Developments in regulatory principles: the UK experience

Simon Cowan[1]

Abstract

Developments in the principles of utility regulation are assessed in the light of UK experience. The incentives-rent extraction model of price regulation, multi-product pricing and access pricing, rules for spot markets, structural regulation and the problem of regulatory commitment are discussed using a unified theoretical framework.

Prepared for the conference on "Regulating network utilities in the EU" held at Nuffield College, Oxford, July 3-4, 2000.

[1] Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UL, UK. e-mail: simon.cowan@economics.ox.ac.uk

1. Introduction

The UK, along with Chile, was the pioneer of privatization of state-owned utilities in the 1980s, and has served as a model for subsequent privatizations and new forms of regulation elsewhere in the world. As experience has been gained and practical lessons have been learned the theory of regulation has also developed. In this paper I assess some recent theoretical developments in regulation in the light of UK experience. The choice of topics is motivated by UK practice, but the purpose is not to claim that there is a tight mapping between observed features of actual regulation and the theory.[2] Instead I present some key ideas using, for the most part, a unified modelling framework. Armstrong *et al.* (1994) and Laffont and Tirole (1993) present more detailed surveys of regulation theory.

For analytical purposes a useful distinction is between the regulation of conduct and the regulation of industry structure – see Kay and Vickers (1988). The paper is divided into three main sections. In Section 2 we examine the regulation of conduct. The focus is on the regulation of the prices of natural monopolies, but the design of rules for competitive spot markets is also addressed. The recent literature on access pricing is covered in some detail. In Section 3 we discuss the role of industry structure in regulation. Section 4 covers the role of commitment in regulation, and Section 5 concludes.

2. Regulation of prices

2.1 Setting price levels for natural monopolies

Price-cap regulation is now the standard method of regulating utilities in the UK, and the model developed here has been copied in many countries. Professor Stephen Littlechild was asked by the Department of Industry to examine possible schemes for the regulation of the profitability of British Telecommunications (BT) when its privatization was planned. In his report (Littlechild, 1983) Professor Littlechild argued that the focus of regulation should be on prices and not profits, and that US-style rate-of-return regulation generated poor incentives for cost efficiency and encouraged regulatory capture. He recommended RPI – X regulation for prices in the markets

[2] For an interesting discussion of the role of economists in policymaking see Faulhaber and Baumol (1988).

where BT faced no competition. An index of prices is allowed to grow by at most the rate of growth of the Retail Prices Index, which is the main measure of consumer price inflation, less a predetermined X factor initially set at 3 percent for BT. Thus real prices had to fall by at least 3 percent each year. The X factor remained fixed and was independent of observed costs and profits during the five-year period before the cap was reviewed. Price caps in practice are characterized by: (i) relatively long periods between formal price reviews; (ii) a commitment by the regulatory agency to avoid resetting the price cap during the lag between reviews in spite of new information being available during that period; (iii) some flexibility to choose relative prices as long as the price index satisfies the RPI $- X$ constraint. The intention is to provide incentives for the firm to cut its costs before the next price review occurs. These incentives exist because cost cutting does not entail immediate price reductions.

The main insight of the theory behind price caps and related incentive mechanisms is that when the firm has more information about its own costs than the regulator it will only act efficiently if it is given incentives to do so. In general, though, the provision of incentives for cost reduction requires that the firm be allowed to earn supernormal profits. If the regulator requires the firm to cut prices whenever it reduces costs then excess profits can be minimized, but costs will be higher than otherwise. The regulator determines how observed costs should be reflected in prices to trade off the objectives of cost minimization and profit (or rent) elimination.

The model that follows is an adapted version of the model of Laffont and Tirole (1986, 1993).[3] The consumer has a unit demand for the product and always purchases. Consumer surplus is $V = U - p$ where U is gross utility and p is the price. The cost function has two elements. The first part is the accounting cost, $c = \theta - e$, where θ is an exogenous cost parameter and e is cost-reducing effort. The regulator observes c but not its individual components. The distribution of θ is common knowledge and the highest possible value of θ is $\overline{\theta}$, while the expected value is $E(\theta)$.[4] The second part of the cost function is the cost of effort, $F(e)$, where $F(0) = 0$, $F' > 0$

[3] An alternative model involves a trade-off between providing incentives to the firm and providing insurance. Armstrong *et al.* (1994, Chapter 2) explore this model, which has similar results to those presented here.

and $F'' > 0$. The regulator maximizes a weighted sum of consumer surplus and profits, $W = U - p + \alpha[p - c - F(e)]$ where α, the weight on profits, satisfies $0 < \alpha \leq 1$. Subtracting and adding total costs gives

$$(1) \qquad W = U - c - F(e) - (1 - \alpha)[p - c - F(e)].$$

Equation (1) shows that the regulator wants to minimize total cost, $c + F(e)$, and, when $\alpha < 1$, to eliminate the rents that the firm earns. With full information the firm is ordered to set the level of e that minimizes $c + F(e)$ and receives a price that just covers total costs, so both objectives are achieved. But with asymmetric information about θ and e there is usually a trade-off.

The regulatory instrument is a function relating the price to observed operating cost: $p(c)$. For simplicity I restrict attention to linear functions, so $p = k + \beta c$, where β is the cost passthrough coefficient. A pure price cap is a special case where $\beta = 0$, while rate-of-return regulation is characterized by $k = 0$ and $\beta = 1$. Two constraints must be satisfied. First the regulator must take account of the way the firm optimizes. Profits are $\Pi = k + (\beta - 1)[\theta - e] - F(e)$, so the firm's choice of e is characterized by the first-order condition $1 - \beta = F'(e^*)$. This determines an effort function $e^*(\beta)$, with $de^*(\beta)/d\beta = -1/F'' < 0$. Greater cost passthrough reduces effort.

The second constraint is that the firm must be willing to participate whatever the value of θ that is realized. This implies that profits should be at least zero when the exogenous cost parameter takes on its highest feasible value, $\bar{\theta}$. Profits in this case are $\Pi(\bar{\theta}) = k + (1 - \beta)[\bar{\theta} - e^*] - F(e^*)$. Since the regulator doesn't want to leave unnecessary rents k is set so that $\Pi(\bar{\theta}) = 0$. Using the implied value of k the general expression for profits is $\Pi(\theta) = (1 - \beta)[\bar{\theta} - \theta]$ and $E(\Pi) = (1 - \beta)[\bar{\theta} - E(\theta)]$, so the regulator has to ensure that $\beta \leq 1$. As β rises the firm puts in less effort and costs rise, but expected rents fall.

[4] Since we are restricting attention to linear incentive schemes the only characteristics of the distribution function of θ that must be known are the highest possible value of θ and the expected value.

We now characterize the solution to the regulator's trade-off between cost efficiency and rent extraction. Expected total costs are $E(\theta) - e^*(\beta) + F(e^*(\beta))$ and the cost to the regulator of the expected rents is $(1 - \alpha)(1 - \beta)[\bar{\theta} - E(\theta)]$. Choosing β to minimize the sum of these two terms, and noting that $1 - \beta = F'(e)$, we find the optimal value of β to be:

(2) $\beta^* = (1-\alpha)[\bar{\theta} - E(\theta)]F''(e^*(\beta))$.

If the value defined in (2) exceeds 1 then the regulator sets $\beta^* = 1$. Three factors determine the optimal value of the passthrough coefficient. First, the closer α is to 1 the less is the regulatory concern about rents and the lower β^* can be. Second, if the range of uncertainty about θ rises, i.e. $\bar{\theta} - E(\theta)$ increases, the regulator wants to set a higher value of β^* to reduce rents. Third, if the marginal cost of effort increases sharply (F' is high) the cost of providing incentives for cost efficiency increases and the regulator raises β^*. Price-cap regulation is a special case that applies when $\alpha = 1$ so the regulator is indifferent to rents and can concentrate on offering full incentives for cost minimization. Rate-of-return regulation is a special case where $\beta^* = 1$, so the firm has no incentive to reduce costs, but rents are minimized. In the more general intermediate case there is some degree of cost passthrough. In the UK some cost passthrough (within the regulatory period) was allowed in the formulae by which the gas, electricity and water industries were regulated. Of course at the time of the formal price review price levels are adjusted to take account of realized and anticipated cost changes, so no price cap scheme is ever "pure" in practice.

What options does a regulator have to improve the trade-off between the promotion of cost efficiency and the minimization of rents? Two standard answers are yardstick competition and franchising. Yardstick competition is feasible when there are several similar regionally separated firms. In the UK the regional water companies and the electricity distribution companies have been regulated using versions of yardstick competition. The report by Stephen Littlechild on the regulation of privatized water companies (Littlechild, 1986) recommended the use of yardstick competition, and the subsequent Water Industry Act 1991 established a formal role for the use of comparative information by the regulator.

The classic statement of the theory of yardstick competition is Shleifer (1985). The essence of the Shleifer model can be shown in our framework. Suppose there are two separate firms, 1 and 2, and that their cost parameters, θ_1 and θ_2, are perfectly correlated. The regulator then immediately improves his information position relative to the case where the cost shocks are independent. If $c_2 > c_1$ the regulator knows that firm 2 has chosen a lower effort level than 1, and 2 can be penalized for this. The regulator should relate the allowed price for one firm to the other firm's accounting cost. For firm 1 the price is $p_1 = F(e^{**}) + c_2$, which gives it the incentive to invest in the first-best effort level, defined by $1 = F'(e^{**})$, since it cannot influence its own price. Effort costs are just covered. Firm 1's profits are $\Pi_1 = F(e^{**}) + \theta_2 - e_2 - \theta_1 + e_1 - F(e_1) = 0$ since $e_1 = e_2 = e^{**}$ and $\theta_1 = \theta_2$. Thus yardstick competition ensures that costs are minimized and rents are eliminated in spite of the information asymmetry. More generally if the cost shocks are not perfectly correlated the regulator can set the price as a linear function of the two accounting cost levels and the trade-off between rent extraction and cost reduction is partially improved.

The second way to improve the trade-off is to use competition for the market, or franchising. Auctioning the right to supply monopoly services has been used for the train operating services in the UK, for the water sector in France and for infrastructure projects in many developing countries. The key point is that in an auction of an incentive contract the regulator is offering a prize with potential rents attached. Firms compete to earn those rents, and competition will dissipate the excess profitability. In a Chadwick-Demsetz auction potential suppliers compete over the price that they are prepared to charge customers (and perhaps also over quality). Suppose for simplicity that each firm knows the values of both its own and the other firm's θ variable and that $\theta_2 > \theta_1$. The regulator fixes a value of β in advance and the optimal choice of effort for the winning firm is e^*. The winning firm is the one that offers the lower fixed part of the price function, k.

Firm 2's lowest possible offer is $k_2 = (1 - \beta)(\theta_2 - e^*) + F(e^*)$. Firm 1 wins the contract by setting k_1 just below this level. Its rent is $\Pi_1 = (1 - \beta)(\theta_2 - \theta_1)$. The equivalent rent in the case where firm 1 was regulated without an auction would be $(1 - \beta)(\bar{\theta} - \theta_1) \geq (1 - \beta)(\theta_2 - \theta_1)$. Auctioning the contract achieves the same level of efficiency but lower rents – the trade-off is improved. An implication is that the greater the number of firms the better the outcome for the regulator, since the likely gap between the top two θ levels will be smaller. Similarly the more correlated are θ_1 and θ_2 the smaller the expected gap between them and thus the smaller the rent for the winner. As the correlation tends towards unity the regulator will reduce β towards zero (which induces productive efficiency) and rents will be eliminated. The identical outcome would occur if neither firm knows its rival's value of θ and the auction is awarded on a second-price sealed bid basis. In practice in the UK the role of franchising has been relatively small, probably because the investment requirements of the firms that have been privatized have been very large, and franchising is easier to apply to the management of assets than to their expansion.

2.2 Relative Prices and Multiproduct Issues

Many regulated utilities produce more than one distinct product or serve multiple markets. What is the optimal price structure for a multiproduct firm that will remain a monopoly? In telecommunications the balance between line rentals, long-distance call prices and local call prices is important. In railways it is crucial that the relative price of peak and off-peak services is appropriate, while the geographical structure of prices is important in the energy industries. The theoretical answer is that Ramsey prices, which maximize consumer surplus subject to a profit constraint, are optimal. Suppose that there are two products and that consumer surplus is $V(P_1, P_2)$ where P_i is the price of product i. The derivative of consumer surplus with respect to P_i is $-Q_i$, i.e. -1 times the quantity demanded of i. Denoting profits by $\Pi(P_1, P_2)$ the optimal prices are characterized by the tangency condition:

$$(3) \quad \frac{\partial V / \partial P_1}{\partial V / \partial P_2} = \frac{Q_1}{Q_2} = \frac{\partial \Pi / \partial P_1}{\partial \Pi / \partial P_2}$$

and the profit constraint. Can the firm be given incentives to choose the Ramsey prices? If the regulator knows the consumer surplus function then the answer is "yes".

The firm would choose prices to maximize profits subject to the constraint that $V(P_1, P_2)$ equals a target level, which generates prices that satisfy (3).

In practice the regulator is unlikely to be able to compute consumer surplus. Vogelsang and Finsinger (1979) recommended an ingenious mechanism that ensures that the firm eventually sets Ramsey prices in spite of the regulator's lack of information about costs and demands. Using a second subscript to denote the time period the period t constraint is

$$(4) \qquad \sum_{i=1}^{2} P_{i,t} Q_{i,t-1} \leq C_{t-1}$$

which means that the cost to the consumer of buying the previous period's quantities at the new prices must be no greater than the firm's total cost last period. In the limit the repeated application of constraint (4) drives profits to zero and Vogelsang and Finsinger show that equation (3) is satisfied, so Ramsey prices are achieved. The model does depend crucially, however, on the firm ignoring the effect its current decisions have on the level of future constraints. Sappington (1980) shows that a firm acting strategically might choose to inflate costs early on in the process in order to increase rents in the future, and that the welfare effects of this behaviour can be very severe.

The main reason the Vogelsang-Finsinger mechanism has problems is that actual costs are mentioned in the constraint. The price constraints that are applied to BT and to the water companies in England and Wales are of the form

$$\sum_{i=1}^{n} P_{i,t} Q_{i,t-1} \leq (1 + RPI - X) \sum_{i=1}^{n} P_{i,t-1} Q_{i,t-1}$$

where RPI is the percentage growth in the Retail Prices Index. A Laspeyres index of prices is capped, and because the firm cannot affect either RPI or X it has no incentive to manipulate its costs. Vogelsang (1989) shows that in the limit this type of Laspeyres constraint leads to Ramsey prices, though the rents will only be extracted from the firm if X is set appropriately.

In a recent paper Armstrong and Vickers (2000) use mechanism design techniques to analyze multiproduct regulation when there is asymmetric information about costs and demand. They assume that the regulator has no information on realized demands so schemes such as the Vogelsang-Finsinger one defined in (4) are not feasible. When there is uncertainty about costs it is optimal to give discretion over relative prices to the firm (since both the firm and the regulator want prices to reflect realized costs), but with uncertainty over demand there should be no discretion if demand shocks are multiplicative (these do not alter elasticities, so Ramsey prices can be implemented without decentralized decision making by the firm), but some discretion if uncertainty about demand is additive.

A final point to note is that contestable markets theory provides guidance on the regulation of prices when an incumbent monopolist is threatened by entry (see Baumol *et al.*, 1982, and Vickers, 1997). In particular the price for a particular product line the price should lie in between an incremental cost floor and a stand alone cost ceiling if both predatory pricing and cross-subsidy are to be prevented. This issue has been particularly important in the telecommunications market in the UK, where entry is free but BT still has a dominant position in many market segments.

2.3 Access pricing

The access pricing issue arises when a firm, *N*, that operates the network as a natural monopoly also offers services over the network that may be subject to potential competition. Competitors at the retail stage need access to the natural monopoly services of *N* in order to sell to final customers and the question is: what is the correct price for these services? Early cases where this was an issue were (i) the electricity industry before privatization, where independent power producers were allowed to supply electricity to the transmission company, which itself owned most generating capacity and was unwilling to buy the rival power at a reasonable price and (ii) interconnection between the new long-distance rival, Mercury, and the incumbent, BT, in the telecommunications market, where the initial determination allowed Mercury to complete calls via BT's network at a price close to marginal cost. Interconnection remains a critical issue for telecommunications, and will become important in the water industry as competition via common carriage develops there.

The electricity, gas and rail industries have been subjected to vertical separation, which changes the access pricing issue. There is no incentive for a separate network firm to favour one supplier rather than another, and the issue instead is how to recover the fixed costs of the network from charges for the use of the network.

The access pricing issue can be thought of as a special case of multi-product pricing, where N has two types of customer, final consumers and intermediate goods producers, who consist of the rival retailers and its own retail division. Here we present a version of the model of Armstrong *et al.* (1996). Suppose initially that N is vertically integrated firm and has no competition downstream. The marginal cost to N of a unit of network services is b and its marginal cost in the downstream (retailing) sector is c. The regulator has full demand and cost information and, in the absence of public subsidies or two-part tariffs, sets the retail price equal to average cost: $\overline{P} = b + c + F/Q(\overline{P})$ where $Q(\overline{P})$ is retail demand at price $\overline{P}$ and F is the fixed cost in the network sector. Now a competitive supplier enters the retail sector. The competitor produces an identical product, acts as a price taker and has marginal cost in the retail sector of $MC(s)$ which is increasing in its supply, s, and satisfies $MC(0) < c < MC(Q(P))$. The latter condition implies that it is socially efficient to have the competitor doing some (but not all) retailing.

The competitor chooses its output level, s, where its marginal retail cost plus the access charge paid to the network firm, a, equals the retail price, $\overline{P}$. So

(5) $$MC(s) = \overline{P} - a.$$

The competitor's supply is increasing in its margin $\overline{P} - a$ since the marginal retail cost rises with the quantity supplied. Suppose the regulator leaves the retail price at $\overline{P}$ and fixes the access charge. Since consumers pay $\overline{P}$ whoever supplies them they are unaffected by competition and the optimal access price is the one that minimizes total retail costs. Cost minimization requires that the marginal retail costs of the rival and of N are equal:

(6) $$MC(s) = c.$$

Equations (5) and (6) jointly imply that the optimal access charge is

(7) $a^* = \overline{P} - c.$

This is the Efficient Component Pricing Rule (ECPR) of Baumol (1983) and Willig (1979). It has been applied in the New Zealand telecommunications industry, and one interpretation of the access price regime applied to BT from 1992 to 1997 (known as the Access Deficit Contributions scheme) was that it was based on the ECPR (Laffont and Tirole, 1996). A useful way to think of the ECPR is that the marginal cost to N of providing a unit of access is the direct marginal cost, b, plus the opportunity cost of the lost profit on the retail sales, $\overline{P} - b - c$. Adding these gives a^*. Because N is fully compensated for its lost retail profits it still breaks even. Indeed because $\overline{P}$ was originally set to cover average costs the amount by which a^* exceeds b is exactly the average fixed cost $F/Q(P)$. Note that the same outcome holds if N is split into a network firm and a retailer facing the same access charge as the rival and a fixed retail price of $\overline{P}$. Now the regulator sets a to cover N's average costs, so $a^* = b + F/Q(\overline{P})$.

If N is not regulated, will it have the incentive to choose a^*? In general the answer is no. N chooses a to maximize its profits from access. The revenue from selling access is as and the variable cost is bs. In addition the network firm loses retail profits of $(\overline{P} - b - c)s$. So the net profit from access is $[a - (\overline{P} - c)]s(\overline{P} - a)$. Clearly N wants to set a above the ECPR level, $\overline{P} - c$. The social cost of such profit maximization is that the marginal sales that the network firm makes are produced with less efficiency than marginal sales of the rival. Indeed if N and the rival retailer were allowed to merge there would be productive efficiency as the merged firm would maximize its profits by allocating retail sales to minimize costs.

The analysis that justified the ECPR assumed that the retail price must remain fixed at $\overline{P}$. We now assess the consequence of relaxing this assumption. The firm has two types of customer and standard second-best arguments imply that the distortions that are necessary to generate revenue to fund the network's fixed costs should be spread across the two markets, and not all concentrated on final customers. The

Ramsey pricing structure is characterized by an access charge *exceeding* the ECPR level defined in (7), while P is below $\overline{P}$. The firm is able to cut the retail price using the extra profits generated from its access customers. Thus the Ramsey price structure has the network part of N more than covering its costs, while the retail side operates at a slight loss. Naturally if the network and retail businesses of N are separated the Ramsey structure is not sustainable. In practice, however, regulators, while acknowledging the theoretical arguments in favour of Ramsey prices, have not sought to apply them in practice because the informational requirements are excessive.

The analysis so far has assumed that N and the rival produce perfect substitutes. If instead they sell products that are imperfect substitutes in demand then there is a case for reducing the access charge *below* the ECPR level. Suppose, for example, that the demand for a new telecommunications service is completely independent of local call demand, but the new service needs access to N's local lines. In this case N loses no profit from offering access (as long as there is no congestion on the local lines) and the optimal value of a is b, the marginal cost of providing the network services. In general the more independent in demands the products are the closer the optimal access charge is to marginal cost, b.

Finally a recent development in the theory of access pricing has been the analysis of reciprocal access, which applies to telecommunications networks that need jointly to interconnect with each other because of network externalities (see Laffont *et al.* (1998) and Armstrong (1998). How should the access charge be set in this case? An important insight of this literature is that if the firms are free to negotiate an access charge that this will be above the socially efficient level. Suppose that there is symmetry, so that the sum firm 1 pays to firm 2 for call completion is the same as the revenue it receives from 2. At first sight one might imagine that firm 1 is indifferent between any access price. But the access charge is a component of marginal costs, and the higher is the symmetric access price the more each firm will want to raise its final price to customers. In the end the firms could achieve a collusive outcome simply through the use of the access price. The socially optimal access charge will be below the ECPR level in this case to offset the positive mark-ups that oligopolists set, and could even be below the marginal cost of providing the network service, b.

2.4 Establishing upstream spot markets

Spot markets have been created in the electricity and gas sectors in the UK. California, Chile and Norway have undertaken similar reforms. Markets for instant delivery are necessary to ensure that energy systems balance. Particularly in electricity there is a need for minute-to-minute coordination between the transmission company and the power generators. The first example of such a spot market in the UK was the Electricity Pool, a centralized arrangement run by the transmission company that determines which generators are called on to run. Rival generators announce their capacities and prices for each period of the next day, and the Pool constructs a supply curve. The price paid to all generators is the price of the marginal producer. In addition generators receive "capacity payments" which are higher the more likely it is that there will be a power failure and the higher the estimated cost that customers bear in the event of such a failure. If the market for generation is competitive one would expect that this system would generate efficient outcomes, with price in the daily auction equal to marginal cost and the capacity payments providing long-run incentives for investment.

In practice the Pool has worked less efficiently. Standard game theory would suggest that when there are only a few bidders, each with a portfolio of plants with different marginal costs, they will be tempted to restrict capacity or to raise the price of their marginal plants. This will reduce the likelihood that the marginal plants are called on to generate, but will enhance the profits of their plants with lower marginal costs that are almost certain to be called on. In an early analysis of the operation of the Pool, Green and Newbery (1992) modelled the game as one where the competing generators offered supply functions, and showed that the duopolistic structure of the generating market was likely to lead to inefficient outcomes.

Recently the regulatory authorities have moved to promote new trading arrangements in electricity (and gas trading has been reformed). The new trading arrangements are designed to promote more flexibility by allowing trading to take place outside the Pool, by encouraging the growth of futures and forward markets and by allowing demand-side bidding, while maintaining a short-run balancing market. One oddity, though, is the proposal that in the balancing market bidders should receive the amount of their bid rather than the price of the marginal bidder. Wolfram

(1999) uses the auction theory to criticize the introduction of discriminatory pricing. Changing the rules of the game will alter bidding behaviour, and generators, such as the nuclear companies, who currently bid close to their own marginal costs, will be tempted to increase their bids. The point to note here is that the theory of mechanism design can help regulators to establish rules of such spot markets which ensure that the participants have incentives to behave in ways that enhance the objectives of regulators.

3. Regulation of industry structure and entry conditions

3.1 Industry structure

It is commonly argued that regulation is easier when the natural monopoly parts of a firm are separated from the competitive parts. In the usual case the natural monopoly business is a network that rivals must use in order to reach final customers, as in the access pricing case considered in Section 2.3, so vertical separation is called for. BT and the water companies were not restructured at privatization but the electricity and railway industries have been vertically separated, and the incumbent in the gas market broke itself up a decade after privatization. If information is symmetric, however, it is difficult to see why vertical separation would help. We saw in 2.3 that with the retail price fixed at $\overline{P}$ vertical separation generates exactly the same outcome as allowing vertical integration while using the ECPR (equation 7) to determine the access price.

When an incumbent is vertically integrated it is likely that the regulator will find it difficult to determine the optimal access price, since cost information is likely to be noisy, and the incumbent has the incentive to allocate as many costs as possible to the network part of the business. Suppose that the regulator can observe $b + c$ when the firm is integrated, but cannot split this cost up into the network component, b, and the retail part, c. The firm reports its value of c (and thus of b). Let this cost report be $\hat{c}$ while the true value is c. The regulator sets the retail price on average cost pricing principles so $\overline{P} = b + c + F/Q(\overline{P})$ and the access charge is determined using the ECPR so $a = \overline{P} - \hat{c}$. If the firm announces the correct value of c then it earns profits of zero.[5] By announcing $\hat{c} < c$ the firm does not alter the retail price but it pushes up

[5] Profits are $\Pi = (P - b - c)[Q(P) - s(P - a)] + (a - b)s(P - a) - F$. Substituting in for P and a yields $\Pi = 0$.

the access charge and thus increases its profits. The asymmetry of information allows the firm to set the profit-maximizing access price, thus generating productive inefficiency in retailing. If vertical separation allows the regulator to observe both b and c then he sets $P = a + c$ and $a = b + F/Q(P)$. Productive efficiency in retailing is achieved (and rents are minimized) so the result is superior in welfare terms to the vertically integrated case. If the costs of break up are smaller than the welfare gain from the improvement in the regulator's information caused by separation then it is worthwhile.

Two other aspects of industry structure have also been of concern to regulators. In the electricity industry the UK Government decided to alter *horizontal* structure at the generation stage. The Central Electricity Generating Board was split into three companies, a nuclear company with 20 percent of capacity, and two fossil-fuel generators with 50 percent and 30 percent of capacity. The nuclear company played no role in determining prices in the Electricity Pool since it bid zero to ensure that its plants always ran, so effectively a duopoly was created. The conditions for collusion could hardly have been better. The theory of non-cooperative collusion (see Tirole, 1988, Chapter 6, and Shapiro, 1989) suggests that it is sustainable when the number of players is low, there is a homogenous product, frequent purchases are made by customers and when the firms have spare capacities so prices can be driven down in the punishment phase. Even if the firms are acting non-cooperatively standard theory suggests that price-cost margins are inversely related to the degree of concentration in the industry. Green and Newbery (1992) suggest that welfare would have been significantly higher if five competing generators had been created at the outset.

The ability to apply yardstick competition is dependent on *regional separation* of firms. In the water and electricity distribution industries this regional separation existed in any case before privatization. There might be a trade-off between keeping firms regionally separated to provide comparative information and allowing them to merge to gain the benefits of economies of scale and scope. In the water industry all non-trivial mergers have to be referred to the Competition Commission, which must take account of the impact of a proposed merger on the regulator's ability to make comparisons.

3.2 Liberalization of entry conditions

The final aspect of structural regulation is the licensing policy of the regulatory authorities. When competition is feasible the question is should the regulator all free entry or should the rate of entry be controlled? Initially in telecommunications UK regulators opted for slow liberalization. Only one competitor, Mercury, was licensed to compete with BT, and simple resale of capacity leased from BT at wholesale rates was not allowed. Following a review of the duopoly policy in 1991 entry was fully liberalized. In electricity and gas entry into the market for supply to large customers was liberalized early on in the process but was not effective initially because of the lack of attention to access prices and conditions.

One argument for allowing limited entry is that entry can lead to excessive duplication of fixed costs when the firms sell homogenous products. Mankiw and Whinston (1986) present the theoretical argument for this. Entry of a new firm benefits consumers through reducing prices but harms existing firms as the entrant steals some of their profitable business. At the margin when there are sunk costs, so price exceeds marginal cost in free entry equilibrium, it is better to have slightly fewer firms. This result is not robust, however, to changes in the assumptions. Restricting entry might facilitate collusion. New firms might introduce different products that provide extra consumer benefits. Entrants might be more efficient than the incumbent in utility markets. The alternative view that entry should be fully liberalized has dominated since the end of the duopoly policy for telecommunications. Apart from the non-robustness of the Mankiw and Whinston result a free entry policy can be justified on the basis of contestability theory or from an Austrian standpoint that competition *per se* is best.

A new entrant in a utility supply market has to offer a significant price discount relative to the incumbent because consumers face switching costs. The ability to take a competitive supply might require the customer to install new equipment, such as a new telephone or a meter, to use a new telephone number, to remain at home in order to have the meter read when the switch takes place, or to key in extra digits to access a new provider. If the incumbent uses complicated and frequently changing nonlinear tariffs then the customer faces another switching cost because price comparison is more difficult. Entrants will have to offer significant price discounts to attract customers. Evaluation of the effectiveness of competition should take allow for the switching costs that are incurred. In addition new entrants have to spend heavily on advertising make customers aware of their existence and to persuade them to switch. Advertising is an endogenous sunk cost in the sense of Sutton (1991) and equilibrium market structures might not be characterized by large numbers of competitors. The more advertising is necessary to win and to retain customers and the greater the degree of price competition that is expected between the rivals the more concentrated the market structure is likely to be.

4. Regulatory Commitment

When regulators cannot commit to future prices there is a hold-up problem – firms might not invest optimally. Utilities' assets have long lives and few alternative uses and it is tempting for politicians to adjust previous agreements. We can show the problem with a simple variation of the model of section 2.1. Suppose now that there is no asymmetry of information (both sides know θ and e) so there is no incentive constraint. The first-best solution has the firm choosing e^{**}, which minimizes total costs, and the regulator setting the price equal to $\theta - e^{**} + F(e^{**})$ to cover all costs. In the new game the firm chooses its effort level and incurs the costs of effort *before* the regulator sets the price. The problem is that in a one-shot game the regulator takes the sunk costs, $F(e^{**})$, as given. Once the effort costs have been sunk the firm will participate as long as operating costs, c, are covered. The regulator's objective is to choose p to maximize $W = U - p + \alpha[p - c] - \alpha F(e)$ subject to $p \geq c$. For $\alpha < 1$ the solution is $p = c$ and the sunk costs are not covered. Anticipating this outcome the firm chooses not to invest in any effort at the first stage and costs are excessively high. This underinvesment problem is caused by the inability of the regulator to commit to a price level in advance of the firm choosing its sunk costs.

There are two solutions to the under-investment problem. First, even if it is impossible for the regulator to commit to a price schedule it might be in his interests to avoid expropriation if the game lasts indefinitely. Salant and Woroch (1992) have a model where the regulator allows the firm to recover both operating costs and the incremental sunk costs incurred in that period. The firm in turn increases its effort level (or capital stock) to approach the optimal level (e^{**}) asymptotically. The fact that the firm always has further investment in effort that it can undertake prevents the regulator from wanting to renege on the price agreement, because the firm will refuse to do any more investment if the regulator has cheated. Salant and Woroch show that the equilibrium of this game can be approximately efficient. The second solution is to develop institutions that effectively create commitment. Levy and Spiller (1996) discuss the importance of the institutional endowment of a country in determining the degree of commitment. At a minimum the regulatory agency needs to act independently of the government (though there must be some accountability), so the agency should not simply be part of a government ministry. The parallel with the argument for an independent central bank should be clear. Levy and Spiller point out that it is also important that the politicians are not able to change the rules of the game arbitrarily, so the way the legislature works, the degree of independence of the judiciary and the availability of appeal mechanisms are also important. It can be argued that the UK is relatively fortunate since its institutional endowment is sufficiently strong that the danger of underinvestment is low.

5. Conclusion

In this paper I have presented a selected review of regulatory principles and related them to UK experience. Where are regulation and the theory of regulation likely to go from here? There are at least three issues where there is a need for more work. First, while price cap regulation seems provides good incentives for operating cost efficiency, it is not clear that it provides optimal investment incentives for utilities. Developing robust mechanisms for investment without encouraging gold-plating will be important. Secondly there is likely to be more attention paid to relative prices in the network parts of the utilities than to price levels, which in many cases have been reduced significantly because of the incentives provided by price cap regulation. Thirdly regulators are already coping with the issue of managing competition in retail markets while protecting existing customers of incumbents from excessive pricing. If equilibrium market structures are going to be characterized by a small number of competitors, perhaps supplying multiple utility services, regulators will also have to be aware of the dangers of collusion.

References

Armstrong, M., Doyle, C. and Vickers, J. (1996) "The Access Pricing Problem: A Synthesis", *Journal of Industrial Economics*, vol. 44, pp. 131-50.

Armstrong, M. (1998) "Network Interconnection in Telecommunications", *Economic Journal*, vol. 108, pp. 545-64.

Armstrong, M. and Vickers, J. (2000) "Multiproduct Price Regulation Under Asymmetric Information", *Journal of Industrial Economics*, vol. 48, pp. 137-60.

Baumol, W. (1983) "Some Subtle Issues in Railway Regulation", *International Journal of Transport Economics*, vol. 10, pp. 341-55.

Faulhaber, G.R. and Baumol, W.J. (1988) "Economists as Innovators: Practical Products of Theoretical Research", *Journal of Economic Literature*, vol. 26, pp. 577-600.

Green, R. and Newbery, D. (1992) "Competition in the British Electricity Spot Market", *Journal of Political Economy*, vol. 100, pp. 1089-99.

Kay, J. and Vickers, J. (1988) "Regulatory Reform in Britain", *Economic Policy*, vol. 7, pp. 285-343.

Laffont, J-J., Rey, P. and Tirole, J. (1998) "Network Competition: I Overview and Nondiscriminatory Pricing" and "Network Competition: II Discriminatory Pricing", *RAND Journal of Economics*, vol. 29, pp. 1-37, 38-56.

Laffont, J-J. and Tirole, J. (1986) "Using Cost Information to Regulate Firms", *Journal of Political Economy*, vol. 94, pp. 614-41.

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

Laffont, J-J. and Tirole, J. (1996) "Creating Competition Through Interconnection: Theory and Practice", *Journal of Regulatory Economics*, vol. 10, pp. 227-256.

Levy, B. and Spiller, P.T. (1996) "A Framework for Resolving the Regulatory Problem", in Levy, B. and Spiller, P.T., eds., *Regulation, Institutions and Commitment*, Cambridge, UK, Cambridge University Press.

Littlechild S.C. (1983) *Regulation of British Telecommunications' Profitability*. London: Department of Industry, HMSO.

Littlechild S.C. (1986) *Economic Regulation of Privatised Water Authorities*. London: HMSO.

Mankiw, G. and Whinston, M. (1986) "Free Entry and Social Efficiency", *RAND Journal of Economics*, vol. 17, pp. 48-58.

Salant, D. and Woroch, G. (1992) "Trigger Price Regulation", *RAND Journal of Economics*, vol. 23, pp. 29-51.

Sappington, D. (1980) "Strategic Firm Behaviour under a Dynamic Regulatory Adjustment Process", *Bell Journal of Economics*, vol. 11, pp. 360-72.

Shapiro, C. (1989) "Theories of Oligopoly Behaviour", in Schmalensee, R. and Willig, R., eds., *Handbook of Industrial Organization*, Vol. 1, Amsterdam: North-Holland.

Shelifer A. (1985) "A Theory of yardstick Competition", *RAND Journal of Economics*, vol. 16, pp. 319-17.

Sutton, J. (1991) *Sunk Costs and Market Structure*, Cambridge, MA: MIT Press.

Tirole, J. (1988) *The Theory of Industrial Organization*, Cambridge, MA: MIT Press.

Vickers, J.S. (1997) "Regulation, competition and the structure of prices", *Oxford* Review *of Economic Policy*, vol. 13(1), pp. 15-26.

Vogelsang, I. (1989) "Price Cap Regulation of Telecommunications Services: A Long-Run Approach", in M.A. Crew, ed., *Deregulation and Diversification of Utilities*, Boston, Kluwer.

Vogelsang, I. and Finsinger, J. (1979) "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms" *Bell Journal of Economics*, vol. 10, pp. 157-71.

Willig, R. (1979) "The Theory of Network Access Pricing", in H.M. Trebing, ed., *Issues in Public Utility Regulation*, Michigan State University Public Utilities Papers.

Wolfram, C. D. (1999) " Electricity Markets: Should the Rest of the World Adopt the UK Reforms", University of California Energy Institute POWER Working Paper No 69. http:\\www.ucei.berkeley.edu/ucei