

**REGULATORY INSTRUMENTS AND THEIR EFFECTS ON
INVESTMENT BEHAVIOUR**

Phil Burns and Christoph Riechmann

Frontier Economics

October 5 2003

DRAFT – not to be quoted without the authors' permission

Keywords: Utility Regulation; Networks; Electricity; Gas; Railways.

JEL-Classification:

Contact Details:
Frontier Economics
150 Holborn
London
EC1N 2NS

Tel: 00 44 20 7611 9494

Fax: 00 44 20 7611 9495

E-mail: philip.burns@frontier-economics.com

REGULATORY INSTRUMENTS AND THEIR EFFECTS ON INVESTMENT BEHAVIOUR

Abstract:

Regulatory instruments have long been understood to have a powerful effect on investment, and part of the motivation for introducing higher-powered regulatory regimes and contracts was to reduce incentives for inefficiency and over-investment (gold plating) inherent in cost-plus regulatory schemes. In practice, the mix of incentives and the institutional framework that make up a higher powered regulatory regime can also lead to unintended distortions on investment behaviour. In this paper we examine the key drivers of investment behaviour, and provide some examples of how these drivers have affected behaviour in practice. We conclude with a set of key areas and inter-relationships that are at the core of a regulatory settlement, and therefore need to be designed appropriately to drive efficient behaviour.

I Introduction

When considering regulated industries, we increasingly focus on the naturally monopolistic network industries that have not, for reasons of their technology and cost structure, been exposed to the pro-liberalisation initiatives that have usually accompanied utility sector reform. Investments are needed to develop and maintain these networks, and the investments themselves have an ongoing value for many years after the investment has been made.

While historically, there have been concerns about over investment, today there are increasing concerns about under-investment. The public debate circles around the level of investment, but it turns out that investment in itself is not an objective for either producers or consumers. Producers are interested in returns that exceed their own cost of capital. They are interested in investments, because investment, among other things, can drive the value of the regulatory asset base on the basis of which their allowed returns are calculated. Consumers are primarily interested in the level of service quality. They have an indirect interest in investment, because it is a key driver of quality – among other drivers. In this regard consumers favour high investments. For them, the observed level of investment can also serve as a leading indicator for future levels of service quality. On the other hand investments add to the cost base which is ultimately to be borne by consumers. To this end consumers disfavour investment.

Neither producers (network operators) nor consumers draw direct value from investment. However, investment is a key driver of the surplus they extract from economic activity. The key question we address in this paper is how to approach the task of balancing these often-diverging interests.

In this paper we examine the key drivers of investment behaviour in regulatory regimes, and provide some examples of how these drivers have affected behaviour in practice. We conclude with a set of key areas and inter-relationships that are at the core of a regulatory settlement, and therefore need to be designed appropriately to drive efficient behaviour.

In doing so, we focus on the specific instruments used by regulators in the setting of price and quality regimes. We do not formally nest these instruments within a general model of regulatory credibility and commitment, since to do so would obscure some fundamental conclusions for little purpose. It is obvious that a bad regulatory instrument incredibly implemented would not lead to the same outcomes as a bad regime that is credibly implemented, but it is by no means clear what the outcomes would be in the first case, as they would depend upon other features of the regulatory arrangements. The construction of the full set of counterfactual cases may be of theoretical interest, but lies outside the principal purpose of this paper, which is to highlight how regulatory instruments can affect investment behaviour. We do, however, draw attention to commitment problems as they arise.

II Investment and changing regulatory paradigms

Incentives to over-invest and poor incentives to reduce efficiency are usually associated with low powered regulatory regimes. Averch and Johnson (1962) showed that strong incentives exist to expand the rate base if the utility was allowed a rate of return in excess of the cost of capital. This point was re-iterated by Baumol and Klevorick (1970) who added that attention should also be focused on the reduced incentives to improve overall cost efficiency under the classic US style of regulation. The massive building programs of the US utilities in the late 1960s and early 1970s tend to indicate that this effect prevailed in practice. In such a world of over investment, high service quality is a (to some extent unintended) side product of low powered regulation.

In order to eliminate these problems either a bureaucratic regulatory procedure is required to provide regulators with the power to strike out some investments in a manner that is not opportunistic (and which therefore would not deter future, efficient, investment); or, if such a mechanism does not exist, regulators are in danger of signing off on investment plans in advance of the investment being made, which inevitably leads to a blurring of accountability in the management of the firm, and ongoing micro-management of the company's activities by the regulator. As such, the

remedies to deal with the over-capitalisation problem under low powered incentive regimes carry significant costs of their own.

Higher powered incentive mechanisms were developed in the 1980s outside of the US, largely to escape from these problems. In practice, higher-powered forms of regulation raise their own specific issues with respect to investment behaviour. The mix of incentives and the institutional framework that make up the overall regulatory package, can lead to a variety of outcomes. The purpose of this paper is to draw out the lessons for incentive design, for both regulators and companies.

III Incentive regulation and output delivery

If the general problem with low-powered incentive regulation is that it encourages gold-plating of the system, low effort, and excessive output delivery, the problem with incentive based regulation is that it can encourage excessive cost reduction, to the detriment of the provision of output. In the UK, the first example of this was the behaviour of British Telecom after privatisation, which was criticised for allowing its public telephone boxes to degenerate into a state of disrepair, since it received no profit incentive to keep them operating normally.

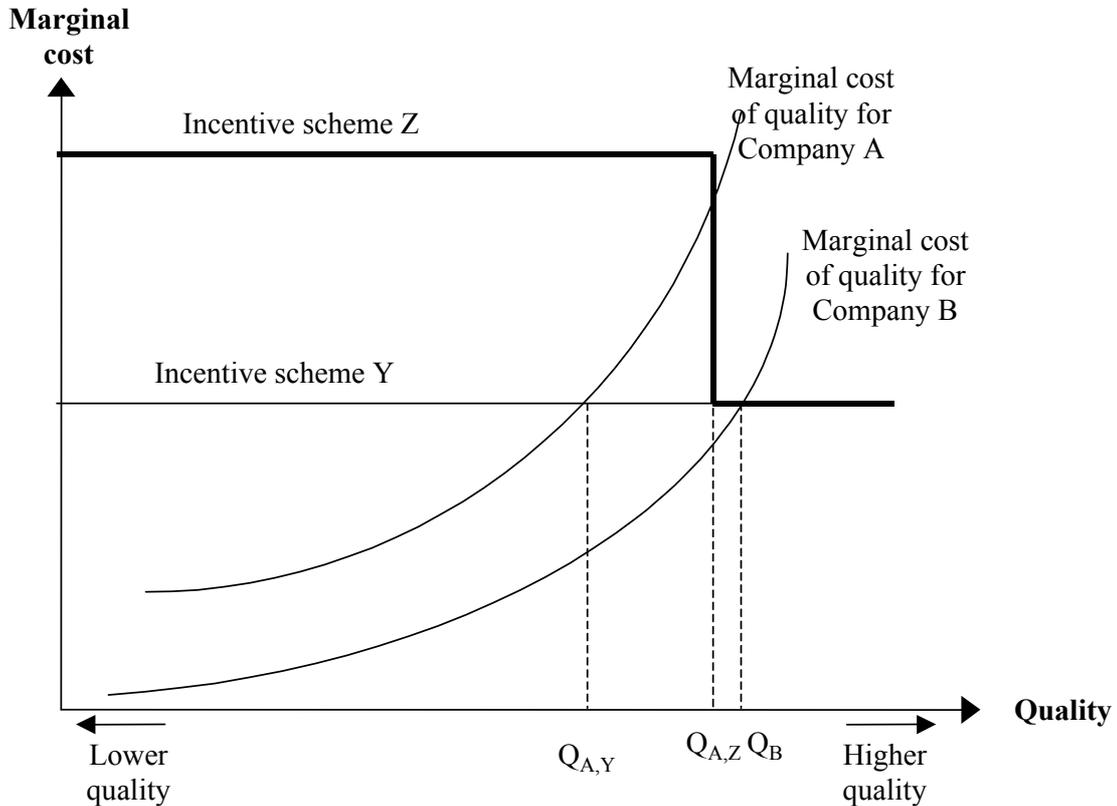
Under incentive based regulation, the provision of service quality is decentralised to the business, within a framework of payments and penalties (which may be defined to a greater or lesser degree of clarity), and the monitoring of that level of service is undertaken by the regulator (to a greater or lesser degree of accuracy). Where the implied payments and penalties are low, or are unclear, it can be expected that the incentives for cost reduction will dominate the incentives for output provision. On the other hand, if the implied payments are very high, then incentives for greater service provision will dominate incentives for cost reduction. In practice, regulators should aim to balance incentives to achieve output provision that is socially acceptable, at efficient levels of cost. It is important to recognise that all regulatory regimes can be characterised in this way – it is not necessary for a formal, explicit service regime to exist for there to be incentives on the part of operators to wish to provide a certain level of service quality.

The problem with the relationship between service quality and investment spend is that there is no doubt that the relationship exists, but that the strength of the relationship and the timing of the relationship is not well understood, either by the firm or the regulator. The concern that quality regulation seeks to address is that investment could be reduced in response to cost incentives, but the effect of this on quality will not be felt for a number of years, which could be beyond the horizons of the present management of the business.

Ideally, therefore, quality regulation should expose current management to the implications of its investment decisions even if these may arise in the future. This could, for example take the form of a regulatory regime based upon both a set of indicators of the current level of service quality provided by the business, and leading indicators of the level of quality that can be expected in the future, given the current state of the assets. The regulatory regime based on this information could be a mixture of price signals (fines and receipts) and discrete regulatory initiatives, such as the removal of the operator's licence.

Figure 1 illustrates a stylised regulatory regime applied to two businesses A and B, each with different marginal costs of quality, which is presumed to be observable on both a contemporaneous and leading indicator basis. A constant marginal payment, Y , embodied either within the price control or within a separate regulatory instrument, leads to the firms choosing the level of quality where the marginal costs and payments are equalised. Incentive regime Z embodies different incentives – there is a very high fine for quality levels below a certain level, which could either be a very high monetary payment, or regulatory exhortations, or the withdrawal of the operator's licence, for example. This regime encourages the companies to converge on some pre-determined, administratively set, level of quality.

Figure 1: Quality incentive schemes



It should be acknowledged that even if a formal quality regime is not adopted, an informal regime applied within the general price control review can be characterised in precisely the same way and would require precisely the same information to implement the regime. A regulator that sets a performance target for a business will have to permit a level of revenue through the price control that enables the business to fund the investment to meet that target, and the degree to which the revenue is conditional, either formally through the price control formula or through other regulatory actions, determines the incentive power. Consequently, the approach described here is general.

In practice, the application of this approach therefore requires a number of key pieces of information:

- Current indicators of quality

- Future indicators of quality
- The performance targets
- The marginal cost of quality
- The explicit or implicit payment rates at different quality levels

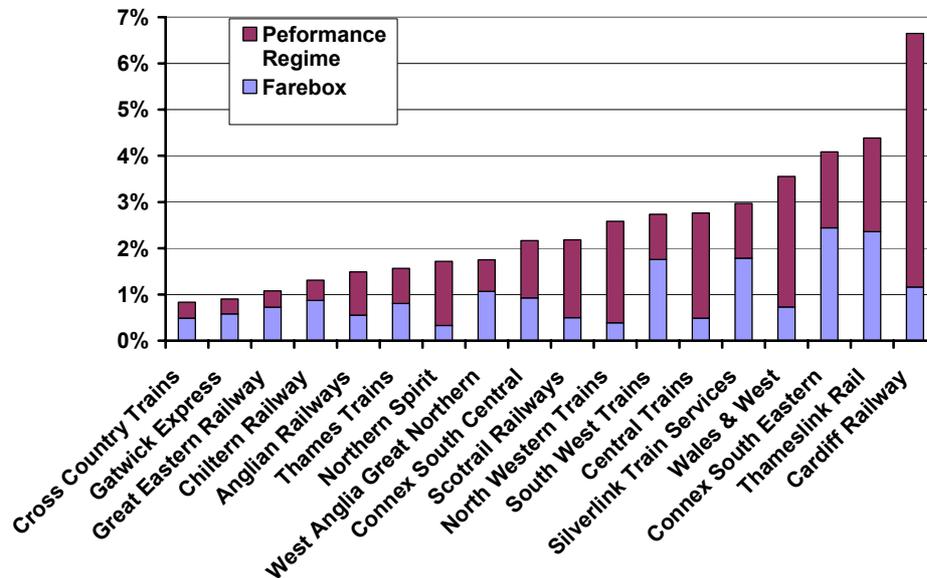
Problems have emerged where either the implicit and explicit payment rates have been low, relative to the marginal cost of quality, and where there are either no future indicators of quality in the regime, or where indicators have existed, they are not appropriate. We discuss each in turn.

Low payment rates

In the rail sector in the UK, both the train operating companies (TOCs) and the network operator (previously Railtrack, now Network Rail) received revenue from a number of sources. TOC revenue is derived from the farebox plus any payments received from the performance regimes. Under the performance regimes, each TOC is assigned a target for reliability and punctuality, with payments made or received if the targets are not met or exceeded. Consequently, the farebox revenue is driven by growth in passenger numbers (which would be partly affected by quality), whilst the revenue from the performance regimes is driven by quality. The only other profit pressure that TOCs must consider in choosing their quality performance is that if quality falls to unacceptably low levels then it is at greater risk of not having its franchise to run the service renewed.

Since 1999, the rail industry has been in a state of turmoil, and it is difficult to unravel the power of different incentive effects in the overall political and economic environment. However, for three years up to 1999, the franchised TOCs operated under a system of regulation that was relatively free of these effects. The figure below indicates the financial benefit to the TOCs if they were to run a perfectly punctual and reliable service, as a proportion of their total income (as of 1998/9), given the parameters of the performance regime and the assumptions used by the SRA in evaluating the effect of performance on farebox.

Figure 2: Maximum overall incentives towards performance as % of total annual income



Source: Frontier Economics calculations based on SRA data

It is clear that the maximum incentive payments available represent a relatively small proportion of TOC total income, especially if the assumed farebox effect of performance fails to materialise. This may indicate that the positive benefits an operator achieves from perfect quality are relatively low performance payments and probably a rather weak farebox effect. In contrast, the penalties for performing moderately badly are a loss of reputation and, ultimately, the loss of the franchise. In other words, it is possible that operators are incentivised to avoid bad performance rather than strive for excellent performance.

The regime applied to Railtrack was somewhat different. It was a regulated network monopoly business, whose incentives to invest could expect to be driven by the performance regime specific to Railtrack (which was similar in character to the one applied to the TOCs) and the regulatory treatment of investment at the price control review. As far as the strength of the incentives in the regime are concerned, the maximum payout to Railtrack in 1998/9 could have been £150m (equivalent to 6% of turnover). By far the biggest source of revenue for Railtrack was, however, the

revenue from track access charges set by the Office of the Rail Regulatory (ORR), and the incentives for investment provided through the price control would swamp those provided through the performance regimes. If the regulatory regime applied to Railtrack was high powered, with a strong emphasis on cost reduction and little monitoring of investment, then the overall regime for quality improvement could be characterised as low powered. If on the other hand, the regulatory arrangements tended to encourage effective investment spend through effective monitoring of investment and asset quality, then the incentives to invest in quality are stronger. As it turned out, at the outset of the regulatory arrangements, the monitoring of Railtrack's investment spend was limited. The licence condition dealing with investment monitoring (section 7) was ineffectual and subsequently had to be tightened considerably. Consequently, for most of the early period after privatisation, the incentives to improve quality were extremely weak compared to the incentives to cut costs and increase profitability.

The consequences of the weak regulatory regime to encourage reliability and punctuality are seen in the table below, which indicate that quality fell between 1996 and 1999.

Table 1: Average delay minutes per period			
	1996/97	1997/98	1998/99
	Minutes	Minutes	Minutes
Railtrack	560,384	551,836	609,044
Operator	306,261	369,852	447,787
Total	866,645	921,688	1,056,831

Source: OPRAF

No future indicators of quality in the regime, or inappropriate indicators

In the rail example given above, an additional problem with the incentive package was the absence of leading indicators of quality against which the businesses could also be judged. If indicators of the health of the network in particular had been

formalised within the regulatory regime, then this would have provided an additional impetus to operate the network reliably and safely.

Where leading indicators of quality do exist (and they tend to be informal in nature), they tend to be based upon the investment that has been undertaken. This can create inadequate incentives for capital cost efficiency because regulators are attempting simultaneously to regulate costs and quality by setting a single variable: the firm's investment programme, which is seen as both a contemporaneous and leading indicator of quality. In these cases, there is evidence that network businesses are becoming more capital-intensive (and less flexible in their response to regulatory incentives) as a result. If quality can be subjected to a separate output-based regime, even a crude one, this dual treatment of capital can be avoided.

A type of future indicator scheme that relies upon information regarding the quality of the assets (as opposed to the level of investment, *per se*) does apply in the water sector in the UK. The companies must submit to the regulator satisfactory asset management plans, on which it is possible for the regulator to judge the ongoing health of the network assets. No explicit system of fines and rewards are based upon these plans, but the regulator has powers to impose an enforcement order on a company - allowing a special administrator to be appointed to run the company - if it is failing to meet any of the conditions set-out in its licence, which include the obligation to produce satisfactory periodic underground asset management plans.

Quality regulation within a comparative regulatory model

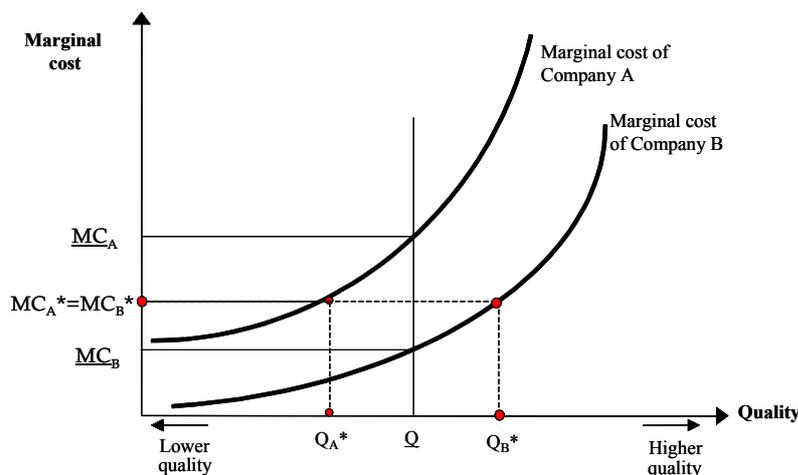
Before leaving the general question of quality regulation within an incentive based regime, it is useful to consider briefly how quality regulation sits within a comparative system of regulation. Many regulators in the UK and in Europe use comparative information to inform a variety of regulatory decisions. However, quite clearly, any yardstick regime applied to cost, must sit consistently with quality regulation.

First, it is well understood that firms differ by their exogenous characteristics. This has led regulators to recognise differences in price levels, whilst at the same time

requiring convergence on some efficiency standard. The implication of this observation is that at one extreme price levels could converge so long as quality performance were allowed to differ, whilst at the other extreme, regulators could focus on uniform quality standards across an industry, but must be prepared to let prices differ to fund the differential costs associated with quality standardisation.

These choices are illustrated in Figure 2, which represents the cost functions of network operators in a favourable (Company B) and in a less favourable (Company A) environment. If the regulator essentially provides each firm with the same marginal payment for quality, on the grounds that the marginal valuation of quality is broadly similar across the two areas, then quality differences will result, but prices will be converged, all else being equal. If, however, the regulator insists on a uniform quality level Q , then costs and prices will vary significantly. Moreover, it is likely that if customer marginal valuations are similar, then company A will be over-delivering on quality, relative to the customers' willingness to pay, whilst company B, will be under-delivering on quality.

Figure 3: Finding the optimal level of quality



This simple example suggests the following implications for the design of quality incentive schemes:

- Where monetary incentive schemes are applied these should be differentiated to reflect local operating environments
- Unless there is compelling evidence that the marginal willingness to pay differs across companies, the emphasis should be on cost and price convergence rather than quality convergence, but within upper and lower quality bounds beyond which the payment scheme is ineffective. Otherwise there is a risk of a deterioration of quality, where the results could be either excessively high quality or unacceptably low quality
- A comparative scheme applied to both cost and quality is likely to be very complex to administer and would most likely lead to multiple equilibria, and increase uncertainty facing the operators.

Summary

Complex incentives for quality improvements are likely to result in perverse outcomes, because of the considerable uncertainties regarding the costs and benefits of improved quality.

If the test of a quality regime is not whether it produces the “right” quality, however defined, but whether it allows high-powered incentives to reduce total costs without endangering quality provision, then such a regime can be simple yet effective.

Regulatory consistency and commitment are important. Regulated firms could take an appropriately long-term view of capital expenditure if they had the same level of confidence in the regulatory penalties or rewards for quality performance far into the future.

Regulators face institutional constraints on their ability to make binding commitments but more consistency and informal commitment to both a price-setting and quality performance regime over the longer term could help to remove concerns

about short term, opportunistic, failure to deliver investment programmes. This would enable a more consistent approach to be taken between promoting operating and capital efficiency.

IV Incentive regulation and choice of inputs

Incentive-based regulation will typically involve the re-setting of price controls at some point. At that point in time, the regulator will usually seek to transfer some of the efficiency gains made by the company to customers. The regulator is also likely to re-set prices on the basis of some efficiency analysis of operating and investment costs that have previously been incurred. Both of these regulatory decisions can impart incentives to distort the business' input mix.

Treatment of efficiency savings

Companies are rewarded for making savings compared to their operating expenditure projections. A saving in operating expenditure in any year is calculated as a reduction in the level of recurrent operating expenditure, that is, we assume that once made a saving should be maintained indefinitely. A £1mn saving in operating expenditure has, using a 7% discount rate, a present value of £14.3mn, some proportion of which will be retained by the company.

Under most existing regulatory arrangements, companies are rewarded for making savings against their capital expenditure budgets through their RAB (regulatory asset base). For the years between price controls, prices are set on the basis of an opening value together with a stream of projected capital expenditure figures. If a regulated firm is able to save £1mn compared to its projected investment in a given year, then it earns the rate of return plus a depreciation payment on the £1mn saved for the years until the next price review. At each review, the RAB and depreciation payments are recalculated on the basis of actual investments, and the benefit of any capital expenditure saving passed through to customers. On this basis it is straightforward to calculate the proportion of operating and capital expenditure savings retained by the company under different regulatory retention periods. In the table below, we show illustrative calculations for a recurring operating expenditure

saving, a one off capital expenditure saving and a recurring capital expenditure saving.

If both operating and capital expenditure savings are recurring then it is clear that an operating expenditure saving is more rewarding for the companies than a similar reduction in capital expenditure. However, if most capital expenditure reductions are one off, rather than recurring, while operating expenditure savings are recurring, then this conclusion no longer holds and instead we would conclude that the strength of the incentive to make efficiency savings is comparable for both operating and capital expenditure.

The nature of capital expenditure savings, when compared to operating expenditure savings that we assume to be recurring, is therefore central to determining whether companies need to be given stronger incentives to make capital expenditure reductions. However, it is clear that a one off reduction in operating expenditure is worth more to a regulated company than a one off reduction in capital expenditure. It is this difference in the strength of incentives that makes the capitalisation of operating expenditure a profitable strategy.

Table 2: Shares of a £1mn efficiency saving in operating and capital expenditure retained by the company under different retention periods¹

Retention period (years)	Operating expenditure (one off)	Operating expenditure (recurring)	Capital expenditure (one off)	Capital expenditure (recurring)
2	100%	13%	17%	2%
3	100%	18%	24%	4%
4	100%	24%	31%	7%
5	100%	29%	38%	11%
6	100%	33%	43%	14%
7	100%	38%	49%	18%
8	100%	42%	53%	22%
9	100%	46%	58%	26%
10	100%	49%	62%	30%
11	100%	52%	66%	34%

Notes: Using a rate of return/discount rate of 7%. For operating expenditure, calculations are based on a recurring saving of £1mn per year. For capital expenditure, calculations are based on a one of saving of £1mn in a given year, for an asset with an assumed lifetime of 40 years. We also assume that prices are reset to actual costs at the end of the retention period. If the assumed asset life were 20 years then the figure for a one off capital expenditure saving would be 47% and the figure for a recurring saving would be 13%, assuming a 5 year retention period.

The role of benchmarking on efficient input use

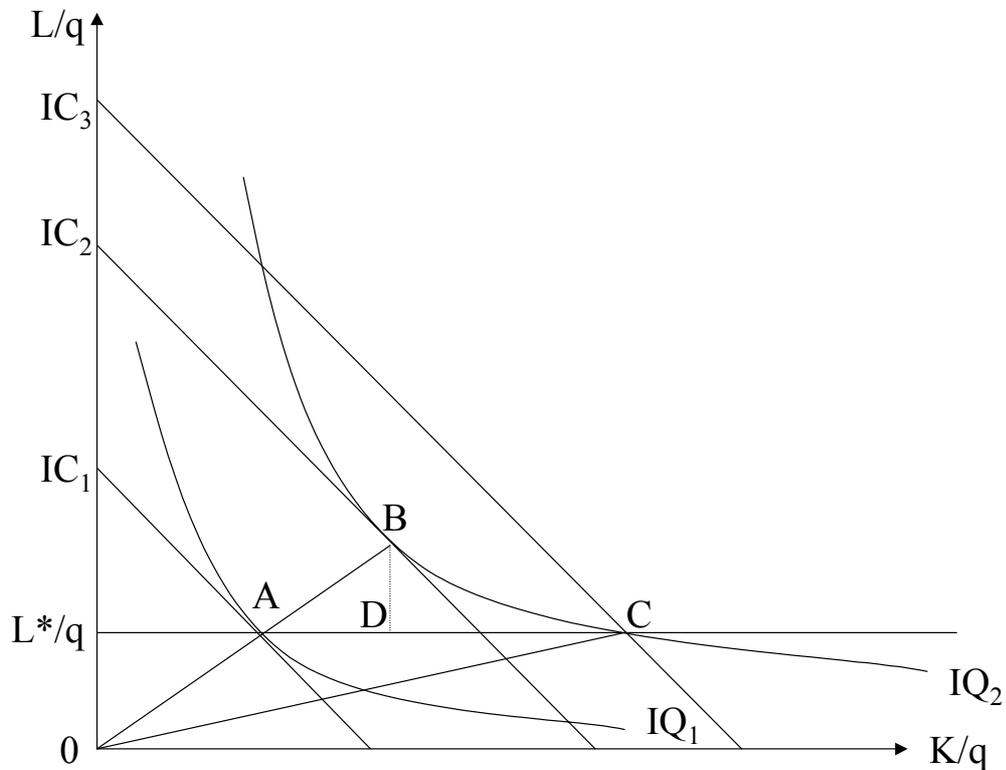
This capitalisation problem is exacerbated if the benchmarking of operating and capital expenditure is inconsistent. Burns and Davies (1999) highlighted a problem in the regulation of the electricity distribution businesses in the UK. If revenue is allowed on a yardstick basis (unrelated to the firm's own costs) for one expenditure category but set on the basis of the firm's own costs for another, RECs have an incentive to transfer costs from the "yardstick" category to the "firm specific" category.

¹ The retention period does not need to be equal to the length of the regulatory review cycle, as incentive payments can be structured such that some benefit continues to be retained by the company beyond the regulatory period in which savings are made

This is illustrated in Figure 4, where we show two firms, 1 and 2, who both produce the same amount of output, q . Firm 1 is more technically efficient because it is producing the same output as 2 but is using less of both capital and labour to do so. Thus we can say that both firms are allocatively efficient, but only firm 1 is technically efficient and firm 2 is technically inefficient – its efficiency is given by the ratio $0A/0B$. Now consider a stylised version of Offer’s regulatory practice in the 1993 price review:

- the regulator decides to set labour (or operating) costs per unit on the basis of a yardstick, L^*/q ;
- it signals that he will recognise investment in determining the regulatory asset base; and
- it monitors both firms to ensure that they have spent up to their regulatory asset base.

Figure 4: Capitalisation in response to regulatory signals



Firm 1's incentives do not change. Firm 2 could choose to improve its efficiency by moving to point A, or could simply substitute capital for labour by moving to point C, claiming to meet its operating expenditure target and fulfilling its capital expenditure commitments. An ill-informed regulator may be inclined to believe this, particularly, if the outputs appear to have been maintained. Alternatively, if there is no slack in the regulatory constraint, then firm 2 may be able to reduce genuine investment in order to capitalise its operating expenditure. This would affect the outputs produced, but since the main output to suffer would be quality, and the impact would be postponed for some time, then the regulator may not be able to detect a diminution in quality, without the benefit of leading indicators of quality (which at that time Offer did not have).

There are three important points to notice from this analysis. First, the inefficient firm has the most incentive to engage in this type of activity, since it has most to lose from a reduction in allowable operating costs to L^*/q . The more efficient the firm, the less incentive it has to substitute capital for labour because it is closer to the benchmark level of operating costs.

Second, this strategy pays off even if the allowed return is well below the true cost of capital. This is because the incentive on the business is to protect revenue, not to make a return (although that would be a bonus). The Averch-Johnson effect exists only when the return is greater than the cost of capital. In this case, the inefficient firm has an incentive to gold-plate even if the return is lower than the cost of capital.

Third, if we informally extend this analysis over time, it is obvious that the regulator will create an arbitrary system of rewards and penalties amongst the businesses if does not act against this behaviour. Put simply, if an inefficient firm obtains the same rewards as a more efficient firm, then over time the efficient firms will be less inclined to expend effort to reduce costs, because they can obtain the same rewards by adopting the inefficient businesses' strategy of capitalisation.

Finally, thus far, we have concentrated on the case where firm 2 behaves inefficiently by engaging in excessive investment. Our other definition of capitalisation is simply where the firm reports a higher level of capital intensity

without actually making inefficient decisions. In this case, firm 2 induces no extra inefficiency but simply protects its revenue by hiding its inefficiency through its reporting policy.

In their paper, Burns and Davies provided a number of indicators of capitalisation, and at the 2000 distribution price control review, Offer made adjustments to the cost base of the businesses to reflect the capitalisation that had occurred, and began work to develop better investment and quality incentives – the Information and Incentives Project (IIP).

V Incentive regulation and joint products

The problem of distortion of inputs can be extended to consider the problem that output provision can be distorted if two outputs that depend on the same costs are subjected to separate regulatory regimes. Where the incentives are balanced, the operator naturally wishes to minimise the joint costs of providing both outputs. Where, however, only one output is subjected to incentive based regulation, it will seek to minimise the costs associated with that output, even if that means the joint costs are not minimised.

Positive externalities

The experience of NGC in the UK is an interesting case study of the balance of incentives across services. At the privatisation of the electricity industry in the UK in 1990, National Grid (NGC) was created to own and operate the high-voltage transmission system. In any electricity system, there are certain costs that depend on how well the transmission system is constructed and operated – such as the additional cost of “constrained on” generation (plant required to operate because of transmission constraints) and “ancillary services” (a range of services with no common international definition, but essentially relating to paying generators to operate in a manner that assists the operation and security of the system). These costs were recovered as the “transport” component of “Uplift” payments, levied on all generators and electricity suppliers (retailers) in the electricity pool.

Initially, NGC simply passed these costs through and therefore received no financial benefit from any cost reductions. This was in contrast to other activities undertaken by NGC, notably those associated with the maintenance and development of the network recovered through use of system charges, which were subjected to high powered incentive regulation. In response to rapid increases in costs, particularly of constrained-on generation, and recognising that NGC had significant influence over some uplift costs, Offer (the electricity regulator) introduced an incentive scheme, the Transmission Services Scheme (TSS), first coming into effect in 1994/95².

The scheme was based on sliding-scale incentive regulation. First, Offer set a target for transmission service costs. Then, if NGC's costs were below this target in any one year, uplift payments were reduced less than proportionately, allowing NGC to retain a proportion of the saving as additional profits. Similarly, if costs exceeded the target, uplift payments increased less than proportionately, exposing NGC to a reduction in profits. A cap on total incentive payments or penalties was established: cost variations beyond the limits implied by this cap were simply passed through one-for-one. The sliding-scale scheme is illustrated in Figure 5 below.

Initially, the scheme specified that NGC retained 30% of any cost savings (to a maximum of £25m, about 5% of the target cost) and paid 20% of any cost over-run (to a maximum of £15m). Targets were revised annually³.

The effects on transport uplift costs were impressive, as Figure 6 illustrates. The rapid increases in cost over the first four years of the Pool's existence were reversed. In particular, the "operational outturn" cost category (principally constrained-on generation) fell by more than half comparing the full years before (1993/94) and after (1995/96) the introduction of the scheme.

² In addition, incentives were introduced into the payments made *by* NGC to the providers of ancillary services, abandoning the previous approach of simple remuneration of cost.

³ Note that sharing ratios and caps changed as the scheme developed. Later schemes also provided for different sharing rules and profit/loss caps for different components of TSS costs.

Figure 5: Allowed revenues and costs under NGC's TSS

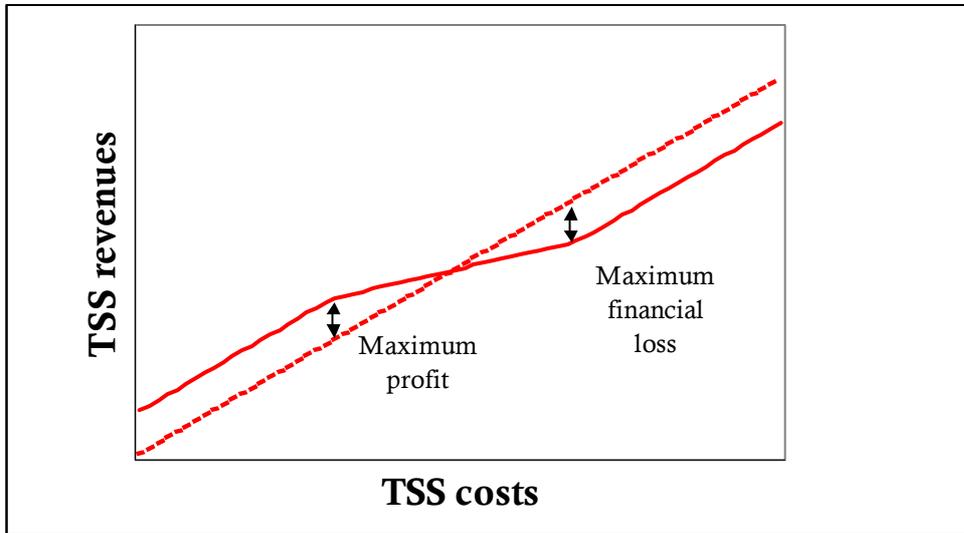
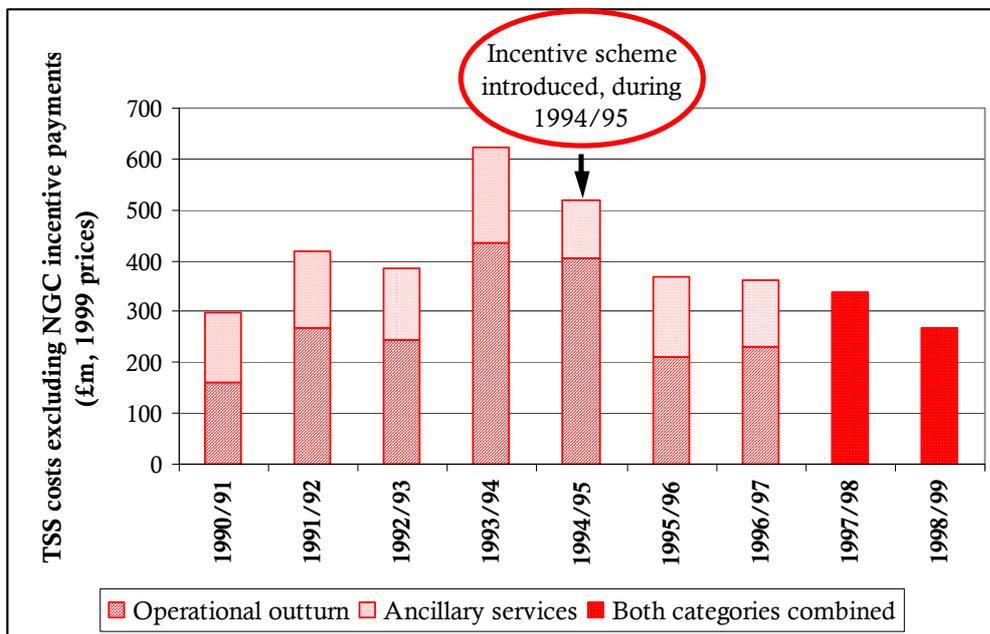


Figure 6: Uplift costs



Source: OFGEM documents such as NGC System Operator Incentives, Transmission Access and Losses Under NETA, consultation document (1999), available at http://www.ofgem.gov.uk/temp/ofgem/cache/cmsattach/1447_nsodec.pdf

Clearly, NGC gained substantially from the scheme. The incentive payment in 1994/95 was £22.5m and in 1995/96 £25m (in nominal prices)⁴. However, the reduction in the two cost categories shown above from 1993/94 to 1995/96 was almost £260m, about 4% of the total cost of wholesale electricity in England & Wales. Thus, although NGC might appear to have been the principal beneficiary, the benefits to customers were substantially greater. Furthermore, by exposing its ability to manage these cost categories, NGC provided Offer with information it could use to set tougher targets in the future. Target costs in 1998/99, for example, were only £215m – less than half the 1994/95 target.

The scheme was subsequently extended to losses. More generally, Offer (now OFGEM) regarded the scheme as a successful demonstration of the power of sliding-scale regulation (see Laffont and Tirole (1993) and Burns, Turvey and Weyman-Jones (1995) for separate analyses of these types of arrangements). For example, OFGEM subsequently introduced sliding scale incentives for the gas pipeline operator, Transco (now part of National Grid Transco) and some seven sliding scale mechanisms apply to different activities.

These regulatory evolutions reflect the more general point that for some network businesses - and electricity transmission is a good example - the operator needs to be incentivised to minimise the combined cost of transmission and generation, which does not necessarily mean minimising the cost of just transmission. For example, in liberalised wholesale electricity markets, interconnectors serve the extremely valuable function of widening a geographic market and consequently weakening the market power of national incumbents. It is important that regulatory arrangements provide the incentive for the network operator to develop interconnector capability, or to increase capacity at certain points on the system if this has the benefit

⁴ It is worth noting that the first payment was below the maximum available while the second was at the maximum. Even in this second year, NGC's performance did not result in costs substantially below the levels at which the maximum payment was triggered. If this had happened, it would be reasonable to conclude that incentives were not the principal driver of the cost reduction. The receipt of two payments close to the maximum suggests (but does not prove) efficient response to incentives on NGC's part.

of reducing upstream costs by relieving transmission constraints between regions and countries.

VI Incentive power and the periodicity effect

As a result of the cyclical nature of regulatory proceedings, companies retain a greater share of the benefit of cost reductions made early in a regulatory period, rather than later, i.e. the year in which a saving is made in a regulatory review alters the period over which any benefit is retained by the company. This is illustrated in Figure 7 and Figure 8, both of which assume a regulated firm faces a review every 5 years and that no steps are taken to address the periodicity problem.

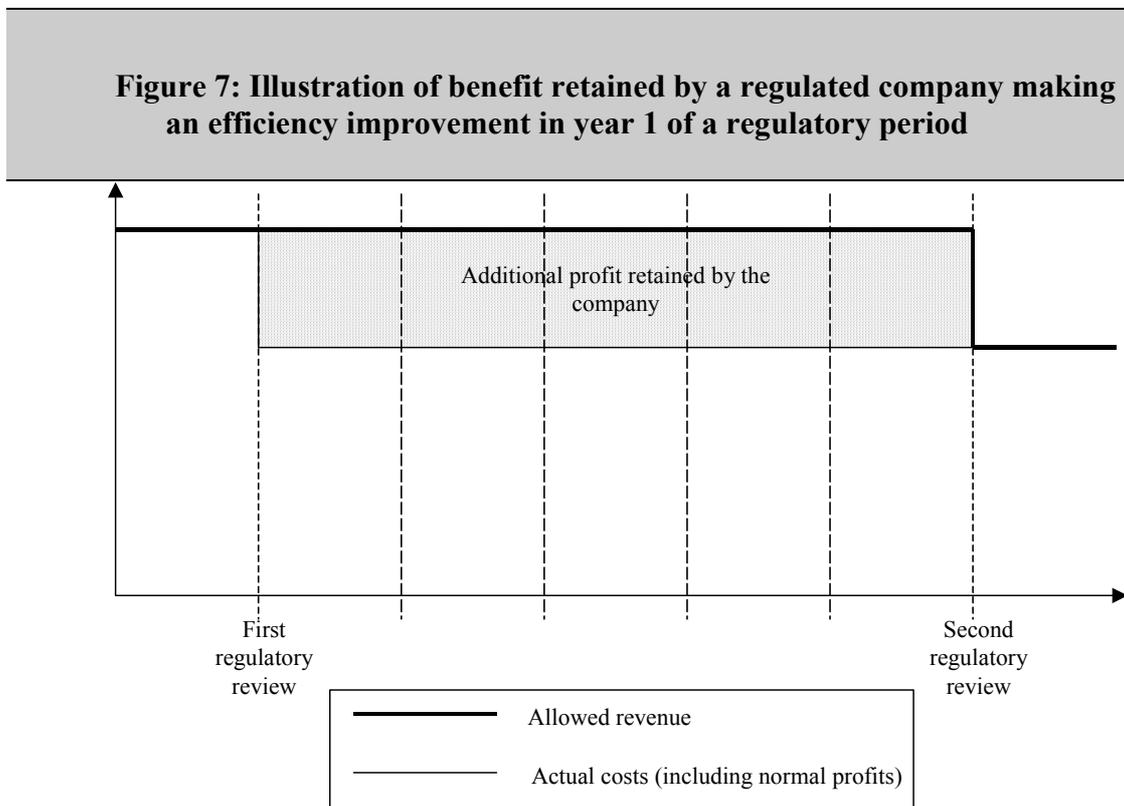
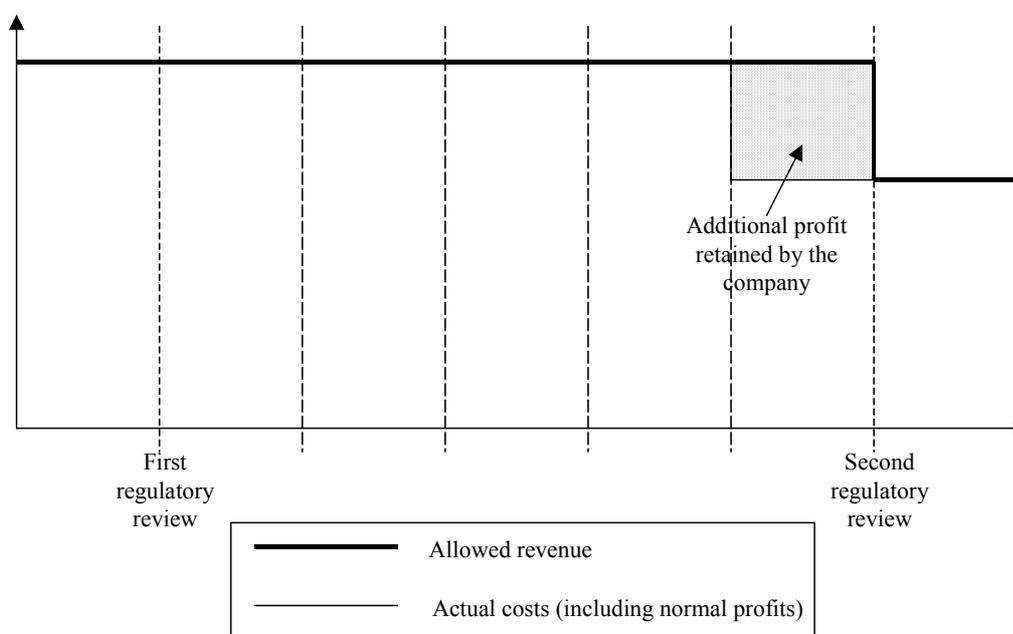


Figure 7 shows the path of costs and revenues for a company that makes an efficiency improvement in the first year of a regulatory period⁵. As a result, the

⁵ In practice, this assumes that the company makes the saving on the first day of the first year.

company retains the benefit of this saving for up to five years, before its prices are adjusted by the regulator at the next review, passing the benefit through to customers.

Figure 8: Illustration of benefit retained by a regulated company making an efficiency improvement in the final year of a regulatory period



In contrast, Figure 8 shows the path of costs and revenues for a company that makes an efficiency improvement in the fifth and final year of a regulatory period⁶. In this case, the company retains the benefit of the efficiency improvement for at most one year, rather than five. This comparison demonstrates that companies have weaker incentives to make efficiency savings as they progress through a regulatory period. This is likely to have two adverse effects.

- The timing of cost reductions will be influenced by the regulatory timetable (where companies are not allowed to retain savings for a fixed period of time), rather than the firm's own assessment of its business. Presumably, this

⁶ In practice, regulatory reviews are often conducted on data pertaining to the penultimate year of a regulatory period as this is typically the timeliest data that is available. This does not alter the principles explained in the discussion above.

will be less efficient than a timetable driven solely consideration of the actual costs and benefits of making an efficiency improvement.

- Variation in incentives encourages “gaming”. Cost savings can be delayed, hidden or transferred between years. If successful, gaming is bad for customers because it results in higher prices than necessary. Even if unsuccessful, it increases the difficulty of carrying out regulatory reviews as it makes the assessment of efficient costs more difficult.

There is some evidence to suggest that regulated companies in a number of sectors have responded to these signals, delaying savings that could be made in the last few years of a regulatory period and making rapid reductions in costs in the years immediately after a review. The table below shows the results of Malmquist productivity analysis undertaken by Frontier Economics. The productivity estimates are based upon a model in which the input is cost (either operating or total cost) and the outputs are kWh, maximum demand, number of customers, and proxies for the environment such as the number of transformers and the length of mains. Quite clearly, there was a boost to productivity building up to the first year of the price control that ran between 1995 and 2000, and by 1997/8, which was the last year on which the regulator had data to inform the 2000-2005 price cap, many companies were reporting reductions in productivity. Most of this reflected the reporting of particular costs, that Ofgem were able to unwind, but more generally, it was clear that the companies were saving further cost reductions until after the review. Evidence on cost performance since the review indicates further significant productivity improvement.

Table 3: Productivity change over time amongst the DNOs

	Operating expenditure productivity	Total cost productivity
<i>91-92</i>	1.81	0.99
<i>92-93</i>	6.14	3.49
<i>93-94</i>	1.36	1.66
<i>94-95</i>	3.67	1.65
<i>95-96*</i>	27.59	13.7
<i>96-97</i>	7.07	2.82
<i>97-98**</i>	-7.72	-4.19
91-98	4.85	2.66
92-88	5.4	2.95

**Year following a price review*

*** Year preceding a new regulatory announcement*

VII Lessons for regulators

In the previous sections of this report, we have outlined how regulatory instruments can affect investment and output delivery in an incentive based regulatory model. In doing so, we hope to have shown that what falls under the general label of incentive based regulation can, depending upon how the parameters of the regime are set, induce too much or too little investment; too much or too little output provision; distort input usage; and distort the timing of investment spending and efficiency gains more generally.

In a sense this is unsurprising. Regulation is a problem of asymmetric information and setting the parameters of the regulatory system is unlikely to result in both an optimal level of cost and service quality. However, through appropriate regulatory design and collection and monitoring of relevant cost and asset data, it should be possible to at least avoid some of the worst imbalances in incentive packages. These imbalances come at a welfare cost and a distributional cost. The welfare cost arises through the inappropriate use of resources. The distributional

affect arises because the essence of incentive based regulation is that companies must be induced to reveal information or elicit effort through the provision of profit incentives. If the wrong incentives are put in place, then payment is made for information of no value, a cost that falls directly upon the consumer. Consequently, the quality of the incentive regime is not an abstract matter, but one of significant economic importance.

Output based regulation should form the basis for the incentive regime

It is important to begin by establishing that output-based regulation is the appropriate basis for the incentive regime. In general, businesses exist to provide services to customers. They do not exist to incur costs, although of course that is a natural consequence of service provision. Regulated businesses are of course no different in this respect to other businesses.

Moreover it is also true that in general if the price of the output increases, the firm is induced to supply more of it, and *vice versa*. For a regulated business, there must be a financial incentive to continue to provide outputs, and this incentive exists through a number of mechanisms – through explicit quality regimes as well as through the price control determination process. Consequently, an output based regime mimics, to some extent, the value for money offering made by firms in competitive markets.

The counterfactual case is one where regulation is input based, that is firms are rewarded for the inputs they use, either directly, or because an input (e.g. investment) is used as a proxy for output. If the value of investment incentives increases with the amount of investment undertaken, the consequence will be high levels of inefficiency. It is questionable whether quality would increase in such a scheme because only inputs, not outputs would be monitored.

The freedom the firm has to choose its level of output can be limited by regulatory instruments

Even though the incentive regime should be output based, this does not necessarily mean that the firm should be given unfettered choice over the level of service provision. This approach was effectively adopted in the rail sector in the UK to disastrous effect. Instead the regulator can choose a mixture of price and non-price incentives to ensure that output provision is locked in to some socially acceptable range. The focus of the incentive regime shifts to the minimisation of costs, subject to that output range being met.

Output based regulation requires a proper definition of outputs

Part of the reason why quality regulation is so under-developed, and the experience has been so mixed, is because the quality indicators have not been available. As a consequence of the problems of using investment as an indicator of quality however, some regulators (e.g. Ofgem in the UK) are beginning to assess whether quality indicators can be developed. The difficulty, as noted earlier is that quality indicators should not just measure current output performance, but also expected future output performance given the quality of the assets in place.

Consequently, regulators need to work harder to establish these indicators of performance. Having done so, different regulatory instruments can be applied to different types of information. For example, it may be possible to apply marginal payment schemes to some easily measurable outputs (e.g. minutes of electric power lost) within a cap and collar scheme, but if the interpretation of an asset report is more subjective, then it should be dealt with under non-price mechanisms. The point is that the regulated business is reminded of its obligations to ensure continued output provision into the longer term. A number of approaches exist to allow the regulator to critically assess the information it has been presented with, including independent audit examinations, shadow network modelling and so forth.

As regulation evolves from its focus on cost minimisation to one of network integrity and output delivery, the proper definition of outputs that can form the basis of incentive regulation is the major task ahead for regulators and companies.

If a comparative incentive regime is used, it should not apply to both costs and service provision, but only to one of these

Yardstick competition has significant efficiency benefits, but the application of a yardstick mechanism to both costs and quality leads to an inherent circularity in the regulatory arrangements, and multiple equilibria. In these situations, regulators should focus on whether they wish to promote cost reduction, subject to quality safeguards that cannot be breached, or quality enhancement on a comparative basis, with costs determined as a residual. In our view, the former approach is likely to be less complex to administer, provides clarity to companies and yet at the same time provides the necessary safeguards on output provision.

Incentive based regulation should avoid distortion of the firm's input choices

Regulators should seek to evaluate and reward operating and capital efficiency on the same basis in order to avoid distorting input use. One way of doing this is to develop a total cost model of the businesses to be benchmarked, and uprate prices on the basis of total costs. The construction of a total cost variable has been described in Frontier Economics (2003)

Regulation should promote balanced incentives across different regulated products

This should ensure that total cost is minimised, not just the cost of producing the output with the highest powered regulation.

Regulation should promote balanced incentives over time

In order to avoid companies distorting the timing of efficiency gains, the regulatory regime should ensure that the NPV of any efficiency gain is the same

irrespective of when in the regulatory cycle it is made. Such an approach has been adopted by Ofwat in the UK.

VIII Conclusions

With the introduction of high powered incentive regimes attention needs to be paid to the development of investment. Historic and low-powered incentive regimes contained an inherent incentive to (over)invest. Incentive regulation that ignores investment or at least the quality of supply could lead to a deterioration of service quality. While this has not yet become an issue in industries where there is initially overcapacity and where there are long asset lives, there have been problems in implementing incentive regulation where the initial starting point was an already weak infrastructure.

This paper has demonstrated that the design of incentive mechanisms to promote efficient investment behaviour is by no means straightforward, and inadequate attention to the problem can result in serious distortions and problems. The key lesson from this paper is that the construction of the package of incentives should be approached with care, and must be accompanied by the development and effective monitoring of meaningful output indicators.

REFERENCES

Averch, H., and L.L. Johnson, 'Behaviour of the firm under regulatory constraint', 1962, *American Economic Review*, 52, pp 1056-69

Baumol, W. and P. Kleverick, 1970, 'Input choices and rate of return regulation: an overview of the discussion'. *Bell Journal of Economics and Management Science*.

Burns, P., and J. Davies, 1998, Regulatory incentives and capital efficiency in UK electricity distribution businesses, *CRI Technical Paper No. 12*.

Burns, P., R. Turvey, and T.G. Weyman-Jones ,1998, Behaviour of the Firm under Alternative Regulatory Constraints, *Scottish Journal of Political Economy*

Frontier Economics, 2003, *Balancing Incentives*. A report for Ofgem

Frontier Economics, 2003, *Dealing with uncertainty*. A report for Ofgem.

Laffont, J.-J. and J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*. MIT Press.

National Audit Office, 2002, *Pipes and Wires*.

Ofgem, *NGC System Operator Incentive Scheme 2003/04 – 2005/06: Initial Consultation Document*, 68/02, 18 October 2002.