Regulating Infrastructure:
The Impact on Risk and Investment[*]

Graeme Guthrie[†]

Victoria University of Wellington

June 24, 2005

[†]School of Economics and Finance, PO Box 600, Victoria University of Wellington, Wellington, New Zealand. Ph: 64-4-4635763. Fax: 64-4-4635566. Email: graeme.guthrie@vuw.ac.nz

Regulating Infrastructure:
The Impact on Risk and Investment

Abstract

The last thirty years have witnessed a fundamental change in the regulation of infrastructure industries. Whereas firms were subject to rate of return regulation and protected from entry in the past, now they face various forms of incentive regulation, competition is actively promoted by many regulators, and both regulators and the firms they regulate must often confront rapid technological progress. This paper surveys the literature on the investment implications of different regulatory schemes, highlighting the relevance of modern investment theory, which puts risk and intertemporal issues, such as the irreversibility of much infrastructure investment, center stage. It discusses the impact on regulated monopolists' investment behavior of key regulatory characteristics, namely the price flexibility allowed by the regulator, the length of the regulatory cycle, and the costs the regulator will allow the firm to recover at future regulatory hearings. It also considers the impact of competition, especially the situation where a vertically integrated firm has its operation of a bottleneck asset regulated, on investment by regulated firms and their competitors.

Regulating Infrastructure:
The Impact on Risk and Investment

1 Introduction

The standard static analysis of an unregulated monopolist is that it sets prices too high and production too low. Price regulation is supposed to raise welfare by lowering the prices the firm charges, inducing it to raise production. Ultimately consumers care about prices and service quality, so why then does so much of the debate surrounding regulation focus on investment? First, because investment is crucial to both prices and quantities in the long run, and because delayed investment can have enormous welfare costs.[1] Second, because the investment in regulated industries required to provide these goods and services involves vast sums.[2] Third, because regulation has a substantial impact on investment.[3] For these reasons, a survey of the relationship between regulation and investment is timely.

The last thirty years have witnessed a fundamental change in the regulation of infrastructure industries. Before this transformation began, regulated firms were typically subject to rate of return regulation, were protected from entry, and operated in environments with slow technological progress. The 'cost-plus' nature of such regulation was blamed for high costs and indifferent quality, problems that rate of return regulation has difficulty overcoming due to information asymmetries. Regulators began introducing various forms of incentive regulation, which were designed to improve firms' incentives for cost reduction by separating regulated-price setting from measurement of firms' actual costs incurred. In some industries, technological change made competition more viable than before, encouraging regulators to relax (and even remove) entry restrictions from the late 1970s onwards. Now, regulated firms face various forms of incentive regulation, competition is actively promoted by many regulators, and both regulators and the firms they regulate must often confront rapid technological progress.

---

[1]For example, Hausman (1997) argues that regulatory indecision delayed the introduction of cellular telephones in the US for seven to ten years, and that the cost to consumers was in the range of $31–50 billion each year (in 1994 dollars). However, Hausman's approach has its critics (Pakes, 1997).

[2]For example, Fay and Yepes (2003) estimate that the world-wide demand for infrastructure services exceeds available capacity to such an extent that $370 billion of new investment is required annually over the five year period to 2010, plus a further $480 billion annually for maintenance expenditure; more than half of this $850 billion total is estimated to be required in low and middle income countries.

[3]Alesina et al. (2005) find that deregulation led to greater investment in a panel of 21 OECD countries during the period 1975–1996, with the most important factor being the extent of entry liberalization. For example, if their measure of overall regulatory stringency falls from its third quartile value to its first quartile value, annual investment rises on average by an amount equal to 2.5 percent of the value of the capital stock.

Developments in two regulated industries — the electricity and telecommunications sectors in the US — illustrate the close relationship between the nature of a regulatory regime and the investment behavior of the firms subject to that regime. In both cases, changes in regulation have often been followed by changes in investment behavior. Further, the regulatory reforms have sometimes been driven by explicit investment considerations. A common theme is the tension between allocative and dynamic efficiency when segments of vertically integrated industries are opened up to competition.

Two aspects of the US electric power industry are illustrative. First, events in the generation sector demonstrate the risks in infrastructure investment due to long construction times and long asset lives. Construction of many nuclear power plants begun in the 1970s was not completed until the late 1980s. The high oil prices and high demand growth forecasts that motivated their construction did not materialize and many plants experienced substantial cost overruns. Regulators responded by allowing the affected firms to recover only part of the construction costs. Disallowances totalled approximately $19 billion and utilities subsequently cancelled many plans to build more nuclear plants (Lyon and Mayo, 2004).

Second, the 1992 Energy Policy Act opened up the generation sector to competition, and six years later the Federal Energy Regulatory Commission (FERC) required the vertically integrated owners of transmission assets to provide access to rival generators at cost-based prices without discrimination.[4] However, not only are the components of the interconnected transmission network owned by multiple firms, but these utilities are regulated by their respective state public utility commissions. Combined with the inherent complexities of operating deregulated electricity markets, this has made for slow progress. Under the reforms proposed by the FERC in 1999, transmission networks of different utilities are to be operated by independent system operators, with new transmission assets being provided by investors whose payoff would be the congestion rents associated with these assets. However, there has been strong opposition from many states, and the exact form that transmission pricing will take is still being determined. Meanwhile, investment in new transmission capacity has lagged growth in electricity demand and in new generating capacity, so that the success or failure of whatever form of transmission pricing emerges from this process will be determined largely by whether it can induce the investment needed to overcome the rapidly-growing congestion.

The recent history of the US telecommunications industry provides another example of the

---

[4]The prospect of (lower) competitive prices meant that electric utilities would be unable to recover some of the costs of their existing assets (including many non-nuclear power plants). Estimates of these 'strandable' costs range from $100 billion to $200 billion. However, when the FERC ordered open access to transmission assets it also reaffirmed that "legitimate, prudent and verifiable" strandable costs could be recovered from departing customers (75 FERC ¶ 61,080, p. 5).

issues addressed in this survey. I focus on three fundamental changes in regulation and market structure, starting with the divestiture of AT&T in 1984, which separated the regional Bell operating companies from the manufacturing and long-distance components of AT&T and prevented the so-called 'Baby Bells' from subsequently entering either of these markets. In the ten years following the divestiture, the resulting competition in the long-distance market saw AT&T's market share (measured in minutes of inter-state calls) fall from 84 percent to 59 percent (FCC, 1998) and entrants develop their own long-distance networks. At the state level, telecommunications regulation evolved from rate of return regulation to incentive regulation. The process was gradual and resulted in a wide variety of different forms of incentive regulation (Greenstein et al., 1995; Sappington, 2002). Investment in network 'modernization' (such as fiber-optic cable deployment) by local exchange carriers is greater under incentive regulation than rate of return regulation, with some forms of incentive regulation being more effective at promoting investment than others, and greater competition leading to greater modernization (Greenstein et al., 1995; Ai and Sappington, 2002).

The Telecommunications Act of 1996 required incumbent local exchange carriers to lease their bottleneck facilities to entrants (known as 'local loop unbundling' after the 'local loops' that connect customers' premises to the local exchange carrier's office). In return, the incumbents would be allowed to enter the long-distance market once there was sufficient competition in their local markets. Access prices for leasing the incumbents' facilities, set by state regulators in line with a cost measure developed by the Federal Communications Commission (FCC), are widely regarded to make seeking access more profitable than building rival facilities (Hausman, 1999; Pindyck, 2003). Entrants responded by leasing the incumbents' facilities and building their own complementary assets in staggering numbers.

In February 2003, the FCC announced that incumbents would not have to unbundle new fiber-optic networks, although they would continue to have to unbundle existing phone lines.[5] A federal appeals court upheld the first of these rules in March 2004, saying that "[u]nbundling would skew investment incentives in undesirable ways,"[6] but struck down the FCC's plan to continue mandatory unbundling of other network elements. The FCC did not appeal this decision, which was soon followed by announcements by the incumbents that they would be increasing investment in their own local networks and by the main long-distance operators that they would

---

[5]In 1992, the FERC imposed mandatory unbundling on the liquefied natural gas (LNG) industry, requiring vertically integrated owners of pipelines to provide non-discriminatory open access to competitors. However, in December 2002 it announced that new import terminals would not have to be unbundled, noting that industry participants believed the "...open access requirements were having the unintended effect of potentially deterring investment in new LNG facilities in the United States." (101 FERC ¶ 61,294, para 24)

[6]From the US Court of Appeals (DC Circuit) Opinion on United States Telecom Association v. FCC, 359 F.3d 554 (2004, p. 46).

be reducing their involvement in local exchange competition.[7] More recently, in early 2005 two of the successors to the regional Bell operating companies announced plans to acquire the main long-distance operators (MCI and AT&T are being acquired by SBC and Verizon respectively). The experiment with service-based competition in US telecommunications would appear to be coming to an end, to be replaced with an industry dominated by a much smaller number of vertically integrated firms.

The experiences in these two industries suggest that regulation can have a significant impact on investment behavior. For example, the increased investment in long-distance telecommunications infrastructure following the AT&T divestiture suggests that regulation reduces investment, while the cross-sectional variation in local telecommunications investment suggests that the magnitude of the investment response depends on the specific regulatory regime. Furthermore, the changes in electric utility investment following regulators' actions in shifting the risk of cost overruns from consumers back onto shareholders suggests that the way in which price regulation allocates risk is an important feature of any regulatory regime. The difficulties that the FERC and the FCC have had with their respective open access regimes highlights the tension between allocative efficiency, which motivates opening up some sectors of a regulated industry to competition, and dynamic efficiency, which motivates the concerns about investment in the remaining regulated sectors.

The principal aim of this survey is to highlight the different ways in which the regulation of infrastructure industries impacts on the investment behavior of firms operating in those industries. It draws on developments in economics and finance that advance understanding of the features of different regulatory regimes that influence firms' investment. The subset of the regulatory economics literature surveyed here includes two different approaches to the problem that reflect the backgrounds of the authors involved.[8] Microeconomists typically focus on the investment incentives generated by different regulatory structures, sometimes at the expense of properly modelling risk and intertemporal issues. Financial economists have approached the

---

[7]For example, on July 19, 2004, Verizon (the successor to the regional Bell operating companies Bell Atlantic and NYNEX) announced plans to roll out a new fiber-optic network, enabling it to compete with cable operators by selling both broadband and cable television services. On July 22, 2004, AT&T announced that it would no longer be competing for residential customers.

[8]Another important subset of the regulatory economics literature builds on the standard principal-agent model and considers the design of optimal regulatory policies when the firm has information that cannot be observed by the regulator. This literature, which has been criticized by Crew and Kleindorfer (2002) for assuming that regulators have access to information unavailable to them in practice, derives policies that enable the regulator to induce the firm to use its private information in a way that benefits society as a whole rather than the firm's shareholders exclusively. The standard textbook presentation is provided by Laffont and Tirole (1993), while more recent work is surveyed by Armstrong and Sappington (2003). The present survey focuses on how *actual* regulation affects investment, rather than on 'optimal regulation' as studied in this part of the literature.

problem differently, using models that better incorporate risk, but the price of this greater realism is that analysis of incentive effects is more difficult. When viewed in isolation, these two approaches each give an incomplete picture of the relationship between risk and regulation and investment incentives. This survey will hopefully encourage cross-disciplinary work, since the two approaches offer definite synergies for the study of regulation and its impact on investment.

Another objective of this survey is to highlight the relevance of modern investment theory to studies of regulation. This theory puts risk and intertemporal issues, such as the irreversibility of much infrastructure investment, center stage. While these features do not play an important role under pure rate of return regulation — risk is shifted onto consumers and the cost-plus nature of this regulation means that firms are guaranteed cost recovery, so that they would never want to reverse investment even if reversal was possible — all of this has changed under modern incentive regulation. The prospect of increased competition, along with an increased likelihood of cost disallowances under incentive regulation, means that regulated firms are no longer guaranteed cost recovery on their, largely irreversible, investments. Incentive regulation has shifted risk back onto shareholders, so that risk plays an important role in firm investment. Models are therefore needed that properly specify the risks, technologies, and investment requirements facing regulated firms, and allow examination of the subtleties of investment behavior, such as the timing of investment. This survey focuses on analyzing the interplay of regulation, investment timing, irreversibility and risk, while pointing out the limitations of other approaches when analyzing the modern regulatory environment.

The paper is divided into three distinct parts. The first part (Sections 2 and 3) sets up the paper by characterizing different regulatory regimes and describing the investment environment. The second part discusses the impact on regulated monopolists' investment behavior of key regulatory characteristics, namely the price flexibility allowed by the regulator (Section 4), the length of the regulatory cycle (Section 5), and the costs the regulator will allow the firm to recover at future regulatory hearings (Section 6). The third part (Section 7) extends this analysis to cover issues of importance when regulated firms face competition. It focuses on the situation where a vertically integrated firm has its operation of a bottleneck asset regulated, describing the impact of regulation on investment by this firm and by its competitors in downstream markets. Section 8 concludes.

2   Classifying regulatory regimes

Regulators have adopted a wide array of different regimes, all designed to restrict the behavior of firms. Each regulatory regime can be interpreted as a sequence of contracts between the regulated firm and the regulator, who acts as an agent for consumers (Goldberg, 1976), with each

contract lasting until the next formal regulatory hearing. A small set of simple characteristics distinguishes these contracts: the amount of freedom the regulated firm has in changing prices and choosing its investment program between hearings; the timing of these hearings; and the cost information used when the regulator sets regulatory parameters.

2.1 Price flexibility

Rate of return regulation allows the regulated firm the least flexibility in setting its prices. Despite its name, under rate of return regulation the regulator sets the *prices* for all of the goods and services provided by the regulated firm, not its rate of return. The firm cannot change these prices until they are officially raised or lowered by the regulator, usually requiring another hearing.[9] Exceptions to this rule include so-called fuel-adjustment clauses, which allow the firm to pass a specified fraction of the cost of changes in fuel or other input prices onto consumers without the need for a formal regulatory hearing.

In contrast, under price cap regulation all that is set is a cap on the price of a specified basket of the firm's goods and services. Price caps are often allowed to grow at a specified rate, usually the rate of price inflation less a margin to allow for efficiency gains (hence the name RPI-X regulation, where X is an adjustment that can be changed at regulatory hearings). This makes it easier for the regulator to commit to revise the price cap relatively infrequently. A regulated firm with N distinct goods can typically choose any set of prices $\{p_{1,t}, \ldots, p_{N,t}\}$ in period t that satisfies

$$\sum_{n=1}^{N} q_{n,t-1} p_{n,t} \leq \hat{R}_t, \tag{1}$$

where

$$\hat{R}_t = (1 + i_t - x) \sum_{n=1}^{N} q_{n,t-1} p_{n,t-1},$$

$q_{n,t-1}$ is the quantity of good n sold in period $t-1$, i_t is the rate of inflation in period t, and x is the required 'efficiency gain'. Numerous other possibilities have been considered. For example, AT&T faced a price cap of the form (1) in the early 1990s, but with prices in a base year $(p_{n,0})$ replacing those in the previous year $(p_{n,t-1})$ in the definition of $\hat{R}_t$, while Vogelsang and Finsinger (1979) proposed restricting prices using (1) with the cap equal to the firm's actual

---

[9]Although prices are set so that the firm is expected to earn a rate of return specified by the regulator, unanticipated events following a hearing result in a realized rate of return that differs from the rate chosen by the regulator. The realized rate of return can be higher or lower than this 'allowed' rate of return. This is the main reason why the Averch and Johnson (1962) model of rate of return regulation, in which the regulator imposes an upper bound on the firm's realized rate of return on its physical capital, is not considered in this survey. This famous paper generated an enormous literature (see, for example, the survey by Baumol and Klevorick (1970)), but does not capture all of the relevant details of rate of return regulation.

cost in the previous period:

$$\hat{R}_t = C(q_{1,t-1}, \ldots, q_{N,t-1}).$$

These price caps are all subject to intertemporal manipulation by firms. For example, Foreman (1995) shows how a firm subject to the price cap in (1) can drive down demand in one good in one period, thereby lowering the weight on the price of that good in the following period and allowing the firm to charge almost any price of its choosing. Sappington and Sibley (1992) demonstrate similar behavior for the cap imposed on AT&T, while Sappington (1980) and Hagerman (1990) explain how firms can manipulate the Vogelsang-Finsinger price cap.

A major driving force behind the introduction of price caps was the incompatibility of rate of return regulation with competition. Braeutigam and Panzar (1989) illustrate some of the problems that can arise when firms subject to rate of return regulation operate in a mix of competitive and non-competitive markets. For example, firms have an incentive to misreport cost allocations and underproduce in the non-competitive markets, allowing them to overproduce in the competitive markets. Price caps have alleviated these concerns, since, by omitting prices of its goods in the competitive market from a firm's price cap, regulators can restrict prices in non-competitive markets and leave competitive forces to restrict prices in competitive markets.[10] The greater price flexibility also gives regulated firms greater ability to compete with unregulated entrants.[11]

An alternative to price cap regulation that has been used to regulate telecommunications firms in the US (Greenstein et al., 1995, Appendix 2) and electricity transmission grid owners in England, Wales, and Australia (Cowan, 2002, p. 182) is to cap total revenue. Under such schemes the firm can choose any set of prices $\{p_{1,t}, \ldots, p_{N,t}\}$ that satisfies

$$\sum_{n=1}^{N} q_{n,t} p_{n,t} \leq \hat{R}_t, \tag{2}$$

where $\hat{R}_t$, the maximum revenue the firm is allowed to earn, is outside the firm's control. This has the same form as (1), except that the left side is the price of the basket of goods sold by the firm in the *current* period. Thus, a firm subject to a revenue cap may have to reduce prices in response to rising demand, while a price cap elicits no such response. In practice, excess revenue is returned to customers in the form of a 'consumer dividend', so that the cap is achieved ex

---

[10]In practice, regulators have often restricted the ability of firms to rebalance tariffs under price cap regulation by setting price caps on individual baskets of goods and services, so that a firm faces several different price caps.

[11]In fact, Abel (2002) finds that US telecommunications incumbents subject to price cap regulation have fewer competitors than those subject to other forms of regulation, suggesting that price cap regulation might make it so much easier for firms to compete with entrants that entry is deterred. Iozzi and Fioramanti (2004) show that firms can use price caps as a commitment device, manipulating the price cap rule so that the firm is effectively committed to setting a low price.

post. Thus, revenue caps are often known as revenue-sharing schemes. Regulators have also adopted so-called profit-sharing schemes, in which revenue on the left side of (2) is replaced with some measure of profit.

2.2 Timing of regulatory reviews

The timing of regulatory hearings is a critical characteristic of any regulatory regime. Under rate of return regulation as practiced in the US, public utility commissions usually set hearing dates after the regulated firm files for a hearing, although consumers can put sufficient pressure on a regulator to initiate a hearing. In contrast, price cap regulation is typically reviewed at regular dates in the UK, typically four or five years apart.[12]

Joskow (1974) argues that under rate of return regulation (at least as practiced in the US in the 1960s and 1970s) regulators are primarily interested in limiting nominal price rises. Firms' requests for price reductions are generally approved without a formal rate of return review, while requests for price increases are subjected to greater scrutiny and the new prices are set so that the firms earn 'allowed' rates of return on their investments in physical capital. Thus rate of return regulation effectively imposes a lower bound on firms' rates of return, equal to the rate that will be set at a formal hearing. Such hearings will only be requested (and then only by the firm) once the realized rate of return is less than the rate of return that will be allowed at such a hearing.[13] Between hearings, firms will earn higher rates of return.[14] If consumers are sufficiently organized, their ability to pressure the regulator into calling a hearing effectively imposes an upper bound on the firm's rate of return, in which case the rate of return fluctuates between the two bounds. When the lower bound is breached, the firm requests a price rise, resulting in a hearing; just before the upper bound is breached, the firm will request a price reduction, thus avoiding the need for a consumer-initiated hearing that would likely result in an even greater price reduction. Thus, the timing of hearings is determined (endogenously) by the evolution of the firm's profitability.

The experience in the US electric utility industry supports Joskow's model of rate of return regulation. Of all 363 rate cases involving US electric utilities in the period 1948–1978, 350 were

---

[12]Some price cap schemes explicitly incorporate flexibility in the timing of reviews. For example, AT&T had the option to apply for a review of its cap (Laffont and Tirole, 1993, p. 18); in the UK, water utility price caps have a ten year cycle, unless either side requests a review after five years (Cowan, 2002, p. 180).

[13]Nowell and Shogren (1991) argue that firms wait past the point where their rate of return is below what they would receive from a new hearing in order to reduce the uncertainty of a regulatory surprise — since delaying increases the information revealed by actual costs and therefore reduces the magnitude of any regulatory error in setting prices.

[14]Joskow (1973, 1974) reports evidence that regulated US electric utilities that requested hearings during the 1960s were earning lower rates of return prior to the hearings than those firms that were not seeking hearings.

initiated by the firms and only 13 by public utility commissions (Braeutigam and Quirk, 1984); of all 96 US electric utility rate reviews during the period 1980–1984, firms waited on average 241 days (with a standard deviation of 177 days) before filing for a price review (Atkinson and Nowell, 1994). This suggests that, at least during the period of these studies, firms typically chose the timing of formal hearings. Joskow (1973) examines the behavior of seven electric utilities in New York state during the period 1961–1969 and finds that in any given year a firm is more likely to request a price increase if the growth in earnings per share is low, if earnings are low relative to interest payments, and if it is earning a rate of return below the regulator's most recently approved rate of return. Roberts et al. (1978) find that the first two factors explain the requests for price increases made by four electric utilities in Florida during the period 1960–1976.

Under the revenue-sharing regime described in Section 2.1, firms keep all revenue below a specified threshold, and must share a specified proportion of revenue above this threshold with customers. This effectively institutionalizes the hearings that occur under rate of return regulation — under rate of return regulation, high levels of revenue trigger a hearing at which the regulator will lower the firm's prices, while under revenue-sharing regulation, prices are effectively adjusted automatically, without the need for a formal hearing. Of course, there will be further differences in practice. For example, lags in the regulatory process will allow a rate of return regulated firm to keep some revenue above the threshold, whereas the excess will actually be returned to customers under revenue sharing.

2.3   Cost measures

When regulatory parameters are set at a hearing, the regulator's aim is usually to allow the firm enough revenue to cover its costs, and no more. However, regulators have adopted many different definitions of costs, especially with regard to the cost of physical capital. I discuss some of the implications in this section.

2.3.1   Different components of cost

The costs that influence regulatory settings can be decomposed into three types: operating costs ($\hat{C}_{t+1}$ in period $t+1$), a return on the firm's capital ($\hat{r}_t \hat{B}_t$, where $\hat{r}_t$ is the allowed rate of return and $\hat{B}_t$ is the firm's 'rate base'), and depreciation of the firm's capital ($\hat{D}_t$). All of these quantities are specified by the regulator.[15] Prices are set such that expected revenue in period $t+1$, $E_t[R_{t+1}]$, satisfies

$$E_t[R_{t+1}] = E_t[\hat{C}_{t+1}] + \hat{r}_t \hat{B}_t + \hat{D}_t. \tag{3}$$

---

[15]To simplify the presentation, here I assume that hearings are held each period. Endogenous hearing dates are covered at the end of this section.

Financial markets determine the market value of the firm in period t as the present value of the stream of net cash flows $R_{t+s} - C_{t+s} - I_{t+s}$ (for all $s \geq 1$), where C_{t+s} denotes the firm's actual operating costs in period $t + s$ and I_{t+s} denotes the firm's investment expenditure in the same period. The market value of the firm in period t, F_t, is equal to

$$F_t = \frac{E_t[R_{t+1} - C_{t+1} - I_{t+1} + F_{t+1}]}{1 + r_t}, \tag{4}$$

where F_{t+1} is the market value of the firm in the following period and r_t equals the expected rate of return investors earn from buying the regulated firm as a going concern; that is, it equals the firm's cost of capital. However, many different combinations of operating cost, rate base, allowed rate of return, and depreciation lead to the same value of $R_{t+s} - C_{t+s} - I_{t+s}$, and hence to the same market value of the firm.[16] In particular, suppose that the regulator sets the rate base equal to $B_t = F_t$ this period and $B_{t+1} = F_{t+1}$ next period, where F_t and F_{t+1} are the market values resulting from the regulatory settings in (3). If the regulator calculates depreciation according to

$$D_t = B_t - E_t[B_{t+1} - I_{t+1}],$$

uses the firm's actual operating costs as the operating cost measure, and allows a rate of return equal to the firm's cost of capital, then the firm's allowed revenue is

$$E_t[R_{t+1}] = E_t[C_{t+1}] + r_t F_t + (F_t - E_t[F_{t+1} - I_{t+1}]). \tag{5}$$

Comparing (4) with (5) shows that the market value of the firm will also equal F_t with these regulatory settings.[17] For this reason, throughout the remainder of this survey I proceed as though the regulator chooses regulatory settings in such a way that the firm's market value equals the regulator's chosen rate base — the regulator effectively chooses the firm's market value.[18] The rule for setting the rate base is thus the key cost measure, and is discussed in more detail in Section 2.3.2. When combined with the rules for determining the timing of regulatory

---

[16]For example, raising the allowed rate of return will not affect outcomes if the rate base is simultaneously reduced by an appropriate amount.

[17]Greenwald (1984) proves the result that the market value of the regulated firm equals its rate base if and only if the allowed rate of return equals its cost of capital. Schmalensee (1989) shows that, provided the allowed rate of return equals the firm's cost of capital, the net present value from investment equals zero for any depreciation schedule that fully depreciates the firm's assets over their service life (that is, $\sum_t I_t = \sum_t D_t$). The regulator is therefore free to choose any depreciation schedule, as long as the allowed rate of return equals the firm's cost of capital. There exists a considerable literature comparing different depreciation schedules on the basis of the prices (for example, Baumol, 1971) and investment decisions (for example, Rogerson, 1992) that they induce. However, Biglaiser and Riordan (2000) show that the depreciation rules used in practice are not sufficiently flexible to induce efficient prices.

[18]Regulators may not always present the results of their hearings in this way, but sometimes the focus on market valuations is explicit. For example, the share prices of regulated electric utilities in the UK rose significantly

hearings and the firm's price flexibility, it describes the most important features of a regulatory regime.

When regulatory settings are normalized in such a way that the market value of the firm equals the firm's rate base, the allowed rate of return $\hat{r}_t$ equals the firm's cost of capital. It therefore rewards investors for bearing the risk associated with the cash flow $R_{t+1} - C_{t+1} + B_{t+1} - I_{t+1}$. The appearance of allowed revenue, R_{t+1}, and the revalued rate base, $B_{t+1} - I_{t+1}$, shows that regulatory settings can affect the allowed rate of return, an issue that is pursued in Section 2.3.3. (For example, the use of two-part tariffs reduces the sensitivity of revenue to quantity variations.) Thus, regulated prices and the allowed rate of return must be determined simultaneously; it is inappropriate to set prices based on an exogenously-determined allowed rate of return.[19]

When the timing of hearings is endogenous, the allocation of risk between the firm and its customers is altered. At one extreme, if the firm can request a formal hearing at any time but customers are not sufficiently organized to pressure the regulator into calling a hearing, then the option to request higher prices imposes a floor on the firm's market value. This allows shareholders to keep positive shocks, and shift some of the risk of large negative shocks onto customers. It increases the value of the firm relative to an otherwise comparable firm with exogenous hearing dates.[20] At the other extreme, if consumers are able to pressure the regulator into calling a hearing, but the firm cannot exert such pressure, then shareholders must bear all of the risk of negative shocks, while sharing the risk of positive shocks with customers. It reduces the value of the firm relative to an otherwise comparable firm with exogenous hearing dates. The overall impact of endogenous hearings on the value of the firm, and therefore its cost of capital, depends on the ease with which consumers can pressure the regulator.[21]

---

following the announcement in August 1994 of new price caps, which were to come into effect in April of the following year. Partly in response to this share price performance, the electricity industry regulator revised the terms of the price caps in March 1995 (Faure-Grimaud, 2002).

[19] Marshall et al. (1981) use the Capital Asset Pricing Model (CAPM) to analyze the endogenous determination of risk in a regulated environment. Their analysis is extended to a wider range of regulatory regimes, in a way that allows for investment to be irreversible, by Evans and Guthrie (2005a).

[20] A firm with the option to choose its own hearing dates can behave like an otherwise equivalent firm with exogenous hearing dates simply by requesting hearings so that their timing coincides with those of the second firm. The freedom to choose hearings at other dates adds to (or at least does not subtract from) the value of the firm.

[21] Brennan and Schwartz (1982a) and Lewellen and Mauer (1993) use the continuous time framework of contingent claims valuation to analyze the impact of endogenous hearing dates on a regulated firm's value and cost of capital.

One of the most important questions that regulators need to consider when choosing regulatory settings is the level of the rate base, since this determines who pays for the firm's assets (across states of nature, and across time).[22] Greenwald (1980, 1984) suggests that three conditions should be imposed on the rate base. First, there should be no windfall profits to the firm at its inception; that is, the rate base must equal zero immediately before the firm enters business. Second, the rate base must always be nonnegative (an ongoing participation constraint reflecting the limited liability of the firm's shareholders) and no greater than the market value of an unregulated monopolist (in order to be feasible). Third, the rate base must meet the 'capital attraction' standard, which stems from the Federal Power Commission v. Hope Natural Gas decision discussed in Section 6.2 and requires that the regulated firm should always be able to attract capital for investment. Greenwald (1980) interprets this as requiring that investment can be financed without either a gain or a loss for existing shareholders; that is, investment should have a zero net present value (NPV).

The rate base often associated with traditional rate of return regulation is the actual cost of the firm's physical capital (adjusted for depreciation). However, since this gives firms carte blanche to incur unnecessary costs, safe in the knowledge that their customers will have to pay, regulators typically only allow firms to recover costs deemed to be prudently incurred. Alternatively, regulators may only allow firms to recover the cost of 'used-and-useful' assets, which are those assets judged by the regulator to be necessary to service customers' demands. That is, ex post demand information is used by the regulator. Sometimes, other ex post information is also used. For example, when only 'avoided costs' are allowed by regulators, firms are only allowed to recover the costs that the firm has been able to avoid because of its investment in a particular asset. The US electric utility industry witnessed large cost disallowances in all three categories in the 1980s, totalling more than $19 billion (Lyon and Mayo, 2004).

A variety of different rate bases have been suggested for use under incentive regulation. Common themes are that the rate base should reflect the market value of the firm that would eventuate from some sort of 'competitive' process and that the rate base should be independent of the firm's investment decisions. An important example of such a regime is provided by pricing based on the TELRIC (total element long run incremental cost) cost measure and its variants, which is widely used by telecommunications regulators.[23] Under this regime, prices are based on the cost structure of an efficient cost-minimizing firm with an optimally-configured network

---

[22]This is especially important when firms face competition, as the movement of customers between firms can result in situations where the shareholders and remaining customers of a firm must bear the cost of assets built to meet the demand from customers who have since departed to rival firms.

[23]Mandy (2002) offers a detailed description of TELRIC pricing as implemented by the FCC.

built with the current technology, subject only to the constraint that the locations of key parts of the network are determined by the regulated firm's existing network.[24] In TELRIC pricing, and other pricing based on benchmark costs, the relevant cost is that faced by a hypothetical firm that replaces the regulated firm at the date of the hearing, and is able to use the latest technology when building its infrastructure, even if that technology was not available when the regulated firm built its assets.

There is an ongoing debate concerning the measurement of capital costs under incentive regulation. The simplest suggestion is that the rate base should equal the optimized replacement cost (ORC) of the firm's assets, which is defined as the cost of replacing the firm's assets with new ones optimally configured to meet demand (Johnstone, 2003). This is how much it would cost a hypothetical replacement firm to provide the same level of service as the regulated firm. The TELRIC cost measure is one example of this approach. However, many authors (for example, Salinger (1998), Hausman and Myers (2002), Mandy and Sharkey (2003), and Evans and Guthrie (2005a)) argue that the appropriate standard is that a replacement firm should be able to break even assuming that the same standard is used at all future hearings; that is, the hypothetical replacement firm recognizes that its future prices will be set in line with the costs of a future hypothetical replacement firm, and so on. Hausman and Myers (2002) argue that US railroad regulators fail to allow for these future replacement firms, resulting in cost measures that are unrealistically low.[25]

Computation of the 'hypothetical cost' rate base is a far more complicated process than that of incorporating a firm's actual costs in its rate base.[26] Because the cost calculation is forward-looking, implementation requires a dynamic model that fully takes into account future risks and the irreversible nature of investment.[27] In such an environment, excess capacity allows the regulated firm to avoid some future investment expenditure, and the value of this saving should

---

[24] "The total element long-run incremental cost of an element should be measured based on the use of the most efficient telecommunications technology currently available and the lowest cost network configuration, given the existing location of the incumbent LEC's wire centers" (Title 47 of the Code of Federal Regulations, §51.505). It thus represents a compromise between the costs that the incumbent will actually incur and those that a greenfields operation would incur.

[25] Railroad tariffs are regulated if the regulator deems that the railroad is a dominant provider, where dominance is determined to exist if revenue is more than sufficient for a hypothetical replacement firm to invest and break even (Hausman and Myers, 2002).

[26] In practice, complicated engineering models are used to determine the composition and cost of a hypothetical efficiently-configured network. Allowing for the appearance of future hypothetical replacement firms, which is not done in practice, would further complicate the modeling involved. Salinger (1998), Hausman and Myers (2002), and Evans and Guthrie (2005a, 2005c) illustrate the complexities involved.

[27] Implementation would be much simpler if investment were costlessly reversible, since then the regulated firm could redeploy capital if the hypothetical replacement firm had lower costs.

be included in the hypothetical cost rate base (Baumol and Sidak, 2002; Evans and Guthrie, 2005a). That is, assets that would not be built by a replacement firm may still have some value, and this value should be included in the rate base. This leads to a rate base that equals the present value of the future investment expenditure the firm can avoid because of its ownership of its existing assets.[28]

In practice, when price caps and other incentive regulation schemes are reviewed, the new settings will not necessarily reflect only the hypothetical lowest costs. The revised price caps will likely reflect past cost savings, as this is necessary for the regulatory regime to be politically sustainable — consumers and politicians will not tolerate large profits for long periods of time (Braeutigam and Panzar, 1993). Possibly anticipating these commitment problems, the legislation supporting the Kansas Telecommunications Agreement prohibits the regulator from conducting an "earnings audit" for the purposes of reviewing the price cap (Weisman, 2002, footnote 21). However, even without such restrictions, experience suggests that prices are not adjusted to fully reflect actual costs. For example, telecommunications firms and water utilities in the UK were allowed to keep some supernormal profits at the first review of their price caps (Cowan, 2002, p. 180).

2.3.3   The cost of capital

The cost of capital depends on how the regulatory regime allocates risk between shareholders and consumers. A regime that shifts (systematic) risk from shareholders to consumers leads to a lower cost of capital, but the benefits of the resulting lower prices to consumers are (at least partly) offset by the greater risk that they must bear.

A crucial determinant of a firm's cost of capital is the cost measure adopted by the regulator. Rate of return regulation immunizes shareholders from shocks to long-term cash flows because prices will be adjusted at future hearings to allow the firm to recover its costs. For example, if prices are held fixed between hearings and prices set at the next hearing reflect the actual costs incurred by the firm, then investors bear the consequences of any shocks to operating costs up until the next hearing (since the prices set at the previous hearing did not anticipate these shocks and cannot be changed), while consumers bear the consequences of the impact on post-hearing operating costs (since the prices faced by consumers after the next hearing will reflect the information contained in the operating cost shock). The risk of demand shocks will be allocated in the same way. In contrast, the use of benchmark costs under incentive regulation means that shareholders typically bear much more of the risk of shocks to long-term cash flows. For instance, under incentive regulation a permanent negative shock to demand will be reflected

---

[28]Greenwald (1984) provides an early definition of a similar rate base.

in a lower rate base at the next regulatory hearing since it reduces the amount of physical capital required by a hypothetical replacement firm; under rate of return regulation, the rate base would be unaffected unless past investments were reevaluated using the used-and-useful criterion. Similarly, if the regulator uses the replacement cost of physical capital, rather than its historical cost, then the firm is exposed to the risk of capital price shocks (Evans and Guthrie, 2005a; Salinger, 1998). However, in terms of their impact on the rate base, demand risk and capital price risk are subtly different. Demand risk is a 'real' risk, since it exists regardless of the regulatory settings. That is, it must be shared between investors and consumers, with the allocation determined by the regulator.[29] In contrast, capital price risk is partly a 'created' risk, since it affects stakeholders only if the regulator revalues the firm's assets at each hearing; if the regulator does not revalue the rate base, then there is no capital price risk to allocate.

The other characteristics of regulatory regimes — the degree of price flexibility and the rules for determining the timing of hearings — also affect the cost of capital. For example, the fuel cost adjustment clauses (FACs) introduced and extended in the 1970s allowed many US electric utilities to pass higher fuel costs onto customers automatically, avoiding the need to wait until a future regulatory hearing. By shifting operating cost risk onto consumers, FACs reduce the firm's cost of capital (Clarke, 1978).[30] The frequency of regulatory hearings also affects the cost of capital, but the relationship depends on the form of regulation. Under rate of return regulation, investors only bear the consequences of shocks *until* the next hearing — investors are exposed to less risk if hearings are held more frequently, so that the cost of capital will fall (Brennan and Schwartz, 1982a).[31] In contrast, under incentive regulation, investors only bear the consequences of shocks to replacement cost *after* the next hearing, when the rate base is recalculated — investors can thus be exposed to more risk if hearings are held more frequently, so that the cost of capital will rise (Evans and Guthrie, 2005c).

Relative to traditional rate of return regulation, firms subject to incentive regulation usually have greater price flexibility between hearings, which lowers their cost of capital, and face more

---

[29]However, regulation can affect the magnitude of demand risk by, for example, opening an industry up to competition.

[30]Consistent with this, Clarke (1980) finds that the use of FACs by 39 large US electric utilities between 1965 and 1974 lowered their cost of equity capital, with the effect being much greater for firms using oil and gas as a fuel source than for firms using coal. However, Golec's (1990) study of 79 US electric utilities over the period 1969–1983 suggests that FACs have a statistically insignificant effect on firms' cost of equity capital.

[31]Prager (1989) analyses the impact of various aspects of rate of return regulation on the borrowing rate for 100 privately-owned US electric utilities in 1979, and finds that the most important determinants of the firms' borrowing rates are the regulators' treatment of construction work in progress (the rate is lower when work in progress is included in the rate base), regulatory delay (the rate is higher when there is a long delay before regulatory hearings), and the use of interim rate relief (the rate is lower if firms are allowed to raise prices in anticipation of an upcoming hearing).

risk concerning future values of their rate base, which raises their cost of capital. They face regulatory hearings less often, which mitigates the higher rate-base risk. Because of the opposing effects, the overall impact of the form of regulation on the cost of capital is largely an empirical question. Unfortunately, empirical research into the impact of regulation on risk is complicated by the fact that, because of the lengthy hearing processes involved, many regulatory events are anticipated by the market. Indeed, market expectations are likely to change gradually over the whole hearing process. This makes it difficult to estimate the impact of changes in regulatory regime on the systematic risk of the affected firms' share prices. In particular, event studies have great difficulty in precisely identifying the effects of regulatory changes (Binder, 1985).

One possible solution to this difficulty is to focus on the effect of electoral surprises. For example, Grout and Zalewska (2005) estimate the systematic risk of regulated firms' equity in the UK in the period 1993–2000, and focus particular attention on the period after the new Labour government was elected in 1997. They find a statistically significant reduction in the cost of equity capital of regulated firms during the period when the new government was considering switching from price cap regulation to profit-sharing regulation, a move which would have reduced the short-term cash-flow risk, and possibly rate-base risk, borne by investors. For example, the CAPM beta falls by somewhere between 0.2 and 0.4, depending on the model estimated.[32]

3   Investment

Now that the regulatory setting has been described, I turn to the nature of regulated firms' investment environments.[33]

3.1   Features of physical capital

Firms can invest in order to maintain the status quo (such as replacing existing physical capital that has reached the end of its physical lifetime) or to change their operating environment. The latter form of investment can be further decomposed into two broad categories: investment

---

[32]One shortcoming of this approach is that there is no specific alternative regulatory regime that is being factored into prices during the period of policy uncertainty. However, comparisons of specific regimes are complicated by the difficulty in controlling for non-regulatory factors that influence regulated firms' costs of capital. For example, Alexander et al. (1996) compare asset betas of regulated firms in different countries and industries and find that systematic risk is greater under price cap regulation than under rate of return regulation. However, because each firm's beta is measured against the host nation's stock market index, they are not comparable across nations. Other cross-sectional studies have similar shortcomings.

[33]Throughout this section, references are given to papers that explicitly model (usually in the context of regulated firms) the aspects of investment discussed here.

to reduce operating costs and investment to increase demand. Illustrating the first category, a vertically integrated electric utility can invest in transmission equipment to reduce line losses and thereby reduce its fuel needs. Similarly, by investing in new capital with low operating costs, a firm may be able to retire old capital with high operating costs (Biglaiser and Riordan, 2000). The second category can range from investing in improving the quality of an existing service to innovative investment that makes new services available to customers.

When investing, firms face the risk of fluctuations in the cost of a project and in the value of the completed project. Cost shocks can arise due to input price uncertainty that is external to the firm's investment activity, or to project-specific technical shocks such as shocks to the time required to complete a project (Pindyck, 1993). Both sources expose the firm to the risk of cost overruns, which may be disallowed by regulators at subsequent hearings. Payoff shocks can arise due to input price shocks, demand shocks, and regulatory policy shocks. The latter range from unanticipated changes in regulatory parameters, such as the allowed rate of return applied to a firm's rate base, to fundamental policy changes, such as the opening of the US electricity generation sector to competition (Woroch, 1998).

These risks are especially important because investment by the firms considered in this survey is largely irreversible. This can arise because the physical capital deployed by a firm cannot economically be recovered and used elsewhere (for example, the cost of recovering fiber-optic cable from underground may outweigh its resale value). Even if it can be physically recovered, the capital may be industry-specific, in which case its resale value will be low in precisely those states of nature where its owners would like to recover their investment. When a firm invests in a project, it is exposed to the risk that the cash flows generated by the project may be insufficient to cover the project's cost. If the investment could be costlessly reversed, then the firm could redeploy its capital if cash flows are low — the only risk the firm would face is that the project's cash flows would be less than the opportunity cost of capital during the time it takes to redeploy the capital. However, when the investment is irreversible the firm's inability to redeploy the capital makes the potential losses much greater.

The risks associated with irreversibility are compounded by three characteristics of infrastructure assets. First, long lead times expose firms to the risk that economic conditions change during construction, leaving a completed asset which is under-utilized. A notable example is provided by the US nuclear power industry during the 1970s and 1980s (Pindyck, 1993). Nuclear plants took anywhere from six to 16 years to complete in the US. Construction of many plants began when electricity demand was growing rapidly and alternative technologies were relatively unattractive (due mainly to high oil prices). By the time many projects were completed, the forecast demand growth had not materialized and oil prices had fallen, contributing to the cost

disallowances described in Section 1. Second, their long physical lifetimes expose the owners of infrastructure assets to the risk that regulatory settings may change well before projects' costs have been recovered by the firms. Although there are differences across industries, typical infrastructure assets have physical lifetimes that greatly exceed the length of the regulatory cycle. This creates commitment problems for regulators, as discussed in Sections 6.1 and 6.2. Finally, economies of scale make it expensive for firms to make the small additions to capacity that would enable them to mitigate some of the risks described above. Instead, firms take on large risks by making large lumpy investments, often in advance of demand.

3.2 Investment choices available to firms

Firms often have flexibility regarding the timing of their investment — investment is rarely a now-or-never proposition. They will only invest when the payoff from doing so exceeds the opportunity cost of investing, where this opportunity cost includes the value of any timing options destroyed by the act of investing. As well as the investment timing decision, firms must decide what to build. For example, different technologies will offer trade offs between capital costs and operating costs, and can also expose the firms to different risks. For example, in the electricity generation sector, combined cycle gas generation plants have lower capital costs and higher fuel costs than traditional coal-fuelled generators; in the telecommunications industry, a firm may favor cheap but irreversible technology (fixed landline networks are relatively cheap but the investment is largely irreversible) or expensive reversible technology (cellular phone technology is easier to redeploy).[34] Firms also have considerable flexibility during the construction process. They can vary the rate of construction (Brennan and Schwartz, 1982b; Teisberg, 1993, 1994) and even the nature of the project. For example, in the wake of cost disallowances for nuclear power plants in the US, the developers of the almost-completed Zimmer 1 nuclear plant in Ohio converted it to use natural gas as a fuel, while construction of many other partly-completed plants was abandoned.

However, some aspects of this investment flexibility can be restricted by regulators. Regulatory approval may be needed before construction can begin (for example, gas facilities require FERC approval), but this restricts, rather than eliminates, the firms' investment flexibility. In many cases, regulated firms operate subject to a common carrier obligation that forces them to meet all demand at regulated prices. Such obligations are often motivated by beliefs that services may be necessities that should be available to all on the grounds of equity, that complete access to essential services stimulates economic development and growth, and that there are significant

---

[34]Kandel and Pearson (2002) model the investment policy of an unregulated monopolist that can combine irreversible and reversible capital when faced with volatile demand.

positive externalities that an unregulated market would fail to incorporate. Another restriction, that is common under incentive regulation, is that the firm must meet minimum quality standards.[35] Other restrictions on investment can be more short-term in nature, such as when regulators relax regulatory settings on the condition that firms undertake specific investments. For example, in Vermont the regulator agreed not to hold hearings during the period 1987–1991 if New England Telephone invested $284 million in network modernization (Sappington and Weisman, 1996, p. 74; Greenstein et al. 1995, p. 192 and Appendix 1).

Despite these restrictions, regulated firms still have considerable investment flexibility, especially under incentive regulation. In particular, the investment possibilities are much richer than the simple quantity decision typical of many static analyses of regulated firms' investment behavior. Therefore, analysis of the impact of regulation on investment must consider a variety of investment measures: timing, quantity, type, and so on.

3.3 Value-maximizing investment

I assume that regulated firms adopt policies that maximize their market value.[36] This assumption is now standard, but in the literature that emerged from the Averch-Johnson model of rate of return regulation, many other firm objective functions were suggested by regulatory economists, including total revenue, the physical quantity of the firm's output, and the realized rate of return (Baumol and Klevorick, 1970).

Firms that seek to maximize their market value will invest only when the net present value (NPV) of the incremental cash flow from investing is positive. This calculation must include all relevant cash flows so that, for example, when the firm has investment flexibility the NPV calculation must include the cost of all options destroyed by the act of investing. This observation lies at the heart of the real options approach to capital budgeting (Dixit and Pindyck, 1994). For example, when a firm has the option to delay investment, it will only invest once the present value of the cash flows generated by a project exceeds the sum of the required investment expenditure and the value of the delay option destroyed by investment (Dobbs, 2004). Similarly, when investment in physical capital exhibits increasing returns to scale, the firm invests to the point that the gain from investing in an additional unit of capacity (which derives from the ability to

---

[35]If a regulated firm is able to retain its cost savings, then it may be motivated to lower quality in order to reduce costs and thereby raise its profits. If it is subject to a common carrier obligation and the investment required to meet demand has a negative NPV, then the firm has an incentive to allow quality to fall if this would reduce demand and eliminate the need for the investment.

[36]In practice, the separation between ownership and control of regulated firms creates the possibility of agency conflicts between a firm's management, shareholders and bondholders that can result in firm actions that deviate from market-value maximization (Jensen and Meckling, 1976). While the corporate finance literature on this topic is enormous, it has not been the focus of regulatory economists' investigations.

delay the next act of investment) equals the cost (which derives from the possibility that demand will fall, leaving the firm with un-used capacity). The value-maximizing investment policy will have the firm making occasional lumpy investments, often in advance of demand (Evans and Guthrie, 2005c).

In a static setting, a social planner charged with maximizing the sum of producer and consumer surplus would order production at the level where price equals marginal cost. An unregulated monopolist produces at a level where price exceeds marginal cost, therefore supplying quantities of goods and services that are lower than the socially-optimal levels. Because production is too low, unregulated monopolists under-invest from the point of view of overall welfare, although they use a cost-minimizing combination of inputs given their chosen level of output. An important role of regulation is to lower price below the monopoly level (or, equivalently, raise production) without unduly distorting the monopolist's choice of inputs. In an intertemporal setting, the social planner would not invest as soon as the present value of the future flow of producer and consumer surplus equals the required investment expenditure since investment destroys valuable timing options. However, an unregulated monopolist waits even longer than this, since it bears the entire cost of the investment, but must share the resulting surplus with consumers. As in the static setting, here regulation can improve welfare if it can move investment towards the social optimum.

This brings me to the central question addressed in this survey: How do regulatory settings affect investment? I proceed in roughly chronological order, starting by considering the impact of whether or not the firm can change its prices prior to the next hearing (Section 4), then considering the role that the timing of the next hearing plays (Section 5), before analyzing the importance of expectations regarding what the regulator will decide at the next hearing (Section 6).

4 Can the regulated firm change its prices between hearings?

The first characteristic of a regulatory scheme described in Section 2 is the degree of price flexibility retained by the firm between regulatory hearings. This section shows how capping a firm's output prices can alter its investment behavior in a quite different way from simply setting the firm's output prices directly as occurs under rate of return regulation. Indeed, allowing the firm to retain some price flexibility can improve dynamic efficiency, relative to traditional regulatory price setting.

The potential for price caps to affect investment behavior is illustrated using a stylized version of a model due to Dobbs (2004). A monopolist faces risky future demand and has the

option to make an irreversible investment in additional capacity. If the firm invests this period, it runs the risk that demand will fall and the value of the additional revenue it can collect from the extra capacity will be insufficient to cover its cost. If the firm waits until future demand is known before deciding whether or not to invest, then while the expansion is completed it must meet the higher demand using the existing capacity. However, the ability of an unregulated firm to raise its output price in this situation partly offsets the cost of waiting for the new capacity to be built, encouraging delayed investment. In this case, a suitably designed price cap can improve investment timing.

In this model, the firm faces the inverse demand function $P = P(Q)$ in the current period. Next period demand will either permanently grow to $P_h(Q)$ or permanently fall to $P_l(Q)$, with each outcome being equally likely.[37] The firm currently has physical capital with capacity K, and can undertake investment costing I that will increase total capacity to $K + 1$ the period after construction begins. Each unit of capacity allows the firm to meet one unit of demand, and demand functions are such that an unregulated monopolist maximizes its net revenue flow by setting its output price so that all available capacity is used. Operating costs equal zero. In order to illustrate the main ideas, I assume that the demand shock is such that if the firm waits, its value is maximized by investing next period if and only if demand rises.

If the firm is unregulated and invests this period, its market value equals

$$F_{\text{invest}} = KP(K) - I + \frac{\frac{1}{2}(K + 1)P_h(K + 1) + \frac{1}{2}(K + 1)P_l(K + 1)}{r},$$

where r is the discount rate. If it delays the expansion decision until next period, the firm's market value equals

$$F_{\text{wait}} = KP(K) + \frac{1}{2}\frac{1}{1 + r}\left(KP_h(K) - I + \frac{(K + 1)P_h(K + 1)}{r}\right) + \frac{1}{2}\frac{KP_l(K)}{r}$$

this period, since it only invests next period if the high demand state eventuates. If the firm waits until it knows whether construction is justified, then it misses out on the early gains, but if it invests early to make sure it gets those gains, then it may find demand has fallen by the time the project is complete. The firm invests this period if and only if $F_{\text{invest}} > F_{\text{wait}}$. Comparing F_{invest} and F_{wait} shows that the firm will invest this period if and only if

$$E[X_i] > rI + \frac{1}{2}\left(I - \frac{X_l}{r}\right), \tag{6}$$

where $X_i = (K + 1)P_i(K + 1) - KP_i(K)$ is the incremental cash flow from having the additional capacity in state i (either h or l). By investing this period, the firm receives additional cash

---

[37]Simple binomial models such as this are often used to illustrate the consequences of irreversibility and uncertainty in the real option literature. See, for example, Dixit and Pindyck (1994, Chapter 2).

flows of X_i next period; the left side of (6) is the expected gain to the firm, However, there are two associated costs, corresponding to the two terms on the right side of (6). First, the firm must incur the investment expenditure sooner. Second, because investment is irreversible the firm is exposed to the risk that the low-demand state will eventuate and that it will regret, ex post, investing early; the second term on the right side of (6) is the expected cost to the firm of not being able to reverse its early investment decision. If X_l is low, so that expansion in the low-demand state is especially unprofitable, then the cost of investing early is high, and the expected incremental cash flow from investing must be high in order for early investment to be value-maximizing. This is an example of the 'bad news principle': the incremental cash flows generated by investment must be sufficient to compensate the firm not only for its investment of capital, but also for any bad news which may arise following the irreversible investment decision (Bernanke, 1983).

However, (6) does not generally result in socially-optimal investment timing — the firm waits too long before investing because it incurs all investment expenditure, but receives only the producer surplus, while a social planner would trade off the investment expenditure against the sum of consumer and producer surplus.[38] Investment can be accelerated, and the welfare losses reduced, with the use of a suitably-designed price cap. In the case of the simple example here, if the price cap $\hat{P}$ is less than $P_h(K)$ and greater than both $P_h(K+1)$ and $P_l(K)$, then regulation raises X_h, by preventing the firm from setting a particularly high price if demand grows ahead of capacity, without affecting X_l. That is, it can raise the expected gain from immediate investment (the left side of (6)) without affecting the cost (the right side of (6)).[39] This encourages earlier investment.[40] The key is to choose a price cap that will not be binding in the low demand state — when demand is low, the firm will want to set a low output price anyway, so the cap will not be binding — but will bind when demand is high and capacity is at its original level. By restricting the firm's price increase in this state, the price cap makes it more expensive for the firm to delay investment, and thus encourages earlier investment.[41]

---

[38]Because the unregulated firm always sets prices such that it operates at full capacity, total surplus is maximized, conditional on capacity. Setting a lower price would not affect realized demand (it would just result in rationing), and just transfer surplus from producer to consumers. Setting a higher output price would lower realized demand and also lower total surplus. Thus, the key determinant of welfare is the firm's investment behavior.

[39]Equivalently, it lowers the value of F_{wait} and has no effect on F_{invest}.

[40]Note that if the regulated firm delays investment and demand subsequently grows, then it will have to ration demand in the second period, not selling to customers in excess of capacity. Once the expansion is complete in the third period, it will be able to meet demand. Thus, while the price cap will induce the firm to invest earlier, it also creates the possibility of quantity rationing.

[41]If the price is capped at a high level, it will not bind in any state, and thus will have no impact on investment timing. If it is set at a moderate level, it has the effect described above. If set at a low level, then it may raise the

22

The price cap accelerates investment in this example because it has its biggest impact when demand is high. However, regulatory lags make such schemes too difficult to implement if the regulator tries to set all prices itself — demand conditions will likely change before the required regulatory hearings are finished. Price caps provide one means of implementing the required state-dependent prices that does not suffer from a regulatory lag. Direct price-setting, in contrast, is a blunt regulatory instrument. This suggests that price caps will be more useful in industries with rapid technological change, volatile demand, and other factors that make regulatory lags especially costly.

So far there has been little empirical research that isolates the effect of price flexibility on investment timing. One possibility is to compare the investment behavior of firms subject to price caps and rate case moratoria (in which tariffs are fixed at current levels for a specified period) during times of low inflation, because then the principal difference between the two regimes is the flexibility to rebalance prices under the former. Greenstein et al. (1995) provide such a comparison in their study of US telecommunications firms' network modernization programs during the period 1986–1991. They find little difference between the modernization programs of firms subject to price cap regulation and those operating subject to a rate case moratorium, suggesting that greater price flexibility has little impact on investment in network modernization.[42]

There has also been relatively little theoretical research into the impact on investment behavior of allowing the firm some degree of price flexibility.[43] Many authors examine investment behavior under, for example, price cap regulation, but the role of intertemporal price flexibility is not crucial to these analyses. While the topic would seem to be understudied, perhaps price flexibility does not have much direct impact on investment behavior. Even if this is so, price flexibility undoubtedly has a significant indirect impact — greater price flexibility allows price regulation to coexist alongside competition, and it allows regulatory hearings to be held less frequently. Section 7 will show that forcing the regulated firm to face competition significantly alters investment behavior and, as the following section shows, holding hearings less frequently can enhance investment efficiency. Thus, price flexibility has, at least, an indirect impact on regulated firms' investment behavior.

---

right side of (6) as well, and perhaps even lead to later investment. This non-monotonic relationship is apparent from Dobbs' (2004) model.

[42]However, as described in Section 3.2, rate case moratoria often come with conditions on investment, so it is possible that drawing this interpretation from the work of Greenstein et al. would be to make the mistake of comparing voluntary investment under a price cap with mandated investment under a rate case moratoria.

[43]Perhaps this is because dynamic, stochastic models such as Dobbs (2004) are needed to reveal the investment implications. Dobbs (2004) is an excellent example of the sort of research this survey aims to promote.

The second characteristic of a regulatory scheme described in Section 2 is the rule determining the timing of regulatory hearings. This section describes various ways in which this rule influences the firm's investment behavior. Section 5.1 shows how holding hearings at fixed, and relatively infrequent, dates — as is usually the case under modern incentive regulation — can lead to more efficient investment behavior. However, Section 5.2 shows that a sub-optimal periodicity in investment timing results. Allowing the firm to determine the timing of hearings also distorts investment, as shown in Section 5.3.

5.1   *Allocating the benefits of cost reduction*

Consider the problem faced by a firm that is considering making an irreversible investment that, if successful, will result in lower operating costs.[44] The firm benefits from lower operating costs, and therefore greater net revenue, until the next regulatory hearing. The long-term consequences of the investment depend on what happens at the hearing.[45] Suppose that a proportion α of cost savings is passed onto consumers after the next hearing, when a proportion γ of the firm's investment expenditure is added to the firm's rate base. (In the case of traditional rate of return regulation, $\alpha = \gamma = 1$; pure incentive regulation corresponds to $\alpha = \gamma = 0$, since all cost savings are retained by the firm, and the firm's actual investment does not influence the regulator's chosen rate base.) If the investment costs I and permanently lowers operating costs by ΔC, then the firm's incremental cash flows resulting from the investment are $-I$ immediately, ΔC per period until the next hearing, and $r\gamma I + (1-\alpha)\Delta C$ per period afterwards, where r is the discount rate. If the next hearing occurs after T periods, then the present value of the gain to the firm resulting from the investment is

$$-I + \sum_{t=1}^{T} \frac{\Delta C}{(1+r)^t} + \sum_{t=T+1}^{\infty} \frac{r\gamma I + (1-\alpha)\Delta C}{(1+r)^t} = \frac{\Delta C}{r}\left(1 - \frac{\alpha}{(1+r)^T}\right) - I\left(1 - \frac{\gamma}{(1+r)^T}\right). \quad (7)$$

In contrast, consumers gain nothing in the short-term, but face an altered output price after the next hearing: the lower operating costs reduce the cost to consumers by $\alpha\Delta C$, but the greater

---

[44]The early literature on regulatory lag typically ignored the irreversibility of investment. Examples include Davis (1973) and Elton and Gruber (1977), who use partial adjustment models in which a short regulatory lag is represented by more rapid adjustment, and Bawa and Sibley (1980) and Brennan and Schwartz (1982b), who develop dynamic models in which the timing of regulatory hearings is stochastic.

[45]The early literature on this subject contains two divergent views. For example, Baumol and Klevorick (1970) recognize that the firm receives the benefits of investment during the period up until the next hearing, but none afterwards. In contrast, in Bailey and Coleman (1971) the firm incurs a loss until the next hearing, after which the benefits of investment can be received.

rate base raises the cost by $r\gamma I$. The present value of the gain to consumers is therefore[46]

$$\sum_{t=T+1}^{\infty} \frac{\alpha \Delta C - r\gamma I}{(1+r)^t} = \frac{\Delta C}{r} \left(\frac{\alpha}{(1+r)^T} \right) - I \left(\frac{\gamma}{(1+r)^T} \right).$$

The total gain from the operating cost-reducing investment, $\Delta C/r - I$, is shared by consumers and investors: when α is high, consumers receive a larger share of the cost savings; when γ is large, they bear a larger share of the investment outlay; when hearings are held more frequently (T is smaller), more of each component is allocated to consumers. Thus, the regulator has three different policy instruments that it can use to allocate the surplus between the firm and its customers. The discussion in this section focuses on the time T between hearings, while the implications of α and γ are discussed in Section 6.[47]

Provided T is positive and $\alpha = \gamma$, the firm will invest if and only if the project is socially optimal. At one extreme, rate of return regulation achieves this by passing all costs onto customers ($\alpha = \gamma = 1$). In this case, raising T increases the share of the surplus allocated to the firm, giving it a greater incentive to invest in cost reduction.[48] At the other extreme, pure incentive regulation achieves this by setting prices independently of actual costs ($\alpha = \gamma = 0$). More generally, the firm will only invest if

$$\frac{\Delta C}{r} \geq I + \frac{(\alpha - \gamma)I}{(1+r)^T - \alpha}. \tag{8}$$

In particular, the firm will not invest in some socially-optimal cost-reducing projects if $\alpha > \gamma$.[49] At first glance, therefore, it appears that efficient investment results if actual costs (both operating and capital) are passed onto consumers after a lag. However, as Section 5.2 will show, this will not generally be the case.

In some situations it is not possible for operating cost-reducing expenditure to be included in the rate base. Consider, for example, a one-off investment of 'effort' that permanently reduces operating costs but cannot be observed by the regulator — it can observe the actual operating costs incurred by the firm, but not the investment expenditure. In effect, the regulator must set

[46]This assumes that demand is unaffected by the price change. More generally, the altered output price will affect demand, and therefore add an additional term. See Bailey (1974), who examines the effect of T on the quantity of investment in the special case where $\alpha = 1$ and $\gamma = 0$; that is, where all cost benefits are passed onto consumers at the next hearing, but the rate base is not adjusted to reflect the firm's investment.

[47]Baumol (1967) and Myers (1972) are early advocates of the use of a fixed period between formal hearings as a means of inducing socially-advantageous investment by the regulated firm.

[48]However, as explained in Section 2.3.3, under rate of return regulation holding hearings less frequently raises the firm's cost of capital, implying higher regulated prices.

[49]Mayo and Flynn (1988) examine the R&D expenditure of investor-owned US electric utilities in the years 1975 and 1983. Consistent with (8), they find that R&D expenditure is higher when the allowed rate of return (observed at the most recent regulatory hearing) is high, and when R&D expenditure is included in the rate base; that is, high values of γ increase investment.

$\gamma = 0$. If the firm is allowed to keep the full operating cost reduction ($\alpha = 0$) then it will exert the socially-optimal amount of effort to reduce operating costs: this is one motivation for price cap regulation. In contrast, if the full operating cost reduction is passed onto customers ($\alpha = 1$), then the firm only undertakes the effort to reduce operating costs if the present value of the savings over the remainder of the current regulatory cycle exceeds the cost. Under-investment in cost reduction results. The situation is more complicated when the lower operating costs last only as long as effort is exerted, because once the effort has been exerted, the regulator realizes the potential for cost reduction and can pass the gains onto customers in the form of lower prices. If the regulator does so, it will permanently reduce the output price at the first hearing after the firm starts exerting effort to reduce operating costs. Thus, the firm only benefits from the reduced operating costs for the remainder of the current regulatory cycle, but it will incur the effort costs indefinitely. This also leads to under-investment in cost reduction.[50]

5.2 Implications for investment timing

The discussion in Section 5.1 assumes that investment is of the now-or-never variety. If the firm can choose the timing of its investment program, then the timing of the investment will be affected by the regulatory process.[51] For simplicity, consider the special case of (7) where all costs are passed onto consumers at the next hearing; that is, $\alpha = \gamma = 1$ and the firm's net present value if it invests T periods before a scheduled hearing equals

$$\left(\frac{\Delta C}{r} - I \right) \left(1 - \frac{1}{(1+r)^T} \right).$$

Since this is increasing in T whenever $\Delta C > rI$, the investment payoff is maximized by investing immediately after a hearing, as this maximizes the length of time during which shareholders enjoy the benefits of lower operating costs. Thus, some investments will be delayed until after an imminent hearing, while others will be brought forward in time. For example, suppose that hearings occur every T years and that the next one will occur t years in the future. The firm should invest immediately if and only if

$$\left(\frac{\Delta C}{r} - I \right) \left(1 - \frac{1}{(1+r)^t} \right) > \frac{1}{(1+r)^t} \left(\frac{\Delta C}{r} - I \right) \left(1 - \frac{1}{(1+r)^T} \right),$$

since the left side is the net present value from investing immediately and the right side is the net present value from waiting and investing immediately after the next hearing. This inequality holds if and only if

$$t > \frac{\log \left(2 - \frac{1}{(1+r)^T} \right)}{\log(1+r)}.$$

---

[50] For a simple formulation of this 'ratchet effect' in a general setting, see Weitzman (1980).

[51] Biglaiser and Riordan (2000) model this formally.

Thus, investment is likely to occur in the first few years of the regulatory cycle, but not in the later part of each cycle.

A similar result holds when hearings are triggered by high realized rates of return. If the firm is already operating close to the upper bound on its rate of return then it may not invest in reducing operating costs, since it will not get to keep the benefits for long. Instead, the firm will wait and invest when the actual rate of return is lower, since this allows it to keep the benefits of reduced operating costs for longer. As in the case of hearings held at fixed dates, the firm also wants to invest early in the regulatory cycle; the difference is that the length of the cycle is endogenous.

Of course, the timing effect will be minor if construction times are much longer than typical regulatory cycles. On the other hand, if the regulator cannot observe the firm's cost function, the firm still has the option of delaying *implementing* the lower operating costs until after the next hearing. Such behavior is reminiscent of Sweeney's (1981) suggestion that regulated firms can maximize their value by implementing innovations that lower operating costs over multiple regulatory cycles; that is, rather than implementing all improvements at once, firms spread them out over time.[52]

In summary, holding hearings infrequently gives the firm the greatest incentive to invest in socially-optimal cost-reducing investments, since this awards the firm a relatively high proportion of the social benefits, but increases the distortion of investment timing. At one extreme, when hearings are held frequently, firms have little incentive to invest; at the other extreme, when hearings are held infrequently, firms have (i) a strong incentive to invest, and (ii) a strong incentive to invest immediately after a regulator hearing.

5.3   Hearing timing flexibility

If the regulated firm can choose the timing of formal hearings, it possesses a potentially valuable risk-shifting device that can affect its investment behavior.[53] Suppose, for example, that the cost of the project described in Section 5.1 is risky and that actual costs are passed onto customers at regulatory hearings (that is, $\alpha = \gamma = 1$). This allows the firm to shift some of the consequences of higher-than-expected project costs onto consumers (by requesting a regulatory hearing to raise output prices), while keeping all of the benefits of lower-than-expected project

---

[52]The specifics of the price-setting process can influence investment timing in other ways. For example, Pint (1992) argues that the common practice of using a 'test-year' to measure costs gives firms an incentive to over-invest in capital during such test-years. However, Pint's model exaggerates the distortions induced by the timing of price reviews because she assumes that investment is costlessly reversible — in her numerical simulations, the regulated firm increases its capital stock more than tenfold in the test year, before reducing it back to its previous level the following year. Such fluctuations in capital stock are clearly not feasible in practice.

[53]See Burness, Montgomery and Quirk (1980) for analysis of this issue.

costs for itself (by not requesting a hearing, and thereby leaving prices at their current levels). The regulated firm therefore has an incentive to take on excessively risky projects, in much the same way as limited liability creates a conflict between shareholders and bondholders (Galai and Masulis, 1976; Jensen and Meckling, 1976).[54]

To see how this works, suppose that the project will cost either I_l or I_h, each with probability $1/2$, and that the cost will not be known until the project is complete. Further suppose that $rI_l < \Delta C < rI_h$, so that the firm will only request a hearing if the project turns out to be high cost. Thus, if the project proves to be low cost, the firm's investment payoff is

$$V_l = -I_l + \frac{\Delta C}{r}.$$

If the project proves to be high cost, the firm requests a hearing and, after T periods have elapsed, prices are adjusted.[55] The firm receives the benefit of the reduced operating costs while it waits for the process to be completed, but then all costs are passed onto customers; from (7), the firm's investment payoff equals

$$V_h = \left(\frac{\Delta C}{r} - I_h \right) \left(1 - \frac{1}{(1+r)^T} \right).$$

It follows that the ex ante investment payoff is

$$E[V] = \frac{1}{2} V_l + \frac{1}{2} V_h = \left(\frac{\Delta C}{r} - E[I] \right) + \frac{1}{2(1+r)^T} \left(I_h - \frac{\Delta C}{r} \right).$$

In particular, the expected payoff to the firm exceeds the expected social payoff by an amount that reflects the value of the firm's option to shift downside risk onto its customers. A mean-preserving spread in the project's cost raises the value of this option without altering the project's expected social payoff. The option is less valuable when the processing lag T is large, since then the firm must bear a greater proportion of the downside risk. Awarding the firm this option leads to over-investment when project cost is uncertain — the firm will invest in any project for which

$$\frac{\Delta C}{r} > E[I] - \frac{1}{2(1+r)^T} \left(I_h - \frac{\Delta C}{r} \right).$$

In particular, the firm will invest in some projects for which the present value of cost savings is less than the expected cost of the project. The distortions will be greater in industries that have more volatile capital costs.

---

[54]The ability to file for a price increase protects the firm from cost overruns, inducing it to favor high-risk projects, while the threat of consumers pressuring for a price reduction limits the firm's ability to profit when projects come in under-budget, inducing it to favor low-risk ones. Which effect dominates will depend on the coordination costs facing consumers and their ability to observe the firm's cost structure.

[55]The regulatory 'processing lag' imposes a delay between a firm requesting a hearing and the hearing actually being concluded (Joskow, 1974). Based on a sample of all 96 US electric utility rate reviews during the period 1980–1984, Atkinson and Nowell (1994) report that on average 261 days elapse between a firm filing for a rate review and the regulator approving a new set of prices. The regulatory lag had a standard deviation of 92 days.

This over-investment can be mitigated somewhat by including avoided, rather than actual, costs in the rate base.[56] Suppose that there exists an alternative project capable of achieving the same reduction in operating costs, but with certain cost I_f satisfying $I_l < I_f < I_h$. Therefore, if the firm requests a hearing after investing in the risky project, the rate base will be increased by I_l if the project turns out to be low cost, and by I_f if it turns out to be high cost.[57] That is, an amount equal to the lowest cost, ex post, of achieving the reduced operating costs is added to the rate base. Therefore the firm will not request a hearing if the project turns out to be low cost, so that the investment payoff remains

$$V_l = -I_l + \frac{\Delta C}{r}.$$

Assuming $rI_f > \Delta C$, the firm will request a hearing if the project turns out to be high cost (since then the increase in prices due to the rate base change dominates the reduction in prices due to the operating cost savings), so that the investment payoff becomes[58]

$$V_h = -I_h + \sum_{t=1}^{T} \frac{\Delta C}{(1+r)^t} + \sum_{t=T+1}^{\infty} \frac{rI_f}{(1+r)^t} = \left(\frac{\Delta C}{r} - I_h\right)\left(1 - \frac{1}{(1+r)^T}\right) - \frac{I_h - I_f}{(1+r)^T}.$$

The last term reflects the fact that the firm must now bear some of the consequences of cost overruns — specifically, it cannot pass capital costs in excess of the (ex post) cheapest project onto customers. The firm invests provided $E[V] > 0$, which is equivalent to the condition that

$$\frac{\Delta C}{r} > E[I] - \frac{1}{2(1+r)^T}\left(I_f - \frac{\Delta C}{r}\right). \tag{9}$$

The extent of over-investment has fallen because the use of an avoided cost standard has reduced the value of the firm's hearing timing option (I_h has been replaced with I_f on the right side of (9)).[59]

The discussion in this section reveals two ways in which the implementation of price cap regulation can improve investment efficiency. First, the automatic inflation adjustment makes it easier for the regulator to commit to less frequent hearings, so that more of the benefits of cost-reducing investment will be allocated to shareholders, giving firms a greater incentive to undertake such investment. Second, if price caps are reviewed at fixed dates, distortions induced by endogenous hearings will be reduced. Offsetting these gains, the length of the regulatory cycle raises the possibility that the timing of investment is adversely affected.

---

[56]See Lyon (1991, 1995) for a model of this policy.

[57]The same outcome will be achieved if the regulator pre-approves expenditure of no more than I_f on the project.

[58]If $rI_f \leq \Delta C$, the firm will not request a hearing, so that it bears all of the upside and downside risk, and the over-investment problem is eliminated. However, customers then receive none of the project's benefits.

[59]Consistent with this, several US electric utilities cancelled partly-completed and over-budget nuclear power plants in the 1980s, in the wake of large cost disallowances for cost overruns on other completed plants (Pindyck, 1993).

6   What costs will the firm be allowed to recover at the next hearing?

The final characteristic of a regulatory scheme described in Section 2 is the cost measure used by the regulator when setting prices and other regulatory parameters. The key issue is whether or not the regulator can commit to allowing the firm to recover the cost of its sunk capital. Section 6.1 describes the adverse consequences for investment when the regulator cannot commit, while Section 6.2 describes some ways in which these undesirable outcomes can be avoided. Finally, Section 6.3 considers the implications for investment if the regulator can commit to using the various cost measures described in Section 2.3.2.

6.1   Regulatory opportunism

When assessing different regulatory schemes, the usual welfare measure is the sum of consumer and producer surplus. However, regulators will not necessarily adopt this as their objective function. They are subject to pressure from the managers of the firms they regulate, their customers, and other groups, and a regulator will raise its utility by reducing the pressure exerted by these groups. It is through this process that a regulator's preferences reflect the interests of the various pressure groups.[60] It is usual to suppose that the regulator maximizes some function of consumer surplus and producer surplus such as

$$U = CS + \alpha PS, \tag{10}$$

where $\alpha \leq 1$ is a nonnegative constant, CS is consumer surplus, and PS is producer surplus. The relative influence of customers and firms will determine the size of α. Bower (1981) suggests that regulators (especially at state level) will attach a relatively low weight to producer surplus since they focus on the interests of individuals within their jurisdiction and firms are more likely to have foreign owners than foreign customers. Offsetting this influence is the possibility that regulators' post-retirement employment opportunities, perhaps as consultants to firms that they previously regulated (the so-called 'revolving door'), induce them to weigh the firm's profitability more heavily (Che, 1995; Salant, 1995). Regulators' preferences are also influenced by the mandates chosen by legislative authorities. If members of a legislature are dependent on a firm's customers for electoral support, then the result may be regulation that attaches greater weight to consumer surplus than to producer surplus (Baron, 1988); if they are dependent on the firm for campaign funding, regulation may attach greater weight to producer surplus.

---

[60]For example, Evans and Garber (1988) model the preferences of a regulator subject to pressure from the regulated firm's management and the firm's customers, and allow for the prices of some goods to be more politically sensitive than others. They derive restrictions that the regulator will impose on the firm in order to maximize the regulator's utility.

When $\alpha < 1$ in (10), the regulator has an incentive to transfer surplus from producers to consumers. The simplest way to effect such a transfer is to set prices that are just sufficient to induce the firm to carry on in business, since then consumers do not need to pay the cost of the firm's sunk assets.[61] If prices are set any lower, the firm will exit the industry and future surplus is lost. If prices are set any higher, then the regulator is not maximizing its objective function — some surplus could be transferred from the producer to the consumer, where it is valued more highly by the regulator.

The most extreme case, which is commonly discussed in the literature, occurs when the regulator can make no commitment to the firm regarding recovery of sunk costs. As a result, the regulator will change the firm's output prices immediately after the firm invests. The regulator will fund inputs which have not yet been sunk, but will not fund those inputs (such as existing physical capital) that have already been sunk. Anticipating this behavior, which prevents it from recovering its sunk costs, the firm will not invest in the first place (Blackmon and Zeckhauser, 1992).

The situation is not quite so pessimistic when the regulator can commit to holding prices fixed for a finite period of time, corresponding to the time remaining in the current regulatory cycle, but cannot commit to what happens at the next hearing. For example, suppose the regulator has set output prices that will remain fixed for T periods. The firm invests in order to maximize its market value, which equals

$$F_0 = -I_0 + \sum_{t=1}^{\infty} \frac{R_t - C_t - I_t}{(1+r)^t} = -I_0 + \sum_{t=1}^{T} \frac{R_t - C_t - I_t}{(1+r)^t} + \frac{1}{(1+r)^T} \sum_{t=T+1}^{\infty} \frac{R_t - C_t - I_t}{(1+r)^{t-T}}. \quad (11)$$

An opportunistic regulator will reset prices at the lowest possible level that just induces the firm to participate; that is, at the next hearing, the present value of future revenue will exceed the present value of future expenditure by an amount equal to the salvage value S_T of the firm's assets at the time:

$$\sum_{t=T+1}^{\infty} \frac{R_t}{(1+r)^{t-T}} - \sum_{t=T+1}^{\infty} \frac{C_t + I_t}{(1+r)^{t-T}} = S_T.$$

Assuming the firm cannot influence demand, using this expression to simplify (11) shows that the firm will invest in order to minimize

$$I_0 + \sum_{t=1}^{T} \frac{C_t + I_t}{(1+r)^t} - \frac{S_T}{(1+r)^T}.$$

---

[61] Another way to achieve this is to allow entry into the industry. This is especially attractive to the regulator when a scheme of price cap regulation is already in place, since then the regulator has no obligation to ensure the firm earns a reasonable rate of return on its sunk capital. Under traditional rate of return regulation the regulator would have to allow the firm to increase its prices in markets where it does not face competition to compensate for the lower prices forced by competition in markets opened to entry (Lehman and Weisman, 2000).

If investment is completely irreversible, so that the salvage value is zero, the firm simply minimizes the present value of expenditure over the life of the current regulatory contract, subject to common carrier obligations and any other restrictions on its investment policy. This illustrates the key problem — a lack of regulatory credibility induces myopic behavior by the firm. The resulting distortions can be severe. For example, the firm has a strong incentive to delay cost-reducing investment until the next regulatory cycle, when the firm will be compensated for its expenditure; if the investment occurs now, the benefits of lower operating costs will be passed onto customers at the next hearing, but the firm will not be compensated for the sunk capital. If the firm does invest, it will favor a series of sequential investments over a single larger, cheaper investment if doing so allows the firm to delay some investment until after the next hearing. Thus, the prospect of regulatory opportunism means that the firm will not fully exploit economies of scale in investment. Similarly, the firm will favor reversible rather than irreversible investment, since the resulting capital will have a higher salvage value. More generally, the firm will favor projects that require low sunk costs at the expense of greater ongoing, and therefore avoidable, operating costs (Blackmon and Zeckhauser, 1992). These responses all deviate from the cost-minimizing (that is, socially-optimal) investment program.

6.2   Limiting regulatory opportunism

If the regulator cannot commit to any particular future regulatory policy, it can choose only from those policies that are ex post optimal; that is, policies that do not compensate the regulated firm for its sunk costs. The socially-suboptimal investment behavior described in Section 6.1 results. Thus, welfare will generally be enhanced if the regulator can be prevented (or at least discouraged) from acting opportunistically.

The extent to which a regulator can transfer surplus from a firm to its customers is determined by the legal system within which it operates. For example, in the US judicial precedents have established that a regulated firm is entitled to a "fair return upon the value of that which it employs for the public convenience."[62] This "fair" return is not applied to all of the firm's assets, just those that are used to meet demand. Exactly what constitutes a "fair" return is not specified. In fact, the US Supreme Court has shied away from determining exactly what is a fair rate of return. Its most decisive ruling is that "...the return to the equity owner should be commensurate with returns on investments in other enterprises having corresponding risks. That return, moreover, should be sufficient to assure confidence in the financial integrity of the enterprise, so as to maintain its credit and to attract capital".[63] Combined with the

---

[62]From the US Supreme Court Decision of Smyth v. Ames, 169 US 466 (1898).

[63]From the US Supreme Court Decision of Federal Power Commission et al. v. Hope Natural Gas Co., 320 US 591 (1944, p. 603).

used-and-useful condition, this limits the ability of the regulator to reduce allowed returns.

However, Levy and Spiller (1994) argue that for regulation to be credible it is necessary to constrain the action of both the regulator and the government (to stop it changing the regulatory framework). Different countries have approached this in different ways. In the US, the separation of judiciary from government, well-developed regulatory procedures, and the Constitution combine to restrict opportunism. In Chile, the pricing formula is actually specified in legislation, and the fragmented legislature makes it difficult to change the law. In the UK, legislation can be changed easily with a change in government, so that firms' rights are specified in contracts that are upheld by an independent judiciary. However, in transitional economies, these commitment-enforcing possibilities may not be available, so that the threat of opportunism will potentially discourage investment.[64]

Even if regulators cannot be prevented from behaving opportunistically, such behavior can be made more costly. One possibility is to raise the cost to individual staff of regulatory authorities by allowing regulators to work (later) for firms that they have previously regulated — that is, allow the 'revolving door' to operate. This increases the long-term costs to a regulator of transferring surplus to consumers without affecting the short-term gains. For example, Salant (1995) shows that allowing managers and regulators post-retirement opportunities, in the form of consulting income, that require ongoing cooperation between firms and regulators allows some socially-optimal projects to be undertaken that would not otherwise be possible. However, one cost of allowing such post-retirement opportunities is that they can lead to firms investing in projects that are not socially optimal; that is, revolving doors can encourage over-investment.

Another approach is to make the transfer of surplus from the firm to its customers more difficult. For instance, when a regulator cannot make binding commitments about the use of its policy instruments (specifically, the extent to which it audits the regulated firm's costs), complex administrative procedures that make it costly for the regulator to gather such information can act as a commitment device. That is, by making its own information-gathering costly, the regulator can effectively commit to not using such information ex post (Sappington, 1986). Similarly, commitment is enhanced when multiple regulators are charged with regulating a firm. Although this can result in (static) allocative inefficiency, it can improve regulatory commitment by making renegotiation more difficult (Martimort, 1999). In countries where in-

---

[64]For economies in transition, this suggests that privatization might involve selling utilities to foreign government-owned firms since the government owners will be in a stronger bargaining position than a (diffuse or concentrated) non-government shareholding. Anticipating this stronger position, and the resulting lower likelihood of regulatory opportunism, such government-owned firms will be able to offer higher prices during the privatization process. This might explain why all privatized Hungarian electric utilities were sold to powerful electric utilities (many of them government-owned) in neighboring countries (Newbery, 1999, pp. 66–72).

frastructure firms are government-owned, the cost of opportunistic behavior can be raised by encouraging widespread ownership of utility stocks by a process of privatization, since then voters will pressure governments not to reduce the value of their investments through opportunistic regulatory behavior. Widespread initial ownership can be achieved by underpricing the shares when firms are privatized and by placing barriers that discourage investors from subsequently reducing their shareholding. Examples of such barriers include preferential tax treatment of dividends and other benefits received by long-term shareholders (Biais and Perotti, 2002; Schmidt, 2000).

In addition, the regulated firm itself can raise the cost to the regulator of transferring its surplus to customers. In particular, trigger strategies can deter regulators from acting opportunistically, and support an equilibrium in which the regulator compensates the firm for its sunk capital. This is achieved by the firm threatening to stop investing if the regulator does not allow the firm to recover its cost of capital (and the regulator threatening to not allow the firm to recover its cost of capital if the firm does not invest adequately). The regulator compares the short-term gains from reducing prices and the long-term consequences of the firm ceasing investment. The firm faces a similar trade-off. These two threats are often sufficient to prevent the players from acting opportunistically (Gilbert and Newbery, 1994; Salant and Woroch, 1992).[65]

Legal restrictions on regulatory behavior, such as a requirement that all assets that are used-and-useful be included in the rate base, enable equilibria that exclude opportunistic behavior to exist in a wider range of economic environments. They restrict the gains that the regulator can make from deviating — even after a deviation, the regulator must pay a return on capital that is being used. However, restricting the regulator too much can make these equilibria more difficult to achieve because they reduce the regulator's ability to punish the firm for under-investing — if the regulator is forced to compensate the firm for all its sunk capital, whether or not it is used-and-useful, then the firm has little disincentive to under-invest.

Whether or not such an equilibrium can be achieved depends on the properties of the industry being regulated, as the ability of a regulated firm to deter opportunistic behavior by the regulator varies across industries. For example, the rapid growth in demand in telecommunications, together with rapid technological change, means that firms will have to undertake investment

---

[65]Models of trigger strategies supporting a cooperative equilibrium often assume that both parties (the regulator and the firm's manager, in this case) have preferences over the outcome in all infinitely-many future periods. However, Salant (1995) shows that some such equilibria can be achieved if both parties have finite terms, provided that those terms do not coincide. Equilibrium investment then follows a cycle, with the firm investing early in the careers of its successive managers, and capital recovery being allowed early in the careers of the firm's successive regulators. However, the finite terms of managers and regulators have a price — not all socially-optimal projects will be undertaken.

in the future, strengthening their threat to respond to opportunistic behavior by withholding investment. In contrast, similar threats made by electric utilities operating in eastern Europe would not have the same force, since the combination of reduced industrial output during transition and the potential for greater energy efficiency means that relatively little investment is needed in the foreseeable future (Newbery, 1999, p. 64).

The investment behavior of a firm facing a regulator behaving opportunistically, as described in Section 6.1, may also be appropriate for a firm facing a regulator who is not (currently) behaving opportunistically. Under-investing, for example, strengthens the firm's position if it ever needs to implement its punishment strategy.[66] Similarly, slowly increasing capacity, using a technology with a high salvage value, or one with a rapid rate of depreciation, all strengthen the firm's threat, and therefore help to support an equilibrium in which the regulator does not behave opportunistically. In fact, anything which requires ongoing expenditure will enhance the firm's threat. The key observation is that the *possibility* of opportunism will affect investment even when the possibility is not realized.

6.3 Investment when the regulator can commit to capital cost recovery

Section 6.1 showed what happens when the regulator does not compensate the firm for its sunk investments. The current section focuses on the situation where the regulator compensates the firm for its past investments, using some measure of the cost of its sunk investment, either the firm's actual cost or some benchmark measure of cost. The central issue is how the regulator responds to information revealed after the firm's investment decision is made. The discussion is separated into two parts, each focusing on a different type of information. Section 6.3.1 considers the response of the regulator to shocks that affect the project's outcome, such as demand shocks, while Section 6.3.2 deals with shocks that do not affect the built project directly, but which nevertheless affect any ex post assessment of the project.

The key ideas can be illustrated using a simple binomial model in which the firm can invest in a project, either immediately (period 0) or next period. If the firm invests in the project, a regulatory hearing is held in period 2. Investment is irreversible and can be undertaken at most once. Construction takes one period, after which the project generates a perpetual cash flow (which might derive from reduced operating costs, additional demand serviced by the asset, and so on). The project's outcome depends on the state of nature, which is revealed in period 1. The project generates cash flow of x_h in the high-payoff state and x_l in the low-payoff state, and each state occurs with probability 1/2. Prior to the hearing the firm receives the full cash flow

---

[66]The firm can add credibility to its threat to punish the regulator for deviating from the equilibrium strategy by issuing debt — highly levered firms will be more willing to punish opportunistic behavior by regulators since shareholders make the decisions but enjoy limited liability.

of x_i for $i \in \{h, l\}$. At the hearing, the firm's rate base is increased by an amount B_i^n, where n is the period in which the firm invested and $i \in \{h, l\}$ distinguishes the high-payoff and low-payoff states. The project costs I_0 to build in period 0 and $I_1 \in \{I_h, I_l\}$ to build in period 1, depending on which payoff state occurs. Finally, suppose that the model parameters are such that if the firm delays the investment decision until period 1, its value is maximized by only investing in the high-payoff state.

The payoff from immediate investment equals

$$F_{\text{invest}} = -I_0 + \frac{E[x_i]}{1+r} + \frac{E[x_i + B_i^0]}{(1+r)^2},$$

since the project generates additional cash flow of x_i for the firm in periods 1 and 2, and rB_i^0 in each subsequent period. Waiting one period to decide whether or not to invest generates the payoff

$$F_{\text{wait}} = \frac{1}{2} \left(\frac{-I_h}{1+r} + \frac{x_h + B_h^1}{(1+r)^2} \right),$$

since the firm invests in the high payoff state but not in the low payoff state. Immediate investment is therefore value-maximizing if and only if $F_{\text{invest}} > F_{\text{wait}}$. This condition holds if and only if[67]

$$E[x_i] + \frac{E[B_i^0 - B_i^1]}{1+r} + (E[I_i] - I_0) > rI_0 + \frac{1}{2} \left(I_l - \frac{x_l + B_l^1}{1+r} \right). \tag{12}$$

This condition reflects the benefits and the costs of investing early. There are three (potential) benefits from investing early, corresponding to the three terms on the left side of (12). First, the firm receives the cash flow x_i in period 1. Second, it 'locks-in' the rate base B_i^0, rather than the alternative, B_i^1. Third, it spends I_0 on the project, rather than I_i. Offsetting these gains are two costs of investing early, corresponding to the two terms on the right side of (12). First, the firm must finance the project one period sooner, costing rI_0. Second, the firm cannot reverse its investment decision if bad news (that is, the low-payoff state) eventuates.

6.3.1 Shocks to investment outcomes

Post-investment shocks can make a good investment decision ex ante look like a poor decision ex post. For example, projects may not deliver the anticipated cost savings or generate the anticipated demand, or may not be needed due to changes in industry conditions that occur during

---

[67]Notice that when viewed in a static setting (that is, when investment is a now-or-never proposition), the firm's investment behavior depends only on the expected values of x_i and B_i^0. When the firm's investment timing options are included, its investment behavior depends on both the expected values and variability of x_i and B_i^0. The role of volatility is thus important when considering the investment response of firms to different rate base rules.

construction. If regulators base their rate base calculations on such post-investment information, then the increment to a firm's rate base can be much smaller than the investment outlay. The firm is protected from this sort of hindsight review if regulators ignore post-investment information.

I start by considering the response of the regulator to shocks that affect the project's outcome, such as demand shocks. To focus on this issue, I assume that there is no construction cost risk (that is, $I_0 = I_1 = I$) and that the rate base increment does not depend on the investment date (that is, $B_i^0 = B_i^1 = B_i$). In this case the investment condition, (12), reduces to

$$E[x_i] > rI + \frac{1}{2}\left(I - \frac{x_l + B_l}{1 + r}\right). \tag{13}$$

If the regulator adds all assets to the rate base at their actual cost, so that $B_h = B_l = I$, then the cost of investing early is relatively low, because the only bad news that the firm may face is that the cash flows during the period until the next hearing are lower than expected; any bad news past that date is borne by customers. In contrast, if only the cost of used-and-useful assets is added to the firm's rate base, B_h and B_l will reflect the state of demand at the time of the hearing. In the case of the demand shocks considered here, the regulator would disallow some of the firm's investment expenditure in the low-payoff state as, with the benefit of hindsight, a smaller, cheaper asset could have been built. Compared to the actual cost case, B_l will be smaller, so that the right side of (13) will be higher — the used-and-useful criterion shifts risk from consumers back to the firm's shareholders. In particular, by placing more of the burden of bad news onto shareholders, it induces the firm to invest later than an otherwise identical firm that faces an actual cost rate base. More generally, firms that anticipate the possibility of cost disallowances will show greater caution when selecting projects that, while attractive ex ante, may not appear attractive to regulators at the next regulatory hearing. Firms will only invest when it is likely that, ex post, the project will be deemed used-and-useful; that is, they will favor projects that have a relatively low potential for bad news.[68]

When making their investment decisions, firms will impose a more stringent test on projects when shareholders are exposed to a greater risk of bad news. That is, some projects will be undertaken by a firm if their actual cost is added to the firm's rate base but not if ex post information is used in determining the rate base increment. For example, projects with long lead times are subject to a greater risk of negative demand shocks during the construction period. Such projects are less attractive than ones with shorter construction periods — and less

---

[68]Saphores et al. (2004) find similar behavior in their analysis of the investment decision of a firm facing a costly and time-consuming regulatory process with an uncertain outcome. They model environmental regulation, but their arguments apply in any situation when a regulator uses post-investment information when evaluating investment projects.

exposure to demand shocks while under construction — and so are less likely to be undertaken by firms with a rate base calculated according to the used-and-useful criterion (Teisberg, 1993, 1994). Similarly, these firms will be less willing to undertake projects that favor scale over flexibility, or that involve greater sunk costs relative to non-sunk ones, than firms not exposed to the risk of cost disallowances. In summary, the move away from pure rate of return regulation to a system in which much greater use is made of ex post information when making rate base decisions should have induced a change in the *type* of investment undertaken by firms.

This is consistent with anecdotal evidence from the US electric power industry — firms moved away from building large nuclear plants with long lead times to, for example, building combined cycle gas generation plants that have a smaller scale and shorter planning and construction lead times. Partly this change reflects technological advances in electricity generation, but it is also likely to have been motivated by the large cost disallowances experienced in the 1980s; the technological advances themselves might have been motivated by the cost disallowances. However, recent empirical research by Lyon and Mayo (2004) casts doubt on whether these disallowances were interpreted by the industry as signalling a change in regulatory regime. They analyze the impact of cost disallowances on 132 US electric utilities' investment behavior during the period 1970–1991 and find that firms that suffered a disallowance reduced their investment, while other firms in the same state only reduced their investment if they operated nuclear power plants. If a particular disallowance was interpreted as a change in regime (from guaranteeing recovery of all costs to guaranteeing recovery only of the cost of used-and-useful assets, for example), all firms in the same jurisdiction would be expected to reduce investment; if it was interpreted as simply the result of an anticipated ex post assessment of an investment decision made in the face of uncertainty, of the sort analyzed in this section, no firms would reduce investment; if, on the other hand, they believe that the disallowance is the regulator's response to imprudent investment, only the firm directly affected will reduce investment. Thus, Lyon and Mayo's evidence suggests that these disallowances were perceived as a regulatory response to imprudent investment, rather than as a change in regulatory regime.

6.3.2 *Shocks to projects not pursued*

We now turn our attention to shocks that do not affect the built project directly, but which nevertheless affect any ex post assessment of the project. For example, a negative shock to capital prices does not affect the project's actual outcome, but may lead a firm to regret investing early since it would have been able to invest at lower cost had it waited. The potential magnitude of such bad news can influence a firm's investment behavior when investment is irreversible, whether or not the firm is regulated, by altering the value of delaying investment. However,

regulation can provide a second channel through which such risk affects investment behavior — it can influence the regulator's ex post assessment of the completed project, and therefore the addition to the rate base approved by the regulator. More generally, if regulators allow firms to recover only those costs that their past investments allow the firm to avoid, then regulation exposes firms to the risk of shocks to the projects that they could have pursued, not just to the projects that were actually undertaken.

To illustrate these possibilities, I consider the effects of a regulator's use of historical and replacement cost rate bases on investment timing, using another special case of the model described at the start of Section 6.3.[69] In this case the project's construction cost is risky, with $I_h < I_0 < I_l$, so that the high-payoff state corresponds to a low construction cost. Suppose that at the hearing in period 2 the rate base is raised by the replacement cost of building the project; that is, $B_i^n = I_i$ regardless of the period n in which investment actually occurs. Then early investment is value-maximizing if and only if

$$E[x_i] + (E[I_i] - I_0) > rI_0 + \frac{1}{2}\left(I_l - \frac{x_l + I_l}{1+r}\right). \tag{14}$$

If the project's cost is expected to rise over time, the term in brackets on the left side is positive, reflecting the incentive the firm has to invest early, when the project is relatively inexpensive. The more positive the trend in project cost, the greater the incentive to invest early.

If the rate base is raised by the actual (that is, historical) cost of building the project, so that $B_h^0 = B_l^0 = I_0$ and $B_i^1 = I_i$, early investment is value-maximizing if and only if

$$E[x_i] + \frac{r}{1+r}(E[I_i] - I_0) > rI_0 + \frac{1}{2}\left(I_l - \frac{x_l + I_l}{1+r}\right). \tag{15}$$

The only change to the condition in (14) is that the coefficient on the expected change in construction cost is smaller, reflecting the fact that by valuing assets at their historical cost the regulator allows the regulated firm to 'lock-in' its rate base by investing early. If construction costs are expected to rise, comparison of (14) and (15) shows that this encourages the firm to delay investment and invest when the rate base will be higher, reducing the incentive to lower expected investment expenditure by investing early; that is, early investment is less attractive when an historical cost rate base is used than when a replacement cost rate base is used. In contrast, if construction costs are expected to fall, this encourages the firm to invest early while the rate base is relatively high, reducing the incentive to lower expected investment expenditure by investing later; that is, early investment is more attractive when an historical cost rate base is used than when a replacement cost rate base is used.

---

[69]See Evans and Guthrie (2005b) for a formal model of the different impacts of historical and replacement cost rate bases on investment timing.

6.3.3 The overall impact of different cost measures on investment timing

From Section 3.3, an unregulated monopolist generally invests too late from the point of view of overall welfare. A way of inducing earlier investment by the regulated firm is needed. How can a regulator raise the value of F_{invest} relative to F_{wait}?

One possibility is to shift the risk of shocks to project outcomes from the firm's shareholders to its customers, thereby reducing the value of the bad news term in (12), encouraging earlier investment. However, shareholders will find it easier than the firm's customers to diversify these risks, so that shifting risk onto shareholders is more sensible in terms of (static) risk allocation. Another possibility is to compensate the firm more for early investment than late investment. If capital prices are declining over time, this can be achieved by calculating the firm's rate base using the historical cost of its assets. Combinations of these policies are possible. For example, the firm could be allowed to recover the historical cost of only those assets that are used-and-useful. The overall effect of a particular rate base on investment timing depends on the nature of the industry being regulated. Evans and Guthrie (2005b) show that valuing the rate base at the historical cost of the assets can induce a regulated monopolist to invest sooner than an unregulated one when the project's cost is expected to fall over time and shocks to replacement cost and demand are not too strongly positively correlated. In the telecommunications industry, where volatility is high and capital prices are falling over time, using historical cost could potentially accelerate investment quite significantly, because of the magnitude of risks that would be shifted onto customers and of the greater incentive to invest early and ensure a relatively high rate base. In contrast, water utilities face far fewer risks and are not subject to the same degree of technological change, so that the choice of rate base will have a much smaller impact on investment behavior.

An alternative way to accelerate investment is to compensate the firm for the delay option destroyed when it invests early. This could be achieved by adding a real option premium to the historical cost of the firm's assets.[70] Several authors (notably, Hausman (1999) and Pindyck (2003)) have argued in favor of such a rate base, pointing out that when the firm invests its *total* cost is the sum of the investment expenditure and the value of real options destroyed by the act of investment. These authors have argued that the regulated firm should be compensated for this total cost, rather than just its investment expenditure, so that the rate base should include the (opportunity) cost of real options destroyed by investment.

However, regulators have moved in the opposite direction — shareholders are now exposed to more of the downside risk from investment, replacement cost is used increasingly to inform rate

---

[70]The issue typically arises in requests that the allowed rate of return equal a rate that exceeds the firm's cost of capital by an amount reflecting the value of real options destroyed when the firm invests.

base decisions, and regulators show no sign of accepting arguments that firms be compensated for the total opportunity cost of investment (such as the destruction of timing options). All of these decisions suggest slower investment by regulated firms. However, there has been one other important change — the relaxation of entry restrictions — that, in some circumstances, can accelerate investment. As I show in the next section, exposing a regulated firm to competition can promote investment in some situations, and discourage it in others.

7 Access regulation

Many regulated industries have a vertical structure containing some components that are natural monopolies and other components that are not. In the past, firms operating in such industries were typically vertically integrated and protected from entry in the potentially competitive segments. However, regulators have moved to encourage competition in segments that are not natural monopolies. For example, in the electricity industry, generation and retailing are typically opened up to competition, while transmission and distribution, which have stronger natural monopoly characteristics, continue to be regulated. In the telecommunications industry, the market for long-distance calls has been opened up to competition for many years. The wired local networks that connect customers have been regarded as natural monopolies in the past, although advances in telecommunications technology are likely to change that in the future.[71]

However, vertically integrated firms can potentially use their monopoly power in one segment to strengthen their competitive position in the segments opened to entry. For example, a vertically integrated firm may be able to strengthen its competitive position in an imperfectly-competitive downstream market by raising its rivals' costs (Salop and Scheffman, 1983). If the upstream price is unregulated, it can achieve this by charging its rivals a high price for the essential input; if the upstream price is regulated, it can potentially engage in non-price discrimination (by restricting access, lowering quality, or taking similar actions that effectively raise its competitors' costs).

Regulators have responded in two ways. One approach involves forced ownership separation, which prevents firms operating in monopoly segments from also operating in the sectors that are opened up to entry. The most famous example is provided by the AT&T divestiture in 1984, which separated the regional Bell operating companies from the manufacturing and long-distance components of AT&T and prevented them from subsequently entering either of these markets. Similarly, British Rail was split into almost 100 separate companies in 1994 and subsequently privatized. One firm, Railtrack, owned the track, and sold access to competing train operators

---

[71]See Joskow (2000) for more on the process of deregulating the US electric power industry and Economides (2005) for a recent assessment of the deregulatory process in the US telecommunications sector.

at regulated prices. The relevant issues when regulating the operator of the monopoly segment are similar to those discussed in Sections 4–6, but with consumers replaced by downstream firms.

If economies of scope make vertical ownership separation undesirable, an alternative approach to structural separation, which allows the firm to remain vertically integrated but opens all sectors other than the natural monopoly ones to new entrants, can be followed. The vertically integrated firm is required to provide access to its facilities in the monopoly segments to the entrants at regulated prices. The terms of access to the bottleneck facilities are typically regulated in an attempt to restrict the sort of behavior described above.[72] Again, perhaps the most famous example is in the US telecommunications sector. The Telecommunications Act of 1996 required incumbent local exchange carriers to lease bottleneck facilities under their control to competitors (known as 'local loop unbundling'), allowing the latter to offer customers advanced telecommunications services such as broadband access. The FCC set access prices based on the TELRIC cost measure described in Section 2.3.2. In return, the incumbents would be allowed to enter the long-distance market once there was sufficient competition in their local markets. A second example is provided by the FERC's attempts to reform the US electricity sector — electric utilities will be able to remain vertically integrated, but their transmission assets will be grouped together on a regional basis and operated by an independent operator that is required to set non-discriminatory transmission prices. However, many states have resisted the FERC proposal and this, combined with the difficulties in designing a transmission pricing regime that gives investors incentives to make efficient investments in much-needed transmission assets, has meant that progress has been extremely slow.

In this section I discuss the new issues that arise when regulating access to bottleneck facilities owned by a vertically integrated firm, concentrating on the implications of access regulation for equilibrium investment, especially investment timing.[73] I analyze the problem backwards through time, since the behavior of firms prior to the first investment in infrastructure will depend on what they anticipate will happen after investment. Therefore, the analysis starts in Section 7.1 by considering competition between firms, and how the terms of access regulation affect subsequent investment by the investment leader's rivals, after one firm has invested in infrastructure. Section 7.2 then considers the implications for competition prior to investment.

---

[72]Rate of return regulation is unsuitable when information asymmetries allow the firm to allocate some of the costs of operating in the downstream market to its operations in the upstream market. These costs would then be recovered from the regulated prices charged upstream, giving the firm a cost advantage in the competitive segments.

[73]Much of the access pricing literature has focused on the consequences of access prices for downstream competition, rather than on the implications for investment behavior. For example, see Armstrong (2002) and Vogelsang (2003) for surveys of access pricing in the telecommunications industry, where two-way access (in which rival firms require access to others' networks in order to provide services to their own customers) is also important.

Competition after the asset is built

It is helpful to think of the different forms of competition between a vertically integrated incumbent and a downstream entrant as lying along a continuum. At one extreme — pure service-based competition (SBC) — the entrant does not need to make an investment in physical capital in order to compete, instead using the incumbent's own assets. For example, energy retailers operating in deregulated electricity markets do not need to invest in electricity infrastructure: they can purchase electricity from generators and carry it to their customers by seeking access to the electricity transmission and distribution networks. At the other extreme — pure facilities-based competition (FBC) — the entrant builds its own facilities so that it can operate independently of the incumbent. Between these two extremes, the entrant uses some of the incumbent's assets as inputs into its own production process, but must also undertake its own investment in physical capital. The range of intermediate cases is exemplified by the local loop unbundling mandated by the 1996 Telecommunications Act, which forced incumbent local exchange carriers to lease unbundled *components* of their networks. By leasing various packages of these components, entrants could potentially engage in a wide variety of different forms of competition, requiring different levels of investment.

Entrants need to undertake a sequence of largely irreversible investments in order to move along the continuum. Investment will typically be delayed at each step due to the interaction of irreversibility and uncertainty. However, potential entrants have various starting points that affect their ability to enter; for example, cable television firms would find it relatively easy to enter their local telecommunications market since their existing facilities mean that less investment is required to enter. I will use a highly stylized example in this section, in which the entrant must sink I_{SBC} in physical capital in order to undertake SBC and an additional I_{FBC} in order to undertake FBC. I_{SBC} represents investment in complementary assets, such as long-distance networks for a firm seeking access to a local telecommunications network. If the incumbent invests at date T_0, then the entrant chooses the dates $T_S \geq T_0$, when it initiates SBC, and $T_F \geq T_S$, when it initiates FBC, in order to maximize its market value

$$F^{\text{Follower}}(T_S, T_F) = E\left[\frac{-I_{\text{SBC}}}{(1+r)^{T_S}} + \sum_{t=T_S+1}^{T_F} \frac{X_{\text{SBC}}^{\text{Follower}}}{(1+r)^t} - \frac{I_{\text{FBC}}}{(1+r)^{T_F}} + \sum_{t=T_F+1}^{\infty} \frac{X_{\text{FBC}}^{\text{Follower}}}{(1+r)^t}\right], \quad (16)$$

where $X_{\text{SBC}}^{\text{Follower}}$ and $X_{\text{FBC}}^{\text{Follower}}$ denote the net cash flow received by the entrant during the periods of SBC and FBC, respectively.[74,75] The strategy adopted, including the timing of investments

---

[74]Gans and King (2004) argue that investment is encouraged by so-called 'access holidays', which guarantee the incumbent an interval during which the terms of access will not be regulated. If the access holiday lasts H periods, then the entrant chooses T_S and T_F subject to the constraints $T_S \geq T_0 + H$ and $T_F \geq T_0$.

[75]Fixed investment dates will not generally be optimal in the presence of uncertainty. Rather, the entrant will

and the nature of the assets built, depends on the terms of access available. In particular, the regulator can influence I_{SBC} by specifying the level of access that the incumbent must offer potential competitors; it can influence $X_{\text{SBC}}^{\text{Follower}}$ by specifying the price that competitors must pay for such access.

The investment flexibility available to the entrant exposes the incumbent firm to considerable risk. After the incumbent has invested in an asset, entrants can wait to see if an innovation is successful before they enter. Therefore, the incumbent will find its profits reduced in good states by entrants seeking access. Risk-sharing is therefore asymmetric, with the incumbent bearing all of the down side risk, but the incumbent and entrants sharing the upside risk. However, the incumbent has made irreversible investments, and so must bear the opportunity cost of capital in both good and bad states. This asymmetry is compounded if the entrant can obtain access for short periods, since then it has the option to only operate in profitable periods (Hausman, 1999; Pindyck, 2003). In the absence of access regulation, access seekers would either have to commit to paying for access for long periods of time (so that they bear some of the risk of bad outcomes), or prices for short-term access would incorporate a premium for the risk imposed on the incumbent (so that they pay for the abandonment option granted them by the incumbent). The incumbent needs to earn a high ex post return on successful projects to compensate it for the ex ante risk that the projects will be unsuccessful. However, a regulator will find it difficult to allow such high ex post returns since the public only sees a successful project (Gans and King, 2004). From the discussion in Section 6.1, incumbent investment will be low when faced with such commitment problems.

There has been a fierce debate about the appropriate level of access prices, especially in the telecommunications sector.[76] A widely-held view is that, by raising $X_{\text{SBC}}^{\text{Follower}}$ in (16), a low regulated access price accelerates SBC but delays the access seeker making the investment required to enter into FBC (and thereby delays any associated welfare gains from lower congestion and any product variety that results from different platforms, for example); if prices are sufficiently high, then FBC may be the only feasible entry strategy.[77,78] It has even been argued that as

choose a stopping rule that determines the timing of investment as a function of the prevailing state of nature.

[76]For example, Hausman (1999) and Pindyck (2003) argue that TELRIC-based access prices in the US telecommunications industry do not compensate incumbents for the asymmetric risks they must bear under mandatory local loop unbundling.

[77]Woroch (2004) derives this result formally using a deterministic model of investment competition, but he does not explicitly model downstream competition. Hori and Mizuno (2005) obtain the same result when demand is stochastic and investment is irreversible.

[78]Bourreau and Doğan (2005) argue that an unregulated incumbent will set an access price that is too low from a social welfare point of view, as this induces the entrant to delay investing in its own facilities, which is to the incumbent's advantage. In their model, FBC (unlike SBC) allows the entrant to compete by offering a different quality product. As the quality differential strengthens the entrant's competitive position, the incumbent sets a

long as access prices are based on the costs of a hypothetical efficient firm, then an entrant will never build its own facilities, since the entrant's own costs cannot be less than those of an efficient provider.

However, the precise impact of the access price on the entrant's investment behavior depends on the form of downstream competition.[79] When the incumbent considers increasing its downstream sales, it anticipates reduced downstream sales by the entrant (by an amount, not necessarily one-for-one, that depends on the form of retail competition), and therefore reduced access revenue. For example, if the access seeker reduces demand by λ units when the access provider increases its own retail sales by one unit, then the upstream cost to the access provider of increasing its downstream sales by one unit equals $(1-\lambda)c_u^i + \lambda a$, where c_u^i is the incumbent's upstream marginal cost and a is the access price. This is the sum of the cost of producing the net increase in output, $(1-\lambda)c_u^i$, and the opportunity cost of the lost access revenue, λa. As a result, the incumbent acts as though its upstream cost is a weighted-average of its actual cost and the access price — the greater the 'displacement ratio' λ, the greater the weight attached to the access price. Depending on the level of access prices and the form of downstream competition, this may allow the entrant to reduce the incumbent's upstream cost advantage by seeking access. Suppose the entrant's upstream marginal cost is c_u^e if it does not use the incumbent's facilities. Then seeking access reduces the entrant's upstream cost differential if $c_u^e - c_u^i$ is greater than $a - ((1-\lambda)c_u^i + \lambda a)$, which occurs if $c_u^e > \lambda c_u^i + (1-\lambda)a$. That is, when contemplating seeking access the entrant compares its own upstream cost with a weighted-average of the incumbent's upstream cost and the access price. When the displacement ratio is high, the weight on the incumbent's upstream cost is high, so that the level of the access charge has little impact on the comparison.[80]

7.2 Competition before the asset is built

I now focus on the impact of access regulation on the access provider's investment behavior, especially the timing of its investment.[81] The analysis is complicated by the strategic interaction

---

low access price, making FBC relatively unattractive to the entrant.

[79]The discussion in this paragraph is based on Sappington (2005).

[80]For example, in the special case with Hotelling competition downstream (where $\lambda = 1$), the entrant engages in SBC if it has an upstream cost disadvantage (thereby eliminating the incumbent's advantage) and in FBC if it has a cost advantage (thereby maintaining its cost advantage). Allowing access at any price (that allows both firms to be financially viable) induces the efficient make-or-buy decision by the entrant.

[81]Regulatory economists have followed two different approaches here. The first extends work by Fudenberg and Tirole (1985) and Katz and Shapiro (1987), who use deterministic models to analyze innovation timing games between firms; their results have recently been extended by Hoppe and Lehmann-Grube (2005). Examples of this strand of the regulatory economics literature include Riordan (1992), Gans (2001), and Woroch (2004). The second approach applies real option theory, especially the model of duopolistic competition described by Dixit

between firms, which may or may not be competing to be the first to invest in facilities.

In the stylized model of entrant behavior in Section 7.1, there are three distinct phases after the leader invests: if the leader invests at date T_0, it earns (regulated) monopoly profits of X_{Mono} between dates $T_0 + 1$ and T_S, when the follower enters into SBC; these profits fall to $X_{\text{SBC}}^{\text{Leader}}$ during the period of SBC, from $T_S + 1$ until the follower enters into FBC at T_F; during the period of FBC, from date $T_F + 1$ onwards, the leader earns profits of $X_{\text{FBC}}^{\text{Leader}}$. Thus, the leader's market value at date 0 is

$$F^{\text{Leader}}(T_0) = E\left[-\frac{I^{\text{Leader}}}{(1+r)^{T_0}} + \sum_{t=T_0+1}^{T_S} \frac{X_{\text{Mono}}}{(1+r)^t} + \sum_{t=T_S+1}^{T_F} \frac{X_{\text{SBC}}^{\text{Leader}}}{(1+r)^t} + \sum_{t=T_F+1}^{\infty} \frac{X_{\text{FBC}}^{\text{Leader}}}{(1+r)^t} \right]. \quad (17)$$

The follower's market value is given by (16).

By investing first, a firm receives monopoly profits until the follower enters into SBC. Furthermore, during the period of SBC it receives access revenue from its rival, as well as any other benefits of incumbency. It may also receive some benefit from being the incumbent during FBC, such as a stronger competitive position due to customer lock-in, brand loyalty, and so on. In contrast, by investing second a firm is able to exploit any technological advances that may lower the cost of investment or allow the firm to offer customers higher quality goods and services. Moreover, the firm gains more information than its rival before it invests, thus allowing the firm to avoid investing in unsuccessful innovations.

Two different types of equilibria typically arise, which are distinguished by whether the follower strictly prefers to invest second or is indifferent between investing first or second. In the first case, investment is a 'waiting game', since one firm is content to wait and allow its rival to invest first. As a consequence, the presence of a competitor does not affect the leader's investment timing, except through its impact on $X_{\text{SBC}}^{\text{Leader}}$ and $X_{\text{FBC}}^{\text{Leader}}$. In the second case, investment is a 'preemption game', since the leader cannot delay investment without the rival firm investing first. Here the presence of a competitor accelerates the leader's investment.

The discussion of access regulation is separated into two parts, corresponding to the two different types of equilibrium.

7.2.1   Access regulation when investment is a waiting game

Investment competition will become a waiting game when the firms involved have markedly different investment costs. It is often the case that one firm has an intrinsic cost advantage that ensures it will be the first to invest in a facility.[82] For example, a firm's cost advantage may arise from easements, which give it the right to use others' property for specified purposes,

---

and Pindyck (1994, Chapter 9), to the access pricing problem (Hori and Mizuno, 2005).

[82]In some situations there will be only one firm that is ever able to invest in the facility, and SBC is the only entry option available to potential competitors. This case is particularly relevant given that, at this relatively

not available to potential competitors; an incumbent telecommunications network will find it cheaper than an entrant to upgrade to provide broadband access.

I consider the leader's investment behavior in three situations: (i) when entry by the follower is prohibited; (ii) when only facilities-based entry is allowed; (iii) when service-based entry is also allowed, with the terms of access regulated. In the first case, the firm bears all the cost of its investment, but shares the flow of surplus with its customers, so that the firm will invest too late from the point of view of overall welfare if its output prices are not regulated. If they are regulated, then the results of Sections 4–6 apply. In the second case, if price regulation is removed once entrants appear, and $X_{\text{Mono}} > X_{\text{FBC}}^{\text{Leader}}$, then the threat of future FBC reduces the leader's investment payoff and will generally induce it to delay investment longer than in the first case. In the third case, the leader's investment timing and the effect on overall welfare depend on the specifics of access regulation.

When investment is a waiting game, the natural investment leader invests at a date T_0 that maximizes $F^{\text{Leader}}(T_0)$, recognizing that the follower's entry dates T_S and T_F will depend on T_0 via the follower's value-maximization problem in (16). Because the preemption constraint is not binding, when analyzing the impact of access regulation on investment timing only the impact of regulatory parameter changes on the leader's value-maximizing investment date needs to be considered. Lowering the access price impacts on the leader's payoff function through two channels: (i) the profit flow during the period of SBC, $X_{\text{SBC}}^{\text{Leader}}$, is likely to fall (but the precise impact depends on the form of downstream competition as discussed in Section 7.1); (ii) the follower may enter SBC sooner (T_S falls) and FBC later (T_F rises). The overall impact on the leader's payoff, and on its choice of T_0, depends on which effect dominates.

In some cases the prospect of future entry by competitors unambiguously delays investment by the access provider. One such situation occurs when SBC is the only form of competition possible, and the access seeker does not need to make any irreversible investments to enter. Then SBC begins as soon as the leader invests and FBC never eventuates, so that the second effect above is irrelevant. If the access terms are such that the leader's profit falls during SBC, then the incumbent delays investment (Woroch, 2004). Raising the access price leads to earlier investment (Gans, 2001).

However, the level of access prices is just one regulatory parameter relevant to access regulation. Another is how access prices vary over time. If the cost of constructing an asset is expected to fall if investment is delayed, then setting access prices in line with the historical cost of the asset, rather than adjusting them to match changes in the replacement cost of the

---

early stage of the deregulation process, access regulation has most often been applied to existing assets. In this case, the relevant investment involves future maintenance or expansion of an existing asset.

asset, can accelerate investment, as explained in Section 6.3.3. Guthrie et al. (2003) show that, for various forms of downstream competition, the greater dynamic efficiency from setting access prices according to the asset's historical cost, rather than its replacement cost, outweighs any loss of allocative efficiency. This suggests that TELRIC pricing of access is not socially optimal.

7.2.2 Access regulation when investment is a preemption game

When investment is a preemption game, the leader's investment timing is actually determined by the follower's payoffs from leading and following. Specifically, the leader will invest immediately before the first occasion that the follower would like to preempt its investment.[83] For example, if firm 1 is the leader in equilibrium, then it will invest at the earliest date T_0 that satisfies

$$F_2^{\text{Leader}}(T_0) = F_2^{\text{Follower}}(T_0), \tag{18}$$

where F_2 is the market value of firm 2. While firm 1's market value would be higher if it could invest later (while still being the first firm to invest), this is not possible as any delay will result in firm 2 investing first.

Regulation affects the leader's investment timing by altering the follower's preemption incentive. For example, capping output prices at a level that binds only when a firm is a monopoly provider reduces X_{Mono} in (17) but does not affect the other profit flows. It thus lowers the preemption incentive and leads to later investment.[84] In situations where the race to invest first otherwise leads to investment occurring too early to maximize welfare, such regulation can raise welfare.[85]

Access regulation affects investment timing by influencing the smallest solution T_0 to (18). As in Section 7.2.1, I consider the leader's investment behavior in three situations: (i) when entry is prohibited; (ii) when only facilities-based entry is allowed; (iii) when service-based entry is also allowed, with the terms of access regulated. As before, in the first case the firm will invest too late from the point of view of overall welfare if its output prices are not regulated. Now, however, when entry is allowed the threat of preemption by the follower typically induces earlier investment by the leader. Access regulation can slow investment by mitigating preemption incentives — this can be socially beneficial if unregulated competition induces the leader to

---

[83]Notice, though, that even in the preemption game the follower may delay investment quite some time after the leader invests. It is the *threat* of early investment by the follower that induces the leader to invest early; once the leader has invested, it will often be the case that the follower will optimally delay investment.

[84]Recall the discussion of Dobbs (2004) in Section 4. He finds that a price cap accelerates investment, since it reduces the high profits the firm can earn while it waits to invest. Here a price cap reduces the profits the firm can earn after it invests, while it waits for the rival to follow.

[85]This is modelled formally by Riordan (1992), who uses a deterministic model in which FBC is the only possible entry strategy (that is, $T_S = T_F$).

invest before the socially-optimally date.

Higher regulated access charges raise the leader's profitability, and lower the follower's profitability, during the period of SBC, with the precise impact depending on the form of downstream competition. However (from Section 7.1), higher access charges may also delay the start and accelerate the end of SBC. If the first effect dominates, then raising access charges makes investing first relatively more attractive than investing second, thereby strengthening the preemption incentive of both firms, and leading to earlier investment by the leader; SBC will occur later, while FBC will begin earlier.[86] Thus, the level of access charges can be used by a regulator to influence investment timing. For example, if unregulated competition would lead to investment occurring sooner than is socially optimal, then regulating lower access charges can raise welfare.[87]

As is clear from Sections 7.2.1 and 7.2.2, much work remains to be done in this important area. First, the impact of access price levels on investment is not yet fully understood, even in the relatively simple situations described here. Second, even less is known about the overall impact on welfare. For example, even if higher access prices would accelerate investment, is this necessarily good for welfare? Third, even those models that incorporate uncertainty in investment payoffs do not adequately capture the learning option that firms acquire by investing second. The approach followed by Hori and Mizuno (2005) illustrates the possibilities offered by real option modeling. However, while these models can easily incorporate the dynamics, investment irreversibility, and uncertainty that are all essential features of regulated competition, they typically assume that the follower 'loses' the race to invest — the leader earns monopoly profits for a period after it invests and shares the market with the follower after the latter invests. In contrast, for the sorts of investment opportunities that are relevant in discussions of regulated competition (for example, investments that offer new products or create existing products with new technology) the follower learns valuable information about the success of the innovation just from observing the outcome of the leader's investment. The possibility to learn valuable information by investing second raises $F^{\text{Follower}}(T_0)$ — the follower will only enter into SBC if the expression inside the square brackets in (16) is large, and the avoidance of negative values raises the overall expected value. Likewise, because the follower will only enter if the

---

[86]Hori and Mizuno (2005) model competition between two firms that make irreversible investments in the face of uncertainty. They show that the effect of higher access charges on the profit flow during SBC dominates their effect on the duration of SBC; that is, a high access charge accelerates the leader's investment.

[87]Using a deterministic model in which SBC begins immediately after the leader invests and FBC never occurs, Gans (2001) derives access charges that induce a preemption game that results in socially-optimal investment timing. However, Hori and Mizuno (2005) show that social optimality cannot be achieved by usage charges alone when entrants can also engage in FBC.

innovation is successful, the leader's payoff in (16) will fall. The two effects combine to lower the preemption incentive, and thereby relax the pressure on the leader to invest early. This 'learning by building' feature of investment timing is not yet well treated in the real option literature.

8 Conclusion

The origins of the relationship between regulation and investment can be found in the investment flexibility that firms, even regulated ones, enjoy. They use this flexibility to maximize their market value and, because regulation affects the impact of investment on firms' market values, regulation alters the investment choices that firms make. In the long run, this investment influences the quantity and quality of the goods that firms produce.

A common theme of Sections 4–6 is that regulation affects firms' investment behavior by altering the allocation of risk between their shareholders and customers, and that it does this in part by specifying the costs that firms are allowed to recover when prices are adjusted at future hearings. Under traditional rate of return regulation, the firm is guaranteed to recover all of its costs, provided they have been prudently incurred. This shifts much of the risk of doing business onto a firm's customers, encouraging investment. In contrast, shareholders bear much more of this risk under modern incentive regulation, discouraging investment.[88] Furthermore, the precise impact of regulation on investment varies with the industry involved. For example, shifting from rate of return regulation to a system of incentive regulation will have a dramatic impact in an industry where the risks associated with investment are great (due to, for example, large technology shocks, longer project-construction times, and more volatile demand) but a smaller impact in other industries.

Another key issue is the credibility of the regulatory regime, although this is less problematic in the US than in many other countries. If regulatory opportunism is possible, then firms will invest either in anticipation of opportunistic behavior or in an attempt to deter such behavior. In either case, firms will try to delay as much investment as possible until after the next regulatory hearing, and overall welfare will be relatively low. In the first case, this reduces the losses that the firm will incur if a regulator does not allow it to recover its sunk costs. In the second case, this allows the firm to threaten a cessation of necessary future investment unless past cost-recovery commitments are met. Once more, a firm's precise investment response depends on the nature of the industry. For example, the need for ongoing investment in a rapidly growing industry

---

[88]However, it is not clear exactly what cost measures regulators use when they *reset* price caps and other forms of incentive regulation. Because these forms of regulation are relatively new and the associated regulatory hearings are held relatively infrequently, there have been relatively few price cap reviews to analyze. To complicate matters, many incentive schemes were initially regarded as spanning a transition to deregulated markets, so that past practices might not be representative of what firms could expect from ongoing incentive regulation.

makes it easier for a firm to deter regulatory opportunism without excessive investment delays.

As discussed in Section 7, several extra dimensions are added to the problem when vertically integrated firms are exposed to competition in some segments: the incumbent faces new risks, such as the loss of customers to competitors, and industry investment is now determined by multiple firms, so investment by competitors must also be considered. Opening up some segments to competition can have a dramatic impact on the incumbent's investment behavior, typically accelerating investment if rivals can preempt the incumbent's investment and slowing it down if they cannot. Regulating the terms of access to the incumbent's bottleneck facilities can alter this behavior, but more research is needed before the possibilities are fully understood. Although there is an emerging literature on investment in this new regulatory environment, economists are struggling to keep up with the changes. Almost ten years have passed since the Telecommunications Act transformed telecommunications regulation in the US and economists still do not have a thorough understanding (theoretically or empirically) of how local loop unbundling affects investment. Understanding of the investment response to electricity transmission pricing is even less developed. More study of access regulation and its impact on investment behavior, especially investment timing, is needed.

References

Abel, Jaison R. 2002. "Entry into Regulated Monopoly Markets: The Development of a Competitive Fringe in the Local Telephone Industry," *J. Law Econ.*, 45, pp. 289–316.

Ai, Chunrong; and David E. Sappington. 2002. "The Impact of State Incentive Regulation on the U.S. Telecommunications Industry," *J. Regul. Econ.*, 22:2, pp. 133–160.

Alesina, Alberto; Silvia Ardagna; Guiseppe Nicoletti; and Fabio Schiantarelli. 2005. "Regulation and Investment," *J. Europ. Econ. Assoc.*, forthcoming.

Alexander, Ian; Colin Mayer; and Helen Weeds. 1996. "Regulatory Structure and Risk and Infrastructure Firms," World Bank Policy Research Working Paper 1698.

Armstrong, Mark. 2002. "The Theory of Access Pricing and Interconnection," Chapter 8 in Martin E. Cave; Sumit K. Majumdar; and Ingo Vogelsang (eds.), *Handbook of Telecommunications Economics (Vol. 1: Structure, Regulation and Competition)*, North-Holland.

Armstrong, Mark; and David Sappington. 2003. "Recent Developments in the Theory of Regulation," in *Handbook of Industrial Organization (Vol. 3)*, Elsevier Science.

Atkinson, Scott E.; and Clifford Nowell. 1994. "Explaining Regulatory Commission Behavior in the Electric Utility Industry," *Southern Econ. J.*, 60:3, pp. 634–643.

Averch, Harvey; and Leland L. Johnson. 1962. "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, 52:5, pp. 1052–1069.

Bailey, Elizabeth E. 1974. "Innovation and Regulation," *J. Public Econ.*, 3, pp. 285–295.

Bailey, Elizabeth E., and Roger D. Coleman. 1971. "The Effect of Lagged Regulation in an Averch-Johnson Model," *Bell J. Econ. Manage. Sci.*, 2:1, pp. 278–292.

Baron, David P. 1988. "Regulation and Legislative Choice," *RAND J. Econ.*, 19:3, pp. 467–477.

Baumol, William J. 1967. "Reasonable Rules for Rate Regulation: Plausible Policies for an Imperfect World," in A. Phillips and O. E. Williamson (eds.), *Prices: Issues in Theory, Practice, and Public Policy*, Philadelphia: University of Pennsylvania Press, pp. 108–123.

Baumol, William J. 1971. "Optimal Depreciation Policy: Pricing the Products of Durable Assets," *Bell J. Econ. Manage. Sci.*, 2:2, pp. 638–656.

Baumol, William J.; and Alvin K. Klevorick. 1970. "Input Choices and Rate-of-Return Regulation: An Overview of the Discussion," *Bell J. Econ.*, 1:2, pp. 162–190.

Baumol, William J.; and J. Gregory Sidak. 2002. "The Pig in the Python: Is Lumpy Investment Used and Useful?" *Energy Law J.*, 23:2, pp. 383–399.

Bawa, Vijay S., and David S. Sibley. 1980. "Dynamic behavior of a Firm Subject to Stochastic Regulatory Review," *Int. Econ. Rev.*, 21:3, pp. 627–642.

Bernanke, Ben S. 1983. "Irreversibility, Uncertainty, and Cyclical Investment," *Quart. J. Econ.*, 98, pp. 85–106.

Biais, Bruno, and Enrico Perotti. 2002. "Machiavellian Privatization," *Amer. Econ. Rev.*, 92:1, pp. 240–258.

Biglaiser, Gary; and Michael Riordan. 2000. "Dynamics of Price Regulation," *RAND J. Econ.*, 31:4, pp. 744–767.

Binder, John J. 1985. "Measuring the Effects of Regulation with Stock Price Data," *RAND J. Econ.*, 16:2, pp. 167–183.

Blackmon, Glenn; and Richard Zeckhauser. 1992. "Fragile Commitments and the Regulatory Process," *Yale J. Regul.*, 9, pp. 73–105.

Bourreau, Marc; and Pinar Doğan. 2005. "Unbundling the Local Loop," *Europ. Econ. Rev.*, 49, pp. 173–199.

Bower, R. S. 1981. "Impact of Regulation on Economic Behavior: Discussion," *J. Finance*, 36:2, pp. 397–399.

Braeutigam, Ronald R.; and John C. Panzar. 1989. "Diversification Incentives Under "Price-Based" and "Cost-Based" Regulation," *RAND J. Econ.*, 20:2, pp. 373–391.

Braeutigam, Ronald R.; and John C. Panzar. 1993. "Effects of the Change from Rate-of-Return to Price-Cap Regulation," *Amer. Econ. Rev., Papers and Proceedings*, 83:2, pp. 191–198.

Braeutigam, Ronald R.; and James P. Quirk. 1984. "Demand Uncertainty and the Regulated Firm," *Int. Econ. Rev.*, 25, pp. 45–60.

Brennan, Michael J.; and Eduardo S. Schwartz. 1982a. "Consistent Regulatory Policy Under Uncertainty," *Bell J. Econ.*, 13:2, pp. 506–521.

Brennan, Michael J.; and Eduardo S. Schwartz. 1982b. "Regulation and Corporate Investment Policy," *J. Finance*, 37:2, pp. 289–300.

Burness, H. Stuart; W. David Montgomery; and James P. Quirk. 1980. "Capital Contracting and the Regulated Firm," *Amer. Econ. Rev.*, 70:3, pp. 342–354.

Che, Yeon-Koo. 1995. "Revolving Doors and the Optimal Tolerance for Agency Collusion," *RAND J. Econ.*, 26:3, 378–397.

Clarke, Roger G. 1978. "The Impact of a Fuel Adjustment Clause on the Regulated Firm's Value and Cost of Capital," *J. Finan. Quant. Anal.*, 13:4, pp. 745–757.

Clarke, Roger G. 1980. "The Effect of Fuel Adjustment Clauses on the Systematic Risk and Market Values of Electric Utilities," *J. Finance*, 35:2, pp. 347–358.

Cowan, Simon. 2002. "Price-Cap Regulation," *Swedish Econ. Policy Rev.*, 9, pp. 167–188.

Crew, Michael A.; and Paul R. Kleindorfer. 2002. "Regulatory Economics: Twenty Years of Progress?" *J. Regul. Econ.*, 21:1, pp. 5–22.

Davis, E. G. 1973. "A Dynamic Model of the Regulated Firm with a Price Adjustment Mechanism," *Bell J. Econ. Manage. Sci.*, 4:1, pp. 270–282.

Dixit, Avinash K.; and Robert S. Pindyck. 1994. *Investment Under Uncertainty*. Princeton University Press, Princeton, NJ.

Dobbs, Ian M. 2004. "Intertemporal Price Cap Regulation Under Uncertainty," *Econ. J.*, 114, pp. 421–440.

Economides, Nicholas. 2005. "Telecommunications Regulation: An Introduction," in Richard R. Nelson (ed.), *The Limits and Complexity of Organizations*, Russell Sage Foundation Press, New York, 2005

Elton, Edwin J.; and Martin J. Gruber. 1977. "Optimal Investment and Financing Patterns for a Firm Subject to Regulation with a Lag," *J. Finance*, 32:5, pp. 1485–1500.

Evans, Lewis; and Steven Garber. 1988. "Public Utility Regulators Are Only Human: A Positive Theory of Rational Constraints," *Amer. Econ. Rev.*, 78:3, pp. 444–462.

Evans, Lewis T., and Graeme A. Guthrie. 2005a. "Risk, Price Regulation, and Irreversible Investment," *Int. J. Ind. Organ.*, 23, pp. 109–128.

Evans, Lewis T.; and Graeme A. Guthrie. 2005b. "Backward- versus Forward-Looking Regulation: Consequences for Investment and Welfare," Working paper, Victoria University of Wellington.

Evans, Lewis; and Graeme Guthrie. 2005c. "Incentive Regulation of Prices when Costs are Sunk," *J. Regul. Econ.*, forthcoming.

Faure-Grimaud, Antoine. 2002. "Using Stock Price Information to Regulate Firms," *Rev. Econ. Stud.*, 69, pp. 169–190.

Fay, Marianne; and Tito Yepes. 2003. "Investing in Infrastructure: What is Needed from 2000 to 2010?" World Bank Policy Research Working Paper 3102.

Federal Communications Commission. 1998. "Long Distance Market Shares: Second Quarter 1998," Industry Analysis Division, Common Carrier Bureau, Federal Communications Commission.

Foreman, R. Dean. 1995. "Pricing Incentives Under Price-Cap Regulation," *Info. Econ. Pol.*, 7, pp. 331–351.

Fudenberg, Drew; and Jean Tirole. 1985. "Preemption and Rent Equalization in the Adoption of New Technology," *Rev. Econ. Stud.*, 52:3, pp. 383–401.

Galai, Dan; and Ronald Masulis. 1976. "The Option Pricing Model and the Risk Factor of Stock," *J. Finan. Econ.*, 3:1/2, pp. 53–81.

Gans, Joshua S. 2001. "Regulating Private Infrastructure Investment: Optimal Pricing for Access to Essential Facilities," *J. Regul. Econ.*, 20:2, pp. 167–189.

Gans, Joshua S.; and Stephen P. King. 2004. "Access Holidays and the Timing of Infrastructure Investment," *Econ. Rec.*, 80:248, pp. 89–100.

Gilbert, Richard J.; and David M. Newbery. 1994. "The Dynamic Efficiency of Regulatory Constitutions," *RAND J. Econ.*, 25:4, pp. 538–554.

Goldberg, Victor P. 1976. "Regulation and Administered Contracts," *Bell J. Econ.*, 7:2, pp. 426–448.

Golec, Joseph. 1990. "The Financial Effects of Fuel Adjustment Clauses on Electric Utilities," *J. Bus.*, 63, pp. 165–186.

Greenstein, Shane; Susan McMaster; and Pablo T. Spiller. 1995. "The Effect of Incentive Regulation on Infrastructure Modernization: Local Exchange Companies' Deployment of Digital Technology," *J. Econ. Manage. Strategy*, 4:2, pp. 187–236.

Greenwald, Bruce C. 1980. "Admissible Rate Bases, Fair Rates of Return and the Structure of Regulation," *J. Finance*, 35:2, pp. 359–368.

Greenwald, Bruce C. 1984. "Rate Base Selection and the Structure of Regulation," *RAND J. Econ.*, 15:1, pp. 85–95.

Grout, Paul A.; and Anna Zalewska. 2005. "The Impact of Regulation on Market Risk," *J. Finan. Econ.*, forthcoming.

Guthrie, Graeme A.; John P. Small; and Julian Wright. 2003. "Pricing Access: Forward-Looking Versus Backward-Looking Cost Rules," Working paper, Victoria University of Wellington.

Hagerman, James. 1990. "Regulation by Price Adjustment," *RAND J. Econ.*, 21:1, pp. 72–82.

Hausman, Jerry A. 1997. "Valuing the Effect of Regulation on New Services in Telecommunications," *Brookings Pap. Econ. Act.*, Micro. 1997, pp. 1–38.

Hausman, Jerry A. 1999. "The Effect of Sunk Costs in Telecommunications Regulation," In: Alleman, James, and Eli Noam (eds). *The New Investment Theory of Real Options and its Implications for Telecommunications Economics*, Kluwer Academic Publishers, Boston, MA.

Hausman, Jerry; and Stewart Myers. 2002. "Regulating the United States Railroads: The Effects of Sunk Costs and Asymmetric Risk," *J. Regul. Econ.*, 22:3, pp. 287–310.

Hoppe, Heidrun C.; and Ulrich Lehmann-Grube. 2005. "Innovation Timing Games: A General Framework with Applications," *J. Econ. Theory*, 121, pp. 30–50.

Hori, Keiichi; and Keizo Mizuno. 2005. "Access Pricing and Investment with Stochastically Growing Demand," Working paper, Ritsumeikan University.

Iozzi, Alberto; and Marco Fioramenti. 2004. "Strategic Choice of the Price Structure and Entry Deterrence under Price Cap Regulation," *Bull. Econ. Res.*, 56:4, pp. 333–352.

Jensen, Michael C.; and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *J. Finan. Econ.*, 3:4, pp. 305–360.

Johnstone, D. J. 2003. "Replacement Cost Asset Valuation and Regulation of Energy Infrastructure Tariffs," *Abacus*, 39:1, pp. 1–41.

Joskow, Paul L. 1973. "Pricing Decisions of Regulated Firms: A Behavioral Approach," *Bell. J. Econ. Manag. Sci.*, 4:1, pp. 118–140.

Joskow, Paul L. 1974. "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Regulation," *J. Law Econ.*, 17, pp. 291–327.

Joskow, Paul L. 2000. "Deregulation and Regulatory Reform in the US Electric Power Sector," Chapter 4 in Sam Peltzman and Clifford Winston (eds.), *Deregulation of Network Industries: What's Next?* AEI-Brookings Joint Center for Regulatory Studies.

Kandel, Eugene; and Neil D. Pearson. 2002. "Option Value, Uncertainty, and the Investment Decision," *J. Finan. Quant. Anal.*, 37:3, pp. 341–374.

Katz, Michael L.; and Carl Shapiro. 1987. "R&D Rivalry with Licensing or Imitation," *Amer. Econ. Rev.*, 77:3, pp. 402–420.

Laffont, Jean-Jacques and Jean Tirole. 1993. *A Theory of Procurement and Regulation.* Cambridge, Massachusetts: MIT Press.

Lehman, Dale E.; and Dennis L. Weisman. 2000. "The Political Economy of Price Cap Regulation," *Rev. Ind. Organ.*, 16:4, pp. 343–356.

Levy, Brian; and Pablo T. Spiller. 1994. "The Institutional Foundations of Regulatory Commitment: A Comparative Analysis of Telecommunications Regulation," *J. Law Econ. Organ.*, 10:1, pp. 201–246.

Lewellen, Wilbur G.; and David C. Mauer. 1993. "Public Utility Valuation and Risk Under Incentive Regulation," *J. Regul. Econ.*, 5, pp. 263–287.

Lyon, Thomas P. 1991. "Regulation with 20–20 Hindsight: 'Heads I Win, Tails You Lose'?" *RAND J. Econ.*, 22:4, pp. 581–595.

Lyon, Thomas P. 1995. "Regulatory Hindsight Review and Innovation by Electric Utilities," *J. Regul. Econ.*, 7, pp. 233–254.

Lyon, Thomas P.; and John W. Mayo. 2004. "Regulatory Opportunism and Investment Behavior: Evidence from the U.S. Electric Utility Industry," *RAND J. Econ.*, forthcoming.

Mandy, David M. 2002. "TELRIC Pricing with Vintage Capital," *J. Regul. Econ.*, 22:3, pp. 215–249.

Mandy, David M.; and William W. Sharkey. 2003. "Dynamic Pricing and Investment From Static Proxy Models," *Rev. Network Econ.*, 2:4, pp. 403–439.

Marshall, William J.; Jess B. Yawitz; and Edward Greenberg. 1981. "Optimal Regulation Under Uncertainty," *J. Finance*, 36:4, pp. 909–921.

Martimort, David. 1999. "Renegotiation Design with Multiple Regulators," *J. Econ. Theory*, 88:2, pp. 261–293.

Mayo, John W.; and Joseph E. Flynn. 1988. "The Effects of Regulation on Research and Development: Theory and Evidence," *J. Bus.*, 61:3, pp. 321–336.

Myers, Stewart C. 1972. "The Application of Finance Theory to Public Utility Rate Cases," *Bell J. Econ. Manage. Sci.*, 3:1, pp. 58–97.

Newbery, David M. 1999. *Privatization, Restructuring, and Regulation of Network Industries.* Cambridge, Massachusetts: MIT Press.

Nowell, Clifford, and Jason F. Shogren. 1991. "The Timing of a Rate Request by a Regulated Firm," *Southern Econ. J.*, 57:4, pp. 1054–1060.

Pakes, Ariel. 1997. "Valuing the Effect of Regulation on New Services in Telecommunications: Comment," *Brookings Pap. Econ. Act.*, Micro. 1997, pp. 39–48.

Pindyck, Robert S. 1993. "Investments of Uncertain Cost," *J. Finan. Econ.*, 34:1, pp. 53–76.

Pindyck, Robert S. 2003. "Mandatory Unbundling and Irreversible Investment in Telecom Networks," NBER Working Paper No. W10287.

Pint, Ellen M. 1992. "Price-Cap versus Rate-of-Return Regulation in a Stochastic-Cost Model," *RAND J. Econ.*, 23:4, pp. 564–578.

Prager, Robin A. 1989. "The Effects of Regulatory Policies on the Cost of Debt for Electric Utilities: An Empirical Investigation," *J. Bus.*, 62:1, pp. 33–53.

Riordan, Michael H. 1992. "Regulation and Preemptive Technology Adoption," *RAND J. Econ.*, 23:3, pp. 334–349.

Roberts, R. Blaine; G. S. Maddala; and Gregory Enholm. 1978. "Determinants of the Requested Rate of Return and the Rate of Return Granted in a Formal Regulatory Process," *Bell J. Econ.*, 9:2, pp. 611–621.

Rogerson, William P. 1992. "Optimal Depreciation Schedules for Regulated Utilities," *J. Regul. Econ.*, 4, pp. 5–33.

Salant, David J. 1995. "Behind the Revolving Door: A New View of Public Utility Regulation," *RAND J. Econ.*, 26:3, pp. 362–377.

Salant, David J.; and Glenn A. Woroch. 1992. "Trigger Price Regulation," *RAND J. Econ.*, 23:1, pp. 29–51.

Salinger, Michael A. 1998. "Regulating Prices to Equal Forward-Looking Costs: Cost-Based Prices or Price-Based Costs?", *J. Regul. Econ.*, 14, pp. 149–163.

Salop, Steven C.; and David T. Scheffman. 1983. "Raising Rivals' Costs," *Amer. Econ. Rev.*, 73:2, pp. 267–271.

Saphores, Jean-Daniel; Eric Gravel, and Jean-Thomas Bernard. 2004. "Regulation and Investment Under Uncertainty: An Application to Power Grid Interconnection," *J. Regul. Econ.*, 25:2, pp. 169–186.

Sappington, David E. M. 1980. "Strategic Firm Behavior Under a Dynamic Regulatory Adjustment Process," *Bell J. Econ.*, 11:1, pp. 360–372.

Sappington, David E. M. 1986. "Commitment to Regulatory Bureaucracy," *Info. Econ. Pol.*, 2, pp. 243–258.

Sappington, David E. M. 2002. "Price Regulation," Chapter 7 in Martin E. Cave; Sumit K. Majumdar; and Ingo Vogelsang (eds.), *Handbook of Telecommunications Economics (Vol. 1: Structure, Regulation and Competition)*, North-Holland.

Sappington, David E. M. 2005. "On the Irrelevance of Input Prices for Make or Buy Decisions," *Amer. Econ. Rev.*, forthcoming.

Sappington, David E. M.; and David S. Sibley. 1992. "Strategic Nonlinear Pricing Under Price-Cap Regulation," *RAND J. Econ.*, 23:1, pp. 1–19.

Sappington, David E. M. and Dennis L. Weisman. 1996. *Designing Incentive Regulation for the Telecommunications Industry.* Cambridge, Massachusetts: MIT Press.

Schmalensee, Richard. 1989. "An Expository Note on Depreciation and Profitability Under Rate-of-Return Regulation," *J. Regul. Econ.*, 1, pp. 293–298.

Schmidt, Klaus M. 2000. "The Political Economy of Mass Privatization and the Risk of Expropriation," *Eur. Econ. Rev.*, 44, pp. 393–421.

Sweeney, George. 1981. "Adoption of Cost-Saving Innovations by a Regulated Firm," *Amer. Econ. Rev.*, 71:3, pp. 437–447.

Teisberg, Elizabeth Olmstead. 1993. "Capital Investment Strategies Under Uncertain Regulation," *RAND J. Econ.*, 24:4, pp. 591–604.

Teisberg, Elizabeth Olmstead. 1994. "An Option Valuation Analysis of Investment Choices by a Regulated Firm," *Management Sci.*, 40:4, pp. 535–548.

Vogelsang, Ingo. 2003. "Price Regulation of Access to Telecommunications Networks," *J. Econ. Lit.*, 41:3, pp. 830–862.

Vogelsang, Ingo; and Jorg Finsinger. 1979. "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms," *Bell J. Econ.*, 10:1, pp. 157–171.

Weisman, Dennis L. 2002. "Is There 'Hope' for Price Cap Regulation?" *Info. Econ. Pol.*, 14:3, pp. 349–370.

Weitzman, Martin L. 1980. "The 'Ratchet Principle' and Performance Incentives," *Bell J. Econ.*, 11:1, pp. 302–308.

Woroch, Glenn A. 1988. "Regulatory Risk, Investment, and Welfare," *Rev. Ind. Organ.*, 3:3, pp. 75–99.

Woroch, Glenn A. 2004. "Open Access Rules and Equilibrium Broadband Deployment," Chapter 13 in *Frontiers of Broadband, Electronic and Mobile Commerce*, Gary Madden and Russel Cooper (editors), Physica-Verlag.