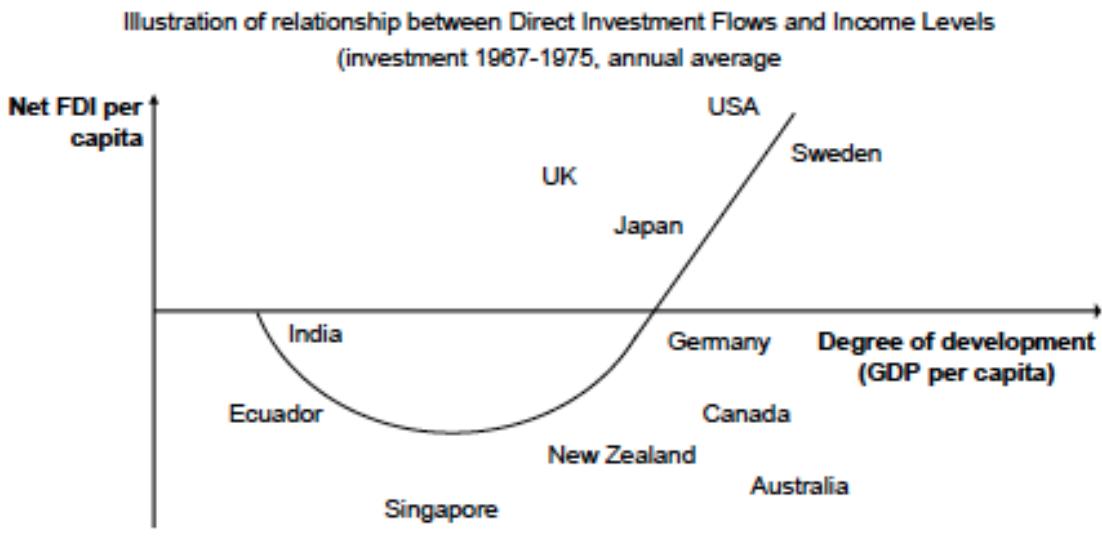
- 2) **Dynamization:** the eclectic paradigm is a static model and it does not consider the dynamics, the changes over time. These dynamics are very important in the internalization processes: history matters a lot.

In 1993 Dunning began to tackle this issue.

- 3) **Countries' development path:** how do FDIs induce changes in the overall system? How are FDIs affected by such changes?

The pattern of FDI is affected by the stage of development path of the country and its position in terms of inward and outward FDI.

The country's pattern and speed of development is related to the inward FDI and to the strength of its MNCs as witnessed by its outward FDI.



The Scandinavian School

The only dynamic model of internationalization before the 70s is Vernon's: the producer who wants to meet the demand abroad, first exports the new products and later starts direct production in the foreign country.

A group of Swedish economists and management academics probed further into the dynamics of internationalization → Johanson and Wiedersheim-Paul, 1975; Johanson and Vahlne, 1977.

They studied the organization of **sales and marketing** in one or more foreign countries. They also analyzed the stages of the internationalization process and they found out it is a time sequence that develops in a logical and linear pattern.

The model explains two patterns in the internationalization of the firm:

- ✓ **First** pattern of internationalization: increasing involvement into the same country. The firm's engagement in a specific foreign market/country, which develops according to an **establishment chain**;
- ✓ **Second** pattern: spread of internationalization from one foreign country to others. The involvement into **several** foreign countries, which proceeds in a time sequence related linearly to the psychic distance from the home country.

Some elements that are **common** to these two patterns are:

- **Incremental** decisions;
- **Dynamic and linear** commitment:
 - from small to large commitment;
 - from one stage to the next;
 - from one foreign country to several.

Johanson and Vahlne (1977) distinguish between variables related to the internationalization:

- ❖ **State aspects:** elements of the situation as it is at the time of the analysis and as the result of investment:
 - Resource commitment to the foreign market (e.g. training of local labor force):
 - amount of resources already committed;
 - degree of commitment.
 - Knowledge about foreign markets and operations (particularly important is the experiential knowledge!):
 - General knowledge related to marketing;
 - Market-specific knowledge.
- ❖ **Change aspects:** results and decisions that extend through time:
 - Performance of current/present business activities;
 - Decisions to commit resources.

First pattern: involvement in any single foreign country will proceed cautiously and in accordance with the following stages in the establishment chain:

- **Exports** via agents/independent representatives;

- Setting up of **sales subsidiaries**;
- Setting up of **production subsidiaries**.

→ The outcome of one decision constitutes the input of the next one. It is a series of incremental decisions.

Second pattern: also here the sequence is **dynamic** and **linear**.

The psychic distance is defined as the sum of factors preventing the flow of information from and to the market (for example, differences in languages, education, business practices, culture and industrial development).

The internationalization stages develop from a country with little psychical distance compared to the home country to farther countries.

Note that the psychic distance is correlated with the geographic distance.

There are some **critiques** to the Scandinavian model: first of all it is applicable only to:

- Early stages of internationalization;
- Firms originating in small industrialized countries (such as Sweden);
- The manufacturing industry.

While this model is not applicable to the so-called “born global”; in other words, to those companies that do not go through the stages of internationalization but operate in many countries from the very beginning (YOOX for instance).

Evolutionary theories of the MNC

In **Neoclassical** economics, the firm's resources are taken as given and the subject matter of economics is their **allocation** within and outside the firm.

However firms grow and develop their resources and capabilities:

- ❖ **Penrose** (1959) characterizes the firm as a bundle of resources and competencies;
- ❖ **Nelson & Winter** (1982) focus on how competencies evolve over time.
→emphasis on **learning** and how history matters;
- ❖ **Cantwell** (1989) and **Kogut & Zander** (1993) adopt evolutionary theories in the context of the MNC.

Cantwell's theory

Cantwell aims at explaining the **competitive advantages** of MNCs (why some firms are more successful than others) and their interaction with the localities.

The firm does not have ownership advantages ex-ante, but it is able to generate and develop its ownership advantages through its **strategic behavior**.

The main sources of competitive advantages are **innovation** and **technology**.

By operating in many countries – characterized by diverse knowledge and innovation contexts – MNCs can acquire knowledge from the localities and use it to further their innovative activities.

MNCs are involved in two types of networks:

- 1) **Internal** network, between the various units of the firms spread in a variety of geographies;
- 2) **External** networks, between units of the firm and suppliers/distributors, consumers and partners in collaborative ventures.

→ The external networks enable to **acquire knowledge** from the external environment and this knowledge is **incorporated and spread** to the MNC via its internal network.

MNCs' innovation activities generate **spillover** effects to the locality and the industry as a whole.

As several firms locate their innovation activities in the same locations, positive agglomeration effects emerge and develop (**external economies of agglomeration**).

At the same time, investors benefit from the **favorable technological environment** that develops in the locality.

The accumulation of technology leads to **high productivity**, **high growth** and **high incomes per capita**.

High income per capita generates **high levels of demand**, thus attracting high levels of investment and **new waves of innovation**.

→ **Cumulative causation process**

→ Location advantages are not exogenous, but endogenously created by the innovation and location strategies of firms combined with the spillover effects of their activities.

According to Cantwell, these are the **critiques** to the Vernon's IPLC:

- Even though there are leader firms in innovation, it is **not possible to extend the concept to countries**; no single industrialized country can be designated as leader in the way Vernon did with the US;
- Innovations are **not necessarily located in the home country** of the parent company;
- Innovation is **linked to production** (while for Vernon is mostly linked demand and consumers);

- Innovation **spills from product to product**, while for Vernon, innovation refers to a new product, not to a general activity involving a variety of products and processes within the firm.

Kogut & Zander's theory

Kogut and Zander aim at explaining the role played by **knowledge** in the boundaries of the firm; in other words, the extent to which the MNC will develop via **international activities** or via **external contractual arrangements**.

The firm is seen as a social community that creates and **transforms knowledge** into economically rewarded products and services.

Knowledge is largely **tacit**, uncodifiable and unteachable; thus it is **easier to transfer** the knowledge within the firm than across organizations.

The limits to the firm are set not by **market failure** (transaction costs and risk of opportunistic behaviors), but by the firm's **efficiency in acquiring knowledge**.

Knowledge in Kogut & Zander:

- Is seen as the main source of **competitive advantage**;
- As long as it is tacit and uncodifiable by definition, it is **difficult to imitate**, thus it is an advantage on which the firm can further build up without fears from rivals' **imitation**;
- As long as it is **cumulative, older knowledge** is codifiable and **more easily transferable** outside the boundaries of the firm;
- Established technology is not a public good; it is **transferable at a cost** that depends on the accumulation of experience and on learning about codification procedures.

INNOVATION AND TECHNOLOGICAL CHANGE

may not achieve the . An in the :

- : to achieve a given amount of output, we are able to use a lower level of input
- : starting from the same level of input, we are able to produce more output

These two types of efficiency are based on a **static dimension**. They do not consider that economy is evolving over time.

We have to talk about **dynamic efficiency**, which means **improvement and generation** over time of **new products and production techniques** in order to save more cost.

We will be able to achieve these things tomorrow if we are able to make some investments today, in particular **R&D investments**. An economic system is efficient as long as it allows a good rate of new product generation and new production techniques over time.

The first economist recognizing the importance of dynamic efficiency was **Schumpeter** in two books "The theory of economic development" and "Capitalism socialism and democracy". These two books put **innovation** as the **main engine of economic growth**. It is a real breakthrough with respect to the neoclassical tradition, whose main idea was more labor and capital mean more output.

The first contribution showing innovation-related differences was **Solow**. He did not use sophisticated tools and econometrics. He decomposed economic growth recognizing that 87% economic growth in US was caused by technical change and innovation. It was the milestone putting innovation at the center of the stage in the economic analysis.

$$Y = A(t)F(K, L) = A(t)K^\theta L^{(1-\theta)}$$

↓ ↓ ↓ ↓
Output Technology Capital Labor
Cobb-Douglas with constant return to scale

Totally differentiating with respect to time and dividing by Y:

$$\frac{\Delta Y}{Y} = \theta(\Delta K/K) + (1-\theta)(\Delta L/L) + \Delta A/A$$

↓
Rate of economic growth ↓
Technical change

Innovation has become the industrial religion. Business sees innovation as the key to increase profits and market share. All the firms think that innovation is an important driver for success in markets. Therefore, the new purpose is to **foster innovation with all means at disposal**.

For instance, research and innovation are placed at the center of the **Europe 2020 strategy**. This includes the headline objective of increasing spending on R&D to 3% of GDP by 2020. In 2000, the European goal was to create the most dynamic and competitive knowledge based economy.

For 2020, the purpose is to create a smart, sustainable and inclusive growth based on innovation. European Union wants to increase the R&D spending; it is an input to reach innovation as output. Now, Europe is very far from this objective.

Economics of innovation

Economic of innovation is important stream embracing ***all studies concerning innovation***: it studies the nature, the characteristics, the determinants and the consequences of innovation and its diffusion at a level of firm, sector and economic system.

Schumpeter was the first one separating three processes of innovation. It is called the ***trilogy of Schumpeter***: invention, innovation and diffusion.

- ***Invention*** is the materialization of an intuition. It may be scientific or not, driven by economic signals or not (exogenous in nature). Often it is a process of stumbling over and largely random in nature: someone wants to invent something and finds something completely different.
- ***Innovation*** is the transformation of an invention in a new service/product. It is influenced by market signals, such as the pursuit of profit. It addresses new, often latent, needs on the part of consumers or firms.
- ***Diffusion*** is the spread process of invention/innovation into the economy. Schumpeter stressed that diffusion is slow: innovators may find difficulties to penetrate the market with their innovations.

Ex-ante it could be difficult to understand the best purpose of the technology invented in order to transform it in an innovation.

Schumpeter was the first providing a ***taxonomy of innovation***:

1. ***Product and process innovation:***
 - Product innovations refer to the creation of new goods and new services.
 - Process innovations refer to the development of new technologies for producing goods or new ways of delivering services.
2. ***Radical and incremental innovation:***
 - Technological breakthrough that might create new markets.
 - Incremental innovation means small adjustments over existing products.
3. ***Drastic and non-drastic innovation:***
 - Drastic innovation causes a great saving so that the company is able to lower the price below the price of competitors, being an unconstrained monopolist.
 - Non-drastic innovation gives to the innovator a cost advantage but it is not an unconstrained monopoly power. The price remains greater than the price of competitors.

Schumpeter mark I

It refers to the **paleo-schumpeterian entrepreneur**. The individual is capable to **transform an invention into an innovation** through an **entrepreneurial act**. The main actor of innovation is the entrepreneur.

He is a person with peculiar characteristics that wants to create innovation with not only economic purpose but also an objective of **realizing its personality**. Economics play a role in the expectation of **supra normal profit** arising from being the first in the market.

If there is no extra-profit as a pay-off of innovative activities, the innovation rate is low. If economy does not allow the entrepreneur to gain extra-profit, the economy would reach a lower rate of innovation.

However, the extra-profit generated by innovation leads to **imitation**. Over time, the entrepreneurial rent of the innovator is eroded by competition from imitators. There is a need of **patents and intellectual property rights regulation** in order to stimulate innovation.

The higher the number of entrepreneurs in the economy, the higher the dynamic efficiency. It is the process of **creative destruction**, which means competition through innovation results in high level of dynamic efficiency. Innovation comes mostly from **small enterprises, new firms and entrepreneurs in general**.

Schumpeter mark II

It refers to the **neo-schumpeterian vision** (1942). It was the period of high innovative performance of the capitalist system, characterized by the presence of **large oligopolistic companies**. Big companies were really gaining momentum in that period and the R&D process became a routine. Therefore, large companies were in a good position to perform R&D and innovation benefitting from economies of scale. He understood that **new firms were suffering increasingly from market imperfections** and they did not have the funds to perform R&D. Thus, in **concentrated markets, innovation comes mainly from big companies**.

The debate is about the relationship between innovation and market structure, and about the role of the public policy.

Three questions about Schumpeter theories:

- 1) **Which is the market structure favoring innovation?** There are two different views:
Schumpeter vs Arrow. The Arrow's view is that the more a market is **perfectly competitive**, the more there will be innovation in the market. The process innovation will cannibalize the profit gained by the monopolist. Before innovation, the monopolist was already gaining some profits, so the profit coming from innovation is replacing something already existing. Therefore, monopoly provides less incentive to innovate than competitive industry because of the **replacement effect**.

The two views may be compatible: incentives to invest in R&D may be superior the more market are competitive but this does not mean that companies in perfectly competitive

markets will be able to invest (lower margins, more market imperfections...). Incentives to invest in R&D are different from capabilities to do so.

The relation between **intensity of competition and innovation rate** has an inverted U shape. It depends on the type of innovation we are dealing with. Therefore, **new firms introduce radical innovation and large companies introduce incremental innovation.**

- 2) **How much should innovation affect market structure?** We have to introduce the **concept of appropriability**: in order to create a treasure (innovation) there should be a mechanism that allows the innovator to get the most of it. The mechanisms are:

- a. **Exogenous instruments** (intellectual property rights and patents): through regulation of property rights, the innovator gets the right to exclude others to use the innovation without authorization.

The patent is the right to use on an exclusive basis the innovation. The company have to pay some fees on a regular basis. There are **some requisites needed to get a patent**: there are subjects eligible for patents, the invention should be new, no obvious (not trivial), and devoted to an industrial use (commercial interest).

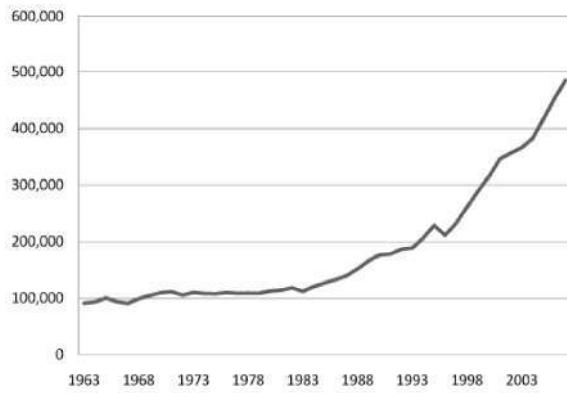
- b. **Endogenous instruments** (imposed by innovator itself): secrecy, lead time (being always ahead of competitors), complementary investments to protect innovation (brand, advertising...).

Often patents are not effective because not always a technology is easy to codify: when you ask for a patent, you have to fill a form **explaining the technology**. It is not easy to translate the physical technology on paper. Moreover, patent means to **diffuse innovation**: there are information externalities spread in the business world. Other firms may **invent around** this information coming out with something similar. It is a problem of reverse engineering to create a sort of substitute for innovation. It is difficult and very expensive to demonstrate that other companies have infringed the patent.

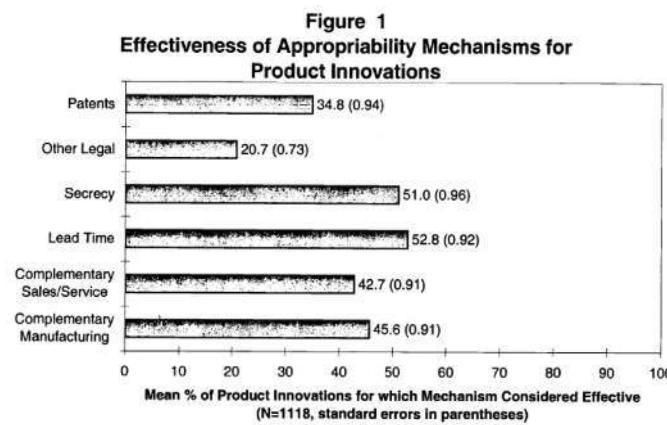
Historically, there were two different views on the use of exogenous instruments:

- c. **Incentives view**: they grant a monopoly power but they are necessary to ensure dynamic efficiency.
- d. **Openness view**: they create balance between dynamic and static efficiency.

This debate has been historically solved much more in favor of the incentives view rather than the openness view. However, the openness view is gaining momentum because the number of patents is increasingly growing and patents are no more used with a defensive purpose but with an aggressive attitude. I patent because I do not want other firms to innovate. Patents together with network externalities create a lock-in effect where innovation is based on cumulativeness and interconnectedness of knowledge. This leads to monopolistic power especially when innovation is based on cumulativeness. Lock-in means that all consumers are locked-in a specific technology (ex. Pc software). Therefore, we had a boom of the number of patents with an exponential trend, even if they are only one instrument to defend innovation.



The Cohen, Nelson and Walsh's study on appropriability tools demonstrated that firms do not consider patents so effective in protecting their own innovation.



Patent might be used in order to prevent others from innovate: they are **dangerous in terms of dynamic efficiency**. Technology is very complex and the number of patents on the same product is increasingly growing. This is the **patent thicket phenomenon**. It is impossible for one firm to have all the patents. Thus, having a patent on a specific and key component blocks the others innovation and give more bargaining power in a licensing transaction.

It is a "**Rembrandts in the attic**" attitude: I do not know if this technology will be useful in the future but I patent. This attitude is more possible for large firms with a lot of money because obtaining a patent is costly.

There are also weak patents: stupid patents. Weak patents can be used to reduce innovation: large companies can use weak patents not for protecting innovation but for preventing others from innovation.

- 3) **Does public policy have a role in the innovative process?** Yes, It has a role in patents system and a wider role. The innovative activity in economy is strongly influenced by several forces: competition policy, regulatory and law regimes, patent system. They all contribute to shape the interested dynamic, without forgetting the role of the institutional and cultural context. The public policy has a role in two areas:

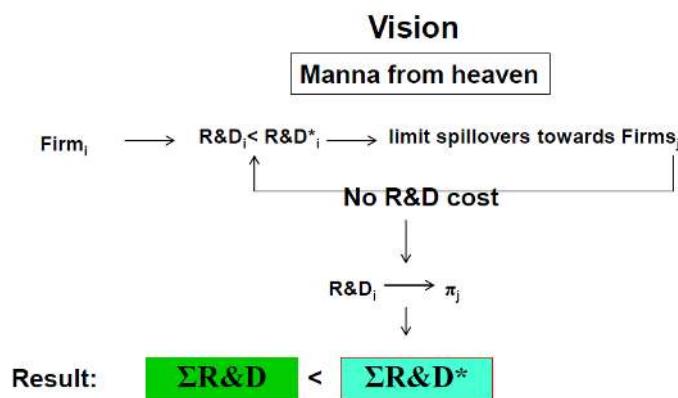
- It has the power to set the **basic rules of the game**: best conditions to favor innovation.

- In the short run, ***it sustains R&D activities*** (subsidies). Private firms may invest less than the social optimum for two reasons: ***spillovers*** (externalities) and ***capital market imperfections*** (presence of asymmetric information between firms and investors).

Spillovers lead to a global under provision of R&D expenditures. The primary output of R&D investments is the knowledge of how to make new goods and services, and this knowledge is not rival-use by one firm does not preclude its use by another. To the extent that knowledge cannot be kept secret, the firms undertaking the investment cannot appropriate the returns to the investment in it, and therefore such firms will be reluctant to invest, leading to the under provision of R&D investments in the economy.

The argument relies on the stylized fact that any formal and informal mechanism to protect innovation is only partly efficient at the very best.

There are some critiques to the spillover: the existence of spillover is generally accepted but it does not justify a policy intervention. The vision is that a firm reduces R&D investments to prevent spillover from another firm but this has an impact on the profit of the other firm. Therefore, the sum of R&D expenditures from both the companies will be lower if one firm decides to reduce the investment.



This is the problem of ***absorptive capacity***: the firm has to invest in order to exploit the R&D of the other (to combine the two R&D). If the two R&D are complementary, we are not sure of the results.

Concerning ***the capital market imperfections***, they are adverse selection and moral hazard.

Regarding the ***adverse selection***, there are R&D intensive knowledge activities on which firms know (much better) than external investors the technicalities of the project and its odds of success (inherently uncertain and risky).

Firms are reluctant to reduce asymmetric information for appropriability hazards and even if this is the case, investors may lack competencies in order to fully understand and evaluate the R&D project.

Investors do not invest or charge a “lemon” premium (Akerlof1970, QJE) for R&D investments respect to ordinary investments.

The cost of external capital for R&D is higher than the internal owing to the lemons' premium (with the risk that only inexperienced "kamikaze" firms are willing to accept investment conditions, that once "failed", further make reluctant investors to finance R&D projects).

The problem is more severe for high-tech start-ups because of the greater shortage of internal funds the absence of a track record as a signal.

Regarding the ***moral hazard***, given the inherent riskiness and uncertainty surrounding R&D projects, it is more difficult for external investors discerning whether a failure was due to managerial misconducts (tendency of managers to spend on activities that benefit them rather than the firm) or not. Managers may take advantage of this greater uncertainty. Investors may simply refrain from investing also due to the scarce "collateral value" of investment.

Given the inherent riskiness and uncertainty surrounding R&D projects, if managers are risk averse (more than shareholders), they may refrain from undertaking potential profitable R&D projects that appear along the way (and will never be known by shareholders and investors). In this case the long-term viability of the firm might be suffering.

Technological change and innovation: the institutional view

The theory of technological change and innovation follows the Schumpeterian patterns that are technological regimes.

Schumpeter Mark I:

- Entrepreneurial regime
- Creative destruction
- Radical innovation
- High opportunities
- Low appropriability
- Low cumulativeness
- Lower entry barriers
- New firms are likely to dominate the innovation activities
- New firms are more likely to survive

Schumpeter Mark II:

- Experienced incumbents regime
- Creative accumulation
- Incremental innovation
- Lower opportunity
- Higher appropriability
- Higher cumulativeness
- Higher entry barriers
- Incumbents are likely to dominate the innovation activities
- Lower survival rate of the new firms

The **institutional theory** can be seen by two different perspectives:

- Economic, by North (1990): institutions are divided in two broad groups, formal and informal;
- Social, by Scott (1995): institutions are defined as “cognitive, normative, and regulative structures and activities that provide stability and meaning to social behavior”.

Both these concepts describe institutions as “**the rules of the game**”.

Market conditions are not sufficient to describe environment for business activity, as well as for technological change and innovation.

The institutional approach complements market characteristics with other formal (for example, regulations and laws) and informal (social norms, public opinion, code of behavior...) contextual characteristics.

This theory is used for studying different contexts for innovation like IT, manufacturing, national innovativeness at large, environmental innovation and so far so forth.

The institutional view

Introduction

It is based on the work of William J. Baumol, 1990.

Entrepreneurship is argued to be an explanation for both historic slowdowns and great leaps in economic growth. It is also argued to be the main driver of technological change and innovation.

This is attributed to the “spirit of entrepreneurship” or rather rate of entrepreneurship, which waxes and wanes for unexplained reasons.

If that was the truth, the policy makers would have no guidance to reawaken that spirit and by that, would have no means to affect technological change or economic growth.

It is a pessimistic view and essentially incorrect.

Entrepreneurship has been present since the beginnings of the civilization and it will probably prevail in the future.

What changes in the environment (i.e. the rules of the game), the reward structure in the economy, which heavily determines how the entrepreneur acts at a given time and place.

It also dictates the ultimate effect on the economy via the allocation of entrepreneurial resources.

Hence, the role of policy in entrepreneurship dynamics can better influence the type of entrepreneurship activities, rather than its supply.

Policy should try to induce a more felicitous allocation of entrepreneurial activities, by adjusting the rules of the game.

The prevailing rules that affect the allocation of entrepreneurial activity can be observed, described, and ideally modified and improved.

Nevertheless, the rules of the game change rather slowly and studying them and their influence are rather cumbersome.

Hence, a historical illustration is needed to encompass all the main economic periods and places important for innovation and its diffusion.

These evidences show that relative rewards to different types of entrepreneurial activity have varied dramatically, and that this seems to have had profound effects on patterns of entrepreneurial behavior.

In turn, that suggests that reallocation can have a considerable influence on the prosperity and growth of an economy, as well as on innovation and technological change.

Conceptual framework

The institutional view relies on the Schumpeterian analysis (Schumpeter I in particular) → entrepreneurial regime, where entrepreneurs are likely to dominate the innovation activity.

The institutional view is an extension of that model to encompass the allocation of entrepreneurship – that bypasses its main shortcoming – policy insights.

The Schumpeterian concept of innovation deems innovation to take various forms besides mere improvement in technology.

There are five different types of innovation according to the Schumpeterian concept:

1. **Product innovation:** the introduction of a new good or service – one with which consumers are not yet familiar, or one of a new quality level;
2. **Process innovation:** the introduction of a new method of production – one not yet tested in the concerned manufacturing branch;
3. **Resource innovation:** the conquest of a new source of supply of raw materials or half-manufactured goods, irrespective of whether this source already exists or whether it has first to be created;
4. **Market innovation:** the opening of a new market, that is a market into which the particular manufacturing branch of the country in question has not previously entered, whether or not this market has existed before;
5. **Industrial organization innovation:** the carrying out of the new organization of any industry, like the creation of a monopoly position (for example through trustification) or the breaking up of a monopoly position.

The institutional view expands Schumpeter's list to encompass allocation of entrepreneurial resources and its role in creating public welfare. It adds aspects of quality of the activity or value added to the society → **productive/unproductive** entrepreneurial activities.

Why? Because entrepreneurs are typically defined to be persons who are ingenious and creative in finding ways that add to their own wealth, power and prestige. Not all of them will be overly

concerned with how much an activity that achieves these goals adds to the social product and public welfare.

Moreover, benefits that the economy derives from its entrepreneurial talents depend substantially on the allocation of this resource between productive and unproductive entrepreneurial activities.

There is the need for aligning wealth/power creation activities (personal goals) and constructive/innovative activities (public welfare goals).

Finally, three main propositions about the institutional view:

1. **The rules of the game** that determine the relative payoffs to different entrepreneurial activities **change** dramatically from one time and place to another;
2. **Entrepreneurial behavior changes** direction from one economy to another in a manner that corresponds to the variations in the rules of the game;
3. **The allocation of entrepreneurship** between productive and unproductive activities can have a profound effect in the innovativeness of the economy and the degree of dissemination of its technological discoveries.

Historical evidence

Ancient Rome, 8th century BC – 5th century AD:

- Romans were openly seeking for power, prestige and wealth
- However, the wealth and prestige were not pursued through participation in industry or commerce
- Honorable status was earned only through one of the three acceptable sources of income:
 - Landholding – typically absentee
 - Usury – action of lending money at unreasonably high rates of interest
 - Political payments:
 - booty: gaining good through looting;
 - indemnity: using exemption from legal responsibility for gaining profits;
 - provincial taxes and money extraction
- No corruption, but rather wrong structural values in society
- On the contrary, commerce and industry were operated mainly by freedmen
- Freedmen were former slaves who bore a social stigma for life
- Slavery was a channel to advancement for the lower class:
 - Lower class → slave → respect of owner → freedom with personal fortune → investments in commerce and industry
- The Roman reward system, although it offered wealth to those who engaged in commerce and industry, offsets this gain through the attendant loss in prestige
- There are evidence on the vigor of innovative activity in that society too:
 - The museum at Alexandria was the center of technological innovation in the Roman Empire. By the 1st century BC, that city knew of virtually every form of machine gearing that is used today, including a working steam engine. But these were used only to make what amounted to elaborate toys. For instance, the steam engine was used to open and close the doors of a temple.

- The Romans also had the water mill. This may well have been the most critical pre-18th century industrial invention because it provided the first significant source of power rather than human and animal labor. However, it was not widely used before the 5th and 6th century.
- Even though evidence of Roman technical stagnation is only erratic, the stagnation in technological change was real.
- At least some part of the explanation is to be found in the **ancient world's rules of the game**, which encouraged the pursuit of wealth but severely discouraged its pursuit through the exercise of **productive entrepreneurship**.

Medieval China, 6th – 16th century AD:

- At that time in China the monarchs commonly claimed possession of all property in their territories
- When they were in financial crisis, confiscation of the property of wealthy subjects living on their territory was entirely in order
- This led those who had resources to avoid investing them in any sort of visible capital stocks, which was a substantial impediment to economic expansion
- Additionally, high social standing was denied to anyone engaged in commerce or industry, even to those who gained great wealth in the process
- The avenue to success laid in governmental positions, and only a few could gain the access through national examinations
- Corruption was immense too, as people had to earn more than modest salaries that the country was providing them
- Even governmental officials were very prone to corruption
- Enterprises have been subject to impediments deliberately imposed by the officials or they were nationalized if successful, which in the long run kills even the slightest attempts of innovation
- What was chiefly lacking in China for the further development of capitalism was not a mechanical skill or a scientific aptitude, nor a sufficient accumulation of wealth, but scope for individual enterprise:
 - ✓ No individual freedom and no security for private enterprise
 - ✓ No legal foundation for rights other than those of the state
 - ✓ No alternative investment other than landed property
 - ✓ No guarantee against being arbitrary penalized from officials or against intervention by the state
 - ✓ Overwhelming prestige of the state bureaucracy, which put out from the start any attempt of the middle class to be different, to become self-aware and fight for an autonomous position in society
- **The rules of the game** have been heavily biased against the acquisition of wealth and position through Schumpeterian behavior

The Earlier Middle Ages: 9th – 10th century AD:

- ❖ Wealth and power were pursued primarily through military activity (war)
- ❖ Land and castles were the medieval forms of wealth and they were pursued and wars can be interpreted as **pursuit of economic goals**

- ❖ Nevertheless, the medieval nobles were not purely economic men
- ❖ Many of the turbulent barons enjoyed fighting for their own sake, and success in combat was an important avenue to prestige in their society
- ❖ In England, with its institution of primogeniture (the exclusive right of the eldest son to inherit his father's estate), the only socially acceptable choice for younger sons was often warfare and fight as means to make their fortunes, and in some cases they succeeded spectacularly
- ❖ These endeavors had noteworthy entrepreneurial characteristics and profound innovation, which were used mainly in the battles:
 - Introduced stirrup as a requisite for effective horses-related tactics
 - Improved military tactics and strategy tactics
 - Improved organizational schemes
 - Castle building evolved from wooden to stone structures and from rectangular to round towers, which could not be made to collapse by undermining their corners
 - Armor and weapons became much more sophisticated with the introduction of the crossbow, the longbow, and, ultimately, artillery based on gunpowder
- ❖ Despite the vast innovation, they were created by an unproductive entrepreneurship
- ❖ An individual who pursues wealth and/or prestige through the forcible appropriation of the possessions of others surely does not add to the national product
- ❖ Military violence marked the period in terms of unproductive entrepreneurial activity and economic slowdown

→ "The Dark Ages"

The Later Middle Ages, 11th – 13th century AD:

- Pacification efforts by the church and more organized governments
- Revival of towns
- Establishment of a labor force and protection of their rights
- Flowering of new activities like architecture and engineering
- **Water-driven mills** were the most profitable activity:
 - technological advancement
 - Cistercian monks were the entrepreneurs and the drivers of the technological change (no clear explanation about why them)
- The water mills marked a minor industrial revolution, as a remarkable source of productive power:
 - monks' mills, their extensive lands, and their large flocks are reported to have brought scale economies and extraordinary **financial returns**
- The water mills were used for an astonishing variety of tasks, in combination with different mechanical devices and sophisticated gear arrangements
- Industries other than agriculture were made possible
- **The rules of the game** have offered substantial **economic rewards** to exercise productive entrepreneurship

→ "The High Middle Ages"

The Fourteenth century, 14th century AD:

- The High Middle Ages were suddenly ended due to the war (the Hundred Years War, between France and England)
- Considerable increase in military activity
- Payoffs have favored more inventions designed for military purposes
- A pervasive business enterprise of this unhappy century of war was the group of greedy troops who roamed Europe, supported the side that could offer the most attractive terms; in breaks between fighting, when unemployment threatened, wandered about thinking up military enterprises of their own, at the expense of the general public
- Renewal of rewards to military enterprise played a significant part in the **slowdown of technological change and economic growth**

The Early Rent Seeking Ages, 16th – 18th century AD:

- ❖ Entrepreneurship took less violent forms
- ❖ Military activity was fading away
- ❖ Entrepreneurial activities usually involved various types of rent seeking, a prominent type of unproductive entrepreneurship
- ❖ Land granting was the most lucrative business
- ❖ Long history – started in the 12th century
- ❖ Picked up in the 15th century and lasted until 18th century and the industrial revolution
- ❖ **The rules of the game** and the related reward system was **not** fostering productive entrepreneurship

The Industrial Revolution, 18th century AD – present:

- The industrial revolution that started in the 18th century and continues today has brought the industrialist and the business persons generally a degree of wealth and respect probably unprecedented in human history
- In general, the industrial revolution brought up the right reward system for entrepreneurial activities.

Entrepreneurship and the rules of the game

If entrepreneurship is the imaginative pursuit of position of power (wealth and prestige), with limited concern about the means used to achieve the purpose, then we can expect **changes in the structure of rewards** to modify the nature of the entrepreneur's activities.

These changes can sometimes be even drastic.

The rules of the game can then be a critical influence helping to determine whether entrepreneurial behavior will be allocated predominantly to activities that are productive (innovative and constructive), unproductive or even destructive.

Unproductive entrepreneurship today

- Monopoly chasing using (hostile) takeovers
- Rent seeking activities like litigations
- Tax evasions
- Misuse of legal gambits to appropriate profits

The **problem** is that entrepreneurs think in an unproductive way, while still being innovative. These activities are tempting, as they can yield profits with least efforts.

What can policy makers do?

If reallocation of entrepreneurial effort is adopted as an objective of society, it is far more easily achieved through changes in the rules that determine relative rewards than via modification of the goals of the entrepreneurs and prospective entrepreneurs themselves.

Policy changes can be made quickly and profoundly.

Contemporary entrepreneurship

Sustainable development represents the main paradigm and goal today.

There are three pillars: **economic, social and environmental**.

The entrepreneurial activity has been an important force for social and ecological sustainability. However, entrepreneurs are compelled to environmentally degrading behavior due to the divergence between individual rewards and collective goals for sustainable development →prisoner dilemma or green prison.

Entrepreneurs can liberate from the green prison by altering or creating the institutions – norms, property rights, and legislation – that establish the incentives of competitive games.

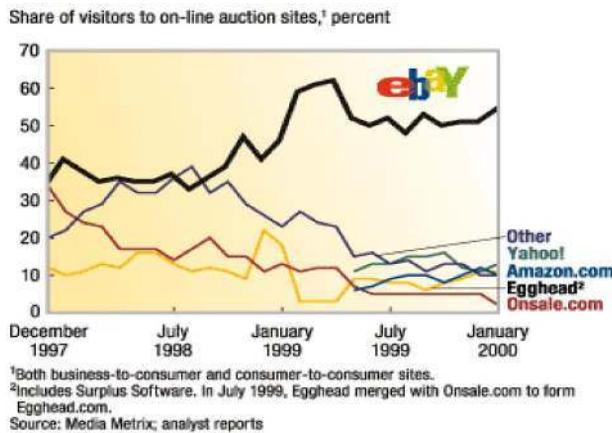
Conclusion

It is not supply or rate of entrepreneurship that policy makers can influence profoundly, but rather the **type of entrepreneurship**.

Firms are, by definition, striving for profit maximization.

Policies makers should create **the rules of the game** such that profit is maximized through **productive entrepreneurship**, rather than unproductive or destructive entrepreneurship.

NETWORK ECONOMICS

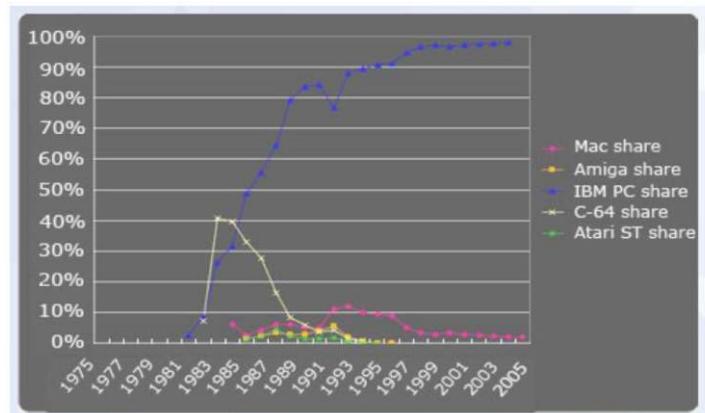


This is the market share evolution of online auction sites. At a certain point of time, one firm took the lead and the others started losing market share. Why Ebay? Because Ebay was more capable to establish itself as a market leader. It delivered a service considered better by consumers. The crucial points are the particular characteristics that allow a firm to emerge as a leader. The **leadership position is driven by the market characteristics.**

Market share by definition depict a relative picture. The market in absolute terms grew a lot: Ebay has a relative leader position in a huge market. The Ebay leadership position has been quite stable over time.

The **initial phase of the market** is the crucial phase to understand who will be the leader.

A similar case is the case of IBM.



Again, at a certain point of time, one platform started to gain momentum and market share while the alternatives started losing space. The leadership position is still in place today.

In order to understand this concept, we could introduce the **garbage bin example**.

In the town, there are six guys that like playing videogames: Alan, Bud, Charlie, David, Eliah, and Frank. Their utility function is characterized by externalities: their preferences depend on the intrinsic value of the PlayStation plus a bonus that depends on the number of other guys owning

the same type of PlayStation (why? because they can exchange games, challenge friends, and have a greater variety of games).

$$U_i = X_k + w * N_k$$

X_k = intrinsic value,

N_k = number of consumers owning type k playstation,

w = parameter (let us say, $w=0.2$)

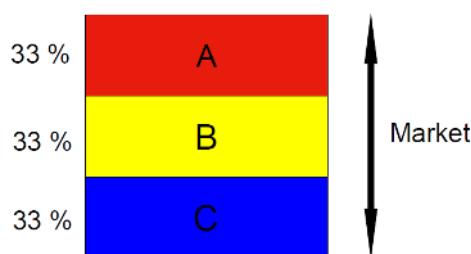
There are three types of PlayStation (a,b,c). It is a **standard war**, which means different firms are competing to establish their products at the expense of the products of competitors: each PS producer invested in R&D and managed to patent some product features. Moreover, each invested in marketing and advertising.

Yesterday, the guys received the catalogues, illustrating technical features (pre-emption). Therefore, tonight the guys will read the catalogues, assign their preferences (intrinsic value X_k) and make their consumption decisions. They will assign different preferences to different playstations, since they value different characteristics.

When they fall asleep, their structure of preferences is:

	A	B	C	Pref
Alan	0.4	0.6	0.5	B
Bud	0.6	0.5	0.4	A
Charlie	0.5	0.4	0.6	C
David	0.4	0.6	0.5	B
Eliah	0.6	0.5	0.4	A
Frank	0.5	0.4	0.6	C

This is only X_k , since there is no externality until the first consumer purchases the first PS. The potential market structure is:



Charlie's cat is very hungry on the monday morning and happens to turn the garbage bin upside down. Charlie wakes up suddenly, it is 8 am. The other guys in town keep on sleeping. Charlie thinks

"since I am awake, I might go to the games shop and buy my playstation". He goes and his purchase decision is obviously to buy C.

From the moment Charlie makes his purchase, the utility of the other customers change.

	A	B	C	Pref
Alan	0.4	0.6	0.7	C
Bud	0.6	0.5	0.6	A/C
Charlie	0.5	0.4	0.6	C
David	0.4	0.6	0.7	C
Elijah	0.6	0.5	0.6	A/C
Frank	0.5	0.4	0.6	C

The second customer is very likely to buy C. If he does so, the third will certainly buy C and the market will be locked in. C will win and take all the pie. The **first purchase decision** changes dramatically the utility of all the other consumers. To the intrinsic value of one product, we have to add that now there is one customer (w^*N_k). This increases the **value of the product**. Now, there are much more probability that a second consumer will buy the same product. If the second consumer buys the same product, the market is locked in this product. There is no more space for competing for the other producers.

Because of this dynamic, this type of market are called "**the winner takes all" market**". If you are in the second place, you are the first of the losers. Market has not the second best price. In these markets, you need luck. It is not sufficient to do all the right moves.

In these markets, there is a **pre-emption tendency**, which means time to market is a very important competitive feature.

Open questions:

- What if the product has the worst technology: all the system is locked in the wrong technology. The risk that the worst technology wins over the others always exists. It is very difficult to find the real case that testify that this occur.
- What about company A and B? They can do very few things inside that market in order to reverse the situation. They can open up a **new competition by innovating** through a breakthrough, something radical.
- Are companies happy with such a high-risk profile of competition (winner takes it all)? They do not want to incur in a standard war. There is always a tendency to avoid this type of war **coordinating before launching the new product**. This dynamic has a **coopetition logic**: collaboration before launching the product + competition in the downward side of the market. They agree on a **common single standard** and then they compete in the market in order to establish their product instead of the others. Therefore, it is difficult to observe the dynamic of "winner takes it all".

We can notice that some accidental things happening in a given moment of time can have a long lasting effect on the market. Initial conditions matter a lot to define the outcome.

Network externality

According to the definition, a good exhibit a network externality when the ***positive change in the utility*** a consumer derives from it raises as ***the number of consumers purchasing the same product increases***. This is the case of telephone, email, hardware and software, party and so on.

Products may exhibit a network externality from their ***technical features***. It is not a psychological matter.

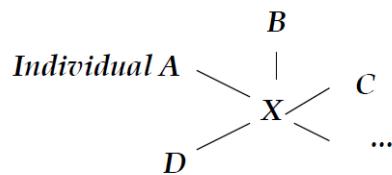
There are also ***negative network externality***: if one more consumer enters the network, it has negative effect on the utility of others. This is related to a ***congestion phenomenon***.

Now we assume to refer to a positive network externality.

Two main typologies of network externality:

- ***Direct externality***: the value of the good increases automatically as the number of users increases. Generally, the good/service has no intrinsic value: it derives all the value from the possibility of making people communicate.
- ***Indirect externality***: the value of the good increases as the number of users increases not automatically, but only in the presence of economies of scale in production because of a greater offer of complementary products (or a better quality) of the good. With the number of users, also the offer increases. One user has a greater variety of goods at disposal. Software is characterized by indirect externality. In this case, often the good has an intrinsic value.

Direct externalities generate a ***two-way network*** (telephone), which are all the networks where service AB is different from service BA: they are two different goods.



The ***value*** of the network is a ***function of the active links inside the network***.

In a two-way network composed by n knots, there are ***n(n-1)*** ***potential links***. The entry of a new user produces a positive externality on existing users, adding 2n new potential links.

The ***Metcalfe law*** is an empirical rule to estimate the value of a 2-way network. If a network is composed by n users and each user assigns a value to the network which is proportional to the number of users, the value of the networks is $V = n^2 - n$. If n is large, the value is $V = n^2$.

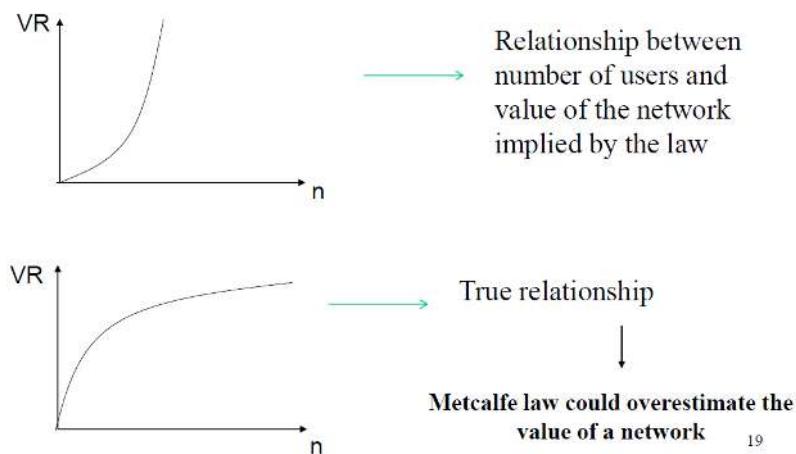
$$w = v(i)f(n) \quad i = 1, \dots, N$$

$$f(n) = n$$

$$v(i) = c$$

$$\text{VR} = \int_0^n v(n)f(n)dn = \int_0^n cn dn = \left(\frac{c}{2}\right)n^2$$

The problem is that this law overestimates what is the true value of the network.



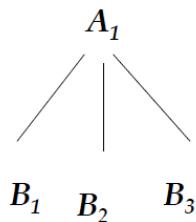
The true relationship is concave because the marginal contribution is decreasing as the network becomes larger.

An example of bad application of the Metcalfe law in the real world was the ***optical fibre in US***. It was a high-tech commodity. In the 90s, the expectation was a very high growth of the industry. Financial analysts provided high rating to companies with the greater networks. The result was a high excess capacity. Consumers used only the 2% of the network available, leading to a price war, negative profits for most of the companies and failures.

Metcalfe law was probably the most well-known empirical rule for providing an estimation of the value of a network. Other measures have been proposed and are used to provide this estimation:

- **Sarnoff law:** the value of the network is proportional to the number of users (very used for broadcasts).
- **Reed law** (value of the social network): given a network composed by N individuals, the value of the network is proportional to $2^N - N - 1$, which is the number of subgroups (communities) potentially formable at its inside.

Indirect externality generates a ***one-way network***: when one of AB or BA is unfeasible, or does not make economic sense, or when there is no sense of direction in the network so that AB and BA are identical, then the network is one-way.



It is the hardware-software paradigm: the greater the usage of a hardware, the greater its attractiveness in terms of developing software. An increase in the number of software further raises the attractiveness of the hardware and increases the number of new adopters and so on. It is called ***bandwagon effect*** or ***positive feedback***.

Ebay was able to trigger a bandwagon effect at the expense of competitors.

IBM-Intel-Microsoft vs Apple case:

IBM dominated the market but they did not believe that there were a market for personal computers. This opened possibilities to other companies (startup → Schumpeter). The most important entrepreneur was Steve Jobs with Apple.

Apple II was launched into the market meeting the expectation of many consumers. The market was very small but it was big enough to change the view of IBM. The problem was that IBM did not have anything new in order to counterbalance the success of Apple II.

IBM tried to make some agreements with the providers of fundamental parts of computers. The most common operating system was CP/M and the microprocessor was 8-bit. IBM wanted to enter into the market doubling the microprocessor capacity to 16-bit. IBM was forced to look for an agreement with an external part.

Apple adopted a closed standard strategy: he wanted to control everything about the computer. IBM was forced to adopt an open standard strategy. IBM made an agreement with Intel, which was the main provider of microprocessors.

Therefore, IBM looked for an operating system for this new microprocessor. Bill Gates bought the operating system MS/DOS and the rights. He made this system available for IBM.

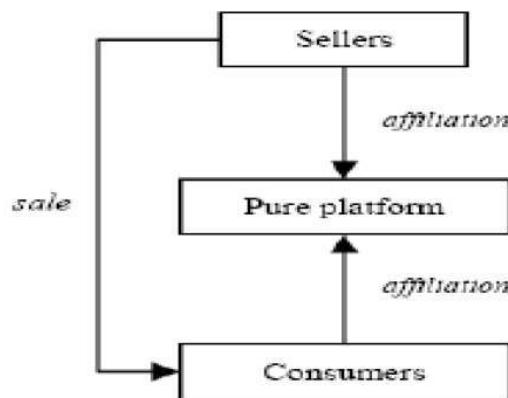
This is the crucial point: ***IBM agreed that Bill Gates could sell without an exclusive agreement***. He retained the property rights. The garbage bin element in this story is the choice of selling the system without property rights and the choice of Bill Gates as supplier. It was the very short time changing the story.

The bandwagon effect element is the following: ***IBM was a highly reputable name*** for consumers and software developers. Developers started to write software only for IBM and consumers were more attracted by IBM even if Apple had the first mover advantage.

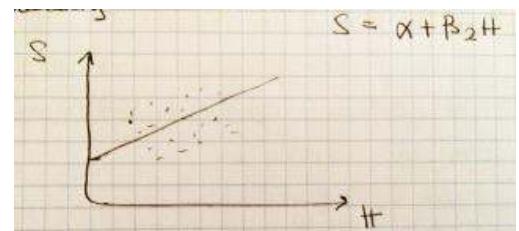
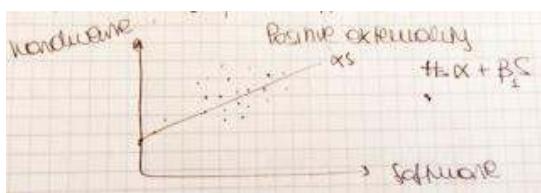
IBM is not even producing personal computers anymore. The platform became a standard in the market but then it has to face new competitors: HP, Dell (clones). Their computers adopted the same Intel platform.

Three remarks:

- 1) **Bandwagon vs negative feedback:** if in the market, there are other competitors and a company is experiencing a bandwagon effect, someone else is experiencing a negative feedback. The lower is the number of users of a platform, the lower is the incentive for developers of complementary goods (software) to create contents for such platform. The lower is the attractiveness of the platform, the lower is the number of users using the platform. It is a negative network feedback.
- 2) **The two-sided markets** are those where a provider enable interactions between end-users, and try to get the two sides on board. There is a provider creating an interaction between sellers and consumers (ex. Ebay, online auctions, credit cards, crowdfunding). The party that exert a greater externality takes the affiliation charge.



- 3) **Empirical estimates of network externalities:** many scientific studies on network externalities. For example, there is the **complementary goods approach** deriving a system of equations and using the number of software available as a variable in hardware adoption regressions and vice versa.



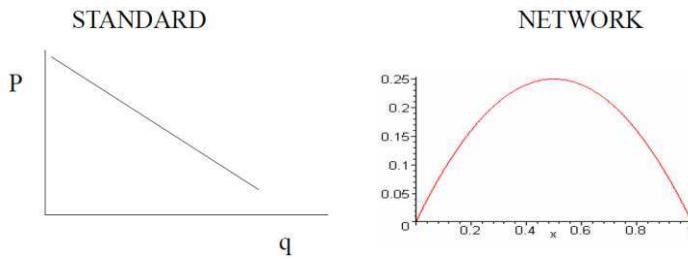
There are also:

- The hedonic approach: the installed base is treated as a product characteristic that will have a positive effect on prices if there are network effects.
- The adoption approach: the installed base at t-1 carries a positive expected sign in the adoption or diffusion equation at t.
- The timing approach: firms with higher expected network effects will adopt a technology earlier and proxies network benefits by the number of potential users of the technology.

Innovation in a network market: the start-up problem

Different shape of a demand curve in a network market derive from the intrinsic characteristics of a network good. With respect to the standard demand curve, the demand curve for a network good does not depend only on price but also on the **agents' expectations about total consumers** of the good. Indeed, the value of what you want to purchase depends on the choice of other consumers. Key implications:

- **Network failure:** if there are very pessimistic expectations about a network good, it will be unsold. In a one-way network, this failure is possible due to the **chicken-egg paradox:** consumers will buy a hardware if there are many software available for this hardware, but the same is for software developers. They write a software only if a great number of people has already bought the hardware since they are potential buyers: if nobody makes the first move, the network goof will be unsold (regardless the price). It is common that one single company owns both the sides.
- The **aggregate demand curve** for a network good could be **upward sloping**: as the network increases, the value of the network good increases and so the reservation price of consumers increases as well. If a firm is able to influence the expectations of potential consumers about the penetration that the good will have (and so its value in the future), it is totally plausible that as the expected size of the network increases, consumers are willing to pay higher price for joining the network.



- Demand presents a **critical mass effect:** for any given price charged by the firm, it is the minimal amount of consumers, who join the network and are satisfied of this choice. Any larger amount will trigger **the bandwagon effect**, any smaller amount will bring to a **network failure**.

Stylized example: firm A has invented and patented a network good (exhibiting direct externalities) and is going to commercialize it under a monopolistic regime.

Trade-off

- High price and a high number of people to convince
- Low price and a small number of people to convince

Suppose:

- $p = 1\text{€}$; at that price a representative agent may be willing to buy the good if it can be enable to communicate with (say) 100 people.
- $p = 100 \text{ €}$; at that price one may be willing to buy the good if it can be enable to communicate with (say) 1.000.000 people.

If the choice of the firm is $p = 1$, it has to convince 100 people. If the choice of the firm is $p = 100$, it has to convince 1.000.000 people.

Whatever the choice, if the firm wants to sell the good it has to attract the critical mass (which is increasing in price).

If $p = 1\text{€}$ but it convinces only 99 individuals, someone of these individuals will be unhappy (someone who gives to the good a value of 1€ only if 100 individuals had joined the network); so he will leave; the network size will shrink to 98 individuals and so on.

If $p = 1\text{€}$ but it convinces more than 100 individuals, e.g. 101, the value of the network good raises, and some more agents will want to buy the network good, network size becomes 102, and the value of the good raises and so on.

The aim is attracting the critical mass in order to generate the bandwagon effect.

Aggregate demand curve for a network good: consider a network of interest for N potential agents. They are indexed by x , which is uniformly distributed on the interval $[0,1]$. Each agent faces the binary decision of whether to buy a good or not. The good exhibit positive network externalities.

If the agent buys: $U = (1 - x) \cdot n^e - p$

If the agent does not buy: $U = 0$

Since a continuum of potential consumers exists in the interval $[0,1]$, there will be therefore a particular consumer, indexed by x^* , such that she is indifferent between buying and not buying the good. This consumer is found by:

$$0 = (1 - x^*) \cdot n^e - p$$

Or

$$x^* = \frac{n^e - p}{n^e}$$

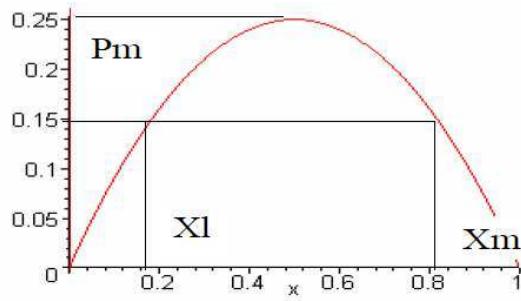
Hence all the consumers that have a higher willingness to pay for the service ($x \leq x^*$) will buy the good, while all the agents that have a lower willingness to pay ($x > x^*$) will not.

Assuming $n^e = n = Nx^*$, the inverse aggregate demand curve for a network good becomes

$$p = (1 - x^*) \cdot Nx^*$$

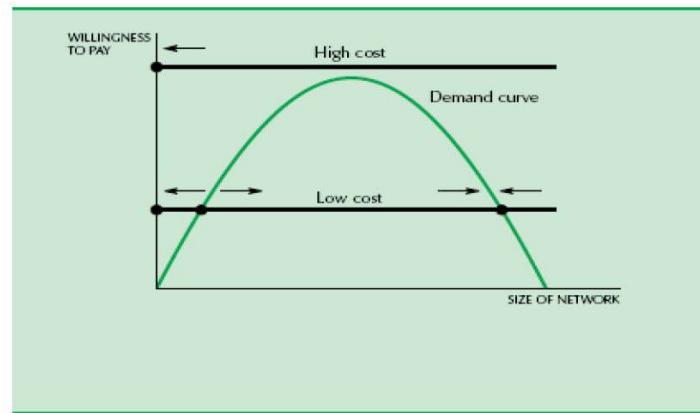
There are three possible situations:

- Network of zero size (network failure caused by pessimistic expectations)
- Critical mass → unstable equilibrium
- Network of large size → bandwagon effect



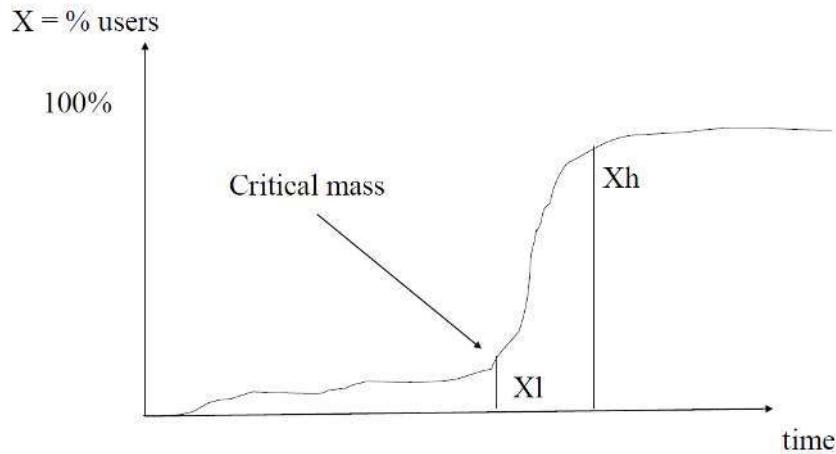
If price is greater than p_m , there is no market for the network good.

It is plausible to assume that when **people are willing to pay more** than the cost of the good, the size of the **market expands** and, when they are **willing to pay less**, the **market contracts**. Geometrically this is saying that when the demand curve is above the supply curve, the quantity goes up and, when it is beneath the supply curve, the quantity goes down.



When the cost is high, the only equilibrium implies market of size zero. As the cost goes down, other equilibria become possible.

Looking at a dynamic framework, the typical path of a network good is a logistic curve highlighting three phases:



Until the point X_l , we have the **launch phase**. After the critical mass, we have the **rapid growth phase** (due to the positive feedback) when the penetration rate raises dramatically in a short period.

Then, there is the **maturity phase** (universal service policies) when it is difficult to convince people to join the good because they have a very low willingness to pay

If a producer sets a price in the upward part of the curve and then is not able to convince the right quantity of consumers, it will be a failure; otherwise, it will generate a bandwagon effect.

The **typical path of an unsuccessful good** is a **straight line very close to zero**. There are examples of goods never reaching the critical mass. Invention does not imply immediately innovation. There is a tension between searching for dynamic and static efficiency.

Strategic policies

They are a way to solve the startup problem:

- **Compatibility with competitors**, which means to share the effort being careful about the free-riding problem. If all companies want to reach the critical mass, there is an incentive to free ride the effort of others.
- **Promotional prices** (very often below costs) to reach a bandwagon effect.
- **Freeware to trend setter, tolerate piracy**
- **Advertising**: it has to convince people also that the product is convincing other people.
- **Invoke public aid**: positive externalities are solved by subsidies.

The role of the state in the launch phase of the product may be risky in terms of dynamic efficiency. In the maturity phase, goods become **merit goods** having a positive effect on the social welfare. The recognition of a merit good enable the state to establish policies to promote it, which are the **policies for universal services**.

Strategic choice between compatibility and incompatibility of products

Giving a definition of **compatibility**: two products are defined compatible if they “can work together”: the output of a given brand can be used by other brands. If this is the case, we say that **different brands adopt the same standard**. In the opposite case, we say that products are incompatible.

There is a **downward compatibility** when a new release of a product is compatible with the old one, but it is not the other way around. It exists because otherwise there is not incentive for people to migrate to the new release of the product.

A product is **one-way compatible** if it can work with a product of another rival brand but it is not the other way around. It exists because it is a survival strategy offering compatibility with the product of the market leader.

In order to understand the consequences of a strategic choice, we have to focus on the **characteristics of a network market**. The success of a standard will depend on the capacity of this standard to **solve the start-up problem** attracting the critical mass. If a firm chooses non-compatibility, it has to know:

- There might be a natural tension of the market towards **monopoly**

- **Early adopters** are extremely important: it is fundamental the capability to attract these adopters because they could represent the critical mass generating the bandwagon effect and eliminating the competition.
- **First mover advantage:** there is a technology-absorbing barrier. The more variety, the less the risk of a “winner takes all” market.

Seeing that formally, we could consider two incompatible goods A and B. Population ($N=1$) is formed by agents a who prefer A and agents b who prefers B.

The utility of a is $v(x)$ if he buys A and $v(x)-\alpha$ if he buys B.

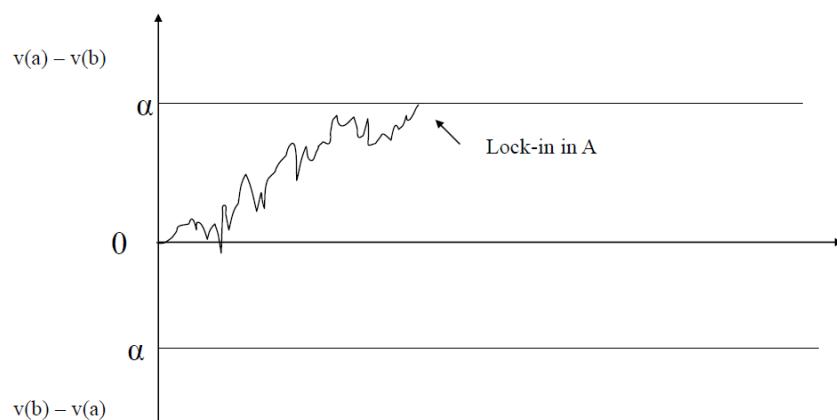
The utility of b is $v(x)$ if he buys B and $v(x)-\alpha$ if he buys A.

We assume that market is perfectly competitive and firms charge the same price. Moreover, we suppose that being the only one to acquire the good will bring zero benefit.

Monopoly will be an equilibrium if $v(1)-\alpha \geq 0$.

Now we assume a sequential entry of consumers into the market. consumers b will buy A only if $v(a)-\alpha > v(b)$ and consumers a will buy B only if $v(b)-\alpha > v(a)$.

The **achievement of a critical mass in a standard war** is the following:



Early adopters are more important (for reaching the critical mass) the higher the network externality and the lower the love for variety. The same logic has to be applied to the first mover advantage.

The model also makes self-evident the risk that ***an inferior technology could emerge*** as winner of standard war. This is a real possibility from a ***theoretical point*** of view, but (fortunately) on an empirical ground is very difficult to ascertain that such risk has ever become reality.

The only controversial case is the one of ***Qwerty vs Dvorak keyboard***. The keyboard is a network good: if a keyboard is common, people know better how to use it. In this case, there is a high degree of path dependence because we are using this system invented 150 years ago even if there are better alternative systems.

According to the definition of ***path dependence***, which equilibrium will emerge in an (economic) system will depend on a ***limited number of initial conditions and decisions made***. Small modification to these initial decisions and conditions will lead to completely different outcomes.

Dynamic processes exhibiting these characteristics are called ***non-ergodic or path dependent ('history matters')***.

In order to know the final outcome, we have to know the whole history. There are different degrees of path dependence:

- 1) ***Path dependence 1°***: a minimal form of path dependence is present whenever there is an element of ***persistence or durability in a decision***. For example, an individual does not alter his consumption of housing services every day in response to changes income or relative prices.
- 2) ***Path dependence 2°: information is always imperfect***; a second circumstance is always possible. When individuals fail to predict the future perfectly, it is likely that ex ante efficient decisions may not turn out to be efficient in retrospect. You may build a house without knowing that five years hence a sewage treatment plant will be built nearby, lowering property values and the neighborhood amenities available.
- 3) ***Path dependence 3°***: given network externalities, ***decision taken in the past by others*** influences the decision that agents have to take today. This can lead to Pareto dominated (rather stable) equilibria for coordination problem. You know a sewage plant is going to be built but build a house nearby anyway since all of your friends are buying houses there and you value being part of that neighborhood. You would rather buy a house away from the sewage plant, and so would your friends, but you and your friends are somehow unable to coordinate your actions.

There are ***two forces towards standardization*** and compatibility:

- Firms have to bear ***high costs in R&D*** to develop a standard and in large measure they are sunk because highly specific.
- Note also that ***marketing expenditure*** and coordination complementors costs can be extremely high when a firm choose incompatibility and opt for a go-it-alone strategy.

Therefore, it is common that firms choose ***compatibility***. This may occur with a formal process of international organizations to recognize the standards. Moreover, firms may choose an ***open standard strategy*** with very low licensing fees for the patented technology they developed or a ***coopetition logic***. Coopetition policies mean that they collaborate upstream and compete downstream.

The logic of open standard could be explained through ***a game-theoretic approach***.

There are 2 firms, A and B, having at disposal two different strategies: lead or follow. There are two possible situation:

a) Battle of sexes game

A\B	Lead	Follow
Lead	(3,3)	(6,4)
Follow	(4,6)	(0,0)

There are two different Nash equilibria: we do not know which firm will be the leader. We know that in this case there will be no standard war, they will find an agreement.

- b) Changing the payoff, the leader gains 8 and the follower only 2.

A\B	Lead	Follow
Lead	(3,3)	(8,2)
Follow	(2,8)	(0,0)

There is only one Nash equilibrium: lead/lead. ***They prefer to fight rather than to cooperate.***

The difference between the two games derives from the ***greed for profit*** of the firm that develops the standard. This is an example in which firms are better off if cooperative or not too greedy. Firms can switch from the second to the first game if the leading firm:

- Sets very low licensing fees for the use of technology (policy of open standard)
- Reduces absorption costs of the follower

Therefore, firms can play the first game choosing an open standard strategy with agreements between leader and follower.

Taking a broader perspective, there is a ***tradeoff in decision***:

- **No compatibility** leads to a standard war with lower probability of success. However, if the firm is able to reach the critical mass before the others, it will take all the pie. A possible result is to obtain a big slice of a small pie with high risk of network failure.
- **Compatibility** leads to higher probability of success in terms of sales (reaching the critical mass) because the market has indirect network externalities. There is a single standard in the market and no doubts on how for example to write a software. The possible result is to take a small slice of a big pie because there is high competitive pressure: there are other firms adopting the same standard. There is high confidence by the agents, low risk of being locked in the wrong technology and low uncertainty by complementors.

There are four possible strategies:

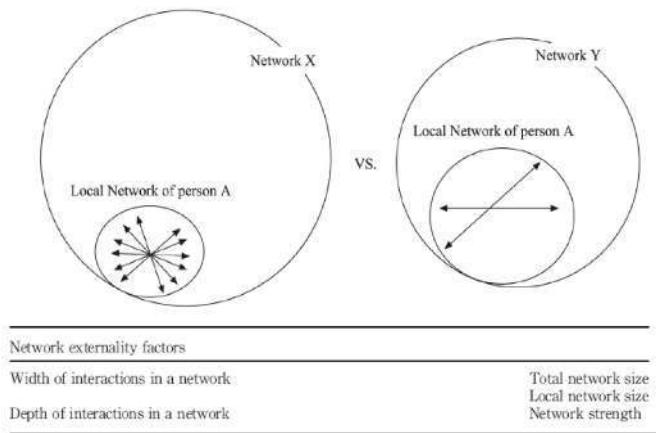
	Closed std	Open std
Lead (development)	SPONSOR/DEFEND	GIVE AWAY
Follow (adoption)	LICENSE IN	CLONE

- **Sponsor/defend** consists of developing a standard and restriction of the use towards concurrent firms (high licensing fees)
- **Give away** consists of the possibility for the concurrent firms to use standard without restrictions or with low licensing fees.
- **License in** means the use of a standard developed by another firm.
- **Clone** means the use of an open standard without restrictions.

	Closed standard	Open standard
Lead	<ul style="list-style-type: none"> + protected market + high margins + high share +/- license earnings - low chance of winning - little external support - high costs - small niche likely 	<ul style="list-style-type: none"> + high chance of winning + large market + broad external support + shared costs - low share - low margins + high competition +/- license earnings
Follow	<ul style="list-style-type: none"> + proven market + possible alliance - secondary position - licence fees - absorption costs - small niche likely 	<ul style="list-style-type: none"> + high chance of winning + large market + license fees - high competition - absorption costs

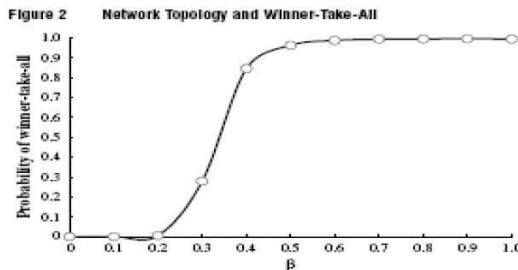
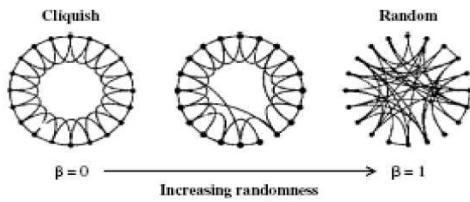
It is not always true that when there are firms with **different products** there is always a **winner takes it all equilibrium**. When there are **externalities** and there is a **lot of variety in the market**, different products can coexist.

There are **some elements preventing the winner takes it all**, one of them is the **community of interests**. This occurs when the randomness is very low and the probability of winner takes it all is very low.



When there are **different network products**, but we are interested in **communicating with the same people**, there is not the winner takes it all. There can be different networks coexisting over time.

If I want to communicate with more people the **randomness increases and winner takes all is more likely to occur**.



In reality, considering which network join, we do not always consider the size of the network. We can consider our friends or other groups of people. The perspective change leading to different decisions.

Games are perishable; this allows firms to open new competition.

This is a dynamic competition. When firms choose no compatibility, the most rationale strategy may be completely irrational. This is related to the ***all in strategy***: you have to be available to put all money in this strategy showing to consumers that you are very committed even if in reality you think it will not be a success. The most rational strategy is to signal an irrational commitment to win the race at all costs. This is an irrational behavior. This is the same of ***collusion in oligopoly***: irrational behavior is the only way of making collusion stable. The ideal price is the monopolistic price higher than marginal cost but, within a collusion, the fixed price is the monopolistic price minus a certain quantity, maintaining anyway the price above the marginal cost. One way to obtain a reasonable result is to show an irrational behavior threatening competitors fixing a price equal to the marginal cost (perfect competition) in order to punish price cutters regardless the consequent ruinous cost.

Some factors positively influence the standard war:

- Size
- Financial resources
- Commercial strength (marketing and distribution channel)
- Brand and reputation

A clear example is the IBM-Apple case.

Cases on standard wars

1) VHS vs Betamax

There are three actors: Sony, Panasonic and Toshiba.

This case is an example of how small firms defeat big ones.

In 1970: Sony and Panasonic, together with JVC, developed a technology for registering scientific events (VCR). This technology was good: it had good commercial success but it was not made for the large consumer market because it was expensive and large. It was a patent-sharing agreement: they both had intellectual property rights. However, both of them wanted to produce a **video recorder for the large market**.

Characteristic	Year	VHS	Betamax
Playing time	1975		1 h
	1976	2 h	
	1977	4 h	2 h
Programming delay	1977	24 h	
	1978	7 days	
	1982	15 days	15 days
Video quality			higher
Size & weight	1977	46x32x16cm, 13,5 kg	53x12x19 cm, 19,2 kg
Other characteristics	1978	Fast/Slow Motion	
	1983	Hi-fi Audio	

In 1975: Sony produced **Betamax**. It had a good engineering design. It was compact and small with good quality of image and sound, and it enabled the fruition of cassettes of **one hour**.

In 1976: JVC/Panasonic entered with its own standard, the **VHS** that performed worse than Betamax with lower quality and design. However, it enabled the fruition of cassettes of the length of **two hours**.

At the beginning, **it was not a network good**. They were thought in order to allow the time shifting with the possibility of pre-recorder a show. There were not network externalities.

Even if a technology is strong, it could be difficult to find a commercial use. Sony soon realized that **despite of its first mover advantage**, as soon VHS entered the market, more consumers were attracted by the new standard. This is because the preferred use is to record movies lasting more than one hour.

Sony tried to push the length of the cassette to **2 hours** and in 1977 entered the market but **Panasonic** launched a recorder that enable fruition of cassettes of **4 hours**.

Indeed, another important use was the recording of sport events, without network externalities. These events last much more than two hours.

Therefore, the community of VHS consumers grew more with respect to Betamax community. This product **became a network good** because the home video industry started. The movies studios with copyrights fought this industry. However, home video industry started with an **entrepreneurial act** by Andre Blay with the 20th century fox and the right to produce 50 movies. He **advertised the production** of the pre-recorded movies. He decided to use the **format of VHS** because this was a larger potential market for his product.

Another segment of this industry was the ***rental video shop***, by an entrepreneurial act. Because of the low availability of movies for Betamax, this entrepreneur chose to rent only movies with the ***VHS standard***.

The ***community of VHS started to grow*** and there were more incentive to enter the market ***producing in the VHS format***. This led to more ***incentives for consumer to choose this standard*** and so on. VHS standard won the standard war.

**SALES MARKET SHARES FOR THE US VCR MARKET,
1978–1986**

Year	Market Share (%)						
	By VCR Brand					By Format	
	RCA (VHS)	Sony (Beta)	Panasonic (VHS)	Fisher (VHS)	Magnavox (VHS)	VHS	Beta
1978	25.5	28.8	14.3	—	6.8	56.8	43.2
1979	34.1	38.1	4.3	—	3.8	55.0	45.0
1980	29.3	28.8	9.3	—	3.9	60.0	40.0
1981	28.0	14.2	15.3	—	4.5	72.4	27.6
1982	22.0	13.0	17.8	2.0	4.0	74.4	25.6
1983	16.0	7.0	15.0	5.0	4.7	81.9	18.1
1984	16.0	6.5	14.0	6.0	4.0	82.2	17.8
1985	13.8	4.8	12.1	7.7	4.0	87.8	12.3
1986	12.0	3.1	11.3	6.8	4.0	91.4	8.6

The picture above shows the market share for the video recorder market. The market share by format follows the described dynamic and the winner takes all the configurations. Looking at the market share by brand, Panasonic was the leader of the standard but its market share was ranging between 9 and 15. It never gained a very strong position because ***Panasonic was a very small firm with respect to Sony***. Therefore, it was forced to adopt an ***open standard strategy*** enabling other producers to use the standard. Indeed, its production capability was not enough. It had to face ***higher competitive pressure*** and gained a ***little slice of a big pie***. It is the same situation of the case of IBM vs apple.

2) DVD vs DIVX

This case explains ***the use of vaporware***, which is the ***strategic use of pre-announcement***. In the network, it creates ***many expectations*** having a very tangible effect.

Toshiba, Sony, Panasonic all agreed on the **DVD** launched for the first time in September 1996.

DVD is a network good with indirect externality. They agreed with movie studios the ***format of the films***. Only with ***coordination***, you can solve the ***chicken-egg paradox***.

The product was commercialized by Best buy. DVD had a good market acceptance but for some movie studios (Disney and Universal) the product was not enough protected from piracy. Circuit city, a competitor of Best buy, ***announced that an alternative product*** would have been soon available in the market, the **DIVX**. DIVX will provide ***one-way compatibility to the DVD*** but it could also enable the fruition of DIVX disks. It had an ***encryption technology***: if you want to decrypt the video, you have to buy a code valid for two days. It was announced in 1997 and launched in 1998. It ***exit the market*** in 1999.

There was a **big problem** leading to the negative result: Circuit city was a retailer, it did not have manufacturing capability. It had to establish some **partnerships with producers**. It was not able to **attract on an exclusive basis the producers of content**.

Some studies have explained the **effect of the DIVX story** (announcement, entry and exit) **on the sales of the DVD**.

$$\begin{aligned} \text{LSALES} = & \beta_0 + \beta_1 \text{LPRICE} + \beta_2 \text{LSOFT} + \beta_3 \text{BOA} + \beta_4 \text{DIVX} \\ & + \beta_5 \text{ENTRY} + \beta_6 \text{DEMISE} + \beta_7 Q^2 + \beta_8 Q^3 + \beta_9 Q^4 + \varepsilon. \end{aligned}$$

Independent Variable	Logarithmic specification (1)	
	Coefficient	T-Statistic
CONSTANT	11.71	1.69
LPRICE/PRICE	-1.70	-0.79
LSOFT/SOFT	0.25	1.18
BOA	5.55	3.74
q^2	0.31	1.82
q^3	0.47	3.36
q^4	0.61	5.75
DIVX	-0.23	-1.61
ENTRY	0.057	0.50
DEMISE	0.015	0.17
No. of observations	39	
Durbin-Watson	1.77	
Adjusted R^2	0.95	

They ran a regression and understood that the product has a network externality. The only thing making a **negative effect** is not a tangible thing, but the **pre-announcement**. It was the **only negative statistical coefficient**.

In order to understand how it was possible, we have to take the typical evolution of a network good (logistic curve). The **DIVX is the launch phase**: the critical mass not achieved. Once arrived at the second phase, the entry, there was not yet room for competition. **DVD** was already able to **reach the critical mass** and exploiting the **bandwagon effect**.

The **one-way compatibility was not a clever move**: there was no incentive for movie producers to move to the new format. Consumers saw that all the movies were available on DVD, therefore, they did not have incentive to buy DIVX. The two elements reinforce each other. It was better to force the movie producers to make a choice.

3) DVD vs BLU-RAY

It is a standard war between **Toshiba and Sony**. This case shows how agreements may change in this industry. Hardware producers have to make important **agreements** with software producers in order **to solve the chicken-egg paradox**.

Before the standard war, they tried to make an agreement in a **coopetition logic**. This attempt failed for three reasons:

- Format were different from a technological point of view.
- Both the companies have already made important investments with a great part of sunk costs.
- High absorption costs for the follower firm in order to incorporate the technology of the other firm in its own product.

The products were launched and commercialized. The DVD was priced at 499\$ and the blue-ray at 999\$. Prices were very high compared to previous technology.

Blu-ray had **higher performances** in terms of capacity storage of the disk. The quality of sound and image was not inferior to the DVD. The higher price was justified.

Blu-ray was incorporated in the PS3. Toshiba made agreements with Microsoft to commercialize Xbox with the incorporated DVD. This **move of the PS3 changed the market**. The DVD lose its leadership position. Until 2006, Sony was more able to **attract movie studios on an exclusive basis**.

HD DVD	Blu-ray	HD DVD	Blu-ray
			
			
			
			
			
			

Now, three companies sponsored only blu-ray. In 2007, Toshiba realized the importance of having contents on an exclusive basis attracting Paramount and Dreamworks. However, it was a **very expensive move**.

The **market is sluggish**: both companies had the possibility to produce a comparable number of movies with their format. The **situation was unstable** and it was clear that the market would evolve in a winner takes all market.

In 2008, **Warner bros** (biggest studio in terms of revenue, 18% of the overall market) decided to turn **only to the blu-ray format**. This choice was taken because blu-ray was more developed and had a higher market share. There is the possibility that Sony offered some money to WB.

It is similar to the Betamax and VHS case: the potential market was larger for the VHS.

Sony understood that there was **no room for the DVD** and it had won the standard war.

It is **important to attract software** and when launching a new standard it is a good move to **incorporate the new product in an already existing and successful product**.

Switching costs and technology replacement

Which factors determine the migration of consumers from an old to a new network technology? The capacity of a new network good to replace the old one is **related to the switching costs**.

If one consumer has to decide whether to migrate or not to a certain technology, in a standard context, he will consider **price and performance of the new product**. In network markets, we have to consider other two factors:

- **Individual switching costs** mostly originated by learning

- **Expectations of consumers** about the **other consumers behavior**

The presence of **individual and collective switching costs** may lead to very **strong lock-in effect**.

If two users should decide if they want to shift from the old hardware platform to the new one, the payoff of each user depends on the decision of the other user due to network externalities.

1/2	New technology	Old technology
New technology	α, α	γ, δ
Old technology	δ, γ	β, β

With $\alpha > \delta, \gamma$ and $\beta > \delta, \gamma$.

The Nash equilibria are (new,new) and (old,old).

It is a very inefficient situation because these two equilibria are:

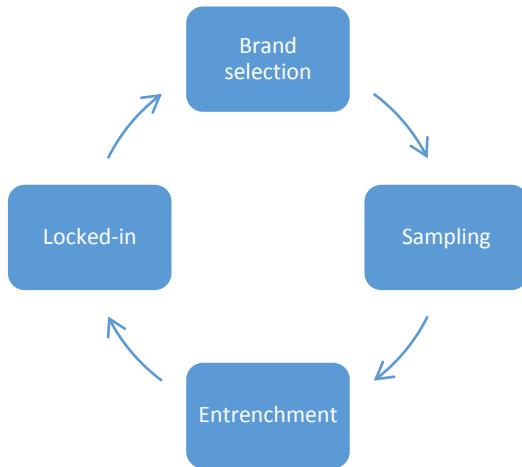
- **Excess inertia when $\beta < \alpha$** [(old,old)<(new,new)]: the moment of changing is very delayed.
- **Excess momentum when $\beta < \alpha$** [(old,old)>(new,new)]: the move from the old to the new technology does not make sense because they benefit more from staying in the old technology.

One equilibrium is Pareto dominated by the other. **Inertia is much more inefficient than momentum**: momentum is very subjective. Inertia is more possible. It is very difficult to have a situation when bad technology defeats a good one, at least in the long-run: it is only theoretical. Some technologies struggle in order to impose themselves, for instance Minitel vs Internet, Windows vs Linux, AM vs FM standard for radio transmissions.

Individual switching costs are characterized by **network externalities**, especially in the ICT market. They are:

- **Learning costs**: learn to use the new technology.
- **Search costs**: know if there is a new technology available.
- **Loyalty costs**: psychological costs, feeling comfortable in a technology.

According to Shapiro and Varian, the **lock-in cycle** is the following. It is related to the learning switching costs. They applied it to the ICT and software industry. When we have to choose a software:



Firstly, we perform a brand selection through the sampling, a free trial of one month. Learning to use it, we are locked-in. When a new version is available, we choose the more advanced version of the already used technology. In this way, producers can exploit the lock-in effect charging ***higher prices***.

Microsoft has intellectual property rights on the operating systems, but it ***has a dominant position also due to the inertia***.

Very often in network markets, ***competition is*** not within the market but ***for the market***. A firm able to ***win a standard war*** gains a very ***dominant position*** winning ***all the market*** in a quite stable situation. ***Market shares*** are often a ***poor indicator*** of the degree of competition in a network market.

However, it is not everlasting. Other firms can ***open a new war trying to launch in the market a breakthrough***. It is a real ***Schumpeterian competition***, where stable monopolistic positions are interrupted by innovation. Microsoft has to be on guard because competition can come from ***unpredictable sectors***.

An example is the Browser war mid 90s: Microsoft entered this market just to defend its dominant position in the operating systems. It is the same move of the Xbox.

INDUSTRIAL AND COMPETITION POLICY

It is a micro-perspective on policy makers' intervention.

It is called ***industrial policy*** and there are two streams of intervention:

- ***Industrial (supporting) policies***
- ***Competition policies***: they refer to two types, anti-trust and regulation

The point is what makes a ***good policy intervention***. The answer is related to ***static (allocative and productive) and dynamic efficiency***. A good policy intervention should deal with all types of efficiency.

However, there could be some ***trade-offs between these types of efficiency***: for example, the cost of protecting innovation generates allocative inefficiency, but innovation generates dynamic efficiency.

Making a brief introduction on ***industrial (supporting) policies***, they are any policy that affects a subset of firms and industries differentially from the remaining group of firms and industries. Any tax, subsidy, trade and other policy measure that affects some specific firms and industries differently can be considered an industrial policy intervention. For example, support to young innovative companies, a government subsidy to the automobile industry for research and development, an imposition of an import quota to protect the automobile industry from import competition, establishment of a consortium and so on.

Concentrating on competition policy, the assumption is that ***competition between firms is good for economy***. The objective of this ***policy is to favor competition***.

The main differences between antitrust and regulation are:

- ***Antitrust*** is mainly an ***ex-post intervention*** applied to ***all firms*** and sectors of the economic system. The functions are to ***punish anti-competitive behaviors*** put in place by firms and ***prevent the formation of artificial dominant positions*** in the market with negative effects on allocative efficiency. This is an ex-ante area referring to merger regulation. If there is a merger, the antitrust has to approve it.
- ***Regulation*** is mainly an ***ex-ante intervention***: they put into the system some ***rules and obligations*** that firms operating in the market have to fulfil. It does not refer to all firms and markets but it refers to a ***specific industry***, for example markets characterized by natural monopoly. The functions are to identify the natural monopoly and to impose conducts to firms so that competition (and its results) can be mimicked.

Antitrust

Antitrust has three areas of intervention:

- ***Agreements between companies that restrict competition***: cartels or other unfair arrangements in which companies agree to avoid competing with each other and try to set their own rules. It is ***collusion*** between firms for example on price.

- **Abuse of the dominant position:** a major player tries to squeeze competitors out of the market using for example predatory pricing. Firms are punished in some cases putting the survival of the firm at risk. The dominant position is not punished. Only the abuse of dominant position is punished.
- **Mergers and acquisition:** they have to achieve approval of antitrust.

Referring to the **Cournot duopoly**, collusion means that **firms coordinate each other** and agree to produce **half of the quantity** that would be served **in monopoly** (green line). Quantity produced is lower and profits for companies are higher. Antitrust has to prevent this kind of behavior.

STRATEGIC VARIABLE: Quantity

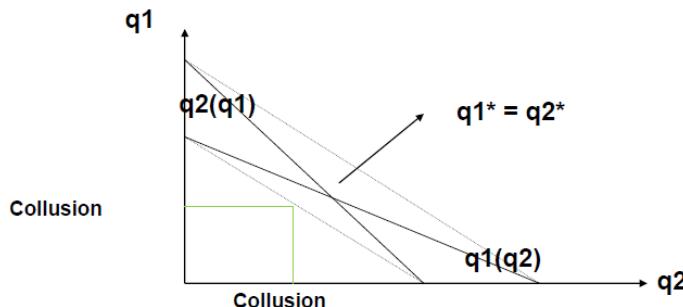
$$p = a - b(q_1 + q_2)$$

$$\Pi_1 = pq_1 - cq_1 = [a - b(q_1+q_2)]q_1 - cq_1$$

$$\Pi_2 = pq_2 - cq_2 = [a - b(q_1+q_2)]q_2 - cq_2$$

$$d\Pi_1/dq_1 = a - 2bq_1 - bq_2 - c = 0 \longrightarrow q_1^* = [(a - c) / 2b] - (1/2)q_2^*$$

$$d\Pi_2/dq_2 = a - 2bq_2 - bq_1 - c = 0 \longrightarrow q_2^* = [(a - c) / 2b] - (1/2)q_1^*$$



Collusion is possible when:

- It is **convenient** for firms participating to the cartel (agreements for prices and quantities).
- The **cartel is sustainable** over time.

All firms want to collude: half monopoly profits is anyway greater than profits in the duopoly. It is always convenient. Regarding the **sustainability**, it depends on the **frequency of interaction between firms** and on how much **firms are likely to play the same game over time**.

Frequency of interaction changed a lot with the use of internet: the firm may change price with just a cost of a click and frequency of interaction is much higher. Collusion is easier because they can understand better the consequences of their strategy through internet.

The game of companies over time is the structure of the **prisoner game**. If the game is **one shot**:

1/2	Cheat	Collude
Cheat	1,1	3,0
Collude	0,3	2,2

Pareto efficient solution is 1,1 even if it is dominated by 2,2. The solution is to cheat.

Nevertheless, **if the game continues over time** and we assume **firms collude until the opponent is colluding**. We assume a **punitive strategy**: once the opponent cheats, the firms cheat as well for all time.

If we take the probability p that this game will be played over time ($0 < p < 1$).

The payoff of the firm in case of cheating is:

$$3 + 1p + 1p^2 + 1p^3 + \dots + 1p^n = 3 + \frac{1p}{1-p}$$

The payoff of the firm in case of colluding:

$$2 + 2p + 2p^2 + 2p^3 + \dots + 2p^n = 2 + 2p \cdot \frac{1}{1-p} = \frac{2}{1-p}$$

If the **payoff of collusion** is **greater than the payoff of cheating**, firms will choose collusion. In other terms, p has to be greater than $\frac{1}{2}$.

What affects the probability p:

- Stable **market structure**: no entry of other firms.
- **Obsolescence of the product**: the higher is the obsolescence of the product, the lower is the probability that this game will be played over time and a lower probability of collusion.

Concerning the **abuse of dominant position**, we refer to the **predatory pricing**:

In order to detect the **presence of an abuse**, the authorities have to **take some steps**: the first step is to **ask to themselves if there is a dominant position**. The answer is related to the **relevant market**: a **relevant product market** comprises all those products having similar characteristics, prices and intended uses so that they are seen by consumers as interchangeable or substitutable. A **relevant geographic market** comprises the area in which the firms concerned are involved in the supply of products or services and in which the conditions of competition are sufficiently homogeneous.

Antitrust has to **understand** if there is a **demand side substitutability** and **supply side substitutability**. In practice, they often refer to a **specific test**: the **SSNIP (Small but Significant Non-transitory increase in price) test**. It is an intuitive test but its implementation requires the use of **econometric tools and several data**. It is statistic, you are **not sure of results**. The steps of the test are:

- Starts with smallest possible market and ask if 5% price increase would be profitable for a hypothetical monopolist: if it is not profitable, this means that the monopolist does not have enough power to raise the price because of substitutability.
- Add the next closest substitute to the relevant market and repeat the test.
- Process continues until the point is reached where a hypothetical monopolist could profitably impose a 5% price increase.
- The market is defined.

Once defined the relevant market, antitrust has to **detect if there is an abuse of dominant position**. Decision is based on:

- o **Market share** (generally 40% in EU, 50% in US)
- o **Length in time of that market share** (persistence)
- o **Differences in market shares with second competitors** (high HI)

Examples of abuse:

- Predatory pricing: depriving smaller competitors of customers by selling at artificially low prices they cannot compete with.
- Obstructing competitors in the market (or in another related market) by forcing consumers to buy a product which is artificially related to a more popular.
- Offering special discounts to consumers who buy all their supplies from the dominant company.
- Making the sale of one product conditional on the sale of another product.

For being **a predatory price**, a price should follow three **criteria**:

- 1) Price set by the dominant firm is below the average variable cost. The dominant firm will gain **negative profits in the short run** because it should have the **reasonable expectation** that this behavior will lead to the **exit of competitors**. Therefore, the firm in dominant position will gain **positive profits over the long run**.
- 2) **Recoupment loss**: profit gained over the long run has to be **greater than the profit lost in the short run**.
- 3) **Predatory intent**: it is necessary to have **tangible proofs**. For example, in the Microsoft case: emails were a proof.

Is the **abuse necessary bad for social welfare**? There is a **threshold to identify the dominant position**, but you are not completely sure of the decision.

In case of predatory pricing, you have some rules to identify it but there is a **certain degree of uncertainty**. For example, AVC is not always clear and recoupment loss is uncertain by definition.

Indeed, **keeping out the opponents** is a very **essential element of competition**. Everything is surrounded by uncertainty. You have to **punish only behaviors that do not go in favor of consumers**. Actions against the consumer benefit are the **only certain things** in the process.

This logic has led to **different views and implementations of the antitrust**: **Chicago antitrust school** had some doubts about the antitrust capacity of increasing the social welfare. The idea is that antitrust legislation may end up harming rather than benefiting the social welfare. According to the neoclassical view, in markets, there is not a structure of performance to apply. There are some elements related to the structure of the industry and not to firms.

For the Chicago school, the **structure of the market depends on the performance of the firms**. Monopoly is often a transient position, so leaving market free to operate ensures dynamic efficiency. Antitrust may also be used in a **Tonya Harding-style competition**, which means asking the intervention of antitrust just because another firm is performing better.

Specific practices considered anticompetitive may not be anticompetitive. It is a risk to punish them. **Firms may refrain from pricing aggressively to avoid any risk of being accused** of predation. **Predatory pricing is very unlikely**:

- It is very expensive
- Acquisition can be a less expensive way to eliminate competitors.

- If the prey exit easily, this means that some others could also easily enter into the market. Barriers to entry are low.

Two different strategies of implementation of the antitrust:

- **Rule of reason:** since it is difficult to detect anticompetitive behavior, antitrust should **apply rules case by case**. They do not have a general validity. The problem is the higher **degree of subjectivity** in the market and **lack of transparency**: antitrust has too much discretionary power captured by firms. Moreover, there is a **little predictability to market players**: there is the probability to incur in some errors considering behaviors anticompetitive.
- **Rule of role:** need to have **some general rules** that can be applied to states. At the beginning of EU, it was necessary to let the states implement their own rules. Then, the EU integration process started because of the need of more valid rules. Globalization affected the antitrust; therefore, US and EU became similar.

Regarding the **merger regulation**, the starting point is that **dominant position is never punished**. To understand why, look at the following example.

- Suppose that demand is given by: $P = 120 - Q$ and all firms have constant marginal cost of $c = \$80$
- Let one firm have innovation that lowers cost to $c_M = \$20$
- This is a **Drastic** innovation. Why?
 - Marginal Revenue curve for monopolist is: $MR = 120 - 2Q$
 - If $c_M = \$20$, optimal monopoly output is: $Q_M = 50$ and $P_M = \$70$
 - Innovator can charge optimal monopoly price ($\$70$) and still undercut rivals whose unit cost is $\$80$

Should be the innovator blocked? No, because innovation is fundamental for dynamic efficiency, even if the firm will turn into a monopolist.

Dominant position has to be punished when firms artificially make it up. There are **horizontal** (between competitors in a particular market) and **vertical mergers**: much of the debates are around **allocative and productive efficiency**. This is because it is impossible to assess the outcome of the merger in terms of dynamic efficiency and innovation. The **focus is on horizontal mergers**, because vertical mergers do not necessarily lead to a lower number of competitors in the market. moreover, the vertical mergers may improve a company efficiency by better coordinating their different production stages.

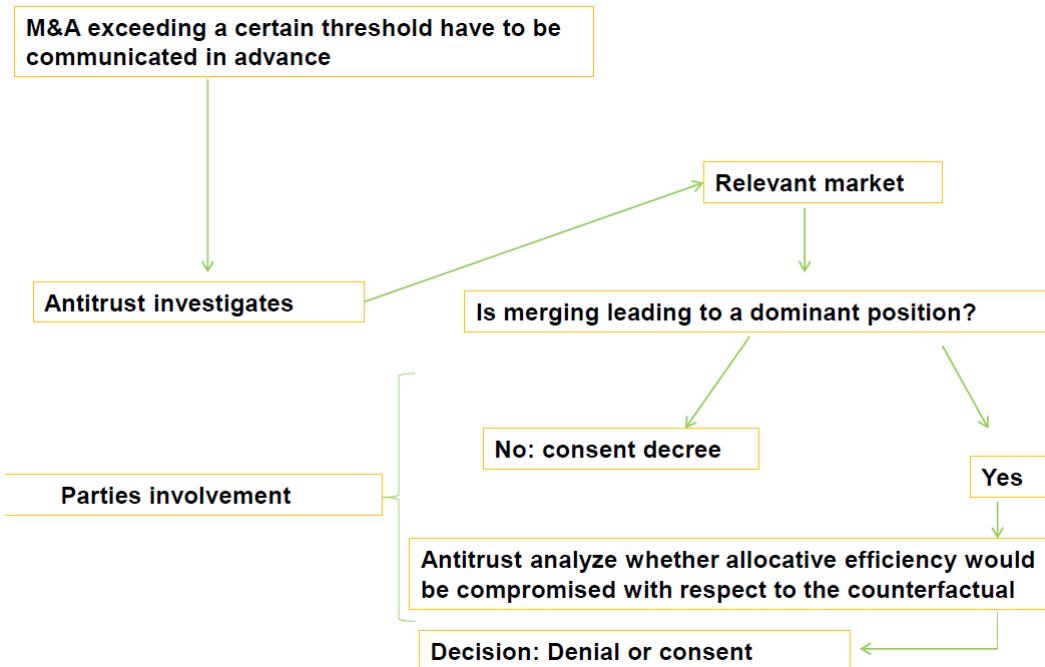
Horizontal mergers lead to a **potential allocative efficiency reduction** but also **a greater productive efficiency** (lower costs of producing goods). This productive efficiency could be translated to the market with the possibility of increasing the allocative efficiency.

Not all merging activities go under the radar of antitrust. **Only all activities involving big actors** are investigated by antitrust. **Smallness is excluded** because the gains of productive efficiency are higher when small entities merge due to economies of scale and scope having a positive effect also on allocative efficiency. Moreover, mergers between small companies do not constitute a dominant position. Instead, they can compete with larger companies through mergers in a win-win attitude.

Is largeness punished automatically? No, because **mergers between large firms may not lead to a dominant position**. It is not automatically true that there is a problem of market power: if there are

no barriers to entry, dominance does not translate into market power. We cannot exclude that there are **economies of scale and scope** due to **synergies** between the companies. Deciding to approve the merger, **antitrust has to consider that productive efficiency could be translated in more allocative efficiency**.

There is a procedure of merger regulation:



Regulation

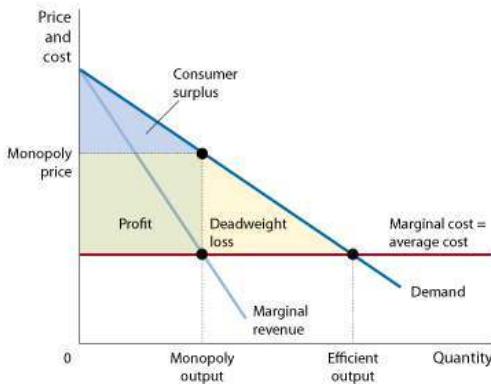
Tariff regulation refers to **natural monopoly**.

Natural monopoly refers to the concept of **subadditivity of cost**. The sufficient but non-necessary condition to subadditivity of cost is that the **derivative of average cost on quantity is <0**.

However, it depends not only on the supply side but also on demand side considerations, especially on the **dimensions of the market**. If the market demand is on the **left of the MES** there is a natural monopoly, but if it is **very far on the right** there is not a natural monopoly. In an intermediate point (on the right but not far from the MES), we could be in a natural monopoly because the subadditivity of cost is verified.

The **liberalization process in public service utilities** began because the demand was growing, shifting DD on the right far from the MES, while technological process has moved the MES closer to the origin. Therefore, in production and sales competition was opened.

If you allow a firm to produce all the quantity, the firm will behave like a monopolist from a pricing point of view. The **first best price solution in terms of social welfare is $MC=p$** . Suppose that people that want to buy only one unit of the product compose DD. In a monopolistic situation, if their willingness to pay is equal to the MC they are out of the market and there is deadweight loss.



Giving the production to a single firm is more efficient from a productive point of view, but there is a problem in terms of allocative efficiency. Therefore, the state can produce the product itself or **force the regulated firm to fix a price equal to the marginal cost**. The problem with this strategy is that it implies **negative profits**. Thus, the **second best solution is to fix price equal to the average cost**.

How to understand whether to push a first or second best solution?

There are two factors influencing the decision:

- The **demand elasticity** affects the deadweight loss, in the graph on the base of the triangle).
- The **fixed cost** affects the height of the deadweight loss triangle.

The **regulator's road map** starts with the question: **is there a natural monopoly?** In other terms, is the MES large relative to D? Is competition possible? If the answer is no, this means the regulator does not have to regulate. **If the answer is yes**, there is **room for regulation**. How to do it depends on the answer of another question: **is deadweight loss at second best relatively to first best solution tolerable?** **If the answer is no**, we are very much in a situation of **high elasticity of demand/high fixed cost** and the **second best solution is not a valid option**. The regulator should **regulate to achieve the first best solution or near first best solution**. There are three option to do it:

- **External subsidy** offered to the firm in order to produce at the first best
- **First grade price discrimination**
- **Non-linear tariffs** (second price discrimination)

If the answer is yes, there is another question: **there can be competition for the market?** Demsetz competition, contestable market or monopolistic competition? **If the answer is no**, the regulator should **regulate to achieve the second best or near the second best**. If the answer is yes, regulator has to implement competition in the market.

Explaining better, the options are:

- External subsidy: the state provide the amount to the firm
- **Two-part tariff** is a mean to gain more profits in price discrimination. Here, it is a mean to achieve the first best solution: a two-part tariff consists of a **fixed fee** and a **variable part depending the quantity consumed**.

$$T = f + pq$$

Pq is provided by the consumers.

$$f = \frac{F}{N}$$

With N numbers of consumers.

Consumers can amortize the fixed fee on the quantity. This means the consumer can pay a relatively higher price per unit but facing less expenditure in absolute terms, because he consumes less. On the contrary, he can sustain a relatively low expense per unit consuming more (high expense in absolute terms).

There are two constraints: incentive compatibility constraint and participation constraint.

- **Peak-load pricing:** it means to fix a higher price during the peak of demand (enable the firm to recoup the fixed cost) and a lower price during off-peak periods. The idea is that consumers may select himself choosing to consume in a certain period.
- **Third price discrimination:** it refers to the possibility for the regulator to find some exogenous characteristics that allow the firm to select consumers on the basis of their willingness to pay. For consumers with a higher willingness to pay than the second best solution, the firm has apply the second best solution. For consumers with a lower willingness to pay than the second best, you apply the first best solution. The fixed costs suffered by the firm are covered by the higher prices fixed to the group with higher willingness to pay. It is a theoretical possibility because it is very difficult to segment consumers according to the willingness to pay.

Concerning the different types of competition, there could be:

- **Monopolistic competition** (intermodal competition): it means that there could be a degree of competition produced by **substitute products** (car vs train). The regulator should try to make this competition stronger, for example showing the alternatives to consumers. It is a non-invasive way of regulating. The firm cannot fix a monopolistic price, it is forced to fix a price near to average cost.
- **Demsetz competition:** there is no assumption on the sunk costs in the market. If the number of firm delivering the good is quite high, the possibility to collude is very low and these firms have access to the same inputs (they have the same production function), a possibility is to make the **firms compete in order to have the right of serving the market**. They compete for the market and not within the market. The **winner will be the firm that fixes a price similar to the average cost of producing the good**. The winner has to serve the market with the lower price possible. It a way to achieve a **second best solution**. It is difficult to meet these assumptions in reality. The relationship between **franchisor and franchisee** has to be regulated by a contract: the winner has to assure in the contract that it will deliver a certain product at a certain price. If the winner is opportunistic, it can exploit the incompleteness of contracts for example delivering lower quality. Often the two parties give right to a third party to judge about possible problems between them in the fulfilment of the contract. This makes all the process more bureaucratic. In practice, distribution of gas in Italy refers to this logic.

From theory to practice, there is a problem of **asymmetric information**. The perfect information of regulator assumption is never met. The regulator does not know the average cost and the marginal

cost curve. It is a problem of **hidden information** and the regulated firm has no incentive to reduce the asymmetric information.

The total cost of the firm is $CT(q, \theta)$, a function of the quantity and of a parameter theta known only by the firm. If the firm is sustaining an average cost with a low theta, it has an incentive to show to the regulator that its average cost is the one at a higher theta. In this way, the firm can fix a higher price gaining an extra-profit.

The **regulator may know that the regulated firm has no incentive to reveal the true theta**. Therefore, he will impose **a price that is in the middle of the two** (price with higher and lower theta). This generates an **adverse selection** problem: by setting this price, the regulator is increasing price for the **firms behaving correctly**. The **hidden action** is generated by the fact that the regulator does not know if the firm is trying to be efficient from a productive point of view. He cannot control the **degree of effort** of the firm in order to product the good.

The difficulties in translating this mechanism in practical terms imply a regulation that puts **some incentives** for firms to be the more efficient possible.

The regulation solves the problem of asymmetric information. Its dynamics refer to three methods:

- **Cost-based regulation**: the regulator pays all expenses taken by the regulated firm in order to reach productive efficiency. An example is the **cost-plus regulation** (rate of return)
- **Incentive-based regulation**: regulator tries to incentive the regulated firm to provide the maximum effort to gain productive efficiency. In this way, there is no incentive to hide the cost structure. An example is the **price cap**.
- **Benchmarking-based regulation**: regulator tries to compare the cost of the regulated firms trying to make compete them in an indirect way. An example is the **yardstick competition**.

Especially the two last regulations solve the information asymmetry.

The ideas behind these regulations are:

1. **Cost-plus**: the regulator regulates the revenue of the regulated firms according to the following formula:

$$TR = \frac{cost + s \cdot RAB}{\text{Tariff} \cdot \text{quantity}}$$

Where TR is the total revenue, cost refers to the operating costs (labor, material, services) and annual capital depreciation (infrastructure, network, machineries), s is the rate of return on capital and RAB is the regulatory asset base. The RAB considers the operating costs (labor, material, amortization...).

The regulator defines tariff to allow volume of **revenues equal to the sum of incurred costs**, which include **a fair remuneration** for the invested capital.

Multiplying by s is made in order to **capture the opportunity cost** of investing money in that regulated business rather than in other alternative activities. The regulator has to refund the regulated firm also of this cost. The rate of return is generally the **WACC** (at least 6/7%).

This **regulation is applied over time**: the regulator has to determine the operating cost of the firm (RAB) and the s in order to apply the tariff. The procedure starts with the

determination of the firm costs that can be recovered, basing on firm accounting of the previous year. Basing on these costs and on the expectations about the future demand, the total revenue is determined. Price structure is proposed to the firm, which approves it, and prices are implemented.

What are the **problems with this type of regulation**? What are its **implications** in terms of productive, allocative and dynamic efficiency (we refer to the capacity to improve the quality of the service and to provide the incentive to make investments in order not to deteriorate it)?

The regulated firm has **no incentive to search for productive efficiency**: it knows that whatever will be the expenses to provide the service it will be refunded.

Concerning the **allocative efficiency**, tariff is the minimum price with respect to observed costs. Therefore, allocative efficiency is enhanced.

About **dynamic efficiency**, the regulated firm has incentive to overinvest in providing services, because every investment increases RAB boundary. There is more to be multiplied by the rate of return and therefore, there firm gains more total revenue (Averch and Johnson call this phenomenon "gold network"). Investments aims at ensuring remuneration and are not necessarily efficient.

The regulated firm achieves a financial-economic equilibrium: costs are covered and investors and creditors are remunerated. However, there is a reduction in the organizational and technological dynamics: no incentives to implement best practices and to adopt innovative systems.

An additional problem is that it has **high administrative costs**: the procedure has to be repeated very frequently in time involving lot of costs from the regulator point of view.

This type of regulation **does not really take care about the information asymmetry**. It is still applied, when the regulator does not suffer from information asymmetry because the technology is well known. Thus, it applies to **very mature sectors** and in sectors where **efficiency gains have already been realized**. Most regulatory agencies moved to the price cap.

2. **Price cap**: a price cap simply sets a maximum allowed **inter-temporal path for the price** of a specific product. The rules for the path are set in advance and only depend on factors that are beyond the control of the regulated firm. The regulated firm sets the price for the service with the only **requirement of not exceeding the ceiling** imposed by the regulator for a certain **regulatory period**.

The regulated firm maintains its natural objective of maximizing profits. The basic price cap rule is:

$$p_t^{MAX} = p_{t-1}^{MAX} \cdot (1 + RPI\% - X\%) \\ \text{con } t = 1 \dots n \text{ and } p_0 \text{ computed}$$

At the beginning of the regulatory period, the regulator makes an **accounting analysis and cost estimation**. The cap can be assumed equal to the most recent price, or to costs measured on a sample of similar firms, or defined repeating the **cost plus estimation**.

The cap will be valid for a certain regulatory period (normally 3 or 5 years).

The length of the regulatory period allows the firm to keep **gains from productive efficiency** improvements and, hence, creates **incentive to adopt efficient behaviors**.

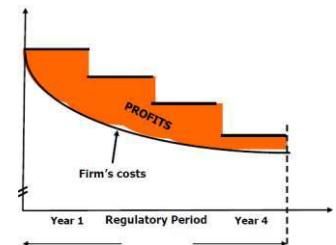
The regulator starts from the **current tariff** applied to the service. At time zero, it applies the **cost plus**.

The price is regulated by the formula and the firm cannot charge the consumers exceeding this price. In particular, RPI is the retail price index (rate of inflation). The X% is the expected productive efficiency gain that the regulator thinks the regulated firm can achieve year by year.

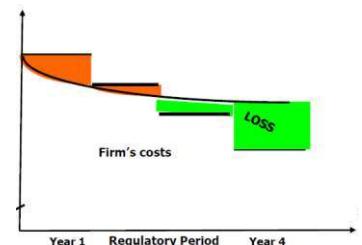
After the liberalization process, the **|X| is greater than RPI**.

Analyzing the time dynamics, at time zero the firm applies the price cap. As X is greater than RPI, it is a decreasing cap over time. Two possible scenarios:

- The regulated **firm meets the expectations** and makes more decreasing the costs much more than expected. It has an incentive in doing that because all the additional **productive efficiency gains** translate into **more profits**. One dollar saved is one dollar of profit.



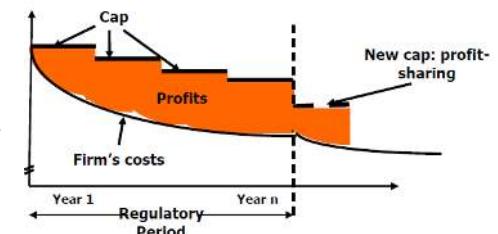
- The **firm does not meet the expectations** of the regulator incurring losses. The firm could incur in **losses** if the cost reduction occurs at a rate lower than X. Here there is the **risk of failure** of the regulated firm and this situation is undesirable for any party: citizens, regulated firm and regulator. These are essential services. Indeed, consumers prefer spending more than missing the service. The fact that X exceed the RPI has few exceptions.



What are the **problems**? What are the **implications** in terms of allocative, productive and dynamic efficiency? Starting from **productive efficiency**, we have a plus because firms have a great incentive to search for it. At the end of the regulatory period, part of benefits passes to consumers, and part is a premium to the efficient behavior of the firm.

About **allocative efficiency**, we have a potential drawback because during the regulatory period costs are not measured and prices are normally higher than the firm costs. However, compared to the cost plus, we have a **win-win gain** in terms of productive and allocative efficiency. The likely dynamic of tariff with a cost plus regulation is a straight line where observed cost covers price. The observed cost will remain stable: there is no incentive to reduce cost.

In case of profit, the regulator should go again towards the regulated firm, applying a cost plus regulation and everything will be replicated. The exception is that the starting price of the subsequent period will be not at the same level of the end of the previous period, but at a lower level. This is the **profit-sharing rule (50-50%)**: in the revision year, the new base is not equal to the costs the firm had over time



but the price-cost difference is distributed to consumers and firm. The firm has profit but there is a part of **consumer surplus**. A percentage of the margin is included in the new base of the cap, while the residual is left to the firm.

There is a weaker incentive, for which if the firm has to decide whether to push more effort, it has to choose at the beginning of the regulatory period because it provides more profit. Indeed, at the end of the regulatory period, the firm knowing the regulator's behavior slows down its path towards efficiency. This is ***the Ratchet effect***. In order to alleviate this problem, the regulator applies the profit sharing: it does not delete all the productive efficiency gain splitting it between consumers and regulated firms.

A great advantage of price cap relies on the ***low administrative cost*** of implementation, just because the cost plus takes place once and then it is valid for a long period.

The regulated firm obtains continuous productivity gains and adopts ***new organizational practices and technologies***.

The implications in terms of ***dynamic efficiency*** are a ***mixed incentive to invest***: great incentive to make cost reducing investments and low incentive to undertake quality-enhancing investments. In order to correct this dynamic efficiency problem, the formula can be augmented by ***the parameter Q (quality)***. The firm improves quality and this leads to a higher price. The firm deteriorating the quality of the product will be punished subtracting the factor Q. Moreover, the formula can be augmented by a ***term I (investments)***, allowing to recoup the investments.

$$p_t^{MAX} = p_{t-1}^{MAX} \cdot (1 + RPI\% - X\% +/ - Q\%) - I\%$$

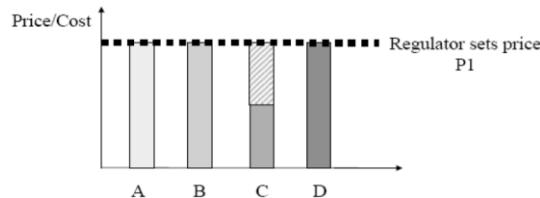
Moreover, the formula can consider the ***pass-through cost***: the regulator can modify the Price cap formula to pass-through those costs on which the firm has no direct control, for example the cost of input.

X% is generally 3% or 5% but the higher X% the lower the profit. The regulator does not want to put high X% because there is the risk of failure of the firm. The regulator should set ***X% high in the first phases*** and decreasing it until the end of the regulatory period because it is difficult to become efficient over time. When ***X = RPI***, we collapse in a ***cost plus situation*** because price is constant over time. The X% is the field of battle between the regulator and the regulated firm: the risk is that the attitude of pushing the social welfare decreases- this is ***regulatory capture***: the regulator define X% to favor the firm and not the consumer.

The financial risk is the ***variable remuneration for investors*** due to the costs covered only in case of productive efficiency gains.

3. ***Yardstick competition***: it is a model of regulation based on the principle of comparative competition between firms. The assumption is that if firms operate under similar condition, they should have in principle similar costs. A famous application was the breakup of the Bell system in '80s. The AT&T, company devoted to the long distance telecommunications. It created 7 companies, monopolies in a local area. It applied a price covering the average of the costs.

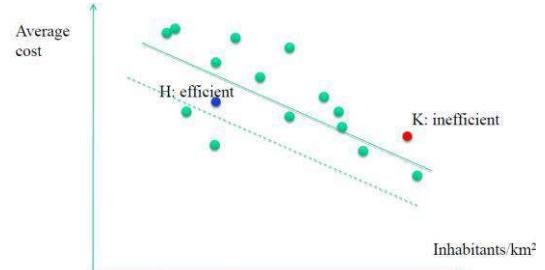
Given the cost of firms, the regulator applies a price, which covers only the **cost of the average firm**. There will be firms realizing some profits because able to reduce costs at a level equal or lower than the average cost set by the regulator, but others realizing some losses. In general, applying this mechanism to all firms, we get an overall cost reduction and consequently a price reduction. In theory, it is true for firms operating in **identical contexts and in absence of collusion**.



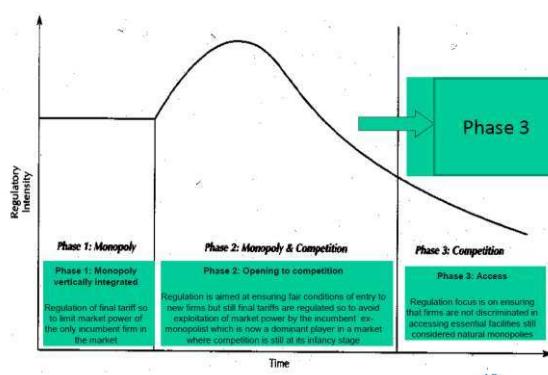
It has the **same implications of the price cap regulation**. Concerning the **productive efficiency**, the more they lower the cost, the higher the profit. Imposing separation between price and cost, we obtain a mechanism with high incentivizing power. The saving translates into profits. There is a **potential allocative inefficiency** because the price is not joint with the cost of the firm. When the exercise is repeated, the cost of the firm will be reduced and the regulator can charge a lower price. Same implications in terms of dynamic efficiency: similar method with premium in case the specific firm is able to improve the quality.

Two problems:

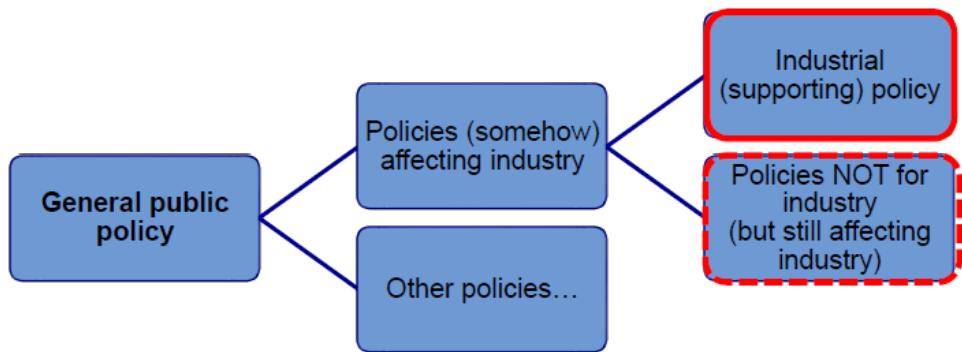
- Cost of the firms are affected by some **unobserved and uncontrolled factors**: you can control some factors affecting the cost but not everything. In practice, firms operate in different contexts, which generate different unit costs. The solution is to use econometric/statistical methods identifying efficient firms according to the population density. They are able to evaluate to what extent exogenous variables affect costs. Maybe a firm is more efficient than another is even if its average cost is higher.
- There could be a very different situation in terms of cost between the firms. For more inefficient firms, it requires a lot of effort to reduce cost with the **risk of failure**. It is better to have the service at a higher price rather than not having the service.



Regulation intensity path overtime



INDUSTRIAL POLICY: APPROACHES AND EXPERIENCES



Public policy is a guide to governmental action with regard to a class of issues. It has to identify and address the public problems implementing and monitoring laws and regulations. It deals with a dynamic, complex and interactive system and can be at a local, national or international level. In general, it has to promote macroeconomic stability with fiscal and monetary instruments.

Industrial policy is any policy that affects a subset of firms and industries differentially from the remaining group of firms and industries. Any tax, subsidy, trade and other policy measure that affects some specific firms and industries differently can be considered an industrial policy intervention.

The main characteristics are:

- Coordinated, focused, conscious effort by government
- Using an array of policy tools (legislations)
- Encourage and promote a specific industry or sector
- Alter industrial structure or production structure toward sectors
- Promote productivity based on growth, industrial growth, efficiency, competitiveness and so on
- Offer better prospects for economic growth than market equilibrium
- Typically focused on manufacturing
- Can apply to agricultural or service sectors too

Historically, three pillars are common for all successful industrial policies:

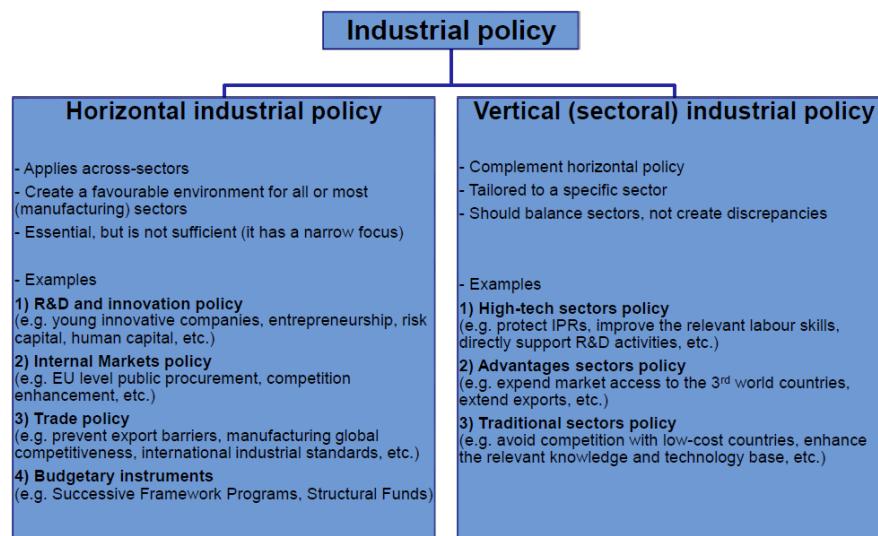
- Strong and credible state that coordinates the developmental priorities
- Strong and credible state that intervenes in the economy in order to structure the economic environment
- Strong and credible state that cooperates with the economic players in order to generate synergies

State is the main pillar.

The main issue of industrial policy is the governmental failure. Solving the market failure with governmental involvement is tricky because governments can fail too. Governmental failure is the most expensive one because it results in lower overall welfare.

Governments often lack information and capability to design effective industrial policies inviting rent-seeking and unproductive behaviors because it is difficult to measure and determine the selection criteria.

Implementation of industrial policies depends on political and institutional system of a country. Governments need to save the declining industries to save employment supporting them through low return of investments.



In the post war period, we have seen three broad eras:

- 1940s-1960s: industrialization is recognized as necessary for development. Market failures would prevent this from happening automatically, especially in developing countries. There is a need of industrial policy, particularly infant industry protection, state ownership and state coordination. → Traditional hands-on industrial policy
- 1970s-1990s: there were practical obstacles to industrial policy. There was the fear of government failure. Therefore, the focus was on trade liberalization (exports), privatization, attracting FDI and minimum government. → Market driven (laissez-faire) approach
- 2000s-present: government is fostering market system again building institutions and facilitating coordination. The “how” rather than the “what” of industrial policy is important. There is a focus on flexibility of industrial policy, and innovation and technological upgrading. → The rejuvenation of (vertical) industrial policy

Is industrial policy good?

Positive evidence:

- Western governments' interventionist responses to the 2008 Financial Crisis
- The success of many East Asian economies, most recently China, is associated with industrial policy

Negative evidence:

- Poorly-designed industrial policies risk having worse outcomes than the market failures they seek to address
- Shortages of transparency and technical capacity among policymakers in low-income countries make poorly-designed industrial policies likely

The question is how to design and implement it?

We have to understand what kind of industrial policies is effective:

- Challenge: Align industrial policy with resource base and level of development (Lin, 2010)
- Dilemma: Should industrial policies target:
 - Country's current comparative advantage industries? (e.g. agriculture, retail, telecoms, etc.)
 - Higher-productivity industries that are not competitive in the short-term? (e.g. high-tech, ICT, biotech, etc.)
- Empirical findings (McKinsey, 2011):
 - Invest in currently competitive, non-high-tech industries
 - Accumulate human and physical capital
 - Reinvest it over time in more productive industries

Does political economy affect industrial policy? For example, in developing countries market failures are more severe. The problems are:

- Institutional voids
- Lower government's capabilities and commitment
- Corruption, inconsistency, short-term planning, etc.
- Lower sophistication of industrial policies
- Governmental failure

Therefore, political economy affects industrial policy.

It is not clear if industrial policy serves the poor. In fact, growth is not inevitably good for the poor. Industrialisation and manufacturing may offer great economic opportunities for the poor in the medium to long term, but not necessarily in the short-term.

The ways to solve this problem are:

- Complement with special measures for the poor
- “Inclusive industrial policies”
- Focus on labour-intensive industries
- Support to agricultural productivity and rural household enterprises

An example of industrial policy is the green innovation: fostering green innovation through three industrial policy pillars:

1. Promoting frontier innovation

- Supply-side policies:
 - Public R&D investments
 - Support for early-stage development (entrepreneurship)
 - International cooperation
 - Protection and enforcement of intellectual property rights
- Demand-side policies:
 - Public procurement
 - Standards and related regulations
 - Adoption incentives and direct subsidies

2. Promoting catch-up innovation

- Mostly the same as promoting frontier innovation, plus more collaborative and sharing policies:
 - Open trade and investment policies
 - Open source and open business model encouragement
 - International mobility of researchers and innovators
 - Voluntary patent pools and collaborative mechanisms

3. Building innovation capabilities

- Education and skills development
- Public-private collaboration stimulation
- Linking to the global networks, value chain and knowledge
- Business environment improvement
- Infrastructure improvement
- Governance improvement

Successful industrial policy development: South Korea



Industrial policy in Europe

- Started in the **mid-20th century**
- European Coal and Steel Community – ECSC (1952) & Euroatom (1958)
 - 6 members (BE, FR, IT, LU, NL, W-DE)
 - Reduce coal production
 - Increase the steel production by improving the production system
 - Success: outcome and cooperation
 - Interventionist in nature: financial measures, targets, quarterly charged prices, penalties and fines for non-compliance, etc.
 - But more political in essence and very focused
- Antecedents of the European Union
- However, still no explicit legal basis before 1972, which hampered the scale-up

Giving a retrospective:

- **1950s - 1960s**
 - Production driven by the industrial policy (only some sectors)
 - Public and private efforts combined
 - Development of knowledge, new technologies, new foreign markets, etc.
- **1970s - 1990s**
 - Industrial policy fell out of fashion (inefficient due to the lack of resources)
 - Evolution of the economy was left to the markets – i.e. large MNEs with mainly short-term profit goals – competition policy was dominant
 - Industrial policy lost sensitivity
- **2000s – present**

- Serious acknowledgement that industrial policy should be rejuvenated
- Promotion of innovation, knowledge and research (FR7 Framework), entrepreneurship, diffusion of more environmentally friendly products, etc.
- However, this is still a sluggish process

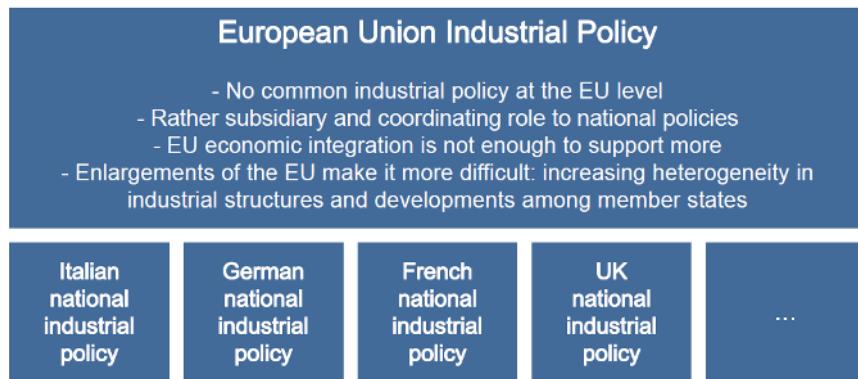
It has a legal basis:

- EU structure requires every policy to have a legal basis from the Treaty
- *Treaty* = International agreement, protocol, convention, pact, etc.
- Treaty of Paris (1951) and Treaty of Rome (1957) were the first: they were intended for the European Coal and Steel Community and the related narrow industrial policy
- Maastricht Treaty (1991), Nice Treaty (2001) and Lisbon Treaty (2007): first complete industrial-policy based treaties
- Problems
 - Very formal and bureaucratic system
 - Require all member states with very different political and economical backgrounds and relationships to agree
 - Instrumented through other policies and measures, increases complexity
 - Sluggish, stagnant and inefficient

The main issues with the legal basis are:

1. Industrial policy is mainly market (competition)-oriented:
 - Good for competitiveness of manufacturing industry
 - Prevents sectoral interventions
2. No detailed/direct policy instrument is declared
 - Aim pursued through other policies, measures and activities
3. All decision-making procedure on specific measure in support of action taken in the Member States should be discussed and specifically clarified
 - No uniform decision-making procedure
 - EU industrial policy reach is limited – it can only help some member states with specific measures
4. Synchronising national and EU level industrial policies
 - National industrial policies might be considered as unfair by the EC
 - Highly subjective and strongly influenced by politics

Having an overview on industrial policy in Europe:



Need of creating a common industrial policy at the EU level:

- Conditioned on creation of political legitimate institutions that can intervene and guide the EU industrial policies
- Goals should be to
 - Eliminate of the competition between states at the economical level
 - Create EU champions, not just national champions

Pro-arguments:

- ✓ EU companies prefer to cooperate than to compete
- ✓ EU champions would have a positive impacts on overall innovation
- ✓ EU champions would strengthen the EU industrial structure
- ✓ EU champions would strengthen the power of EU and attract FDI

Problem: EU currently make it difficulties for companies to merge across borders.

In the recent years:

- Guide principles for the Community industrial policy (1990):
 1. Openness - Guarantee proper functioning of market competition inside and outside the Community
 2. Horizontal – Abandoning sectoral policies
 3. Subsidiarity – Community tackles only tasks that cannot be done better at the national level
- Legal frame was set, but industrial policy was not flourishing
- Industrial policy was not the priority of the EU economic policies

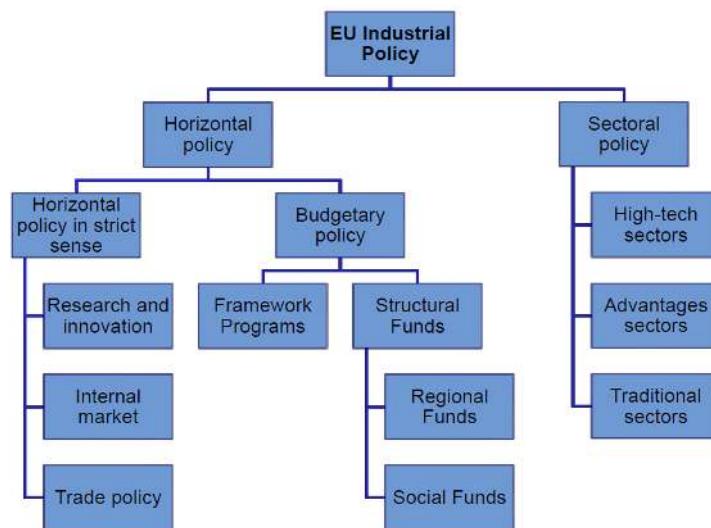
In 2000s - rapid progression

→ The return of the industrial policy

- More concrete activities, implementation and monitoring
- Industrial policy was made one of the most important 7 initiatives under “Europe 2020” strategy
- Monetary *Euro* Union - Internal market created opportunities

- Main goals:
 1. Smart growth - Knowledge and innovation, ICT, R&D
 2. Sustainable growth - Resource efficient, greener economy, environment
 3. Inclusive growth - High employment, social and territorial cohesion, bottom of pyramid and welfare business activities

The structure of industrial policy in Europe is:



Summary of the current EU level industrial policy:

- Recent re-focus on (more productive) manufacturing industries
- Creating favourable environment for EU manufacturing industry
- Complement to national policies (mainly a subsidiary and coordinating, yet increasing role)
- Predominantly horizontal in nature: openness, horizontal and subsidiarity are the main guiding principles
- Opposes strong sectoral intervention, which differentiate it from traditional industrial policy
- Sectoral policies are derived from the principles of the horizontal policy
- Aims of the EU industrial policy are pursued indirectly through other policies, which increases complexity
- Still relatively inefficient

Advantages:

- ✓ Jump-starts certain sectors of public interest (e.g. climate change & cleantech)
- ✓ Fixes financial systems (high-tech sector access to capital problem due to intangible assets)
- ✓ Fixes different market failures
- ✓ Fosters investments in growth-rich instead of in the non-tradable sector (e.g. real estate), which happens in case of Laissez-faire policies
- ✓ Improves competitiveness of some sectors on global market (e.g. against China that has growth-enhancing sectoral policies)

- ✓ Prevents countries from developing only upstream R&D and services, and outsourcing manufacturing to cheap unskilled-labour countries
- ✓ Creates “knowledge spillovers” that benefit more than one sector

Disadvantages:

- Distorts free competition
- “Picking winners”
- Exposes governments to abuse their position (e.g. regulatory capture)