

No. 54

Regulation & Contracts for Utility Services:
Substitutes & Complements

Lessons from UK historical experience

Jon Stern

REGULATION AND CONTRACTS FOR UTILITY SERVICES: SUBSTITUTES OR COMPLEMENTS?

LESSONS FROM UK HISTORICAL EXPERIENCE

Jon Stern

London Business School and NERA

ABSTRACT

It is frequently suggested that regulation by contract can effectively substitute for regulation by a specialist regulatory agency for utility service industries. We examine these arguments and consider both legal aspects and the experience of the UK in the nineteenth and early twentieth century. We consider in detail the UK's problems with reliance on franchise contracts in the railway and electricity industries and in other utility service industries. We conjecture that regulation and contracts are complements for network industries rather than substitutes. The existence of a regulatory agency allows for better and simpler contracts, which are easier to monitor, enforce and revise. This is what would be expected from the theory of incomplete contracts. We demonstrate that UK historical experience is strongly consistent with this view.

1. Introduction

In much of the discussion about the economic and legal framework for infrastructure utilities (such as electricity, telecommunications, railways, etc), the establishment of an independent regulatory agency is seen as an alternative to contracts – or vice-versa.¹ A particular theme is whether or not the use of contracts can reduce or even eliminate the need for a regulatory agency. A good example of such discussions is the recent paper by Littlechild (2002) in which he argues that competitive bidding for long-term contracts can have significant merits over regulation in the context of electricity distribution.

This paper argues that it is misleading and unhelpful to consider utility regulation as a *substitute* for contracts. In all regulated utilities, contracts play a major role but regulation provides elements that cannot readily be provided by contracts. In particular, the presence of a regulatory agency operating under a defined regulatory process provides an ordered and low cost way of handling substantial renegotiations. This is a crucial support for private investment in long-lived, capital intensive industries with sunk costs. In addition, regulation has major advantages over contracts where there are large numbers of small consumers and also where access and pricing rules have to provide for multiple users of monopoly networks.

¹ This paper has benefited greatly from the comments by seminar participants at the DAE Cambridge Regulation Seminar, the LBS and the University of Sussex. I am particularly grateful for comments and encouragement from Chris Bolt, Saul Estrin, Stephen Littlechild, Shira Lewin-Solomons, David Newbery, Mark Schankerman and Ralph Turvey. The paper and the views expressed in it are, however, solely the responsibility of the author.

These issues have been present since the railway age of the 1840s so that economic history provides useful evidence. The topic also raises questions about the legal as well as the economic role of contracts and their limitations. The way in which contracts have evolved in legal terms and the legal status of regulatory agencies is relevant not just in historical terms but also to the current relationship between regulation and contracts in developed and developing countries, as will be discussed below.

Much of the economic literature in this area draws on problems of incomplete contracts in the face of uncertainty and the related issue of the advantages and disadvantages of vertical integration and these will inevitably play a significant role in what follows². Critical issues here are major information uncertainties (eg on potential technical progress rates of industries in the early stages of development) and how best to organise the development of physical infrastructure networks over which services are provided. This has regularly raised difficult problems of the costs and benefits of centralised versus decentralised expansion; the potential for competition *in* the market and *for* the market; the advantages and disadvantages of vertical integration between network and services; how best to relate prices to costs and – always – how to achieve an acceptable balance between the economic interests of consumers and investors. Information uncertainties and incomplete contracts may reveal a need for explicit regulation; but, achieving effective regulation is difficult and regulation is also incomplete and imperfect.

The need for co-ordination of network development and the question of price evolution in rapidly developing infrastructure industries is graphically demonstrated in the history of the UK railway industry and the emergence of the electricity industry in the mid-late nineteenth century. Piecemeal, bottom-up developments based on individual contracts created major problems and economic inefficiencies. For the railways, the need for a regulatory entity to supplement the contracts was recognised by some observers before 1840. Unfortunately, the level of understanding both of the economic issues and of regulatory design was understandably limited so that the attempted regulatory solutions adopted were seriously flawed and failed to provide an effective regulatory supplement to the contractual arrangements. Nevertheless, as well as providing a fascinating story, these early examples provide important insights into the underlying economic problems with which we are still having difficulty - in developed as well as in transition and developing countries.

The questions of whether and how sound and effective regulatory institutions can be created and sustained to support private investment in infrastructure was at the heart of nineteenth century UK debates on the appropriate economic framework for the railways and no good answer was found. We have learnt much over the last 100 years about how to do utility regulation but there are still major regulatory design failures – in developed as well as in developing and transition economies. Nevertheless, finding effective regulatory solutions remains critical to strategies for increasing the role of private investment and competition into these industries.³

For developing and transition economies (DTE's), introducing private investment in infrastructure utilities depends heavily on the existence of effective regulatory agencies. However, introducing private investment, let alone liberalising and unbundling the industries, becomes much easier on resource availability grounds if contracts can be substituted for

² See, for instance, Williamson (1985) and Hart (1995). See also Hart (2003).

³ See Stern and Holder (1999) and Noll (2000).

regulatory agencies⁴. In addition, there is the related issue of whether regulation is primarily concerned with contract enforcement (which implies minimising or eliminating regulatory discretion); or whether effective regulation inevitably involves the use of accountable and bounded discretion (eg in agreeing changes in investment and pricing plans)⁵. Similar questions also arise in developed countries. This is demonstrated by the recent UK debates and decisions over the relative role of regulatory agencies versus contracts (a) in the organisation of the major investment programme for the London Underground railway infrastructure and (b) the framework for the UK railway system consequent on the commercial failure of the privatised Railtrack network company⁶.

This paper is primarily concerned with the limits to contracts and their uses from a legal perspective and as demonstrated from nineteenth century UK economic history. The rapid expansion in recent years of independent regulatory agencies with non-trivial amounts of (bounded and accountable) discretion is best seen as a response to the problems of relying solely on contracts for infrastructure development. That is not to argue that regulatory agencies can substitute for the use of contracts – this was not the case in nineteenth century Britain and is not the case today. Countries in which contract enforcement is problematic are highly unlikely to be able to sustain effective and independent regulatory agencies. The paper will, however, argue that some aspects of utility service industry organisation and supply raise issues that unavoidably need regulation to supplement contractual arrangements. The point of this paper is therefore to use UK historical experience, primarily in railways and electricity, to help identify under what circumstances regulation via an independent regulatory agency is unavoidable - or at least seriously advisable. A companion paper will discuss contemporary concerns.

The structure of the paper is as follows. In section 2, we discuss contracts and their role in a market economy including some legal implications. In section 3, we discuss the historical record on what regulation has done to supplement contracts in the area of utility services, with particular reference to nineteenth century UK experience with railways and other utility service industries and, in section 5, we discuss the pre-1930s electricity industry. Section 6 provides a short economic appraisal of the UK experience in these industries and the paper ends with some concluding comments.

⁴ See Domah, Pollitt and Stern (2002) and Stern (2000).

⁵ Levy and Spiller (1994) is the classic exposition of the need to minimise or eliminate regulatory discretion while Stern (1997) argues for the virtues of (and the necessity for) accountable regulatory discretion as, very eloquently, does Foster (1992).

⁶ The issues listed in this paragraph are discussed in detail in a forthcoming companion paper. See Stern (2003).

2. The Role of Contracts in a Market Economy: Legal Implications

To understand why the Victorians and later generations have found it necessary to introduce regulation by explicit agency to supplement (and if necessary override) private law contract frameworks for utility service industries, we need to consider the role of contracts in a market economy and what they can do and their limitations. We do so in this section.

In this and subsequent sections, the key distinction that we wish to draw is between:

- (a) Regulation by Contract
Franchise or other long-term supply contracts (including network access contracts) whether between private companies or a private company and a state owned entity, *where there is no regulatory or equivalent agency providing external review and revision of terms* (particularly on prices); and
- (b) Regulation by Agency
Utility service providers operate within a framework which provides ex ante regulation *by external agency* for at least some elements of supply, whatever contracts may be in place. This is typically (but not always) associated with private investment and competition on non-monopoly elements.

2.1 Contracts and Commercial Behaviour

It has long been known that the percentage of contract disputes that are taken to the courts is minimal⁷. Indeed, the evidence seems to be that the use of contracts as a sanctioning tool to enforce agreements is much less important than their use to summarise understandings. In this perspective, for business transactions (eg between companies), contracts may best be thought of as providing a framework of agreement that can be used as a reference point in all subsequent discussions – both for positive developments and when things go wrong. In other words, in practice, contracts are used much more to summarise commercial understandings rather than to provide the basis for legal action in the event of one or other party failing to meet their obligations under the contract. Indeed, commercial contracts between companies appear primarily to codify relations and agreed areas of *trust*.

The perspective above is set out and developed in a major, recent legal exploration of contract law by Hugh Collins⁸. Collins summarises as follows the paradox of the very limited role of contracts in commercial transactions relative to their massive role in legal and economic theory:

“On the one hand, most political, economic and social theory suggests that in a market economy the law of contract provides a fundamental mechanism of social order.... On the other hand, the evidence from empirical studies of contractual behaviour indicates the marginal and sometimes socially disintegrative effects of the law of contract. Consumers who purchase defective products almost never vindicate their legal rights in the courts.... Businessmen seem not to be concerned about the legal validity of their transactions, do not insist upon their legal rights, and when a

⁷ See Macaulay (1963) and much subsequent literature.

⁸ See ‘Regulating Contracts’ by Hugh Collins (1999). Most of the rest of this section is based on the highly thought-provoking analysis in this book.

dispute requires adjudication, they avoid the ordinary courts like the plague.... There is also evidence that whole swathes of the most vibrant parts of the economy such as the financial and banking sectors function almost entirely within their own regulatory system that renders the law almost irrelevant.”⁹

Some of the points that Collins raises in more detailed discussion of the role of contracts usefully amplify the very limited role that contracts have as a *sanctioning device* within market economies – particularly in the context of repeated transactions between firms and/or individuals.

Among the key points he makes are:

- The ability of markets to exist and to continue function effectively in the absence of contracts or legal frameworks in surprising circumstances (eg Accra street trade, informal consumer credit markets in New York) – markets without state backing or support;
- The role of trust in economising on contractual complexity;
- The existence of many other sanction mechanisms besides enforcement of contract law;
- The role of implicit obligations eg in “club markets” or towards consumers;
- The role of reputation and long-term commercial interests in providing non-contract solutions in circumstances where problems arise.

For economists of any persuasion, it is not particularly surprising that most breaches of contract do not lead to legal action. Given the costs of legal action in terms of costs of legal proceedings, time delays and management costs – as well as public revelation of details of private dealings and conversations – the expected benefits from taking legal action to enforce contracts or obtain damages will only rarely outweigh the expected costs. This is even more likely to hold for contractual disputes between a company and individual small consumers.

What is perhaps more surprising, and more useful, is the notion of the role of trust as crucial to the understanding of the role of contracts in economic transactions. There is considerable evidence that contractual disputes are most likely to end up before the courts when there has been a major breakdown in trust between the parties.¹⁰ In other words, taking legal action over a contract declares a loss of trust between the parties and a desire from one or both parties to terminate commercial relations permanently. Such actions often seem designed to obtain revenge as much as compensatory damages.

The notion of contracts as a way of creating and reinforcing trust between parties is a particularly constructive way of thinking about the relations between economic agents and their legal embodiment. This particularly applies to agents of comparable economic size and power (eg manufacturers and wholesalers/retailers or producers of intermediate goods and final producers). Indeed, this perspective fits more naturally into modern micro-economic theory with its emphasis on reputation, repeated transactions and the need to sustain incentives for continued supply and investment by commercial partners¹¹. Such topics are a major staple in game-theoretic micro-economic analysis.

⁹ See Collins op. cit, p. 5.

¹⁰ Macaulay (1963), Beale and Dugdale (1975) as examples.

¹¹ See Williamson (1985), Hart (1995) and many others.

Particularly noteworthy is Collins' discussion of the role of "market conventions" which incorporate "unwritten laws" accepted by both parties but not formally included in contracts¹². This not only relates closely to the micro-economic literature on the behaviour of firms discussed above; it also relates directly to the importance of all players in the regulatory game knowing and observing the "rules of the game" which is at the core of the notion of 'informal accountability' in utility regulation¹³. In both cases, unwritten understandings appear to play a major role in how formal agreements (contracts, laws, regulatory codes) are applied in practice. This is typically with a great deal of interpretation and discretion rather than in a literal way with punishment or sanctions applied for all (or most) breaches.

One response to the arguments set out above - particularly the existence of markets without contract backing - is that they may be fine for *exchange* markets like informal street markets, but that they cannot support complicated production relationships, particularly where large amounts of long-lived capital are involved. It is hardly surprising that in those circumstances contracts are more prevalent and more detailed. However, there is still a major difference between:

- (a) contracts that are negotiated primarily as a basis of incorporating a long term commercial understanding and providing the basis for regular updating (such as private sector "open book" contracts); and
- (b) contracts which are primarily designed to provide effective sanctions against poor performance by other parties¹⁴.

Contracts of type (a) are very common in the commercial world, including utility service related contracts such as PPA's (power purchase agreements). However, many PPA and other infrastructure service contracts, particularly for foreign investors building facilities in developing countries, appear to be type (b) contracts. In the light of the analysis above, it is not surprising, firstly, that they have frequently failed to provide a continuing basis for viable operation; nor, secondly, that many have dissolved fairly early in their life in distrust and mutual acrimony (viz IPP programmes in Indonesia and Pakistan). Type (b) contracts are more likely to be clearly constructed on the basis of mistrust and rights of redress in bad outcomes rather than how best to use and enhance commercial opportunities and build on good outcomes¹⁵.

The emphasis on contracts primarily embodying trust does not mean that they are irrelevant in providing sanctions if things go seriously wrong between the parties - either in legal or in economic terms. Collins provides good theoretical and practical evidence to support the notion of contracts as providing a reference point for working out solutions to problems or mistakes, even if they are not solutions obtained by legal process in the courts.¹⁶ In addition, the ability to go to the courts does provide some long-stop sanction against gross misbehaviour or breach of trust. However, even here the position is much easier between companies (or other agencies) of comparable power. This is shown by the difficulties of using the courts to provide effective and

¹² See Collins op cit, p. 29 and Beale and Dugdale (1975).

¹³ See Stern and Holder (1999) for a discussion of this concept (eg in the way that regulatory processes can increase or reduce the effectiveness of regulatory agencies.)

¹⁴ Note, though, that regulatory regimes have difficulties in providing effective sanctions against poor performance. But, they do have greater possibilities of providing incentives to avoid it.

¹⁵ See Stern (2003) for a discussion of this issue in the light of some model PPA contracts for different countries.

¹⁶ See Collins op cit Chapter 5.

prompt redress in cases affecting large numbers of small consumers (viz the thalidomide case and others since).

The discussion above has concentrated on UK law. The same issues arise, however, in many countries and similar types of solution appear to be result, taking on the appropriate legal form in different countries.

2.2 Contracts, Private Law and Regulation

The discussion in the previous section has implicitly taken the conventional economic assumption that contracts are classic examples of *private* law and, as such, is to be clearly distinguished from regulation by agency. This appears to be the implicit assumption in Littlechild (2002) and is also the basis for the critique of regulatory institutions by the Chicago economics and law proponents¹⁷.

This strong distinction is strongly challenged by Collins who argues that, since the late nineteenth century, private law has increasingly included elements of regulation within its domain in areas such as consumer contracts, employment contracts, etc. He argues strongly that many aspects of regulation can (and are in practice) handled in private law and that this works very well when the parties for whom regulatory concerns arise are those who negotiated the original contract.

The relevant set of contracts could include long-term contracts either (a) between private companies or (b) concession contracts between national or local governments or state-owned companies/agencies and private companies. For infrastructure industries, the former can include network infrastructure access and use agreements (eg third party access arrangements and telecom interconnection agreements) as well as oil and gas production contracts, power purchase agreements, etc. The latter can include long-term concession or franchise contracts between state or local governments or state-owned companies/agencies and private companies (eg road and bridge concessions, electricity distribution franchises, cable television franchises, etc)¹⁸.

In this perspective, the key distinction between regulation by private law and regulation by agency or by government is that private law regulation “differs from other kinds of business regulation ... simply in the generality of rules which only become specific when applied by private ordering and adjudication”.¹⁹ In other words, it is not that contracts and private law cannot cover aspects of public policy and regulation; rather it is that it is hard (and very costly) for them to do so when many parties are involved.

Private law is much less efficient at resolving problems when there are issues of general standard-setting, monitoring and enforcement. Hence, on this view, it is unsurprising that regulation by utility agencies concentrates:

¹⁷ See, for instance, Posner (1992)

¹⁸ We use the terms *concession* or *franchise* contracts interchangeably to denote long-term (eg 20-year or longer) contracts with control of the assets of the business for the duration of the contract. The distinction between concession and franchise contracts is that, under concession contracts, the ownership of the assets always remains with the national or local government and reverts to them at the end of the concession, although the concession may be retendered for a fixed period or permanently. Franchise contracts can be a prelude to permanent privatisation.

¹⁹ Collins op cit, p. 59.

- (a) on small consumers (prices, quality of service, etc) and
- (b) on network access rights and prices.

Private law contracts between large consumers and electricity or other utility supply companies can adequately handle “regulatory” issues.

One of the key features of the 30-year contract for the London Underground electricity distribution network and Seeboard examined in Littlechild (2002) is that London Underground is the sole customer. This is a feature that is extraordinarily uncommon with electricity or other networks and, from a legal perspective, makes it unusually appropriate among network electricity contracts to be handled by a private law contract.

If private law and contracts increasingly including public interest elements, should we also retain the view that utility (or other) regulation by agency is part of public law and outside the remit of private law? Collins argues convincingly that Ofgem, Oftel, etc embody major elements both of public law (eg judicial review of actions by the regulator) with major elements of private law. There is no question that regulatory agencies can impose particular obligations in contracts involving regulated companies so that the “commercial relations [of regulated companies] have been constituted within a confined space set by considerations of the public interest”²⁰. Nevertheless, UK electricity consumers do (now) purchase their electricity on the basis of contracts with supply companies; they do have contracts with mobile phone service providers, gas suppliers, train operating companies, etc. In addition, the electricity suppliers purchase electricity on the basis of contracts, but have to operate under transmission and distribution use of system agreements that must be approved by Ofgem.

However, given the mixture of public and private law elements, it is not surprising that the boundary between contracts and regulation is far from clear-cut. For example, in the UK, codes such as the natural gas Network Code operate as contracts between the network operator (Tranco) and the shippers. Codes and changes to the Code must be approved by the regulator (Ofgem), but Ofgem cannot directly initiate Code changes; in addition, appeals against Ofgem approved Code changes can only be made to the courts under judicial review. This is in contrast to Tranco’s regulatory licence where Ofgem can propose licence changes but appeals can be made on substance to the Competition Commission. This overlap can cause problems in assignment of changes, particularly where market mechanisms are being developed. Hence, in developing capacity auctions for Tranco, the auction design was set out in the Code whereas the requirements for auctions, reserve prices and auction quantities were set out in the Tranco licence²¹.

The example above illustrates the kind of difficulties that can arise when regulatory provisions and contracts are combined, but they appear to be awkwardnesses with which parties can live rather than fundamental difficulties. Although they increase transaction costs, the alternative of relying on multilateral negotiation between network operator and users seems to be a much more difficult and higher cost route. However, the underlying issue is probably whether the regulatory powers given to Ofgem and other regulatory agencies provide a bias towards interventions with poor benefit-cost payoffs. That depends on the other checks and balances in the regulatory system.

²⁰ See Collins op cit, p. 307-8.

²¹ I am grateful to Chris Bolt for pointing this out (personal communication).

Summarising this section, we find that the legal framework discussed above suggests that in utility service industries what we should *expect* to observe is that:

- (i) most transactions involving licensed utility companies are carried out on the basis of private law and contracts (which may include regulatory elements); *but that*
- (ii) contracts with small customers are likely to include requirements laid down by the regulator; *and that*
- (iii) general requirements, standards, codes and provisions are approved, monitored and enforced by the regulator under powers derived from general legislation.

This is what I would argue we do observe, not just in the UK, but in the US, the EU and other OECD countries²². How far it is (or can) be replicated in developing countries is discussed in a companion paper.

²² Including German telecoms but not (yet) for German electricity and natural gas. See Brunekreeft (2002) for a recent discussion of the German electricity industry and its non-regulation.

3. Utility Contracts and Regulation for the Railways in Nineteenth Century Britain

Regulatory agencies are not a recent nor a US invention. In the UK, railway regulation emerged in 1844 and, more formally, in the 1870s as an attempt to deal with the serious problems that arose with leaving the private financing and management of railway services to long-term contracts²³. The US Interstate Commerce Act of 1887 was largely modelled on (an idealised version of) the UK 1844 Railways Act.

The start of the modern railway age is usually dated from 1830 when the Liverpool and Manchester Railway started regular operations following the passing of an authorisation Act in 1826. Railways were the first modern infrastructure *industry*. In particular, there were three key features about railways which created a perceived need for explicit regulation by agency. These were that:

- **Railways typically provided a bundled service.** Railway companies usually (but by no means always) provided the trains and train services as well as the tracks over which the trains ran – unlike roads or canals.
- **Railways experienced major unanticipated (and unanticipatable) changes in costs as the industry developed.** Because of very rapid technical progress, UK railways were subject to very rapid rates of productivity growth for the first few decades of their development after which productivity growth slowed down rapidly and then stagnated after 1900.
- **Railways were the first example of an integrated complex industrial network** with all the problems that created in terms of network externalities, sufficient versus wasteful investment, network access and pricing, etc.²⁴

Policy-makers first had to grapple with these issues in the period 1830-50. This, of course, was long before the distinction between marginal and average costs had been worked out and also before there was any clear understanding of natural monopoly and its relationship to scale economies. Indeed, these concepts were developed later in the century, not least in response to the problems of railway development and pricing.

In practice, the critical problems that arose with long-term franchise contracts in UK railways were:

- (i) Considerable difficulties in finding procedures within the contracts that allowed for rate reviews and, hence, progressive divergence between prices and costs – both in general and on specific rates; and
- (ii) The inability to impose any common standards.
 - for accounting practices and undue discrimination between customers; or

²³ This section draws heavily on the excellent survey in Chapters 1 and 2 of Foster (1992)

²⁴ For a discussion of efficiency and costs in the nineteenth century UK railway and other infrastructure industries, see Foreman-Peck and Milward (1993).

- over common standards to consumers (eg service obligations, quality of service, health and safety, etc).

The results of this were:

- a) Increasing difficulties in achieving rate changes;
- b) An inability to stop the padding of the asset base, or to exert any downward pressure on costs;
- c) An inability to handle increased horizontal or vertical integration within railways or between railways and shipping or road transport; and
- d) An inability to investigate or control undue discrimination in pricing.

In the 1840s, it was far from obvious to legislators and policy-makers that railways (or at least the tracks) had substantial monopoly power or what was the nature of that monopoly power. From 1844 to 1914, various attempts were made to handle these issues with more explicit regulation by agency attempting, ultimately unsuccessfully, to address the gaps left by relying on contract law alone.

During the late nineteenth century, other UK utility service industries involving private investment (notably town gas, rubbish collection, water and later electricity) also introduced regulatory devices within a contractual framework to try to overcome the inherent problems of providing these services within a purely private law contractual framework. These attempts had mixed success and municipalisation became prevalent from around 1890. We discuss these other industries and the experience with the emerging electricity industry in the next Section.

3.1 Before the Railways – Roads, Bridges and Canals

In the UK, infrastructure facilities– typically roads then canals, railways, etc – were constructed and managed under *franchise contracts*. Thus, in the eighteenth and nineteenth centuries, each new UK toll road would be constructed by private investors authorised under a specific parliamentary act. This act provided a contract enforceable in the courts and specified the main conditions for the franchise.

Typical conditions for the franchise were:

- A 21-year franchise (renewable by passing a new act);
- Dividend payout limitation of 10%;
- Toll rates specified in the Act

For trunk roads and canals, the costs were very largely up-front construction costs with relatively low current costs (maintenance and related). As a result, franchise contracts of this type were an effective way of supporting private investment in these facilities – particularly in a low inflation era. In addition, network effects were limited or absent and joint costs were low.

The point of the tolls was to allow the investors to recoup and earn a return on their original capital outlay plus allowing them to cover (low) ongoing costs eg for maintenance, etc. For neither roads nor canals did the constructor provide the transport service (stagecoaches, barges, etc) - the latter were competitive businesses with many players and low entry costs. The investments were for infrastructure *facilities* not infrastructure *industries* like railways, electricity, gas, water, etc and there limited network externalities. Hence, franchise contracts involving a simple version of rate of return regulation was well suited to the road and canal investment issues of the time and could generally be handled by contract, using the courts as a backup if necessary. The period 1780-1820 was, of course, a time of low inflation so that the tolls fixed in nominal terms for 21 years were unlikely to have needed resetting to cover general cost and price rises in the way that they would have done in the twentieth century.

It is no co-incidence that private investment in roads in particular has continued to be based on concession or franchise contracts (eg various European countries and major road construction in the 1990s in Latin America). Regulatory agencies for roads seem to be non-existent – there appears not to be a need for them²⁵. There can, of course, be problems when profits are much higher or lower than expected (eg because traffic forecasts either grossly under-predict or over-predict usage and/or there are major construction cost overruns). There can also be problems from opportunistic behaviour concerned with franchise bidding and re-entering. However, establishing costs and profitability is reasonably straightforward so that variable franchise duration provisions (as on the UK Dartford Bridge contract) can usually handle losses or very high returns - provided sufficient concern by government and contractors over reputation and goodwill²⁶.

For the UK, private investment under franchise contract in roads, bridges and canals died out during the nineteenth century as roads became more of a network and road investment became more of a national responsibility. But, it has returned successfully since the 1980s eg with the Dartford Bridge and subsequent PFI (Private Finance Initiative) and successor schemes²⁷.

The franchise contract model is also useful for other infrastructure *facilities* (eg airport terminals) but failed to provide a reliable or enforceable regulatory framework for privately financed railway services (or later US and other countries' railways) in the nineteenth and early twentieth centuries²⁸. Indeed, UK railways provide an excellent example of the Levy and Spiller contention that state ownership/nationalisation is the fallback regulatory option. This appears even more true after the failure of the 1990s UK railway privatisation of Railtrack²⁹.

²⁵ But see Newbery and Santos (1999) for a suggestion for road pricing by a regulatory agency in the UK.

²⁶ Renegotiation of rates is also possible in extremis. Poor franchise design, tendering, monitoring and enforcement can lead to high levels of renegotiation eg with many of the Latin American toll-road contracts. For further discussion see Guasch, Laffont and Straub (2002) and Stern (2003).

²⁷ Road concession contracts in Western Europe by no means always use private finance. In 1998, the road concession companies were all more than 50% publicly owned in Austria and Norway as were over 85% of French and Italian road concessionaires. See Bousquet and Fayard (2001.)

²⁸ The franchise model has been widely applied under the guise of concession contracts in France, Francophone and other countries for a wide range of infrastructure facilities where the state wishes to retain ownership, at least in the long term eg for water and sewage systems, electricity distribution networks, etc.

²⁹ Network Rail, the successor to Railtrack plc has now been formally re-established as a company limited by guarantee, with a management executive accountable to a large board of stakeholders (over 100). It has not been formally renationalised and its borrowings are not included in the government's balance sheet.

3.2 UK Railway Regulation: The 1844 Act and the Preceding Debate

For railways, the road franchise model via individual parliamentary acts was the original basis for authorising the construction of new railway lines³⁰. It was, however, recognised fairly early that franchises of rather more than 21 years were required to allow investors to recover their outlay given (a) the capital intensity of the operations and (b) the sizeable operating costs.

The other issue was that early observers, notably James Morrison in 1836, recognised that there were significant economies of scale (or at least significant economies of density) in railways so that costs – and prices - should fall as demand rose³¹. We can now recognise the additional scale economies arising from network effects due to the impact of the later growth in interconnection between railway lines. Morrison also realised that cost reductions would be augmented by technical progress in what was a technically very young industry.

In addition to his realisation of the special technical characteristics of railways and their economic importance, Morrison also recognised the incentives to pad costs and indulge in excess investment under dividend limits. (Dividend limits provided an attempt at rate of return regulation but one which could easily be avoided eg by “stock watering” ie issuing more shares at par to founder investors.) Both of these sets of influence led him to consider the promotion in parliament of the establishment of an independent railways regulatory commission with the power of regular review. However, in the end, he suggested review by parliament - not least to avoid arguments about retrospective amendment of contracts.

Standard economic arguments would now normally suggest that these issues be handled by LRMC (long run marginal cost) pricing – at least for the track element - with maximum prices set to generate an agreed expected rate of return. This concept was not available to Morrison who instead suggested regular reviews of final prices (with single not two or multi-part tariffs) on the basis of trends in actual costs and increases in demand. However, modern regulatory economics would still call for periodic and regular price and investment reviews, which Morrison realised were essential.

The first practical attempt at modern utility service regulation was in the 1844 Railways Act. The original bill, promoted by Gladstone, was quite radical over its proposals for price regulation. The main thrust of the bill was to enforce price reviews by parliament as the fallback condition for renewal of rail franchises. These price reviews were to be derived from assets revalued on something approaching MEA (modern equivalent asset) lines. The fallback, if a review was refused or rejected, was reversion of the franchise to the state. But, in practice, this fallback was unworkable as a credible threat since it amounted to franchise cancellation without compensation.

The eventual version of the 1844 act greatly watered down the price review and particularly the asset valuation provisions in the original bill. In addition, the act was seriously flawed because:

However, in practice, it has de facto been effectively renationalised in terms of the framework within which it operates.

³⁰ Common licences only emerged in the 1870s and 1880s, primarily for gas and electricity.

³¹ James Morrison was a Liberal Member of Parliament. In the 1830s and 1840s. He was a self-taught businessman rather than an economist but was close to some economist contemporaries eg McCullough. See Chapter 1 of Foster op cit for a fascinating discussion of Morrison’s thinking and subsequent debates.

- it exempted pre-1844 railways;
- it provided no standard accounting requirements or information disclosure obligations; and
- there were no measures to prevent undue discrimination.

The lack of information disclosure or standard accounting requirements was particularly damaging. There was no obligation on railway companies to publish any prices until the 1870s and this allowed prices to diverge very substantially from costs, not least because of railway companies' abuse of their market power in the railway freight and related markets.

In this context, it is worth noting that the 1983 Littlechild Report on UK telecom regulation explicitly advocated that a prohibition on price discrimination and/or mandatory publication of prices by BT (the dominant incumbent) be included in BT's licence to prevent anti-competitive abuses for consumers as well as potential competitors³². Had such conditions been imposed on UK railways in the 1840s, the performance of the nineteenth century UK railway industry might well have been very different.

In fact, the 1844 Railways Act is usually remembered because it is the first modern example of collective action to enforce quality standards and a universal service obligation (USO). Under the act, railway companies were obliged to run a "Parliamentary train". This was a regularly running train with an average speed of at least 12 mph and with a ticket price for third class passengers of no more than 1 penny per mile. Operators were obliged to run this train on Sundays as well as weekdays and so the Act provided the basis for Sunday excursion trains.

The 1844 Railways Act may have failed as an attempt to introduce effective utility regulation but it, the previous debate and the earlier bill all recognised (to a greater or lesser extent):

- (i) The difficulties in providing major cost reviews and price changes via contract procedures; and
- (ii) The need for collective imposition of common standards via public law - including USOs - where there were many customers, particularly small (household) customers.

Subsequent developments in UK railway regulation attempted to deal with these weaknesses while preserving its operation in the private sector but failed to do so.

3.3 The Emergence and Development of the Railways and Canals Commission

During the second half of the nineteenth century, there was increasing agitation against the high level of railway rates and, particularly from freight users, about the plethora of rates. By 1887, Great Northern railways had 13 million rates and North Western had 20 million and there was widespread price discrimination. Rates were particularly high where railways had merged with (or had no competition from) shipping or canals. In addition, there was widespread growth of costs from excessive investment and padding of capital employed – not least because it was for

³² See Littlechild (1983), para 4.14.

consumers to demonstrate via contract procedures in the courts that rate rises were unjustified. So long as the 10% dividend payout was unbreached, companies were at perfect liberty to raise rates³³.

Various UK mid-century commentators such as Galt pointed out how rail freight and passenger rates were significantly lower (25-50%) in Belgium than in the UK. A major element in this was that UK companies had a variety of incentives to over-invest in line expansion and multiple tracks as well as to increase their physical and financial capital base. One of the main differences between Belgian and British rates was that construction costs seemed to be significantly higher in Britain. The dividend limitation scheme – which acted as a feeble rate of return regulatory tool – was a significant contributory factor to the bias towards excess investment volumes and costs in the mid-nineteenth century era of growing demand and high railway investment returns. In consequence, Galt and others argued for nationalisation of the railways with lower fares on the model of the Post Office. Many others (eg Jevons) called for more rigorous regulation³⁴.

The response to these arguments and continued pressures from freight users was the establishment in 1873 of the Railways and Canals Commission. This ‘railway regulatory agency’ persisted in various guises until 1956.

It was decided to institute a commission because:

- Parliament could not provide a review body with continuity - and the threat of review by Parliament under the 1844 Act had demonstrably failed;
- the Board of Trade (the relevant Ministry) was rejected because it had insufficient judicial character; and
- the courts were rejected as they were deemed to have insufficient practical knowledge or administrative facility.

The 1873 commission did not resolve the regulatory problems because:

- (a) there were still very limited obligations on companies over publishing rates and no common accountancy requirements or other informational obligations;
- (b) the onus of proof remained on consumers to demonstrate why general or specific rate increases should not be allowed; and
- (c) the commission (and its successors) only considered rates on individual lines not averages or indices of rates.

The failure of the 1873 arrangements led to a beefing up of the Commission so that, in 1887, the Commission was uprated to have the authority of a High Court. It was presided over by a judge who had two lay assessors, while cases were argued by barristers using standard High Court legal procedures and rules of evidence and with verbatim records of the proceedings. This was the closest the UK has come to adopting what is thought of as the US model of regulation.

³³ For further details, see Foster op cit chapter 2.

³⁴ See Foreman-Peck and Milward op cit pp 21-23 for a discussion and references.

However, it was also a failure because none of the underlying problems had been resolved while, in addition, the costs of bringing a case were now very high.

These failures led 1894 to the reversal of the onus of proof for rate rises so that, after 1894, the railway companies had to demonstrate to the Commission why they should be able to raise rates. This looks much more like a modern regulatory regime – except that it still applied to individual rates. Unfortunately, the impact was to ossify rates and to increase the incentives for collusion and merger of rail companies. Rates ossified because:

- (i) it was difficult and expensive to prove the case for rate increases;
- (ii) as modern road transport developed from 1900 (and particularly after 1920), the price elasticity of demand for rail rose sharply so that railway companies were worried that revenues would not increase from rate increases, particularly on freight; and
- (iii) temporary reductions in rates were effectively made impossible following a 1907 appeals case in the High Court³⁵.

The 1894 act made it significantly harder to raise railway rates just at the point where railway operating costs started rising sharply rather than falling as they had done until the 1890s.

The 1894 Act also, however, seriously advanced the process by which railway companies had significant costs imposed on them which they were not allowed to recover by rate increases. This started in 1894 with the Railways Inspectorate being given the power to impose safety obligations at the expense of the companies. From 1894-1914, the companies had progressive cost increases imposed on them over

- safety standards
- wages and conditions of work
- service obligations such as “workmen’s trains”.

In consequence, by the 1920s and later, the railway companies were trapped in a costly and ineffective rate of return regulatory system as well as strong safety and USO obligations, the costs of which could not be recovered in prices. This was on top of having been effectively bankrupted by the failure to raise rates during the inflationary burst of World War 1.

3.4 The Economic Lessons from Nineteenth Century UK Railway Regulation

The system created for UK regulation in the late nineteenth century represented a serious attempt at creating a modern regulatory regime – but the outcome seems to resemble the worst nightmares

³⁵ See Cain (1978), p.89.

of Chicago economist critics of regulation. Not surprisingly, particularly with the continued development of lower cost road freight and passenger transport, this classic regulatory failure ended up with the nationalisation of the railways in 1947.

The point to emphasise, however, is that all the attempts at establishing regulatory agencies for the UK railways in the nineteenth century were failures of regulatory design. Their designers recognised the inherent problems of relying solely on long duration franchise contracts and private law remedies for railway regulation. In particular, they recognised the inability to review and revise the prices in the initial contracts or to impose general conditions without the assistance of an external independent regulatory body.

It is not (pace the Chicago school) that contracts and the private law provisions would have done better – for most of the period, that was the system effectively in place. Rather, it is because the regulatory regime could and should have been better designed. In particular, the UK government should have been more thorough-going over imposing proper and regular regulatory review by an independent agency rather than making the concessions it did to the supposed sanctity of contracts. Pace Littlechild 1983, this would have required proper information disclosure on prices and a prohibition on undue discrimination and other predatory practices. Ironically, imposing more thorough-going information disclosure and regulatory review might well have allowed the survival of contracts eg between railway companies and large freight customers.

A major incentive for the earliest railways was the construction of lower cost alternatives to compete with canals eg the 1830 Liverpool to Manchester railway competed against the Duke of Bridgewater's canals. Hence, in the absence of price regulation, railway companies set prices against what the market could bear ie the price of horse-drawn stagecoaches, carts, canal barges, etc. But, railway costs were much lower so that such prices allowed significant profits. Hence, by the 1850s, the only real threat to prices was the threat of new entry from competing lines but this threat was low, particularly as the system matured. This was primarily because the incentives for excess track expansion investment meant that, in most cases, new entrants faced the threat of a rates war that they would almost certainly lose – particularly when it was so difficult to obtain price comparisons. Hence revenue sharing, price collusion and merger dominated over competition. Where competition did exist, (eg London to Liverpool from steam ships), freight rates were lower, but competition was the exception rather than the rule, particularly when the railway companies vertically integrated through the purchase of canal and dock companies.

If the weak price comparisons and an absence of effective price regulation allowed a bloated, high-cost and high-priced UK railway system in the nineteenth century, it also hastened its decline in the twentieth. When costs started rising sharply and inflation became more prevalent, there was no mechanism by which the higher wage and other costs could be passed through into higher rates so that railway companies share prices fell by 30 per cent between 1896 and 1911³⁶. But, after 1900, there was growing and real competition – not from within the railways but from lorries, cars, buses, etc. Aside from any difference in operating costs, the absence of road pricing meant that railways found it increasingly difficult to compete because well over 50% of their costs were track-related.

In consequence, after 1920, railway companies found it virtually impossible to raise rates without lowering total receipts because of demand responses. The railway companies and shareholders that prospered in the nineteenth century because of cost advantages and an absence of regulation

³⁶ See Foreman-Peck and Millward op cit, p 82.

became progressively and rapidly squeezed in the twentieth century because of the absence of any regulatory agency to relate prices to costs as competition from other transport modes intensified. This profit squeeze was worsened by the problems of an over-extended and high-cost network.

Foster concludes that the real missed opportunity in UK railway regulation was, firstly, rejecting an independent regulatory commission (as originally proposed by Morrison in 1836) in favour of parliamentary review; and, secondly, the watering down of the 1843 bill by parliament³⁷. In hindsight, it is difficult to disagree even though it is far from clear whether it was politically feasible at the time or whether, in practice, it would have worked like a modern regulatory agency.

Curiously, the Transport Tribunal, the successor to the Railway Rates Tribunal and the Railways and Canals Commission, continued in existence after nationalisation and was only abolished in 1956. As such, the UK regulatory commission for railways must be, in its various guises, one of the longest-lived specialist regulatory agencies. Its final rate hearings in 1955-56 had 44 days of hearings and took over 1,000 pages of evidence³⁸. It is (understandably) an almost forgotten body but it was an important early model for a utility services regulatory agency - and clearly the closest that the UK has come to the US regulatory commission model, based on formal legal procedures.

³⁷ See Foster *op cit.*, p 34-35.

³⁸ See Foster *op cit.*, p. 59.

4. Contracts and Regulation in Other UK Utility Service Industries in the Nineteenth Century

The deficiencies in nineteenth century UK railway regulatory arrangements were reflected in the arrangements for other utility service industries. For them, largely unsuccessful attempts were made to introduce regulatory elements into a contractual framework. In this section, we will discuss (i) household rubbish collection and (ii) the town gas industry.

4.1 Household Rubbish Collection in Nineteenth Century London

The largest part of mid-nineteenth century household rubbish collection was the residue of coal-burning fires, stoves etc – dust and ashes, hence “dustmen” as the British term conventionally used for rubbish (garbage) collectors. In 1861, all but one of the London Vestries used private contractors to collect domestic refuse but this was progressively replaced by Vestries doing their own collection - ie municipalising the service³⁹. The need to do so again primarily arose from the inability in contract law to impose common standards for services with large numbers of customers.

The fundamental problems were:

- (i) there was no effective institutional mechanism to enforce service obligations to households in the contracts between the Vestries and the contractors; and
- (ii) to win contracts, contractors had an incentive to reduce wages to minimise costs. This led to growing reliance by the dustmen on the receipt of “voluntary” tips from households, which further reduced the quality of service.

A good example of the regulatory failure that drove local authorities to take over domestic refuse collection from contractors is given in the following, splendidly orotund, quotation from the Islington Medical Officer of Health in 1897, reported by Turvey as follows:

“I cannot but express my decided opinion that in every respect, and especially as regards the almost fabulous number of applications [ie complaints to the local authority about the quality of service], the Contractors exhibited a total disregard of the obligations involved in their contracts, which were, during the whole period most obviously, and even scandalously, disregarded. One would think that the utmost bounds of the consciences of these contractors must have been reached in contemplation of the provoking and persistent negligence they were displaying towards the inhabitants.”⁴⁰

4.2 Gas: The Sliding Scale

The same regulatory problems affected the emerging UK gas (town gas) and later the electricity industries in the late nineteenth century as affected the railways. They were handled better but not sufficiently so as to maintain the viability of private finance over municipalisation or, eventually, nationalisation⁴¹.

³⁹ See Turvey (1996) for a fascinating discussion on which this section is largely drawn.

⁴⁰ See Turvey op cit, p.3.

⁴¹ For these industries, nationalisation was largely due to the need for central co-ordination to achieve economies of scale and construct the networks that would enable the establishment of modern, technically

The key differences over the railway regulatory framework were:

- (i) general franchise terms could be granted from 1875 under the Gas and Water Facilities Act;
- (ii) the sliding scale, a precursor of RPI-X, attempted to link price rises to cost reductions within a dividend limitation framework; and
- (iii) share capital increases had to be sold by tender or auction rather than by the issue shares to existing shareholders at par.

The first known example of a UK sliding scale arrangement was in 1866 for the gas supply in Sheffield.⁴² The way the sliding scale worked was that dividend increases were only allowed if charges to consumers were reduced relative to a prescribed standard price; conversely, increases in prices required a fall in dividends. But, this was still done via contract clauses written into the authorising statutes so that there was no local or national agency to monitor or enforce the arrangements. The adoption of a sliding scale regime for franchises was also voluntary, although there was considerable and effective pressure on companies to include them as the basis of their franchise.

In economic terms, the sliding scale had the following defects:

- The terms of the sliding scale did not bite for young industries experiencing rapid technical progress – this applied to the gas industry in the late nineteenth century and electricity after 1900;
- Limiting price increases by dividend limitation became irrelevant as competition from electricity set limits on what gas companies could charge without losing customers; and
- The sliding scale was set in nominal terms not in real terms so that the scheme could not handle inflationary circumstances.

In addition, there was still no process of regulatory review of the standard price. Although the reduction in costs had slowed down after around 1875, cost reductions continued, albeit at a slower rate. This led to pressures for reductions in standard rates which were imposed by parliament on 28 companies between 1898 and 1915 so that the sliding scales were (in the words of Pigou) “... not substitutes for but complements to a system of periodic revision of the licence terms....”⁴³

These problems (together with the fact that they were embedded in franchise contracts, the basic terms of which were difficult to rewrite) meant that the sliding scale died rather than evolved into something that could have provided a better basis for price regulation. The rise of more powerful

efficient electricity and (natural) gas industries viz. the establishment of a state-owned electricity transmission network in 1926.

⁴² See Foster op cit pp 61-2 and Turvey (1995) for further discussion

⁴³ Pigou as quoted in Turvey (1995), p.5. Turvey’s paper contains a rich and fascinating discussion of the sliding scale, its economic incentive effects and its relationship to modern RPI-X style price cap regulation

municipal agencies couple with the problems over sliding-scale price regulation led to considerable municipalisation of gas sales in Britain particularly in the 1870s and from 1890-1910⁴⁴. (The same was true for water and later for electricity).

Some municipalities set prices as low as possible to benefit their consumers but most set them on a commercial basis to gain revenues for subsidising other activities and providing local public goods. This was particularly true for towns with a weak revenue base and high population growth⁴⁵. Between 1851 and 1881 nearly one-third of new gas companies were municipally owned and, by 1905, 33%, of net output of gas companies was supplied by publicly owned (municipal) companies. This trend was fostered by the legal right of municipalities to buy out any network utility after 21 years.

Nevertheless, municipalisation was far from universal as, in 1939, 60% of gas undertakings were privately owned⁴⁶. For gas, nationalisation in 1949 was not primarily a failure of price regulation. Rather, it was a rationalisation of the patchwork of local monopoly franchises of varying kinds of ownership to introduce standardisation and boost efficiency. However, to the extent that efficiency and productivity gains are the outputs of effective regulation by agency, it can still be argued that UK gas nationalisation was the response to previous regulatory failures.

⁴⁴ See Foreman-Peck and Millward, *op cit*, p 172-3.

⁴⁵ See Foreman-Peck and Millward, *op cit*, p 180-191. Cf the German Stadtwerke where profits from electricity and gas distribution and (until recently) retail supply margins on long-term contracts with large wholesale energy suppliers are used to subsidise urban transport and other services

⁴⁶ See Foreman-Peck and Millward, *op cit*, Table 5.1.

5. The Early Twentieth Century: Electricity

Electricity is the classic example of an industry with extremely rapid technological progress, but which was dependent for its expansion on the roll-out of monopoly network facilities; firstly for local low voltage distribution and later for high voltage transmission. The early history of UK electricity shows the challenges of developing the industry via franchise contracts in the absence of a regulator – as well as the potential risks from premature, mistaken regulatory choices.

5.1 The Initial Development of the UK Electricity Industry 1870-1920

Electricity was first produced and marketed for lighting. It emerged as a (high cost and low reliability) competitor to gas in the 1870s. Indeed, the first parliamentary act was the 1882 Electric *Lighting* Act and many of the early electricity companies were named electricity lighting companies. Electricity was initially used almost entirely for lighting – commercial and public lighting as well as some limited sales to richer households in dense conurbations. This was the position until after 1895 with first traction (mainly urban tramways) and then, after 1900, industrial power uses becoming the main growth areas. Only in 1909 did UK industrial power sales exceed lighting sales.⁴⁷

In 1890, most generating stations had only 0.1 MW of capacity and the price of electricity was over 7d per kWh. By 1912, the average size of new generators was 2 MW and average price was 3.6d per kWh, with lower prices for less peaked industrial users relative to lighting users. For instance, by 1894, the Manchester municipal electricity company was selling power to industrial users at 1.5d per kWh. The improvements in the cost and price of electricity were accompanied by (and to some extent caused by) the rapid development of electricity using equipment such as DC motors for traction and AC polyphase motors for industrial machinery. The combination of technical progress in electricity supplying and electricity using equipment allowed an increase in generation load factors from around 10% in the early 1890s to 23% in 1907⁴⁸.

These trends led to growth in total annual UK electricity sales from 38 GWh in 1895 to 180 GWh in 1900, 645 GWh in 1905 and 1,975 GWh in 1913. Between 1895 and 1913, the sales per head of population rose from 1kWh to 42 kWh. The share of industrial use rose from zero in 1895 to 18% in 1905 and 51% in 1913⁴⁹. This was truly a explosive rate of growth but one that was very largely based on bottom-up expansion of local, vertically integrated generation and distribution companies with little or no interconnection between municipalities (unlike Germany and the US where larger regional integrated supply areas developed earlier)⁵⁰.

5.2 Regulation and Franchise Contracts in the Early UK Electricity Industry

The regulatory problems that emerged in electricity supply were not so much the cost and price revelation and review problems of the nineteenth century railway industry but issues of technical standardisation (eg the choice between AC and DC as well as the choice of AC frequency),

⁴⁷ See Hannah (1979) Chapter 1 and Table A.1. Hannah's statistics and discussion for the period up to 1914 derive primarily from Byatt (1978), which puts into book form an unpublished D. Phil thesis of 1962.

⁴⁸ See Hannah Chapter 1 and Statistical Appendix pp 426 -33 which quotes estimates taken from Byatt.

⁴⁹ See Byatt op cit, pp 98.

⁵⁰ NESCo (the Newcastle-upon-Tyne Electric Supply Company) was the notable exception and we discuss this below.

transmission development, interconnection, wholesale trade as well as scale economies and mergers. These were only addressed after 1918 with the establishment of the Electricity Commissioners in 1919 and (much more significantly) with the establishment of the state-owned Central Electricity Board (CEB) in 1926 which led to the development of the National Grid.

The 1882 Electricity Lighting Act allowed companies (private and municipal) to establish electricity companies and gave them the powers to break up streets to do so. It also required maximum prices to be set, but there was much less use of the sliding scale than in gas. A major feature of the 1882 Act was that it allowed municipalities the right to buy out privately owned franchises after 21 years at the written-down value of the capital. The original proposal was for a franchise period of 15 years but the House of Lords amended this to 21 years. (21 years was the standard late eighteenth century franchise length for toll roads and canals as well as that initially proposed by Morrison and Gladstone for railways in 1844.) In 1888, a subsequent Act increased the franchise period to 42 years.

During the 1890s, there was a wave of municipalisation with many towns either buying out private companies (often at favourable prices relative to book value) or starting their own companies. By 1903, municipalities accounted for over two-thirds of all connections to public supply mains and, among large towns, only London, Newcastle, Preston, Bournemouth and Northampton were supplied by private companies⁵¹. Municipal companies enabled the reaching of the first wave of economies to scale in generation as well as in supply eg via sales to municipally owned tramways and to local industry which increased load factors.

By 1917 there were 600 licensed electricity undertakings in the UK. The Haldane-Merz Committee's recommendations to combine these into 6 regional companies was vigorously opposed by the municipalities and defeated as was the proposal to establish District Electricity Boards which would take over all existing generation and sell bulk power at a standard price to local distribution and supply companies⁵².

In general, the municipally owned companies were operated on a commercial basis. They were profitable and both their costs and prices were comparable to those of the private companies⁵³. Municipal companies do not seem to have pursued major cross-subsidisation of local consumers, although, for 1921-22, the average revenue per unit of electricity sold by municipal companies to lighting and domestic customers was around 16% lower than for privately owned companies while the average revenue per unit sold to industrial and other customers was about 20% higher⁵⁴. Some of this difference may, however, be due to compositional effects and regional differences. Municipal companies also had to obtain permission (originally from the Board of Trade, after 1919 from the Electricity Commissioners) for new loan issues.

In consequence, problems over price-cost alignment and review were much less severe in electricity in its early development stage than for railways because of:

- Limited franchise periods with effective municipal buyout options;
- Open and published maximum price clauses;

⁵¹ See Hannah op cit p.22.

⁵² See Hannah op cit, pp 63-70.

⁵³ See Foreman-Peck and Millward, Chapters 5 and 6.

⁵⁴ See Foreman-Peck and Millward, Table 6.2.

- The ability to compare prices and other aspects of performance with other companies (public and private) – implicit yardstick regulation; and
- The need for municipalities to justify additional borrowing for investment to a central agency.

The regulatory framework did not, however, impose strong pressure on costs because the companies all operated local monopoly franchises. For private firms as well as for the municipals, the incentives were to retain their profitable local monopoly, prevent new entry, resist mergers and avoid or limit bulk power purchases from outside their area. The incentives were to resist network interconnection between areas. The result was to make the development of regional or other larger companies much more difficult, as was found in the vigorous resistance to the consolidation proposals of the 1917 Haldane-Merz and Williamson Committees and the 1919 Electricity (Supply) Bill.

It was not impossible to establish regional companies as shown by the development of the privately owned NESCo (Newcastle-upon-Tyne Electric Supply Company) which operated the largest integrated power system in Europe pre-1914. NESCo developed the use of high voltage transmission (20,000 volts) to support sales to their own distribution companies, which covered 1,400 square miles of Tyneside and Teesside, as well as bulk power sales to other distribution companies (eg Middlesborough and Tynemouth). The benefits of the economies of scale are shown by the fact that the average price of NESCo power in 1905 was 1.03d/kWh as opposed to around 4d/kWh nationally. This resulted in major consumer benefits, not least to new electricity-using industries such as the chemical plants which were attracted to the area⁵⁵.

However, NESCo was the exception and there was no effective regulatory pressure on companies in other areas to take similar actions to reduce their costs. In general, after 1905, the (42 year) franchise areas based on municipal areas were increasingly too small to support the new, larger optimal scale generating stations and supply areas. Transmission was under-developed until the centralised CEB was put in place in 1926 to fill the gap. Franchised generation and distribution companies, operating under the protection of their franchise contracts, remained as powerful local monopolies able to resist imposed change from outside. This was true not just before 1914 but remained a major problem until the establishment of the CEB in 1926 and a significant problem from then up to nationalisation in 1947.

The presence of these powerful UK local franchise holders able to resist the development of wider, integrated regional markets as in the US and in Germany is well known. However, regulatory design factors may be important in this.

The UK electricity franchise contracts did not allow for renegotiation of maximum prices. Pre-1914, low inflation and rapid technical progress meant that there was no need for price increases. After 1919, franchise holders could appeal to the Electricity Commissioners for price increases to reflect higher costs but all other terms of the franchise contract were left unaffected. In general, the Commissioners' powers were heavily qualified and they had no authority to make price increases or other approvals (eg sanction for new generating plant) conditional on other actions.

⁵⁵ See Hannah op cit, pp 28-33.

In contrast to the UK position where franchisee's rights were heavily protected, US franchise contracts often made *explicit provision for renegotiation in response to changes in circumstances*, subject to arbitration or reference to an independent committee. These independent committees could take the responsibility for monitoring service quality. These arbitration or review committees gradually evolved into state Public Utility Commissions with substantial power to extract concessions from the utility as a condition of maintaining their franchise without competitive entry⁵⁶.

The absence of any explicit contract review powers in the UK meant that there was no equivalent provision for explicit franchise review to change the requirements on franchise holders or even to trade changes in return for concessions eg on price rises. The (bounded) wide US franchise review possibilities led to the development of a regulatory framework that supported continued private investment. Conversely, the absence of such explicit regulatory review powers in the UK meant that weak pressure on costs and efficiency led to the 1948 nationalisation as a centralised way of obtaining the benefits of the economies of scale and scope that more effective regulation might have achieved in the 1920s and '30s.

5.3 Standardisation and the Choice of Voltage and Frequency

One of the most interesting regulatory challenges is the question of network standardisation. This had been an issue with railways (eg on choice of track width) whereas, it was not a major issue for town gas and other locally supplied services which did not require interconnection between areas. However, for electricity it was crucial and complex.

When and how to impose common network standards is a critical issue, which typically requires regulatory intervention even when largely left to the private sector (viz US telephone industry). For the UK, standardisation came relatively late and slowly. This again seems to reflect the weakness of regulatory arrangements to impose (or renegotiate) conditions on local franchise companies.

The first issue was the choice between AC and DC systems. Many of the “lighting only” electricity companies established in the early 1900s were DC only. Although AC increasingly replaced DC as other electricity was used for other purposes, as late as 1937/8 there were still 1.1 million DC customers, only 36% were supplied at the standard voltage.⁵⁷

The second issue was the choice of voltage at which supply was received. Pre-1914, this varied widely with companies supplying at or around 200 or at 240/250 volts (as opposed to the US standard of 110/115). As late as 1926-27 – seven years after the establishment of the Electricity Commissioners – there was still huge variation. There was a standard voltage of 230 but only 23% received power at that voltage. For the rest, 32% were supplied in the range 220-240 volts, 37% at 210-215 volts, 4% at under 200 volts and another 4% at over 240 volts. As late as 1937/8, only about half of consumers were on the standard voltage.

⁵⁶ See Newbery (1999), p. 23 for a description of the evolution of US municipal regulation.

⁵⁷ See Hannah op cit pp 196-97 and also for subsequent statistics reported in this section.

The delays in achieving voltage standardisation led to significant additional costs both for the electricity supply industry (via equipment costs and the need for extra equipment) as well as for manufacturers of electricity-using equipment. Indeed, these extra costs must have slowed the rate at which appliances were installed for households. As late as 1939, only two-thirds of homes had electric service and many of these had a lighting-only service and many had only one plug socket. The ownership of appliances was surprisingly low with only around 45% of homes having an electric iron, 25% a vacuum cleaner, under 20% an electric fire and 10% or less having an electric kettle or cooker or wash boiler/heater. Regulatory action to enforce earlier standardisation (eg in the 1920s) would surely have provided major efficiency benefits to industry and major direct and indirect consumer benefits⁵⁸.

However, there are regulatory costs from standardising too early and this is shown in the history of frequency regulation⁵⁹. Pre-1914, most electricity was transmitted at a 50 Hz frequency, but the NESCo area used a 40 Hz frequency. In 1924, there were still 17 different frequencies in use. 50 HZ was used in most industrial areas (and was adopted as the standard) but 40 Hz was still used throughout the North-East and around Manchester, 85 and 100 HZ were used in London and some of the south coast and 25 HZ in South Wales, the Midlands and Clydeside.

Standardisation on a single frequency was essential for interconnection and, in particular, for the establishment of a transmission grid connecting all areas to allow gains from trade and economies of scale. Again, frequency standardisation was critical for reducing manufacturing and investment costs both of electricity supply industry equipment but also electricity using equipment.

NESCo's early decision to standardise at 40 Hz turned out to be premature. An early regulatory decision to standardise on that would probably have locked the country into an inefficient standard – a period of experimentation and competing standards may well have been useful. However, by 1917 and the Merz-Haldane Committee, it was clear that action to enforce standardisation was necessary. This was only achieved in the 1930s following the establishment of the CEB and its collaboration with the Electricity Commissioners on the construction of a national Grid.

Significantly, there was early standardisation of the voltage to be used by the National Grid at 132 Kv. The CEB and the Electricity Commissioners made the decision to standardise on this right from its initial construction in the 1930s. This is in striking contrast to the problems of achieving standardisation of distribution level voltage and frequency – or even of AC or DC – at regional and local level. Standardisation of the others was only achieved after nationalisation. The delay represents a major regulatory failure arising from the inability of the electricity regulatory agency to intervene in or renegotiate franchise contracts.

⁵⁸ See Hannah op cit, p. 208.

⁵⁹ See Hannah op cit pp 38-9 and 88-9.

6. Contracts and Regulation in UK Utility Service Industry History: An Economic Interpretation

The UK relied very heavily on franchise contracts for the development of railways, town gas, local services and electricity (but not for telephony which was nationalised over the period 1905—12). The discussion of these franchise contracts makes it clear that they were very far from complete contracts. In all cases, we are in a world of *highly incomplete contracts*, not least because of unforecastable technical progress and cost reductions during the life of the franchise contracts.

6.1 Incomplete Contracts and Franchise/Concession Contracts

This is important since economic theory as well as economic experience suggests that the need for renegotiation mechanisms derives crucially from contract incompleteness. The greater the degree of contract incompleteness, the greater is the likelihood that there can be Pareto improving negotiations. For infrastructure service industries, this effectively means the use of explicit regulatory agencies to allow and facilitate contained and bounded renegotiation in the light of new information⁶⁰. This is critical for industries with high investment costs in sunk assets, particularly when there is good reasons to expect the arrival of unforecastable new information on costs, efficiency growth, economies of scale and scope, etc.

Recent empirical modelling by Guasch, Laffont and Straub (GLS) based on incomplete contract theory shows that the probability of (fundamental) contract renegotiation was significantly reduced for 1990s Latin American private investment concession contracts in roads and water⁶¹. The existence of a regulatory agency significantly reduced the probability of renegotiation by:

- (i) leading to better designed contracts (eg clauses adapted to the type and circumstances of the specific concession);
- (ii) improving the quality of enforcement⁶².

In the discussion of the results, GLS also point to the role of an external regulator in

- Improving monitoring of the contracts as well as commitment of both sides;
- Enhancing “learning by doing” in contract design, enforcement and operation; and
- Encouraging simpler, more transparent contracts with fewer objectives (and trade-offs between objectives).

Indeed, GLS end their paper by pointing to the role of a regulatory agency in allowing *more incompleteness* in infrastructure concession contracts while maintaining a low risk of contract

⁶⁰ See Guasch, Laffont and Straub (2002), pp 6-7 and their references to previous literature on incomplete contracts by Hart and others.

⁶¹ Concession contracts are equivalent to the franchise contracts of UK nineteenth and early twentieth century experience discussed earlier

⁶² See Guasch, Laffont and Straub, op cit pp 21-26.

failure. They suggest that this can best be achieved via a pre-existing legal framework set out in primary law, including explicit regulation by a pre-existing agency, to support short, concession-specific contracts⁶³. This last is supported by the result (reported in Guasch, 2001) that the probability of fundamental concession renegotiation was 61% where the regulatory framework was specified in the concession contract but only 18% where the regulatory framework was specified in a law. For water concessions the ratio was 92% to 23%⁶⁴.

In other words, the GLS analysis and results point strongly towards *the complementarity between contracts and regulation rather than their substitutability*. The existence of an external regulatory agency appears to make franchise/concession contracts significantly easier to design, write, negotiate, monitor, enforce and revise, even in circumstances of relatively low and predictable technical progress (1990s road and water investment and operations). This leads to the conclusion that ***for infrastructure industries, the existence of an external regulatory agency enables the use of better contracts and results in superior outcomes for private investors and consumers.***

Of course, this benevolent result is far from guaranteed. It assumes that the regulation in question is well-designed, has effective legal backing, has good governance, is properly and predictably enforced, etc. But, regulatory failure is far from unknown either in nineteenth century Britain or in developing and developed countries more recently. Establishing and sustaining effective utility service regulatory agencies can be very difficult. Hence, although a regulatory agency *is potentially able* to provide a reduction in transaction costs that *allows* for superior outcomes than relying on contracts alone, there is absolutely no guarantee that this will happen in practice in any country.

It is clear from nineteenth century British and more recent experience in other countries that regulatory failure may fail to improve the contract-alone outcome for utility/infrastructure services in sustaining private investment. However, the evidence is also that utility service franchise contracts in the absence of a regulator are frequently not sustainable in the longer run (eg collapsing to state-ownership) or result in inferior outcomes (nineteenth century UK railways). It is therefore difficult to argue that having a regulatory agency to supplement contracts is likely to worsen the prospects of success for private investment. It may fail to improve them, but the failure rate of concession contracts in the absence of an explicit external regulator is so high as to be highly unlikely as to worsen the prospects for viable private investment.

We have demonstrated in previous sections the failure of the UK franchise contracts in the nineteenth century either in the absence of regulation or with weak regulation that failed to provide for effective review and change. Similarly, of the Latin American concessions examined in GLS, over 40% had been renegotiated by 2000 – three-fifths within the first three years of a 15-30 year concession. For water concessions, the average time before renegotiation was only 1.7 years.

⁶³ See Guasch, Laffont and Straub, op cit p 26.

⁶⁴ See Guasch (2001), pp 25-27.

6.2 Incomplete Contracts and UK Infrastructure Industry History

The perspective given by incomplete contract theory is also highly relevant for nineteenth century Britain.

Governments, legislators and company managers were operating with some new industries whose likely development was very unclear. There was also a much more limited understanding of the issues of natural monopoly as well as of contract and regulatory design. The design of simple toll road and canal franchises in the non-inflationary world of the late eighteenth century provided little preparation for the problems created by the arrival of the railway age.

For railways, as the first modern infrastructure industry, it must have been extremely hard (if not impossible) in the 1830s and '40s to foresee the magnitude and duration of technical progress, scale economies and network externalities and their impact on railway costs. It must have been similarly extremely hard if not impossible to foresee the progressive decline in the rate of growth in railway TFP (total factor productivity) after 1870, or the (probable) negative rate of TFP growth after 1900⁶⁵.

In these circumstances, any franchise contracts were likely to be seriously incomplete. This suggests that an effective external regulatory agency was highly necessary to supplement the rail franchise contracts eg to provide for regulatory review as information on cost trends became more available - as Morrison and a few others intuitively realised at the time. In addition, an incomplete information perspective is useful to examine the damage caused by the failure to enforce information disclosure on the railway companies.

It is extraordinary that there was no obligation on nineteenth century railway companies to publish their prices. Even when the Railways and Canal Commission was established in 1873, there was very limited requirement on railways to publish their rates. As late as 1912, it was impossible to establish the average rate for freight or the average fare for passengers. In consequence, the railways had enormous power to exercise price discrimination, particularly relative to freight users. By 1887, there were 13 million different rates on the Great Northern Line and 20 million on the London and North Western⁶⁶. This was hugely destructive to the effective working of the market.

By the 1870s, collusion between railway companies coupled with vertical integration between railway companies and canal, port and shipping companies meant that the railway companies had great bargaining power relative to users. Rates could be tailored to maximise profits according to the degree of competition from competing lines (if any) or other transport modes. Rates could be tailored by region, by distance, by type of goods carried - or by almost anything. The absence of price transparency was a significant impediment to voluntary contracting by railway service users and providers and helped ensure that regulation had to be more intrusive when it did develop.

The effective limit on railway prices was set by what could be charged by horse-drawn carts and coaches (adjusted for the costs of time). In the nineteenth century, that allowed the railway companies to earn very high returns which were hardly affected by the dividend limitation rules. In the twentieth century, the pricing limit was set by competing road transport. The progressively

⁶⁵ See Foreman-Peck and Millward op cit, p.88.

⁶⁶ See Foster op cit, p 46 and pp 47-51.

lower costs of lorries and cars (together with the absence of road pricing) meant that the railways were unable to compete.

The lack of price publication was, not surprisingly, accompanied by an absence of reliable comparable information on costs. There was little effective legislation to enforce common accounting standards on railway companies so that it was impossible either to measure the rate of return on capital or the capital base. Given the degree of monopoly power of the railway companies, not surprisingly, this resulted in the late nineteenth and early twentieth centuries in:

- (i) very high construction costs relative to Belgium and other countries;
- (ii) over-extension of the network (which was to cause major problems in the twentieth century);
- (iii) declining productivity growth; and
- (iv) declining profitability.

It is, unfortunately, not possible to compare the construct the counter-factual of a nineteenth century UK railway industry with a good quality twentieth (or twenty-first) century standard regulatory agency – and, of course, there are major questions about the feasibility of creating and sustaining such an agency in mid-nineteenth century Britain. Nevertheless, it is difficult to believe that the results in terms of network size and design, costs, quality, etc would not only have been far superior at the time but also would have left a far superior legacy.

On the basis of experience in other countries (and given the relative size of track costs and the absence of road pricing), it is unclear whether UK railway nationalisation could have been avoided, but there should have been significant economic efficiency gains if only a properly defined regulatory process had been available with which

- (i) to obtain and publish comparable information on costs and prices; and
- (ii) to review and revise the franchise contracts.

For electricity 50 years later, some of the same issues arose. In particular, there was the issue of trying to devise appropriate long-run contracts for a highly capital intensive industry whose technical development (and rates of innovation and technical progress) were so hard to predict. The range of uses of electricity could not have been predicted from the high-cost lighting only businesses of the 1870s and '80s. In addition, there were the difficult issues of when and how to standardise electricity system on several dimensions (AC or DC), voltage and frequency.

The franchise contracts for local suppliers failed to deal with these issues – in fact, the effective resistance of local (mainly municipal) franchisees to changes of terms worsened the problem. Again, the inability of the UK government to establish a regulatory agency with powers to intervene with the terms of the original franchise contracts caused significant economic costs to be incurred. The original contracts were highly incomplete and became progressively less satisfactory but there was no market-driven or regulatory method available with which to improve them. Nationalisation was the only way of resolving the dilemma. Indeed, nationalisation to provide standardisation and integrated development may well have been (in this case and in others) the necessary precursor to the 1980s privatisation with competition.

The contrast between the pre-1939 lower voltage supply and distribution stand-off and the development of the National Grid is stark. After 1926, a single entity working closely (eg on loan approvals) with the Electricity Commissioners was able to construct quickly and at reasonable cost a high voltage network that significantly increased benefits from trade and allowed the effective utilisation of scale economies. The CEB was a public sector body but private investment plus effective regulatory oversight coupled with simple, revisable contracts might well have been as effective.

7. Concluding Comments

The understanding of the failures - and limited successes – of nineteenth and early twentieth century UK utility regulation are important not just in their own right but also for major current policy issues. Questions of uncertainty and how to design a framework for private sector investment and/or finance of major infrastructure or utility service projects remain difficult and are highly controversial today. This is true both of developed countries (eg the London Underground PPPs) and for developing and transition economies (electricity and other utility service industry reforms).

The nature of the uncertainties may have changed from the technological uncertainty of the early railway and electricity industries to issues more related to economic and institutional uncertainty. However, the underlying problem is the same – how to provide an effective framework for capital-intensive industries with sunk assets, no second-hand markets and powerful incentives for strategic behaviour by governments and companies. To handle these issues, we suggest that the best perspective is an incomplete contracts approach, which recognises that regulation by external agency can provide a beneficial supplement to whatever contract framework is appropriate.

Of course, not all regulatory agencies can or will meet this standard. It requires transparent and accountable regulation, with bounded discretion, which the UK failed to establish in the nineteenth and early twentieth centuries and which many countries find it hard to establish and sustain today.

These current-day issues, particularly as they affect developing and transition economies, will be discussed in a forthcoming companion paper.

March 2003.

References

- Beale, H. and Dugdale, T. 1975. *Contracts between Businessmen: Planning and the Use of Contractual Remedies*. *British Journal of Law and Society* 2.
- Bousquet, F and Fayard, A. 2001. *Road Infrastructure Concession Practice in Europe*
- Brunekreeft, G. 2002. *Regulation and Competition Policy in the Electricity Market: Economic Analysis and German Experience*. Mimeo. Department of Applied Economics, University of Cambridge.
- Byatt, I.C.R. 1979. *The British Electrical Industry 1875-1914*. Oxford. The Clarendon Press.
- Cain, P.J. 1978. *The British Railway Rates Problem*. *Business History*, Vol. 20, No. 1, 87-99.
- Collins, H. 1999. *Regulating Contracts*. Oxford. Oxford University Press.
- Domah, P., Pollitt, M.G., and Stern, J. 2002. *Modelling the Costs of Electricity Regulation: Evidence of Human Resource Constraints in Developing Countries*. CMI Working Paper No.11 and LBS Regulation Initiative Discussion Paper No 49.
- Foreman-Peck, J. and Millward, R. 1994. *Public and Private Ownership of British Industry 1820-1990*. Oxford. The Clarendon Press.
- Foster, C.D., 1992. *Privatization, Public Ownership and the Regulation of Natural Monopoly*. Oxford. Blackwell.
- Guasch, J.L. 2001, *Concessions and Regulatory Design: Determinants of Performance – Fifteen Years of Evidence*. Mimeo. World Bank and University of California, San Diego.
- Guasch, J.L., Laffont, J.J. and Straub, S. 2002. *Renegotiation of Concession Contracts in Latin America*. USC Law School CLEO Research Paper CO2-22.
- Hannah, L. 1979. *Electricity before Nationalisation: A Study of the Development of the Electricity Supply Industry in Britain to 1948*. London. The Macmillan Press.
- Hart, O. D., 1995. *Firms, Contracts and Financial Structures*, Oxford, Oxford University Press.
- Hart, O.D. 2003. *Incomplete Contracts and Public Ownership: Remarks and an Application to Public-Private Partnerships*. *Economic Journal* Vol 113, No 486, 69-76.
- Levy, B., Spiller, P.T., 1994. *The institutional foundations of regulatory commitment: a comparative analysis of telecommunications regulation*. *Journal of Law and Economics and Organisation* 10 (2), 201-246.
- Littlechild, S. 1983. *Regulation of British Telecommunications' Profitability*. Report to Secretary of State for Industry. London. Department of Industry.
- Littlechild, S. 2002. *Competitive Bidding for a Long-Term Electricity Distribution Contract*. *Review of Network Economics* Vol.1, No.1, 1-38.

- Macaulay, S. 1963. Non-Contractual Relations in Business. *American Sociological Review* 45.
- Newbery, D.M., 1999. *Privatisation, Restructuring and Regulation of Network Utilities*. The MIT Press.
- Newbery, D.M. and Santos G., 1999. Road Taxes, Road User Charges and Earmarking. *Fiscal Studies*, Vol. 20, No 2, 103-32.
- Noll, R.G., 2000. Telecommunication Reform in Developing Countries. In *Economic Policy Reform: The Second Stage*. Ed. Krueger, A.O. Chicago. University of Chicago Press.
- Posner, R. 1992. *Economic Analysis of Law*. Boston. Little Brown
- Stern, J. 1997. What Makes an Independent Regulator Independent? *Business Strategy Review*, Summer, 67-75.
- Stern, J., Holder, S., 1999. Regulatory Governance: Criteria for Assessing the Performance of Regulatory Systems. An Application to Infrastructure Industries in the Developing Countries of Asia. *Utilities Policy* 8, 33-50.
- Stern, J. 2000. Electricity and Telecommunication Regulatory Institutions in Small and Developing Countries. *Utilities Policy*, Volume 9, 131-157.
- Turvey, R. 1995. The Sliding Scale: Price and Dividend Regulation in the Nineteenth Century Gas Industry. *NERA Topics*, No. 16. London.
- Turvey, R. 1996. Economic Growth and Domestic Refuse in London in *LSE on Social Science: A Centenary Anthology* ed. Sasson H. and Diamond D, London. LSE Books.
- Williamson, O. E., 1985, *The Economic Institutions of Capitalism*, New York, The Free Press.