aLESSANDRO pETRETTO

ECONOMICS OF INSTITUTIONS AND LAW

STUDIORUM UNIVERSITAS
FIRENZE UNIVERSITY PRESS

MANUALI

– 21 –

ALESSANDRO PETRETTO

Economics of Institutions and Law

Firenze University Press
2017

Economics of Institutions and Law / Alessandro Petretto.
– Firenze : Firenze University Press, 2017.
(Manuali Scienze Sociali ; 21)

http://digital.casalini.it/9788864535807

ISBN 978-88-6453-579-1 (print)
ISBN 978-88-6453-580-7 (online)

Cover image: © Mike Clegg | Dreamstime

Peer Review Process
All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Committees of the individual series. The works published in the FUP catalogue are evaluated and approved by the Editorial Board of the publishing house. For a more detailed description of the refereeing process we refer to the official documents published on the website and in the online catalogue of the FUP (www.fupress.com).

This book is printed on acid-free paper

CC 2017 Firenze University Press
Università degli Studi di Firenze
Firenze University Press
via Cittadella, 7, 50144 Firenze, Italy
www.fupress.com
Printed in Italy

Summary

A. Petretto, *Economics of Institutions and Law*, ISBN (print) 978-88-6453-579-1, ISBN (onine) 978-88-6453-580-7, CC BY 4.0, 2017 Firenze University Press

Introduction

This work proposes some lectures[1] for an intermediate course in *Public Economics*. The lectures aim to treat four topics, in many elements related each other: (i) the economics of institutions and political economy; (ii) the economics of law; (iii) the economics of public services enterprises ownership and (iv) the organization of public administration in providing public services.

We may say that these topics are not fully familiar with the standard normative approach to public economics, as followed in the traditional textbooks. Actually, there is not an organic and comprehensive theory about all of them, taken together. Thus, this work could be considered as an attempt in this direction.

As far as the first topic is concerned, *Economic of institutions and political economy*, we analyse, in the first chapter, how institutions arise in an industrialised country and how they are working and the effects they produce. The approach we follow is that one proposed in a recent literature, due to theoretical contributions by Daron Acemoglu and co-authors [Acemoglu (2009, ch. 22), Acemoglu (2013, ch.1), Acemoglu and Robinson (2015.)]. Main Acemoglu's argument is that, although economic institutions are the essential factor shaping economic outcomes, they are themselves endogenous and determined by political institutions and distribution of resources in society. Besides this approach, we treat also some standard issues on positive public economics and political economy referring, in particular, to the voting procedures and the relationships between political processes and economic efficiency [Hindriks and Myles (2013, ch.11) and Benassy-Quéré et (2014, ch. 2)]. Essentially, we move toward a positive theory of public expenditure with reference to some specific models of political economy.

As far as the second topic is concerned, *Economics of law*, by following a specific Acemoglu's reasoning, we may state as the structure of the civil law and particularly the set of enforcing property rights rules provide the main institutions of a country. Therefore, along this approach, in the second chapter, we analyse the basic propositions about the relationships between civil law and economics. The main issues considered in the lectures are those of legal liability for accidents, as a branch of tort law, of the nature and the justification of property rights, of the contracts law, in particular the formation, interpretation and the remedies for their breache. Further issues are those of litigation and suit and of law enforcement, in particular as far as the cases of corruption, tax evasion, anticompetitive collusion are concerned.

In this Law & Economics framework, the seminal contribution we are mainly following is the classical one by Louis Kaplow and Steven Shavell [Kaplow and Shavell (2002), Shavell (2005)].The approach employed for analysing legal issues is typical of the economic analysis, as it uses extensively stilized models and postulates that individuals and firms are forward looking and rational, acting with a view toward the possible consequences of their choices. Moreover, as in the standard welfare economics framework, the social desirability of an action, with legal consequences, is described by the condition of equality between its social marginal benefit and its social marginal cost.

The third topic treated in the lectures, *Economics of public services enterprises ownership*, refers to the economic basis of the choice between public and private ownership of enterprises providing public services. In this respect, we follow, in the third chapter, some relevant issues about Private-Public partnership mainly due to the contribution by Oliver Hart and colleagues [Hart et al. (1997), Hart (2003)]. In this framework, first, we treat the issue of the social convenience of delegating the running of a public facility, like a hospital or a prison, to a private manager instead to a public official.

[1] This work collects the lectures of the course in *Public economics* hold, in the academic years since 2011 to 2015, within the Master in Economic Sciences, at the University of Florence. I should like to thank Lisa Grazzini, for several helpful suggestions and insights in all themes of the lectures, and Lapo Filistrucchi, who devided with me the responsability of the course in *Public Economics*.

A. Petretto, *Economics of Institutions and Law*, ISBN (print) 978-88-6453-579-1, ISBN (onine) 978-88-6453-580-7, CC BY 4.0, 2017 Firenze University Press

Second, we analyse the choice of combining ("bundling") or separating ("unbundling") the two activities of building an infrastructure and managing the relative service. Third, we discuss the governmental choice of the means for financing a Private-Public Partnership contract, taking into account the distinction among works, if they are "hot" (generating positive cash flows) or "cold" (producing insufficient cash flows). Finally, we deal with the cost-benefit analysis of the sale of a public enterprise in order to ascertain the conditions for a socially desirable pricing of shares.
The fourth topic we would intend to analyse in the lectures, the *Organization of public administration in providing public services*, is a very general and wide one, and then we propose some examples. One of the most meaningful institutions of a country is its Health care system. Hence, in the fourth chapter, we propose a theoretycal analysis of the organization of health care, according to the various systems prevailing around the world [Zewifel et al. (2009), Glied and Smith (2011) and Petretto (2013a)]. Accordingly, we analyse the most important health care specificities for the government intervention in this context. Further, we consider the issue of the organization of health service production and provision, particularly the question of separation versus integration of production and of the choice of the systems of payment, i.e. tariffs, to providers.
As a further example of internal organizazion of public adiminstration, in the sixth chapter, we refer to the setup of intergovernmental relationships in a federal country where central government is used to transfer resources to local jurisdictions in order to alleviate the imbalance between expenditures needs and revenues [Hindriks and Myles (2013, ch. 20) and Petretto (2013b)]. The aim of these transfers is to ensure to every citizen the access to reasonably comparable levels of public services within a chosen locality, at a cost in line with what would be paid elsewhere. Therefore, equalization transfers promote horizontal equity by permitting fiscal treatment of identical persons in a federation and by enabling jurisdictions to provide minimum standards of essential packages of public services. In the chapter, we deal with the efficiency implications of such equalization system. The question we investigate is, given the level of a public service output constrained by a minimum standard fixed by the central government, how much of it a local government provides, at what cost and at what level of quality.
By trying to link together these topics, we may remind once again Acemoglu's statement, previously mentioned, by saying that civil law, public services enterprises, health care organization and intergovernmental transfers are institutions, endogenously determined within the economic systems, and, as a such, able to influence meaningfully their welfare performances.

Chapter 1

Economics of institutions and Political economy

1. General issues on economics of institutions

1.1 Institutions and states

There are different views about what we mean with the institutions and what they actually are. According to the *Efficient institutions view*, societies will choose institutions for maximizing social surplus In particular, if enforcing property rights has benefits which are greater than the costs, then a set of enforcing rules are going to represent the main institutions.
Quite interesting is the Political Coase Theorem (PCT), which we may think at the basis of the *Efficient institutions view.* As we shall see later on along these lectures, according to the standard version of Coase theorem, in the economy there should be forces and incentives toward efficient outcomes. An application of the theorem to institutions setting may run as follows.
Let

$$Y=F(X, P)$$

an objective function, where for simplicity we may interprete Y as the GDP, or the rate of increase of it. We suppose that Y depends on X a vector of characteristics of the economic environment of a country, and on P a vector of policies and institutions. Given a set Ξ of feasible policies and institutions, an optimal P^* can be defined as follows

$$P^*(X) \in \arg\max_P F(X, P), \forall P \in \Xi$$

Therefore, an optimal vector of institutions is an element of the set of the feasible vectors maximizing the overall objective function.
The Political Coase theorem states that there are strong forces towards some $P^*(X)$ belonging in the maximal output set. Therefore, following these forces, beneficial modifications, in terms of Pareto improvements, of institutions are always possible and should be pursued. However, there is a systematic evidence there are societies that, instead, choose some P such that $F(X,P)<F(X,P^0)$ for some feasible alternative $P^0 \in \Xi$, which could be also optimal. In other words, there are societies that persistently pursue and realize wrong policies, with significant output and welfare consequences.
One possible and obviuos reason of this somewhat paradoxical behaviour is that, by applying to political framework, the Coase theorem encounters serious commitment problems. If a ruler has political power concentrated in his hands, he cannot commit not to expropriate assets or revenues in the future. The enforcement of property rights, which would encourage investment by agents, requires that the credibility relinquishes political power to some extent. However, according to the bargaining at the basis of Coase theorem, the ruler, giving up some power, has to be compesated for what he could have received using instead this power. However, once lost the initial power, he has no guaratees that he will receive the promised payments in the future.
The conclusion that the Political Coase theorem generally is going to fail suggests that we have to look for other views of institutions.
According to the *Social conflict view,* institutions are chosen by groups controlling power, maximizing their own rents. Of course, the institutions that result by this choice may not coincide with those that maximize total surplus. By the *Ideology view,* as people have different beliefs about what is best for the society, the institutions may derive from the prevailing ideology in a specific time

A. Petretto, *Economics of Institutions and Law*, ISBN (print) 978-88-6453-579-1, ISBN (onine) 978-88-6453-580-7, CC BY 4.0, 2017 Firenze University Press

of history. By the *Incidental institutions view,* institutions emerge not as a choice of economic actors, but simply as incidental consequences of other actions.
It seems that, among these views, the Social conflict one can provide the most convicing explanation of the institutions development in modern societies. Therefore, by accepting it, let us consider the distinction among the variuos types of institutions as a result of social conflicts.
The first distinction is between *economic* and *political* institutions. Among the former ones there are the financial and fiscal systems, the civil law system, the rules for the competition extent, the mechanisms of market regulation, the education systems, i.e. all istitutions devoted to determining the "rules of the game", in particular the degree of property rights and constracts enforcement. Among the latter ones there are the form of government (democracy vs. dictatorship or autocracy), and the extent of constraints on politicians and political elites, i.e. all institutions devoted to determining the "rules of the political arena", by limiting the distribution of power.
A second typology of institutions is that one distinguishing between *formal* (statutory laws, courts, parliamentary system) and *informal* (not determinde by the law) institutions. The last distinction is between *predatory* (bad) and *developmental* (good) institutions, according to their effects on economic and social welfare.
Still ramaining in the context of the *Social conflict view,* let us look at the institutional origins, in particular those ones different to those deriving by the efficient view An explanation is based on the phenomenon of hold-up in investments and benefits. If the productive investments can be undertaken by a group of citizens or producers that are distinct from the "political elites", i.e. the current power holders, they do it only if they receive the benefits of the investments. Some economic institutions should arise for ensuring and securing property rights for non--elites citizen. However, often this is not the case because the elites in power cannot commit to respect the relative property rights. Another explanation of the origins of not efficient institutions is that they may derive from the desire of political elites to protect their power. Some political institutions may derive from the intent to blocking the activity of political losers for changing the status quo. A distinct but similar origin of economic institutions may derive from the intent to blocking or impeding distribution of income changes carried on by groups worse off, i.e. the economic losers.
The istitutions form the states, but there are several conceptions of the state. The first one is the state "as non-actor", according to which a normative intervent to correcting market failure is going to emerge, without any opportunistic behaviour by agents. The second one can refer to the state as "a nexus of cooperation", according to which the coercitive power of state encourages cooperation among agents. A third conception consider the state as an agent of a specific social group; this is a classical marxian theory of state. Another conception is that of the state as a "grabbing hand" institution, i.e. a state controlled by bureaucrats or politicians without an appropriate democratic control. Finally, we may quote the weberian theory of the state, according to which the latter it is shown as "an autonomous bureaucracy".
The essence of Acemoglu's theory is that the equilibrium determination of institutions is reached in a dynamic framework. The two "state variables" of this dynamical game are political institutions and the initial distribution of resources: from the knowledge of these two variables, at time t, it derives all the other variables in the system. While political institutions determine the constraints and incentives of the key actors, in the political sphere, then the distribution of *de jure* political power in society, the initial distribution of resources influences the distribution of *de facto* political power at time t. These two sources of political power, in turn, determines the choice of economic institutions and the future evolution of political institutions. Economic institutions influences investment in physical and human capital and technology and the organization of production. Therefore, from economic institutions we have the economic outcomes, including the aggregate growth rate of the economy and the distribution of resources at time t+1. Although economic institutions are the essential factor shaping economic outcomes, they are themselves endogenous and determined by political institutions and distribution of resources in society.

1.2 Toward a Political economic equilibrium of endogenous institutions

Let us present a scketch of a very simple model to describe this process [Acemoglu (2009, ch. 22)]. The economy is populated by a continuum of three classes of risk-neutral agents: the workers with measure normalised to 1, who supply their labour inelastically. The elite, denoted by *e*, that initially holds political power in the society. There is a total of θ^e elites and they do not take part to the productive activities. Finally, there are θ^m "middle class" agents, denoted by *m*, who are the entrepreneurs in the economy, with access to the production technology. The sets of elite and midlle class are denoted by S^e and S^m, respectively.
Each entrepreneur $i \epsilon S^m$ has access to the production technology for producing the final good *Y:*

$$Y_i(t) = F(K_i(t), L_i(t))$$

where $Y_i(t)$ is the final output produced by entrepreneur *i* and $K_i(t)$ and $L_i(t)$ are the total amount of capitale and labour that he uses in production. Since the total workforce in the economy is equal to 1 labor market clearing at time *t* requires that

$$\int_{S^m} L_i(t)di \leq 1.$$

As in the standard neoclassical model of growth, a fraction δ of capital depreciates.
Suppose that the society has access to four different policy instruments: a linear tax rate on output $\tau(t) \epsilon [0,1]$ and lump sum transfers to each of the three groups (workers, middle-class entreprenuers and the elite): $T^w \geq 0$; $T^m \geq 0$; $T^e \geq 0$. Since lump sum transfers are constrained to be non-negative, they cannot be used as non-distortionary lump-sum taxation. Instead, revenues can only be raised using a linear tax on output.
Let assume that taxes are set before the relevant investment decisions. In particular, the timing of events is such that at each *t,* we start with a predetermined tax rate on output $\tau(t)$, as well as the capital stocks of the entrepreneurs, $K_i(t)_{i \epsilon S^m}$. Entrepreneurs decide how much labor to hire $L_i(t)_{i \epsilon S^m}$ and in the process the labor market clears. Output is produced and a fraction $\tau(t)$ of the output is collected as tax revenue. The political process, i.e. the politically powerful social group, then decide the transfers, $T^w \geq 0$; $T^m \geq 0$; $T^e \geq 0$, subject to the government budget constraint

$$T^w(t) + \theta^m T^m(t) + \theta^e T^e(t) \leq \tau(t) \int_{S^m} F(K_i(t), L_i(t))di$$

where the L.H.S denotes total government expenditures in transfers, and the R.H.S is the predetermiend tax rate times output. Next, the political process announces the tax rate $\tau(t+1)$ that will apply at the next date. Entrepreneurs choose the capitall stock for the next date $K_i(t+1)_{i \epsilon S^m}$ after observing this tax rate, so that they know exactly what tax rate they will face at the next date.
It is evident that the *political economy equilibrium* deriving from this dynamic framework has two sources of inefficiency: the resources devoted to transfers and the distortionary taxation on output.
As far as the first type of distortion is concerned, it derives, as said, from the non-negative constraints of transfers' sequence. If it is the elite who takes the political economy under control, it could be even that he exploits the whole rent: $T^w=0$; $T^m=0$ and then

$$T^e(t) = \frac{1}{\theta^e} \tau(t) \int_{S^m} F(K_i(t), L_i(t))di.$$

Morevover, the tax rate $\tau(t+1)$ is chosen by solving recursively a maximization problem of the "elite value", given the tax announcement last period for to day, $\tau(t)$ and the distribution of capital stocks by the entrepreneurs, $K_i(t)_{i\in S}{}^m$.
As far as the second type of distortion, this emerges from the equilibrium condition of the sequence of capital stock and labour choices by an entrepreneur i according to the maximization of the utility function. This is represented by the discounted sum of his consumption levels, with discount factor $\beta\in(0,1)$, where the consumption, at each time, is the differerence between the value of the production the investment and the wage bill, and plus the transfer:

$$\sum_{s=t}^{\infty}\beta^{s-t}[(1-\tau(s))F(K_i(s),L_i(s))-(K_i(s+1)-(1-\delta)K_i(s))-w(s)L_i(s)+T^m(s)].$$

The first order condition (F.O.C.) w.r.t. the the choice of capital, at time t is as follows[2]:

$$\beta[(1-\tau(t+1))f'(k_i(t+1))+(1-\delta)]=1.$$

$k_i(t+1)$ denotes the capital-labour ratio chosen by entreprenur i for time $t+1$ given the tax rate $\tau(t+1)$, which has already been announced and committed to a the time of the investment decision. We can see as the tax reduces, through the term $1-\tau(t+1)$, the gross rate of return to capital $f'(.)$ and then the investment itself. As a benchmark, we may consider the competitive equilibrium without taxes (and transfers) such that the unique solution of the previous condition is identical to the first best capital-labor ration k^*. With positive taxes, the level of capital-labor ratio is less than k^* (given the strict concavity of $f(.)$).
Given the available tax instruments, the only way the elite can extract rent from entrepreneurs is by imposing distortionary taxes. Thus, the source of "inefficiencies" in this economy is the combination of revenue extraction motive by the political powerful combined with a limited menu of fiscal instruments.

2. Institutions, social choice and voting

Voting is the most commonly employed method of resolving a diversity of view or eliciting expression of preference. Therefore, it is the basis for the institutions working accordingly to the Social conflict view. Voting is used to determine the outcome of elections from local to supra-national level. Within organizations, voting determines who is elected to commitees, and it governs the decision-making of these committees. Voting is a universal tool that it is encountered in all institutional spheres. The natural question to ask of voting is whether it is a good method of making decisions and this question could be considered by two perspectives: that one of achieving of a clear-cut decision and that one about the efficiency of this decision.

2.1 Collective decision-making and imperfect rules: the Impossibility theorem and the Majority rule

Voting is a particular method of social choice, taking a given set of individual preferences and tryng to aggregate them into a social preference. The well-known *Arrow's impossibility theorem* says that there is no way to devise a collective decision-making process that satisfies a few common sense requirements and works in all circumstances [Hindriks and Myles (2013, ch.11)]. If there are only two options, majority voting works just fine, but, with more than two, the method can get into trouble.

[2] As usual, $f(k_t)$ is the production function in the intensive form, according to which the output per man is a function of capital-labour ratio.

One example of a situation where there is not a decision is the so-called *Condorcet Paradox*. Let us suppose there are three voters (1, 2 and 3) with the conflicting rankings over three options: *a, b* and *c* according to the following structure of preferences ($\mathcal{P}$ is the preference ordering). For voter 1: $a\mathcal{P}b\mathcal{P}c$; for voter 2: $c\mathcal{P}a\mathcal{P}b$ for voter 3: $b\mathcal{P}c\mathcal{P}a$.
Every voter has transitive preferences over the three options, i.e. if voter 1 prefers *a* to *b* to *c* then he prefers *a* to *c* and the same is true for the other voters.
Now suppose we use majority to select one of these options. We have that two out of three voters prefer *a* to *b* , while two out of three prefer *b* to *c* and two out of three prefer *c* to *a*. At the collective level, there is a cycle in preference and no decision is possible. Indeed, such collective preferences are intransitive, meaning that the preference for *a* over *b* and for *b* over *c* does not imply that *a* is preferred to *c*. This generation of social intransitivity from individual transitivity is the essence of *Condorcet Paradox*.
The general problem addressed by the Arrow's impossibility theorem is to seek a way of aggregating individual rankings over options into collective rankings. In doing so, difficulties such as the Condorcet Paradox has to be avoided, exluding cycles and finding a *Condorcet winner* (CW).
Let X a set of possible options and define the majority preference with $\mathcal{P}_m$, a CW is an option q such that $q\mathcal{P}_m q'$ for every other option $q' \in X$. The problem is that the existence of a CW requires very special configurations of individual preferences, violating one of the conditions stated by Arrow, according to which a "good" collective choice method should accomadate any possible individual ranking of options
The implication of Arrow's theorem is that any search of "perfect" method of collective decision-making is doomed to failure. Consequently, all collective decision-making must make the most of imperfect decision rules.
When the policy is one-dimensional, so that the options can be put in a transitive order, say from left to right in the politicul spectrum, sufficient, but not necessary, conditions for the existence of a CW are given by the *Median Voter theorem,* where the median voter is the individual whose preference is on the median of preferences distribution. However, there are two further basic assumptions on individual preference structure to satisfy alternatively.
The first one is that of *single-peaked preferences* (S-P), according to which there is always a single preferred option. Let $W(q,\phi_i)$ be the *political preferences function*, i.e. the preferences of individual i as defined on undimensional policy q, where ϕ_i captures differences in preferences across individuals. Let $q(\phi_i)$=argmax $W(q,\phi_i)$ the preferred policy. Preferences are single-peacked if $q'' \leq q' \leq q(\phi_i)$ or $q'' \geq q' \geq q(\phi_i)$ then $W(q'',\phi_i) \leq W(q',\phi_i)$.
As an example, let us immagine a commodity is sold by a number of shops distributed along a road. As the quality is the same, consumers will choose according to the distance from the shops. Each consumer most prefers a location of the shop close to home and ranks the others according to how close they are to the ideal. Hence, there is only a peak in the preference curve.
The second basic assumption is the *single-crossing* (S-C) *property,* according to which if, for $q>q'$ and $\phi'_i>\phi_i$ or for $q<q'$ and $\phi'_i<\phi_i$, $W(q,\phi_i) \geq W(q',\phi_i)$ implies $W(q,\phi'_i) \geq W(q',\phi'_i)$.
The consequence of tht S-C property is that for any two options q and q', with $q<q'$, if the median voter prefer q, then all the voters to the left also prefer q and if the median voters prefers q', then all voters to the right also prefer q'. Therefore, there is always a majority of voters who agree with the median voter and the option preferred by the median voter is a CW. Notice that S-P property, which is an assumption only on preferences, implies S-C property, which is an assumption on preferences and policies, but the converse is not true.
Now we may, more formally, state the following:

Median Voter Theorem (MVT)
With an odd number of voters and a one-dimensional policy space, if the preferences of the set of voters satisfy the single-peakdness property, or alternatively, the single-cross property, then the

preferred option of the median voter is a Condorcet Winner.

An actractive aspect of the MVT is that it does not depend on the intensity of preferences, and thus none has an incentive to misrepresent their preferences. In some sense, honesty is the best strategy for everyone. For a voter to the left of the median, misrepresenting preference more to the left does not change the median and thus the final outcome, where misrepresenting preferences more to the right either does nothing or moves the final outcome further away from his preferred outcome. With the same reasoning, a voter to the right of the median has no incentive to misrepresent his preferences either way. Last, the median gets his most-preferred outcome and thus cannot benefit from misrepresenting his preferences.
Having seen how the MVT, although within the framework of one-dimensionality, may lead to a clearly predicted outcome, we can now inquire if this outcome is efficient. In other words, the question is that if we may say that the MVT, as a collective decision-making according the *Social conflict institutions view*, can be also a decision-making of the *Efficient institutions view*. By answering this question, we are making a parallel between the Median voter theorem and the Political Coase theorem.

2.2 Private provision of public goods, median voter and inefficient outcomes

Consider a population of n consumers. Each consumer has the following utility function: $u^i=U^i(x_i,g)$ where g is the private good consumption and g the non-rival public public good. The latter is produced at a unitary cost c_g, thus the total cost is given by $c_g\, g$.
Now, we define the concept of *preferences policy function* $\varphi^i(g)$ which will be usefull also later one:

$$U^i(y_i - t^i(g), g) = \varphi^i(g).$$

This function gives the utility by consumer i deriving by the level g of the public good, and $g_D^i = \arg\max \varphi^i(g)$ is his preferred (demanded) level. $t^i(g) \le y_i, t^{i'}(g) \ge 0$ is the cost of the specific taxation, i.e. the individual contribution to the expenditure for providing the public good. The specification of function $t^i(g)$ depends on the government budget constraint. Therefore, in order to know something more about the function $\varphi^i(g)$ we have to consider how the total cost is collected and individually distributed; it could be a volontary contribution, a unitary price, a lump sum tax, an income tax and so on. First, consider the case of the public good provision without any role plaied by a public institution, i.e. a purely private provision.

2.2.1 No public institutions and private provision
Let us start by the case of a very low number of consumers, say n=2, $t'(g)=b^i$, i=1,2 is a voluntary contribution such that the individual and aggregate budget constraints are respectively:

$$x_i = y_i - b^i,$$
$$b^1 + b^2 = c_g g.$$

Each consumer i chooses his contribution b^i in a Nash-Cournot framework, i.e. considering as given the strategy of the other player, by maximizing the following preference function

$$U^i(x_i, g) \equiv U^i(y_i - b^i, \frac{b^1 + b^2}{c_g}), i = 1,2.$$

The F.O.C. is as follows:

$$-\frac{\partial U^i}{\partial x_i}+\frac{\partial U^i}{\partial g}\frac{1}{c_g}=0, \text{or } MRS_g^i \equiv \frac{\frac{\partial U^i}{\partial g}}{\frac{\partial U^i}{\partial x_i}}=c_g, i=1,2.$$

This condition provides two reaction functions $b^1(b^2), b^2(b^1)$ and thus two equilibrium value of the contribution: $\widehat{b^1}$, $\widehat{b^2}$ for a level of

If we add for *i*=1,2 we get at Nash equilibrium

$$MRS_g^1+MRS_g^2=2c_g.$$

The efficient outcome g^* satisfies, instead, the standard Samuleson rule, according to which the sum of the marginal rates of substitution between the public good and the private one is equal to the marginal cost:

$$MRS_g^1+MRS_g^2=c_g.$$

Hence, the private, Nash-Cournot, provision is inefficient, implying a too low level of the publc good: $\hat{g} < g^*$.

This is a clear case of inefficiency originated by a lack of institutions devoted to coordinating in someway the private players.

This inefficiency, in the small number of partecipants case, arises out of the bilateral monopoly/duopoly nature of the problem. It is well known that similar problems would arise also in bilateral bargaining over private goods. As such, the problem has more to do with game theoretic considerations than with the nature of the good. When the number of partecipants is large, however, this similarity disappears. It can be demostrated that by sufficiently increasing the number of actors in a market for private goods, the outcome would tend towards efficiency, via the mechanism of perfect competition. With a public good, however, the outcome will diverge further from the efficient one. In markets for private goods, when the number of partecipants is large, individual actors become price-takers while, in public good markets, people become quantity-takers, thus no agent feels that he can influence the amount of public good which is made avilable. The rational agent, therefore, will attempt to "free ride" on the public goods supplies of the others and the consequence will be gross underprovision of public goods.

Let us consider the case *n*>2, *n* large enough. Now the condition of private contribution becomes:

$$MRS_g^i \le c_g, i=1,...n.$$

The inequality is written in recognition of the fact that some agents may choose to devote no resources for contributing to the public good provision. This would be the case if the cost of the public good is large relative to any individual's direct benefit.

If we assume that the condition holds as an equality for $m \le n$ agents, we have in equilibrium,

$$\sum_{i=1}^{n} MRS_g^i \ge mc_g,$$

with the strict inequality holding if at least one of the remaining *n-m* individuals place positive value

on the last unit of the public good.
The opportunity for bargaining which is, even in principle, present in the small numbers case virtually disappears when the number of agents is large. If the bargaining costs were proprortional to the number of agreements, they would increase with the square of the size of the bargaining group. Hence, it is unlikely that we would obtain a level of public good, which exceeds that one indicated by the individual condition of contribution choice: $MRS_g^i \le c_g$.

2.2.2 Majority voting

For considering the new scenario which allows, for n>2 citizens, the quantity of public good to be provided by a majority vote, let us suppose the expenditure of the public good is shared equally among the consumers. In other words, we have an individual cost (tax) to each consumer given by $t^i = t = \frac{c_g g}{n}$.

With gross income y_i, a consumer can purchase private goods to the value of $x_i=y_i-t$ after paying the head tax t for the public good. This provides an effective price of $p_g \equiv \frac{c_g}{n}$ for each unit of the public good and the following level of utility

$$U^i(y_i - p_g g, g) = \varphi^i(g).$$

This preferences policy function $\varphi^i(g)$ is initially increasing in g, when $\frac{\partial U^i}{\partial g} > \frac{\partial U^i}{\partial x_i} p_g$, i.e. when the marginal utility of the public good is greater than the marginal cost of it in terms of reduction of private good consumption. At $\frac{\partial U^i}{\partial g} = \frac{\partial U^i}{\partial x_i} p_g, \varphi^i(g)$ reaches the maximum. After that point the function $\varphi^i(g)$ becomes decreasing in g. Given this structure of preferences, consumeres i, i=1,..n can be numered so that their preferred levels of public good satisfy the ordering $g_D^1 < g_D^2 < .... < g_D^n$, thus satisfying both single-peakedness and single-crossing properties. As the policy space is one-dimensional (the level of g), if n is an odd number, the median voter theorem ensures that the consumer with the median preference for the public good will be decisive in the majority vote. The median preference belongs to the consumer at position (n+1)/2 in the ranking. We label the median consumer as m and denote their chosen quantity of the public good by g_D^m. As said, this outcome cannot be manipulated by nobody by misrepresenting the preferences. Indeed, g_D^m solves the optimization problem $\max_g U^m(y_m - p_g g, g) = \varphi^m(g)$, whose F.O.C. is

$$MRS_g^m \equiv \frac{\frac{\partial U^m}{\partial g}}{\frac{\partial U^m}{\partial x_m}} = p_g \equiv \frac{c_g}{n}.$$

In words, when the marginal rate of substitution between the public good and the private one by the median voter is equal to the effective price.
In contrast, the efficient outcome satisfies the standard Samuleson rule

$$\sum_{i=1}^{n} MRS_g^i = c_g$$

or

$$MRS_g^a \equiv \frac{\sum_{i=1}^{n} MRS_g^i}{n} = p_g.$$

Therefore, majority voting leads to efficient provision of the public good only if

$$MRS_g^m = MRS_g^a,$$

i.e. the median voter marginal rate of substitution is equal the that one of the mean of the population of voters. There is not reason to expect that it will, so it must be concluded that majority voting will not generally achieve an efficient outcome[3].
Note that if

$$MRS_g^m < (>) MRS_g^a$$

we cannot say if too little (much) public good is provided. We will explore, in the following section, this issue by means of a more articulated political economy equilibrium of public spending. Moreover, we will find a relationship between median voter preferred level of the public good with income distribution in the society, as underlined in section 1.1, arguing a dynamic equilibrium determination of institutions.
In conclusion, either private provision nor majoritity voting in general leads to outcomes which are coherent with an efficient institutions framework. One could ask if this failure can derive from the excessive semplicity of the decision making of both the solutions of public good provision problem. In the following section, we are going to examine more complex and realistic decision making processes but verifying that there is a prevailing tendency towards inefficient outcomes.

3. Theories of public sector and social conflict

In this section, we present some models explaining how political processes can conduct to inefficient outcomes in the level of public spending and taxation. In other words, they are some simple examples how non-benevolent institutions, rised in the equilibrium of a dynamic game reflecting the social conflict view, may turn out in the failure of the Political Coase theorem.
Within the Social conflict view of institutions, we present four types of conflict in the society:

(i) between rich and poor people about the redistribution policies carried on by the government;
(ii) among different groups or localities on the level of groups-specific public goods;
(iii) between the government and uninformed citizen/taxpayers;
(iv) between the cabinet of the government and bureaucrats, heads of governmental departments.

3.1 A Political economy model of public expenditure and income distribution

This model tries to capture the conflict in public preferences between those who wish to have higher expenditure and those who wish to limit the burden of taxes[4]. It will turn out that the risolution of this conflict shows as the size and composition of actual public spending reflects the preferences of majority of citizens as expressed through a political process. The equilibrium level of public expenditure can be related to the income distribution and more precisely the growth of government is related to the rise of income inequality.
Consider an economy of n consumers whose income y_i, i=1,..n, fall into the range [0, y_i^{max}]. The government provides a public good g that is financed by a proportional income tax, at a rate τ, which

[3] Actually, the number of voters is meaningful, as if n is very large, the distribution the marginal evaluations tends to the Normal one, for which mean and median concide.
[4] The model is extension of a model exposed in Hindriks and Myles (2013, ch. 5).

has the feature of a lump sum tax. Thus the private good consumption is $x_i = (1-\tau)y_i$. The utility of consumer i with income y_i is $u^i(x_i,g)$, which we suppose as follows

$$u^i(x_i, g) = (1-\tau)y_i + b_i(g).$$

The function $b_i(g)$ represents the benefit obtained from the public good and it is assumed to be increasing, $b_i'(g)>0$, and concave, $b_i''(g)<0$. Let μ be the mean income level in the population of consumers, $\mu \equiv \frac{\sum_j y_j}{n}$, so the government budget constraint is (we suppose that the unitary cost of the public good is 1 €):

$$g = \tau \sum_j y_j = \tau n\mu.$$

Suppose now that $b_i(g)=b(g)$ all i, so the benefit function is the same for population. Using this budget contraint, a consumer $\{i, y_i\}$ will enjoy utility from provision of a quantity g as follows

$$\varphi^i(g) = (1-\frac{g}{n\mu})y_i + b(g).$$

This is the *preferences policy function* w.r.t g and the preferred level of public good provision for the consumer i is given by the first order condition of its maximization:

$$\frac{\partial \varphi^i(g)}{\partial g} \equiv -\frac{y_i}{n\mu} + b'(g) = 0$$

$$=> b'(g) = \frac{y_i}{n\mu}$$

This condition relates the marginal benefit of an additional unit of the pubilc good $b'(g)$ to the marginal cost attributed to consumer i, $y_i/n\mu$. The quantity of the public good demanded by the consumer, g_i^D depends on their income relative to the mean.

As the marginal benefit of g is decreasing, $b''(g)<0$, the preferred public good level is decreasing as income rises. Indeed, with a proportional income tax, the rich pay a higher share of the cost of public good than poor do. Hence, public good provision will disproportionally benefit the poor. To see this implicetely differentiate the previous F.O.C.:

$$\frac{dg_i^D}{dy_i} = -\frac{-(1/n\mu)}{b''(g)} < 0.$$

Note as, instead, the desired level of pubblic good is increasing with the mean income:

$$\frac{dg_i^D}{d\mu} = -\frac{(y_i/n\mu)}{b''(g)} > 0.$$

Let resolve the disagreement over the desired level of public good by majority voting. In the context of this model, all consumers would prefer the level of public good to be as close as possible to their preferred level. Given any pairs of alternatives, consumers will vote for that which is closest to their preferred alternative. The alternative that is closest for the largest number of consumers will receive maximal support. There is in fact only one option that will satisfy this requirement: the option

preferred by the consumer with the median income. One-half of the electorate, above the median income, i.e. the rich, would like less public good and the other half, below the median, i.e the poor, would like more public good. Any alternative that is better for one group would be opposed by the other group with opposite preferences.
The political equilibrium g^* determined by the median voter, is then the solution to:

$$b'(g^*) = \frac{y_m}{n\mu}$$

where y_m/μ is the income of the median voter relative to the mean.
It is empirically sustained that the distribution of income is skewed to the left with median income lower than the mean one, therefore $b'(g^*)<1/n$ (while efficiency requires $b'(g^o)=1/n$). Since the marginal benefit decreases as public good provision increases, the political equilibrium level of public good increases with income inequality as measured by the ratio of median to mean income.
Accordingly, more inequality (a lower ratio) would lead the decisive median voter to require more public spending. Government activities are perceived as redistribuive tools, throughout social security, poverty alleviation programs and public employment too. Because its nature, and the interaction with the tax system, the demand for redistribution will increase as income inequality increases.

3.2 Parliamentary public spending decision and special interest politics

This model analyses the consequences for the society of the conflict among different groups of citizen-voters on the level of public goods provision[5].
Consider a society with K distinct but symmetric groups (we may for instance think at localities), each having a continuum of members with a mass of unity. Let the utility of the individual j, belonging to a specific group of the society i, u^{ji}, i=1,..K, depends on his consumption x^{ji} and on the consumption of a public good specific to a group of the society g^i:

$$u^{ji} = x^{ji} + \psi^j b(g^i)$$

where ψ^j is an idiosyncratic weight that individual j assigns to the benefit of the public good g^i consumed only by group i. We assume that in each i, ψ^j is distributed on an interval according to the symmetric distribution $F(\psi^j)$, with Exp(ψ^j)=1. All individuals receive the same income y, pay taxes T_i to finance the public good and consume the remainder:

$$x^{ji} = x_i = y - T_i = y - g^i.$$

Therefore, both taxes and public good provision are group specific and decided within groups. It could be the case of local taxes that finance local infrastructures. The utilitarian optimum in this setting implies as F.O.C. that the average marginal benefit in the group equals the marginal social cost of unity, namely

$$1 = b_g(g^i) \equiv \frac{db(g^i)}{dg^i}$$

which is the Samuelsonian rule within the group i.
By inverting the function of marginal benefit $b_g(g^i)$ we get the same level of public good in all groups:

[5] This is a syntethic re-elaboration of the orginal model by Persson (1998). See also Persson and Tabellini (2000, ch .1) and Benassy-Quéré et al. (2014, ch.2).

$T_i = g^* = b_g^{-1}(1), i = 1,...K.$

If, for instance, the benefit function is such that:

$b(g^i) = A[g^i]^{\xi}$, where $0 < \xi < 1$, it turns out as follows

$$b_g(g^i) = A\xi[g^i]^{\xi-1} = 1.$$

Consequently, by inverting the function, we obtain:

$$g^i = g^* = [A\xi]^{\frac{1}{1-\xi}}.$$

Since g^* is an outcome beating any alternative in a pair-wise vote is also a *Condorcet winner* so this allocation could be implemented by direct voting in each group. In this specific setting (symmetric distribution of preferences, with mean equal to median) we would have a mojority vote outcome also Pareto efficient, as in the *Efficient institutions view*[6].

To illustrate the political process of special-interest politics, we will however switch to the case when the public good is financed by an economic-wide pool of tax revenue, with equal contribution from each group. With a common lum-sum tax T the government budget cosntraint is $TK=\Sigma_i g^i$. In this case, if there are no spillovers, there is a negative externality stemming from the needs to finance other groups'public gkoods without benefiting from them. For group k this means:

$$x^{jk} = x^k = y - T = y - \frac{\sum_i g^i}{K}$$

If group k is able to generate a unilateral expansion of g^k, its marginal cost would thus be $1/K$, far below the marginal social cost, if K is large. A possible outcome is that one coming from pork-barrel-politics, according to which a generalized overspending approaching the "universalistic" solution will emerge:

$$\tilde{g} = b_g^{-1}(1/K) = [KA\xi]^{\frac{1}{1-\xi}}$$

However, let us immagine this decision rests with a parliament where each locality has a representative and the preference are exogenous. Each representative tries to maximize the utility for his constituency l, which is

$$R^l = y - \frac{\sum_i g^i}{K} + \psi^l b(g^l).$$

As a benchmark (extrem) case, let us consider the situation where a group a is able to act as a dictator, extracting all the revenues from taxation to financing only its public good. Hence,

$$g^d = TK, g^l = 0,\ l \neq d.$$

In this case, it turns out as follows:

$$R^l = y - \frac{g^d}{K}, l \neq d$$

$$R^d = y - \frac{g^d}{K} + \psi^d b(g^d)$$

[6] Indeed, it is known from the theory of fiscal federalism that, without spillover effects and any tax distortions, the "correspondence principle" should be applied. According to this, the jurisdiction determining the order of provision of each public good should include precisely the set of individuals that consume it.

Suppose now, as a milder case of specific-interest power, a "budget process" in a legislative session modelled as the following sequence of events:
(1) One of the representatives is to be an agenda setter: $l=a$;
(2) Representative a makes a policy proposal given by the vector $g=(g^i)$;
(3) The legislature votes on the proposal. If it gets simple majority, i.e. it collects at least $K/2$ votes from the other legislators (clearly a votes always for his proposal), g gets implemented. If not, a default outcome, with $g^S=(g^i)=T=0$, gets implemented.
The chosen agenda setter g knows that every legislator not getting at least as high payoff from his proposal g as from the default policy g^S will vote no. This requires the following "incentive compatible constraint":

$$R^l(g^l) - R^l(g^S) = \psi^l b(g^l) - \frac{\sum_i g^i}{K} \geq 0.$$

Thus the agenda setter maximizes his preference R^l s.t. the previuos incentive constraint, holding for a mojority coalition M, including at least $K/2$ other legislators. We might show (Persson 1998) as the F.O.C of this constrained maximum is

$$\psi^a b_g(g^a) = \frac{1}{K - \sum_{i \in M} \frac{1}{\psi^i b_g(g^i)}}$$

$$\psi^i b_g(g^i) = \left(g^a + \sum_{i \in M} g^i \right) / K = T, i \in M$$

$$|M| = K/2; g^i = 0, i \notin M$$

The intution of the equilibrium is as follows. To get support from other legislators, a has to spend tax revenue in their districts, resources that could otherwise have been spent in his own district. This opportunity cost means that a will not spend more than necessary. First, he will chose a minimum winning coalition, composed of $K/2$ other legislators; and the districts whose legislators are outside of the winning coalition get no public good at all, even though they bear the cost of taxes. Second, for the members of the winning coalition, a spends only as much as is necessary to satisy the incentive constraint with equality, so that the members are barely as well off as with the default policy. Third, he will pick the legislators with the $K/2$ highest values of ψ^i to be included in the majority, as these are the cheapest to buy off.
This allocation is clearly inefficient, as the agenda setter's district gets more public good than in the social optimum and the districts outside of the majority certainly get less.

3.3 Imperfect information by voters and government abuse of power

The model describes the case of abuse of power by the government, which arises from the lack of information available to voters[7]. Also in this case there are greedy bureaucrats which exploite their monopolistic power to extract rent. The central issue is then how to set incentives that encourage the government to save costs, given the available information. The citizen-voters could prefer to pay the bureaucrats something for limiting their temptation of opportunistic beaviors.
Consider a situation in which the cost to the government of supplying a public good can vary, being the unit cost either low, at c_l, or high, at c_h and let $\Delta \equiv (c_h - c_l)$. The benefit to the public, a representative

[7] This is a re-elaboration of the "Agency model" by Hindriks and Myles (2013, ch. 5).

consumer, from a level g of the public good is, as usual, represented by the concave function $b(g)$. The net benefit, i.e. the net surplus, given by $x=u-y$, is $b(g)-T$, where T is the tax paid to the government for the public good provision. The benefit to the government for providing the public good is the difference between the tax and the cost: $T_i-c_ig_i$, $i=h,l$. In this context, we may model the relationship between the governement and representative taxpayer as a Principal/Agent game with hidden information and adverse selection.

When the public is informed about the level of cost of the government, the quantity of public good will be chosen to maximize the net benefit subject to the government budget constraint $T_i=c_ig_i$. For cost c_i, the public net benefit, with the government budget constraint satisfied, is $b(g_i)-c_ig_i$. The taxpayer will demand a level of public good such that the marginal benefit is equal to the marginal cost, so

$$b'(g_i^*) = c_i,$$

and will pay to the government $T_i = c_i g_i^*$.

Now assume that the public cannot observe whether the government has cost c_l or c_h. The government can benefit by misrepresenting the cost to the public, with the intent to exagerate the cost to adding expenditures that benefit the government, i.e. the politicians and the bureaucrats supporting it, but not the public. When the cost is high, the government cannot exagerate. When the cost is low, the government is better off pretending the cost is high to get the tax T_h for the amount g_h of the public good instead of getting T_l for producing g_l. Indeed, it is

$$T_h - c_l g_h = c_h g_h - c_l g_h \geq T_l - c_l g_l = 0$$

Therefore, the benefit of this misrepresenting is $T_h - c_l g_h = g_h \Delta$.

To eliminate this temptation, taxpayers must pay an *informational rent*, i.e. an extra amount $R>0$ to the government, in excess of its cost when the government pretends to have the low cost. Since the truly high-cost government cannot further inflate its cost, the public pay $T_h = c_h g_h$ when the government reports a high cost. If the reported cost is low, the taxpayers demand the amount g_l of public good defined by the condition $b'(g_l)=c_l$ and pay the government $T_l=c_lg_l+R$, where R is exactly $g_h \Delta$, the extra revenue the government could have made if it had pretended to have high cost. To give a government with a low cost just enough revenue to offset its temptation to pretend to have highier cost, it is necessary that $R = (c_h - c_l) g_h = g_h \Delta$.

This is the rent required to induce truthful revelation of the cost and have the provision of the public good equal to that when the public is fully informed.

It is possible for the taxpayers to reduce this excess payment by demanding that the high-cost government supply less than it would with full information. Assume that cost is low with propability p_l and high with probability $p_h=1-p_l$. The optimal level of g_h is obtained by maximizing their expected benefit subject to the government telling the true, i.e.

$$\begin{aligned} \max_{g_h} \quad & p_l[b(g_l) - T_l] + (1 - p_l)[b(g_h) - T_h] \\ & T_l = c_l g_l + R \\ & T_h = c_h g_h \\ & R = g_h \Delta, \end{aligned}$$

then by maximizing w.r.t. g_h the following function

$$p_l[b(g_l) - (c_l g_l + \Delta g_h)] + (1 - p_l)[b(g_h) - c_h g_h].$$

The F.O.C. is:

$$b'(\hat{g}_h) = c_h + \frac{p_l \Delta}{1 - p_l} > c_h$$

This quantiy of $\hat{g}_h$ is lower that that with full information, g_h^*, because the marginal cost is greater. The distortion of the quantity demanded from the high-cost government increases the more likely is the government to have low cost, i.e. the greater is $\frac{p_l}{1-p_l}$. The distortion results from a simple cos-benefit argument. It trades-off the benefit of reducing the rent, which is proportional to the difference $c_h - c_l$, and the probability p_l that the government is of the low-cost type against the cost of imposing the distortion of the quantity on the high-cost government that occurs with probability 1- p_l.
Therefore, if the government is truly low cost, it need not be given the high tax. However, to eliminate the temptation for cost inflation, taxpayers have to provide the government just enough of the rent as a reward for reporting truthfully, when its cost of public services is low.

3.4 A model of inefficient budget determination and Spending review

This model describes the inefficiency due to the conflict between the government as a whole and greedy bureaucrats running the decisions of the former throughout the activities by governmental and ministerial departments[8]. As it will be clear, also this is a case of inefficiency due to some lack of information between economic agents within the institutions.
Each department is headed by a bureaucrat, with some politician as sponsor, who obtains at time *t* a benefit from the public good supplied $g(t)$, but with a cost from the effort to efficiently organize the department, $e(t)$. Formally, the head of the department chooses his strategies to maximize the net benefit as follows

$$\max_{\{g(t),e(t)\}} \quad U[g(t),e(t)] - C[g(t),e(t)].$$

In the net benefit function $U(.)$ is a utility function with $\frac{\partial U}{\partial g} > 0, \frac{\partial U}{\partial e} < 0$. Indeed, for the head of the department increasing output means more power and prestige, while the managerial effort has disutility effect. $C(.)$ is the cost function with a positive marginal cost $\frac{\partial C}{\partial g} > 0$, while the cost decreaes with managerial effort, $\frac{\partial C}{\partial e} < 0$.
The first order condition requires that marginal benefits are equal to marginal costs:

$$\frac{\partial U}{\partial g} = \frac{\partial C}{\partial g}, \text{ and } / \frac{\partial U}{\partial e} / = / \frac{\partial C}{\partial e} /.$$

Given the standard properties of cost function, the F.C.O. implies an interior solution for *e*, then at the optimum is $e^*(t) < e_{\max}$. Consequently, the chosen effort is not the maximum achievable: $C(g^*(t), e^*(t)) > C(g^*(t), e_{\max})$.

Government does not observe *e* , thus even if he knows the cost function, $C(.)$, and observe ex-post the level of it, it does not know the efficient level of the department cost, along which to formalize a budget for time *t*+1. Let us suppose that at time *t* the head of the department has been able to obtain

[8] The model extends one proposed by Hindriks and Myles (2013, ch. 5).: t

a budget equal to the actual cost: $B(t) = C(t) \equiv C(g^*(t), e^*(t))$.
However, at time *t*+1, the head of the department claims a higher budget in such a way

$$B^c(t+1) = (1+\eta)B(t)$$

where $\eta > 0$ is the rate at which the department inflates its budget claim.
This parameter could be well the same for all the other departments. Such a rule represents a straightforward mechanical method of updating the budget claim, even if, as in this case, is inefficient: "...last year's budget is taken and a little more added". Now the government, reunited in a meeting of cabinet, takes these bids and proportionally reduces them to reach the final allocation, taken for granted a certain level of inefficiency of each department. The agreed budget may be as follows:

$$B(t+1) = (1-\gamma)B^c(t+1) = (1+\eta)(1-\gamma)B(t)$$

where 0<γ<1 is the rate at which the cabinet deflates each budget claim. Thus, the department receives $(1+\eta)(1-\gamma)C(t)$, and the term $(1+\eta)(1-\gamma)$ describes the dynamics of the budget in the time, i.e. if the change from *t* to *t*+1 of the government transfer budget is positive or negative.

The uninformed cabinet can control in someway the inefficiency of the department if $\gamma > \frac{\eta}{1+\eta}$, as the budget is decreasing. It would obtain a perfect control when

$$\gamma = 1 - \frac{1}{(1+\eta)(1+\sigma)}$$

where σ is such that $C(t) = (1+\sigma)C(g^*(t), e_{\max})$.

In this case the cabinet could reconduct the department at the minimum cost for supplying the public good: $B(t+1) = C(g^*(t), e_{\max})$.
However, given the lack of information it is difficult for the cabinet to obtain this result as the head of the department can have a strong political support within the government itself. One way to try to reduce the bureaucratic power of the head of the department and the consequent cost-inflation is to develop "Spending review" procedures. These are rules imposed to each department for pursuing more efficient performance, enforced by monitoring and controlling activities "in the field", better carried on if entrusted to independent Authorities or Commissions. They should also be supported by a consistent set of incentives, of both *monetary* and *carrier concern* nature.
For instance, an incentive-related budget determination could establish a remuneration to the head following this scheme

$$B(t+1) = \tau + C(t),$$

where the budget allowed at time *t*+1 is directly transferred to the head of the department, but without takinge into account of his claim $B^c(t+1)$. It is, instead, equal to the actual cost at time *t* (the only information the Authority has to know) and a bonus (or a penalty) specified in such a linear form[9]:

$$\tau = a - bC(t).$$

[9] This incentive-compatible transfer mechanism has been initially formulated by Jean Jacques Laffont and Jean Tirole [see the description of the L-T model by Benassy-Quéré et al. (2014, p. 98)].

Hence, the budegt at time t+1 is

$$B(t+1) = a + (1-b)C(t),$$

where $1 \geq b \geq 0$ is the coefficient of incentive-power of the scheme.
If $b \to 1, B(t+1) = a$, the budget is just equal to the constant term of the bonus function, i.e. is a "fixed price transfer", independent to the actual cost sustained by the head of the department, who thus is boosted to contain the latter, being residual claimant of the amount a-C.
If, instead, $b \to 0, B(t+1) = a + C(t)$, so the budget is a "cost-plus transfer", as the head of the department receive the constant term of the bonus and is totally assured from the risk of the cost. As a consequance, he has no incentive to cost-containing.
$b \to 1$ might be considered by the Authority the best remuneration mechanism. However, it may happen that a costly quality of g is related to the chosen mechanism: generally, the quality tends to be improved (reduced) with a cost-plus (fixed-price) system. Therefore, a power coefficient such that $1>b>0$ can provide a better (second best) intermediate solution.

Chapter 2

Topics on Economics of Law

1. Civil law, liability and incentives

Civil law is one of the main institutions of modern societies and Economic analysis of law is devoted to treat two basic questions about civil legal rules. The first one is following a positive approach and refers to the effects of legal rules on the behaviour of relevant economic agents; the second one is working in a normative context as it refers to the evaluation in social terms of these effects. The approach employed is that used in economic analysis generally, as it emphasizes the use of stilized models and it postulates that individuals and firms are forward looking and rational, acting with a view toward the possible consequences of their choices.
Moreover, the welfare economic framework is adopted to assess social desirability. Thus, the social optimality of an action with legal consequences is described by the condition of equality between its social marginal benefit and its social marginal cost. Other non-economic approaches often leave the criterion of the social good and choice unclear or substantially implicit. Note that the economic approach to law, although modern and quite recent, has a long tradition as we may say it is going back up since 18th century with the seminal works by Cesare Beccaria and Jeremy Bentham.
The sections of this chapter will cover five basic legal subjects. We begin with the basic concept of liability, a general issue that can be easily illustrated considering the case of accidents. Second, we discuss property law, the nature and the justification of property rights, how they are acquired and transferred. We treat in some details the problem of the concflict on the property itself when the rights are not well defined and assigned. Third, we examine contract law, so the formation, interpretation of contracts and remedies for their breach. The basic issue refers to the concept of incompletness of contracts. The forth topic concerns civil litigation and the last one considers the public enforcement of law, the magnitude of the sanction and their deterrence effects. For providing meaningful examples, we analyse the economic implication of some cases of law violation like corruption, tax evasion and anticompetitive collusion among firms.
It will turn out clear as all these topics are specific applications of the economic theory of incentives.
Let us start with the branch of tort law referring to legal liability for accidents, a means by which society can reduce the risk of harm by threating potential injurers with having to pay for the harms they causes. Liability is frequently viewed as a device for compensating victims of harm, but if we consider that insurance can provide compensation more cheaply than the liability system, we derive that the primary social function of the latter is the provision of incentives to prevent harm.
The aim of the Economics & Law approach is to analyse the link between liability and incentives to reduce risk when the injurer and the potential victim is without contractual relationship each other. We consider two basic rules of liability:

1. The *Strict liability rule* (SLR), according to which the injurer must always pay for the harm he causes, and
2. The *Negligence rule* (NR), according to which the injurer must pay only when he is found negligent. In Italy, this corresponds to the so-called "good father of family" paradigm.

We are going to consider three cases. In the first one, we have that the injurer, carrying out a fixed activity, harms a victim (*unilateral accident*). What we want to establish is the level of care expenditure by the injurer himself to avoid or limit the harm. In the second case, also the victim can do something to reduce the harm (*bilateral accident*) due to a fixed activity by the injurerer and we have to ascertain the level of care of both injurer and victim. In the third case, we have a unilateral

A. Petretto, *Economics of Institutions and Law*, ISBN (print) 978-88-6453-579-1, ISBN (onine) 978-88-6453-580-7, CC BY 4.0, 2017 Firenze University Press

accident, but now the injurer produces an output measured in quantitative terms and we have to find the level of this activity together with level of care expenditure by the sole injurer (*unilateral accident with endogeous activity*).

Let us examine the first case and let us suppose that parties are risk-neutral. Further, we denote with x the level of expenditure care, and with $p(x)$, $p'(x)<0$, the probability of accident. Finally, with h we denote the (certain) harm inflicted to the victim.

From the social point of view the optimal level of care expenditure, x^*, is given by minimizing the expected costs, i.e.:

$$\min_{(x)} ExpC = x + p(x)h.$$

The F.O.C. is

$$-p'(x^*)h = 1.$$

i.e. the marginal benefit in terms of harm reduction is equal to the marginal cost of 1 euro.

Of course, without any rule of liability, the injurer simply chooses $x^0 = 0$. In SLR, instead, injurer chooses $x^0 = x^*$, as its objective becomes just to minimize *ExpC*, because of the expected compensation $p(x)h$ to be added to x. Therefore, the F.O.C. remains the same as before.

In NR, given the "negligence threshold" $\hat{x}$ the injurer will have to pay h if $x < \hat{x}$, but will not have to pay anything if $x \geq \hat{x}$. If $\hat{x} = x^*$ i.e. if the negligence limit is put exactly at the optimal level, in both cases, injurer will choose x^*, which is the minimum level of care leaving him without any compesation. However courts need to be able to calculate x^*, in addition to observe h. If courts cannot do it and can only make some estimation, $\hat{x}$ may be greater or lower than x^*. Given this informational requirement, we may say that, in the unilateral type of accidentes, NR is less efficient than SRL.

Now consider the case of bilateral accident, where also the victim can invest in care to reduce the harm. y is the correspondent ex.penditure. Therefore, the probability function now becomes $p(x, y)$, $\partial p/\partial x < 0, \partial p/\partial y < 0$. The social goal is pursued by solving as follows

$$\min_{(x,y)}[x + y + p(x, y)h]$$

The F.O.Cs of optimal expenditure $x^*>0$, $y^*>0$ are

$$-(\partial p/\partial x)h = 1;$$

$$-(\partial p/\partial y)h = 1.$$

With SLR, contrary to the decision by injurerer ($x^\circ = x^*$), the victim choice is $y^\circ = 0 < y^*$. There is no incentive to invest in care, given the complete insurance guaranteed by SRL system. However, in many systems a mechanism to limit this moral hazard effect is pursued with a "defense of contributory negligence". According to this, an injurer is liable for harm only if the victim's level of care is not negligent i.e. $y \geq \hat{y}$. In this case, $y^\circ = \hat{y} = y^*, x^\circ = x^*$. It could be the unique equilibrium, if the courts, knowing $p(x, y)$, are also able to calculate the optimal care level and to establish $\hat{y} = y^*$.

At the same conditions about the knowledge by the courts, in NR, the game reaches directly the same unique equilibrium.

Now let us consider the third case, again a unilateral accident. Here only the potential injurer can prevent the harm but he has also to decide how much to produce. Hence, both level of care and level of activity are to be determined. Let z be an activity level, e.g. number of times the injurer engage in his activity, for instance a scale of output. $b(z)$ is the benefit from activity z, say a profit function. $x+p(x)h$ is the expected unitary cost of care, equal to the expected harm each time an injurer engages in his activity. Social goal is to maximize the net benefit, solving the following program

$$\max_{(x,z)}[b(z) - z(x + p(x)h)],$$

As in the first case of unilateral accident, also here we reach $x^*>0$ given the F.O.C. is still

$$-p'(x^*)h=1.$$

While the condition for optimal level of activity $z^*>0$ is that marginal benefit is equal to the expected unitary social cost

$$b'(z)=x^*+p(x^*)h.$$

In SLR, the injurer is boosted to reach the optimal values, as its objective function is the same as social one.

In NR, instead, as seen before, it can be $x^\circ=x^*$, but it will happen that $z^\circ>z^*$. Indeed, injurer will escape liability by choosing x^* and he will choose z in such a way

$$\max_{(z)}[b(z)-zx^*].$$

Thus the F.O.C. is as follows

$$b'(z)=x^*.$$

The injurer's cost of raising his level of activity is only his cost of care x^* which is less than the social cost: $x^*+p(x^*)h$. The negligence is defined in terms of care alone and often for courts it is difficult to determine not only z^*, but also estimate z^0.

2. Property law

2.1 Property rights and the theory of firm

A first justification for property rights (PR) is that, where they absent, individuals would spend time and effort trying to take things from each other and protecting things in their possession. Consequently, they would often find themselves involved in conflict. Enforcement of PR by the state, while involving its own costs, reduces these serious disadvantages that would be incurred in the absence of PR. Further, in the absence of PR, individuals would face the possibility that their property would be taken from them but they may also make the same to others. In this sense, protection of PR protects people against risk.

A system of PR allows things to be transferred freely and probably they will tend to be allocated to those who value them most. The ability to transfer things is indirectly necessary for enjoing economics of mass production and specialization of labour, as when a large quantity of a good is produced by a single entity, the output will have to be distributed or transferred to many other individuals, and the entity will also often need to obtain inputs from other parties. In addition, transferability of property of assets allows it to be used as collateral, so contributing to the credit markets functioning.

Note that, in any case, PR are not strictly related to market working and functioning in terms of well-being, as the benefits of PR enforcement could be enjoyed also under a centrally planned economy. For example, incentives to work can be provided by paying workers on the basis of effort, even if a state enterprise owns what they produce.

Property rights can be devided in (i) basic rights, composed of particular possessory rights and (ii) right to transfer these rights. What we commonly conceive of as "ownership" of something entails both a large swath of possessory rights and associated rights to transfer them. Among the former ones, we have the right to build on land, plant on it, and so forth, under most contingencies and to the infinite future. The division of possessory rights may be valuable when different parties derive different benefits, although sometimes there are also some disadvantages that could boost toward a concentration. In general, the combination of possessory rights and rights to transfer may promote efficiency throught the incentive to invest.

Ownership of separate productive assets is often consolidated, namely is held by a single entity, the firm. The issue is developed by the modern theory of firm due to authors like Coase, Willimamson

and Hart [Aghion and Holden 2011]. The discussion refers in ascertaining if it is better to have a consolidated form of PR, by means of the vertical integration of productive activities, or separate property rights and activities. In other words if it is better to build a firm or instead running market contracts among all the involved agents.

The main advantage of consolidated ownership of assets is that it reduces transaction costs because internal transfers of goods and services may be accomplished by command, eliminating the need for negotiation and bookkeeping expenses. However, consolidated ownership may lead to a diluition of incentives to work, in comparison to the situation where each individul owns the assets he uses in production. Firms can combat this problem if they can observe individuals' efforts by penalizing shirking; if not, i.e. in a moral hazard context, they can tie compensation to measures of output. Another advantage of consolidated ownership is that it enables a firm to avoid breakdowns in bargaining that would occur under separated ownership due to asymmetric information. For instance, a seller of a factor input might overstimate its value to the next-stage producer and demand too much for it, stymieing an efficient transfer. With consolidated ownership, efficient transfers can be directly ordered. Further, consolidated ownership can alleviate problems of inadequate investments in assets. Indeed, an asset owner may not have a sufficient incentive to make a relationship-specific investment because he anticipates that his gains will be partially expropriated by the owner of a complementary asset, at the time when he is to put his asset in use. However, if both assets are owned by the same party, the problem of expropriation of the gains from the relationship-specific investment in the first asset will be mitigated, and investment in it should be more efficient. However, the other individual's incentive to invest in what otherwise would have been asset may be duiled if the first party owns both assets; thus, consolidate ownership does not necessarily improve investment incentives overall. Additionally, it may sometimes be possible, under separate ownership of assets, to guarantee that investments in them be sufficient by making a contract to that effect; but it requires that investments be observable.

In any case, it is useful to note that separate ownership, combined with sufficiently encopassing contracts, may be indistinguishible from the consolidated of ownership of assets by firms. Conversely, firms themselves can be understood to consist of a set of contracts, as it happens with a corporation that is a particular contract among its shareholders.

Let us formalize the esposed arguments by an example from Aghion and Holden (2011).

Consider a relationship between a buyer (*B*) and a seller (*S*) of an intermediate good, called simply INPUT. *B* can use INPUT to produce a final good, called GOOD, which can be sold to a consumer, which values it at *v*. *S* can make a privately costly investment that makes INPUT cheaper to produce, for instance a new producing machine. If *S* makes the investment, at a cost of *5*, then INPUT can be produced at 10, other wise it costs 16 to produce. *B* can make a privately costly investment, which makes GOOD more valuable to the consumer, for instance an enhancing quality-producing machine. This investment also costs 5. If *B* makes the investment then v=40; otherwise v=32. Only *B* can make the revenue enhancement investment and only *S* can make the cost reduction investment, for instance because thay have different human capital characteristics.

B and *S* would like to write a contract that specifices that each party should make its respective investment, because that leads to the total surplus in the relationship to be 40-10-5-5=20 that is GOOD is sold for 40, it costs 10 to produce, and *B* and *S* each incur an investment cost of 5. However, suppose that the two parties cannot contract on INPUT, nor can they contract on a cost-sharing rule ex-ante, nor can they contract on the investment. As we will specify more accurately in the following section, we are facing an *incomplete contract. S* and *B* will have to bargain about the price that *B* pays to for INPUT after the investment stage. Suppose that *B* and *S* are *nonintegrated* so that at the bargaining stage they split whatever surplus is generated 50:50. In this situation, it is not convenient for *B* to invest. If *B* does invest, he will bear a private cost of 5, but gets half of the increase in surplus of 40-32=8, or 4. Similarly, *S* bears a cost of 5 by investing, but gets an increased pay-off of (16-10)/2=3 in the bargaining. So S, too, will not invest. When neither *B* nor *S* invests, totale surplus is

then 32-16=16. Now consider the case *B* and *S* are *vertically integrated*, with *S* owning *B*'s machine for producing GOOD. *S* no longer needs to bargain with *B*. Hence, *S* will receive the entire increased surplus from investing in cost reduction, i.e. 16-10-5=1, and thus will be prepared to invest. However, *B* will not invest as he will get none of the benefit of making GOOD more valuable. *S* cannot compel *B* to invest, nor contract on *B* making the investment. Total surplus is thus 32-10-5=17; this is larger than under nonintegration, so forward vertical integration is desirable.
In fact, *B* ownership (backward integration), does better still. Now *B* invests, but *S* does not, getting total surplus of 40-16-5=19. This is not so good as if contracting was possible (complete), that would yield a surplus of 20, but it is better than the other possible ownership structures. What makes *B* ownership preferable to *S* ownership is that *B*'s investmennt is relatively more important at the margin than *S*'s. Both cost 5, but *B*'s has a benefit of 40-32=8, whereas S's has a benefit of 16-10=6.
The example highlights two basic propositions of the theory of firm: (i) asset ownership can help to mitigate inefficiencies that would otherwise arise from underinvestment in productive activities; (ii) the party whose marginal investment is more productive should own the asset.
Actually, *B* ownership emerges as the equilibrium ownership structure as it maximizes joint economic surplus.

2.2 Conflicts in the use of property: The tragedy of the Commons

We analyse two cases of inefficient use of a scarce resource in the absence of PR: the case of the fish in a lake and the case of the space in a car park. Both cases emphasize not only the usefulness of property but also of the correct assignment of it.
As far as the first example is concerned, let n be the number of equal boats, each hired at cost c (rent) for fishing in a lake. Let $f(n)$ be the amount of fish caught, in a given time, by a boat. The boats are "strategic substitutes", given that the total fish in the lake is in some way, at least in the short time, fixed: $f'(n)<0$.
Therefore, fish are non-excludable goods but with a degree of rivalrness. $p\,f(n)$ is the total revenue of a boat, where p is the price of a fish sold in the market. Let w be the opportunity cost of running the activity by an angler and π his profit.
In equilibrium of free entry, the number of boats n^0 is such that each boat's profit is equal to zero:

$$\pi = p(n^{\circ})f(n^{\circ}) - c - w = 0.$$

The social welfare is given by the consumer willingness to pay for the total fish available in the market $v[nf(n)]$, minus the cost of catching them. It can be expressed in term of the number n of boats as follows:

$$W(n) = v[nf(n)] - n(c+w).$$

By differentiating $W(n)$ w.r.t. n we obtain:

$$W'(n) = [f(n) + nf'(n)]v'(nf(n)) - c - w$$

or, given that in market equilibrium, $p = v'(nf(n))$,

$$W'(n) = npf'(n) + \pi.$$

In equilibrium we have seen that $\pi=0$. Thus $f'(n)<0$ implies that in equilibrium $W'(n)<0$. As the social optimal number of boats n^* is such that $W'(n^*)=0$, it means that $n^* < n^0$, i.e. that there are too many boats operating in the lake. Consequently, there is an excess of exploitation of the scarse resource as a common good.
This inefficiency is due to the lack of property rights on any portion of the lake. If a PR would assigned to a fisherman, he could invest in his sector for creating new fish population, without being afraid to have the relative benefits expropriated by other "free rider" fishermen. If, however, transactions costs are too high and other technical obstacles arise, in absence of an appropriate distribution of property rights, the inefficiency can be mitegated by some correctives. Particularly, giving to a regulator the task of defining a system of Pigouvian taxes per boat, so increasing their costs, or fixing quota of

circulating boats, i.e. imposing by law $n^0 = n^*$, if the latter is known.
The amount of a Pigouvian tax for each boat is $p\,f'(n^*)$, accordingly we obtain

$$W'(n^\circ) = p(n^\circ)[f(n^\circ) + n^\circ f'(n^\circ)] - c - w = 0, \text{then } n^\circ = n^*.$$

As it is intuitive, a set of Pigouvian taxes corresponds to applying a *Strict liability rule* with a compensation equal to the expected harm (the damage), while fixing quota is a solution coherent with a *Negligence rule*.
Let us consider now the conflict in the use of property arising for the provision of a service like a car parking.
Let Q be the number of cars that could enter the parking. With $\Pi = p\,Q$, we denote the total benefit by the service, and with $p=a-Q$ the marginal benefit (along the demand function).
If there is no ownership and no price setting, the service is freely provided, hence $p=0$ and $Q=a$. Actually, there is no difference between parking in or out the garage, so the outcome is inefficient.
Now, let us suppose there is one exclusive owner getting the total benefit (surplus), as a profit (we suppose no costs of organizing the car park). By maximizing the total profit/benefit, i.e. solving the program

$$\max_Q \Pi = pQ = (a - Q)Q,$$

we obtain the following first order condition, giving the equilibrium monopolistic quantity:

$$(a - Q) - Q = 0, Q = a/2; p = a/2.$$

We reach an efficient solution according to which, being all individuals equal, they pay the same price equal to the marginal benefit.
Notice how in this case, the monopoly has not social costs; indeed, the aggregated surplus is maximized because prices are personalized. It is a particular case of perfect discriminating monopoly (or *first degree price discrimination*), extracting the whole consumer surplus. In terms of the Economics & Law paradigm, the fairnes of total surplus distribution is a problem to face in a second time, through lum sum transfers and taxes.
Let us compare this solution with the case of separated ownership and let immagine two owners, *A* and *B*, selling independently the service (two separated entries to the garage). The model becomes a Cournot duopoly. Consequently, we look at its Nash equilibrium: Each owner tries to maximize, by choosing the quantity, its pay-off, taking as given the output of the other player.
Thus owner *i, i=A,B* solves the following program:

$$\max_{Q_i} \Pi_i = pQ_i = (a - Q_i - Q_j)Q_i$$

Thus, the reaction functions become, for *A*:

$$Q_A = (a - Q_B)/2,$$

and for *B*

$$Q_B = (a - Q_A)/2.$$

In Nash equilibrium, solving the equations system, we have:

$$Q_i = \frac{[a - (a - Q_i)/2]}{2}, i = A, B$$

$$Q_A = Q_B = a/3; p = a/3.$$

As $Q = 2a/3 > a/2$ in Nash duopoly equilibrium there is an excess of exploitation, i.e. too many cars intend to enter the garage, with a clear worsening of the quality of the parking. Also in this case we can control the inefficiency by increasing the price over the marginal benefit (a Pigouvian tax from *a*/3 to *a*/2, i.e. *a*/6) or introducing quantity restraints on the available space, given a total number of cars of $Q=a/2$. Both solutions have informational requirements shortcomings. However, the First best outcome might be reacheable by having an exclusive owner, i.e. by defining and distributing, in case

of negligibe transaction costs, the property rights in order to assign them to only an agent able to maximize the social surplus.

2.3 Conflicts in the use of property: Externalities and market inefficiency

An externality is present whenever some economic agent's welfare (utility or profit) is directly affected by an action of another agent (consumer or producer) in the economy, without any mediation by the price system, i.e. with no explicit compensation for the effect. In this case, even Competitive equilibrium (*CE*) fails to reach Pareto-efficiency (*PE*). To see this let us consider a simple model, with two individuals 1 and 2 and two commodities *x* and *z* [Hindriks and Myles (2013, ch. 8). The first one is a consumption good acting as numeraire and the second one is a private consumption good generating an externality. Thus the utility functions for the two individuals are:

$$U^1 = x^1 + u_1(z^1) - v_1(z^2)$$
$$U^2 = x^2 + u_2(z^2) - v_2(z^1)$$

$u_i(z^i) > 0, u_i'(z^i) > 0$ is the utility to *i* in directly consuming good *z,* while $v_i(z^j) > 0, v_i'(z^j) > 0$ is the damage to *i* from the consumption of good *z* by *j*.
In *CE* the following conditions are satisfied

$$u_h'(z^h) = 1,$$
$$x^h + z^h = \varpi^h,$$
$$p_x = p_z = \frac{\partial U^h}{\partial x^h}, h = 1, 2$$

together with the total resources constraint

$$x^1 + x^2 + z^1 + z^2 = \varpi^1 + \varpi^2 \equiv W.$$

For both goods marginal utility is equal to marginal cost of 1 € and this equals the market price. Further, all the resources are fully employed. However, this set of conditions do not satisfy *PE*. The latter can be obtained in a utilitarian (cooperative) framework by maximizing the total utility taking into account the total resources constraint. That is

$$\max_{\{x^h, z^h, h=1,2\}} \quad U^1 + U^2$$
$$\text{s.t.} \quad x^1 + x^2 + z^1 + z^2 = W.$$

Hence, the F.O.C.s are

$$u_i'(z^i) - v_j'(z^i) = 1, i, j = 1, 2.$$

The R.H.S. is still the marginal cost equal to 1 €. The L.H.S is the social marginal benefit of consumption z^i which is lower than the private one $u_i'(z^i)$. In other words, with negative externality the consumption of *z* should be lower than that decided in *CE*. Symmetrically, in case of positive externality.
In order to correct the consumer prices to contemplate the social cost of externality and to have the coincidence of Pareto and competitive equilbirum conditions, we may introduce a Pigouvian taxes system in this way:

$$u_1'(z^1) = q_1 = 1 + t_1 = 1 - v_2'(z^1)$$
$$u_2'(z^2) = q_2 = 1 + t_2 = 1 - v_1'(z^2)$$

For applying Pigouvian taxes there must be a different tax rate for each externality-generating good for each consumer. This is a very strong information requirement: the state must know at least the expected harm for inducing injurers to act optimally by balancing the cost of precaution against the reduction of expected harm that would be brought about. An alternative way is to build a system of licences by legislating that externalities can only be generated up to the quantity permitted by licenses held.
The difficulty here is the same as with *Negligence rule:* How to know ex-ante the optimal level? In any case, if the licences are tradeable, a sort of market of permits may arise and work.

2.4 Solving the conflicts by bargaining: The Coase theorem

The inefficiency of the free competition with an external effect derives from a lack of property rights or an unadequate distribution of them. The idea of copying a market system to reach *PE* for distributing property is that of the bargaining theoryzed by the famous Ronald Coase's theorem (CT). We have two linked Propositions for declining it:

Coase Theorem
a) *If property rights are clearly defined and transferable and transactions costs are negligible, bargaining among parties reaches an efficient solution*.
b) *The efficient outcome of the bargaining can be reached independently how property rights are ex-ante assigned*.

In order to explain the working of parts a) and b) of CT let us consider the bargaining solution in case of two firms, one polluter, *A* (damaging) and one pollute, *B* (damaged).
The technology of the polluter is given by the following cost function:
$C_A = C_A(q_A)$, with $C'_A(q_A) \equiv PMC_A$, as the private marginal cost.

A'profit is given by $\Pi_A = p_A q_A - C_A(q_A)$; so the chosen level of output q_A^0, according to a competitive behaviour, is such that $p_A = C'_A(q_A^0)$.. In equilibrium, as Π_A is at the maximum, the marginal profit is zero: $\frac{\partial \Pi_A}{\partial q_A} \equiv MP = 0$. The technology of the pollute *B* is given by the following cost function:
$C_B = C_B(q_B, q_A)$; $\frac{\partial C_B}{\partial q_B} > 0, \frac{\partial C_B}{\partial q_A} > 0,$
so *A* generates costs to *B*, whose profit is $\Pi_B = p_B q_B - C_B(q_B, q_A)$, and $\frac{\partial C_B}{\partial q_A} = -\frac{\partial \Pi_B}{\partial q_A} \equiv MD$ is the marginal damage inflicted to *B*.
$PMC_A + MD$ is the social marginal cost of output *A*, which, in order to reach *PE* , must be linked to the social marginal benefit, SMC_A , which in the model is equal to the market price of output A.
Notice what may happen in terms of distribution of property rights. If it is the polluter *A* to have the right to pollute how much he wants, the chosen level is the profit maximizing one: $q_A = q_A^0$. If, instead, the right of not being polluted is fully given to *B*, the outcome is q_A =0. However, in neither case *A* and *B* are respectively satisfied; each one would like to find an agreement with the other party.
We have a bargaining on q_A^* if totale profit, $\Pi_A + \Pi_B$, is maximized, i.e. when this first order condition is satisfied

$$p_A - C'_A(.) - \frac{\partial C_B}{\partial q_A} = 0, \text{ i.e. } MP = MD,$$

so the polluter may compensate the pollute or viceversa, without any gain at the margin.
Clearly at bargaining equilibrium we have $0 < q_A^* < q_A^\circ$ and $p_A = SMC_A$, i.e. the *PE* condition.
What is the interpretation of CT in terms of economics of law? The normative interpretation of the theorem by Chicago School of law is as follows: If transaction costs are negligible, the market absorbs the law, so what it remains to law is only the task to promote efficiency facilitating contracts. Then the civil law is working as simply a lubricating oil of the exchanges system. However, it is also clear that when CT fails, by relaxing some hypotheses, like that one on negligible transaction costs, the role of the civil law becomes relevant. In particular, the law pursues the tasks of defining "appropriately" the allocation of property rights to economic agents and disciplinating the extension of themselves.

3. Contracts

3.1. Nature and functions of contracts

The contracts are the main institutional instruments for the working of the market. Thus, it is possibile to establish a relationship between contract theory (CT) and welfare economics (WE). In both cases, we look for benchmark situations: In CT we refer to the notion of complete contracts and negligible transaction costs, while in WE we refer to a complete markets and no externalities framework. CT defines the exchange pay-offs for the parties and the total surplus, while WE considers prices as the best allocative devices for decetralizing the economic choices. In CT the conditions for maximizing the total surplus of the exchange are described by the Coase theorem, while in WE we look at the relationship between Pareto efficiency and pure competition, through first and second fundamental theorems. In CT the convenient transaction is such that the marginal benefit and the marginal cost of an activity are equal. In WE the condition is the same, with the further implication that, in perfect competition, the condition requires the price to be equal to the marginal cost. In CT the surplus distribution between the contractual parties is, at a first stage, irrelevant; in WE this statement is confirmed, as the First best is based on the paradigm of conceptual and instrumental separation between efficiency and equity.
The methodological parallel is even clearer by taking into account that CT defines a *Pareto efficient contract* if it is impossible to modify it in a manner that raises the expected utility of both the parties. In this case, we say that a contract is *mutually beneficial*. Further, a contract is *constrained Pareto efficient* if it also satisfies, in a asymmetric information setting, some incentives constraints. In WE these circumstances define a Second best framework.
The main function of a contract is constraining parties to promises, against the incentive to opportunistic behavior. The contract, once signed and validated, has the effectiveness of the law, as contracts are assumed to be enforced by a tribunal. The possibility to be called before a court gives the incentive to enforce the contract, by providing a credible threat, and, given this feature, a contract provides a strong incentive to invest. Thus, an economic activity is working nicely if carried on under the shadow of the law, even if there are other factors boosting correct behaviors, like ethics, social norms and reputation.
Although enforceable, contracst are desirable even if substantially imperfect because they are significantly incomplete. The incompleteness means that contracts leave out all manner of variables and contingencies that are of potential relevance to parties and they also fail to employ included variables in a mutually beneficial manner. We have already encounter the incomplete nature of a contract, in section 2.1, where we were analysing the cost-benefit of vertical integration of firms. Recall that such a problem arises because the two parties involved in a relationship, where a seller provides an intermediate good to a buyer, cannot contract on features of this input, nor can they contract on a cost-sharing rule ex-ante, nor can they contract on the investment for enhancing the

quality of the final good or reducing the cost of producing the input. Given this incomplete nature contracts are guaranteed by a less-than-rigorous enforcement.
There are three general important reasons for the incompleteness of contracts. First, there are the expenditure needed for writing complete contracts in terms of transaction costs. Second, some variables (effort levels, technical production difficulties), although observables by the parties, cannot be verifiable by tribunals or verifiable only at high cost. Third, the expected consequences of incompleteness may not be very harmful to contracting parties, thus an incomplete contract can be well satisfactory for all.
With incompleteness it results quite important the interpretation of contracts that may fil tribunal's gaps, resolve ambiguity and overrid literal language. This can benefit the parties by easing their drafting burdens or reducing their need to understand contracual details. Given a method of interpretation, parties will choose contracts in a constrained-efficient way.
When parties breach a contract, they often have to pay damages in consequence. We distinguish the *Damage measure system*, where the formula governing what they should pay can be determined by the tribunal or it can be stipulated in advance by the parties themselves. In this case, the prospect payment of damages provides the incentives to perform contractual obligations and promotes enforcement of contracts. As alternative use of damage measure, there is the *Specific performance system*, requiring a party to satisfy, in any case, his contractual obligation. Clearly, specific performace have problems of monitoring and controlling parties' effort levels and the quality of production concerning the contract.
Parties often have the opportunity to renegotiate their contracts, and this will always occur when inefficiency would otherwise result.

3.2 Contract formation: search effort

An important aspect of contract formation is the effort individuals devote to it, i.e. the time and resources they expend for searching contractual opportunities. It is useful to compare the socially ideal and the privately desired degree of search effort. There may be either to much or too little search for contracts, depending on circumstances. On one hand, as the return that a person can obtain by the surplus division is less than the surplus itself, the search activity, in principle, tends to be socially inadequate. On the other hand, if a contract formation by one party tends to deny others from making contracts, the search effort can be even excessive. Let us analyze formally this cost-benefit framework.
Let e the effort a person decides to devote to search for a contract partner. There is one suitable partner, whom the person will discover with a probability, increasing function of the effort: $p(e)$, $p'(e)>0$. If the person discovers the partner, a contract will be made, resulting in a surplus s, and the person will obtain, say, a fraction γ of this, so his net pay-off from search will be as follows:
$\Phi_{se} = p(e)\gamma s - e.$
If the person does not find the partner, the partner will make a contract with another party, and that will produce surplus t. The socially optimal amount of search is determinated by maximizing the followiong net social pay-off

$$W_{se} = p(e)(s-t) - e,$$

since the ex-post social pay-off to the person's finding the partner is s-t.
Clearly, the person will search too much (little) if

$$\Phi_{se} > (<) W_{se},$$

or

$$p(e)(\gamma s - s + t) > (<)\, 0, \text{and then if } \gamma > (<)\, 1 - \frac{t}{s}.$$

Note that if $t>s$ the person will search too much if he can obtain a positive, even negligible, fraction of the surplus of the contract.

The nature of the comparison between the private and the socially desirable degree of search effort raises questions about whether social authorities could obtain the information needed to formulate a corrective policy. Actually, in the real world, it turns always more important the role of internet for searching and easily finding partners for many types of contracts.

3.3 The efficient performance of a contract

Let us suppose that symmetrically informed risk-neutral parties - a buyer, *B*, and a seller *S* - enter into contracts, and that the only variables of concern are the value of performance and the production cost. *S* faces uncertain production cost *c* which will learn before he decides whether to produce. *v* is a certain value of performance to *B*. The Pareto efficient outcome is for *S* to produce if and only if $c<v$. To be more specific, in a complete contract, with terms for all contingencies, performance would be required if and only if the production cost of the seller is lower than the performance value for the buyer. Further, a change in the terms of contract price would compensate a party, for agreeing to alter a term from any iniatially considered contract under which perfomance does not occur, if and only if $c<v$. Then the contract is a mutually beneficial transaction producing a positive surplus v-c, which will be distributed according to the price of the transaction, *p*.

In the absence of *contract enforcement*, then there would be too little production because *B* would only pay *S* for actual delivery of the good and cannot guarantee the price. In particular, if *S* would obtain a fraction $\tilde{a}$ (his bargaining strenght and $1-\tilde{a}$ that one to *B*) of the surplus from a transaction, he would obtain a price $p = \tilde{a}v$ Thus, *S* would decide to produce only when $c < \tilde{a}v = p$ rather than whenever $c<v$.

If there is *contract enforcement* and the parties cannot renegoziate before *S* decides whether to produce, and if *c* is verifiable by the tribunal, the parties could write a "complete contract" specifying performance if and only if $c<v$. The parties would want a *damage measure d* for breach of this contract to be sufficiently high to induce performance when $c<v$, and thus any *d* exceding *c* would work.

If *c* is not verifiable, the parties are able to write an "incomplete contract" specifying:"…. *S* shall deliver the good to *B*, who will pay *p* at the outset…..", accompanied by damages *a* for *S* breach. Under such a contract, S will perform when $c<d$ and will commit breach otherwise. If the *expectation measure* is employed, i.e. $d=v$, *S* will perform if and only if $c<v$, so that performance will be efficient. If $d>v$ there will be excessive performance, as there will be if there is *specific performace*. If $d<v$, there will be little performance.

There could be a post-breach mitigation behavior of *B*, in the sense that he could mitigate the consequences of the breach by searching for alternative suppliers and the like. Let *z* be the mitigation expenditure of *B* to raise his ex-post-breach value, say $w(z), w'(z) > 0.$ Efficiency requires *B* to choose *z* to maximize $[w(z) - z]$ and first order condition requires as follows:

$$w'(z^*) = 1.$$

If *y* is the gross value of *S* performance to *B*, then we can define *v*, the net value of performance, as $v = y - [w(z^*) - z^*].$

Thus, expectation damages for breach should equal this *v*, not the gross value *y*. And if the damages equal v, then *B* will choose z^* if he is the victim of the breach, and the net value of the performance will actually be *v*.

Now consider the case of renegotiation. If the parties can renegotiate their contract after *c* becomes

known but before *S* decides whether to produce, then, given symmetric information, it is natural to suppose that there will always be Pareto efficient performance, regardless of *d*. If *B's* value *v* is uncertain as well as *S* cost *c,* the major difference in the outcome is that, since *v* cannot be prescribed as damages in the contract, *v* must be verifiable for the expectation measure $d=v$ to be applied by the tribunal (*c* is not still verifiable).
However if *c* is verifiable, and *v* is not, Pareto efficient performance can be achieved by constructing the contract so that *B* will commit breach by refusing to pay for performance when it would be inefficient. Let the price $p>0$ be paid at performance and let damages for B breach be $d=p-c$ i.e. S' profits. Then *B* will breach and refuse performance whenever

$$v - p < -(p - c) \text{ or } v < c.$$

3.4 Reliance investment during the contract period

The existence of a contract may have an important implication. It can boost the parties to invest in some activity, taking for granted the realization of it: the so called "reliance investment". In general, the buyer, *B*, makes reliance investment, trusting on the delivery of the good, by anticipating some costly actions, e.g. an advertising campaign for the commodity he is planning to produce and sell by transforming the input provided by the seller, *S*. However, the seller's costs are uncertain, hence *c* is a random variable characterized by a probability distribution with density function $\phi(c)$, and cumulative function $\Phi(c) = \int \phi(c)dc.$
Let *r* be the *B*'s reliance investment, and $v(r)$ the value of the contract performance given *r*. Clearly, this value is increasing on the reliance investment: $v'(r)>0$.
B chooses *r* before *S* learns *c* and decides about producing. For *S,* once known $v(r)$, the efficient produciont decision, i.e. the best contract performance, is when $c<v(r)$. For *B*, instead, the rational action is such that the reliance investment level solves the following stochastic maximization problem

$$Max_r \int_0^{v(r)} [v(r) - c]\phi(c)dc - r.$$

The F.O.C. is

$$v'(r^*)\Phi(v(r^*)) = 1,$$

where

$$\Phi(v(r^*)) = \int_0^{v(r)} \phi(c)dc$$

is the probability of getting the value *v* of the performance, given r^*. In words, the expected marginal return is equal to 1, the marginal cost of *B*'s relience investment.
Note that the marginal return to reliance investment is only a contingent return. Indeed, the investment pays off only with probability $\Phi(v(r))$ when $c< v(r)$ i.e. when the efficient production is realised.

3.5 Asymmetric information and incentives

Two general types of incentive problems must be considered when information is not distribuited symmetrically among the parties of a contract. One is the *hidden information* and the other the *hidden action.* Let us consider employment contracts as presented by Bolton and Dewatripont (2005, ch.1). The first problem (*adverse selection*) refers to a situation where the employee may have private information about his inability or unwillingness to take on certain tasks. That is, the information about some relevant characteristics of the employee are hidden from the employer. The second problem (*moral hazard*) refers to situations where the employer cannot see what the emplyee does, whether he works or not, how hard he works, how careful he is.

3.5.1 Adverse selection

For modelling the first asymmetric information problem, consider an employer who contracts with two types of employees, a "skilled" employee and an "unskilled" one, and who does not know which is which. According to the so-called *Revelation principle,* it is optimal for the employer to consider offering only two employment contracts, one destinated to the skilled employee and the other to the unskilled one, but to make sure that each contract is *incentive compatible*. This will be the case if each type of employee wants to pick only one contract, that destinated to him.

Let the employer utility function be $U[a\theta(1-l)-t]$ and the employee utility function $u(\theta l+t)$, where (1-*l*) is the employee time sold to the employer, and *l* is the time the employee keeps for himself, *t* is the monetary/output transfer from the employer to the employee. *a* is a positive constant, that we assume >1 (the time is more efficient when sold) and θ measures the "unit value of time" or the skill level of the employee. The state of nature θ is learned *privately* by the employee before signing any contract; he knows whether he is skilled with a value of time θ_H or unskilled, with value of time $\theta_L < \theta_H$. The employer knows only p_H the probability of facing a skilled employee, and conversely $p_L = 1 - p_H$. The relevant reservation utility is $\hat{u}_H = u(\theta_H)$, when facing a skilled employee, and $\hat{u}_L = u(\theta_L)$ when facing an unskilled employee.

If the employer could also learn the employee's type, he would simply offer in state θ_j a contract with a transfer $t_j = \theta_j$ in exchange for all his work time, i.e. 1-l_j =1. Such a contract would maximize production efficiency and since the employees's *individually rationality constraint,* $u(t_j) \geq u(\theta_j)$, would be binding under this contract, it would maximize the employer's pay-off.

If employee productivity is, instead, private information, the employer, given the offer of wage contract $t_j = \theta_j$, would respond in any case by pretending to be skilled to get the higher wage rate θ_H. In this situation the only contracts the employer can offer the employeee are those offering a total payment of *t(l)* in exchange for (1-*l*) units of work. The Revelation principle gives the key semplification for finding the optimal contract, as it says that all the employer needs to determine is a menu of two "points contracts": (t_L, l_L) and (t_H, l_H), where by convention (t_j, l_j) is the contract chosen by type *j*. As each type of employee would pick only one point on the full schedule *t*(*l*) anyway, the employer may as well pick that point directly, if it is incentive compatible. Hence, type θ_H must prefer contract (t_H, l_H) over (t_L, l_L), and type θ_L contract (t_L, l_L) over (t_H, l_H).

Thus, the optimal menu of employment contracts under hidden information can be represented as the solution to the optimal contracting problem under complete information:

$$\max_{(l_j, t_j)} \; \{p_L U[a\theta_L(1-l_L)-t_L] + p_H U[a\theta_H(1-l_H)-t_H]\},$$

subject to the reservation contraints (RC)

$$u(l_L\theta_L + t_L) \geq u(\theta_L) \equiv \hat{u}_L \text{ and } u(l_H\theta_H + t_H) \geq u(\theta_H) \equiv \hat{u}_H$$

but, given the incomplete information, with two additional *incentive contraints* (incentive to say the true, IC)*:*

$$u(l_H\theta_H + t_H) \geq u(l_L\theta_H + t_L)$$

$$u(l_L\theta_L + t_L) \geq u(l_H\theta_L + t_H)$$

The solution to this constrained optimization problem will produce the most efficient contract under hidden information, a second best contract as it in general results in less efficient allocations than under complete information. An inefficiency due to the informational monopoly power of an agent/party.

3.5.2 Moral hazard

Formally, to introduce hidden actions into a standard employment with uncertainty, we may suppose

that the amount of time e=1-l worked by the employee is private information. In addition, the employees chooses the action e before the state of nature θ_j is realized and that this action influences the probability of the state of nature. When the employee chooses action e, the output for the employer is simply θ_H with a probability $p_H(e)$, $p'_H(e)>0$, and θ_L with probability function $p_L(e)=1-p_H(e)$. Notice that e is interpreted as "effort" and more effort produces more output, at cost e for the employee. Since effort is not observable, the agent can be compensated only on the basis of realized output θ_j. The employer is thus restricted to offering a compensation contract $t(\theta_j)$ to the employee. Also the employer must now take into account the fact that e will be chosen by the employee to maximize his own expected pay-off under the output-contingent compensation scheme $t(\theta_j)$. Formally, the effort must solve the following optimization problem

$$e \in \arg\max_l \quad \{p_L(e)u[t(\theta_L)+l]+p_H(e)u[t(\theta_H)+l]\}$$

Therefore, when the employer chooses the optimal compensation contract $[t(\theta_j]$ to maximize his expected utility, he must make sure it is in the employee'se best interest to supply the right level of effort e. Hence, he must solve the following optimization problem:

$$\max_{t(\theta_j)} \quad \{p_L(e)U[\theta_L - t(\theta_L)]+p_H(e)U[\theta_H - t(\theta_H)]\}$$

s.t. both the (RC) constraint:

$$p_L(e)u[t(\theta_L)+l]+p_H(e)u[t(\theta_H)+l] \geq \bar{u} = u(1)$$

and the (IC) constraint, requiring e be a solution of the previous employee's utility maximization problem.

As in contracting problems with hidden information, when the action supplied by the employee is not observable the employer must take into consideration not only the employee's individual rationality constraint (RC) but also his incentive constraint (IC).

4. Litigation and suit

4.1. Private incentive to suit and socially optimal level of suits

Civil litigation is the bringing of lawsuit by private actors to enforce their rights in the area of civil law. The expense involved in the operation of legal system has some economic implications to be considered. First, it is useful to compare the private incentive to sue with the socially optimal suit. As a general matter, the plaintiff will sue when the cost of suit c_P is less than the expected benefit from suit. For simplicity, we assume that, if suit is brought, the plaintiff obtains as a judgment a certain amount h equal or near to harm suffered. Thus the plaintiff, will sue if and only if $c_P < h$. Note that if there is only a probability p of winning that amount, a risk neutral plaintiff will suit if and only if $c_P < ph$. If the plaintiff is risk averse he would be less likely to sue, because the cost of suit is in some way higher than c_P.

Generally, the private incentive to sue is fundamentally misaligned with the social optimal incentive to do so, given the social costs and benefits of suit. The divergence between social and private costs implies a socially excessive suit, as a plaintiff, in contemplating bringing suit, bears only his own costs and does not take into account the defendant's costs or the state's costs that suit will engender. On the other hand, there is a difference between the social and private benefits of suit that can either lead to a socially inadequate level of suit or reinforce the cost-related tendency toward excessive suit. Namely, the plaintiff does not recognize as a benefit to himself the social benefit of suit, its deterrent effect on the behavior of injurers. But he considers his private benefit, the gain he would obtain from prevailing.

Let us consider a little formalization. Suppose *Strict liability rule* is applied. As stated victims will sue if and only if $c_P<h$. Let q the probability of harm h if suit is not brought and q' the probability of harm when suit is brought. Further let c_D the defendant's litigation costs and c_S the state's cost. x are the precautionary expenditures the injurer is induced to make if there is suit. Of course, $x>0$ implies $q'<q$.
Hence, suit will socially worthwhile if and only if

$$q'(c_P + c_D + c_S + h) + x < qh$$

or

$$q'(c_P + c_D + c_S) < (q - q')h - x$$

In other words, suit is socially worthwhile if the expected litigation costs are less than the net deterrence benefits of suit. It is clear that the conditions of private and social incentive to suit are different. In particular, whether victims will sue does not depend on the costs of the defendant c_D and of the state c_S. Morevover from the social point of view the benefit is not only h what the vectim receives as damages award. This must be weighted with the difference of the two probabilities (q-q') and reduced by the precautionary expenses, x. Therefore, victims might sue when suit is not socially optimal and victims might not sue even when suit would be socially optimal.
Under *Negligence rule* the conclusions are qualitatively similar but the problem of excessive suit is less likely. If a victim would not sue a non-negligent injurer because he would know that he would lose, it becomes always socially optimal for victims to bring suit against negligent injurers, however great the legal costs of suit would be. In this case, injurers have the incentive to act non-negligently. Therefore, in this extreme case, there will be no suits for negligence and then no legal costs to be borne. However, given this strong deterrence there might be a problem of too few suits.
If instead, more realistically, a victim might sometimes bring suit against a non negligent injurer, then legal costs will incurred under *Negligence rule*, and the situation is qualitatively similar to that under *Strict liability rule*. There may be too many suits as well too few.
The implications of the social and private divergence are that the state intervention may be desirable, especially to correct a problem of excessive number of suits, for instance by taxing suit or barring it in some domain.

4.2 Settlement versus trial and the legal fees distribution rule

A settlement is a legally enforceable agreement in which the plaintiff agrees not to pursue his claim further. Let w be the award to the plaintiff, that we suppose to be the amount that would be won. This amount, resulted as an agreement among the lawyers of the parties, can be less than the harm suffered by the plaintiff as victim, h. Let us suppose the parties are risk-nutral. The plaintiff's expected gain from trial, given probability to win p_P, provides the minimum amount he would accept as a settlement: $p_P w$-c_P. Correspondently, the defendant's expected loss from trial, given probability to loose p_D, provides the maximum amount he would pay in settlement rather than go to trial: $p_D w$-c_D .
Consequently, the settlement is possible if and only if:

$$p_P w - c_P \le p_D w + c_D.$$

Therefore, there is a settlement range

$$\Psi \equiv [p_P w - c_P, p_D w + c_D],$$

which a settlement S_P, a payment to the plaintiff, must enter: $S_P \in \Psi$. On the other hand, we have that a settlement range is empty, $\Psi = \varnothing$, and a trial will occur when

$$(p_P - p_D)w > c_P + c_D$$

i.e. when the expected award in the plaintiff's opinion exceeds the expected award in defendant's opinion by more than the sum of litigation costs; the plaintiff is more optimistic than the defendant.

Risk aversion of the parties will generally increase the size of the settlement range and thus presumably will make settlement more likely. If the plaintiff is risk-averse, he will be willing to settle for less than $p_P w$-c_P and if the defendant is risk-averse, he will be willing to pay more than $p_D w$+c_D.
In deciding to settle or not, the rule of legal fees distribution is important. Since now we have considered the so-called *American* (and also continental) *rule*, according to which the parties bear their own legal costs. In the so-called *English rule* (a fee-shifting rule), instead, the loser pays the legal costs of both sides. Fee shifting may be one-way, favoring plaintiff P (defendant pay all if the plaintiff wins) or defendant (the plaintiff pay all if the defendant wins). Fee shifting may increase the chance of trial given the increase of differences in litigants' estimates of expected gains and loss from trial.
Indeed, with English rule, the expected gain by the plaintiff is

$$p_P w-(1-p_P)(c_P+c_D),$$

while the expected loss by the defendant is

$$p_D w+p_D(c_P+c_D).$$

Hence, the English rule tends to reduce the settlement range, which now is

$$\Psi_E \equiv [p_P w-(1-p_P)(c_P+c_D), p_D w+p_D(c_P+c_D)] \subset \Psi,$$

and then it tends to increase the chance of trial if parties are both optimistic about winning and passing on the legal expenses to the other. But it tends also to reduce the chance of trial, especially when the parties are risk-averse, because tends to raise the amounts the parties will spend at trial, as a party's expenditure will only be a cost with a probability rather than with certainty.
In addition, the lawyers as agents of the litigants may play an important role on the amount and distribution of the costs of the litigation. Clients and lawyers are in Principal-Agent relationships, so typical problems of information asymmetry arise. In particular, there is a standard moral hazard problem when the client cannot observe lawyers' effort and lack of legal expertise. To prevent or to reduce this a risk-sharing system, like a fee linked to lawyers' performance, is desirable. However, lawyers are risk-avers and pretend to be compensated at a hourly rate of time spent, without regard to legal outcomes. In any case, lawyer's activity is generally composed by repeated purchases, and then a meaningful incentive role could be playd by the need to improve the reputation within the profession.

4.3 Suit duration and the pathological demand for trials

Since now, we have considered the defendant as a passive player waiting for the decision to sue made by the plaintiff. Actually, he may play an active role by ex-ante deciding if to recognize the tort, paying the victim, or to accept the challange of a trial. Let us introduce two elements that are generally meanigfull for taking this decision: the expected duration of the trial by the defendant and the interest, he has to pay in case of defeat after the conclusion of the suit.
Let n the average time that a trial starts and concludes. Let i_l the "legal" interest rate to be applied for the compensation if, at the end, the plaintiff wins. This interest rate is fixed and specified for a long time by the law or by the court. Let instead r the market rate of interest. The defendant may decide to recognise the tort and pay immediately h or to accept the trial and wait for the result, which, if negative, could be an agreed award to the plaintiff w, which now we suppose to be equal to the harm h.
The defendant would prefer to agree to the suit if the amount he has to pay immediately is higher than the present value of the amount he could be obliged to pay at the end of the trial.

$$h > p_D \frac{h(1+i_l)^n + (ac_P + c_D)}{(1+r)^n}$$

or

$$h(1+r)^n > p_D h(1+i_l)^n + p_D(ac_P + c_D)$$

where $0 \le a \le 1$ is the share of the legal expenditures of the plaintiff that are attributed to the defendant in case of defeat. As extreme applications of fee shifting we have a=0 with the American rule and a=1, with the English rule.

The L.H.S. of the condition represents the opportunity cost, in terms of lost interests in the market, for recognizing the tort and paying immediatly h before the plaintiff sues. The cost depends on the to day market rate of interest because it corresponds to the interests the defendant has to pay now for borrowing from a bank the amount h. The R.H.S represents the opportunity cost of the suit for the defendant. It is an expected cost because it depends on the probability for him to loose.

Let with

$\Gamma(n) \equiv \frac{ac_P + c_D}{h(1+i_l)^n}$

denote, for the defendant, the ratio of the cost of carrying on the suit to the amount he has to pay in case of defeat. The index is clearly decreasing with the average duration of trials: $\Gamma'(n) < 0.$

Hence the defendant will accept the suit if

$$\left(\frac{1+r}{1+i_l}\right)^n \square (r - i_l)^n > p_D(1+\Gamma(n)).$$

Thus, the incentive to accept the challange of the trial increases with the difference between the two rates (r-i_l) and with the expected duration of trial n. Indeed, given i_l, an increase of r makes easier to satisfy the condition. Further, with $r > i_l$, an increase of n, increases the opportunity cost of the strategy "recognizing the tort" more than that of the strategy "accepting the suit".

Symmetrically, these parameters boost the defendant not to offer the plaintiff a settlement at a time n_s-n, because it is likely more convenient to go on and wait for the end of trial. With (r-i_l)>0, it is as the defendant would borrow from the plaintiff the amoutn h at a advantageous cost, given that he can gain the differential of the two interest rates. Whereas the civil laws and the legal institutions allow for sure that $r > i_l$ and are burdened by a high n, somewhat a "pathological" level of demand for trials arises.

4.4 The specificity of the administrative trial

In many countries, the relationships between the Public administration and private parties are regulated by the law following specific administrative proceedings, which is working as a non-cooperative game between a private agent and a bureaucrat in charge for carrying out the relationship. The legal discipline of this game tends to accompany the Public administration by specifying the ends it has to pursue, by limiting the sphere of its choices and decisions, by obliging it to legittimate its evaluations and to provide what it has to in the due time. However, it is possible that, at the end of the proceedings, this is not enough and that the private parties disagree and decide to suit against an administrative court (a "special" one) for protecting own rights and interests, they think have been damaged by an act of a public bureaucrat heading the public office.

By extending to this new context what we showed in the previous sections about ligation, we may say that a private will sue if and ony if the expected cost (the defense against the adminstrative court, g_P plus a fixed access fee, F) is lower than the benefit, b. The latter can be thought equal to the monetary value of the private interest damaged the public adminstration should indemnify, if loser. Generally, in the adminstrative suit, the *shifting rule* is applied, then the private must take into account

that, if loser, he will have to pay also the costs brought by the pubblic administration, g_A. Hence, if we denote with π_P the probability of winning by the private party,

$$V_P \equiv \pi_P b - (1-\pi_P)(g_P + g_A) - F > 0$$

is the condition of the private incentive to suing, as the expected value of the trial, V_P, is positive. This value, as well as the incentive, tends to decrease if, among the suit costs, there is a penalty for the so colled "reckless suit", in case of negative result of the suit itself for the private plaintiff.

As before, we have to ascertain the social value of the suit, given by the costs brought by the public adminstration defendant. First, there is the cost given by the expected expenditure to sustain the activity of the state lawyers, although these costs are possibly underestimated, as financed by fiscal revenues and not directly brought by the public official defendant. Second, there is the amount x of precautionary expenditures devoted to engage an effective administrative proceedings, such that to reduce the risk to lose in case of suit. It is an extra-cost respect to that one normally requested for completing the proceedings and it depends on the degree of risk aversion of the public adminstration. Note as this cost has a different nature respect the precautionary expenditure of a private injurer.

Now, let $\pi_A(x)$ the probability of the public adminstration to loose the trial. Clearly, it is decreasing with x: $\frac{d\pi_A(x)}{dx} < 0$. Consequently, the total cost for the public admnistration is

$$CT_A \equiv \pi_A(x)b + \pi_A(x)(g_P + g_A) + x - F.$$

To this we have to add the external social cost, not perceived by the public bureaucracts in the public office, but that is brought by the collectivity of taxpayers, C_S. Therefore, $CT_A + C_S$ has to be subtracted to the private value of the suit V_P, for getting the net social value, V_S, or

$$V_S \equiv V_P - CT_A - C_S = \pi_P b - (1-\pi_P)(g_P + g_A) - \pi_A(x)b - \pi_A(x)(g_P + g_A) - x - C_S.$$

Hence, the condition for social cost-benefit of the suit is

$$V_S \equiv (\pi_P - \pi_A(x))(b + g_P + g_A) - g_P - g_A - x - C_S > 0.$$

Therefore, the suit implies a *certain cost,* given by the legal expenditures $g_P + g_A$, the precautionary expenditure x, and the external social cost, C_S, plus an *expected cost*, due to the possible gain deriving by the divergence of the expectations, $(\pi_P - \pi_A(x))(b + g_P + g_A)$. Of course, also in this case, $(\pi_P - \pi_A)$ play the role of "optimism", i.e. the difference between the two probabilities. This is generally positive, as the public administration usually consider to loose with a probability lower than that one estimated by the private party, $\pi_P - \pi_A > 0$. However, if $\pi_P = \pi_A$, the social value is negative given the legal expenditures. Consequently, the suit may have a positive social value only with a high and justified optimism by the private parties, able to compensating the certain costs.

In any case, it is different w.r.t. the private value, thus the social value of the trial against the adminstrative court is different from the private one of the potential plaintiff. Actually, the strategy of the admnistrative proceedings is that to increase x in order to reduce the probability of loosing. This, on one hand, can feed the optimism, but, on the other, the social cost increases with x and then the social value remains always negative, although positive the private value. Therefore, we may have an excessive number of suits against the administrative courts.

In conclusion, the economic reasoning about the incentives to suit in front of an administrative court is analogous to that one carried out for the civil law. The main differences are twofold. On one hand, we have the specific behaviour of the defendant, a public bureaucrat in charge as responsible of the proceedings. On the other, we have the administrative discipline regulating and controlling the iter of the admnistrative acts production.

5. Law enforcement

5.1 Rationale for public enforcement

Among the reasons in favour of a public enforcement of law, there is that often victims do not know who caused harm and penalizing wrong doing is difficult. Society tends to rely instead on public investigation and prosecution. Further, private enforcement may induce wasteful effort devoted to finding violators. Public enforcement can efficiently developes a coordination of technologies at large scale. However, the main reason probably is that the state generally does not want to permit private parties to use force needed to gather information.
Also in this case there are two rules of liability: *Strict liability* (SL), the individual is definitively sanctioned, and *Fault-based liability* (FL), he is sanctioned only if his behavior fells below a fault standard. Once determined the liability we have the sanction that may be a monetary fine or prison term. There may be also a combination of the two sanctions.
Economics of law traditionally refers to the expected utility (EU) framework for a decision to commit a harmful action. According to this, a risk-neutral individual will commit the act if that would raise the expected utility taking into account the gain he would derive and the probability form and level of sanction.
Given the total welfare,

$$W = \sum_i ExpU^i ,$$

the enforcement authority maximizes W by choosing enforcement expenditures or equivalently a probability of detection (when it is fixed) and also the sanction. The optimum condition is $pf=h$, where p is probability of the sanction, f is the fine and h the benefit, which corresponds to the inflicted harm. Therefore, in this Benthamian context the optimal fine is given by $f^* = h/p$.
For many reasons, we have to take into account that there is a maximal fine, f_m, due for instance by the total wealth of an individual, so $f^* \le f_m$.
For defining an optimal deterrence rule, we have to consider that the increase probability of deterrence is costly, while increase the fine is not, so it should be $f^* = f_m$. However, the possibilty to do errors in enforcement can require trying to increase also the probability of deterrence. There are two types of errors: (i) an individual who should be be found liable might mistakenly not be found liable with probability $\hat{\alpha}_1$; (ii) an individual who should not be found liable might mistakenly be found liable with probability $\hat{\alpha}_2$.
An individual will commit the wrongful act when his net gain - i.e. the gain g, net of his expected fine if he does commit it - leaves him better off than paying the expected fine if he does not commit it, namely when

$$g - p(1-\hat{\alpha}_1)f > -p\hat{\alpha}_2 f$$

or when

$$g > p(1-\hat{\alpha}_1 - \hat{\alpha}_2)f .$$

The R.H.S. of the condition, i.e. the expected cost of liability, is declining in both the probabilities of errors. Both types of error reduce deterrence and then social welfare. The probability of deterrenced p should be higher to offset these effects, but it is costly in terms of organization of deterrence forces. If the individual is risk-averse, he will commit a harmful act if and only if his expected utility is raised by so doing, and in general he will not be equally deterred by different combinations of sanction and

probability with the same expected value. He will be more deterred the higher the magnitude of the potential sanction in the combination, with the expected sanction held constant.
To see this let propose this formalization. $u(.)$ is the utility of income function of a risk-averse person, y is the income, g is the gain from the act, p is the probability of a sanction f. The person's expected utility if he commits the act will be

$$ExpU = pu(y+g-f)+(1-p)u(y+g).$$

If p falls to kp, where $k<1$, and f rises to f/k, so that the expected sanction is still pf, the person's expected utility becomes

$$Exp\hat{U} = kpu(y+g-f/k)+(1-kp)u(y+g).$$

Differentiating $Exp\widehat{U}$ w.r.t. k yields

$$\frac{\partial Exp\hat{U}}{\partial k} = p\{(f/k)u'(y+g-f/k)-[u(y+g)-u(y+g-f/k)]\}.$$

Given the concavity of $u(.)$, this espression is positive as $u'(.)$ is decreasing.
Hence, the lower is k, the lower is expected utility, and therefore the greater is the deterrence. For instance, a person will be more deterred by the sanction of € 1000 with probability 20% than by the sanction of € 500 with probability 40%. Risk-avers party suffer disutility more than in proportion to increase in the magnitude of sanctions. Note that this means that elasticity in absolute value of $ExpU$ w.r.t. p is less than that w.r.t. f.
In the following sections, we investigate three forms of law violation and of the relative enforcement instruments: corruption, tax evasion and anticompetitive firms collusion.

5.2 Corruption

We talk about proper corruption between public officials and private agents when the latter ones have convinced the former ones to act in favour of them illegittimly. A robust empirical evidence show that, among the main causes and economic explanations of this kind of law violation, there are low GDP, low level of education and democracy. While there is a relationship between public sector dimension - in terms of high level of public expenditure, market regulation and public sector discretionality - and corruption. The existence of strong interest groups influencing public officials' activities by exchanging favors increases corruption. The phenomenon is also linked to the complexity of rules and high level of bureaucracy.
Talking of remedies to corruption, we have to recall, first, that in terms of economic theory, corruption generally increases with imperfect information. State and public officials are engaged in a Principal/Agent relationship, with moral hazard and adverse selection situations, therefore it could be useful to organize a system of economic incentives to reduce corruption, like high wages for high productivity. There are also political institutional remidies for corruption, starting with favouring political competition among parties to get the government, yardstick competition among politicians managing municipalities. In addition, favouring accountability and political responsibility toward citizens/voters/tax payers may contrastat corruption, as developing an independent information flaws and increasing transparency. Finally, an electoral system, able to properly select the preferred politicians, may help to reduce corruption.
In order to explain how an illigittimate coalition between a private (P) and a public official (U) may carried on, we describe a simple model. Let b the benefit of the agreement to P and m the bribe requested by U. The deterrence for the two agents is given by the probability to be discovered, α and to be obliged to pay a sanction for both: s_i. $i=P,U$.
With $\overline{m}$ we denote the maximal bribe P is available to pay:

$$\overline{m} = b - \alpha s_P,$$

i.e. the benefit less the expected cost of his sanction.

With $\widetilde{m}$ we denote the minimal amount *U* is available to accept:

$$\widetilde{m} = \alpha s_U ,$$

i.e. the expected cost of his sanction.
The corruption act will be carried on if and only if

$$\overline{m} > \widetilde{m} \text{ or if } b > \alpha(s_P + s_U).$$

The positive surplus of the coalition, $b - \alpha(s_P + s_U)$ can be, by a Nash bargaining outcome, devided fifty-fifty, thus the consequent equilibrium bribe is as follows

$$m^* = \frac{1}{2}(b - \alpha s_P - \alpha s_U).$$

The deterrence of the corruption is working if the expected total sanction, $\alpha(s_P + s_U)$, is higher than the benefit *b* to *P*. Hence, it must occur a trasparency of the actions of public administration, an effective system of controls, increasing the probability α and somewhat high total level of sanctions for both agents. The distribution of the sanction between *U* and *P* instead does not matter for the efficiency of the enforcement activity.

5.3 Tax Evasion

We need to remind some specific definition before discussing the incentive to evade and the systems to enforce the fiscal law. First, with "tax compliance" we mean the taxpayer effort and availability to pay for the due amount to Fiscal Agency. Second, with "tax gap" we refer to the difference between the potential tax base and the effective (declared) tax base. Third, "tax evasion" is the lost yield, which may be mesured by multiplying the average tax rate with the tax gap. Tax evasion is an illegittimate behaviour. "Tax avoidance" refers to all procedures, given the spaces provided by the fiscal law, to slip from the duty to pay taxes. In particular, "Tax expenditures" are the tax deductions allowed by the fiscal law. Tax avoidance is a legittimate behaviour, at least from the legal point of view. Finally, we have to recall that the tax-payer may be incentivated to produce less tax base (possibly shifting it somewhere) to escape from the excess burden of taxation and tax distortion.
The traditional model to analyse the incentive to evade is the so-called fiscal evasion portafolio choice. According to this, a taxpayer decides how much income to hide and then how much to evade by maximizing this expected utility:

$$\begin{aligned} \max_{(y_e)} ExpU(y_e) &= (1-\alpha)u(y_{NA}) + \alpha u(y_A) = \\ &= (1-\alpha)u[y - t(y - y_e)] + \\ &+ \alpha u[y - t(y - y_e) - ty_e - fty_e] = \\ &= (1-\alpha)u[y(1-t) + ty_e] + \alpha u[y(1-t) - fty_e] \end{aligned}$$

where *A* is "assesment" event with probability α, *NA* is "non assesment" event, with probability 1-α; *y* is gross income of the individual, *t* is income tax rate; y_e is the tax gap or hidden income and ty_e is the evasion; *f* is a proportional fine the tax-evader must pay, if uncovered. $u(y_A)$ and $u(y_{NA})$ are contingent utility functions. As we suppose risk-aversion, *u*(.) is a concave function, so $u' > 0, u'' < 0$.
The F.O.C. is as follows

$$(1-\alpha)u'(y_{NA})t - \alpha u'(y_A)ft = 0,$$

or

$$\frac{(1-\alpha)u'(y_{NA})}{\alpha u'(y_A)} \equiv SMS_{y_{NA}, y_A}(y_e^*) = f.$$

With risk aversion, $y_e = 0$ if $y_{NA} = y_A$, and then if $\frac{(1-\alpha)}{\alpha} = f$. Evasion is instead positive if the probability to be assessed is low w.r.t. the fine:

$$\frac{(1-\alpha)}{\alpha} > f; \text{ or } \alpha < \frac{1}{1+f}.$$

The amount of tax evaded ty_e^* depends negatively by the level and the efficacy of controls, i.e. the probability α, and by the fine, *f*. More controversial is the effect of tax rate *t*. An increase of it increases the net gain in terms of tax evaded, in NA event, but increases the amount of fine, in *A* event.The two effects tend to compensate each other. Moreover, an increase of *t*, decreases the net income and this may increase the risk-aversion of tax-payer effect, which may contain the propensity to evade. However, an increase of tax rate tends also to reduce the labour supply by boosting to look for no-taxed activities supplied in the "black market". By this way, the hidden income, y_e, tends to increase. Empirical research by econometric models and experimenatl economics exercises confirms that an increase of tax rate actually increases evasion.

However, what is the revenue effects of a fight against the evasion by means of controls and fines? Let the control activities given by *x* and the expected net revenue as follows

$$R = t(y - y_e) + \alpha(x)(1+f)ty_e - C(x)$$

where $\alpha(x), \alpha'(x) > 0$ is the probability of a succesfull assessment and also a measure of controls efficacy, and $C(x)$, $C'(x) > 0$, is the cost function for control and assesment activities. The effects of an increase of controls are

$$\frac{\partial R(.)}{\partial x} = t\frac{\partial(y - y_e)}{\partial \alpha}\alpha'(x) + [(1+f)ty_e + \alpha(x)(1+f)t\frac{\partial y_e}{\partial \alpha}]\alpha'(x) - C'(x)$$

And as $\frac{\partial(y-y_e)}{\partial \alpha} = -\frac{\partial(y_e)}{\partial \alpha} > 0$, we have

$$\frac{\partial R(.)}{\partial x} = [(1+f)ty_e - t(1-\alpha(x)-\alpha(x)f)\frac{\partial y_e}{\partial \alpha}]\alpha'(x) - C'(x)$$

We know that in order to have evasion, i.e. $y_e > 0$ it must be $(1-\alpha-\alpha f) > 0$. Hence, given that $\frac{\partial y_e}{\partial \alpha} < 0$, we have that

$$\frac{\partial R(.)}{\partial x} > 0 \text{ if } (1+f)ty_e\alpha'(x) \geq C'(x), \text{or } ty_e \geq \frac{C'(x)}{\alpha'(x)}\frac{1}{1+f}.$$

With a high evasion, the fight against it tends to be more convenient if it is efficient in increasing the probability of assessement, relatively to the marginal cost of it.

As far as an increase of the fine, we have a univoque result:

$$\frac{\partial R(.)}{\partial f} = \alpha t y_e - t(1-\alpha-\alpha f)\frac{\partial y_e}{\partial f} > 0,$$

hence, given that $\frac{\partial y_e}{\partial f} < 0$, an increase of the fine increases for sure the fiscal revenue.

Many considerations and results of this section may be changed, even reversed, if we take into account the possibility of an ethic negative judgement of evasion by an honest taxpayer. The utility function for taxpayer *i* should be changed as follows:

$$ExpU_i = ExpU_i(y_i) - \xi_i ty_{ei}$$

where ξ_i is a honesty indicator. With this utility function, with risk-aversion, we have that

$$y_{ei} > 0 \text{ if } \alpha < \frac{1}{1+f+\xi_i}.$$

Thus, the honesty acts as a supplementary personalized fine, reducing the hidden income of *i*.

5.4 Anticompetitive collusion

Competition and rivalrny have many social desirable effects, as far as the welfare of users and consumers is concerned. Therefore, anti-trust law enforcement should be devoted to limit the main forms of anti-competition practices and behaviours as cartels, joint ventures, mergers, vertical contracts, predatory prices. In this section, we deal with the rise of anticompetitive collusion practice. Collusive agreements may take different forms. Firms might agree on sales prices, allocate quota amog themselves, devide markets so that some firms decide not to be present in certain markets, in exchange for being the sole seller in others, or coordinate their behaviors along some other dimensions. Institutional arrangements to sustain collusion might range from very well organised cartel-like structure where a central secret office takes the main decisions, to situations where firms merely find some form of communication to sustain the agreement. In other situations the agreement can origin a merely "tacit collusion".
Collusive practices allow firms to exert market power they would not otherwise have and artificially restrict competititon and increase prices, thereby reducing welfare. Accordingly, they are proibited by any anti-trust law, and indeed in large part of the anti-trust Authorities' efforts is devoted to fighting such practices. For this reason, it is important to identify the main mechanisms behind collusion, in order to study the factors that facilitate it, and to explain which behavior should be treated as an infringement of the law and which should not.
A collusion is a situation that must be removed if it is explicit and sustained in time. Two elements must exist for collusion to arise and to be stable: *detection* and *punishement*. First, the partecipants of a cartel must be able to detect in a timely way that a deviation has occorred, for instance a firm setting a lower price or producing a higher output than the collusive levels agreed upon. Second, identifying the deviation is not enough: there must also be a punishment, which might take the form of rivals producing much higher quantities, or selling at much lower prices, in the periods after the deviation, thus depressing the profit of the deviator.
Another important issue is that of *coordination*. Firms that are tacitley colluding might arrive at a fully collusive price, but this is just one of the many possible equilibrium outcomes. Under explicite collusion, instead, firms can talk to each other and coordinate on their preferred equilibrium without having to experiment with the markets, which is costly. Furthermore, if there are some shocks which modify market conditions, communication will allow the firms to change to a new collusive price without the risk of triggering a period of punishment.
Let simply model the conditions for collusion to arise [Motta (2004, p.160)]. Let π_i^c and V_i^c respectively be the current profits and the present discounted value of profits that firm *i* receives if it chooses a certain collusive action, given that all firms also collude. Let π_i^d the current profit of firm *i* if it deviates when all other firms take the collusive action and V_i^p the present discounted value of firm's profits in punishement phase, that is in all periods that follow the deviation period.
Denote with $\delta \in (0,1)$ the discount factor, assumed identical for all firms in the industry. The discounted factor can be expressed as $\delta = \frac{1}{1+r}$ where *r* is the interest rate between two periods of time, and therefore the value in today's terms of 1 € that one receives in the following period. Thus, $\delta \to 0,$ corresponds to the case where $r \to \infty$: one € earned in the future is not worth anything in today's terms; at the other extree, $\delta \to 1$ corresponds to the case where $r \to 0$: one € earned in any future period has the equal value as 1 € earned today.
Collusion can arise only if each firm will prefer to play the collusive action rather than deviate from it and be punished, therefore, the following *incentive contraints* (IC) must hold, one for each firm in

the agreement:

$$\pi_i^c + \delta V_i^c \geq \pi_i^d + \delta V_i^p, i = 1,..n.$$

Clearly, the lower the deviation profit one makes relative to the collusive profit, and the lower the profit in the punishement phase, the more likely will be sustained. The harsher the punishment the stronger the deterrent to cheating on the collusive agreement.
The *n* incentive contraints can also be written as

$$\pi_i^d - \pi_i^c \leq \delta(V_i^c - V_i^p), i = 1,...n,$$

which states that the gain from deviating obtained today must be lower than the losses from deviating from the collusive strategy, incurred from tomorrow onwards. Again, this condition must be satisfied for all firms, otherwise one or more deviations will occur and collusion cannot sustained. Finally, another way to express the same incentive constraint is as follows:

$$\delta \geq \frac{\pi_i^d - \pi_i^c}{V_i^c - V_i^p} \equiv \overline{\delta_i}, i = 1,..n.$$

Collusion arises at equilibrium only if the discount factor is large enough, i.e., if it is larger than a certian "critical discount factor", $\overline{\delta_i}$. Only if firms are patient enough will the collusive agreement be sustained. Indeed, if the discount factor is very low, firms do not give importance to what will happen in the future, and they will prefer cheating to reap all the benefit they can today. Hence, the collusion will not arise. If it is not the case, for instance because there is a high long-term interest rate, the anti-trust law should interven removing the cartel and apply a sanction for the collusive firms.
Although, in theory, the identification of the mechanisms, through which collusion acts, and of the factors that faciltate it, is somewhat clear, its practical implications for legal purposes are less straightforward. As, in economic terms, collusion is a situation where prices are "high enough", one could think that to verify the existence of collusion in the legal sense, i.e. of anti-competitive behaviuor, one has to analyse price data, in a given industry, and infer if they are above some threshold levels above which they should be considered collusive. However, for several reasons it would be very difficult in practice for the Antitrust authority to look at market outcomes to decide whether there has been an infringement of anti-trust law.
First, in many circumstances, price data might be not available, and when they are, they might refer to list prices rather than effective prices. Second, even if reliable data existed, there would be probably be disagreement about the meaning of monopoly price in an industry. Sellers might have very different views of what that price would be, and an outside observer could have yet different perceptions. It is well known that estimates of costs differ widely, sometimes even within the management of the same firm. Third, suppose there is an agreement on what monopoly price in the industruy would be: how close to the theoretical monopoly price should sales prices be for them to be judged "too high" and therefore collusive? Fourth, the very principle that firms could be convicted solely because they charge "too high" prices is a dangerous one, and it might open the way for anti-trust interventions whenever firms are successful enough to find consumers willing to pay high prices for their products.
Rather than looking at the *level* of prices in a industry, one might then be tempted to infer the existence of of collusion, i.e. an infringement of law, by analysing the *evolution* of industry prices over time. Courts and antitrust authorities have sometimes been tempted to infer the existence of a collusive illegal behaviour from the fact that sellers charge similar prices over time, the so-called "parallelism", or "conscious parallelism". Nevertheless, to observe that prices move in a similar way is not enough to establishe that firms are guilty of collusion. Common exogenous shocks such that as the increase

in input prices of all the suppliers, or an increase of inflation, or an increase in property prices would probably lead all the sellers to increase price proportionally, without implying that they are colluding. Further, the collusive outcome might arise without firms agreeing or communicating to coordinate their behaviour.

Despite these difficulties, antitrust anticollusion interventions are very frequent in industrialsed countries. Even in the absence of hard evidence, courts are often able to prove infringement of the law by second-guessing the firms' intentions and their motivations.

Chapter 3

Public versus private ownership of public services enterprises

1. The application of the theory of incomplete contracts to public services

In this chapter, we propose three specific topics. First, we consider the issue of delegating the running of a public facility, like a hospital or a prison, to a private manager instead to a public official. Hence, we look at the conditions for establishing which one between the private and the public ownership of an enterprise providing a public service is more socially desirable. Second, we analyse the choice of separating ("unbundling") or combining ("bundling") the activities of building an infrastructure and managing the service using the latter. Third, we discuss the governmental choice of the means for funding a Private-Public Partnership contract, taking into account the distinction between "hot" versus "cold" works. The first ones are public investment assuring with the revenues from tariffs a positive cash flow, while the second ones can generate only a partial cash flow.
All these arguments are applications of the theory of incomplete contracts, treated in previous chapters, to the organization of production and provision of public services. Therefore, it is useful to introduce the argument of the chapter by doing a parallel between the theory of firm, examinated also in section 2.3 of the previuos chapter, and the theory of privatization versus nationalization of economic activities. In the former, two firms A and B have a long-term relationship; one is a buyer and the other the supplier. Thus there are two cases: In the first one A and B sign a arms-length contract, still remaining independent; in the second one, the two firms go toward a merging, with a vertical integration: $A+B$. The theory of the firm analyses the conditions under which, as clarified with the example by Aghion and Holden (2011), the second solution is better off. In the theory of privatization, we consider the government, G, and P, a private firm supplying to G a service as input for public goods production.
Also in this framework, we have two cases: G and P sign a contract (privatization) or G directly buys P (nationalization), hence $G+P$ becomes a unique entity. In both cases, the institutional strategies are devoted to regulate a firm, B or P, for maximizing the benefit of the other A or G, thus both study the role plaied by the ownership. In property right model of vertical integration, ownership serves to elicit appropriate ex-ante investment, particularly in human capital. When A buys B, it has more residual control rights. Consequently, it has a greater bargaining power and earns a greater return, hence it can invest more. B's incentive to invest, instead, falls, once it becomes simply a section of A. In the nationalization versus privatization context, in deciding if G should buy P, it is necessary to consider, as a benefit, the fact that some bureaucrat, now in charge, will invest more, and, as a cost, the fact that the manager, now an employee, will invest less. Symmetrically, if the decision is to sell P.
In the last section of the chapter we are going to deal formally with the *problem of Pricing*, that is the choice of the optimal price sale of a public enterprise.

2. Incomplete contracts and the nature of enterprises ownership

Hart et al. (1997) model discusses the issue of ownership by considering the problem of managing a public facility, like a prison, a school or a hospital. Therefore, we have two cases. In the first one, G owns the facility and employes, with a labor contract, a public official, as a manager, who receives a salary P_0. In the second one, G proposes a delegation contract to a private firm for providing the service by the facility, now paying P_0 as a price.
We suppose that the manager, public or private one, can make two kinds of investment (both non verifiable): one raising quality, i, and the other reducing costs, but also quality, e. The main idea of

A. Petretto, *Economics of Institutions and Law*, ISBN (print) 978-88-6453-579-1, ISBN (onine) 978-88-6453-580-7, CC BY 4.0, 2017 Firenze University Press

the model is that a public employee has little incentive to engage in either activity, because the hold up by *G*, who appropiates all the benefits of the innovations. Private ownership encourages, instead, both kinds of activity, because the manager now is less subject to hold up by the owner. In order to find the solution to the issue private vs. public ownership, we have to verify, in case of private owenership, which of the two effects, due to "good news" (more *i*) and to "bad news" (more *e*), is more important.

Let us formalize benefits and costs of the game. We consider a good with two specifications: the "basic" one, defined ex-ante, in the contract, and the "modified" one, defined at a renegotiation stage, with contigencies realized.

Therefore, the timing is as follows. At time 0, the government and the manager write a contract on a "basic good" delivery and choose the ownership structure. At time 1/2, the manager chooses *i* and *e*. Each innovation leads to a change in quality but initial contract is vague (incomplete), so neither violates it. At time 1, in case of incomplete contract, the agent, having the right, after investment realization, proposes a renegotiation of the contract in order to transform the basic good in the modified good.

The benefit of the service (modified good) is given by

$$B = B_0 - b(e) + \beta(i),$$

where $b(e), b'(e) > 0$, $\beta(i), \beta'(i) > 0,$ are respectively the cost of *e* and the benefit of *i* in terms of quality.

The operative cost of the service is given by

$$C = C_0 - c(e),$$

where $c(e), c'(e) > 0,$ is the cost-containment effect of investment *e*.

Consequently, the manager' overall costs are given by $C+i+e$.

Without renegotiation, if the facility is private, cleraly $e>0$ and $i=0$. If, instead, it is public, $e>0$, and $i>0$, but *G* can realize only a fraction (1-λ) of the net social gain of the two innovations $[-b(e) + c(e) + \beta(i) - e - i]$, beeing λ a measure of weekness of the incentives of *G* employee.

It is useful to derive the benchmark framework, i.e. the First best social optimum, reacheable with a complete contract, where both ownership types are indifferent. We have simply to maximize the total net surplus of trading relationship between the government and the manager, finding the optimal values of $e^*>0$ and $i^*>0$. Formally,

$$\max_{(i,e)}[-b(e) + c(e) + \beta(i) - e - i]$$

implies, as F.O.C.s

$$-b'(e^*) + c'(e^*) = 1;$$

$$\beta'(i^*) = 1$$

The net social marginal benefit of both the investments are equal to the marginal cost of 1 €.

Now, let us look at the incomplete contract equilibrium under private ownership. In this case, the manager owns the facility and a renegotiation takes place on quality innovation, *i*, with split 50:50 of the gains $\beta(i)$; P_0 is the reward to the manager. The latter can reduce costs without government's approval, as the cost $b(e)$ is such only for the government. Hence the pay-off fuction of the government is

$$U_G = B_0 - P_0 + (1/2)\beta(i) - b(e)$$

and the pay-off funcion of the manager is

$$U_M = P_0 - C_0 + (1/2)\beta(i) + c(e) - e - i.$$

The manager chooses $\tilde{e}$ and $\tilde{i}$ as follows

$$\max_{(i,e)}\left[(1/2)\beta(i) + c(e) - e - i\right].$$

Then the F.O.C.s are:

$$c'(\tilde{e}) = 1$$

$$(1/2)\beta'(\tilde{i}) = 1$$

The total surplus of the contract is

$$\tilde{S} = \tilde{U}_G + \tilde{U}_M = B_0 - C_0 - b(\tilde{e}) + c(\tilde{e}) + \beta(\tilde{i}) - \tilde{e} - \tilde{i}.$$

Let us now consider the equilibrium under public ownership. The government owns the facility and a renegotiation takes place over a share λ of both innovations, *i* and *e,* that the government cannot appropriate, $\lambda[-b(e) + c(e) + \beta(i)]$, with a split 50:50 of the gains. The pay-off functions of the government and the manager/bureacrat now are

$$U_G = B_0 - P_0 + (1 - \lambda/2)[-b(e) + c(e) + \beta(i)]$$

$$U_M = P_0 - C_0 + (\lambda/2)[-b(e) + c(e) + \beta(i)] - e - i.$$

The public manager chooses $\hat{e}$ and $\hat{i}$ as follows

$$\max_{(i,e)} \{(\lambda/2)[-b(e) + c(e) + \beta(i)] - e - i\}.$$

Thus the F.O.C.s are

$$(\lambda/2)(-b'(\hat{e}) + c(\hat{e})) = 1$$

$$(\lambda/2)(-b'(\hat{i}) = 1$$

And the total surplus now is

$$\hat{S} = \hat{U}_G + \hat{U}_M = B_0 - C_0 - b(\hat{e}) + c(\hat{e}) + \beta(\hat{i}) - \hat{e} - \hat{i}.$$

Now it is interesting to compare the conditions on *e* and *i* level of the incomplete contracts and also w.r.t. optimal conditions of the complete contract benchmark. Looking at the F.O.Cs it is easy to verify that it turns out as follows:

$$\hat{e} < e^* < \tilde{e}$$

$$\hat{i} < \tilde{i} < i^*.$$

In other words, public owenership invests in both innovations less than private owenership; with respect to the First best, private owenership over invests in cost-containement and public owenership under invest in quality enhancement.

The optimal ownership structure is that one with the largest total surplus; P_0 is only relevant for the surplus division. Renegotiation under symmetric information ensures that all structures yield an ex-post efficient outcome. The only difference between the ownership structures concerns the choice, at time 1/2, of the ex-ante investments *e* and *i*.

Private ownership leads to an excessively strong incentive to engage in cost reduction and to a moderate incentive to engage in quality improvement. Public ownership removes the excessive tendency to engage in cost reduction but reduces also the incentive to quality improvement. The choice depends on which distortion is less damaging.

3. Public-private partnership and Project financing

The issue we now introduce is looking for answering questions as follows: how to build a bridge (or a hospital)? By a separate contract by which a public body, like a municipality, delegates to a private firm just the task of building the infrastructure or by a contract specifying a true partnership between the public body and the firm (Private-public partneship (P-P-P) approach)? More specifically, in the latter contract, often namend as *Project--financing* - or *bundling* - the builder is also the service provider, an operator, while in the former - or *unbundling* - is not, and the public body remains the operator.

As in the previous section, the question is important only if we face incomplete contracts, as with complete contracts the two alternatives are indifferent.

Hart (2003) models this situation by a sequential game, where, in the first stage, there is the action of building the infrastructure and, in the second stage, the action of operating the service. After having

signed a contract referring to one of the two typologies, the private firm can do two types innovations, or investments. The first one, *i*, is defined as *productive investment* because it increases both the quality of the infrastructure and contains the operative costs of producing the service. Note as here, contrary to the case of previous section, *i* has a double effect. The second innovation, *e*, is named as *unproductive investment* because it reduces the operative costs but also damaging the quality of the service.

Let the benefit function from service of the facility given as follows:

$$B = B_0 + \beta(i) - b(e)$$

and the operative cost of the service[10]:

$$C = C_0 - \gamma(i) - c(e)..$$

The First best, complete contract situation, is on picking up $i \geq 0, e \geq 0$ in order to

$$\max_{(i,e)} \quad B - C - i - e =$$
$$= B_0 + \beta(i) - b(e) - C_0 + \gamma(i) + c(e) - i - e.$$

The F.O.C.s are:

$$\beta'(i^*) + \gamma(i^*) = 1,$$
$$c'(e^*) - b'(e^*) \leq 1$$

The first condition implies that at the optimum $i^*>0$, i.e. the social productive investment is positive. The second condition, instead, implies a corner solution where $e^*=0$. The unproductive investment is zero at the optimum as we suppose that the condition is satisfied at the origin, $c'(0) - b'(0) \leq 1,$ so there is no necessity to increase *e* over 0. At the First best, all we need is only a productive investment, being able to both increase quality and reduce costs.

Now let us consider the case of a separated contract to build, or "unbundling", at a fixed remuneration P_0. The builder builds at the cheapest cost the infrastructure, while staying within the terms of the

contract. Note that in the contract it is $i=e=0$. Innovations, if any, are after the contract.

However, at time 0, the builder chooses *i* and *e* to maximize (P_0–i-e), then $\hat{i} = \hat{e} = 0$. No innovations are developed, as the builder cannot appropriate in some way the benefits of them.

Let us now consider the P-P-P, bundling, case. At time 0, the builder, going to became also the operator, chooses *i* and *e* as follows:

$$\max_{(i,e)} \left[P_0 - C_0 + \gamma(i) + c(e) - i - e \right].$$

The F.O.Cs are

$$\gamma'(\bar{i}) = 1$$
$$c'(\bar{e}) = 1.$$

Notice as the benefit function does not enter payoff function of the builder-operator. Therefore, we have

$$\bar{i} > \hat{i} = 0 \ ; \bar{i} < i^*$$
$$\bar{e} > e^* = \hat{e} = 0$$

The conclusion is that in the unbundling case, the builder is not boosted to invest in either investments. He under-invests w.r.t. the benchmark because he cannot internalize quality improvements and operative costs reductions. Therefore, the unbundling contract is carried out only if the quality of the

[10] The benefit and cost functions satisfy the usual property of concavity.

building can be ex-ante well specified and ex-post checked and verified by the public administration with low costs.
In P-P-P bundling case, the builder-operator can internalize the costs reduction by both investments, and does not internalize neither quality improvement or quality reduction, so, w.r.t. the benchmark, he over-invests in *e* and under-invests in *i*. This contract is carried out only if the quality of the service can be well specified in the initial statement, in the sense that there are good performance measures which can be used to reward or to penalize the service provider. Instead, it is the quality of the building to be not well specified.

4. The governmental choice of funding means of a public-private partnership contract

4.1 The choice between tariffs and taxes

Let us suppose that a local government wants to build an infrastructure and to transfer the public service provision to the private enterprise itself through a P-P-P bundling contract, and let examine the way of financing it. The main question refers to the choice on the composition between users' fees and a public subsidy, financed by an earmarked tax [Grazzini and Petretto (2012)].
Let $v>0$ denote the discounted private marginal willingness to pay (MWP) for the project's services, i.e. the value the consumers attribute to the investment and also a proxy of the level of demand. Demand uncertainty is summarized by a probability density function over v, $f(v)$, with c.d.f. $F(v) = \int f(v)dv$. This density is bounded from below by v_{min} and from above by v_{max}.
Let $\Pi(v)$ denote producer (builder and concessionaire) surplus in state v, which, given an exogenous and certain initial cost I (more precisely the present value of operatives and investments costs, known *a priori*) is given as follows:

$$\Pi(v) = R(v) + T(v) - I,$$

where $R(v)$ denotes the present value of user fees revenues, or tariffs, collected by the concessionaire, in each state v, and $T(v)$ denotes the present value of the subsidy it receives. Further, $T(v)$ is financed by taxation and takes the form of a cash payment over-time, contingent on v, to supplement revenue from the project under a Build-Operate-and-Transfer contract, with a minimum revenue guarantee.
Since the concessionaire receives $R(v)$ in state v, the local government receives the difference between the present value of MWP and the present value of revenues of the concessionaire: v-$R(v)$, with $0 \le R(v) \le v$. If the term of the concession is finite and $v - R(v) > 0,$ these funds are used to reduce distortionary taxation elsewhere in the economy. If the term of the concession is infinite, $v=R(v)$.
Since there is a (hard) government budget constraint, 1€ of the repayment is worth $1+\lambda$ (the marginal cost of public funds), while 1 € of taxation costs $(1+\lambda)(1+\zeta).$ The parameters λ and ζ represent two types of distortions. The first one is the cost of getting money from private taxpayers, and the second one captures the administrative costs of the government agency managing public service. $\zeta>0$ means that some of the resources from government to the concessionaire are wasted in the process, for agency problems, or for collecting procedures within the local public organization. The main problem is the possible diversion of funds raised by the earmarked tax to other expenditures purposes outside the project.
The consumers' surplus of the activity of the project, $S(v)$, is given by the difference between users' MWP in state v, and the total amount transferred to the concessionaire, but plus the reduction in distortionary taxes due to the increase in the revenue collected by the government at the end of the

concession (the final repayment). Therefore, the consumers' surplus obtains as

$$S(v) = v - [R(v) + (1+\lambda)(1+\zeta)T(v)] + \lambda[v - R(v)] =$$
$$= (1+\lambda)[v - R(v)] - (1+\lambda)(1+\zeta)T(v)$$

The optimal values of $R(v)$ and $T(v)$ for a benevolent local government are given by the maximization of the social surplus subject to the concessionaire participation constraint, otherwise it is not available to sign the P-P-P contract. Thus, its problem is

$$\max_{R(v),T(v)} \int \left[S(v) + \varepsilon\Pi(v)\right] f(v)dv$$
$$s.t. \quad \int \Pi(v) f(v)dv \geq \bar{u}, \quad 0 \leq R(v) \leq v, \quad T(v) \geq 0,$$

where $0 \leq \varepsilon \leq 1$ the weight that the politician gives to producer surplus in its pay-off function, and $\bar{u}$ denotes the opportunity cost of the concessionaire. By substituting the term for $S(v)$ and $\Pi(v)$,[11] the above program may be re-written as the minimization of net social cost due to the means of funding the project:

$$\min_{R(v),T(v)} \int \left\{(1+\lambda-\varepsilon)R(v) + [(1+\lambda)(1+\zeta)-\varepsilon]T(v)\right\} f(v)dv.$$

According to the solution of the above problem, it turns out that the terms of the P-P-P contract depend on outside parameters such as ε, λ and ζ.

From the minimization problem, it is easy to check that, if $\zeta>0$, user fees are a more efficient instrument for compensating the concessionaire than the subsidy. The cost to society of 1 € of user fees is $1+\lambda-\varepsilon$, while a subsidy costs $(1+\lambda)(1+\zeta)$. However, if the project's social value exceeds I, and user fees revenue is insufficient to compensate the concessionaire, in low demand states, the subsidy becomes beneficial as an insurance repayment, and $\zeta>0$ determines the structure of this optimal risk-sharing contract. When $\zeta>0$, the trade-off faced by the local government is the following. On the one hand, it would like to utilize user fees revenues as much as possible to compensate the concessionaire in order to avoid paying distortionary subsidies. On the other hand, if using only user fees, the concessionaire may run an excessive risk. Thus, an efficient contract should insure the concessionaire against low demand states through subsidies.

4.2 Hot versus cold works

It can be shown that the optimal contract is characterized by a minimum revenue guarantee, m, and a revenue cap, M, with $m<M$. Thus, there are projects such that $M<v$, where the concessionaire collects M in present discounted user fees, while the government collects the remaining v-M. No subsidies are paid and the end of the concession is finite. These projects are called *hot works*.

Then, there are projects such that $m \leq v \leq M$, the concession lasts indefinitely and no subsidies are paid. Indeed, the total revenue accrued to the concessionaire in present value is equal to v, and the government budget is unaffected by the concession. These projects are in some sense intermediate.

Finally, there are projects with $v<m$, the concession lasts indefinitely, and the government grants a subsidy of m-v to the concessionaire.

These are called *cold works*. In these cases, $v_{max} < I$ and then $m=I$, because with $m>I$ the

[11] Dropping all the variables not depending on government's choice.

concessionaire participation constraint holds with slack, and with $m<I$ it cannot be satisfied. Thus, the optimal contract subsidizes the concessionaire in all demand states to ensure that total revenue is equal to the cost of the project.
Then,

$$R(v)+T(v)=I \text{ for all } v,$$

and the government pays a subsidy equal to I-v in each state. This is the present value of earmarked tax for state v. The expected value of the earmarked tax is given by the difference between the fixed cost and the expected present value of MWP:

$$Exp[T(v)] = I - \int v f(v) dv.$$

5. Privatization rationale and sale price

5.1 The social desirability of privatization

The sale on the market of the shares of a public enterprise should occur if and only if a social welfare gain is expected. An approximation of the latter, in the state v, i.e. facing a given level of user's demand, can be represented as follows:

$$\Delta W(v) = \Delta\Phi(v) + \Delta\Psi(v)$$

where $\Delta\Phi(v)$ represents the firm's social value change from public to private ownership and $\Delta\Psi(v)$ represents the social value of revenue from sale. The former is given by the espression

$$\Delta\Phi(v) \equiv VS_p(v) - VS_g(v),$$

where $VSp(v)$ is the value for the collectivity at time t_0, of the firm, when under the control of private shareholders and $VSg(v)$ is the corresponding value of the firm, when still remaining public. Both notions of value come from discounting the flow of expected future social -- direct and indirect -- benefits and costs computed at the shadow prices, as analysed in the the previous section. More specifically, both are given by the weighted sum of consumer and producer surpluses in the two cases. Formally, it would be as follows

$$VS_i(v) \equiv S_i(v) + \varepsilon\Pi_i(v); i = p, g.$$

$\Delta\Phi(v)$ can be positive or negative. If the ownership change is simply a shift from a public monopoly to an unregulated private one, the aim of a public enterprise to get allocative efficient performances, rather simply to maximize profits, will result in $\Delta\Phi(v) < 0$. Consequently, in order to be convenient and to make money, the sale price must be high. If the sale is preceded by a large-scale process of liberalization and competition promotion together with some independent regulation mechanism, the gap between the two values is reduced up to the change of the sign. In this case, with a positive and high $\Delta\Phi(v)$, the convenience of the sale may even occur with also a very low sale price ("underpricing").
The social value of revenue from sale is given by

$$\Delta\Psi(v) \equiv -\varepsilon\Delta\Pi(v) + (1+\lambda)\Delta R(v),$$

The first term of tha R.H.S. is the reduction of profit of the enterprise (producer surplus) suffered by the private owners and is given by P, the sale price, going from the private to the public coffer,

$-\Delta\Pi(v) = P,$ weigheted with the distributive term $\varepsilon, 0 \le \varepsilon \le 1$. The second term is the increase of public coffer from the sale, $\Delta R(v) = P,$ evaluated at the shadow price given by the marginal cost of public funds (1+λ). As a consequence, the social welfare change of the privatization is as follows

$$\Delta W(v) = (VS_p(v) - VS_g(v)) + (1 + \lambda - \varepsilon)P.$$

Therefore, we may summarize the rule of social desirability: *The sale of a public asset is convenient if the effective price P is greater than P_g, the minimum price acceptable by the state, i.e. the price at which the society is indifferent to the transaction*.
Formally, we have

$$\Delta W(v) > 0 \text{ if } P > \frac{VS_g(v) - VS_p(v)}{(1 + \lambda - \varepsilon)} \equiv Pg.$$

5.2 The socially beneficial sale price

Note that, from the previous rule, social welfare always increases, ceteris paribus, with the sale price as

$$\frac{\partial \Delta W(v)}{\partial P} = 1 + \lambda - \varepsilon > 0$$

i.e. one more € from privatization has a value equal to the amount by which the marginal cost of public funds is higher than 1 minus the distributive term ε.
Hence, the state will try to extract from the potential buyers a high price, possibly the highest one, which is the maximum price they are willing to pay, P^*. This is given by the amount at which private buyers evaluate this firm and it is equal to $VP(v)$, the value of the cash-flow net of the reward to the manager. This value, beeing obtained by a purely private calculus, differs from both the firm's values $VS_p(v)$ and $VS_g(v)$, which consider also social evaluations.
Thus the maximum welfare increase from the sale to a potential buyer is:

$$\Delta W^*(v) = VS_p(v) - VS_g(v) + (1 + \lambda - \varepsilon)VP(v)$$

and thus, by substituting, the maximum welfare increase, measured in terms of fiscal revenue, is given by the difference between the maximum and the minimum price:

$$\frac{\Delta W^*(v)}{(1 + \lambda - \varepsilon)} = P^* - Pg.$$

In any case, the effective price will fall within an interval defined by the outcome of transaction and the chosen techniques of sale (*beneficial sale price interval*):

$Pg \le P \le P^*$.

Note that, despite it being advantageous to get the highest price from a given buyer, it does not go with saying that the firm should be always transferred to the buyer offering the highest price, as other external elements of evaluation should be considered.
If the private buyer is chosen by a multi-dimensional auction selection, the quality standard of the service, the volume of planned investments and the employment policy can be evaluated together

with the price of sale. Therefore, the firm can go to a buyer ensuring an ex-post higher social value, $VS_p(v)$, and thus a lower P_g. This would imply a reduction of the lower limit of the beneficial sale price interval.

Further, in general, a higher $VS_p(v)$ is correlated to a lower $VP(v)$, and thus a lower P^*. The latter will be lower with the liberalization of the public service market and with hard competition, dissipating all expected extra-profits. Hence, the upper limit of the interval can also decrease.

Chapter 4

Public versus private organisation in health care economics

1. Health care specificity and government intervention

1.1 Health care services and individual well-being

Health services, like general practitioner and specialised medical examinations, diagnostic tests, admissions to hospital, medicines, are devoted to treat a state of sickness, aiming at a adequate level of individual health. According to the classical notion coined by Amartya Sen, health is a functioning, extensively influencing, as liberty, nourishment and basic education, the individual level of well-being. However, the individual health level does not derive only from the consumption of health services, but mainly from the efficiency of a peculiar household production process, depending on his/her personal life style. The latter process, on the other hand, works in a social context given by several meaningful variables, as the environmental conditions, the society level of knowledge and culture, the distribution of wealth, etc.
Health services are private goods, since they are both rival and excludable, but are not direct consumption commodities, as a loaf of bread or a glass of wine. They are, instead, intermediate goods, working as inputs in the household productive process determining the level of health functioning .
In formal terms, let define as $\mathbf{m}_i = (m_{ik}, k = 1..M)$ the vector of M health services to treat a status of sickness of individual i, indicated by s_i, a measure of need. Then, let define with H_i the level of health reached by the individual, a measure of human capital stock. Consequently, the individual i, i=1,..n well-being may be represented as follows:

$$U_i = U_i(x_i, H_i)$$

$$H_i = h_i(s_i, \boldsymbol{m_i})$$

According to first function, the level of utility achieved by individual i depends on the direct consumption of a private composite commodity x_i, the numeraire, and by the functioning health H_i. According to the second function, i health functioning depends on the initial state of health, s_i, combined with the use of M health services, as described by the production function $h_i(.)$, embedding as well the effects of environmental and social conditions and the life style[12]. Any health service j has a positive, not-increasing, marginal productivity,

$$\frac{\partial h^i(.)}{\partial m_{ij}} \equiv h^i_j(.) > 0, \frac{\partial^2 h^i(.)}{\partial m^2_{ij}} \equiv h^i_{jj}(.) \leq 0, j = 1,...M$$

while the sign of cross-derivative h^i_{jl} depends instead on the relationship of complementarity or substitutability between services j and l . Notice as the marginal productivity formalizes the notion of appropriateness of j on health care of individual i, which is specified by the physician along the medical protocols. Given the "personalised" price, P^i_j, consumer pays out-of-pocket, the amount

[12] As it is clear, in this simplified formalization there is only one period, thus there is no save and no accumulation. Health H_i is obtained just at time t_0.

A. Petretto, *Economics of Institutions and Law*, ISBN (print) 978-88-6453-579-1, ISBN (onine) 978-88-6453-580-7, CC BY 4.0, 2017 Firenze University Press

$P_j^i m_{ij}$ for each health service he purchases, j=1,..M.
In this particular context, the consumer equilibrium is given by solving the following program

$$\max_{(m_{ij})} \quad U_i(x_i, h_i(s_i, \mathbf{m}_i))$$
$$\text{s.t.} \quad x_i + \mathbf{P}^i \mathbf{m}_i = Y_i$$

The F.O.Cs are

$$MRS^i_{j,x} \equiv \frac{U^i_H h^i_j}{U^i_x} \equiv MRS^i_{H,x} h^i_j = P^i_j, j = 1,..M.$$

According to this condition, the marginal benefit of care throughout the service j, $MRS^i_{j,x}$, is a measure of willingness to pay, given by the marginal utility of health (in terms of numeraire), $MRS^i_{H,x} \equiv \frac{U^i_H}{U^i_x}$, weighted with the appropriateness index.

In equilibrium, $MRS^i_{j,x}$ must be equated to the out-of-pocket individual's cost of service, P^i_j, as a fraction of the producer price (marginal cost): $P^i_j = \eta^i_j MC_j$ with $\eta^i_j \leq 1.$

P^i_j, is the opportunity cost for the individual i, so the lower is η^i_j, the lower is the strength of the scarcity signal sent to agent i. The size of the opportunity cost specifically reflects the health care financing system. In general, the fraction is less than one because the service is not paid at full price. According to the prevailing system, the financing of the $(1-\eta^i_j)MC_j$ may derive from insurance premia, or from social contributions, or from the general tax system funding a National Health Service (NHS). In the latter case, P^i_j is near to 0 as it may, at the most, include a moderate co-payment.

1.2 From the rationale of public intervention to health care systems typologies

The modern theory of Public Economics singles out a set of reasons for the public intervention in the organization and provision of health services. The motivations follow both efficiency and equity concerns.

1.2.1 Efficiency and public provision of health care

First, although, as said, health services are private goods, nevertheless they often produce positive externalities, as their benefits spread also to people not directly consuming them. Vaccinations and the large part of preventive medicine provide the most emblematic cases. By consuming only according to private economic calculus, it would determine a social under-provision of the service, thus a public provision for internalizing the effects becomes socially desirable.
Second, health services are often merit goods, for whom a social (paternalistic) preference is adding to, or even substituting, the private one. The merit good argument is linked to imperfect (myopic) individual evaluation of the benefit of a commodity consumption. When a health service has this feature, its consumer price should be subsidized, until, if necessary, the total exemption.
Individuals have imperfect information on the features and the expected benefits of treatments. In general, health services are "experience goods", a specific category of commodities whose quality can be observed only ex-post, while is ex-ante left to a private contract, between the practitioner and the patient, unable to bound the former from inducing the latter to consume what and how he wants for personal aims. This argument might justify specific contractual relationships between the state and general practitioners, in order to increase the confidence by patients and also to reduce search

costs. Such contracts have generally a private nature, in terms of a convention with the unions of the sector, which disciplines the extent of practitioners' duties and their rewards.
The individual, facing the possibility of health care spending, given uncertainty and risk aversion, has a clear incentive to buy insurance. However, in this context we have the well known cases of market failure, due to asymmetric information between insurer and insured[13]. In particular, the phenomena of ex-post Moral Hazard (MH) given by the third party purchasing, TPP, context, and of Adverse selection (AS), generally imply the non-existence or the inefficiency of insurance market equilibria. In these cases a social insurance can guarantee, at some condition, welfare-improving equilibria w.r.t private second best Pareto-constrained equilibria.
In presence of MH-TPP, it can be shown that a risk-sharing equilibrium with partial coverage can be improved in a National Health Service (NHS) organization by a higher coverage and by a system of controls, provided the decentralised entities are effectively constrained by hard budget constraints. However, a system of co-payments may still be desirable for risk-sharing also in a NHS.
In case of AS, it can be proved that a Rothschild-Stiglitz separating equilibrium (RSSE) can be improved by substituting it with an uniform average premium, a Community rating insurance (CRI) with cross-subsidization. Further, this average premium can be also substituted by a system of social health contributions, as payroll taxes, or by specific taxes dedicated to funding a NHS.
A little formalization on private insurance failure can be useful [Zweifel et al. (2009, ch. 6)]. Let us start with an ex-post MH-TPP context. With

$$y = Y - \pi - E + I = Y - \pi - cE,$$

we represent the individual disposable income, equal to gross income Y less the premium π and the health service expenditure E plus the benefit of insurance (reimbursement), $I=(1-c)E$, where $0<c<1$, is the coinsurance rate. Let $u(E,y)$ the status dependent wealth utility, with the usual risk-averse hypotheses (concavity w.r.t. y), so we have: in the "sick status" $u^s = u(E, Y-\pi-cE)$ and in the "healthy status" $u^h = u(0, Y-\pi)$.
$ExpU(E,y) = pu^s + (1-p)u^h$ is the expected utility, where the sick status has probability p and the healthy status $(1-p)$. At the consumer equilbrium, the demand for the service and then the expenditure is a decreasing function of the coinsurance rate: $E(c), E'(c) < 0$. And the premium is a function of the coinsurance rate: $\pi(c) = p(1-c)E(c)$.
Further, the optimal second best coinsurance rate $c^* > 0$ may be given by solving the problem

$$\max_{(c)} \quad pu(E(c), Y - \pi(c) - cE(c)) + (1-p)u(0, Y - \pi(c))$$

whose F.O.C is as follows:

$$pu_E^s \frac{dE}{dc} + pu_y^s \left(-\frac{\partial \pi}{\partial c} - E(c) - c\frac{dE}{dc}\right) + (1-p)u_y^h \left(-\frac{\partial \pi}{\partial c}\right) = 0.$$

From the maximization of the direct expected utility function w.r.t. E we have as F.O.C.:

$$pu_E^s = cpu_y^s$$

Then by substituting, we obtain

[13] It is a specific application of the theory of contracts with asymmetric information, analised in Chapter 2.

$$cpu_y^s \frac{dE}{dc} - pu_y^s \frac{\partial \pi}{\partial c} - pu_y^s E(c) - cpu_y^s \frac{dE}{dc} - (1-p)u_y^h \frac{\partial \pi}{\partial c} = 0.$$

or

$$-(pu_y^s + (1-p)u_y^h)\frac{\partial \pi}{\partial c} - pu_y^s E(c) = 0.$$

Now taking into account that $\frac{\partial \pi}{\partial c} = p(1-c)\frac{dE}{dc} - pE(c),$ and by substituting, we obtain

$$-(pu_y^s + (1-p)u_y^h)[p(1-c)\frac{dE}{dc} - pE(c)] - pu_y^s E(c) = 0$$

$$(pu_y^s + (1-p)u_y^h)E(c) - (pu_y^s + (1-p)u_y^h)(1-c)\frac{dE}{dc} - u_y^s E(c) = 0$$

or, considering the elasticity of the health care expenditures w.r.t the coissurance rate, $\varepsilon_{Ec} \equiv \frac{dE}{dc}\frac{c}{E} < 0,$

$$(pu_y^s + (1-p)u_y^h) - (pu_y^s + (1-p)u_y^h)\frac{1-c}{c}\varepsilon_{Ec} - u_y^s = 0.$$

From this we get

$$\frac{1-c}{c}\varepsilon_{Ec} = -\frac{u_y^s - (pu_y^s + (1-p)u_y^h)}{(pu_y^s + (1-p)u_y^h)},$$

Hence, the optimal c^* is as follows:

$$\frac{1-c^*}{c^*} = \frac{\theta}{\varepsilon_{Ec}}; \theta \equiv 1 - \frac{u_y^s}{pu_y^s + (1-p)u_y^h} < 0.$$

This condition manages a trade-off between the aim of risk-sharing and the aim of controlling the level of over-spending in E, and it is similar to Ramsey inverse elasticity formula of optimal commodity taxation [Hindriks and Myles (2013, ch. 6)]. Indeed, $\frac{1-c^*}{c^*}$ can be thought as the rate of optimal subsidy (w.r.t. margional cost of 1€); symmetrically c^*, the optimal coinsurance rate can be considered as an optimal co-payment, thus positively related to price elasticity. Hence, the co-payment should be high for service with high elasticity and zero for services with zero elasticity (services for chronic pathologies and diseases).

The alternatives to the co-payment, for controlling the MH-TPP over-spending effect, may be the consumption rationing by lengthening the waiting lists, harder budget constraints and taxes increase. All these alternatives might be even more distortionary than the co-payment.

Now let us have a look to a standard AS context. Given two types of individuals with high, h, and low, l, probability of illness, $p_h > p_l$ and with fractions of population respectively of λ and (1-λ). Let π_i, $i=h,l$, be the premium for a contract giving a reimbursement I_i of the health service expenditure E, with $I_i \leq E$. As said, the RSSE, with the share of low risks sufficiently small in order to guarantee its existence, implies a structure of premiums and insurance coverage as follows

$$\pi_h = p_h E;$$
$$\pi_l = p_l I_l \text{ with } I_l < E.$$

A contract with full insurance is proposed and acquired by h-types and a contract with partial insurance is proposed and acquired by l-types.

A social insurance, in terms of mandatory pooling solution, is Pareto-superior to the RSSE equilibrium because both types are fully insured and l-type are better off. It may work with a CRI p^a

as follows

$\pi_h = p^a E = \pi_l; p^a = \lambda p_h + (1-\lambda) p_l$

where there is an implicit cross-subsidization taxes-transfers structure from l to h, $\lambda t_h + (1-\lambda) t_l = 0$, such that

$$t_l = (p_l - p^a)E = -\lambda(p_h - p_l)E < 0;$$
$$t_h = (p_h - p^a)E = (1-\lambda)(p_h - p_l)E > 0.$$

The insurance market failure arises also for those individuals whose probability of illness is near to one (elderly people, poor and socially excluded individuals, chronic invalid individuals) which could be cream-skimmed by private insurers. The coverage of these high risks can be guaranteed only by a social universal insurance system.

1.2.2 Equity and public provision in health care

Equity in health care pursues a notion of equity of outcomes, aiming at guaranteeing the opportunity to reach a given level of health to all individuals, independently on their economic, social and territorial conditions and status. The final objective is to allow human capabilities, such as the recovery of physical functionality, the absence of complications and a adequate life expectation, and to realise at least a decent level of health functioning. In formal term, the health care system should satisfy these two conditions:

$$s_i = s_u \Rightarrow \mathbf{m}_i = \mathbf{m}_u \text{ and then } H^i = H^u \quad \Leftrightarrow \quad h^i(.) = h^u(.)$$
$$s_i < s_u \Rightarrow \mathbf{m}_i \geq \mathbf{m}_u$$

According to the first statement if two individuals, *i* and *u*, have equal initial health status, they should have the possibility to access to the same vector of services and then to be potentially able to reach the same level of health. The sole differences should depend on the parameters of the household production function and then on the life style. According to the second statement, the individual *i*, with a worse initial health status than individual *u* should accede to a vector of services with at least one component strictly greater. Thus, this principle of health equity tends to reduce the differences in the initial level of health along the basic equality of opportunity idea of .more resources for less talent. However, the ways to actually reach, or approximate, this result depend on some feasible intermediate objectives. Indeed, several notions of universalistic provision of health care are actually pursued by the various systems, especially for specifying the prevailing funding criteria.

First, we have the simple notion of *equality of per-capita expenditure*, when the funding of decentralised public bodies is designed to equate per-capita public spending, possibly weighted with socio-demographic variables. Second, we have the *equality in satisfying standardised needs*, when the funding is referred to the notion of Essential levels of health care (ELC) and standardised costs per treatment. Third, it is frequently evoked the notion of *equality of access to services*, essentially meaning that the providers are obliged to guarantee an equal treatments to patients, independently on individual risk (no cream-skimming) and income (no wealth discrimination). Finally, we may remind the concept of *equality of individual payments*, when a uniform per-patient cost for health care treatments is required (no price discrimination). These intermediate objectives are actually attainable, but, could be conflicting each other and, in any case, they obtain only second best equity conditions, as they are imperfect and constrained respect to the first best one (fulfilment of a adequate level of functioning health).

Particularly meaningful is the second notion, referring to the criterion of guaranteeing a essential

package of care covered by public funding.

1.2.3 Industrial configuration failure and public production of health care

The previously mentioned equity and efficiency concerns offer the rationale for public intervention in terms of public provision of health services, i.e. by funding them outside price-setting, independently on the nature and ownership of the chosen providers. However, for some treatments, it may occur situations of industrial configuration failure where is socially desirable to join together public provision and public production as well. When it is convenient to localize services provision in urban areas where the private returns to invested capital are higher and there are economies of scale, local monopolies can easily take place for extracting spatial rents. Elsewhere, with lower expected returns, it may instead turn out a scarce network of providers and then an excess-demand of services.

In these cases, the public sector should be engaged not only to funding the services, but also to providing them directly, in order to support a fair territorial distribution of supply. The inadequacy of supply and the rationale of public production also occurs when, in some territories, the network of private providers, although numerous ones, cannot supply high-quality services. In these cases of inefficient industrial structure, public production, if well organised, is justified both for efficiency and equity aims.

1.3. Health care systems and risk disease coverage

The distinction between public (private) provision and production is crucial in order to analyse and classify the several models of health care. Indeed, the production argument allows us to talk about health care organization of supply, while the provision argument to talk about health care financing of medical expenditures. Both these two items contribute to define the features of a NHS.

We may distinguish two polar organizational cases: the Public system and the Private insurance system. As far as the first one is concerned, there is the purest version, the so called Beveridgean model, which considers full public provision and production of health services, both financed by general taxation. Essentially, the English and Italian NHSs, before the reforms of the last two decades, were following this model. Also the so called Bismarkian model is a public model. It is the original social insurance model financed by health contributions going directly to public Health Funds. In principle, both Beveridgian and Bismarkian models are universalistic ones.

As a Private insurance system we can mention that one presently still working in U.S.A. In this system there is a prevailing structure of private insurances policies and only some public programs for specific categories of patients and limited fractions of population: the Medicare for elderly people (over 65 years-old) and Medicaid for poor people (selected by means testing). This health care system is coherent with the concept of residual welfare state implying a limited coverage of every social risk. The two polar models are now mainly academic ones and statistical outliers. For instance, all European countries are now applying mixed systems, where public and private programs are acting simultaneously for guaranteeing a universalistic provision of health services.

In order to evaluate the different systems, it is crucial to precise what is meant with degree of health care coverage of each one. However, this notion is not unique as it may be referred to: (i) the *extent of coverage*, i.e. the share of population whom health care is guaranteed; (ii) the *depth of coverage*, i.e. the number and the features of services included in the insured package; (iii) the *highness of coverage*, i.e. the fraction of treatment costs directly financed by the insurer or the NHS and then not directly paid by patients. This classification must be taken into account for avoiding confusions in looking for what is simply named a universalistic system.

In order to distinguish the variety of mixed systems we have to consider several features. First, the criteria according to which the general practitioner (the agent who makes the order of purchasing the

service) is assigned to each household. Then, the individual degree of choice of the provider by which to obtain health care services, as diagnostics and specialist treatments and hospital admissions, is meaningful. In this respect, there are systems allowing a high individual freedom of choice and other ones with a rigid assignment by public administration, but, of course, there are several intermediate situations. Another relevant distinction refers to the ways of paying the providers, i.e. if it is allowed an ex-post coverage of the production costs, or it is fixed an ex-ante budget, or it is organized a system of prospective standardised tariffs for each treatment (see infra). Further, the mixed systems may be differentiated according to the forms of organizing the supply of drugs and to the ways their prices are established, and the co-payments and coinsurance rates structure is designed.
According to these criteria, the literature in health economics usually distinguishes three specific types of mixed systems.
The first one is known as the *reimbursement model*, where insurance companies or decentralised health districts (or regional governments) reimburse the patients expenditures, after they have paid, at administered prices, the services to public as well private providers.
The second system is the *integrated model*, where the health district builds up with the providers a unique connected public structure. The patients do not pay for the treatments which are financed by taxes, but they have a limited or no choice where to receive them. The internal hospitals and clinics are financed at costs of inputs or, sometimes, with reference to a fixed budget.
Finally there is the *contractual model* where the funding body, the insurer or the health district, is separated from the providers which, whether public or private ones, are committed and rewarded according to a procurement contract. This is based on prices that in most cases are fixed ex-ante and standardised along the system of Diagnostic related groups (DRG). Consumers-patients have, according to the variety of contractual systems, some freedom of choice of the provider and the general practitioner they desire to engage. Quasi-markets (QM) are a specific typology of the third model, often called managed or internal competition, applied in the last decades in some European countries.

2. Organization of health care production and provision

2.1 Health care industrial organization of production: separation versus integration

The vertical separation of purchasing structures from producing and supplying ones is a necessary but not sufficient condition for creating competition in health care industry. Thus, this institutional design issue must be treated per se, by looking at cost-benefit of vertical integration in terms of production costs level, economies of scale and scope, network and coordination economies. Within this context, health economics can well consider the pros and cons of different organizational systems. Indeed, the transactions between a buyer, e.g. a health district, and seller, e.g. a hospital, can be realized by a market contractual exchange (in case of separation) or by a internal transfer within a unitary body (in case of integration).
Therefore, for modelling the determinants of this choice, we have to consider the usual alternative between hierarchy and market, by comparing the administrative costs of managing a complex structure (the firm), with the costs for signing incomplete contracts with the providers (the market), as we discussed in section 2.1.
Further, a variety of industrial issues matter, like the degree of complementarity of the purchasing and production assets, the rent-seeking behaviours of the several agents, those making the choice and those applying it for the relevant activities, the existence of sunk costs and irreversible investments and the consequent hold-up issue. Of course, all the transaction costs concerns, as the contractual size and complexity, and the time and costs requested for settling the eventual controversies on trial are

also relevant. Thus, by limiting the opportunistic behaviours and ex-ante and ex-post contract inefficiency, the integration will be preferable to a system based on market exchanges and transactions. This will happen when its advantages can exceed the static and dynamic inefficiencies due to huge bureaucratic centralised structures, typical of public administration.
We can formalize the different typologies of I-O schemes as follows. Let us employ these notations:
A = set of institutions, such as national or regional authorities, devoted to regulating a public and/or private production,
B = set of agencies conveying the users demand and needs, and
C = set of the suppliers of treatments and final services.
Now, we have three cases, summorized in the following matrix:

1.Separated model	$A \cap B \cap C \equiv \varnothing$
2.Integrated model	$A \cup B \cup C \equiv \Omega$
3.Semi.integrated model	$(A \cup B) \equiv \Psi \cap C \equiv \varnothing$

In a de-integrated (separated) model, as the Dutch one, the intersection of the three sets is empty as each set interacts with the others by contractual exchanges. In this case - the Quasi-market system (QM) - there is a limited degree of industrial concentration. The tasks of expressing, on the behalf of patients, the demand and of establishing the appropriateness of services and treatments are delegated to a specific Authority, a public Sponsor of citizens. The providers are government or non-profit institutions, but they must be, in any case, appropriately ex-ante selected as reliable preferred providers. The selection procedure aims at controlling and limiting the provider power in the negotiation due to asymmetric information and at assuring that the several necessities and urgencies of users are suitably fulfilled.
In a public-public integrated model, like the one prevailing in some European regimes, the union of the three sets, Ω, assembles all the health care activities. Thus, there is a strong industrial integration that, in the polar version, unifies in a whole structure, a Local health firm, as the ASL in Italy, all the functions of planning, demand rationing, financing, production and supply of services. In milder versions of the model, A and B are joined in a union set Ψ, the intersection of the latter with C is once again empty: Ψ buys services from C. Hence, some hospital firms (AO) may be separated from the ASL, still remaining publicly owned. This semi-integrated configuration is wholly working in U.K, where there are a Health District Authority (HD) separated from the Trust hospitals (TH), which the patients can almost freely choose as favourite.
The main objective of the integration is to improve the capacity of coordination by the planner, by limiting the conflicting interests among the involved agents. This kind of conflict is, instead, considered beneficial by the advocates of the de-integrated model. Further, the integrated model is aimed at limiting the opportunistic manipulation of demand, which is often not objectively determinable, mainly on the social-assistance components.
However, the integrated model has all the shortcomings and defects of the centralised and complex structures, at high level of bureaucratization and high administrative costs, and it suffers the politicians' interference, in all the ways it can occur. Moreover, the model does not give adequate incentives toward the internal efficiency of the institutions whose lack becomes the main cause, together with the so-called soft budget constraint syndrome, of the wide sunk deficits of the decentralised bodies.
Consequently, on looking at the cost-benefit of separation vs. integration, considerations of Political economy should be taken into account. The application of new Political economy models to health economics allow to contemplate the behaviours and the conflicting relationships among citizens and politicians, the public providers and the interests groups, as the pharmaceutical industry and doctors profession.

2.2 Systems of payment to providers: Tariffs for treatments and incentives

Let a tariff for a hospital treatment – under a separated or semi-separated context- be given by the reimbursement of the costs plus a reward to the manager as a linear function of the costs themselves:

$$T = t + C(q, e, \mu)$$
$$t = a - bC(q, e, \mu);\ 0 \le b \le 1.$$

where $C(.)$ is the production cost of the volume of treatments q; $C_q \equiv \frac{\partial C}{\partial q} > 0$ is the marginal cost. The parameter $e \in [e_{\min}, e_{\max}]$ is the cost-containment effort carried on by the hospital manager and $C_e \equiv \frac{\partial C}{\partial e} < 0$ its marginal effect on the level of cost, with $\frac{\partial / C_e /}{\partial e} < 0$, for regulatory concerns. The parameter $\mu \in [\mu_{\min}, \mu_{\max}]$ represents the quality of the treatment and $C_\mu \equiv \frac{\partial C}{\partial \mu} > 0$ its marginal cost, positive as clearly higher quality requires more resources, with now $\frac{\partial C_\mu}{\partial \mu} > 0$. The term b measures the incentive power of the payament scheme[14].

Let us suppose that a HD or an ASL plan establishes a given volume of output to be provided, so $q=q^*$ is inserted in the contract signed with the hospital (TH or AO). Both e and μ are instead not observable by the HD and then are not contractible, while production costs of the TH are only ex-post observable, i.e. when realised. Let then define with $\psi(e,\mu)$ the objective function of the manager of the hospital, where $\psi_e \equiv \frac{\partial \psi}{\partial e} < 0$ represents the marginal cost of cost-containment effort (disutility), with $\frac{\partial / \psi_e /}{\partial e} > 0$, and $\psi_\mu \equiv \frac{\partial \psi}{\partial \mu} > 0$ the marginal benefit of quality in terms of prestige and reputation of the manager, with $\frac{\partial \psi_\mu}{\partial \mu} < 0$.

In a separated or semi-integrated system, a residual claimant provider will tend to choose e and μ to maximize the following function of the residual:

$$R = T - C(q^*, e, \mu) + \psi(e, \mu) = a + (1-b)C(q^*, e, \mu) - C(q^*, e, \mu) + \psi(e, \mu).$$

In the case of a fee-for-service tariff (cost-plus contract), as in both reimbursement and integrated models, we have b=0. Therefore, it is

$$T = a + C(q^*, e, \mu) \text{ and } R = a + \psi(e, \mu)$$

i.e. the residual is independent on production costs, so there is no incentive to contain them. Indeed, theoretically, the optimal level of the hospital manager effort is a corner solution such that $e^*]_{b=0} = e_{\min}$. However, being fully insured, the manager does not exploit the informative monopoly and might give up cream-skimming procedures and provide high quality treatments. Therefore, it might well happen a further corner solution such that $\mu^*]_{b=0} = \mu_{\max}$.

In the case of a fixed per treatment tariff (fixed-price contract), i.e. b=1, as in pure QM contractual model, where there is a prospective DRG payment, it is

$$T = a \text{ and } R = a - C(q^*, e, \mu) + \psi(e, \mu)$$

[14] Note the similarity of this model to that one explained in section 2.3.4.

Since the residual is now decreasing with the treatment production cost, the hospital has an effective incentive to contain it. The optimal level of the effort is given by the condition $/C_e/ = /\psi_e/$ implying that $e^*]_{b=1} > e^*]_{b=0} = e_{\min}$. However, the hospital, bearing the full firm risk, may be induced to cream-skim high illness risks and to restrain the quality level of the service; indeed, in this case, the optimal quality level is given by the condition $C_\mu = \psi_\mu$, and hence $\mu^*]_{b=1} < \mu^*]_{b=0} = \mu_{\max}$.

In more concrete terms, a fixed-price contract, without a reliable ex-ante selection of providers and effective controls on standards, may entail a worsening of treatment quality, with for instance a higher rate of patients' mortality.

This simple model seems to suggest that, in order to reach an adequate standard of quality but satisfying a fixed budget constraint, a mixed systems of cost and risk-sharing payment, with $0<b<1$, might be preferable.

The optimal levels, e^0 and μ^0, are now given by the two conditions

$$b/C_e/ = /\psi_e/$$
$$bC_\mu = \psi_\mu,$$

according to which, it obtains

$$e^*]_{b=1} > e^0 > e^*]_{b=0} \equiv e_{\min}$$
$$\mu^*]_{b=1} < \mu^0 < \mu^*]_{b=0} = \mu_{\max}.$$

Actually, an incentive risk-sharing criterion may be pursued by organising a system of budgeted plans based on a fixed volume of treatments, some proxy indexes of quality, and standardised and verifiable costs, with ex-post revenues abatements in presence of non-fulfilments of the objectives.

Chapter 5

A Political economy approach to local governments choices

1. Modelling local governments choices with equalization transfers

In many federal countries the central government transfers resources to local jurisdictions in order to alleviate the imbalance between expenditures needs and revenues [Hindriks and Myles (2013, ch. 20), Petretto (2013b)]. The aim is to ensure to every citizen the access to reasonably comparable levels of public services within a chosen locality, at a cost in line with what would be paid elsewhere. Therefore, equalization transfers promote horizontal equity by permitting fiscal treatment of identical persons in a federation and by enabling jurisdictions to provide minimum standards of essential packages of public services.[15]

Around the world, in industrialized countries and in less developed countries as well, we may find many applications of *Fiscal capacity equalization* and of *Need equalization*. The former, on the basis of the so-called *Representative Tax System* (RTS), tends to equalize the difference between standard revenue and the effective local one (at a uniform tax rate), while, the latter tends to cover the difference between a standardized local need expenditure, measured on the basis of the so-called *Representative Expenditure System* (RES), and some benchmark. Combinations of RTS and RES are often also applied. In Italy, the Need equalization criterion is applied for some essential regional expenditure items, like health care, social assistance, education and public transit (more or less 80% of total expenditure), while Fiscal capacity equalization criterion is applied for transfers to municipalities.

Equalization systems, as said, are specifically devoted to guarantee horizontal equity but they have also efficiency implications. In this respect economic literature has developed two specific issues. On one hand, it has analyzed the consequences of migration and factor mobility, due to equalization, on productivity of the local firms. On the other hand, the economic literature has deeply discussed the efficiency consequences of equalization in terms of the level of tax rates and public expenditure, taking also into account tax competition phenomena.

In this chapter, we deal with the efficiency implications of Need equalization by looking at the consequences of such transfers on productive costs and quality of local public services provision.

In order to examine this matter, we build up a simple model where the flow of federal transfers to local governments is given by a revenue sharing of a federal tax and a need equalization grant. The latter is specified along a well known RES rule, applied in several federal countries. According to this, the grant is linked to the gap between a need standardized expenditure index and a standard local tax revenue index. Further, we assume, along the political economy theories of state and institutions examined in the second chapter, that local politicians have some preference on cost-inefficiency, as by this means they can acquire political consensus with perquisites and wasteful expenditures. Therefore, they are conflicting with users of public services who want instead high quality services and low local taxes.

We consider a federation with a pre-committed central government and several local governments, not fiscally interconnected each other. Hence, we may simply model a local government that provides to a representative consumer-tax payer a composite local public service, considered, according to a *merit good argument*, as essential by the national legislation. It finances the production costs of

[15] Specific notions of equalization are disciplined by many Constitutional acts, as, for instance, in Canada, Australia, Germany and Italy.

A. Petretto, *Economics of Institutions and Law*, ISBN (print) 978-88-6453-579-1, ISBN (onine) 978-88-6453-580-7, CC BY 4.0, 2017 Firenze University Press

quantity *q* of the service at a quality level *m*, the latter measured by a real number in a closed interval[16], with the following means. First, a surtax at rate *t* on a tax base *Y*, which is also taxed, at rate τ, by the central government. Second, the local government receives also, as a transfer, a share α of the revenues collected by the central government. Thus, the central government tax base is partly shared with the jurisdiction. Finally, there is the equalization grant whose mechanism is to be specified in the following sections[17].
The central government remains on the shadow, having already chosen own tax rate and fiscal arrangements of the equalization transfers, which are then exogenously given. As a consequence, we want, as said, to ascertain the local government choices given these fiscal parameters.

1.1 Consumer preferences

These are represented by the following separable function

$$V = \upsilon(\rho, I) + \varphi(q, m).$$

$\upsilon(\rho, I)$ is an indirect sub-utility function of aggregate tax rate, $\rho \equiv t + \tau$, and initial endowment of resources (untaxable income), *I*. By duality, this derives by maximizing a quasi-concave direct utility function, which depends on a untaxed commodity, the numeraire, and a taxed one, whose value at producer price turns out to be the tax base, *Y*. From now on, *Y* is disposable labour income and then the untaxed commodity is leisure.
Accordingly, by Roy identity[18], $\upsilon_\rho = -\upsilon_I Y < 0$. Moreover, from consumer equilibrium, it can be derived the consumer reaction function to fiscal choices, $Y(\rho), Y_\rho < 0,$ i.e. a declining supply of labour w.r.t the consolidate tax rate.
$\varphi(q, m)$ is a quasi-concave sub-utility function of quantity and quality of the public service, with $\varphi_q > 0, \varphi_{qq} < 0; \varphi_m > 0, \varphi_{mm} < 0.$ Quantity and quality can be complements ($\varphi_{mq} > 0$) as well substitutes ($\varphi_{mq} < 0$); thus the marginal willingness to pay for quality can increase or decrease with the consumption of the service, according to the type. Actually, we may find examples of local public services where a high quality of provision can favour as well discourage an increase of quantity demanded. With $\zeta_{mq} \equiv \frac{\varphi_{mq} q}{\varphi_m}$ we denote the unsigned demand-elasticity of substitution between quality and quantity.

1.2 The local government revenues

As said, the local government obtains funds from three sources: (i) the local taxation, *tY*; (ii) a revenue sharing over the federal tax yield, $\alpha \tau Y$, where $0<\alpha<1$ is the fraction decided by federal government; (iii) a equalization grant, if entitled. Indeed, we consider a gross, vertical, equalization process, by which only poor regions receive a grant, and the total of grants are funded by federal taxation.
Consequently, according to a somewhat general RES rule, the transfer is given as follows

$$G = Max[\beta(N - t^s Y), 0].$$

---

[16] This parameter is clearly anologous to that one of the previous chapter referred to quality of the hospital treatment.

[17] There are no own local taxes but this restriction is only for simplifying the analysis. Indeed, we could consider also a local commodity tax but, if its base would be inserted in the equalization mechanism, the effect of the latter on productive efficiency should have the same qualitative sign.

[18] As with X_y we denote the partial derivative $\partial X / \partial y, \upsilon_I$ is the marginal utility of income and so on.

$0 < \beta \leq 1$ is the equalization rate, N the need lump sum component of the grant, to be explained in the successive sub-section, and t^S is the standard surtax rate, a fiscal policy arrangement. According to this rule, a local government receives a share of the difference between the Need and the local revenues it can collect with reference a surtax rate specified in the equalization rule itself. If the these revenues are higher than the Need, the jurisdition, being a "rich" one, does not get any grant at all.
A simplified RES rule can be applied by constraining the parameters as follows:

$$\beta = 1, t^s = t_0 + \alpha\tau = \frac{N_r}{Y_r}.$$

In other words, for only one jurisdition or region, the richest one $(Y_r = Y_{\max})$, it is $G_r = 0$. All other ones, i.e. $i \neq r$, receive a grant exactly equal to the difference between the need term N, a standardized public expenditure as we'll see later on, and the revenues from local taxation, obtained applying a basic uniform (minimum) surtax rate, t_0, and revenue sharing:

$$G_i = N_i - t_0 Y_i - \alpha\tau Y_i.$$

The revenue sharing rate α is established for exactly allowing the budget equilibrium, without any grant, for the richest local jurisdition[19]. Hence:

$$\alpha = \frac{N_r - t_0 Y_r}{\tau Y_r}.$$

It is clear that t_0 and α are equivalent tax instruments and they are both exogenous for local government: the central government fixes the first one and the second one is obtained as solution of the budget equation of richest region. Hence, actually, only t_0 is the fiscal policy variable[20].
Let us now consider the case of RTS rule, with the Fiscal capacity equalization grant. It is possible to obtain the rule simply fixing the lump sum component of the grant as $N \equiv t^s Y^s$, where Y^S is the standardized (average) tax base. Thus,

$$G = Max[\beta t^s (Y^s - Y), 0].$$

This is the gross (vertical) Fiscal capacity equalization version of the rule. The system can be also net (horizontal) when the transfer (positive as well negative) is

$$G = \beta t^s (Y^s - Y)$$

Hence, in this case, rich regions or municipalitis support the poor ones with a negative subsidy, while central government does not intervene.
Notice that, in both RES and RTS. G is a "matching grant", linearly and negatively related to local tax base, and we will see that this is what mainly matters as far as the incentive to efficiency is concerned.
Summing up we get the following revenues function for a poor region:

$$R = \beta N + \hat{t} Y$$

where $\hat{t} = t + \alpha\tau - \beta t^s$ is the effective local tax rate, i.e. the perceived local rate to which the fiscal distortion at local level is linked. Notice as in the simplified RES rule previuosly considered, it would be $\hat{t} = t - t_0$, and simply $R = N + (t - t_0)Y$, where $t \geq t_0$. The analysis of the following subsections can be easily adapted to this special case.

---

[19] This rule is applied for financing public services provided by municipalities within Landers in Germany.
[20] The fiscal autonomy of each jurisdition can be pursued by means of the other local taxes, however excluded from the equalization formula.

1.3 Needs and costs

If we adopt the RES interpretation, according to which the Needs are measured by the product of standard unitary cost c^S with minimum (essential) standard of output provision, q_E, the grant formula becomes as follows

$$N \equiv c^s q_E.$$

Index q_E can be thought as a synthetic representation of the normatve notion of *Essential levels of health care*, explicitly mentioned by some legislation (see previous chapter section 1.2). The parameter c^S is specific to the considered jurisdiction, and it may be estimated or computed by one of the several RES techniques, e.g. the regression analysis of an expenditure (cost) function.
As far as the production costs of the jurisdiction are concerned, we assume this factorized, quasi-linear, function:

$$C(q,m,\theta;A) = c(q,m;A)\theta q.$$

Hence, the total costs are given by the product of a unitary average cost function $c(q,m;A)$,where A is a vector of demographic and environmental variables influencing the production, with the level of output q inflated by the parameter $\theta \geq 1$. The latter is a variable of cost-inefficiency, an index measuring the impact of perks and wasteful expenditures made by the local politicians and bureaucrats seeking for political consensus and power. Therefore, it is also an index of the incumbent politicians' ability or competence in that jurisdiction[21].
The shape of the cost function is given by the following set of expressions:

$$C_q = \left[c_q q + c(q,m;A)\right]\theta > 0, C_m = c_m \theta q > 0, C_\theta = c(q,m;A)q > 0$$
$$C_{qq} = [c_{qq} q + 2c_q]\theta ? 0, C_{mm} = c_{mm}\theta q > 0, C_{\theta\theta} = 0$$
$$C_{\theta q} = C_q / \theta > 0$$
$$C_{mq} = (c_{mq} q + c_m)\theta \geq (<) 0 \Leftrightarrow \xi_{mq} + 1 \geq (<) 0, \xi_{mq} \equiv \frac{c_{mq} q}{c_m}$$

The marginal costs of quantity, quality and inefficiency index are positive, as all employ scarce resources. The first one may be increasing as well decreasing with quantity as neither c_q or c_{qq} are signed. The second one is increasing with quality as we may reasonably assume $c_{mm} > 0.$ The third one is constant with respect to inefficiency index. The positive sign of $C_{\theta q}$ implies that quantity and inefficiency are cost-substitutes, which seems conceivable. Cmq is not instead signed, depending on the sign of c_{mq}. If ξ_{mq} is higher (lower) than -1, quality and quantity are cost-substitutes (complements). In the latter case, the technology exhibits *economies of scope* in producing output with high quality. Actually, an innovation increasing the standard of quality may save or as well require more resources for producing the service.
We may interpret the standard unitary cost c^S in this way. Let us assume that central government

[21] Notice the difference between the inefficiency parameter θ and the parameter *e,* used, in the previous chapter, to represent the cost-containement effort by the hospital manager. They have an anologous political economy source but the opposite sign.

knows the local cost function $C(.)$, but does not observe the quality locally realized, being able only to estimate the mean value $Em = \overline{m}$ from a probability distribution of quality indexes $\Im(m)$. Environmental features A are observed and employed in econometric analyses for estimating the standard cost. The variable θ is not observed and then not acknowledged in the contract defined by the equalization rule. Therefore, the standard unitary cost may be $c^s = c(q_E, \overline{m}; A)$, which might be lower or higher than the effective unitary cost $c(q,m;A)\theta$, depending on the level of output (returns to scale), the actually realized level of quality and the inefficiency index. However, the former is the most likely (normal) case.

1.4 Government preferences

We suppose local politicians have, as pay-off function, the sum of the utility function of the representative consumer and the following benefit function of extra-costs for perquisites and wasteful expenditures:

$$a\psi(\theta), \psi' > 0, \psi'' < 0.$$

The function $\psi(\theta)$ reflects the preference for cost-inefficiency and $a \geq 0$ shows the degree of non-benevolence or rent taking by local politicians. If a=0, they are perfectly benevolent as rightly accountable. If a>0 they are in some extent rent-takers.

Accountability depends on institutional rules, in particular on the transparency and simplicity of the techniques applied for assessing the Need index and implementing the chosen equalization. If local politicians can hide the exact mechanism of equalization to voters, can limit or avoid the control over their rent-taking activities. This is easier if, for instance, the technique of estimating the Need index is controversial and even too complicate to be understood in the political debate. In these cases, local politicians may shroud on the fog the rent they get by reporting there is a lack of funds due to the formula which is, in their opinion, specifically damaging the region. Another example of a lack of transparency, favoring the opportunistic behavior of politicians, comes from an unclear definition of tax base to which the equalization applies. If, for instance, the tax base is the reported income instead of the earned one, the local politicians may be induced to do not contrast sheltering procedures (tax avoidance and evasion) carried on by taxpayers.

Any institutional reforms able to remove or limit such lack of transparency may, reducing a, improve the efficiency of a local governments in looking for high quality and low costs. More generally, all political reforms improving politicians accountability, as electoral rules towards a more direct selection of majors and regional governors by voters have the same beneficial effect.

2. The Political economy equilibrium

The local politicians choose their strategies knowing the federal government fiscal choices and the consumer reaction function, i.e. the shape of the tax base function. In the following section, we are going to ascertain, first, how the fiscal autonomy is used with the choice of the surtax rate t, for financing extra-costs, and, second, how local politicians are choosing their strategies on m and θ, given the fiscal arrangements α, β, t^S, on the minimum standard q_E and also on parameter a.

2.1 Equilibrium conditions

The equilibrium of local government is obtained by solving the following maximization process w.r.t q, m, t and θ:

$$Max\, W = \upsilon(t+\tau, I) + \varphi(q,m) + a\psi(\theta)$$
$$s.t.$$
$$\beta N + \hat{t}Y = c(q,m;A)\theta q \quad (\lambda)$$
$$q \geq q_E \quad (\mu)$$
$$\mu(q - q_E) = 0$$

The corresponding Lagrangian is the following function[22]:

$$L = W + \lambda[\beta N + \hat{t}Y - c(q,m;A)\theta q] + \mu(q - q_E)$$

The multiplier λ reflects, as usual, the marginal cost of taxation, while the multiplier μ reflects the benefit of the service as a merit good and the cost of strengthening the binding minimum standard constraint. Indeed, by applying the envelope theorem to the maximum function $W^*(\alpha, \beta, t^s, q_E, a)$, we get the responses of the local government pay-off function to the changes in the fixed and exogenous parameters (all, but *a*, decided by the central government):

$$\frac{\partial W^*}{\partial \alpha} = \lambda \tau Y > 0,$$
$$\frac{\partial W^*}{\partial \beta} = \lambda(N - t^s Y) > 0,$$
$$\frac{\partial W^*}{\partial t^s} = -\lambda \beta Y < 0,$$
$$\frac{\partial W^*}{\partial q_E} = \lambda\beta(c^s + \frac{\partial c^s}{\partial q_E} q_E) - \mu,$$
$$\frac{\partial W^*}{\partial a} = \psi(\theta) > 0.$$

Therefore, given the marginal cost of taxation, λ, the maximum local government objective function is increasing with the revenue sharing rate and the equalization rate, and decreasing with the standard tax rate. The sign of the objective function change w.r.t. the minimum standard q_E depends on the comparison between the benefit of alleviating the budget constraint with a higher grant [23], $\lambda\beta \frac{\Delta N}{\Delta q_E}$, and the opportunity cost of allocating resources on production instead to other tasks (e.g. quality as well perks), μ. Finally, of course, the local politicians' pay-off function in equilibrium is increasing with the degree of rent-taking opportunity *a*.

Clearly, in the simplified RES rule case, only the following effects are meaningful:

$$\frac{\partial W^*}{\partial t_0} = -\lambda Y < 0, \frac{\partial W^*}{\partial q_E} = \lambda(c^s + \frac{\partial c^s}{\partial q_E} q_E) - \mu, \frac{\partial W^*}{\partial a} = \psi(\theta) > 0.$$

Now the objective function is decreasing with the uniform tax rate. Notice that, in this case, the politician is indifferent on the level of revenue sharing rate, as all changes of it are compensated by the Need equalization grant.

The F.O.C.s of maximizing the Lagrangean function $L(q, m, t, \theta; \lambda, \mu)$ are as follows:

---

[22] Notice, as the revenue is given by $(t + \alpha\tau)Y$, for a rich region we may insert β=0 in the budget constraint and all the successive espressions.

[23] However, notice that economies of scale, c_q <0 could even reduce the standardised cost and then the grant.

$$(q^*): (\varphi_q + \mu) - \lambda C_q = 0, q^* = q_E$$
$$(m^*): \varphi_m - \lambda C_m = 0$$
$$(t^*): \lambda = \upsilon_I \eta(\alpha, \beta, t^s),$$
$$(\theta^*): a\psi'(\theta) - \lambda C_\theta = 0.$$

The condition for optimal tax rate t^* can be re-written, using the notion of Marginal Cost of Public Funds (Dahlby 2008)[24], as follows

$$MCPF \equiv \frac{\lambda}{\upsilon_I} = \eta(\alpha, \beta, t^s) = \frac{1}{1 - \frac{\hat{t}\varepsilon}{1 - \tau - t}}$$

where $\varepsilon \equiv Y_{1-\rho}\frac{1-\rho}{Y} > 0$ denotes the elasticity of labour supply with respect to the net wage[25].

It is immediate to check that in normal conditons, *MCPF*>1, so it means that 1€ of public expenditure costs to the local society of tax-payers more than 1€, because of the tax distortion due to the substitution effect.

It is interesting to note the difference between this expression of *MCPF* with that one where the local govenment gets revenue only by the surtax, without any transfers by the central government, thus $\hat{t} = t,$ i.e. when $MCPF = \frac{1}{1 - \frac{t\varepsilon}{1-\tau-t}}$.

We cannot say if the latter is higher or lower than the former. The fact is that, in our model, *MCPF* depends on the "effective local tax rate" $\hat{t}$, the perceived local rate by the economic agents. Therefore, the distortion depends not only on federal and local tax rates, but also on the terms linked to the trasfer rule: $\alpha\tau - \beta t^s$. For instance it is immediate to check that an increase in β and/or t^S reduces the tax distortion effect, the contrary with an increase of α. Hence the *MCPF* with revenue sharing and Need equaization grant is higher (lower) than that one in fiscal autonomy if $\alpha\tau > (<)\beta t^s$.

From the F.O.C.s we obtain, by susbstituting and eliminating the Lagrange multiplier λ, the following conditions for *q, m* and θ:

[24] Indeed we have

$$\frac{\partial L}{\partial t} = -\upsilon_I Y + \lambda \hat{t} Y_\rho + \lambda Y = 0$$

Then

$$-\upsilon_I - \lambda \frac{\hat{t}}{1-\rho}\frac{Y_{1-\rho}(1-\rho)}{Y} + \lambda = 0$$

Thus

$$\upsilon_I = \lambda\left(-\frac{\hat{t}}{1-\tau-t}\varepsilon + 1\right).$$

[25] This is inverse to the elasticity of the tax base w.r.t to tax rates when the subnational government is on the upward-sloping section of its Laffer curve.

$$\frac{(\varphi_q + \mu)}{\upsilon_I} = C_q \eta(\alpha, \beta, t^s), q^* = q_E$$

$$\frac{\varphi_m}{\upsilon_I} = C_m \eta(\alpha, \beta, t^s)$$

$$\frac{a\psi'(\theta)}{\upsilon_I} = C_\theta \eta(\alpha, \beta, t^s)$$

The first condition establishes that in equilibrium, the marginal benefit of the output $\frac{(\varphi_q+\mu)}{\upsilon_I}$ (where the "merit good argument" Lagrange multiplier μ must be added to φ_q when the chosen output has to be equal to the minimum standard q_E) is equal to its marginal cost, "corrected" because including the *MCPF*. The second one says that the marginal benefit of the quality $\frac{\varphi_m}{\upsilon_I}$ is equalised to its "corrected" marginal cost. Given the features of the cost function, it turns out that $m^*>0$. The third one says that the optimal level of cost-inefficiency for the local politician is such that the marginal benefit $\frac{a\psi'(\theta)}{\upsilon_I}$ is equal to the marginal cost, i.e. the implied increase of the productive costs correctd by the *MCPF*. Clearly $\theta^*>1$ if $a>0$. In words, the level of cost-inefficiency depends positevely, as expected, on the degree of non-benevolence or rent-taking by the politicians involved in the local government.

2.2 Simple comparative statics

We may conclude the section by verifying the impact of changes of fiscal parameters chosen by the central government on the equilibrium values of quality and efficiency. In order to ascertain easily the effects of changes of α, *β, t*^{S} on m^* and θ^*, we may follow a heuristic partial equilibrium approach, according to wich we assume constant both the marginal utility of income, υ_I, and the elasticity of labour supply, *ε*.

Given the properties of utility and cost functions (in particular $\varphi_{mm}<0$ and $C_{mm}>0$), $MBm \equiv \frac{\varphi_m}{\upsilon_I}$ is a decreasing function and $MCm \equiv C_m \eta(\alpha, \beta, t^s)$ and increasing function on *m*, given *q, t* and θ. Further, (given $\psi''<0$ and $C_{\theta\theta}=0$), $MB\theta \equiv \frac{a\psi'(\theta)}{\upsilon_I}$ is decreasing and $MC\theta$ constant w.r.t. θ, given *q, t* and *m*. Equilibrium values m^* and θ^* are where the marginal benefit and marginal cost functions intersect each other.

Let consider the chosen level of quality, m^*. An increase of the revenue sharing $\Delta\alpha$, tends to increase the marginal cost of *m* as the perceived tax rate $\hat{t}$ increases and then the tax distortion increases the dimension of $\eta(\alpha, \beta, t^s)$: $\frac{\partial \eta}{\partial \alpha} = \tau \frac{\partial \eta}{\partial \hat{t}} > 0.$

Therefore, *MCm* increases, while *MBm* remaining unchanged, and this implies a reduction of the equilibrium level of quality. An increase of the degree of equalization $\Delta\beta$ tends, instead, to decrease the marginal cost of *m* as $\frac{\partial \eta}{\partial \beta} = -t^s \frac{\partial \eta}{\partial \hat{t}} < 0.$ and then it determines, still *MBm* remaining unchanged, an increase of the equilibrium level of quality. The same result obtains with an increase of the standard surtax rate Δt^s as $\frac{\partial \eta}{\partial t^s} = -\beta \frac{\partial \eta}{\partial \hat{t}} < 0.$

Let us consider the level of productive inefficiency, θ^*. An increase of the revenue sharing $\Delta\alpha$, tends to increase the marginal cost of θ, $MC\theta$ given the increase of $\eta(\alpha, \beta, t^s)$, and this, with $MB\theta$ unchanged, implies a decrease of θ. Opposite is the effect of an increase of β and t^S.

Notice that an increase of accountability $-\Delta a$ gives an increase of $MB\theta$ and a decrease of the choice

variable θ. In words, any reform increasing the transparency of the equalization system, and then the accountability of the local political set-up, implies a reduction of cost-inefficiency, without influencing the level of quality.

Interesting enough is also to verify the effects on quality and inefficiency of changes of the essential level (minimum standard). An increase of q_E, if $\mu>0$, implies an increase of q. Consequently, the effect on quality of an increase of the minimum standard of the public service provision is not determined, as the shift of marginal benefit and cost functions depends on the relative shape of marginal utility and marginal cost of quality. Indeed, as seen, the sign of φ_{mq} and C_{mq} is not given a priory. The cost-inefficiency θ tends instead to certainly decrease with an increase of the minimum standard, because of the cost-substitutability between inefficiency index and quantity $C_{\theta q}>0$, which increases the opportunity cost of wasteful expenditures.

References

Acemoglu, D. (2009), *Introduction to Modern Economic Growth,* Princeton University Press, Princeton.

Acemoglu D. (2013), *Political Economy Lecture Notes,* MIT, Cambridge Ma.

Acemoglu, D., Robinson, J.A. (2015), "The rise and decline of General laws of capitalism", *Journal of Economic Perspectives,* 29(1), 3-28.

Aghion, P., Holden, R. (2011), "Incomplete contracts and the theory of the firm: what we learned over past 25 years?, *Journal of Economic Perspectives,* 25(2), 181-197.

Bénassy-Quéré, A., Coeuré, B., Jacquet, P., Pisani-Ferry, J. (2014), *Politica economica, teoria e pratica,* Italian Ed.. Petretto, A, Russo, F.F (eds.), Il Mulino, Bologna.

Bolton, P.Dewatripont, M. (2005), *Contract theory,* The Mit Press, Cambridge, Ma.

Dahlby, B. (2008), *The Marginal cost of Public Funds, theory and applications,* The Mit Press, Cambridge, Ma.

Hart, O. (2003), "Incomplete contracts and public ownership: Remarks, and an application to Public-Private Partnership", *Economic Journal,* 113 (march), C69-C76.

Hart, O., Shleifer, A., Vishny, R.W. (1997), "The proper scope of government: theory and an application to prisons", *Quarterly Journal of Economcis,* 112(4), 1126-61.

Hindriks, J., Myles, G.D. (2013), *Intermediate Public Economics,* II ed., The Mit Press, Cambridge Ma.

Kaplow, L., Shavell, S. (2002),"Economic analysis of law", in A.J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics*, vol. 3, Amsterdam, North Holland, 661-1784.

Grazzini, L., Petretto, A. (2012), Earmarked Taxation for Financing Public Investments Projects, in M. Nicolai, *Primo rapporto sulla finanza pubblica. Finanza pubblica e federalismo*, Maggioli, Rimini, pp. 515 -- 537.

Glied, J., Smith, P.C. (2011), *The Oxford Health Economics handbook,* Oxford University Press, Oxford.

Motta, M. (2004), *Competition policy: theory and practice,* Cambridge University Press, Cambridge UK.

Persson, T. (1998), "Economic policy and special interest politics", *Economic Journal,* 108 (447), 310-327.

Persson, T., Tabellini, G. (2000), *Political economics, explaining economic policy,* The MIT Press, Cambridge Ma.

Petretto, A. (2013a), "On the fuzzy bounderies between Public and Private in Health-Care Organization and Funding systems", *Rivista di Politica Economica,* gennaio/marzo, 326-369.

Petretto, A. (2013b), "Need equalization transfers and productive efficiency of local governments", *Rivista Italiana degli Economisti,* vol. XVIII(1), 25-46.

Shavell, S. (2005), *Foundations of Economic analysis of law,* Belknap, Harvard.

Zweifel, P., Breyer, F., Kifman, M. (2009), *Health Economics,* Springer-Verlag, Berlin.

A. Petretto, *Economics of Institutions and Law*, ISBN (print) 978-88-6453-579-1, ISBN (onine) 978-88-6453-580-7, CC BY 4.0, 2017 Firenze University Press

MANUALI

Biomedica

Branchi R., *Le impronte nel paziente totalmente edentulo*

Branchi R., *Riabilitazione protesica del paziente oncologico testa-collo*

Martinelli E., *Rieducazione del piede. Problematiche biomeccaniche e posturali statiche e dinamiche*

Martinelli E., *Sviluppo del dolore rachideo in gravidanza. Mutamenti della biomeccanica rachidea, problematiche posturali, prevenzione e attività fisica adattata pre e post parto*

Martinelli E., *Rieducazione posturale. Fondamenti per la progettazione della postura*

Martinelli E., *Prevenzione del mal di schiena di origine meccanica con attività motoria e comportamentale. Approfondimenti di patomeccanica e biomeccanica rachidea*

Rossetti R., *Manuale di batteriologia clinica. Dalla teoria alla pratica in laboratorio*

Rucci L., *Testo Atlante di embriologia clinica della Laringe. La chirurgia conservativa compartimentale della regione glottica*

Rucci L., *Clinical Embryology of Human Larynx for Conservative Compartmental Surgery. A Text and Atlas*

Scienze

Bart J.C.J., *Polymer Additive Analytics. Industrial Practice and Case Studies*

Caramelli D., *Antropologia molecolare. Manuale di base*

Lo Nostro P., Peruzzi N., *Spontaneamente. Esercizi Risolti di Chimica Fisica*

Ricci R., *Lezioni di Sistemi Dinamici*, a cura di Lorenzo Barletti, Angiolo Farina, Lorenzo Fusi, Federico Talamucci

Scialpi A., Mengoni A. (a cura di), *La PCR e le sue varianti. Quaderno di laboratorio*

Simonetta M.A., *Short history of Biology from the Origins to the 20th Century*

Spinicci R., *Elementi di chimica* (nuova edizione)

Scienze Sociali

Ciampi F., *Fondamenti di economia e gestione delle imprese*

Giovannini P. (a cura di), *Teorie sociologiche alla prova*

Maggino F., *L'analisi dei dati nell'indagine statistica. Volume 1. La realizzazione dell'indagine e l'analisi preliminare dei dati*

Maggino F., *L'analisi dei dati nell'indagine statistica. Volume 2. L'esplorazione dei dati e la validazione dei risultati*

Magliulo A., *Elementi di economia del turismo*

Petretto A., *Economics of Institutions and Law*

Visentini L., Bertoldi, M., *Conoscere le organizzazioni. Una guida alle prospettive analitiche e alle pratiche gestionali*

Scienze Tecnologiche

Borri C., Pastò S., *Lezioni di ingegneria del vento*

Borri C., Betti M., Marino E., *Lectures on Solid Mechanics*

Cidronali A., Paolo Colantonio P. e Lucci L., *Antenne Integrate Attive*

Gulli R., *Struttura e costruzione / Structure and Construction*

Maddio S., *Introduzione ai Sistemi di Localizzazione Indoor*

Policicchio F., *Lineamenti di infrastrutture ferroviarie*

Umanistica

Bertini F., *Risorse, conflitti, continenti e nazioni. Dalla rivoluzione industriale alle guerre irachene, dal Risorgimento alla conferma della Costituzione repubblicana*

Bombardieri L., Graziadio G., Jasink A.M., *Preistoria e Protostoria egea e cipriota*

Bombi A.S., Pinto G., Cannoni E., *Pictorial Assessment of Interpersonal Relationships (PAIR). An analytic system for understanding children's drawings*

Baldini S., Marini D., Vorrei. *Corso di lingua italiana di livello elementare 1. Libro di testo e libro degli esercizi*

Borello E., Mannori S., *Teoria e tecnica delle comunicazioni di massa*

Brandi L., Salvadori B., *Dal suono alla parola. Percezione e produzione del linguaggio tra neurolinguistica e psicolinguistica*

Coniglione F., Lenoci M., Mari G., Polizzi G. (a cura di), *Manuale di base di Storia della filosofia*

Marcialis N., *Introduzione alla lingua paleoslava*

Michelazzo F., *Nuovi itinerari alla scoperta del greco antico. Le strutture fondamentali della lingua greca: fonetica, morfologia, sintassi, semantica, pragmatica*

Peruzzi A., *Il significato inesistente. Lezioni sulla semantica*

Peruzzi A., *Modelli della spiegazione scientifica*

Sandrini M.G., *Filosofia dei metodi induttivi e logica della ricerca*

Trisciuzzi L., Zappaterra T., Bichi L., *Tenersi per mano. Disabilità e formazione del sé nell'autobiografia*