1 Economic theories of regulation

Johan den Hertog

1. INTRODUCTION

There are two broad traditions with respect to the economic theories of regulation. The first tradition assumes that regulators have sufficient information and enforcement powers to effectively promote the public interest. This tradition also assumes that regulators are benevolent and aim to pursue the public interest. Economic theories that proceed from these assumptions are therefore often called ‘public interest theories of regulation’. Another tradition in the economic studies of regulation proceeds from different assumptions. Regulators do not have sufficient information with respect to cost, demand, quality and other dimensions of firm behavior. They can therefore only imperfectly, if at all, promote the public interest when controlling firms or societal activities. Within this tradition, these information, monitoring and enforcement costs also apply to other economic agents, such as legislators, voters or consumers. And, more importantly, it is generally assumed that all economic agents pursue their own interests, which may or may not include elements of the public interest. Under these assumptions there is no reason to conclude that regulation will promote the public interest. The differences in objectives of economic agents and the costs involved in the interaction between them may effectively make it possible for some of the agents to pursue their own interests, perhaps even at the cost of the public interest. Economic theories that proceed from these latter assumptions are therefore often called ‘private interest theories of regulation’.

Fundamental to public interest theories are market failures and efficient government intervention. According to these theories, regulation increases social welfare. Private interest theories explain regulation from interest group behavior. Transfers of wealth to the more effective interest groups often also decrease social welfare. Interest groups can be firms, consumers or consumer groups, regulators or their staff, legislators, unions and more. The private interest theories of regulation therefore overlap with a number of theories in the field of public choice and may effectively turn into theories of political actions. Depending on the efficiency of the political process, social welfare either increases or decreases. The second section of this chapter discusses these general public and private interest theories of regulation, and the criticisms that have been leveled at them.
The general theories of regulation were to a large extent developed to explain government intervention in a number of sectors of the economy. Important changes have recently taken place in the regulation of fundamental sectors of the economy such as electricity and gas, electronic communications, water and sewerage, postal services and transport (airports and airlines, railways, busses). The services provided by these sectors are often essential for both businesses and consumers. Interruption in the supply of these services will slow down economic activities, bring a stop to interactions taking place and will create risks to life and health in several parts of society. The production technologies in those sectors are such that monopolistic firms naturally arise. For these and other reasons, such industries have in the past either been regulated, as for example in the US, or they have been organized under the control of the state in the form of state-owned enterprises, as for example in the European Union. The privatization and restructuring of the state-owned enterprises in Europe necessitated other forms of control. The third section of this chapter analyzes the processes of liberalization, restructuring and re-regulation. The former state-owned enterprises were often integrated firms which provided the production, distribution and sale of, for example, electricity or telecommunication services. The perception was that competition could be introduced in some of these stages because they were thought to be inherently competitive. Examples are the production and sales stage of electricity or gas supply and the provision and sale of internet and telephone services. The integrated firm thus had to be restructured into separate firms for the different stages in the supply chain. Liberalization meant the lifting of the legal entry barriers for the competitive stages in the supply chain as a result of which markets could develop. As to the network parts of the supply chain, such as the cables, pipes, wires, and tracks competition was not expected to arise, or perhaps only in the long run. After restructuring, those parts of the formerly state-owned enterprises therefore had to be re-regulated. In the third section, I will first review explanatory theories of the processes of privatization, restructuring and liberalization. Second, I will analyze the different types of regulation that have been applied in practice to control the network or monopolistic parts of the industries involved.

This chapter concludes with a look at future developments in regulation.

2. GENERAL ECONOMIC THEORIES OF REGULATION

In the legal and economic literature, there is no fixed definition of the term ‘regulation’. Some researchers consider and evaluate various definitions and attempt through systematization to make the term amenable to further analysis (Baldwin and Cave, 1999; Morgan and Yeung, 2007; Ogus, 2004a). Others
abstain almost entirely from an exact definition of regulation (Ekelund, 1998; Joskow and Noll, 1981; Spulber, 1989; Train, 1997). In order to delineate the subject and because of the limited space, a further definition of regulation is nevertheless necessary. In this chapter, regulation will be taken to mean the employment of legal instruments for the implementation of social-economic policy objectives. A characteristic of legal instruments is that individuals or organizations can be compelled by government to comply with prescribed behavior under penalty of sanctions. Corporations can be forced, for example, to observe certain prices, to supply certain goods, to stay out of certain markets, to apply particular techniques in the production process or to pay the legal minimum wage. Sanctions can include fines, the publicizing of violations, imprisonment, an order to make specific arrangements, an injunction against withholding certain actions, divestiture of businesses or closing down the business.

A distinction is often made between economic and social regulation, for example by Viscusi, Vernon and Harrington (2005). Two types of economic regulations can be distinguished: structural regulation and conduct regulation (Kay and Vickers, 1990). Structural regulation concerns the regulation of the market structure. Examples are restrictions on entry or exit, and rules mandating firms not to supply professional services in the absence of a recognized qualification. Conduct regulation is used to regulate the behavior of producers and consumers in the market. Examples are price controls, the requirement to meet all demand at the regulated price, the labeling of products, rules against advertising and minimum quality standards. Economic regulation is mainly exercised on so-called natural monopolies and market structures with imperfect or excessive competition. The aim is to counter the negative welfare effects of dominant firm behavior and to stabilize market processes. Social regulation comprises regulation in the area of the environment, occupational health and safety, consumer protection and labor (equal opportunities and so on). Instruments applied here include regulations dealing with the discharge of environmentally harmful substances, safety regulations in factories and workplaces, the obligation to include information on the packaging of goods or on labels, the prohibition of the supply of certain goods or services unless in the possession of a permit and banning discrimination on grounds of race, skin color, religion, sex, or nationality in the recruitment of personnel.

The economic literature distinguishes between positive and normative economic theories of regulation. The positive variant aims to provide economic explanations of regulation and to provide an effect-analysis of regulation. The normative variant investigates which type of regulation is the most efficient or optimal. The latter variant is called normative because there is usually an implicit assumption that efficient regulation would also be desirable; for the distinction between positive and normative theories, see the discussion between Blaug (1993) and Hennipman (1992).
This chapter focuses on general explanatory and predictive economic theories of regulation. In this respect two preliminary remarks are in order. First, the mainstream economic literature is implicitly or explicitly critical of public interest theories of regulation. These theories are often thought to be ‘normative theories as positive analysis’ (Joskow and Noll, 1981), implying that the evaluative theoretical and empirical analysis of markets has been used to explain actual regulatory institutions in practice. The public interest theories of regulation are described as rationalizing existing regulations, while private interest theories are discussed as theories that explain existing regulation (for example Ogus, 2004a). According to some other authors, either there is no such thing as public interest theories of regulation or they are a misinterpretation and have lost their validity (Hantke-Domas, 2003; Hägg, 1997). To have a proper discussion on the evaluation and appraisal of economic theories of regulation, it would be desirable to explicitly proceed from evaluation criteria that have been developed and are the subject of debate in the methodological literature on the appraisal of theories (Dow, 2002; Blaug, 1992). Some of these criteria would be, for example, internal consistency, empirical corroboration, plausibility and more. By making the evaluation criteria explicit, the appraisal of economic theories of regulation would become more precise and explicit. The second remark pertains to the concept of regulation. A distinction is often made between legislation and regulation. Usually in legislation regulatory powers are allocated to lower level institutions or officials. The result of the use of such powers by these officials or institutions is then called regulation. Within the perspective of some explanatory theories, the distinction between regulation and legislation does not always add much additional explanatory or predictive value to regulatory theories. The explanatory power of a market failure as a driving force of public interest regulation, for example, does not really depend on whether decision-making powers have been centralized or decentralized. From other perspectives, the distinction is important. The explanatory power of variables like rent-seeking and capture may differ according to the level of regulatory decision-making. According to some theories, delegation may help to prevent inefficient rent extraction by politicians (Shleifer and Vishny, 1998). According to others, that is where the problems of unaccountable regulators begin (Martimort, 1999). In the rest of this chapter, I will use the term regulation irrespective of its centralized or decentralized character and only introduce the distinction between regulation and legislation if this distinction is crucial for the particular hypothesis or theory in question. For a perspective on the use of the term regulation and its meaning in this respect, see Jordana and Levi-Faur (2004).

2.1 Public Interest Theories of Regulation

2.1.1 Introduction
The first group of regulation theories proceeds from the assumptions of full information, perfect enforcement and benevolent regulators. According to these theories, the regulation of firms or other economic actors contributes to the promotion of the public interest. This public interest can further be described as the best possible allocation of scarce resources for individual and collective goods and services in society. In western economies, the allocation of scarce resources is to a significant extent coordinated by the market mechanism. In theory, it can even be demonstrated that, under certain circumstances, the allocation of resources by means of the market mechanism is optimal (Arrow, 1985). Because these conditions frequently do not apply in practice, the allocation of resources is not optimal from a theoretical perspective and a search for methods of improving the resource allocation arises (Bator, 1958). This situation is described as a market failure. A market failure is a situation where scarce resources are not put to their highest valued uses. In a market setting, these values are reflected in the prices of goods and services. A market failure thus implies a discrepancy between the price or value of an additional unit of a particular good or service and its marginal cost or resource cost. Ideally in a market, the production by a firm should expand until a situation arises where the marginal resource cost of an additional unit equals its marginal benefit or price. Equalization of prices and marginal costs characterizes an equilibrium in a competitive market. If costs are lower than the given market price, a firm will profit from a further expansion of production. If costs are higher than price, a firm will increase its profits by curtailing production until price again equals marginal cost. A market equilibrium, and more generally an equilibrium of all markets, is thus a situation of an optimal allocation of scarce resources. In this situation supply equals demand and under the given circumstances market players can do no better. A great number of conditions have to be satisfied for an optimal allocation in a competitive market economy to exist (see generally Boadway and Bruce, 1984).

One of the methods of achieving efficiency in the allocation of resources when a market failure is identified, is government regulation (Arrow, 1970, 1985; Shubik, 1970). In the earlier development of the public interest theories of regulation, it was assumed that a market failure was a sufficient condition to explain government regulation (Baumol, 1952). But the theory was soon criticized for its Nirvana approach, implying that it assumed that theoretically efficient institutions could be seen to efficiently replace or correct inefficient real world institutions (Demsetz, 1969). This criticism has led to the development of a more serious public interest theory of regulation by what has been
variously referred to as the ‘New Haven’ or ‘Progressive School’ of Law and Economics (Ogus, 2004a; Rose-Ackerman, 1988, 1992; Noll, 1983, 1989a). In the original theory, the transaction costs and information costs of regulation were assumed to be zero. By taking account of these costs, more comprehensive public interest theories developed. It could be argued that government regulation is comparatively the more efficient institution to deal with a number of market failures (Whynes and Bowles, 1981). For example, with respect to the public utilities, it could be argued that the transaction costs of government regulation to establish fair prices and a fair rate of return are lower than the costs of unrestricted competition (Goldberg, 1979). Equally, it could be argued that social regulation in some cases would be a more efficient institution to deal with pollution of the environment or accidents in the workplace than private negotiations between affected parties. Regulators would not be plagued by failures in the information market and they could more easily bundle information to determine the point where the marginal cost of intervention equals the marginal social benefits (Leland, 1979; Asch, 1988).

These more serious versions of the public interest theories do not assume that regulation is perfect. They do assume the presence of a market failure, that regulation is comparatively more efficient than other institutions and that deregulation takes place when more efficient institutions develop. These theories also assume that politicians act in the public interest or that the political process is efficient and that information on the costs and benefits of regulation is widely distributed and available (Noll, 1989a). The core of this basic framework is captured in Figure 1.1.

Imagine an unregulated natural monopoly firm supplying public utility services. The firm makes supernormal profits, charges different prices to different consumer groups and does not supply services to high-cost consumers in rural areas. Economic theory predicts an inefficient allocation of resources. Without regulatory intervention, these costs are at their highest at the point where the EL curve intersects with the vertical axe (intersection not visible). Intervening in the market results in a decline in these welfare costs. The stronger the level of intervention, the lower will be the welfare losses in the private sector. The naïve public interest theory of regulation, for example, would explain ‘fair rate of return’ regulation in terms of the presence of the natural monopoly firm. Prices must decline and production increase until societal resources are allocated efficiently. The more complex public interest theories of regulation take the costs of regulatory intervention into account. The more a regulator intervenes in the private operation of the firm, the higher the intervention costs will be (curve IC). The regulator must have information on the cost and demand facing the firm before efficient prices can be determined. There will be compliance costs for the firm in terms of time, effort and resources. It will have to comply with procedures, adapt its administration and
incur productivity losses. Once put into practice, the cost of monitoring firm behavior and enforcement of the regulations arises. It is to be expected that the firm will behave strategically and conceal or disguise any relevant information from the regulator. Furthermore, indirect costs are to be expected. The less profit the firm makes, the lower the effort in decreasing production costs or in developing new products and production technologies. Also less tangible effects are predicted. Regulatory intervention makes private investments less secure: risk premiums rise, investments decline and economic growth will slow down. The regulator is aware of these costs and has several options to choose from: it could, for example, regulate prices or profits or a combination of both. Whichever it chooses, there will be different intervention costs and different consequences for static and dynamic efficiency. The optimal level of intervention ($I_{opt}$) implies trading off resources allocated to increasing levels of regulatory intervention and decreasing levels of inefficient firm behavior. Complicating the policy options further, for politicians there are alternatives to the regulation of prices, profits, service levels, etc. The legislator could also decide to franchise an exclusive right to operate the market or set up a public enterprise to maximize welfare. Again, these institutions require different costs of intervention and have different effects in terms of static and dynamic efficiency or other policy goals. Amongst other things, they differ with respect to informational requirements, administrative costs, the burden for the private sector, including the cost of errors, distributional effects,
governance, accountability, risks of capture and corruption, and more. The public interest theories of regulation thus basically assume a comparative analysis of institutions to have taken place to efficiently allocate scarce resources in the economy.

Equivalent reasoning applies to the field of social regulation. Imagine that lifting weight, for example a patient in a hospital or cement in the construction sector, leads to back trouble or even work disability. Employees are often not very well aware of the risks they run, and even if they are, they will find it difficult to deal with small risks such as 0.0001. The costs involved however, may be considerable in terms of medical costs, lost earnings and risk of injury and pain, and consequences for relatives and friends. Inefficiency in the allocation of resources in the absence of regulation is again depicted by the curve EL. A regulator may decide, for example, to regulate maximum weights. She needs to identify the potential risk involved, how this risk varies with exposure to lower weights and different circumstances. Then the maximum allowable weight lifting must be determined. The regulator knows that increasing levels of intervention or standard setting will increase costs (curve IC). The more detailed and precise, the higher the regulatory costs. The higher the weight standard, the higher the compliance costs will be: more nurses in hospitals and increasing use of capital equipment in the construction sector. Indirect costs will also increase with the level of intervention: there will be a lower ratio of input to output and substitution between now comparatively higher priced labor and capital equipment. Not only will employment decline, but so will the speed of technical change. The setting of the standard lowers the incentive to seek for technologies to further prevent lifting costs below the standard. Again, the regulator is aware of these costs and has several options to choose from. It could set an output or performance standard limiting the number of incidents. It could prescribe an input standard by specifying the use of certain care technologies or machinery. Alternatively, it could set a target standard that imposes criminal liability for certain harmful consequences or it could impose process standards procedures, obliging the firm to identify the risks and deal with them. All these forms of intervention have different intervention costs and compliance costs and different effects in terms of static and dynamic efficiency or other policy goals. The optimal standard or level of intervention depicted in the diagram is \( I_{\text{opt}} \). And again, complicating matters further, for political decision-makers there are alternative institutions to regulation, such as providing the firm and the employees with information and having private law and tort liability deal with any costs involved or, in cases of severe dangers to life and health, a prohibition on the use of certain techniques, equipment or materials.

Four types of market failures can be distinguished as driving forces of public interest regulation. Discrepancies between values and costs of allocated
resources can arise as a result of imperfect competition, unstable markets, missing markets or undesirable market results. Imperfect competition will cause prices to deviate from marginal resource cost. Unstable markets are characterized by dynamic inefficiencies with respect to the speed with which these markets clear or stabilize. These instabilities waste scarce resources. Missing markets imply the demand for socially valuable goods and services for which total value exceeds total cost but where prices or markets do not arise. And finally, even if the competitive market mechanism allocates scarce resources efficiently, the outcomes of the market processes might still considered to be unjust or undesirable from other social perspectives. I shall discuss these market failures in turn, assuming in the first instance that the government acts as an omniscient, omnipotent and benevolent regulator (Dixit, 1996).

2.1.2 Imperfect competition

In the first place, an efficient market mechanism implies certain rules and regulations and assumes that individual property rights are established, allocated and protected and that freedom to contract exists and is enforced (Pejovich, 1979). Not only are the costs of market transactions reduced by property and contract law, the protection of property rights and the enforcement of contract compliance is also more efficiently organized collectively than individually. The freedom to contract can, however, also be used to achieve cooperation between parties opposed to efficient market operation. Agreements between producers to keep prices high and supply artificially low quantities will give rise to prices that deviate from marginal costs. A dominant position by one or a few firms also gives rise to prices departing from marginal costs. Antimonopoly legislation is aimed at maintaining efficient market operation through merger control and by prohibiting anticompetitive agreements or behavior. More important for this review are the special characteristics of certain products and production processes in sectors such as energy, telecommunications, transport, postal services and water. Much of so-called economic regulation relates to these sectors. In order to explain some of the market failures in these fields, I will make use of Figure 1.2, where a simplified market situation is depicted with a typical single product firm in such a sector. The figure shows the declining average cost curve AC of a typical firm and the market demand curve D. Marginal costs are assumed to be constant and equal to \( P_{mc} \). The market demand curve D or average revenue (AR) intersects with the declining part of the average cost curve of the firm, which implies that average cost is minimized if production is concentrated in one firm. If several companies with the same production technology produce the same total quantity, the unit costs of production rise. For this reason, this situation is called a natural monopoly. For the more precise conditions of a natural monopoly, see Baumol et al. (1982).
An example of how such a situation arises is when the production process requires a great deal of sunk capital for the installment of a distribution network. You can think of a network of pipes, wires, cables or railroads to supply gas, electricity, telecommunications, television signals, or transport services respectively. In those cases, the cost per service (average cost) continues to decline as the production of goods and services increases (Baumol, 1977). In an unrestricted market, several failures will appear. First of all, a firm contemplating entering the market will realize that entry will provoke price competition from the incumbent firm, driving prices down to marginal costs. If that is the case, the sunk costs necessary to set up production cannot be recovered and the firm will decide not to enter. If the firm thinks it is more efficient than the incumbent firm, so-called ‘wars of attrition’ will arise, leading to bankruptcy and waste of scarce resources. Second, if the incumbent firm is indeed a natural monopolist, it will maximize profits by equalizing marginal costs and marginal revenues ($P_m$ in the diagram). Under the usual assumptions of inability to price discriminate and inability to prevent arbitrage, the firm will limit production (to $Q_m$) and set profit-maximizing prices to clear the market. As a result, prices will deviate significantly from marginal costs (Train, 1997). Next, the firm knows that if it has invested huge amounts of

Figure 1.2 Natural monopoly and its tradeoffs
sunk capital, it will be vulnerable to expropriation by political decision-makers (Newbery, 2001). Third parties will rationally expect the firm not to exit the market as long as prices are above marginal cost, even though they are insufficient to recoup the fixed costs of the network. As a result of these political risks, risk premiums will rise and the firm will under-invest (Sidak and Spulber, 1998). Furthermore, entry is expected to take place, not only when these markets develop or in highly profitable parts of the market, but also where firms consider themselves to be more efficient than the incumbent firm. Productive inefficiency is the result and cut-throat competition will appear. For a number of these natural monopolies, history has shown such events to take place (Kahn, 1988; Priest, 1993; Vietor, 1989, 1999). Finally, from a social point of view, market outcomes will be discriminatory or unjust. High-cost consumers will not be served, consumers who cannot switch to alternative supply will pay discriminatory high profit-maximizing prices and the firm will make huge profits at the consumer’s expense.

To promote a more efficient and equitable allocation of scarce resources, these natural monopolies are either put under the control of the state, as happened in many European countries, or highly regulated, as for example is the practice in the United States. In the former case, these firms are instructed to maximize welfare instead of profit. In the latter case, regulation consists of stopping entry and enforcing price or profit rules that promote an efficient and equitable allocation of resources (Schmalensee, 1979; Braeutigam, 1989). For example, prices are set equal to average cost by means of rate of return regulation (price \( P_{ac} \) in Figure 1.2) and a price structure is determined such that the firm breaks even. Ideally, the optimal price would of course be a price according to marginal resource cost. This, however, would lead to financial losses, depicted by the difference between \( P_{mc} \) and \( AC_{opt} \). At \( Q_{opt} \) the cost per unit is \( AC_{opt} \), while the revenues per unit are only \( P_{mc} \), which in the diagram is about 30% too low to cover costs. The regulator thus has to choose between the inefficiencies of subsidies or a price that covers cost, but that is not ‘first-best’ efficient. Actual regulations taking into account limitations on information and enforcement are discussed in Section 3.2.1 (Privatization and regulatory reform), and following.

### 2.1.3 Market instabilities

A more efficient allocation of resources may be accomplished not only by a redress of imperfectly competitive markets but also by stabilizing inherently unstable markets. Imbalances within an economy occur at the level of separate markets and on a macro level. In separate markets, destructive or excessive competition may arise, often as a result of long-term over-capacity. The establishment of a new equilibrium may take a long time if the individual market players are in a prisoner’s dilemma. For all market parties jointly, efficiency
is achieved if the existing over-capacity is rationalized. For an individual producer, however, the ‘sunk costs’ may imply that it is rational to wait until other suppliers have withdrawn from the market. Because this consideration applies to all producers, over-capacity can persist for a considerable time. Over-capacity situations may also arise when production capacity is adjusted to demand at peak moments or peak periods. Examples are peak loads in the rush hour (electricity, electronic communication, busses, underground railways and trains), during the harvest in agriculture (trucks) and during the tourist high season (touring cars, aircraft). Figure 1.3 illustrates this instability.

Assume that because of some shortage in supply in a former period, prices have risen to \( P_1 \), which is somewhat above the market equilibrium price \( P_e \). At that price, supply is \( Q_1 \). However, at that quantity, the market can only clear at price \( P_2 \). Suppliers will react to that price by supplying only \( Q_2 \) units in the next period. However, at this quantity, the market will clear only at a very high price, motivating suppliers in the next period to supply increasingly more, etc. Excessive or ruinous competition may also arise in a natural oligopoly setting. In a natural oligopoly situation, productive efficiency is achieved if only a few

![Figure 1.3 Dynamic inefficiency of an unstable market](image-url)
companies supply the market. The small number of companies allows each of them to react to each other’s market strategies. One of the outcomes may be a persistent price war. Effects are that prices may decline below average cost and that price dispersion is increased. Both effects create inefficiencies in the allocation of resources or in consumer decision-making. Furthermore, excessive competition may be detrimental to safety and reliability when consumers cannot observe or verify the quality of goods and services (Kahn, 1988, pp. 172–8). In the past, regulatory practices assumed that situations of excessive competition applied to sectors such as air travel, passenger transport, freight or transport by water (trucks, taxis, shipping). For these sectors, business licensing restrictions were devised and capacity was rationed, sometimes in combination with minimum price regulation. Regulatory theories however, considered the collection of excessive competition rationales for government intervention to be ‘an empty box’ (see Breyer, 1982, pp. 29–32) and regulations might better be explained from a private interest perspective. Recently, modern regulatory theories developed several instances of structural market failures in potentially competitive sectors (Telser, 1978, 1994, 1996; Button and Nijkamp, 1997). Free entry may result in too much entry if costs are sunk in the form of production facilities or marketing expenditures. The increase in competition and the resulting decrease in consumer prices may not add up to the increase in the resource costs that entry requires. There may also be insufficient entry if consumers value product diversity but firms do not enter because profit opportunities decrease with increased entry (Mankiw and Whinston, 1986; Perry, 1984). Apart from the level of individual sectors, imbalances may also occur at a macroeconomic level. Market economies are characterized by a business cycle, the regular alternation of periods of increasing and declining economic activity. In the course of the business cycle, a self-sustaining process develops in the product market that is not compensated by adjustments in the labor market. This arises partly because of lack of information, long-term labor contracts and efficiency wages. Trade-cycle policies can be desirable to prevent temporary disturbances which have permanent effects. For example, specialized investments with limited alternative value in other market segments may be permanently lost in a recession. Also, structural unemployment may arise when unemployed workers lose their skill and motivation. Finally, stabilization of the business cycle may be desirable to prevent the decline of production and employment such that different social groups are unequally affected by the economic rise and fall. But, of course, all these regulations have costs, which have to be compared to the benefits in terms of increases in social welfare.

Traditionally, trade-cycle policies are put into effect together with instruments of budgetary and monetary policy; for an overview of the significance of these instruments and the underlying theories, see Snowdon, Vane and
Wynarczyk (2007). Because these instruments are not directed at specific sectors and can only take effect after some time, wage and price regulation has been developed in some market economies. To combat a wage-price spiral, governments have developed the legal means to freeze wages and prices for a period of between a half to one year, if necessary applicable in designated sectors (Ogus, 2004a, pp. 300 ff and Breyer, 1982, pp. 60 ff). Of course, since market economies have become global, these instruments, while still existing as legal powers, have largely become obsolete in practice.

### 2.1.4 Missing markets

**Information problems and bounded rationality** For a number of reasons, markets may be ‘missing’ for some goods and services for which utility or ‘willingness to pay’ exceeds production costs. Missing markets may be the result of information problems and transaction costs. These problems and costs justify much of the social regulation that efficiently aims to protect the worker, the consumer and society at large. In practice, the full information assumption of a perfectly competitive economy is rarely found. With respect to the information that is available on goods and services in a market, it is useful to make a distinction between ‘search goods’, for which the quality of a product can be determined prior to purchase, ‘experience goods’, for which quality only becomes apparent after consumption of the good and ‘credence goods or trust goods’, for which the quality cannot even be established after consumption (see generally Beales et al., 1982; Armstrong, 2008; Nelson, 1970; Darby and Karni, 1973; Dulleck and Kerschbamer, 2006). Examples of each are the purchase of flowers, second-hand cars and repair firms, respectively. Consumers and workers have an interest in being informed on relevant aspects of their purchases and of their jobs, such as the safety and health dimensions. In a competitive market, employers and producers also have an interest in revealing the relevant characteristics of jobs and products. However, the ‘information market’ is characterized by market failures and these failures spill over to the market for goods and services (Hirshleifer and Riley, 1979). On the demand side, information will be searched for until the marginal cost equals marginal benefit. However, since search intensifies competition, it produces external benefits for uninformed parties. As a result, information is searched for in suboptimal amounts from a social perspective. Furthermore, at a general level, information will be undersupplied. The production of information is costly, but the dissemination is not. If competition drives prices down to marginal distribution costs, it will be difficult to cover the fixed cost of production. Also, the market will undersupply the optimal amount of information when producers are not able to fully appropriate all the revenues from their investments. Examples are investments in the knowl-
edge of the health and safety dimensions of chemical substances. And certainly, information will not be supplied when it is not in the interest of the sector itself to do so, as is the case in situations of collective ‘bads’ such as smoking hazards. When it is not possible to establish the relevant quality dimensions of particular goods or services in advance, purchasers will be prepared to pay an average price corresponding to the average expected quality. Sellers of high quality products will not be prepared to sell at that asking price, and will withdraw from the market. The end-result is that the quality of goods and services will decline, as will the price buyers are prepared to pay (Akerlof, 1970). In this process of adverse selection, high quality goods are driven from the market by low-quality goods. In addition, the asymmetric distribution of information can also give rise to moral hazard in the enforcement of contracts, which means that parties take advantage of their information lead. Examples would be food processors who use poor quality food and lawyers who give unfounded advice.

Figure 1.4 illustrates how valuable products or services may disappear from the market and how a market is finally missing. Imagine high quality car repair firms who are offering their services \( S_1^h \), and drivers who cannot establish the quality of the repairs. Had they known the quality of the services supplied, they would have bought \( Q_1 \) units. Drivers, however, are not informed of the quality of the firms and just know there are high and low quality firms. Visiting a repair firm, a driver expects ex ante the services supplied to be of average quality. They are willing to pay for the services according to the expected average quality, which is determined by the ratio of high and low quality firms in the market. Their actual market demand is thus not \( D_1^h \), but \( D_2^h \). At that price, only \( Q_2 \) units are supplied in the market, so the ratio of high and low quality firms decreases. Eventually, customers will come to learn the change of this ratio in the market and adapt their demand accordingly. The demand for high quality services shifts to \( D_3^h \). Professional firms who supply high quality and high cost services cannot supply these services at the lower market price and will leave the market. Now the supply of high quality services decreases to \( S_2^h \).

The final result is a missing market for high quality services, although high quality suppliers are willing to sell high quality services, which consumers are willing to buy. Of course, suppliers are interested in communicating the quality of their work, particularly if it is higher than average. For example, they will try to create a reputation for delivering high quality services. But that solution only works if consumers purchase these services regularly or if they can profit from ‘hearsay’. For some experience or credence goods, this will not be the case. But even if information markets worked perfectly, some market failures would keep appearing, particularly in the fields of health and safety, but also in other markets for complex products.
Bounded rationality Consumers and workers are assumed to make rational, welfare-maximizing choices. In this respect, a distinction should be made between the rationality of the outcome and the rationality of the decision-making process leading to that outcome. If consumers or workers experience limits in absorbing and evaluating information, they will adapt their preferences and their decision-making processes accordingly. Given those limits, the outcome of the decision-making process may be defined as ‘boundedly’ rational. Further research along these lines has led to the development of a new branch of economics, behavioral economics which incorporates insights from psychology and sociology into economics (Dellavigna, 2009; McFadden, 1999; Rabin, 2002; Mullainathan and Thaler, 2001).

To be able to make the relevant tradeoffs and to decide fully rationally, consumers and workers should be able not only to fully understand all the relevant characteristics of the products and jobs at hand, but they also must be able to understand and evaluate future developments that will have an impact on decisions that are currently taken. According to some scholars, three condi-
tions must be met for rational behavior to occur (Poiesz, 2004). Economic subjects must be motivated to make rational decisions, they must have the capacity to make such decisions and they must have the opportunity to decide rationally. Whether or not these conditions are met will depend on a variety of circumstances such as the characteristics of the products, activities or workplaces involved and the nature of the competitive process. Products and workplaces may, for example, be characterized by the degree of risk involved, health risks, financial risks or other. Research has shown that consumers and workers find it hard to adequately deal with risks (Thaler, 1992; Margolis, 1996). Accidents with products or risks to life and health at work or on the road are usually low probability events of the order of, for example, 0.0001 or perhaps even one in a million. Most people will find it hard to think and act rationally about such events. Small risks are often overestimated and larger risks are underestimated. Preferences appear to be anomalous with respect to decisions concerning the future or uncertainty. Willingness to pay studies reveal that workers or consumers are willing to pay a certain amount to lower health risks by a certain percentage, but at the same time are willing to pay a far larger amount to bring forth an equivalent percentage reduction toward a situation of zero risk. More generally, the following distortions have been noticed in the literature (Dellavigna, 2009; Sunstein, 2002, pp. 33–53):

- people tend to think events are more probable if they can recall an incident occurring (availability heuristic);
- people tend to believe that something is either safe or unsafe and that it is possible to abolish risk entirely: exposure to dangerous substances always implies cancer and natural chemicals are less harmful compared with manmade chemicals (intuitive toxicology);
- certain beliefs become generally accepted because people simply accept other people’s beliefs (social cascades);
- people often focus on small bits of complex problems without looking at causal changes: a ban on asbestos may cause manufacturers to use even less safe substitutes (over-simplification);
- small dangers are easily noticed, benefits are minimized: ‘better safe than sorry’ (perceptual illusion);
- people have an emotional, mental short-cut reaction to certain processes and products: high benefits and low risk tend to become aligned mentally (affect heuristic);
- people’s reactions to risks are often based on the ‘worst case’ outcome rather than on the probability of its occurrence (alarmist bias);
- people worry more about proportions than about numbers: they worry more about a risk of one in a hundred than about a risk of one in a
million, even though the first group consists of, for example, 1000 persons and the second of 200 million people (proportionality effect);

- evidence suggests that people assess their willingness to pay very differently when taken in isolation (just skin cancer) compared to assessments in cross-category comparisons: willingness to pay to prevent skin cancer and to protect coral reefs (separate evaluation incoherence).

The general implication is that ordinary people make mistakes: they rely on mental shortcuts, they are subject to social influences that lead them astray and they neglect tradeoffs (Sunstein, 2002; Breyer, 1993; Viscusi, 1998). The conclusion is that people are only boundedly rational and that consequently the allocation of resources driven by misguided or mistaken decisions will be inefficient (Simon, 1948; Kreps, 1998). These problems may be exacerbated if regulators act upon such preferences to change the allocation of resources.

The problems of bounded rationality, adverse selection and moral hazard may explain the existence of, for example, the introduction of private or public certificates, minimum quality standards for the safety of food, toys, cars etc., licenses and other trading regulations for professional groups such as building contractors, hairdressers and plasterers, and more. By means of these rules, minimum requirements can be set on commercial knowledge, professional skill and creditworthiness, and more so that the transaction costs decline and the information problems are reduced (Leland, 1979; Shapiro, 1986; Zerbe and Urban, 1988). Rules relating to misleading information aim to minimize the cost of moral hazard (Beales et al., 1981; Schwartz and Wilde, 1979). Because of the nature of credence or trust goods, it is difficult to set minimum quality standards precisely in those cases where the risks of moral hazard are high. In such circumstances, legally sanctioned self-regulation can combat the problems of adverse selection and moral hazard (Van den Bergh and Faure, 1991; den Hertog, 1993). Not only do those involved have a vested interest in the maintenance of a minimum quality, they are also better able to formulate and maintain quality rules (Gehrig and Jost, 1995). If purchasers cannot establish quality levels, minimizing search costs by a ban on advertisement may actually be efficient (Barzel, 1982, 1985). Problems of adverse selection and moral hazard arise particularly in insurance markets (Rothschild and Stiglitz, 1976). Insured parties have superior information available with respect to the incidence of risks, but they lack information regarding the quality and independence of intermediaries. In many countries, social legislation is introduced as a reaction to these problems, and rules are established for intermediaries. It is a matter of empirical research to determine if those regulations are more efficient than market solutions, such as developing a reputation, or private law solutions, such as tort liability and self-regulation.

However, as Breyer (1993) has suggested, if preferences are distorted, a
vicious circle of public perceptions, parliamentary action and regulatory methods may arise. In economics, revealed or stated preferences are the main explanatory device and evaluative measure of regulatory practices. Justification of regulation or the explanation of regulation from a market failure perspective becomes difficult if preferences are distorted (Adler and Posner, 2001; Sunstein, 2002). Hence, paternalist theories of government intervention become once more the object of debate and scientific research (Ogus, 2005; Glaeser, 2006; Thaler and Sunstein, 2008).

Externalities and public goods In addition to information failures, prohibitively high transaction costs may also result in missing markets. For example, transaction costs can impede the development of a market for the efficient use of environmental goods. In a market economy, resources are efficiently used when the production of goods is increased until the marginal costs equal the marginal benefits of production. In a market with perfect competition, an individual producer aiming to maximize profit will increase its production until the marginal costs equal the market price. However, an inefficient allocation of resources can arise in the presence of externalities (Meade, 1973). Externalities are cost or utility effects for third parties outside the market interactions where these external effects develop. An often cited textbook example concerns the discharge of waste material by a factory such that downstream drinking water companies must incur the costs of water purification. Because the private costs for the discharging manufacturer differ from the social costs, production will be increased further than would be desirable from the point of view of the efficient allocation of resources. According to the Coase theorem, an efficient allocation of resources can nonetheless result from a process of negotiation in the case of clearly defined property rights and in the absence of transaction costs (Coase, 1960). However, information costs, bargaining costs or enforcement costs may prevent efficient solutions. Negotiation costs can be prohibitively high if several parties are involved in the negotiating process or if elements of strategic behavior are present (Veljanovski, 1982). Furthermore, there may be situations where the aggregated damage is large, while the damage per person may be too small for individuals to organize and participate in any action. Also, it may not even be known who causes the damage, as for example in the case of secret oil dumping at sea, or it may not be known which substances causes the damage (acid rain). The damage may also manifest itself years after exposure, as in the case of cancer from asbestos or future generations may be the main victims. In those cases, it will be difficult to prove causality between the negligent act and damages. Finally, even if held liable, firms may not be able to financially compensate for the damages. In all such cases, a private law failure is added to the market failure and regulation may be the more efficient solution if its costs are lower than the benefits in
terms of welfare loss control (Gruenspecht and Lave, 1989). Examples of such internalization of social costs are safety regulations for items such as automobiles and food, noise levels for aircraft, the obligation to use catalytic converters in automobiles and limits on the emission of hazardous substances permitted.

Markets may also be missing for goods and services with public goods characteristics (Samuelson, 1954). For the supplier of such goods, it is difficult to exclude from consumption those who fail to pay for the good (non-excludability). In the second place, consumption of such goods by one person does not diminish the consumption opportunities for other persons (non-rivalness; Musgrave, 1969). Classical examples of these types of goods are technical innovations, the production of information, the broadcasting of television and radio signals, lighthouses, public order, defense, street lighting, and sea defenses. As a result, public goods are either not produced at all or not in the optimum quantities because of free-rider problems and problems with establishing the ‘willingness to pay’ for these goods (Bohm, 1987). If a supplier has already produced the goods, consumers will be tempted to free ride on the willingness to pay of others since they can no longer be excluded from consumption of the good. To establish the optimum quantity of a collective good, the marginal utility of single increments of this good must be known to all the consumers involved. Because of its non-rival character, the aggregate willingness to pay for marginal units is compared with the marginal costs. When consumers are asked to reveal their willingness to pay for extra units, they will exaggerate or minimize it for strategic reasons. Exaggeration will occur when the willingness to pay for extra units is not linked to actual payment for extra units of the good. Maximization will occur when the financing of the public good is linked to the willingness to pay that was indicated. Because consumers cannot be excluded, there will be a tendency to free ride on others’ willingness to pay, and for strategic reasons individuals will indicate only a modest willingness to pay for themselves. For these reasons, a market economy will not be able to produce such goods in optimum quantities, if at all. Government regulation is necessary to establish the optimum quantity of the goods concerned, and to enforce payment for these goods. Many goods, such as radio and television broadcasting, research and development, education, health care, parks and roads, have a public good dimension. In such cases also, government regulation can theoretically contribute to a more efficient use of resources in an economy.

2.1.5 Undesirable market results

According to public interest theories, regulation can be explained not only by imperfect competition, unstable market processes and missing markets, but also by the need to prevent or correct undesirable market results. In a compet-
itive market economy, participants in the economic process are rewarded according to their marginal productivity contribution. This result of the market process may be undesirable for economic and other reasons. In the first place, it is possible that a redistribution has large incentive effects (Stiglitz, 1994, pp. 47 ff). In the second place, it is possible that an efficient redistribution will increase the general level of economic welfare when dilemmas such as the prisoner’s dilemma impede voluntary transfers (Hochman and Rogers, 1969, 1970). An efficient redistribution might also take place if the marginal utility of income diminishes and satisfaction capacities do not differ widely among people. However, in economics it is conventional to assume the unfeasibility of cardinal measurement of utility and interpersonal utility comparison, so that this last form of efficient redistribution cannot theoretically be justified from an economic point of view (Robbins, 1932). In the third place, it can be argued that bounded rationality offers a number of justifications for correcting market outcomes from the point of view of efficiency: real preferences are not adequately reflected in market behavior and values or their formation is the result of distorting contexts (Sunstein, 1997). Redistributive policies on behalf of drug addicts or other disadvantaged groups often recognize that people do not ‘prefer’ to become disadvantaged in this way. Furthermore, it has been widely established that consumers have social preferences and care about the inequality of market outcomes (Dellavigna, 2009). Finally, the correction of undesirable market results can also be considered desirable for reasons other than economic reasons, such as considerations of justice, paternalistic motives or ethical principles (Ogus, 2005). In that case, tradeoffs may arise between, for example, economic efficiency and equality: the incentive effects of redistribution may result in a decline in the level of individual utility (Okun, 1975). Public interest theories are most often applied by economists to explain regulation as aiming for economic efficiency (Joskow and Noll, 1981, p. 36). In other cases, public interest theories are interpreted more broadly and regulation is predicted to correct inefficient or inequitable market practices (Posner, 1974). According to this last view, regulation might be said to aim for a socially efficient use of scarce resources as opposed to an economically efficient allocation of resources, where economically is interpreted in a narrow way. However, according to some other views, a complete efficiency analysis should be able to include principles and values like corrective justice (Kaplow and Shavell, 2002; Zerbe, 2001), as long as these values and principles consume scarce resources. Examples of laws and rules intended to prevent or ameliorate undesirable market results are a legal minimum wage, maximum rents, cross-subsidies in postal delivery, telephone calls and passenger transport, rules enhancing the accessibility of health care, rules guaranteeing an income in the event of sickness, unemployment, disability, old age, ‘reasonable’ rate of return regulation, and more.
2.2 Criticism of the Public Interest Theories of Regulation

Theories explaining regulation as an efficient solution to market failures, have been criticized from different angles. First, the core of the public interest theories of regulation, the market failure theory, has been the object of criticism. Second, it has been claimed that the hypothesis that government regulation is efficient or effective has been invalidated by empirical research. In contrast to this criticism, it has in the third place been argued that it is impossible to test or refute public interest theories of regulation. Finally, it has been argued that public interest theories are incomplete: the formation of public preferences and the translation of these interests into welfare-maximizing regulatory measures is missing from these theories.

2.2.1 Market failures as model failure

The theory that regulation can be explained as an answer to market failures has been criticized from several points of view (Cowen, 1988; Zerbe and McCurdy, 1999, 2000). First, the conclusion that monopoly power, externalities or any other so-called market failure gives rise to an inefficient allocation of resources can only be understood by assuming a model in which some of the transaction costs involved are absent. The allocation of resources appears to be efficient if transaction costs are included in the analysis (Dahlman, 1979; Toumanoff, 1984). Monopoly power, for example, only appears to have inefficient outcomes. Once transaction costs are taken into account, such as the inability of the monopolist to price discriminate or to prevent arbitrage or the inability of the consumers to organize and negotiate effectively, the market outcome is efficient. The same reasoning applies to externalities. Marginal costs appear to differ from marginal benefits, but once the transaction costs of the market mechanism are taken into account, the market outcomes turn out to be efficient. The cost-minimizing outcome has been attained.

Second, in practice, the market mechanism itself is often able to develop institutions to compensate for any inefficiencies. Firms will devise ways, for example, of highlighting essential quality aspects such as safety or superior performance. Problems of adverse selection are solved by companies themselves by, for example, the issue of guarantees, the use of brand names and extensive advertising campaigns as a signal of quality (Nelson, 1974). The market sector also succeeds in producing goods that have traditionally been characterized as typical public goods, such as lighthouses (Coase, 1974). So-called externalities have been shown to have been internalized by the market itself (Cheung, 1973, 1978). The assumption of market failure when a dominant firm supplies the market is similarly criticized (Demsetz, 1976). Any significant returns could be the result of the superior efficiency of such companies and furthermore, account must be taken of the possibility of
competition for the market (Baumol et al., 1982) as opposed to competition in the market.

Third, a more general criticism of the theory of market failure is its limited explanatory power. An economist generally needs only ten minutes to rationalize government intervention by constructing some form of market failure (Peltzman, 1989). The market failure theory is an inconsistent and unnecessary part of public interest theories of regulation. An adequate explanatory regulatory theory must explain how and why regulation is comparatively the best transaction cost-minimizing institution in the efficient allocation of resources for particular goods, services or societal values (Zerbe, 2001). The concept of market failures does not contribute to that task.

2.2.2 Is regulation efficient and effective?

In the second place, the original theory assumes that government regulation is effective and can be implemented without great cost (Posner, 1974). So the very transaction costs and information costs that underlie market failure are assumed to be absent in the case of government regulation. This assumption has been criticized in both empirical and theoretical research. Theoretical research, the theory of the second best, has demonstrated that the partial aim of efficient allocation does not make the economy as a whole more efficient if unavoidable inefficiencies persist elsewhere in the economy (Ng, 1990). Unavoidable inefficiencies such as dominance in product markets or taxation distort the allocation in the economy at large. Not only is the good concerned produced in insufficient quantities, but also too many resources are devoted to the production of other goods and services in the economy. These distortions also mean that allocation in factor markets is suboptimal. Suppose prices are lower than marginal cost (road congestion) in ways that cannot be controlled by the regulator. Suppose furthermore that a regulator wants to set welfare-maximizing prices in a sector under its control. In that case, it is of little use to aim for allocative efficiency through, for example, price regulation of public transport by setting prices equal to marginal costs. A ‘second-best’ solution would be to supply transport to other sectors of the economy at prices lower than marginal costs until the welfare loss of the price subsidy is equal to the welfare gain of reduced supply in the congested sectors. This would require knowledge by the regulator of the cost and demand schedules of all sectors of the economy. Furthermore, in reality, a great many of these unavoidable inefficiencies exist. They result from external effects, taxation, imperfect competition and flawed information. That renders the achievement of a second-best optimum unfeasible in practice (Utton, 1986), so that even less precise ‘third-best’ rules and policies have been suggested (Ng, 1977). Other theoretical research points to fundamental flaws in policymaking. Accurate predictions on how rules will work, can not be made if regulations
change the behavior of regulatees and the structures in which they operate. Furthermore, the changes in the original situation by rules and regulations will be anticipated by rational actors (Kydland and Prescott, 1977; Boorsma, 1990). Optimal policies would thus become inconsistent. One effect is the preference for uniform and fixed rules rather than discretion for regulators, but other inefficiencies would then result from the heterogeneity of sectors, regions or firms (Kaplow, 1992, 1995; Latin, 1985). But of course the information to devise optimal policies and regulations is not available nor is enforcement perfect (Sappington and Stiglitz, 1987a, 1987b; Viscusi and Zeckhauser, 1979). The results are inefficient rules, imperfect enforcement and incentives for firms and consumers to behave inefficiently (for a criticism of some of this research, see Kelman, 1988). Examples are inefficient safety standards set by regulators, the selection of inefficient combinations of production factors by firms (the Averch-Johnson effect) and the inefficient planning of investment projects (Viscusi, 1985; Baumol and Klevorick, 1970; Sweeney, 1981, and more generally on the interaction between firms and regulators: Laffont and Tirole, 1993). Theoretical research into the efficiency and effectiveness of government regulation gradually developed into theories of non-market failures equivalent to the theories of market failures (Wolf, 1979, 1993; Tullock, Seldon and Brady, 2002). These theories point to non-market failures such as:

- the lack of information on the marginal benefit of regulatory agency activity and the consequential lack of incentives to equate marginal cost with marginal benefit;
- the lack of output valuation or indicators and the consequential lack of incentives to minimize cost or to end regulatory activity;
- the lack of a market for regulatory control analogous to the market for corporate control, with its consequential failure to discipline managers;
- inequalities in the distribution of the agency benefits as a result of capture or bargaining;
- the unavoidability of unintended effects, unexpected side-effects and even adverse effects of regulation.

The exaggeration in some parts of the literature of the supposed inefficiencies of government regulation led Wittman to argue that political markets were in effect efficient (Wittman, 1989, 1995; and criticizing this position, Lott, 1997; Rowley, 1997).

Empirical research into the effectiveness and efficiency of government regulation has also given rise to criticism of the public interest theory. For a general overview of the effects of economic regulation, see Joskow and Rose (1989). Research into economic regulation began with the famous article by
Stigler and Friedland (1962), ‘What Can Regulators Regulate?’, about the effects of price regulation on electricity producers. In this paper, they showed that regulation did not lower rates, that it had an insignificant effect on profits and that price discrimination was not significantly reduced. The paper started an entirely different way of thinking about government regulation. An earlier synthesis of this type of empirical research showed first that the influence of regulation on natural monopolies was slight if not non-existent (Jordan, 1972). In the second place, it appeared that regulating potentially competing sectors such as air traffic and freight resulted in an increase in prices and a restricted number of competitors. Empirical research further demonstrated that regulation prescribed an inefficient price structure in which for the most part certain consumer groups received cross-subsidies (Posner, 1971). Later research into the effects of economic deregulation demonstrated furthermore that mainly consumers, but to some extent also producers, derived benefit on balance from less government regulation (Winston, 1993, 1998). Often it was employees who benefited from regulation. Deregulation increased welfare by 7–9% of gross national product (GNP) and employment increased as well. Social regulation appeared to keep costs and benefits more or less in balance (Hahn and Hird, 1991), although there is also empirical evidence suggesting that much social regulation is poorly targeted or is over-stringent (Sunstein, 1990; Hahn, 1996; Baldwin, 1999; Wilson, 1984). Research also suggested that about one third of the productivity slowdown in a decade resulted from the cost of social regulation (Gray, 1987). A qualifying remark can be made pertaining to economic and social regulation, that it is often difficult, if not impossible, to quantify many of the benefits. For example, it is difficult to put a value on the distributional effects of cross-subsidies or the preservation of a variety of life forms, or to take account of the preferences of future generations. Finally, there are arguments for assuming that even competition legislation is sometimes misused as an instrument of monopolization (Baumol and Ordover, 1985; McChesney and Shughart II, 1995).

2.2.3 Testing the public interest theories of regulation
Public interest theories usually assume that regulation aims to establish economic efficiency. Interpreted in this way, these theories are unable to explain why on occasion other objectives such as procedural fairness or redistribution are targeted at the expense of economic efficiency (Joskow and Noll, 1981, p. 36). On the other hand, when it is assumed that regulation pursues social efficiency, another problem is encountered. Where there is conflict between efficiency and equity, it is impossible for at least two reasons to establish the social efficiency of regulation (Sen, 1979a, 1979b). Such conflicts may arise, for example, when regulators mandate universal service obligations for public utilities, cross-subsidies for certain consumer groups, a
prohibition on using price discrimination, minimum wage legislation or rent control, and more generally the protection of disadvantaged groups. First, in such situations the evaluation of social efficiency is difficult because evaluation standards for the tradeoff between efficiency and equity are not available. No agreement exists regarding the definition of equity in concrete situations (Dworkin, 1981). Second, establishing the social efficiency of regulation requires that economic efficiency and justice be weighed against each other. The theoretically justified and practically usable scale of values that this calls for is not available (Ng, 1985). The absence of generally applicable standards of justice and the lack of insight into the relationship between justice and efficiency renders empirical testing of public interest theories as explanatory theories of regulation impossible. A key problem for public interest theory is that the evaluating, normative theory of economic welfare is being used as a positive explanatory theory of regulation (Joskow and Noll, 1981). Empirical work testing public interest theories relative to private interest theories has concentrated for the most part on the effectiveness and not the efficiency of regulations: are prices lower, is price discrimination absent, is there a decrease in costs, has pollution declined, is the influence of interest group detectable, etc. Examples include Jakee and Allen (1998), Kroszner and Strahan (1999), Tanguay et al. (2004).

2.2.4 The incompleteness of the public interest theories
A final point of criticism is that public interest theories are incomplete. In the first place, the theories do not indicate how a given view on the public interest translates into legislative actions that maximize economic welfare (Posner, 1974). The political decision-making process consists of various participants who aim for their own objectives under different constraints. In contrast to the theoretical model of a fully competitive market economy, it is unclear how the interaction of the participants in the political process will lead to maximum economic welfare. What is lacking is a rational choice theory leading to welfare maximization. Second, a theory of regulation should be able to predict which branches of industry or which sectors should be regulated, and to whom the advantages and disadvantages should accrue. The theory should also be able to predict what form regulation is to take, such as subsidies, restricted entry, or price regulations (Stigler, 1971). Public interest theories appear not to be able to make adequate predictions that are amenable to testing by empirical economic science (Stigler, 1971). Furthermore, facts are observed in social reality which are not well accounted for by public interest theories. Why should companies support and even aim for regulation intended to cream off excess profits? Of course, much normative public interest analysis has been undertaken on the forms of regulation, and not only of economic regulation. For comparison of regulated private enterprises and public enterprises, for
example, see Boardman and Vining (1989); Martin and Parker (1997); Megginson and Netter (2001). This literature generally or cautiously argues that private enterprises are more efficient. There is also the well-known literature on standards versus prices for damages, pollution and other externalities (Ogus, 1998, Cooter, 1998). Usually, the theories predict that economic instruments are more efficient, but in practice we more often find standards to be more efficient. Public interest theories find it difficult to explain these findings. On the choice of instruments more generally, valuable normative studies include Stewart (1981), Dewees (1983), and Trebilcock and Hartle (1982). On the comparison between regulation and liability, see Burrows (1999), Dewees (1992), Dewees, Duff and Trebilcock (1996), Ogus (2007), Shavell (1984a, 1984b), Weitzman (1974), White and Wittman (1983), and Wittman (1977). These studies mostly focus on the efficiency properties of the instruments and often do not explain from a positive perspective which instruments are being used in practice and for what reasons.

2.3 Private Interest Theories of Regulation

After the public interest theory had fallen into disrepute through empirical and theoretical research, the capture theory was developed mainly by political scientists; for a discussion see Posner (1974). This theory assumes that in the course of time, regulation will come to serve the interests of the industry involved. Legislators subject an industry to regulation by an agency if abuse of a dominant position is detected. In the course of time, other political priorities arrive on the agenda and the monitoring of the regulatory agency by legislators is relaxed. The agency will tend to avoid conflicts with the regulated company because it is dependent on this company for its information. In addition, it often does not have unlimited resources, which makes it aware of the costly effects of litigation of its decisions. Furthermore, there are career opportunities for the regulators in the regulated companies. This leads in time to the regulatory agency coming to represent the interests of the branch involved. For an overview of the various strategies available to be applied by agencies and regulated companies, see Owen and Braeutigam (1978).

The capture theory is unsatisfactory in a number of respects (Posner, 1974). First, it is insufficiently distinct from the public interest theory, because the capture theory also assumes that the public interest underlies the start of regulation. Second, it is not clear why an industry succeeds in subjecting an agency to its interests but cannot prevent its coming into existence. Third, regulation often appears to serve the interests of groups of consumers rather than the interests of the industry. Regulated companies are often obliged to extend their services beyond the voluntarily chosen level of service. Examples are transport services, and the supply of gas, water, electricity and telecommunication.
services to consumers living in widely scattered geographical locations. Fourth, much regulation, such as environmental regulation, regulation of product safety and labor conditions, is opposed by companies because of the negative effect on profitability. Finally, the capture theory is more of a hypothesis that lacks theoretical foundations. It does not explain why an industry is able to ‘take over’ a regulatory agency and why, for example, consumer groups fail to prevent this takeover. Nor does it explain why the interaction between the firm and the agency is characterized by capture instead of by bargaining. Recently dynamic capture theories have been developed, explaining the lifecycle of regulatory agencies evolving over time from acting in the public interest to becoming increasingly inefficient and more eager to please private interests (Martimort, 1999). According to these theories, capture is the result of the increasing power of the agency which arises because the agency in its ongoing relationship gets to know the firm better and better. The agency has thus more and more opportunities to pursue its own objective and the political principal can only control this by having more stringent administrative rules and fair and open procedures. These limitations ‘cripple’ the agency and make it receptive to the influence of the regulated firm.

2.3.1 The Chicago theory of regulation

Stigler In 1971, ‘The Theory of Economic Regulation’ by George Stigler appeared. This was the start of what some called ‘the economic theory of regulation’ (Posner, 1974) and others ‘the Chicago theory of government’ (Noll, 1989a). Stigler’s central proposition was that ‘as a rule, regulation is acquired by the industry and is designed and operated primarily for its benefit’. The benefits of regulation for an industry are obvious. The government can grant exemption from antitrust legislation, grant subsidies or ban the entry of competitors directly so that the level of prices rises. The government can maintain minimum prices and restrict entry more easily than a cartel. The government can suppress the use of substitutes and support complements. An example of each is the suppression of transport by trucking to protect the railroads and the subsidization of airports for the benefit of airlines. On the one hand, therefore, demand will arise for government regulation. The political decision-making process, on the other hand, makes it possible for industries to exploit politics for their own ends. For this proposition, Stigler uses the insights of Downs (1957) and Olson (1965). In the political process, primarily interest groups will exercise political influence, as opposed to individuals. Individuals will not participate because forming an opinion about political questions is expensive in terms of time, energy and money, while the benefits in terms of political influence will be negligible. Individuals will only be informed on particular interests as members of an interest group. Democracies
will thus mostly be a platform for interest groups. Some groups can organize themselves less expensively than others. Small groups have an advantage because the transaction costs are lower and the ‘free-rider’ problem is smaller than in the case with large groups. Furthermore, in small groups preferences will be more homogeneous than in large groups. Small groups also have an advantage in that for the same expected total revenue, the revenue per member of the group is greater. The fact that apparently large groups can still be well organized is explained by Stigler through concentration and asymmetry (Stigler, 1974). The large companies in a concentrated branch will see themselves as a small group. In the case of asymmetry in an industry, for example as a result of product diversity or widely varying production techniques, separate companies will wish to prevent unfavorable regulation and will participate in the organization. Stigler’s theory is illustrated in Figure 1.5.

Figure 1.5 illustrates how the amount of interest group action depends on the distribution of costs and benefits for the interest groups. Interest group A has both lower costs of organization and mobilization of its members, for example because it is a comparatively small group, and higher benefits. Interest group B has comparatively lower benefits, for example because they have to be shared among more members.

The result of variation in the costs of organization is that producers organize more readily than consumers. Not only are the costs more modest for

---

**Figure 1.5** Interest group representation depends on costs and benefits

Source: Adapted from Baron (2000, p. 171).
producers, but the cost of regulation will also be spread over more consumers so that the higher price per consumer is too small to justify organization. Politicians aim for re-election. Organized branches can contribute to re-election in two ways: by supplying votes and other resources. Examples of these resources are campaign contributions, chairing fundraising committees and the offer of employment to party members. The larger industries have an advantage in this over smaller branches, unless the smaller branches have something in their favor such as a strong geographical concentration. Politicians will honor the demand for regulation by industries because the opponents do not find it worthwhile to gather information and organize. The conclusion is that regulation is not directed at the correction of market failures, but at effecting wealth transfers in favor of the industry concerned in exchange for political support. In particular, more competitive industries (professions, taxis, agriculture, trucking) will be favored: they have much to gain from regulation and are comparatively easily to organize.

2.3.2 Extensions to the Chicago theory of regulation

Peltzman In the same issue of the *Bell Journal of Economics* in which Stigler put forward his theory of economic regulation, Posner (1971) implicitly supplied the first criticism. He observed that in many cases regulation strongly benefited certain consumer groups. For instance, uniform prices were prescribed for such things as rail transport, the supply of gas, water and electricity, telecommunications traffic and mail distribution. But the costs of supplying these services differ considerably, for example between residential and rural areas. It is more costly to provide network services for rural customers than for urban consumers. Other examples are the supply of drinking water to households, schools and fire services, either free of charge or at a price lower than marginal cost; free rail travel for government workers and military personnel; the supply of electricity to hospitals at less than marginal cost and so on. This phenomenon of internal or cross-subsidization does not fit in well with Stigler’s theory of regulation. Even if other consumer groups are obliged to pay higher than marginal costs for their goods and services, cross-subsidization works against the aim of profit maximization. An explanation of cross-subsidization is provided in an extension to the theory of regulation by Peltzman (1976). He assumes that politicians will choose their policy of regulation in such a way that political support is maximized. It is not likely that regulation will benefit the supported industry exclusively. Some consumer groups will also be able to organize themselves effectively. Moreover, industry organization and information costs are an obstacle to the immediate and total withdrawal of political support in the event of a small decrease in cartel profit. Lower prices benefit consumers, higher prices gener-
According to Peltzman, the core problem for regulators is efficient regulation: what price level should be chosen such that the gain in votes resulting from the wealth transfer just balances the loss of votes resulting from the rise in prices. Figure 1.6 illustrates his theory. The right-hand panel shows costs and revenues for the dominant firm. Marginal cost is assumed to be constant and equal to average cost. The demand schedule is decreasing and marginal revenue declines twice as steeply as average revenue because setting a lower price for additional units implies lower prices for intra-marginal units as well. The welfare-maximizing price is equal to $P_{mc}$, where resource cost $MC$ are equal to $P$. Profit maximization for the firm, however, is where marginal cost equals marginal revenue. This is at price $P_{m}$ and units $Q_{m}$. The left-hand panel tilted profit hill with the consumer welfare-maximizing price $P_{mc}$ and the profits-maximizing price at $P_{m}$. If the firm charged an even higher price, the revenue losses from decreasing quantities would be higher than the revenue increases from higher prices. Two isomajority curves have also been drawn, reflecting combinations of profits and prices that render equal votes for the regulator. The further the M-curves are to the south-west, the higher the political support for the regulator. The M-curves show that successive increases in $P_{mc}$ require increasingly higher profits to hold votes constant. A vote-maximizing regulator will therefore raise prices only to the point where votes gained from the firm by raising price are exactly offset by votes lost from consumers through the higher price. Unlike Stigler’s suggestion, a regulator will not maximize profits for the firm. Rather, the political equilibrium will be at the efficient price $P_{e}$, where it is no longer
possible to raise political support by raising price for the firm. The efficient price is at a point somewhere between the profit-maximizing price and the welfare-maximizing price, where votes lost by raising price are exactly offset by votes won by raising price.

This extended theory explains not only the phenomenon of cross-subsidization, but also predicts which industries will be regulated. These are the relatively competitive branches and the monopolistic branches. In the first case, these branches have a keen interest in regulation and, in the second case, consumers or particular consumer groups have an interest in regulation. Intermediate industries and their customers have only a limited interest in regulation: regulated prices will not differ that much from the market outcomes. Regulatory practices appear to confirm this prediction. Regulated industries are either monopolistic, such as rail transport and telecommunications, or highly competitive, such as freight, agriculture, independent professions and cab companies.

As well as types of industries, the theory of regulation also predicts the form the benefits will take. In principle, transfers can be effected directly through subsidies or indirectly through price or quantity regulation or restrictions on market entry. Stigler originally assumed that regulated industries would favor indirect support. The granting of subsidies would invite entry, so that the per-producer subsidy would be dissipated. In an extension to the theory of regulation, Migué (1977) has shown that the form of transfers is partly dependent on the supply elasticity of the production factors in the industry concerned. In the political market, the public are both consumers and suppliers of production factors. Suppliers of production factors will prefer subsidies when supply is inelastic. The taxation necessitated by the subsidy is distributed over many taxpayers, while the subsidy accrues to a limited group of suppliers of production factors. This extension to the theory of regulation explains why subsidies are granted in sectors such as education, health care, domestic housing and city transport, and why quota systems and price regulation can be found in sectors such as agriculture, airlines, road transport and railways. Similar reasoning explains why polluting companies prefer emission standards (quotas) to taxation (Buchanan and Tullock, 1975). Quotas have the ability to increase profits, while taxation benefits the general public. Environmentalists will also favor standards for their visibility and their ability to generate funds to implement the regulations. Disadvantaged consumer groups are usually less well organized, if at all.

Another extension to the theory of regulation comes from McChesney (1987, 1991, 1997). He sees politicians not as neutral agents between competing private interests directed at obtaining transfers of income. In his view, politicians also try to benefit by putting private parties under pressure. He gives examples in which the US Congress, under the threat of price reductions
or cost increases, forced concessions from private parties. To make such rent extractions possible, politicians encourage private parties to organize. Organization not only enhances the probability of gaining wealth transfers, it also increases the risk of having one’s own surplus threatened and expropriated. Finally, Keeler (1984) has supplemented Peltzman’s model with public interest considerations. In his model, politicians gain not only political support through transfers of income or wealth between interest groups. An increase in economic efficiency, for example by promoting economies of scale and internalizing external effects, increases resources or welfare which can be distributed among producers and consumers. Rational politicians will use these resources to further their re-election.

Becker A further contribution to the Chicago theory of regulation was made by Becker (1983, 1985a, 1985b). He concentrated on effects of competition between interest groups, which he calls pressure groups. As the political pressure increases, political influence also increases and the financial returns from the pressure exerted rise. Some groups are more efficient in exercising political pressure than others, perhaps as a result of economies of scale in the production of pressure, better abilities to counter free-riding, better access to the media, or in other ways. Wealth transfers thus take place from less efficient to more efficient groups through such instruments as price regulation or subsidies. A limit to these transfers exists, however. Wealth transfers are associated with economic losses, which are known as deadweight costs. As a result of these losses, the loss of the least efficient pressure group is greater than the gain of the more efficient pressure group. As the welfare losses become greater, the pressure of the more efficient group will decline because the returns to pressure are lower. At the same time, the pressure of the less efficient group increases with increasing deadweight costs because its potential return on exerting pressure increases. This countervailing pressure limits the possibility of transfers to the more efficient pressure group. It can be deduced from this analysis that politically successful groups are small in proportion to the group bearing the burden of the transfers. The larger the burdened group, the smaller the levy per member of the group and the smaller the deadweight costs. This diminishes the countervailing pressure. The smaller the receiving group, the larger the potential return per member of the group, which serves to increase the pressure exerted. The analysis explains, for instance, why in countries where the agriculture sector is small and thus well organized, it is subsidized, while large agriculture sectors elsewhere, which are difficult to organize, are heavily taxed but with small levies and deadweight costs per member.

In Stigler’s and Peltzman’s view, competitive industries have much to gain from regulation and are in a better position than consumers to bring about
favorable regulation. In practice, such regulation of competitive sectors is rarely seen. The explanation is found in Becker’s theory. Loss of welfare is greater where the elasticity of supply is higher. In competitive sectors, the elasticity of supply is large. The welfare losses and transfer of income resulting from regulation of such sectors will be so large that countervailing pressure by the disadvantaged consumer and producer groups will far outweigh any pressure from the benefited industry. Becker’s analysis of an organized industry is illustrated in Figure 1.7.

Imagine that regulation depends on the amount of expenditure on pressure and that firms have incurred sunk costs in organizing. In a competitive sector, profits are zero and price is $P_{mc}$. If firms have organized, the profits could rise by the rectangle $A$, that is $(P_{m} - MC)$ times $Q_{m}$. That is the maximum amount firms would be willing to spend on obtaining favorable regulation. But the loss to consumers is the transfer to producers (rectangle $A$) plus the triangle $B$. That is, consumers have more to lose and thus an incentive to spend more to prevent regulation. However, consumers also have to incur the costs of organization. Assume the cost of organization for consumers is given by the amount of rectangle $C$. Firms will now succeed in obtaining regulation if their
expenditure is higher than the potential net gain to consumers from preventing regulation, that is, the rectangle A plus the triangle B minus the rectangle C. This makes clear why regulation is less likely in the event of high deadweight losses: consumers have more to gain from preventing regulation. It also makes clear the importance of the size of the group: the larger the group, the higher the organization costs and the lower the amount each member will gain, recalling that A and B are fixed amounts.

It can be further deduced from the analysis that regulation is more likely in branches exhibiting market failures, a result in agreement with the public interest theory of regulation. In monopolistic industries, consumers may obtain larger transfers than the accompanying losses for producers. In competitive industries, on the other hand, the gain of the winners is smaller than the loss of the losers. All other things being equal, more pressure will be exerted in monopolistic industries by the potential winners and less pressure by the potential losers than in competitive industries. Market failure is therefore not a sufficient condition for regulation, such as in the public interest theory of regulation; regulation is also dependent on the relative efficiency of pressure groups in exerting political pressure. In contrast to Olson (1982), competition between pressure groups will not have any negative effects on the growth of the national product and productivity, at least provided that pressure groups are of equal size and provided they are equally efficient in producing pressure. Competition between pressure groups will also lead to the most efficient form of regulation.

Even if under certain circumstances the results of competition among pressure groups is efficient, Becker claims the production of pressure is not. All pressure groups would be better off if they decreased their expenditure on pressure by equal amounts. Various laws and rules directed at limiting the influence of pressure groups can be explained as instruments to limit wasteful expenditure on political pressure.

The Chicago theory of regulation seems particularly suited to the explanation of so-called economic regulation. Social regulation – regulation in the area of consumer protection, safety, environment and health – seems at first sight to be less amenable to explanation by this theory. There are diseconomies in the area of organization, the benefits are divided among many involved parties and the costs of regulation are allocated to concentrated groups. Nonetheless, private interest explanations have also been put forward to explain social regulation (see for example Bartel and Thomas, 1985, 1987; Pashigian, 1984a, 1984b; Tanguay et al., 2004). For example, the application of rules and standards, the dominant form in social regulation, is in the interests of those companies already complying with the standard. Furthermore, companies that produce inputs that are required by the standard also benefit. Also, large companies benefit when it is necessary to comply with administrative
obligations or costly measures. Small companies are driven out of the market and the softening of competition may outweigh the increase in regulatory costs, depending on the elasticity of the relevant variables (Bartel and Thomas, 1987). Legal requirements are above all often differentiated into existing producers and new producers. By setting higher standards on new producers, entry to the market is impeded and competition is restricted (Huber, 1983).

A unified framework integrating several perspectives of the private interest theories of regulation has been devised by Beard, Kaserman and Mayo (2003; 2005).

2.4 Criticism of the Private Interest Theories of Regulation

The theory that regulation can be explained as an efficient mechanism for redistributing wealth to the more efficient interest groups has been criticized from different angles. First, the core of private interest theories of regulation, that private interests translate into transfers in the political market, has been the object of criticism. Second, the hypothesis that regulation promotes mostly private interests has been claimed to have been invalidated by empirical research. Contrary to this criticism, it has been argued that it is very difficult, if not impossible, to test or refute the private interest theories of regulation. Third, it has been argued that private interest theories are incomplete: political actors and an analysis of the interaction between the various actors in the regulation process are lacking from these theories. Finally, rent-seeking theorists criticize the efficiency assumptions of the Chicago theories of regulation.

2.4.1 Transfer of wealth as the driving force of regulation

With respect to the driving forces of regulation, several weaknesses have been identified in the literature. Noll (1989a) criticized the Chicago theories for running the risk of being a tautology. Regulation is always associated with redistribution of wealth. It involves costs and benefits for the different actors involved, such as the lobbying industries, consumer groups, bureaucrats, legislators, regulators, workers, taxpayers, and more. But by establishing who derives the benefits and who carries the cost, it has not been established that these costs and benefits actually drive regulation. Also, early on, Posner (1974) argued that a society concerned with the ability of interest groups to obtain favorable legislation would establish institutions that promote the public interest. Many institutions and features of public policy, such as the independent judiciary or the constitutionally required payment of compensation in eminent domain cases, are more plausibly explained by reference to a broad social interest in efficiency than by reference to the designs of narrow interest groups (Posner, 1974, p. 350). A related criticism is that private interest theories have been shown to be weak in predicting which interest groups
will actually be successful and who will receive the wealth transfer. For example, research has shown that workers mainly derive benefits from regulation, not producers (Bailey, 1986; Winston, 1993). This is an unexpected outcome not predicted by private interest theories of regulation. Similarly, a review of the introduction or amendment of some 20 regulations, using an analysis of stock market data, showed no systematic tendency for increased wealth for the industries concerned (Binder, 1985). A final puzzle is the inability of private interest theories to predict which industries will actually be regulated. There are numerous branches of industry, all of them more favorably placed to organize and obtain transfers compared with the organizational characteristics of consumers. However, only some of these branches are regulated, while others with similar characteristics are not.

2.4.2 Empirical validity and testing of private interest theories of regulation

Private interest theories were mainly developed to explain the findings of ineffectiveness and inefficiencies of regulatory practices. The empirical research started with the seminal contribution by Stigler and Friedland (1962), who argued that regulation did not lower rates, nor reduce price discrimination, nor significantly lower profits. Ironically, this research was later proved wrong about the magnitude of the effect of regulation (Peltzman, 1993), but it still generated a completely new perspective on the effects of government regulation. The analysis should have shown that regulation actually lowered price by about a fourth and thereby caused output to rise by over half. Other research also proved regulated prices to be significantly below the profit-maximizing prices. For electricity, these profit-maximizing prices were estimated to be 20–50% higher than actual regulated prices (Green and Smiley, 1984). Furthermore, regulatory activity in the fields of the public utilities and of health, safety and the environment has come in waves of political activity (Peltzman, 1993). This is difficult to bring in accordance with the static private interest theories of regulation. For an important subfield of regulation, occupational licensing, some and perhaps most of the evidence points to public interest explanations (Zerbe and Urban, 1988; for an opposing view, see for example Becker, 1986; Philipsen and Faure, 2002). See generally, on how to obtain empirical results of the study of regulation and deregulation, Joskow (2005).

Contrary to the above argument, it has also been claimed that it is difficult, if not impossible to test the private interest theories of regulation (Potters and Sloof, 1996). In their comprehensive review of empirical studies, the authors first note the importance of taking both demand and supply considerations into account in the explanation of public policy. The influence of an interest group depends also on the influences of other interest groups and on the importance
of the constituency for the regulator. To determine the relative influence of an interest group it is therefore necessary to practically measure the constituent interests in a theoretically correct way. In the end, the authors conclude that ‘it seems impossible to infer whether voters (the public at large), politicians, or special interest groups are the most influential’ (Potters and Sloof, 1996, p. 430).

2.4.3 The incompleteness of the Chicago theories of regulation

The Chicago theory of regulation assumes that interest groups determine the outcomes of elections, that legislators honor to the letter the wishes of the interest groups and that legislators are able to control regulators. In these theories, the following elements are conspicuously missing:

(a) the motivation and behavior of the various political actors, such as voters, congressmen, legislators, government workers and agencies;
(b) the interactions between the various actors in the regulation process;
(c) the mechanisms through which legislators and regulators serve the interests of the organized industries.

In fact, it is assumed that the operation of the political process of legislation and the administrative process of regulation has hardly any independent influence on the pattern and form of regulation, if at all. This assumption has been criticized from several quarters and several attempts have been made in the literature to fill in the three gaps.

Public choice on regulation

By effectively taking the details of the political process into account, regulatory theory obviously overlaps with the bordering field of public choice. Overviews of theories of interaction between regulation and public choice are given by Romer and Rosenthal (1987), Noll (1989a), Levine and Forrence (1990), Mueller (2003) Frey (1978, 1983) and Frey and Ramser (1986). In particular, the interests of the involved bureaucracy or staff in regulatory systems have also been studied: see Dunleavy (1991) and Wilson (1989). Studies of regulation thus evolve into theories of political action.

Figure 1.8 (1994) illustrates some of the complexities in the political system, while still not even taking into account the impulses of other actors, such as unions, courts, different layers of parliament, and so on.

The theory that only organized interests succeed in obtaining favorable legislation has been criticized by Wilson (1974, 1980, 1989) among others. Not all legislation can be characterized by concentrated benefits and diffused costs. Majoritarian politics in Congress is to be expected when both costs and benefits are widely distributed; antimonopoly legislation is one example. Interest group politics arises if both costs and benefits are concentrated; labor
legislation and railway regulation are examples of this. Client politics is the result of concentrated benefits and diffused costs; examples of this are the protection of professional groups by means of licensing and the subsidizing of companies and branches. A final characterization is entrepreneurial politics, in which the costs are concentrated and the benefits distributed; examples are protection of the environment, of consumers against unsafe products and of workers against industrial accidents and occupational illnesses. This last form of regulation is particularly difficult to bring in line with the Chicago theories of regulation. Figure 1.9 illustrates Wilson’s distinctions (Wilson, 1980).

According to Wilson, interest groups are therefore not the only originators of regulation. Wilson also criticizes the assumption that legislators are able to fully control regulators. In his view, the behavior of the agencies is better understood by an analysis of the motivations of those involved internally. He distinguishes careerists, professionals and politicians and uses this to account for various types of regulation policy. Price regulation, for example, will be simple in structure if set up by careerists and more complex if it is developed by professionals and engineers.

Contrary to Wilson, Derthick and Quirk (1985) assume that the regulation policy of agencies is actually heavily influenced by surrounding forces. They show that agencies honored diffuse interests at the expense of the concentrated interests of regulated industries, by deregulating certain industries. They seek explanations for these developments in the intellectual climate and the pressure
exerted by the President, Congress and the courts on the agencies. In line with the Chicago theories of regulation, Weingast (1981) also sees no independent role for agencies. Changes in the regulatory policies of agencies are a consequence of changes in the preferences of Congress or its commissions. Contrary to Chicago, Weingast shows how a structure-induced equilibrium of policy choice arises in reaction to the instability of majority rule voting. In an equilibrium of congressional committees, the agencies and the interest groups have divergent but comparable goals. Weingast’s model describes how political demands by interest groups are satisfied and how diffuse single-issue groups such as environmental groups and consumer organizations are able to acquire political power at the expense of traditional interest groups such as industry, employees and agriculture.

The assumption that legislators are fully under the control of interest groups is particularly criticized from the perspective of the principal-agent theory. Information on the behavior of legislators and regulators is expensive. This creates room for legislators and regulators to escape the monitoring of voters and interest groups and to act in their own interests or according to ideological preferences (Kalt and Zupan, 1984, 1990). To explain differences between general and special interests and private and public interests, Levine and Forrence (1990) distinguish between types of motivation and two types of political dominance. Depending on what a political actor is aiming for, private and public interests can be distinguished. Private interests are the preferences of political actors with respect to their self-interest. Public interests are preferences related to the interests of others. The private and public interests indicate what a political actor will maximize when there is room to aim for their own preferences. General-interest policy is a policy that should be ratified in the absence of information, organization, transaction and monitoring costs. Special-interest policies should not be ratified by the general polity in the absence of monitoring costs and so on. General interests and special interests

<table>
<thead>
<tr>
<th>COSTS</th>
<th>Concentrated</th>
<th>Diffused</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentrated</td>
<td>Interest group politics</td>
<td>Client politics</td>
</tr>
<tr>
<td>Diffused</td>
<td>Entrepreneurial politics</td>
<td>Majoritarian politics</td>
</tr>
</tbody>
</table>

**Figure 1.9 Types of governmental policies**
Monitoring costs result in ‘slack’ or policy drift, or in other words, scope for the regulator to pursue its own objectives. On the one hand, this discretion may be used to favor special interest groups, on the other hand, it can be used to promote the interests of others. Here again there are two possibilities. In the absence of monitoring and control costs, this policy slack would either be ratified by the general polity or not. The debate on the capture of regulators and legislators concerns the question of who dominates the political process and is therefore less focused on private or public interests but rather on general and special interests. According to Levine and Forrence, general interests will prevail to the extent that slack is reduced. The amount of slack decreases drastically when issues enter the public agenda. Favorable conditions for this are: political competition, special interest organizations, public policy intelligentsia and the news media. A general-interest policy does not otherwise imply that the policy is also efficient. There is no guarantee that majority decisions are also
welfare maximizing. When, for example, the rented sector is a substantial part of the housing market, a policy of rent control will achieve ready approval without any guarantee of efficiency.

2.4.4 Normative rent-seeking analysis as positive theory
A conceptually different type of criticism of the Chicago theories of regulation comes from the Virginia School of Public Choice (Rowley, Tullock, Tollison, McCormick et al.). For an overview of this work, see Tullock (1993), Buchanan, Tollison and Tullock (1980) and Rowley, Tollison and Tullock (1988). In their theories, rent-seeking, the term coined by Ann Krueger (1974), is a central feature. Rent-seeking means spending scarce resources on political action by individuals and groups to obtain monopoly rights or other favors granted by governments. The Virginia School criticizes the Chicago theorists for disregarding the inefficiencies of regulation. Virginians evaluate institutions from the perspective of waste in the allocation of scarce resources. In this respect, these theories are similar to the public interest theories of regulation. Where public interest theories presume regulation to be efficient, the rent-seeking theorists presume regulation to be inefficient. Public interest theories have for that reason been characterized as ‘normative analysis as positive theories’. Analogously, the rent-seeking theories can be characterized as ‘normative rent-seeking analysis as positive theories’. In what has become a classic contribution, Tullock (1967) has shown that the inefficiencies of monopoly consist not only of the Harberger triangle, but also of what has come to be known as the Tullock rectangle. The Harberger triangle shows the deadweight losses from a monopoly (triangle B in Figure 1.11). Next to that, the Tullock rectangle measures the transfer in welfare from consumers to producers as a result of higher monopoly prices (rectangle A in Figure 1.11).

The possibility of rent extraction invites investments by would-be monopolists to obtain favorable regulation. A monopolist is willing to spend at most the ‘Tullock rectangle’ to obtain favorable legislation, thereby dissipating the future rents. But other would-be monopolists also compete for this rent, wasting even more resources. Furthermore, the potentially disadvantaged consumers will spend scarce resources to prevent the creation of such a monopoly. And ex post, a monopolist will spend scarce resources to protect his monopoly rights against possible threats from potential competitors and disadvantaged consumers. Finally, according to Hicks (1935, p. 8), the best of all monopolies is ‘a quiet life’. This led Leibenstein to predict that the monopolist would be x-inefficient and not produce at minimal costs (Leibenstein, 1966; see, however, Stigler, 1976). Estimates of the profit rectangles and welfare triangles thus range from 7 per cent to 50 per cent of GNP; estimates of lobbying costs come on top of that, but are considered to be rather smaller (Mueller, 2003, p. 355).
In Stigler’s and Peltzman’s view, a non-contestable monopolist who has incurred substantial fixed costs has no incentive to obtain regulation since profits are already maximized. However, according to the Virginians, the incentives for regulation remain, though now from the side of bureaucracy and politics. The interests of bureaucrats and politicians can be served by giving certain consumer groups hidden privileges through cross-subsidies (Crew and Rowley, 1988). Furthermore, on the Chicago view, redistributive instruments such as taxation, subsidies or regulation are equivalent with respect either to efficiency or to the precise nature of political equilibrium. The Virginians point, however, to the visibility of such taxes and subsidies, and the waste of scarce resources it provokes. Regulation gives more room to politicians and bureaucrats to put their own objectives into effect. Regulation and taxation also differ in terms of the inefficiencies they create. Traditionally, account is taken only of the welfare losses as measured by the Harberger triangle. Taxation usually has lower welfare losses than regulation, because it concerns a far larger group and the costs per person will be low. It is to be expected, however, that the Tullock rectangle inefficiencies will be larger if taxation is comparatively more transparent as the source of transfers than regulation.

Figure 1.11 illustrates the arguments of the rent-seeking theorists. Not only are firms willing to spend at most the expected increase in profits resulting from regulation (rectangle A), consumers will be willing to spend A + B to prevent regulation. Lobbying efforts will motivate bureaucrats and politicians to have a stake in the wealth transfer and the accompanying taxes will again motivate other interest groups to lobby for regulation of their own sector. Furthermore, employees in regulated sectors will also put in some effort to obtain a part of the transfers. Finally, regulated sectors will be cost inefficient in several ways. The resources used for all these activities are considered to be a waste. These wasted resources can no longer be used for productive investments.

The rent-seeking theorems have been criticized for overestimating the assumed losses of welfare. It is not likely that monopolists will be forced to use the entire Tullock rectangle in order to acquire their monopoly and, furthermore, rent-seeking outgoings have positive effects on welfare (Varian, 1989). A strong criticism comes from Samuels and Mercuro (1984), who judge that the limiting assumptions lead to misleading conclusions and that the normative analyses are too selective and limited to serve as the basis for policy.

2.5 Institutional Theories of Regulation: Independence and Accountability

The debates on who and what influences regulatory policy and the costs and benefits of regulatory capture have led to a number of theories on the design
of regulatory institutions (Levy and Spiller, 1996). According to public interest theories, a regulator pursues general-interest policies, but private interest theories have shown regulators might be captured. There is a tradeoff between independence and accountability (Maskin and Tirole, 2004). Regulators of industries should be independent of Ministers who might use regulatory powers to be re-elected. But if regulators are independent, would they not be captured by the regulated industries? How can they be held accountable?

On the one hand, to pursue general-interest policies, sheltering regulators from the general polity can be desirable. Public utility services are important for the broader constituency and at the same time require a considerable amount of fixed investments. If regulators are dependent on short-sighted politicians, they might not be able to resist the temptation to lower prices to marginal costs. Investors would foresee they would not be able to cover the costs of sunk investments and they would not invest. The independency of agencies is therefore desirable. Independent regulators might also be a source of innovation. Independency implies the possibility of experimenting with new ways of monitoring and regulation. Independency will also motivate politicians to state the mission and task of the agency as clearly as possible. This makes it less likely or possible for the agency to pursue its own goals.

Figure 1.11 Rent-seeking by consumers and producers
On the other hand, if regulators are no longer accountable to politicians, they might be captured and pursue special interests. The regulated firm might capture the agency: it has information, resources, employment opportunities and it can extract agency resources by, for example, litigating agency decisions. Furthermore, a regulator is usually independent with respect to a particular domain or sector. This limited domain makes spillover effects more likely to take place (Hahn et al., 2003). For example, mandating a safer and heavier car increases polluting gasoline consumption and sulfur emissions from higher smokestacks end up as acid rain elsewhere. In the tradeoff between independence and accountability, several aspects of an ideal institutional design have been suggested and they depend in part on the institutional environment in which the agency operates (Levy and Spiller, 1996; Martimort, 1999). For example, it is of little use to strive for independency of regulators when technical expertise or auditing technologies are lacking, as is the case in many developing market economies. In such cases accountability can be predicted to be low. Important variables with respect to the institutional environment are: the independency of the judiciary, the ability to commit to stated promises, the dispersion of powers among regulators or among the regulator and the government, the strength of the bureaucracy, the resources available to the agency, being elected or appointed, tenure for a longer or a shorter period, clearly stated mission and tasks, and so on.

3. LIBERALIZATION, RESTRUCTURING AND RE-REGULATION

3.1 Explanations of Privatization and Deregulation

Once the Chicago theory of regulation had been developed, social developments seemed to refute it. While this theory explained regulation as aiming for transfers of income, at the end of the 1970s and the beginning of the 1980s, many complex rules were dispensed with in a process of deregulation and privatization. This process of deregulation and privatization was mainly concerned with economic regulation of sectors such as transport (airlines and freight), telecommunications, energy and the financial sector. In the US, where these sectors were regulated, deregulation took place and in Europe, public enterprises were privatized. Social regulation actually increased in scale, even though the nature of this regulation changed (more cost-benefit analyses, more risk analyses, more performance standards, fewer specification standards) (Winston and Crandall, 1994). For a description of this process of deregulation and (later) re-regulation, see, for example, Bailey (1986), Hahn (1990) and Kahn (1990) for the USA, Vickers (1991) for the UK.
(1998) and Newbery (2001). The literature offers many reasons and explanations for these developments. Some examples are: the desire to foster efficiency and innovation; to decrease the extent of political intervention; to reduce government borrowing; to increase political popularity; to increase rewards paid to top management; to undermine trade union power; to lower the level of cross-subsidies; to eliminate the inefficiencies caused by uninformed regulators or rent-extracting regulators.

From the theories of regulation discussed above, various explanations can be derived for this process of deregulation and privatization (see den Hertog, 1996; Peltzman, 1989; Keeler, 1984). From public interest theories of regulation, two general explanations of deregulation can be derived. In the first place, it is possible that the cause of market failure is removed by technological or demand factors. For example, through increasing demand for transport facilities, a former natural monopoly may change into a competitive market. Furthermore, technological developments, such as communication via satellite or through wireless facilities instead of by cable, can undermine natural monopolies. A second explanation for deregulation may be the presence of more efficient alternatives to regulation to solve the market failure. New instruments may have been developed such as public franchising or yardstick competition (Kay and Vickers, 1990). It is also possible that better insight exists into the envisaged and non-envisaged effects of regulations; see the literature mentioned above, as well as Baldwin (1990), Stewart (1985), Wilson (1984) and Wolf (1979). Finally, it is possible that theoretical developments, such as, for example, contestable markets, inspire more confidence in the operation of the market mechanism (Bailey and Baumol, 1984).

At least four causes of deregulation can be derived from the Chicago theory of regulation. In the first place, shifts can come about in the relative political power of pressure groups, for example, as a result of the more efficient combating of free-riding, the more efficient use of media or as a result of special entrepreneurship (Ralph Nader). In the second place, deregulation can arise when politically effective groups believe that they can better promote their economic interests in an unregulated market, for example by self-regulation. In the third place, deregulation can be the result of declining profits, so that the political yield of regulation declines. The fixing of prices or the introduction of entry restrictions in sectors consisting of multiple companies, such as airlines or freight, will result in competition taking place in other dimensions of the product. Competition in the area of service, such as the frequency of transport, will result in a decline in profits. In Becker’s view, that leads to decreased pressure from the branch involved and increased pressure from consumers for price reduction. According to Peltzman, politics will seek more fruitful regulation yields. Fourth, deregulation can be accounted for by increasing deadweight costs. These costs increase in the course of time
because substitutes for regulated products are developed and because costly methods of evading and avoiding particular regulations are discovered. The deregulation of sectors such as transport, telecommunications and banking can thus be seen as an echo of the regulation movement of the 1930s. Increasing deadweight costs are in the second place a result of the increasing marginal tax rates in the 1960s and 1970s. According to Becker, this stimulated the pressure on taxpayers who were able to collect more political support than the groups who benefited from social security programs.

According to theories of regulation that have evolved into theories of political action, deregulation can first of all be accounted for by a change in the balance of power of pressure groups. In the second place, structure-induced equilibrium can be disturbed by the actions of political entrepreneurs, such as the chairpersons of regulatory commissions. In the third place, politicians can seek political support for deregulation by providing voters with information about the inefficiencies of regulation. Alternatively, politicians could try to use the complexities of regulatory issues by claiming that economic deregulation would greatly advance economic and social welfare.

A general comparison of deregulation practice and the various theories gives a mixed picture (see Peltzman, 1989; Noll, 1989b). If the public interest theory were generally applicable, deregulation would have taken place sooner. On the other hand, events such as the deregulation of freight are once again difficult to account for with the Chicago theory: in this sector, profit was being made and employees also had much to lose from deregulation. Various circumstances, including political entrepreneurship, were considered to be applicable and have also played a role in practice. In other cases, the US Congress has played no role and the legislation was changed after deregulation was already a fact. Also expectations with respect to the future development of regulation and deregulation are mixed. On the one hand, there are researchers such as Kahn (1990) and Hahn (1990) who are convinced of the relative efficiency of the market mechanism and of regulation mechanisms that support and sustain the market. They see a greater role for government in the area of competition policy and in setting constraints on the functioning of markets, such as in the area of safety. On the other hand, there are commentators, such as Cudahy (1993), who assumes that the process of deregulation will be followed in the downward phase of the business cycle by a phase of renewed regulation. Some alleged disadvantages of deregulation, such as predatory pricing, fluctuating prices and discriminatory prices, insufficient service, increased lack of safety, job insecurity and redundancy for large groups of employees, will ring in a new age of regulation. Regulation and deregulation appear to be dynamic processes, whereby internal forces in one setting will be the cause of a movement toward the other. Interestingly, and in line with such the dynamic processes, new theories have been developed that
explain regulation as well as deregulation in terms of the comparative efficiency of private litigation versus public regulation (Glaeser and Shleifer, 2003).

### 3.2 Restructuring, Competition and Re-regulation

#### 3.2.1 Privatization and regulatory reform

Natural monopoly activities have either been organized in public enterprises, as in most European countries, or they have been regulated by regulatory authorities, as in the USA. When, for several reasons, it was decided that public enterprises in European countries should be privatized, decision-makers of course realized that some form of regulation of the former incumbent monopolist would be necessary, either in relation to the inherently monopolistic parts or for as long as effective competition had not yet arisen. Price-cap regulation was developed for these formerly public enterprises, in contrast to the US method of ‘fair rate of return regulation’ of monopoly firms. Rate of return regulation or profit regulation of private monopoly companies had been developed in the US through a long historical process of court cases against regulatory intervention by government agencies. The outcome of this process was what some characterized as a ‘regulatory contract’ (Sidak and Spulber, 1998). Both the interests of the firm and the interest stated by the regulatory agencies were honored in this contract. For the regulated firms, the result of these court cases implied that they should be able to maintain financial integrity, attract new capital and that investors should be compensated adequately for the risks assumed. On the other hand, they should provide their goods and services on a ‘non-discriminatory’ basis and charge ‘just and reasonable’ prices (Kahn, 1988; Philips, 1969). Prices for different customer groups were decided and investments allowed in so-called rate cases, which could take place yearly if necessary, depending on whether or not costs or profits had risen. Regulatory agencies would in the first stage of such a case decide on the revenue requirement or cost of service. For this reason, rate of return regulation is also called cost of service regulation. In the second stage, the structure of prices for different markets and products (thus, customer groups) is decided. The revenue requirement can be described by the following formula: \( P = E + rB \), where \( P \) is total revenue, \( E \) is expenditure, \( r \) the rate of return on capital investments, and \( B \) the rate base (or regulatory asset value or regulatory asset base). Regulators thus determine the revenue such that it is equal to the sum of operating expenditures (\( E \): OPEX) and capital expenditures (\( rB \): CAPEX); compare point \( P_{ac}, Q_{ac} \) in Figure 1.2. For a gas distribution company, for example, the operating cost would consists of such elements as gas, labor, depreciation, and taxes. The net operating income or cash flow would be the difference between the revenue from selling the gas.
through its network and the operating cost. That amount should be sufficient
to cover the capital cost (CAPEX), a weighted average of the cost of equity
(stocks) and debt (bonds). The rate of return for investors is then determined
as the required net operating income as a percentage of the rate base (the value
of the plant less depreciation). If investors do not receive sufficient returns on
their investments (dividends), prices should go up and the net operating
income (profits) should increase until the allowed rate of return is reached.

This type of regulation can be analysed from different perspectives (How and
Rasmussen, 1982; Thompson, 1991; Posner, 2001). First, it is a comparatively
stable form of regulation: just and reasonable prices are combined with a fair
rate of return for investors, which makes those involved committed to the
outcome (Newbery, 2001). Second, determining the rate level and rate structure
by rate cases, which are not unlike court procedures, is an administratively
cumbersome process of regulation and requires many resources. Third, since
regulators will not risk the firm going bankrupt, they will prefer to set the rate of
return too high rather than too low. This leads to the so-called Averch-Johnson
effect, a productive inefficiency because capital is effectively subsidized and
which makes firms favor capital-intensive production technologies. On the other
hand, this might also contribute to dynamic efficiencies if these technologies
embody innovations. Fourth, the operating expenditures \( E \) are considered to be
primarily the responsibility of the firm’s managers. These costs sometimes
amount to 70–85% of the firm’s total cost and are essentially ‘passed through’
into consumer prices, although of course they must be ‘reasonable’. The fixed
profit rate and the ‘cost pass-through’ practice do not provide managers with
incentives to be cost-efficient. Of course, the period in between two rate cases,
the regulatory lag, allows the firm to make a profit by being more efficient at
given regulated prices. However, managers will also take into account that these
efficiencies will be acknowledged by the regulators in the next rate case.

For those reasons, some incentive mechanisms have been introduced in rate
of return regulation (Trebing, 1968; Joskow and Schmalensee, 1986; Mayer
and Vickers, 1996). An example is so-called ‘sliding scale regulation’, where
firms are allowed to keep some of the extra profits they make by being more
efficient. The actual rate of return would then be determined by the formula
\[ R_a = R_t + h(R^* - R_t), \]
where \( R^* \) is the target rate of return, \( R_t \) the rate of return
at initial prices and \( R_a \) the rate of return at the new prices, with \( h \) a ‘sharing
constant’ between zero and one. For example, if the target rate of return is 12%
and firms are allowed to keep 50% of the extra profit they make (\( h = 0.5 \)), and
the actual rate of return at initial prices is 16% (\( R_t \)), then the actual rate of
return (\( R_a \)) at the new prices would be 14%, sharing the extra profit of 4%
between shareholders and consumers by lowering prices. The benefits are
increased incentives to reduce costs, the sharing of profits and less cumber-
some administrative procedures.
Price-cap regulation  Becoming more efficient was precisely one of the more important objectives of privatization of the public enterprises in Europe (Vickers and Yarrow, 1988; Bishop et al., 1995). If the privatization of public enterprises – the selling of shares – was to be a successful endeavor, investors needed to know which type of regulation would prevail after privatization and they had to be convinced that investing in a formerly public enterprise would be profitable. To do so, a new system of regulation was developed, originally based on the ideas of Professor Littlechild (Bartle, 2003; Beesley and Littlechild, 1989; Rees and Vickers, 1995). In the UK, legislation on privatized utilities determines that regulation should be such that utilities ensure adequate quantity and quality of supply, that competition is promoted, that efficiency and economy are promoted and that the interests of consumers are protected. To accomplish these goals, *price-cap regulation* was instituted (Armstrong et al., 1997). Utility prices were determined according to the formula $RPI - X$, where $RPI$ stands for the retail price index and $X$ for the productivity increases of the production factors employed. If, for example, inflation rises by 5% and productivity increases yearly by 3%, then regulated firms are allowed to maximally increase their prices according to the cost increase, which is 2%. Procedurally, to regulate a firm, a financial model of the firm is devised and consultation, documents published on the internet so that interested parties can react. After this period of consultation, the regulator decides, among other things, on the price-caps, which are either accepted by the regulated firm or contested by appealing to the courts. The entire procedure takes about one to two years, but the decisions on the price formula and thus the X-factor will remain in place for about five years. This means that, if a firm succeeds in lowering its costs more than the productivity increases reflected by the X-factor, its profits will rise equivalently at given prices. Schematically, the X-factor is determined as shown in Figure 1.12 (adapted from Baldwin and Cave, 1999, pp. 229–31).

The X-factor determined at the price review thus reflects the unanticipated efficiency savings of the past, the forecast efficiency savings, desired quality improvements, a desirable ‘glide path’ adapting prices to costs, and more. Price-cap regulation differs in the following ways from rate of return regulation (Beesley and Littlechild, 1989; Newbery, 1998; Sappington and Weisman, 1996; Crew and Kleindorfer, 1996; Baron, 1995, Vogelsang, 2002). Unlike rate of return regulation, price-cap regulation focuses particularly on motivating managerial efforts and investments in decreasing operating costs. For this reason, it is also called incentive regulation. Also, price-cap regulation proceeds from projected efficient cost. Regulators determine these projected costs on the basis of efficiency studies, comparative analysis with other firms (yardstick competition; see below) or from general productivity trends. Because price-cap regulation is only applied to the network or monop-
part of the firm and not to the competitive parts, where competition is expected to discipline the firm, this type of regulation is administratively less cumbersome compared to rate of return regulation. The price-cap furthermore concerns a ‘basket’ of prices which means that the structure of prices is for the most part determined by the managers of the firm. They will set prices to maximize profits, which often also contributes to efficiency. Also, regulators have more flexibility and discretion compared with US regulators: there are more degrees of freedom in setting X than in determining the rate of return. Finally, the length of time between the formal price reviews is four to five years, unlike rate of return regulation where rate cases might take place each year, for example in times of inflation. This could mean that firms earn excess profits for a period of up to five years.

There are also a number of obvious drawbacks from this type of incentive regulation. If it is difficult to observe quality, and quality requires spending resources, the firm may be motivated to decrease costs by decreasing quality, thus improving its profits (Elliot, 2006; Sappington, 2005). The firm might thus invest less in maintenance, reliability, frequency, and the like. In practice, regulators have sometimes included an extra factor in the formula to motivate managers to reach certain quality levels. Prices are allowed to increase more when those levels have been reached. Also, the regulator will gradually come to learn the firm’s realized costs. It may be tempting for regulators not to proceed from projected efficient cost, but from what they have learned the firm can accomplish. But if they do so, the firm will anticipate this and will not be motivated or behave strategically in its efforts to minimize costs. At the end of the regulation period, managers will, for example, postpone investments in efficiency. Of course, regulators should also be committed not ‘to claw back’ any excess profits the firm has acquired by maximizing the difference between its

\[
\text{Revenues} = \text{Expected revenues} = \text{Expected revenues} - \text{Expected costs} = \text{Efficient operating costs (OPEX): labor, energy, inputs, depreciation, taxes, etc.} + \text{Capital costs (CAPEX): plants less depreciation, allowed rate of return on asset base (old and new investments)}
\]

\[
\text{Determination of } X \quad \text{Revenues should equal costs}
\]
revenues and costs. Again, if they do, they will undermine the motivation of the firm’s managers to maximize profits and decrease costs. Finally, price-cap regulation is best suited as an incentive regulation to promote the cost efficiency of firms. But if the primary goal is to have more investments, for example to renew the sewerage network or to install a fiber-optics network, it is not obvious that price regulation is the best instrument to deal with such issues. Investments are more commonly thought to be motivated by profits rather than by prices.

The comparison makes clear that price-cap regulation and rate of return regulation are actually two polar forms of regulation (Laffont and Tirole, 1993; Newbery, 1998; Cowan, 2006; Joskow, 2008). Regulators trade off the objectives of cost efficiency and profit extraction. The regulator does not know the firm’s cost and its potential, nor is the level and composition of the firm’s demand fully known by the regulator. The regulator thus faces a dilemma: it can take away all the firm’s profit, but then managers have no incentive to reduce the operating cost. And furthermore, in order to induce the firm to produce, it has to set a high allowable revenue anyway. But to motivate efforts in cost efficiency, the regulator also needs to let the firm keep some of its profits. Regulators have tried to escape this dilemma by reducing their information disadvantages. For example, they have used benchmarking, yardstick competition and bidding processes to become better informed about the cost opportunities the firm actually has (Sections 3.2.3 and 3.2.4).

There appears to be a general economic problem of regulation here (Newbery, 1998). The regulator has to determine the allowable revenue for the firm. This can specified as \( P = bP^c + (1 - b)C \), where \( P^c \) is a price independent of the firm’s average cost, \( C \). If \( b \) is zero, the allowable price or revenue is equal to the firm’s cost, which compares to rate of return regulation. The firm is thus allowed to cover its cost but no more and it has no incentives to reduce its costs. If \( b \) is 1, the formula reduces to the allowed price being equal to a fixed price, which corresponds to price-cap regulation. The firm is allowed to keep all of its profits. The term \( b \) is the incentive ‘power’ of the regulatory scheme: it is at its highest if \( b = 1 \). For example, assume the firm’s unit cost are €10. The regulator sets the incentive power to zero and lets allowable revenue to be only equal to the firm’s cost of €10. Alternatively, the regulator could set the incentive power term at, for example, 0.7. Let us assume that initially the average costs are equal to the projected efficient cost, \( P^c \). At the efficient cost of €10, the firm can still recover its costs because the allowable revenue is also €10 = (0.7 \times 10) + (0.3 \times 10). However, let us now assume that the firm can reduce its costs by better operating procedures to €9 per unit. The projected efficient costs are still €10, but the actual costs to the firm have decreased to €9. Allowable revenue for the firm now falls to €9.70 = (0.7 \times 10) + (0.3 \times 9). The firm is rewarded for its managerial efforts to reduce its
cost by an additional profit of €0.70 per unit. The consumer benefit of these efforts is €0.30. Of course, given the objectives of the regulator to extract profit and to minimize cost, such a scheme is only efficient if these cost reductions are truly the result of managerial effort and not the result of some exogenous decrease in input prices (see generally Lyon, 1996; Mayer and Vickers, 1996; Hawdon et al., 2007).

3.2.2 Competition among networks

It was mentioned above that the privatization and regulatory reform of public enterprises embodied the perspective of introducing competition to these public utility sectors. From a public interest perspective, the literature offers at least five arguments why deregulation and the introduction of competition might be desirable, despite the fact that the industries involved are characterized by decreasing cost per unit supplied (Joskow, 1996; Economides, 2004; Klein, 1999; Newbery, 2004). First, competitive pressures will force the incumbent firm to lower the X-inefficiencies it incurred during the process of regulation. The resulting cost decreases may be such that lower market prices will result, despite the inherent tendency of unit costs to increase that results from the entry of one or more firms into a ‘natural monopoly’ market. Second, inefficiencies resulting from rent-seeking by the firm or inefficiencies resulting from uninformed or rent-extracting regulators will decrease. Third, since regulated firms are often also active in adjacent but unregulated markets, the introduction of competition will make it more difficult for such a firm to pursue predatory or cost-raising strategies toward rivals. Fourth, if the industry is being restructured (separated), the firm operating the network will have an incentive to profitably increase the capacity and quality of the network, instead of artificially cutting back on capacity to keep out rivals in downstream markets. And finally, depending on the possibility of horizontal (regional) restructuring, the introduction of competitive pressures by means of yardstick competition will become possible.

However, whether or not these benefits will actually materialize crucially depends on the appropriateness of the governance structure in relation to the particular industry or country characteristics involved (Armstrong and Sappington, 2006; Parker, 1999; Levy and Spiller, 1996; Kessides, 2004; Crocker and Masten, 1996). As mentioned earlier, the industries involved may be characterized as network sectors. They are usually integrated sectors of production, distribution and retail, where the distributional part has network characteristics (pipes, wires, railways, runways, etc.). Ideally, if market demand were such that several networks could be supported or that several substitutes were available, the benefits of competition could materialize, given sufficient general antitrust enforcement. More likely, these networks will be of different size, which would then require access regulation to interconnect...
them. For example, in the market for telecommunications several networks are usually available to provide telecommunication services (telephone, internet, television signals). Differences in size may also require asymmetric regulation, for example by obligating only the former monopolist to provide a universal service (Armstrong, 2001; Crew and Kleindorfer, 2002). Difficult tradeoffs would have to be taken into account by sector regulators, for example stimulating competition at the take-off stage of competition through easy access requirements, and later on gradually leveling the playing field when more mature stages of competition have developed. But even the existence of several networks or substitutes does not guarantee that effective competition will take place. Switching costs may effectively lock customers in to existing suppliers, preventing them from changing suppliers even if it would be efficient to do so (Farrell and Klemperer, 2006; Gans and King, 2000; Waterson, 2003). Again, complex regulatory tradeoffs are involved, since regulations to lower switching costs may actually increase entry costs and market prices.

3.2.3 Competition for the market: franchise bidding as an alternative to regulation

In several instances, market demand or technological characteristics are such that competition between networks or substitute products is unavailable. Possible alternatives to competition in the market are competition for the market or the use of yardstick competition. The alternative of competition for the market was discussed as early as the 1960s (Demsetz, 1968) and has since then also been put into practice in such fields as cable television, railway passenger services, local bus transport services, the water industry, the postal sector and highways (Cambini and Filippini, 2004; Otsuka, 1997, Baldwin and Cave, 1999; Littlechild, 2002). The essential idea is to organize a bidding process to allocate an exclusive right (the franchise) to carry out certain activities (broadcasting television or radio signals, delivery of local bus passenger services, operating a distribution network). The benefits, compared with the traditional rate of return regulation are obvious. The government no longer needs information on costs and demand to achieve optimal pricing, a regulatory agency need not be established and the cost inefficiencies resulting from regulation are not present. The most cost-efficient firm is automatically selected in the bidding process. However, as soon as the bidding process is based not solely on price but also on quality, matters become more complicated. Although bidding on items with multiple dimensions is possible, a number of the benefits of the simple price-bidding schemes disappear. Not only is the transparency of the bidding process often diminished, resulting in for example more court cases, but also the enforcement of the franchise contract becomes costly. Quality dimensions (frequency, reliability, security, etc.) are not easily observed and have to be defined, measured and monitored.
This requires an ongoing monitoring institution. Rent-seeking efforts will increase with the increasing responsibilities of the franchising agency. Furthermore, difficult tradeoffs with respect to the duration of the contract have to be made. Short-term contracts encourage compliance with the terms of the contract, but discourage incentives to invest. And long-term contracts, which seem appropriate in the case of network investments, must allow for changes in demand and cost conditions. Not only are such contracts difficult to write, but they also require continued monitoring and they bring with them options to behave opportunistically for both the franchisee and the franchisor. And finally, competition requires a sufficient number of applicants who want to obtain the franchise agreement. However, the incumbent franchisee has a number of advantages that are difficult for rivals to overcome. Not only is she better informed about the market, the franchise process and the franchising agency, the agency itself might prefer to continue to do business with the incumbent franchisee. There’s always a risk in changing the status quo and the possible lost benefits of an alternative provider do not have a ‘residual owner’ who could protest against conservative decisions. In practice, the incumbent often wins again at re-franchising (Nash and Smith, 2006; Dalen et al., 2006; Zupan, 1989).

3.2.4 Vertical separation, horizontal separation and yardstick competition

One possible way to reduce the complexities arising from the desire to franchise integrated public utilities is to separate the different functions of the integrated utility. Different stages in the provision of goods and services can be distinguished, such as for example production, distribution and retail, or building, operating and transfer of the infrastructure. Identifying the benefits and costs of vertical or functional separation is a contentious issue in the literature (Shelanski and Sidak, 2001; Sappington, 2006; Jenkinson and Mayer, 1996, Newbery, 2002). Vertically separating the industry would appear to have a number of advantages, with respect to franchising, but also with respect to regulation. First, if competition among multiple networks is unavailable, it would allow the identification of the network part of the industry that should be subject to regulation or to public franchising. Competition would increase in the other stages. Second, it would make the franchise contract less complicated which, ceteris paribus, would increase the number of applicants. And third, if the network was the bottleneck facility and if that part was to be separated from the production and marketing stages, the incumbent firm would no longer be in a position to strategically abuse its ownership of the network by practicing anti-competitive strategies such as price squeezes, raising rivals’ costs, etc.

Separation also has a number of disadvantages. Coordination between activities at different stages now has to take place by means of contract instead
of by means of internal managerial command-and-control methods. These contracts are difficult to write and enforce. If trains are delayed, is this because of the maintenance of the network or because the train operator performed inadequately? Economies of scale and scope of integration will be lost. The increase in risk for companies using the network will be reflected in fewer participants in the franchise auction and in higher bidding prices. Separation will also diminish the incentives to invest if operating costs increase or if not all revenues from these investments can be appropriated. Furthermore, if the integrated firm is no longer allowed to enter adjacent or downstream markets, competition in those markets will diminish. Finally, vertical separation may give rise to the double marginalization problem. This means that the separated firms at successive stages in the supply chain will now all set their margins on their costs. That leads to higher prices, lower profits and lower welfare. The fundamental tradeoff with respect to vertical separation thus concerns the benefits of competition versus the investments and coordination benefits of integration.

In some industries, such as the water industry, it is for technical reasons not possible to vertically separate the industry into, for example, a competitive production stage and a monopolist distribution stage. However, other such industries (electricity, passenger services) are usually organized regionally, which opens up the possibility of introducing rivalry by means of benchmarking or yardstick competition (Shleifer, 1985; Sobel, 1999; Yatchew, 2001; Cowan, 1997; Parker et al., 2006). If the industry was originally organized nationally, horizontal (regional) separation would be a logical way to introduce this form of competition. The essential idea is to compare the average cost of firms in different regions and correct those costs for individual factors influencing these average costs (population density, ratio of business versus residential customers, environmental factors, etc.). The regulator will then set the price of a firm’s product in a particular region equal to the average costs of all firms, excluding the average costs of this particular firm. If the regulated firm now reduces its own costs, its profits will increase. If prices for all private but regulated firms were set in this way, this would motivate all firms to be as cost efficient as possible. Cost-inefficient firms would be penalized by making losses.

This system, unfortunately, also suffers from a number of drawbacks. First, it is difficult to find comparable firms, because market conditions differ and because of differences in past investments. Second, a sufficient number of firms must be available to make such a comparison. Extending the horizontal separation of the industry too far will result in diseconomies of scale and scope. Third, there is a danger of collusion among the managers of firms to maintain high cost levels. For these and other reasons, another regulatory strategy has sometimes been to keep firms integrated but to demand access to their
essential facilities and to demand interconnection with alternative or rival networks.

3.2.5 Vertical integration and access regulation

An alternative to structural separation of the industry is the introduction of competition by means of liberalization of the industry and the introduction of access regulation. It has been suggested that the most distinctive development in the regulation of telecommunications is the shift from rate regulation to access regulation (Spulber and Yoo, 2005; Tardiff, 2006). Different forms of access to a network can be distinguished, such as, for example, retail access, interconnection access, unbundled access, and wholesale access. Retail access is provided for final customers and concerns the right of end users to obtain the services provided by a network. Often these services are provided at regulated ‘reasonable’ and ‘non-discriminatory’ retail rates. Interconnection access refers to connections between competing networks, providing access to each other’s facilities. Interconnection is necessary to ensure that a customer on network A is able to communicate with a customer on network B. Unbundled access refers to the possibility of competitors using connections in the incumbent’s central offices so that they can directly connect to the customer’s premises. Wholesale access involves providing access to a competitor in order to provide wholesale services. The access could be negotiated and contracted by the incumbent monopolist or it could be mandated by the regulator. You can think of services such as metering and billing, customer premises equipment, internet and data, and more. Liberalization makes possible the entry of competitors who either own and operate their own network or supply services over existing networks. For example, internet service providers may provide access to the internet using the network of the incumbent telecommunication company. On the one hand, by not separating the incumbent firm, important economies of scale and scope can be maintained. On the other hand, several types of costs are involved when competition is introduced by means of access regulation (Spulber and Yoo, 2005). It requires the incumbent firm to transact with competitors and thus to incur transaction costs. It may involuntarily have to increase its capacity. In order to adjust its network capacity, it may have to reconfigure its network and all these changes require adjustments costs. Furthermore, as a result of mandating access, bottlenecks in using the network and congestion costs may appear. These effects mainly occur because the incumbent firm is forced to expand its network and to transact and contract beyond its voluntarily and optimally chosen organizational structure. Regulatory enforcement will also be costly. Since firms will obviously appeal against the terms on which mandated access is negotiated with the incumbent, the telecom regulators will have to decide on difficult pricing issues (Vogelsang, 2003; Sarmento and Brandao, 2007; Laffont and Tirole, 2000).
Digitalization of communication services and the replacement of copper cables by fiber-optic networks that is currently taking place, will change market structures significantly. Telephone, television and internet services are now often provided over the same network, a phenomenon called ‘convergence’. Both cable companies and telecommunication companies will be able to offer these so-called ‘triple play’ services. The prediction is that these future changes in market structures will imply fundamental changes for regulators (De Bijl and Peitz, 2008; De Bijl and Huigen, 2008; Cave, 2006). Where originally the expectation was that effective competition would take place through several competing networks, and that regulators could withdraw from this market, it is now seriously considered that new technologies will drive the market towards a new natural fiber-optics monopolist. If such a network becomes separated from competitive upstream and downstream markets, should governments take the lead in developing these networks, and if commercial parties operate the network, how should they be regulated? These are just a few of the questions that will need to be answered in the near future.

4. OUTLOOK ON FUTURE DEVELOPMENTS OF REGULATION

In the past two decades, regulators have been successful in lowering the operating expenditures of regulated firms. However, it has been noted that massive investments will be needed in regulated sectors in the near future (WRR, 2008). Huge investments are needed to meet future energy demands, to replace ageing sewerage systems and power networks, to protect against climate change (floods), to reduce or store carbon-dioxide emissions, to modernize congested transport systems, to finance renewal of electronic communication systems (fiber-optics), and soon. The question arises of whether current regulatory institutions and practices are up to these challenges. Can we expect regulatory instruments focused primarily on motivating managers to lower operating expenditures to be just as successful in motivating firms to spend vast sums on uncertain, irreversible, long-lived and innovative investments? Can we expect this to happen, particularly where:

- regulators in the past have allowed only ‘prudently’ incurred investment to receive a fair rate of return;
- regulated firms were left with ‘stranded’ assets after the introduction of competition;
- firms were not allowed to merge to an efficient scale, solely so as not to deprive the regulator of a sufficient number of firms necessary for benchmark purposes;
• regulated firms were forced to compete with a hypothetically efficient firm in an optimally configured network that obviously was not stuck with past investments;
• investments are long-lived, but regulatory reviews take place every five years;
• construction time tends to be large and demand uncertain, so that the investments may not cover their costs and recoupment may not be allowed since investments must be ‘used and useful’;
• excess profits, the driving force of most investments, above the allowed rate of return are taxed away at the next regulatory review by raising the X-factor or by requiring higher quality levels of services to be provided; and,
• regulators have a statutory duty to promote the interest of consumers and returns on investments earned on the regulated company’s assets are up to one third of the total price paid by consumers?

In the US, the regulation of public utilities took the form of a ‘regulatory contract’ between investors on the one hand, and society as represented by the regulatory commissions on the other. Decisions by the integrated firms on fundamental business variables, such as direction of investments, investment levels, levels of prices, price structures, were all discussed at public hearings and were decided in so-called ‘rate cases’.

The EU type of regulation was for the most part driven by the desire to make the formerly state-owned enterprises more efficient and to accomplish that task in an efficient way. To those ends, independent regulatory institutions have been developed and several regulatory strategies have been pursued. However, the issue is no longer to make the former state-owned enterprises more efficient, and in an efficient way. The relevant issues are much more strategic and concern combining business investment programs with fundamental tradeoffs, not only between different consumer groups, both now and in the future, but also between different investor groups and between the interests of investors and those of consumers at large. The question arises whether independent regulatory authorities are up to that task, given that they are not accountable to the public nor equipped to evaluate difficult tradeoffs between the interests of different investor groups and different consumer groups. It would seem that legislators and regulators will be facing fundamental challenges and changes in the years to come.

BIBLIOGRAPHY


Buchanan, James M., Robert D. Tollison and Gordon Tullock (1980), Toward a Theory of the Rent-Seeking Society, College Station, TX: Texas A&M University Press.


Kalt, Joseph P. and Mark A. Zupan (1990), ‘The Apparent Ideological Behavior of Legislators:


90 Regulation and economics


Ng, Yew Kwang (1990), Welfare Economics, London: Macmillan.


Economic theories of regulation


Pindyck, Robert S. and Daniel L. Rubinfeld (2001), Microeconomics, New Jersey: Prentice Hall.


Regulation and economics

Economic theories of regulation