Economic regulation principles, history and methods

1 author:

Martin Ricketts
The University of Buckingham
76 PUBLICATIONS **502** CITATIONS

Economic Regulation: Principles, History and Methods

Martin Ricketts   28th May 2005
Department of Economics and International Studies
University of Buckingham, UK

Introduction

Economic regulation is an important instrument of government policy in market economies. We do not talk of regulating planned systems. To regulate implies the exercise of some influence on an activity that is different from total 'control'. It is no accident therefore that the economics of regulation has become increasingly important in recent years as direct state ownership has declined. The perceived failure of central planning has not in itself discredited all government attempts to improve economic performance. Indeed, the idea that the provision of certain limited but crucial regulatory functions by the state is necessary for economic advance is a well-established part of classical liberal (as distinct from libertarian) economic and political theory.

Regulation is not, however, a very precise term. Some forms of regulation are concerned with setting a framework of rules for people to follow in their dealings with each other. In this sense the law of contract or property would comprise part of the regulatory base of the economy. A financial 'regulation', for example, that all companies must disclose price sensitive information to the market by means of a public announcement within a specified time period would be of this 'rule setting' type. The regulation applies to all market participants and is 'end independent' in the sense that no one could possibly know exactly what would actually happen in every case as a result of introducing it. Although the detailed outcomes are not known, the regulation could be defended as a means of making the market process more efficient over the long term.[1]

Other regulations are more prescriptive in form. They instruct people to achieve particular ends, often by the implementation of specified means. An example would be a regulation instructing a firm to install certain machinery in order to reduce pollution or improve safety. Measures of this type do not merely establish the rules governing the behaviour of transactors in the market; they attempt to determine the results or to confine them within 'acceptable' limits. Regulations that forbid transactions below a specified price (for example a minimum wage) or that impose restrictions on hours worked may still leave some room for agreement, but they directly rule out contracts with 'unacceptable' characteristics.

This distinction between regulation as the 'planning' of collectively determined desired ends and regulation as the 'governance' of a continuing decentralised

---

[1] The regulation might also be attacked on the grounds that it forces firms to divulge sensitive plans to competitors earlier than would otherwise be the case and that this might reduce the dynamic efficiency of the market over the long term. It is not the desirability of this regulation, however, but its character that is the point at issue.

market process is important. At the conceptual level it seems clear enough - as clear as the distinction between a referee who fixes the outcome of a game and a referee who simply enforces the rules. In practice, of course, it is sometimes difficult to tell them apart. All referees find themselves accused of rigging the rules or interpreting their provisions to the advantage or disadvantage of particular interests. Nevertheless, the conceptual distinction is at the heart of many of the disputes concerning the role of the state in a market economy. Nee (2002) for example supports Polanyi's (1944, p.140) thesis that a market economy requires administrators to be 'constantly on the watch to ensure the free workings of the system'. He argues that there is a causal link between the emergence of a 'legal-rational bureaucracy' (Weber 1947) and the support of market institutions and applies this idea to the transition of modern day China. Others are more sceptical about the role of bureaucratic agents because they regard them not as abstract rule enforcers but as direct decision-makers with a high potential to distort rather than to assist the allocation of resources.[2]

The following sections discuss in more detail the tension between these two broad traditions in the economics of regulation. The paper starts by contrasting the standard economic theory of regulation with alternative schools of thought. Later sections investigate the problem of 'natural monopoly' and the various methods used in the regulation of the public utilities.

The Public Interest Theory of Regulation

One of the most important consequences of the 'marginal revolution' of the 1870s was the development of a far more rigorous and formal welfare economics than was provided by the preceding classical analysis of economic policy.[3] The new theory was associated with theorists such as Walras, Pareto and Pigou. At its heart was an investigation of the properties of a general competitive equilibrium. In this Walrasian equilibrium all economic agents chose the amount of labour and other factors of production to supply and the basket of goods to consume by maximising their utility at the prevailing perfectly competitive prices. These equilibrium prices were such that no aggregate excess supplies or demands existed in the market for any traded good or service. All markets cleared.

To demonstrate the existence of a general competitive equilibrium was a great intellectual achievement in itself but it was the normative properties of this equilibrium that came to form the basis for economic thinking about public policy to the present day. It could be shown that provided all goods and services were traded and that private benefits and costs were the same as social benefits and costs (there were no externalities or 'spill-overs') a perfectly competitive equilibrium was Pareto efficient. No one could be made better off without making someone worse off. The gains to trade were exhausted. Maximising social welfare required the marginal social benefits of

---

[2] Ricketts (2002) elaborates on these points in the context of Nee's paper.
[3] See Robbins (1952) for a discussion of classical ideas on economic policy.

each and every activity to equal the relevant marginal social costs. Given that a condition for utility maximisation on the part of each consumer was that the marginal benefit of a good or service should be brought into equality with its market price; and given that a condition for profit maximisation on the part of each producer was that the marginal cost of a good or service should be brought into equality with its market price; and given that all consumers and producers faced the same market prices; it is possible to see quite intuitively how this theorem was derived. Equilibrium prices represented marginal benefits to consumers and marginal costs to producers.

The 'First Theorem of Welfare Economics' had very complex repercussions for public policy. It was capable of leading in two very different directions depending upon how it was interpreted. One response to the 'first theorem' was to argue that since perfect competition had such desirable properties, public policy should concentrate on removing impediments to competition. To the objection that competition might result in a very unequal distribution of income a 'Second Theorem of Welfare Economics' could be used to show that any desired distributional result could be achieved through appropriate lump sum taxes and transfers. Thus the recommended policy package was to achieve efficiency through competitive markets and equity through 'non-distortionary' taxes and transfers. In so far as economic regulation appears at all in this system it takes the form of competition policy.[4]

An alternative response to the 'first theorem' was to emphasise the austerity of the formal conditions. No matter how 'competitive' the real economy seemed to those embroiled in its day-to-day tumult, it was obvious that it would never satisfy the requirements of perfect competition. Indeed almost all competitive behaviour paradoxically could be shown to be incompatible with perfect competition. Price shading, for example, immediately contravenes the formal requirement that all contractors are 'price takers'. The introduction of new products or processes confers monopoly power for short periods. Advertising would be redundant in a world of perfect information. If this were not enough, it was also clear that increasing returns to scale prevailed in some industries, markets were incomplete and the assumption of no external benefits or costs was unwarranted.

Viewed from this second perspective, the role of economic regulation was potentially very extensive. Markets did not have the characteristics of the perfectly competitive model and therefore could not be expected to achieve economic efficiency. Regulators were required to correct for 'market failure'. Industries subject to declining costs required a subsidy since the marginal cost of extra output would be below the break-even price. Polluting activities should be taxed according to the external damage inflicted. Monopolies should be instructed to set prices equal to marginal costs. Jointly consumed or 'public goods' such as environmental quality and public information required the intervention of regulators to ensure an efficient level of provision.

---

[4] A more extended description of the first and second theorems of Welfare Economics is provided in an appendix.

In this tradition, therefore, regulators were seen as technicians acting in the public interest and correcting for the failures of the market. It was implicitly assumed that they had both the information upon which to act and the appropriate motivation. This idea of the professional public servant was also a requirement for the macroeconomic regulation that came out of the Keynesian revolution of the 1940s and 1950s. The economy was understood as a system of equations, the public objectives were well defined and the technical task was simply to control the values of a set of policy instruments so as to bring about the best possible result, social welfare maximisation. Regulators were in the Platonic tradition of philosopher kings. Later criticism was to refer to these somewhat elitist assumptions of Keynes as 'the presuppositions of Harvey Road' – the Cambridge road in which Keynes was brought up. Whether based upon general equilibrium theory or Keynesian macroeconomics, however, the role of the regulator is better described as Hobbesian. The sovereign authority is used to impose a solution where the social outcome would otherwise be disadvantageous because individuals are unable to come to enforceable agreements. The law is structured so as to 'minimise the harm caused by failures in private agreements'.[5]

Public Choice and the Theory of Regulation

From the 1960s onwards the 'public interest' approach to economic regulation was subjected to increasing criticism. This was to some extent explained by events. Keynesian macroeconomic regulation had not produced stability although a depression of pre-war intensity was not experienced and it was inflation rather than unemployment that became the problem of the day. Similarly the microeconomic regulation of markets in the United States and the performance of those industries nationalised by the post-war Labour government in the UK did not seem to result in clearly observable efficiency gains. Stigler and Friedland (1962), for example, could find no evidence that electricity prices differed during the inter-war years between those States with regulation and those without, while according to Stigler (1964) the Securities and Exchange Commission had been similarly unable to produce measurable results. When clear effects of regulation were observed, they did not always accord with the expectations of normative economic theory. The regulation of rents, for example, led to investigations of its efficiency costs rather than its benefits[6]. Regulation was sometimes used to keep prices below marginal costs for some groups by setting prices above marginal costs for others, thus using the price system as a redistributive tool in direct opposition to the second theorem of welfare economics.[7]

One response to these observations was to depart from the assumption of a disinterested public official and to model the behaviour of regulators and politicians using the same type of analysis as might be used in the private sector. If managers of public limited companies could be viewed not as

---

[5] This is sometimes called the 'Normative Hobbes Theorem'. See, for example, Cooter and Ulen (1997) pp.89-90.
[6] Olsen (1972).
[7] See, for example, Posner (1971).

traditional profit maximisers but as maximising firm size or revenue[8] it was a small additional step to assume that bureaucrats were budget maximisers[9] or that politicians were vote maximisers[10]. If entrepreneurs in conventional theory were profit seekers they would be expected to lobby for a favourable regulatory environment once they understood that regulators and politicians were open to influence. Producer groups were expected to be particularly powerful because they were in a better position than were consumers for overcoming the 'collective action' or 'free rider' problem.[11] Stigler (1971) used this framework to advance the so-called 'capture theory' of regulation with its startling conclusion that the regulatory system comes to be operated in the interest of the regulated firms rather than the more general 'public interest'. This theory was criticised by Peltzman (1976) for not taking into account the incentive for other groups to resist as they experienced increasing marginal costs of regulation.

The investment of resources in attempts to divert income from other people through the political and regulatory process became known as 'rent-seeking', a term coined by Kreuger (1974). Rents arise when sellers receive more than the minimum that they are prepared to accept or buyers pay less than the maximum they are prepared to offer. It is a feature of a perfectly competitive equilibrium that the sum total of such rents is maximised and that therefore there is no rent on the marginal transaction. The maximum that a buyer is prepared to offer for additional output (its marginal benefit) is equal in a competitive equilibrium to the minimum a supplier is prepared to accept (its marginal cost). Voluntary trading activity that results in mutual gain is therefore a form of rent seeking that creates rents and increases efficiency. The gains to trade take the form of additional rents accruing to the various trading parties. Political pressure that results in the redistribution of rents through the political and regulatory system, however, is a form of rent seeking that is entirely wasteful. It represents a process of rent dissipation rather than rent creation because the aggregate resources invested in laying claim to rents might equal or even exceed the rents available.[12] Buchanan et al (1980) have made the analysis of rent seeking a significant component of public choice theory.

This public choice perspective on the regulatory system led to a re-appraisal of the history of regulatory growth especially in the United States. Even major historical events such as the establishment of the Inter State Commerce Commission (ICC) in 1887, long seen as a means of curbing the monopoly power of the railroad companies, could be reinterpreted as a means by which these same companies could avoid increasingly cut throat competition on long hauls.[13] The Bell Telephone Company during the first decade of the twentieth

---

[8] Baumol (1967) and Williamson (1964). Baumol was the originator of sales revenue maximisation although Williamson's theory is more appealing in that it deals with the concept of managerial slack and the appropriation of rents.
[9] Niskanen (1968, 1971).
[10] Downs (1957); Buchanan and Tullock (1962); Breton (1974).
[11] Olson, M. (1965).
[12] See Tullock (1980).
[13] For example Kolko (1965).

century supported regulation to restrict new entry after the expiry of patents had made it more vulnerable. Airline interests were prominent in supporting the establishment of the Civil Aeronautics Board (CAB) in 1938. Between 1950 and 1974 no application to start a new inter-state airline was granted by the CAB.[14] For many years the Federal Communications Commission (FCC), established in 1934, restricted the development of cable companies in order to protect the advertising revenue of local television stations. In trucking regulation Stigler (1971) explained restrictions on the weight of trucks by reference to the average length of haul on the railroads and the strength of the farming interest. Railroads with a long average length of haul would be less threatened by the advent of motor trucks on farms and would therefore tolerate a higher weight limit.

From the point of view of positive economics and as a means of explaining regulatory outcomes, public choice theory had advantages over the public interest theory of regulation. From a normative point of view the effect was to move attention away from the identification of ideal 'ends' to a discussion of what institutional mechanisms would be relatively good at achieving them. The public interest theory of regulation was a branch of what Demsetz (1969) labelled 'nirvana economics'. Actual institutions should not be compared with the nirvana of a perfect regulatory system achieving a Pareto efficient allocation of resources. They should be compared with achievable alternatives that did not depend upon unexplained changes in human motivation (the 'people could be different' fallacy) or upon unexplained and costless changes in the availability of information (the 'fallacy of the free lunch'). Attention was thus redirected towards the comparative analysis of institutions or alternative 'governance' arrangements. What 'rules of the game' are likely to produce the best results over time that can be achieved in practice? This is the fundamental problem of regulatory economics and forces us to consider 'constitutional' issues about the effects of property rights, procedural rules and individual incentives in regulatory processes.

The New Institutional Economics and Economic Regulation

Transacting is a process by which people come to mutually beneficial agreements. This fundamental observation greatly influences the way we look at economic life. Agreements do not really feature in the theory of competitive equilibrium. In a perfectly competitive world people adjust their own behaviour to a given set of prices through anonymous 'arm's length' trades. They buy and sell but do not search and bargain. There is nothing left to search for and nothing to bargain about. By following the tradition of Edgeworth (1891) rather than Walras, however, and by investigating the way that people contract and then recontract to their mutual advantage a different light is thrown on the ultimate meaning of 'market failure'. If there were unexploited gains to trade, why would free transactors not come to an agreement to secure them?

By framing the question in this way we are led to look for the forces that stand in the way of exchange. Transactors might not be aware of the potential gains

---

[14] For these and other examples see Wilson (1980).

to trade; they might not have secure legal title to the rights over which they wish to transact; they might be unable to police or enforce any agreements arrived at; and they might face a protracted and difficult process of bargaining. In other words, exchange is a process that involves costs and requires supporting legal institutions. Where transactions costs are low and property rights well defined and enforceable, however, private agreements would be expected to achieve most of the gains to trade. Coase (1960) was the first to look at instances of 'market failure' from this perspective. He pointed out that the Common Law already furnished many cases that illustrated this basic idea. In cases of nuisance, the Courts would grant injunctions and then permit private bargaining to achieve an efficient outcome if continuation of the offending activity were of sufficient social value. The proposition that with clearly assigned property rights and zero costs of transacting people will bargain to an efficient allocation of resources is known as the 'Coase Theorem'.

The Coase Theorem applies in conditions that are as rarefied as those underlying a Walrasian general equilibrium. Its implications for the theory of economic regulation are, however, rather different from those derived from static competitive analysis. The normative version of the theorem states that the law should be structured 'so as to remove the impediments to private agreements'.[15] In other words, the role of the regulator is to clarify property rights and reduce the costs of trading them. The Coasian regulator is the facilitator of the trading process and the provider of transactions cost reducing institutions rather than the achiever of ideal end states. However, in a world where transactions costs will always be significant and can never be removed entirely, Hobbesian judgements cannot be avoided and the regulatory process will not be entirely end-independent. The initial assignment of property rights, for example, will have no efficiency consequences where transactions costs are zero. But where trading is costly, it can matter to whom property rights are assigned. If, for example, airport landing rights were allocated collectively to local residents the transactions costs of re-assigning them to airlines would be formidable. The reverse process might be equally costly so that the rights would tend to stay where they were assigned and ex ante judgements about desirable outcomes would be an unavoidable element in the design of a suitable trading system for the allocation of congested air space.

The New Institutional Economics has resulted in more than a simple recognition of the existence and importance of transactions costs in economic organisation.[16] By interpreting observed institutions as contributing to the reduction of transactions costs, attention is drawn to the means by which these institutions develop. Institutions such as money or property can be seen as 'internal' - that is, developed by the repeated interaction of transactors over time in a quasi-evolutionary process rather than 'external' and imposed by a powerful sovereign. Repeated co-ordination games result in the establishment of conventions of behaviour, conventions with which each participant has an incentive to comply and which only later may be codified into formal law.[17]

---

[15] Cooter and Ulen (1997) pp.89-90.
[16] For a review see Kasper and Streit (1998).
[17] See Sugden (1986).

Regulatory arrangements, broadly conceived, are not therefore confined to those introduced by the legislation of the modern state and, indeed, where the state acts it will often be as a substitute for existing private methods.[18] In the sections that follow, however, it is the regulatory activities of the state that are of central concern.

<u>Natural monopoly</u>

Where a single supplier is technically able to serve the entire market at lower total cost than any feasible combination of two or more suppliers, the market is said to be a 'natural monopoly'. Theoretically the condition for natural monopoly is that cost functions are 'sub-additive'.[19] Empirically this means that natural monopoly is associated with industries that require the input of large indivisible amounts of fixed (and usually sunk) capital. Historically, the sectors most subject to natural monopoly have been the 'public utilities' such as gas, electricity, water and telecommunications as well as various forms of transport. At the core of these industries are 'indivisible' assets such as pipes, cables, wires, satellites, rail and road links, waterways and so forth. Even classical liberal economists such as Smith, Mill and (in more recent times) Baumol and Buchanan recognised the desirability of some form of government regulation in the presence of natural monopoly or 'extreme jointness efficiencies'.[20] Smith argued for regulation of the toll roads and Mill discussed the case of canals. We have already noted how the regulation of railroads began in the United States in the 19th century and was extended in the 20th to telecommunications, broadcasting and electricity.

In more recent times, greater attention has been given to network economies in consumption compared with the older focus on simple economies of scale in production. The more users a communications network attracts, for example, the more valuable its services become and, up to some capacity limit, the lower become the costs per unit of service delivered. Economies of scope are also often associated with modern public utilities. These exist where it is less costly to produce different products together rather than in separate firms. These will again ultimately derive from some joint input that can be used simultaneously in the production of several different products. At the retail level, for example, there are economies in metering and billing if consumers deal with a single company for gas, electricity and water. Economies of scope thus favour the growth of diversified rather than specialised firms.

From a Coasian point of view the problem of natural monopoly is the unfeasibility of putting together the contracts that are required for efficient provision. A natural monopolist who could negotiate costlessly and individually with all users for services to be rendered in all future time periods would not

---

[18] The 'lex mercatoria' of the medieval period was a system of rules governing trading that did not depend upon state power. Blundell and Robinson (1999) discuss the possibilities in the modern era of regulation that is voluntary and not dependent upon the intervention of the state.

[19] See the Appendix for a more detailed discussion.

[20] See Baumol (1965) and Buchanan (1975) p.97.

require regulation on standard efficiency grounds. A monopolist who must sink substantial amounts of capital in advance of any such agreements and who has limited ability to discriminate in price, however, will maximise profit by inefficiently restricting access and charging a price in excess of marginal cost. Some form of collective process is required to overcome this problem. The development of co-operative arrangements in which consumers become 'owners' of the natural monopoly enterprise, for example, can occur quite spontaneously.[21] If consumers collectively hold control rights to the assets they can, subject to the inevitable costs and inefficiencies of collective choice processes, ensure that the monopoly operates in a way that serves their interests.

A 'club' is one organisational response to the problem of natural monopoly. It does not have to take the form of 'the state', but over much of the twentieth century the nationalisation of natural monopolies was common nevertheless. In the UK, for example, the gas, electricity, water, rail and telecommunications industries were all under state control by the 1950s.[22] This represented the ultimate Hobbesian solution. Failures in private agreements would be fixed by managers appointed by the state. Managers were specifically required 'to further the public interest in all respects'.[23] This rather vague injunction was refined by technical advice over time to include setting prices related to marginal social costs and undertaking investment projects with positive net present values calculated using a test discount rate.[24] If efficient pricing would lead to financial losses, two part tariffs involving overhead charges were recommended.[25] Effectively the managers were being asked to bring about the efficient outcomes in the natural monopoly sectors that a perfect market might conceptually achieve for the other sectors.

The criticisms levelled during the 1960s and 1970s at the regulatory agencies in the US, mentioned above, were mirrored in the UK by a similar critical response to the performance of the nationalised industries. By the late 1970s official reports and academic studies were questioning the performance of UK public enterprises.[26] In spite of their legal status as 'public corporations'[27] independent of government, managers of the nationalised industries were subject to political pressure from ministers as well as from labour unions and other pressure groups. Prices, cross-subsidies, employment levels and investment plans were often intensely controversial. Where financial losses

---

[21] See Hansmann (1996).

[22] Nationalisation was not, of course, confined to natural monopolies but in the UK eventually included steel, coal, shipbuilding, aerospace and even car manufactures.

[23] Coal Nationalisation Act 1944.

[24] See the White Paper (1967) *Nationalised Industries,* Cmnd.3437.

[25] In conditions of declining average costs – as in the presence of economies of scale, marginal cost will be less than average cost. Thus, setting prices equal to marginal cost, as recommended by textbook theory, would result in financial losses. These losses could be covered by overhead or 'standing' charges unrelated to a consumer's actual consumption. Alternatively the losses would have to be covered by subsidies.

[26] For example, NEDO (1976), Pryke (1981).

[27] Not all 'natural monopolies' took the form of public corporations. The Post Office (which included telephones) was a government department before 1961. Telecommunications was only separated from the Post Office in 1981 with the creation of BT.

occurred it was impossible to tell whether these were a result of efficient pricing or inefficient operations. In response to this problem specific financial and productivity targets were introduced by the late 1970s.[28]

Innovations of this type only served, however, to throw into sharper relief the contracting problems faced by politicians and managers. What incentives did politicians have to determine appropriate targets, to make them explicit and to stick to them over time? What incentive did managers have to provide target-setters with appropriate information or to achieve the targets once set? Were some of the relevant targets inherently non-verifiable and therefore not contractible? These questions implicitly concerned the best response to agency problems and raised the possibility that public enterprise was inherently likely to be less productively efficient than other forms of enterprise.[29] In particular, if the government wished to purchase certain goods and services on behalf of the public, it was unclear why contracting with managers of privately owned assets should be any more difficult than contracting with managers of publicly owned assets.[30] The absence of privately exchangeable control rights in the public sector constrained the type of incentives that could be arranged. The take-over threat, the bankruptcy constraint, the direct intervention of shareholders with a significant personal stake, and the granting of stock-options were all ruled out by state ownership. Many studies during the 1970s and early 1980s attempted to investigate the relative performance of public compared with private enterprise from this property rights perspective.[31]

In cases where public and private concerns could be observed in competition with each other, the latter were usually found to have the edge in terms of factor productivity. This has been confirmed by work in the post-privatisation era that investigates the relative performance of public and private enterprise over a whole range of countries and industries.[32] General conclusions about the organisation of natural monopolies, however, were more difficult to draw because the alternative to monopolistic public enterprise was monopolistic private enterprise. Any advantages of the latter with respect to cost efficiency might not be very great in the absence of competition and might be counterbalanced by the greater exploitation of monopoly power. In other words any improvements in *technical* and *cost efficiency* deriving from privatisation were quite compatible with an overall deterioration in *allocative efficiency*.[33] The important question was whether the advantages derivable

---

[28] See the White Paper (1978) *The Nationalised Industries,* Cmnd.7131.

[29] More recently Shirley (2000) has investigated the problems of introducing performance contracts into the Chinese state-owned enterprise sector.

[30] This is an important line of argument in Shleifer (1998).

[31] Examples are Davies (1971), Forsyth and Hocking (1980), Caves and Christensen (1980), De Alessi (1974), Lindsay (1976) and Millward and Parker (1983).

[32] See for example, Picot and Kaulman (1989), D'Souza and Megginson (1999) and Shirley and Walsh (2000).

[33] *Technical efficiency* requires that output is maximised for any given level of inputs. *Cost efficiency* requires that any given output is produced at minimum cost for the prevailing factor prices. Clearly technical efficiency is a necessary condition for cost efficiency but is not sufficient unless there is no choice between different possible methods of production or technologies. *Allocative efficiency* requires that it is impossible to make someone better off by

from the existence of exchangeable control rights and the value maximising incentives that accompanied them could be combined with some protection against restrictions in output. This protection would have to come from a regulatory system acting as a substitute for public ownership.

In the natural monopoly industries, therefore, the case for privatisation was concerned as much with the incentives faced by politicians, regulators and other interests (and hence about public choice) as it was about managerial incentives. If politicians and industry managers had failed to achieve 'public interest' objectives under nationalisation, what reasons were there for expecting politicians and regulators to do better in the context of privatisation? Politicians would presumably still be pursuing votes and still be subject to political pressure, while regulators could be 'captured' and would have their own as well as the public interest to consider. The difference between nationalised and privatised utilities from a public choice point of view, however, is in the rules that govern possible political intervention. Under state ownership political pressure on managers could be direct but hidden from public view. Under privatisation, managers have fiduciary duties to shareholders, the price of the company's stock acts as a continuing signal concerning financial performance, accounts must be drawn up to meet certain accounting standards, price-sensitive information must be divulged to the market and regulatory intervention must itself accord with procedural rules. These factors might be anticipated to increase the cost and reduce the expected benefit to politicians from trying to influence management decisions. Privatisation will therefore tend to be associated with 'de-politicisation'.[34]

<u>Regulation and Competition</u>

One very significant characteristic of the regulatory system that accompanied privatisation in the UK was the duty placed upon the regulator to encourage competition. This reflected the judgement that privatisation would confer far greater social benefits if it were accompanied by a move towards competitive markets than if monopoly power remained intact. It also had important general consequences for the nature of the regulatory regime. The regulator could be seen not as the enforcer of particular outcomes but as the provider of 'governance' for a market process. This is closer to the Coasian approach to 'market failure' than the Hobbesian one. By encouraging or 'facilitating' competition the regulator is permitting the widest possible scope for the striking of new private agreements. The idea of the regulator as a rule provider for market processes is reflected in the recent merger of the offices for gas and electricity regulation into a single office, the Office of Gas and Electricity Markets (OFGEM).

---

some reallocation of resources without making anyone worse off. Economists sometimes refer to this as *Pareto efficiency*. Cost efficiency is a necessary condition for Pareto efficiency but is not sufficient. As explained in the text, a monopolist could in theory be cost efficient but, if the price set by the monopolist is above the marginal cost of output, potential social gains are available from higher levels of production.

[34] See Boycko et al (1996).

Although regulators can be portrayed as market facilitators the more directly interventionist elements are never far from the surface. The problem is the tension that exists between two distinct views of competition. One emphasises an evolutionary trial and error process of discovery in which contractors are always searching for a decisive advantage over their opponents. This is the 'Austrian' view associated with Hayek (1949, 1978) and Schumpeter (1936, 1943). If regulation is limited to the classic requirements of preventing force or fraud the result is a 'competitive order'. The other view sees competition more in terms of an athletic contest in which closely matched people try to achieve identical ends in identical conditions constrained by highly formal and often extensive sets of rules. Regulation is required to prevent cheating and the resulting system is one of 'ordered competition'.

Intervention in the 'competitive order' puts the regulator in a position analogous to that of the warden of a game reserve who monitors the natural environment and occasionally takes action to maintain the necessary habitat or cull animals that are becoming too dominant. Intervention in a system of 'ordered competition', on the other hand, puts the regulator in a position nearer to that of an administrator of the Olympic Games who monitors compliance with the rules on performance enhancing drugs. Clearly these are rather different conceptions of the regulatory process. One is overseeing the competition of predator versus prey and occasionally 'playing God' to bring about desired changes in the balance of forces. The other is trying to produce an ideal of procedural 'fairness'.

In the case of 'natural monopoly' the whole idea of encouraging competition seems paradoxical since a single producer would appear, by definition, to have a decisive cost advantage. There are, however, several ways in which the aim can be interpreted. The first and simplest is for the regulator to make sure that if, as technology advances and market size grows, activities change from being 'natural' monopolies to becoming potentially competitive, new entry is not impeded by artificial restrictions. This would represent a clearly Coasian policy of removing impediments to trade. Even if natural monopoly conditions continue to exist, the removal of entry barriers can make the market more 'contestable' assuming that the problem of 'sunk capital' is not too serious. Where there are potential entrants able to take the market from an incumbent as soon as the latter tries to exploit a monopoly position by raising prices the market is contestable. Regulatory changes in the airline and bus industries in the late 1970s and 1980s were based upon this idea. [35]

The second way in which competition has been facilitated in the 'natural monopolies' is by a policy of 'vertical disintegration'. As already noted, the natural monopoly element of most public utilities lies in the provision of network assets. In electricity, for example, it is transmission and distribution that are 'naturally monopolistic' rather than generation. Similarly in telecommunications, the provision of wires from the local telephone exchanges to domestic and business premises may be a natural monopoly,

---

[35] See Baumol et al (1982).

but this certainly would not apply to the manufacture of equipment or the delivery of various telephone services using the wires. The duty to encourage competition can therefore be interpreted as a policy of isolating the 'natural monopoly' element from the surrounding potentially competitive activities. Regulation can then be confined, in principle, to the core natural monopoly assets, and this usually entails trying to ensure access on equal terms to all competing users. The 'common carrier' is required to post regulated prices at which access to the network will be granted. This separation between the provision of network assets and their use in delivering services to consumers might be effected without breaking-up the business into its component parts and running them as separate concerns. Regulating a common carrier is more straightforward, however, if the carrier is not itself involved in using its own network. The suspicion that vertically integrated utilities will favour their own 'internal' customers over external users is costly to overcome by regulatory means. This would be especially true if there were real scope economies associated with the integration of network operators and network users.

In trying to achieve 'equal' or 'fair' terms of access of suppliers to a common carrier's network the regulator is approaching most closely the view of competition associated with sporting activities – ordered competition. The problem, however, is that to produce competition of this 'ordered' type at certain points, the regulator interferes in the whole structure of contractual relations. The degree of vertical integration is determined by regulation rather than by 'the competitive order'. It was Coase (1937) who first demonstrated that the scope of the firm could be explained by the existence of transactions costs. The process of contracting within a firm was different from the process of contracting across markets. Sometimes an activity would be more profitably 'managed' within the firm and sometimes the use of an outside supplier using market contracts would be preferred. At the boundary of the firm the cost of undertaking an activity within the firm would be the same as contracting with an outside supplier. Observed forms of business organisation, therefore, were the outcome of a competitive process driven by attempts to economise on transactions costs. The vertical disintegration of the public utilities by regulatory intervention reflects judgements not only about the desirability of ordered competition in network access but also concerning the transactional efficiency of the resulting organisational structure.

Because of the large sunk costs associated with most public utilities, reliance on competitive processes to undermine monopoly power through innovation and new entry has conventionally been regarded as unrealistic in these sectors. Competition, as has been seen above, has had to be artificially contrived. The benefits of this 'ordered competition' are widely seen as outweighing any adverse effects on transactional efficiency. In a few areas, however, transactional issues have featured prominently. In the rail industry, for example, the separation of track and signalling infrastructure from the train operating companies in the UK has been controversial mainly because of safety concerns and the perceived difficulty of assuring proper track maintenance by the use of contract. This case is interesting because the separation of track and train companies has not resulted in much direct competition between train operators over the same routes so that the

conventional justification for functional disintegration does not apply. The potential importance of transactional considerations is reflected in the decision in October 2003 to suspend the contracting-out of track maintenance by the track operator (Network Rail) and to transfer the work to internal units in order to re-establish greater control.

If transactions costs are an impediment to the introduction of 'ordered competition' in some areas, the appropriate regulatory response to the existence of 'Schumpeterian competition' is a significant problem in others. The telecommunications sector, for example, has been transformed since the 1970s by technological developments in mobile telephony, computer technology, fibre-optic cable and satellite communications. In dynamic technological conditions like this the regulator either stands aside entirely or, when trying to 'order' the competition, intervenes in ways more reminiscent of the role of the game warden than that of the referee. Until 1998, for example, in the UK BT was forbidden from offering entertainment services in competition with cable companies although cable companies after 1991 were able to offer telephony services. The idea was to encourage new entry by restricting the ability of BT to respond in a predatory manner. Although justified in terms of encouraging competition, the policy was not based upon the Coasian objective of reducing the impediments to private agreements and allowing the market to operate. Instead, it was a Hobbesian policy of trying to bring about particular results.

Methods of Regulation

Principal and Agent[36]

In the absence of contestability, competition cannot finally solve the natural monopoly problem and the regulation of the core assets remains to be considered. We have already noted the objections to the public interest theory of regulation, which abstracts from information and incentive problems and assumes a well-informed and benevolent regulator able to enforce optimal policies on the regulated firm. An alternative approach sees regulation in terms of a principal-agent relationship and focuses on the problem of moral hazard. Here information asymmetry (the fact that information is not equally available to all interested parties) is assumed to constrain the activities of the regulator. Information about the behaviour of the regulated firm, for example, might not be available or only available at very high cost. The regulator (principal) is still assumed to be pursuing public interest objectives but now does not know the firm's costs, which depend upon its effort as well as random factors. A regulatory 'contract' must be drawn up which elicits compliance with the objectives of the regulator.

Where firms (agents) are risk neutral the classic theoretical answer to the 'unobservability' of an agent's effort is for the agent to receive the entire additional benefit from extra effort. If, for example, a farmer received the value of the harvest he would, unlike an employed labourer, have no reason to

---

[36] See appendix for further discussion of Principal-Agent theory.

shirk. He would, of course, bear the risk associated with vagaries in the weather and other chance factors, but if he were risk neutral or if there existed well developed insurance markets in the relevant contingent claims, a contract assigning the 'outcome' or harvest to the farmer would be optimal. The principal (in this case the farmer's landlord) would receive a fee or rent that would be independent of the outcome. In certain circumstances therefore, the unobservability of effort does not matter. A contract representing a pure 'franchise' arrangement whereby the agent pays a franchise fee to the principal and then keeps the net result would give 'high powered' effort incentives.

In the case of a regulator (principal) wishing to achieve allocative efficiency and negotiating with a monopoly supplier (agent) the equivalent solution is for the firm to receive the full social value of its output while paying a fixed fee to the government for the right to undertake the business.[37] If the firm wishes to maximise its profit it will operate in a cost efficient way without being closely monitored. It will also set the efficient output because, assuming that such a thing can be contrived, its marginal revenue will equal marginal social benefit and profit maximisation requires that this is set equal to the firm's marginal cost. Essentially this scheme aligns the firm's marginal revenue with the social marginal benefit of output and avoids risk bearing costs because the firm is assumed to be a risk-neutral party. The firm therefore behaves like a perfectly discriminating monopolist and its expected profit maximising choices will be socially efficient.

The difficulties with this scheme are not hard to see. It economises on the need for information about the firm's costs of production and effort, but it requires the regulator to pay the firm sums representing the consumers' surplus generated on its output. Clearly if the firm simply receives revenue from its sales to consumers it will behave like a normal monopolist. If it is to behave optimally its marginal revenue must reflect the full marginal social value of its output and this requires that subsidies should be paid equal to the consumers' surplus generated. Thus the regulator is assumed to know a great deal about the marginal social value schedule for the firm's product and the whole approach is thus still in the tradition of the 'public interest' theory of regulation. The scheme also requires the regulator and the firm to bargain over the fixed fee. The regulator might try to arrange the fee such that the firm eventually achieves a competitive return on its assets but the information to calculate this is unlikely to be available. A competitive award of the franchise to the highest bidder would be one way of proceeding but this also involves transactions costs (discussed further below) and implies that all the bidders have the same understanding of the social marginal benefit of output as the regulator. The size of the fixed fee will determine the firm's eventual profit but not the efficiency of its operations under this scheme. If distributional considerations are important to the regulator and not just efficiency, however, the negotiated size of the fee will be important.[38]

---

[37] See Loeb and Magat (1979) and Baron and Myerson (1982).
[38] For further discussion see Ricketts (1998).

Rate of Return Regulation

Actual regulatory systems have evolved over time and have been influenced only gradually by the theoretical literature on optimal contracting. The original model for the regulation of public utilities was established in the United States and was based upon control of the rate of return on capital. Where information about costs is hidden from the regulator, the disadvantages of this system are substantial from a contracting point of view. The regulatory 'contract' has incentive effects similar to a commercial 'cost-plus' agreement. Unlike the principal-agent model outlined above, which made the firm's return dependent upon its own actions, rate of return regulation offers the firm the opportunity to earn an 'acceptable' return and thus undermines the incentive to operate in a cost efficient way. In particular an 'unacceptable' rate of return can as easily be made 'acceptable' by increasing the use of capital as by reducing prices and profits. It is the cost-plus nature of rate of return regulation or cost of service regulation that creates these incentives detracting from cost economy. The tendency of regulated companies to adopt highly capital-intensive production methods and generally to 'inflate the rate base' is known as the Averch-Johnson effect.[39]

Finding out about costs in a regulated company becomes a central problem for rate of return regulation and draws the regulator into detailed discussions about production methods. Under these circumstances the danger of 'capture' is increased and this has figured prominently in the US literature as mentioned in section 3. More indirect methods of generating information and providing incentives include the use of 'yardstick competition'. Here the regulator compares information on costs and performance between firms operating in different regions. Attempts to mislead the regulator about cost conditions would then require all firms to co-ordinate their responses. In the UK water industry, for example, the existence of sufficient comparators has been an important factor underlying regulatory opposition to mergers between regional water companies.

RPI minus X Regulation

The post-privatisation method of regulation introduced in the UK took the form of a price cap rather than a rate of return cap. The system emerged from an official report by Littlechild (1983) into the regulation of BT. A regulated price is set for a specified service or weighted average basket of services for a given period of time (usually five years). The regulated firm is then permitted to increase this price by the rate of increase in the retail price index (RPI) minus a negotiated factor X to represent anticipated productivity improvements in excess of the national average.[40] Occasionally, as in the

---

[39] See Averch and Johnson (1962). Further explanation of this effect is provided in the appendix. The disincentive to operate in a cost efficient way is mitigated to some degree by the existence of regulatory lags. If it takes some time for regulators to notice and respond to unacceptably high rates of return, the firm can take advantage of this period to profit from more efficient operations.

[40] National average productivity growth is, of course, reflected in the inflation rate and therefore the RPI.

case of gas, the formula would be extended to allow for changes in the prices of specified inputs that represented a large component of the regulated firm's costs and that were outside its control.

As a form of contract the RPI minus X system is similar to a 'fixed price' arrangement in which the buyer is assured of the result while the supplier takes the risk and receives the rewards from additional effort or from the discovery of cost reducing innovations. It is therefore closer in its incentive properties to the principal-agent contract discussed earlier and avoids the inefficient Averch-Johnson effects that are associated with rate of return regulation. It would not be expected to achieve a fully efficient result, however, because, unlike the Loeb and Magat (1979) case, the firm receives the regulated price for its output and not (except by a fluke) a sum equal to the marginal social value.

Littlechild (1983) originally conceived the RPI minus X system as a means of preventing the privatised utilities from exploiting their monopoly positions in the short to medium term, while competition was encouraged in the longer term (as discussed in section 6). As a long-term means of regulating natural monopoly, however, there is a danger that the incentive properties of RPI minus X are weakened. The problems relate to the process of 're-negotiating' the contract at the end of each regulatory period. If the new price cap and the new value of X are related to existing profits and to rates of productivity improvement achieved in the past, incentives are compromised.[41] As the end of the regulatory period approaches, firms will prefer to delay productivity improvements that might adversely affect the coming regulatory bargain. More generally they will use their information advantage over the regulator to try to exaggerate their costs. Changes in the nature of the product or in the contents of the regulated 'basket' of services imply that a central negotiating problem will be to define the 'output' to which the price cap applies.

Beesley and Littlechild (1988) argue that because of the fixed regulatory period, and the fact that each bargain is forward looking rather than merely adjusting to past events, RPI minus X regulation gives greater scope for bargaining than does rate of return regulation. In this bargaining environment capital investment in very long-lived and specific assets becomes a particular problem if price reviews occur too frequently. Firms may worry that, after the investments have been made, future prices will be set at levels that do not yield a sufficient return - the reverse of the 'capture' problem. Political pressure on regulators to keep prices low or to 'claw back' past profits make the firm vulnerable. This can lead to firms financing investment using higher levels of debt rather than equity because regulators cannot ignore obligations to bondholders, at least (as in the UK) where they are required to permit the regulated firms to finance their proper activities. The regulator may promise not to behave in an opportunistic way towards shareholders but the inability of governments (and hence regulators) credibly to commit very far into the future

---

[41] The disadvantages of fixed review periods have been countered in the UK by the use of 'glidepaths' and (in the water sector) a rolling five-year window so that firms always have time to adjust to regulatory decisions.

is widely recognised as a significant problem. Indeed, it is not even clear that a promise not to intervene *within* a regulatory period can be relied upon.[42]

Profit Sharing Regulation

RPI minus X regulation is in principle designed to give high-powered incentives and to encourage profit-seeking behaviour. In practice, however, profits in excess of some limit are subject to heavy popular criticism and this has led to suggestions for incorporating an explicit element of profit sharing into regulatory contracts. A simple tax on the profits of otherwise unregulated utilities would transfer some monopoly profits to the government but would do nothing to change pricing and output decisions. More complex 'sliding scale' profits taxes in which the tax rate falls as the output of the monopolist rises can be devised to induce firms to lower prices or increase output.[43] Alternatively a system of price regulation might be modified to incorporate profit sharing with consumers through automatic price adjustments. Such schemes are in the tradition of 'public interest' regulation in the sense that efficiency and distributional objectives are explicitly being pursued by the designers, but the work of public choice and institutional economists is recognised by the introduction of political and other constraints on the policy process.

Franchising

By seeing regulation as a form of contracting over a specified period of time between an agent of government and a private supplier of monopoly services, the close connection between certain types of regulation and franchising is apparent. It was Demsetz (1968), updating the work of Chadwick (1859), who suggested that 'competition for the field' could substitute for 'competition within the field' and that this would reduce the requirement for detailed rate of return style regulation of utilities. In a Demsetz auction, bidders compete with each other to supply the entire market at a certain price for a given franchise term. The franchise is awarded to the bidder offering the lowest price. From the point of view of incentives the outcome of a Demsetz auction is similar to price cap regulation, but with the cap set by a competitive process rather than by bilateral bargaining. It could also be seen as a way of making the market 'contestable' at specified intervals even if sunk costs prevent the market being contestable on a more continuous basis.

The organisational choice between rate of return regulation and the Demsetz franchise auction involves many of the considerations raised by Coase (1937) in his discussion of the transactional features of firms versus markets. Regulation has firm-like features such as long-term associations, a continuous process of information gathering, monitoring of behaviour, and a susceptibility to managerial moral hazard. Franchise arrangements are characterised by

---

[42] Helm (2001, p.478) comments that government is 'not a very credible fixed-price/fixed period contractor', while Stiglitz (1998) argues that 'limitations on the ability to make commitments' are often responsible for the difficulty of implementing reforms that seem close to Pareto improvements.

[43] See for example, Burns et al (1998).

more arms' length relations, significant transactions costs but also higher-powered effort incentives. An analogy might be between a managed chain of shops (monitored and regulated by the centre) and a franchise chain (with contracts negotiated with the centre but with more local discretion). Franchising is primarily therefore a means of addressing the problem of moral hazard. The marginal profits from extra effort go to the franchisee and this provides strong incentives to find cost-reducing methods and to operate existing facilities efficiently. Awarding a franchise competitively by asking for price quotes also has the advantage of revealing to the regulatory authorities information about cost conditions in the industry.

Franchise auctions are, however, themselves subject to hazards and costs. Where, for example, the nature of the service to be provided is complex, subject to technical change and difficult to measure, it may not be possible to specify in a contract precisely what is required of the franchisee. Bidders may be asked to specify themselves the nature and standard of the service that they can supply along with the prices they would charge were they to win the contest. In these conditions the possibility of opportunistic behaviour is always present. If the true quality of bidders is hidden from the franchisor ex ante the danger of adverse selection – the use of low quality suppliers because higher quality ones have been driven from the market – is always present.[44] A further possibility is that bidders might shade prices below cost in order to win the contract but plan to re-negotiate more favourable terms later. This is more likely to occur when non-transferable capital is accumulated so that 'first-mover advantages' accrue to the incumbent. Forcing a franchisee into bankruptcy and changing to a new one might be more costly than simply re-negotiating terms. The incumbent challenges the franchisor to 'waste' the firm specific capital already in existence.

In general, long term franchise agreements involving the use of large amounts of durable, specific and possibly non-transferable capital in unpredictable market conditions are costly to negotiate and enforce. Agreements will often require re-negotiation as circumstances change and unforeseen contingencies arise. As a result, the franchising solution to natural monopoly can begin to look very similar to rate of return regulation as time advances. As Williamson (1985, p.350) noted, franchise bidding can lead to an administrative apparatus little different from the one associated with the regulation for which it was an intended substitute.

Concluding Comments

A major theme running through this paper has been the contrast between regulation as a means of enforcing rules and regulation as a means of attaining collectively determined ends. There is increasing recognition that institutional as distinct from purely technological factors play an important role

---

[44] Laffont and Tirole (1986) derive optimal contracts under conditions of hidden information concerning the efficiency of firms but where ex post costs are observable. Their results suggest a contract in which the supplier announces an expected cost and is reimbursed for a proportion of cost overruns. By offering a 'menu' of such contracts, firms will self-select such that the most efficient opt for the 'fixed-price' end of the spectrum.

in economic development.[45] In particular the rule of law, the existence of an uncorrupt and competent bureaucracy, an independent judiciary, along with political stability and accountability are all significant elements. These are closely related to the first (rule enforcing) concept of regulation. Where regulation turns into state planning, however, economic development is adversely affected. Jalilian et al (2003), for example, use a measure of regulatory quality which includes 'quantitative regulations, price controls and other interventions in the economy' (p.11). Higher quality regulatory governance is associated with a lower incidence of these types of direct intervention. In econometric tests, higher quality regulatory governance was shown positively to affect the growth performance of an economy.

A second major theme has been the political and interest group pressure leading to an expanding regulatory state and to increasingly detailed intervention. In the field of the public utilities emphasis on the promotion of competition has helped to establish regulators as rule enforcers or 'game wardens' rather than planners, but even here it is often difficult to distinguish disinterested adjudication from the fixing of a result.

A final theme has been the development of regulation as a form of contracting. This is an extension of the old public interest approach that allows for imperfect information and the resulting transactional hazards. In the absence of public ownership, intervention to provide public goods or to control natural monopoly requires the implementation of some form of contracting. The extension of state 'contracting' to an ever-wider range of activities, however, is simply to impose collective judgements about resource allocation by means of contract with private suppliers instead of by state ownership and administration. The negotiator of contracts on behalf of the state and the regulator of decentralised market processes are undertaking conceptually quite distinct activities.

<u>Appendix</u>

<u>The First Fundamental Theorem of Welfare Economics</u>

This theorem states that in the absence of externalities, or equivalently with a complete set of markets, a perfectly competitive equilibrium is Pareto efficient. In a perfectly competitive market, all contractors face the same set of prices for goods, services and factors of production and all prices are taken as parametric - that is they are not influenced by the behaviour of any individual participant in the market. An equilibrium set of prices is one at which utility and profit maximising behaviour on the part of all contractors results in universal market clearing. There is no excess demand or supply at these equilibrium prices in any market. A Pareto efficient allocation of resources is one in which it is impossible to make one person better off without making

---

[45] See North (1990,1991)

another person worse off. This is equivalent to saying that there are no mutually advantageous exchanges, agreements or 'deals' available.

An intuitive discussion of the theorem is provided in the text. Gains to trade derive from divergences between consumers in relative valuations of goods and services. If, for a given allocation of resources, person A values (i.e. is prepared to pay for) additional units of a particular good more than person B values the same good, there will be mutual gains available from transferring some of the good from B to A at a price somewhere between their two valuations. At such a price, Person B will receive more than his valuation of the good he is supplying while person A will pay less than her valuation of the good she is demanding. Both will be better off as a result of the trade. The efficiency properties of a perfectly competitive equilibrium ultimately derive from the fact that individual maximising behaviour leads people to equate their marginal willingness to pay for each good to its market price. Since all contractors face the same set of prices, the competitive equilibrium must imply that all contractors have the same marginal willingness to pay and that no scope for Pareto improvements based upon diverging marginal valuations can exist.

The Second Fundamental Theorem of Welfare Economics

Many different allocations of resources will have the property of being Pareto efficient. In particular, highly unequal allocations as well as highly egalitarian ones may be in the Pareto set of socially efficient allocations. The Second Theorem states that each and every one of these allocations can be supported as a competitive equilibrium, given certain technical assumptions about preferences. This theorem is a central pillar of the standard 'first-best' approach to public policy. Ideally, lump sum taxes and transfers that do not interfere with the efficiency properties of competitive markets should be used to pursue redistributional aims. Equity and efficiency can be separated as policy objectives with suitable tools aimed at each target. Measures that distort prices and interfere with the operation of the competitive market should not, according to this view, be introduced in the interests of achieving redistributional objectives. Whatever the desired distributional objectives, their achievement will not require the destruction of the competitive market providing that suitable non-distortionary policy instruments can be found.

Natural Monopoly and Sub-Additivity

A cost function is sub-additive if $C(Q) < \sum_i C(q_i)$ for all q_i such that $\sum_i q_i = Q$. This simply expresses the essential attribute of 'natural' monopoly that it is less costly to produce output Q using a single firm than by allocating the same output between a set of two or more firms. A production process that is subject to pervasive economies of scale will give rise to a sub-additive cost function. In these circumstances it is possible to double output by using less than double the inputs and there is always a potential cost advantage to concentrating production in a single unit. Even after economies of scale have

been exhausted and average costs begin to rise there will still be a range of output over which a single firm will be a lower cost producer than two or more separate firms. Note, however, that this theory is based upon the assumption of cost-minimisation and pays no attention to matters such as 'managerial slack', rent seeking, or 'X' inefficiency[46] that might be associated with monopoly.

<u>The Theory of Principal and Agent</u>

Contracting would not be a problem if information were costlessly available to all parties. In fact, however, much information is costly to collect and may be distributed unequally (or asymmetrically) across the contractors. A particular difficulty arises if the actions or 'effort' of one party to a contract (the agent) is unobservable to the other party (the principal).

Clearly it is not always necessary for effort to be observable. If, for example, both parties know the results of the agent's effort, and if the connection between effort and result is a simple causal one, a contract specifying the required results or 'outcome' will be efficient. Either the agent successfully achieves the specified results or he does not. If he achieves the results he gets paid an agreed fee. If he does not achieve the results, he does not get paid. If I agree to pay a gardener for mowing the lawn, it is not necessary for me to observe his actions providing the result is verifiable at low cost.

Greater difficulties arise when the simple causal connection between effort and outcome is broken. Outcomes might to some extent reflect chance factors outside the control of the agent. In this case it would be necessary to draw up a 'state-contingent contract' in which the outcomes promised by the agent were dependent upon the occurrence or non-occurrence of certain chance factors. In this way the risks and rewards could be suitably distributed between the principal and agent. The gardener promises to mow the lawn if it is fine but is otherwise free to contract elsewhere. I promise to pay a fee, even if it is wet, for any costs incurred by the gardener in committing to mow my lawn. Clearly these arrangements require that the 'state of the world' (wet or fine) is verifiable at low cost. I will be unhappy about paying a fee to an absent gardener if (in my opinion) it is fine.

The classic analysis of principal and agent concerns the case where both the state of the world and the effort of the agent are unobservable – or at least unverifiable at low cost by a court of law. In this case the contract can only mention the actual results, which will be the outcome of both the agent's effort and pure chance. Consider the case of a risk-neutral principal and a risk-averse agent. A contract that specifies a given payment or fee to the agent irrespective of the result will mean that the principal takes the entire risk. From a risk-sharing point of view this is efficient. The principal is risk neutral and is therefore the efficient risk bearer. The problem is that the agent has no incentive to exert effort. He gets paid the same whatever happens. This is the traditional 'wage' or 'time rate' arrangement. At the other extreme we might

---

[46] The term 'X-efficiency' was coined by Leibenstein (1978).

envisage a contract in which the principal receives a given payment independent of the actual result. This is the pure franchise contract by which the agent pays a fixed fee to the principal and keeps whatever remains of the actual result. From a risk sharing point of view this is inefficient. The agent takes the entire risk although, by assumption, he is the risk averse party. On the other hand, the contract induces a high level of effort because any additional output resulting from higher effort accrues to the person exerting it (the agent).

An optimal contract will share the result between the principal and the agent and will represent a compromise between the two extreme positions described in the previous paragraph. Starting from the position where the agent receives a 'wage' there will be advantages to 'higher powered effort incentives'. These are provided by tying the agent's reward more closely to the actual results. The 'wage' element is reduced and a 'share' in the result is introduced. Risk sharing losses are incurred but, at first, greater effort incentives outweigh them. At a certain point, however, the gains from moving to yet higher-powered effort incentives will just be counter-balanced by the losses incurred by transferring risk from a risk neutral to a risk averse party. This will be the optimal contract.[47]

If the agent is assumed to be risk neutral, the 'franchise' arrangement turns out to be fully efficient. This is because there are no risk sharing losses incurred by loading risk onto the agent, while effort incentives will also be efficient – the agent will work up to the point at which his or her marginal expected return equals the marginal cost of effort. This is the case discussed in the text.

<u>The Averch-Johnson (A-J) Effect</u>

The A-J effect is an increase in the capital intensity of production methods that is predicted to result from the introduction of rate of return regulation.

In figure A1 an unregulated monopolist producing output q_m and facing competitive factor markets for labour and capital will use L_m units of labour and K_m units of capital. This is the cost-minimising technology to adopt. Total costs are equal to C_m. If the firm were to use a different technology it would incur higher costs and hence its profit would fall. If profit is measured on the vertical axis below the origin, a curve can be traced showing profit for each technology adopted by the firm as indicated by the associated input of capital. The curve is drawn on the assumption that output remains constant at q_m. Thus, the unregulated firm will achieve maximum profit of π_m.

The effect of rate of return regulation can be seen by noting that it introduces a regulatory constraint linking profit to the quantity of capital used. In the figure, the area $\pi \leq RK$ represents this constraint, which will presumably be binding as an equality if the firm maximises its profit. R represents the maximum acceptable rate of return on capital expressed as a 'rental rate' per

---

[47] For a fuller explanation see Ricketts (1986).

unit. Under this regulatory constraint the firm will operate with K_R units of capital and L_R units of labour. Its costs will rise to C_R and its profits will fall to π_R. Notice, however, that maintaining the old factor proportions would have caused profits to decline much further to π^*.

Rate of return regulation is thus expected to result in cost inefficiency. Regulated firms will no longer have an incentive to use cost minimising production methods even when they are assumed to have the objective of profit maximisation.

Figure A1. The A-J Effect
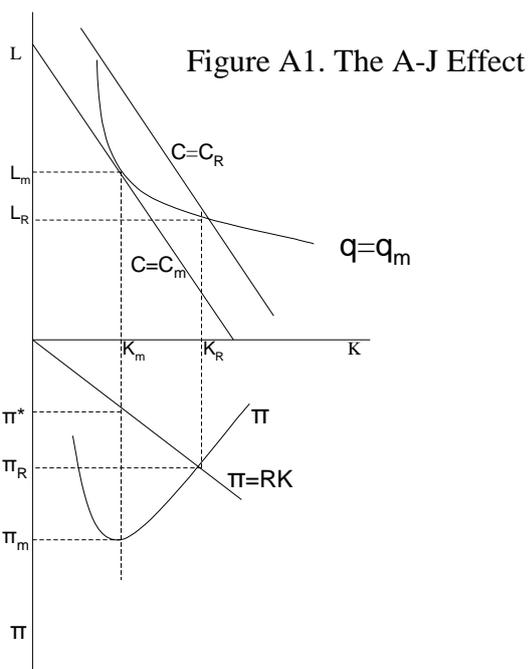
The analysis in figure A1 focuses on the substitution of capital for labour and takes output as constant. Clearly the greater the elasticity of substitution between capital and labour (the more gentle the curvature of the isoquant) the larger will be the effect of rate of return regulation on the technology used and the smaller will be the reduction in the monopolist's profit. With a zero elasticity of substitution (L shaped isoquants) rate of return regulation could not induce the substitution of capital for labour.

References

Averch, H. and Johnson, L.L. (1962) 'Behaviour of the Firm under Regulatory Constraint', *American Economic Review,* Vol.52, pp.1052-69.

Baron, D.P. and Myerson, R.B. (1982) 'Regulating a Monopolist with Unknown Costs', *Econometrica,* Vol.50, pp.911-30.

Bartle, I. (ed.) (2003) *The UK Model of Utility Regulation,* CRI Proceedings 31, University of Bath.

Baumol, W.J. (1965) *Welfare Economics and the Theory of the State,* Bell & Sons, London.

Baumol, W.J. (1967) *Business Behavior, Value and Growth,* Harcourt, Brace and World, New York.

Baumol, W.J., Panzar, J.C. and Willig, R.D. (1982) *Contestable Markets and the Theory of Industry Structure,* Harcourt Brace Javanovich, New York.

Beesley, M.E. and Littlechild, S.C. (1988) 'The Regulation of Privatized Monopolies in the United Kingdom', *RAND Journal of Economics,* Vol.20, pp.454-72.

Blundell, J. and Robinson, C. (1999) *Regulation without the State,* Institute of Economic Affairs, London.

Boycko, M., Shleifer, A. and Vishny, R.W. (1996) 'A Theory of Privatisation', *Economic Journal,* Vol.106, No.435, pp.309-319.

Breton, A. (1974) *The Economic Theory of Representative Government,* Chicago: Aldine.

Buchanan, J. (1975) *The Limits of Liberty,* University of Chicago Press.

Buchanan, J. and Tullock, G. (1962) *The Calculus of Consent*, University of Michigan Press, Ann Arbor.

Buchanan, J., Tollison, R.D. and Tullock, G. (1980) (eds.) *Toward a Theory of the Rent Seeking Society,* Texas A and M University Press, College Station, Texas.

Burns, P., Turvey, R. and Weyman-Jones, T.G. (1998) 'The Behaviour of the Firm under Alternative Regulatory Constraints', *Scottish Journal of Political Economy,* Vol.45, No.2, pp.133-57.

Caves, D.W. and Christensen, L.R. (1980) 'The Relative Efficiency of Public and Private Firms in a Competitive Environment: The Case of the Canadian Railroads', *Journal of Political Economy,* Vol.88, No.5, pp.958-76.

Chadwick, E. (1859) 'Results of Different Principles of Legislation and Administration in Europe: Of Competition for the Field, as Compared with the Competition within the Field of Service', *Journal of the Royal Statistical Society,* Vol.22, p.381.

Coase, R. (1937) 'The Nature of the Firm', *Economica,* Vol.4, No.16, pp.386-405.

Coase, R. (1960) 'The Problem of Social Cost', *Journal of Law and Economics,* Vol.3, No.1, pp.1-44.

Cooter, R. and Ulen, T. (1997) *Law and Economics,* Addison-Wesley, New York.

Davies, D.G. (1971) 'The Efficiency of Public versus Private Firms: The Case of Australia's Two Airlines', *Journal of Law and Economics,* Vol.14, No.1, pp.149-165.

De Alessi, L. (1974) 'An Economic Analysis of Government Ownership and Regulation: Theory and Evidence from the Electric Power Industry', *Public Choice,* Vol.19, No.1, pp.1-42.

Demsetz, H. (1968) 'Why Regulate Utilities?' *Journal Of Law and Economics,* Vol.11, No.1, pp.55-65.

Demsetz, H. (1969) 'Information and Efficiency: Another Viewpoint', *Journal of Law and Economics,* Vol.12, No.1, pp.1-22.

Downs, A. (1957) *An Economic Theory of Democracy,* Harper and Row, New York.

D'Souza, J. and Megginson, W.L. (1999) 'The Financial and Operating Performance of Privatised Firms during the 1990s', *Journal of Finance,* Vol.54, No.4, pp.1397-1438.

Edgeworth, F.Y. (1891) 'On the Determinateness of Economic Equilibrium', reprinted in Edgeworth, F.Y. (1925) *Papers Relating to Political Economy,* Macmillan, London, pp.313-319.

Forsyth, P.J. and Hocking, R.D. (1980) 'Property Rights and Efficiency in a Regulated Environment: The Case of Australian Airlines', *Economic Record,* June, pp.182-5.

Hansmann, H. (1996) *The Ownership of Enterprise,* Harvard University Press, Cambridge Mass.

Hayek, F. (1949) *Individualism and Economic Order,* Routledge and Kegan Paul, London.

Hayek, F. (1978) 'Competition as a Discovery Procedure', in *New Studies in Philosophy, Politics and the History of Ideas,* Routledge and Kegan Paul, London.

Helm, D. (2001) 'Making Britain More Competitive: Regulation and Competition Policy', *Scottish Journal of Political Economy,* Vol.48, No.5, pp.471-487.

Jalilian, H., Kirkpatrick, C. and Parker, D. (2003) 'The Impact of Regulation on Economic Growth in Developing Countries: A Cross-Country Analysis', Discussion Paper.

Kasper, W. and Streit, M.E. (1998) *Institutional Economics: Social Order and Public Policy,* Edward Elgar, Cheltenham UK.

Kolko, G. (1965) *Railroads and Regulation 1877-1916,* Princeton University Press, New Jersey.

Krueger, A. (1974) 'The Political Economy of the Rent Seeking Society', *American Economic Review,* Vol.64, pp.291-303.

Laffont, J-J. and Tirole, J. (1986) 'Using Cost Observation to Regulate Firms', *Journal of Political Economy,* Vol.94, No.3, Part 1, pp.614-641.

Leibenstein, H. (1978) *General X-Efficiency Theory and Economic Development,* Oxford University Press, New York.

Lindsay, C. (1976) 'A Theory of Government Enterprise', *Journal of Political Economy,* Vol.84, No.5, pp.1061-1077.

Littlechild, S. (1983) 'Regulation of British Telecommunications' Profitability', Report to the Secretary of State, Department of Industry, London. Reprinted in Bartle (2003).

Loeb, M. and Magat, W. (1979) 'A Decentralised Method for Utility Regulation', *Journal of Law and Economics,* Vol.22, pp.399-404.

Millward, R. and Parker, D.M. (1983) 'Public and Private Enterprise: Comparative Behaviour and Relative Efficiency' in Millward, R. et. al. (eds.) *Public Sector Economics,* Longman, London, pp.199-274.

National Economic Development Office (1976) *A Study of UK Nationalised Industries,* HMSO, London.

Nee, V. (2002) 'The Role of the State in making a Market Economy', *Journal of Institutional and Theoretical Economics,* Vol.156, No.1, pp.64-88.

Niskanen, W. (1968) 'Non-market decision-making: the peculiar economics of bureaucracy', *American Economic Review,* Vol.58, No.2, Papers and Proceedings, pp.293-305.

Niskanen, W. (1971) *Bureaucracy and Representative Government,* Chicago: Aldine.

North, D. (1990) *Institutions, Institutional Change and Economic Performance,* Cambridge University Press.

North, D. (1991) 'Institutions', *Journal of Economic Perspectives,* Vol.5, No.1, pp.97-112.

Olsen, E.O. (1972) 'An Econometric Analysis of Rent Control', *Journal of Political Economy,* Vol. 80, No.6, pp.1081-1200.

Olson, M. (1965) *The Logic of Collective Action,* Harvard University Press, Cambridge, MA.

Peltzman, S. (1976) 'Toward a more General Theory of Regulation', *Journal of Law and Economics,* Vol.19, No.2, pp.211-240.

Picot, A. and Kaulmann, T. (1989) 'Comparative Performance of Government-Owned and Privately Owned Industrial Corporations – Empirical Results from Six Countries', *Journal of Institutional and Theoretical Economics,* Vol.145, No.2, pp.298-316.

Pigou, A.C. (1938) *The Economics of Welfare,* 4th Edition, Macmillan.

Polanyi, K. (1944), *The Great Transformation: The Political and Economic Origins of Our Time*, Beacon Press, Boston.

Posner, R.A. (1971) 'Taxation by Regulation', *Bell Journal of Economics and Management Science,* Vol.2, No.1, pp.22-50.

Pryke, R. (1981) *The Nationalised Industries: Policies and Performance Since 1968,* Martin Robertson, Oxford.

Ricketts, M. (1986) "The Geometry of Principal and Agent: Yet another use for the Edgeworth Box", *Scottish Journal of Political Economy,* Vol.33, No.3, pp.228-48.

Ricketts, M. (1998) **"**Bargaining with regulators", *In:* Newman, P., ed. *The New Palgrave Dictionary of Economics and the Law,* Macmillan Reference Limited, pp.149-153.

Ricketts, M. (2002) "The Role of the State in Making a Market Economy: Comment on a Paper by Victor Nee" (2000) *Journal of Institutional and Theoretical Economics*, Vol.156, No.1, pp.95-98.

Schumpeter, J.A. (1936) *The Theory of Economic Development,* Harvard University Press, Cambridge, Mass.

Schumpeter, J.A. (1943) *Capitalism, Socialism and Democracy,* Unwin University Books, London.

Shirley and Walsh (2000) 'Public versus Private Ownership: The Current state of the debate', *Policy Research Working Paper No.2420,* The World Bank Group.

Shleifer, A. (1998) 'State versus Private Ownership', *Journal of Economic Perspectives,* Vol.12, No.4, pp.133-50.

Stigler, G.J. (1964) 'Public Regulation of the Securities Market', *Journal of Business of the University of Chicago,* Vol.37, No.2, reprinted in Stigler (1975) pp.78-100.

Stigler, G.J. (1971) 'The Theory of Economic Regulation', *Bell Journal of Economics and Management Science,* Vol.2, No.1, pp.3-21.

Stigler, G.J. (1975) *The Citizen and the State,* University of Chicago Press, Chicago.

Stigler, G.J. and Friedland, C. (1962) 'What can Regulators Regulate? The Case of Electricity', *Journal of Law and Economics,* Vol.5, No.1, pp.1-16.

Stiglitz, J. (1998) 'The Private Uses of Public Interests: Incentives and Institutions', *Journal of Economic Perspectives,* Vol.12, No.2, pp.3-22.

Sugden, R. (1986) *The Economics of Rights, Co-operation and Welfare,* Basil Blackwell, Oxford.

Tullock, G. (1980) 'Rent Seeking as a Negative Sum Game' in Buchanan, J. et al. (eds.) *Toward a Theory of the Rent Seeking Society,* Texas A and M University Press, College Station, Texas.

Walras, L. (1954) *Elements of Pure Economics,* Jaffé, W. (trans.), Allen and Unwin, London.

Weber, M. (1947) *The Theory of Social and Economic Organization,* Free Press, New York.

Williamson, O.E. (1964) *The Economics of Discretionary Behaviour: Managerial Objectives in a Theory of the Firm,* Prentice-Hall, Englewood Cliffs NJ.

Williamson, O.E. (1985) *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting,* Collier Macmillan, London.

Wilson, J.Q. (1980) (ed.) *The Politics of Regulation,* Basic Books, New York.