



NEW ZEALAND INSTITUTE FOR THE STUDY
OF COMPETITION AND REGULATION INC.

Asset Stranding is Inevitable: Implications for Optimal Regulatory Design

Lewis T. Evans*

*New Zealand Institute for the Study of Competition and Regulation
and Victoria University of Wellington, Wellington, New Zealand*

Graeme A. Guthrie

Victoria University of Wellington, Wellington, New Zealand

November 12, 2003

* Corresponding author. Address: ISCR, PO Box 600, Victoria University of Wellington, Wellington, New Zealand. Ph: 64-4-4635562. Fax: 64-4-4635566. Email: lew.evans@vuw.ac.nz

Asset Stranding is Inevitable: Implications for Optimal Regulatory Design*

Lewis T. Evans[†]

New Zealand Institute for the Study of Competition and Regulation,
and Victoria University of Wellington, Wellington, New Zealand

Graeme A. Guthrie

Victoria University of Wellington, Wellington, New Zealand

*We thank participants at an NZISCR seminar and the Advanced Workshop in Regulation and Competition's 16th Annual Western Conference for helpful comments, and Telecom New Zealand Ltd for providing data.

[†]Corresponding author. Address: ISCR, PO Box 600, Victoria University of Wellington, Wellington, New Zealand. Ph: 64-4-4635562. Fax: 64-4-4635566. Email: lew.evans@vuw.ac.nz

Asset Stranding is Inevitable: Implications for Optimal Regulatory Design

Abstract

The irreversibility of much infrastructure investment means that some assets will stop earning revenue before the end of their physical lives; they will be stranded. Under traditional rate of return regulation, firms are guaranteed the ability to recover the costs of investment, insulating them from the consequences of asset stranding. Under modern incentive regulation, firms are allowed to earn revenue just sufficient to cover the costs of a hypothetical efficient firm which provides services at minimum cost, exposing them to the risk of asset stranding. By actively encouraging competition, regulators increase this risk. We suggest two conditions, applicable to both regimes, which must be met if regulation is to be “reasonable”: the regulated firm must not lose value from investment, and it cannot collect more revenue than would the lowest cost alternative provider. This implies that regulated firms should be allowed to earn the riskless rate of return on the historical cost of their assets under rate of return regulation, and a different (generally higher) rate of return on the replacement cost of their assets under incentive regulation. The risk premium depends on both the systematic and unsystematic risk of demand shocks. Since customers bear the risk of asset stranding under rate of return regulation, and shareholders bear this risk under incentive regulation, welfare is higher under incentive regulation as long as customers are more risk-averse than shareholders. We show that when there is a choice between reversible and irreversible technology, there is no price specification under rate of return regulation that will induce the firm to choose the efficient bundle of technology, while under incentive regulation the firm will choose the efficient mix of technologies. That is, incentive regulation allocates the risk of asset stranding efficiently, and also gives firms the incentive to reduce this risk to efficient levels. Finally, incentive regulation has less demanding information requirements than traditional rate of return regulation.

JEL Classification code: G31, L5

Keywords: Regulation, Rate Base, Uncertainty, Sunk Costs, Dynamic Efficiency

Asset Stranding is Inevitable: Implications for Optimal Regulatory Design

1 Introduction

A revolution has occurred in the nature of regulation during the last thirty years, which has seen regimes in which regulated firms were protected from competition but subjected to rate of return regulation largely replaced by ones in which competition is actively encouraged and regulated firms are subjected to price caps or similar forms of incentive regulation. The key change in regulation can be highlighted by considering the treatment of assets which are no longer productive. Under traditional regulation, such assets remain in the rate base provided they are the results of ‘prudent’ investment, and prices are adjusted to maintain sufficient revenue flow for the firm to recover the cost of its assets. In contrast, under incentive regulation unproductive assets are usually removed from the rate base.¹ When investment is irreversible, the firm cannot redeploy the capital elsewhere, and the nature of incentive regulation offers the firm little prospect of recovering its outlay from consumers.² In short, the assets are stranded and the regulated firm must bear the cost. Demand shocks (due to such factors as population movements, consumers switching to new technology, and competition) are likely to grow in importance in the future, meaning that infrastructure firms subject to incentive regulation are likely to face even greater stranding risk in the future.³

Unfortunately, finance theory as applied to regulation has not kept pace with these developments. The traditional approach to determining reasonable rates of return is to set allowed revenues such that the market value of the regulated firm equals the cost of its assets (Marshall,

¹In fact, the rate base is calculated as the cost of the assets required to meet the demand for the regulated firm’s product, where the latest technology and the least-cost asset configuration can be chosen. Thus, even productive assets can be removed from the rate base, to be replaced by lower cost alternatives.

²Irreversibility is a widespread phenomenon, even in industries where physical capital is not especially industry-specific. For example, between 50 and 80 percent of the cost of machine tools in Sweden is sunk (Asplund, 2000), and the market value of physical capital in the U.S. aerospace industry is just 28 percent of its replacement cost on average (Ramey and Shapiro, 2001). Irreversibility is likely to be even greater in most infrastructure networks. Hausman (1999) and Economides (1999) debate the extent of irreversibility in the context of telecommunications.

³For example, Rodini et al. (2002) estimate the degree of substitutability between secondary fixed-line and mobile telephone access using U.S. household data, concluding that they are currently “moderate” substitutes, and that the degree of substitutability will increase in the future. This will result in more volatile demand for fixed-line telephone access, and a greater risk of asset stranding in fixed-line telecommunications networks. Mayer and Vickers (1996) point out that changes in relative fuel costs will affect the optimal configuration of electricity generating companies, imparting volatility into the replacement cost of electricity generation assets.

et al., 1981). Under rate of return regulation, the latter is taken to be the historical cost of the firm's assets. Since this guarantees the firm riskless profit flows, the reasonable rate of return is the riskless interest rate. However, the approach to incentive regulation is different: the regulated firm is allowed just enough revenue to cover the costs faced by a hypothetical "efficient" firm if it was to replace the regulated firm, thereby exposing the regulated firm to the risk that its hypothetical rival's cost structure, and therefore its own allowed revenue, will fall in the future.⁴ Since a large proportion of cost is the cost of capital, incentive regulation exposes the regulated firm to the risk of fluctuations in the replacement cost of its assets. The question of what level of revenue, and, in particular, what rate of return on the asset base, is reasonable in this situation has gone largely unaddressed in the academic literature. In this paper we present a model designed especially for the modern regulatory environment, featuring irreversible investment and uncertainty about future demand, and use it to determine rates of return which are reasonable for infrastructure firms subject to incentive regulation. In particular, we examine the implications of stranding for the calculation of a regulated firm's reasonable rate of return.

We formulate, and apply, two conditions which ensure returns are "reasonable". In keeping with the literature on rate of return regulation, both are expressed in terms of the market value of the firm, which we take to be the present value of the flow of future net revenues, less the present value of future investment outlays. Firstly, we require that the market value of the regulated firm is never negative. A negative market value would mean that the present value of future investment outlays exceeds the market value of the stream of net revenue generated by the firm's (existing and future) assets, giving the firm no incentive to continue in business. This nonnegativity assumption is essential if the regulatory regime is to be credible.⁵ Secondly, we require that the market value of a hypothetical replacement firm is never positive. If this value was positive, such a firm could build a new network to meet the existing demand, allocate funds to finance future investment, and still have some wealth left over from the expected flow of future net revenues; in short, the hypothetical replacement firm could profitably replace the regulated firm. The second condition thus ensures that the firm's revenues are determined by the cost structure of the lowest cost provider.

In our model, the regulated firm has a universal service obligation which forces it to invest whenever demand exceeds the capacity of its network. We start by considering the situation in

⁴For example, the FCC's starting point in its TELRIC calculation is the cost structure of an efficient cost-minimizing firm with an optimally-configured network built with the current technology (Weisman, 2002).

⁵Gunn (2003) calculates prices for a regulated monopolist without imposing this nonnegativity constraint. In his model, the regulated firm will only continue to operate if it is willing to invest retained earnings in projects which will lower the market value of the firm (p. 156). This is inconsistent with value-maximizing behavior on the part of the firm's manager.

which investment is irreversible, so that transitory increases in demand can leave the firm with excess capacity. We consider two forms of regulation. In the first, which corresponds to rate of return regulation, the firm's allowed revenue depends on the actual capacity of its network. We show that the reasonable level of revenue results if and only if the firm is allowed to earn the riskfree rate of return on the historical cost of its network. In the second form of regulation, which corresponds to incentive regulation, the firm's allowed revenue depends on the demand for network connections. We show that for the revenue flow to be reasonable, the firm must be allowed to earn a rate of return which differs from the riskless interest rate, reflecting the fact that investors bear the risk of asset stranding. The systematic risk of demand shocks is one input into the calculation of the appropriate rate of return, but other factors, including the extent of unsystematic risk, are also important.

The universal service obligation imposes an asymmetry on the relationship between the regulated firm and its (potential) customers — customers can abandon the firm's network, but the firm cannot abandon the customers.⁶ Ultimately, the firm must be compensated for granting customers the option to abandon its network. Under rate of return regulation, prices rise when customers leave, so that effectively it is the customers who remain who pay for the abandonment options exercised by their departed counterparts; the firm is compensated, *ex post*. Under incentive regulation, prices do not change when customers leave. Instead, they are permanently higher for all connected customers, meaning that customers pay for their own abandonment options before they are exercised.⁷

Traditional regulation and incentive regulation thus differ in how they allocate the risk of asset stranding. We show that if customers are risk-averse with respect to the amount they pay for network access, then incentive regulation leads to higher welfare. This result arises because under traditional regulation the price of network access fluctuates, with prices low when demand is high, and prices high when demand is low. In contrast, incentive regulation offers less volatile prices. Incentive regulation allocates the risk of asset stranding to the firm's shareholders, who are better able to diversify their risk than the firm's customers.⁸

⁶Hausman (1999) was one of the first authors to recognize that regulated firms give valuable options to competitors and customers.

⁷In our model, customers can join and leave the regulated firm at will. In practice, their contracts with the firm may preclude early abandonment. Hausman and Myers (2002, p. 295) have argued that customers should pay for their abandonment options in the form of higher prices, with the premium added to prices depending on the length of the customers' contracts with the firm. The premium will be highest for short-term contracts.

⁸Beard et al. (2003) use a simple two-period model in which a regulator sets (i) the profit which a firm is allowed to earn in the event that its assets are not stranded, and (ii) the compensation which the firm will receive if its assets are stranded. The less compensation offered, the more profit must be allowed if the firm is to willingly invest in the project. The authors find that full compensation would not be offered by a welfare-maximizing

In practice, although firms subject to universal service obligations do not have the option to choose *whether* to supply, they can often choose *how* to supply. In the second part of the paper we extend our model to incorporate this flexibility by giving the firm access to two technologies when building its network.⁹ Investment using the more expensive technology can be costlessly reversed, while the cheaper technology involves irreversible investment.¹⁰ The cost-minimizing network configuration uses a combination of the two technologies — using too much of the cheap technology raises the risk of asset stranding, but using too little incurs unnecessary costs. We find that traditional regulation induces the firm to build exclusively with one or the other technologies. In contrast, under incentive regulation the firm chooses the efficient network configuration. Thus, not only does incentive regulation allocate the risk of asset stranding efficiently, but it also reduces this risk.

There is a long history of papers applying finance theory to the calculation of reasonable rates of return for regulated firms. Early work is surveyed by Davis and Sparrow (1972). Myers (1972) argues that the expected rate of return on new investment should equal the regulated firm’s cost of capital, measured using the CAPM. Leland (1974) describes a simple static model with expected utility-maximizing investors and defines a “fair return” as a pattern of profits such that the market value of the firm equals the cost of its assets. Marshall, et al. (1981) use the same criterion for determining a fair rate of return but, in contrast to Leland, they use the CAPM to calculate the market value of the firm, and point out that the price-setting decision cannot be separated from the rate-setting decision. Brennan and Schwartz (1982a, 1982b) present a multi-period model of traditional rate of return regulation in which, like Marshall, et al., the risk of the firm is endogenous. They define a reasonable rate of return as one which ensures the market value of the firm equals the value of the rate base, which they take to equal the historical cost of the firm’s assets, at the date of each regulatory hearing.

However, this literature tells us little about modern incentive regulation. The early static models simply cannot incorporate stranding, as this is a dynamic phenomenon. In fact, in these papers the regulated firm is effectively allowed to build a “greenfields” operation, leaving no possibility of stranding. A dynamic model is required to analyze the impact of asset stranding, but even the later models of Brennan and Schwartz, for example, focus on traditional rate of return regulation — they assume that the firm’s net revenue flow equals the product of the regulator. However, they do not look at the risk allocation issues considered in the current paper.

⁹Kandel and Pearson (2002) have recently used a similar model to analyze the influence of irreversibility on the value of investment timing flexibility.

¹⁰Examples of reversible investment might be the use of mobile generators to relieve congestion at locations on electricity networks and the installation of lower-cost lower-quality network elements that require higher annual maintenance.

actual capacity of its assets and a (regulated) stochastic rate of return, so that they only consider what we interpret as traditional rate of return regulation; incentive regulation and the issue of asset stranding are not considered. Recently Hausman and Myers (2002) have analyzed price-setting under incentive regulation, in the context of United States railroads, and argue that the regulator has not considered the impact of investment irreversibility when calculating allowed rates of return.¹¹

One common feature of the existing literature is the argument that regulators should set prices such that the market value of a regulated firm equals the cost of its assets. In line with this earlier work, we find that the market value of the firm equals the historical cost of its assets under reasonable rate of return regulation.¹² However, under reasonable incentive regulation the regulated firm's market value exceeds the replacement cost of its assets whenever the firm has stranded assets. The premium equals the amount by which the value of the regulated firm exceeds the value of its hypothetical replacement, and we interpret it as the value of the incumbent's excess capacity. This excess capacity has value because, while it does not generate any profits in the short-term, ultimately it will reduce the firm's required future investment.¹³ The conclusion is that if incentive regulation is to be "reasonable", then the regulated firm must be allowed to derive some value from its stranded assets.¹⁴

In the next section we formally set up the regulatory framework and derive the implications of irreversible investment for the performance of these schemes and the welfare and prices they imply. In Section 3 we extend the model to the case of alternative reversible and irreversible

¹¹Evans and Guthrie (2003) and Panteghini and Scarpa (2003) analyze the investment timing decisions of firms subject to various forms of regulation, including modern incentive regulation, but do not focus on the calculation of the firms' allowed rates of return.

¹²Note that we assume continuous regulation. If, like some authors, we assumed that prices were reset at discrete intervals, the market value of the firm would deviate from the historical cost of its assets between regulatory hearings. See, for example, Brennan and Schwartz (1982a, 1982b).

¹³Note that this is not planned excess capacity, for which it would be easier to justify rewarding the firm. Rather, it is excess capacity which the firm has acquired almost "by accident", capacity which was once in use but has now been stranded by changes in demand. In particular, the replacement firm would not build this excess capacity into its network. Furthermore, there will often be a positive probability that these so-called stranded assets will become economically viable in the future. Examples include power generators kept in mothballs for use during energy crises, and once stranded networks rendered economically viable by new technology (e.g. ADSL in telecommunications networks).

¹⁴There are several different ways to interpret this result. In this paper, we keep stranded assets out of the rate base, but increase the allowed rate of return to compensate for the risk of asset stranding. An alternative interpretation is to allow assets which are not in use to remain in the rate base (reflecting the fact that they are, indeed, valuable assets), but to apply a lower allowed rate of return. Provided the rates of return are chosen appropriately, the two approaches are equivalent.

technologies and consider the performance of rate of return and incentive regulation for prices and the efficiency of the cost structure of the regulated firm. The final section reviews the findings and comments on potentially useful future investigations.

2 A model of asset stranding

2.1 Setting up the model

We consider a firm which operates an infrastructure asset (referred to as ‘the network’ below). Let x_t equal the number of potential customers at time t , and let s_t equal the maximum number of connections (the capacity of the network), so that the number of customers actually connected to the network at time t is $\min\{x_t, s_t\}$. Shocks to x_t could arise for many reasons: population shifts can cause the number of customers wishing to connect to certain parts of the network to change, even when aggregate customer numbers do not change¹⁵; the arrival of new technology may result in customers abandoning the network for others offering better services, or in new customers being attracted by new services offered by the incumbent; more generally, competition from rival providers may lead to substantial fluctuations in customer numbers. We suppose that the number of potential customers evolves according to the geometric Brownian motion

$$dx_t = \mu x_t dt + \sigma x_t d\xi_t,$$

where μ and σ are constants and ξ_t is a Wiener process. We price contingent claims as though x_t evolves according to the “risk-neutral” process

$$dx_t = (\mu - \lambda)x_t dt + \sigma x_t d\xi_t,$$

for some constant λ satisfying $r + \lambda > \mu$ where r is the riskless interest rate.¹⁶

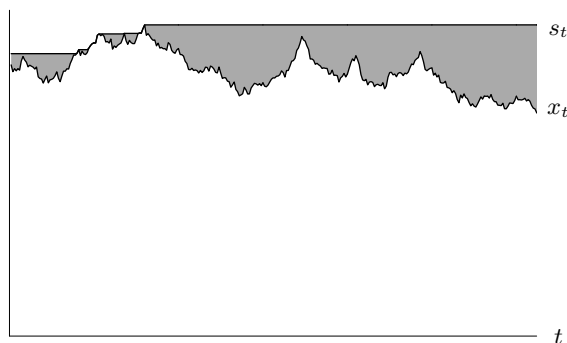
We assume that investment in new connections is irreversible, with each new connection incurring a sunk cost of c which is constant over time. For simplicity, we assume that demand is price insensitive, that connections have an infinite life, and that the network’s operating costs are zero. We suppose that the regulator forces the firm to connect any new customer who wishes to join the network, so the network’s capacity increases over time according to

$$ds_t = \begin{cases} dx_t & \text{if } s_t = x_t \text{ and } dx_t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

¹⁵If customers move from one region to another, but remain connected to the same network, the firm may well have to expand capacity in one region, even when the total number of customers remains constant.

¹⁶We can interpret λ as the risk-premium of an asset with returns which are perfectly positively correlated with changes in x_t . Thus it captures the systematic risk of shocks to customer numbers. For further discussion of risk-neutral pricing see Dixit and Pindyck (1994, Chapter 4).

Figure 1: Adjusting the capacity of the regulated network



Notes. The curve labelled x_t plots the number of potential customers as a function of time, while the curve labelled s_t plots the number of connections. As long as x_t is less than s_t/ψ the firm holds the level of s_t constant. Whenever x_t is greater than s_t the firm builds just enough new connections to ensure all customers can connect to the network. The height of the lightly-shaded region shows the number of connections which are stranded when customer numbers fall below the network's capacity.

The situation is shown in Figure 1. The capacity of the network is constant as long as there is excess capacity. If the network is operating at full capacity, then investment in new connections must occur as soon as a new customer appears. The figure shows what can go wrong — the firm is forced to expand capacity as soon as there are more customers than connections; if customer numbers subsequently fall, the network is left with excess capacity.

2.2 Measuring the network's cost

The regulated firm has lumpy cash outflows, incurred whenever expansion of the network is required, and a continuous flow of revenue collected from customers. The regulator's task is to determine exactly how much revenue the firm can collect, a problem complicated by the fact that the firm's expansion costs are sunk. The level of revenue which is 'reasonable' depends on the cost of the network, but there are many ways in which this cost can be measured. Since all investments in the network are irreversible, the cost measure of most relevance to the incumbent is the present value of all future investment outlays that are required to meet the needs of the network's customers. The precise value will depend on the number of customers currently connected to the network, as well as its current capacity. We denote it by $C(x, s)$. Of more interest to a potential entrant is the cost of replacing the network. Replicating the network at time t requires an initial outlay of cs_t . Since this results in a network with capacity of s_t , the present value of all future outlays, measured immediately after the initial investment is

completed, is $C(x_t, s_t)$. Thus the present value of all costs required to replicate the network is $cs_t + C(x_t, s_t)$. On the other hand, replacing the network at time t with one which is optimally-configured requires an initial outlay of cx_t , since it is optimal to replace the network with one having no excess capacity.¹⁷ As the resulting network has capacity of x_t , the present value of all future outlays equals $C(x_t, x_t)$. The present value of all costs required to optimally replace the network is thus $cx_t + C(x_t, x_t)$.

These three measures of cost all depend on the function $C(x, s)$, which is described in the following proposition.¹⁸

Proposition 1 (Cost function) *If the network currently has x connected customers and capacity of s , then the present value of all future expenditures needed to ensure any customer who wishes to connect to the network can do so is*

$$C(x, s) = \begin{cases} \frac{cx}{\beta-1} \left(\frac{x}{s}\right)^{\beta-1}, & 0 \leq x \leq s, \\ c(x-s) + \frac{cx}{\beta-1}, & s < x, \end{cases}$$

where

$$\beta = \frac{1}{2} + \frac{\lambda - \mu}{\sigma^2} + \sqrt{\frac{2r}{\sigma^2} + \left(\frac{1}{2} + \frac{\lambda - \mu}{\sigma^2}\right)^2}. \quad (1)$$

An immediate consequence of this result is that

$$C(x, s) \leq cx + C(x, x) \leq cs + C(x, s).$$

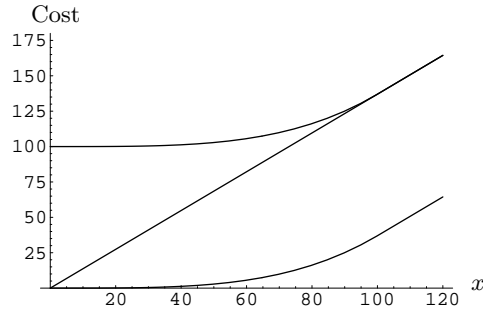
That is, when all current and future costs are considered, it is cheaper to continue to operate the existing network than to replace it with one which is optimally-configured, which is cheaper than simply replicating the network. Figure 2 plots the costs of the three options as a function of x . The bottom curve plots $C(x, s)$ as a function of x , assuming that $s = 100$ connections are already in place. The straight line plots the cost of replacing the network with one which is efficiently-configured, and the top curve plots the cost of replicating the existing network.

Our measures of cost are forward-looking: they reflect both the initial outlay required to build the network as well as the stream of outflows needed to expand the network to meet future demand. Given the irreversible nature of investment in network assets, these cost measures can be significantly larger than the initial cash outlay. Proposition 1 shows that although it costs cx to build a network capable of connecting x customers, the present value of all costs equals $C(x, 0) = \beta cx / (\beta - 1) > cx$. Some typical values of the cost multiplier $\beta / (\beta - 1)$ are given in Table 1. Except when customer numbers are deterministic, the multiplier exceeds unity. The

¹⁷This would not necessarily be the case if we included fixed costs of investment in our model. Then it would be optimal to build excess capacity into the replacement network, although simply replicating the existing network would not generally be optimal.

¹⁸Proofs for all propositions can be found in Appendix A.

Figure 2: Present value of all future expansion costs



Notes. The bottom curve plots $C(x, s)$, the present value of all future investment outlays incurred by the incumbent, when $s = 100$. The middle curve plots $cx + C(x, x)$, the present value of all costs incurred by an entrant which replaces the network with one which is efficiently-configured. The top curve plots $cs + C(x, s)$, the present value of all costs incurred by an entrant which replicates the network. In all cases, $c = 1$, $\mu = 0$, $\sigma = 0.1$, $r = 0.05$ and $\lambda = 0$.

Table 1: Cost multipliers

	$\lambda = 0.00$				$\lambda = 0.04$		
$\mu \backslash \sigma$	0.000	0.050	0.100	0.150	0.050	0.100	0.150
-0.02	1.000	1.055	1.181	1.343	1.020	1.079	1.167
0.00	1.000	1.171	1.370	1.600	1.030	1.111	1.228
0.02	1.667	1.763	2.000	2.325	1.055	1.181	1.343

Notes. If customer numbers follow a geometric Brownian motion with drift μ and volatility σ , and attract a (systematic) risk premium of λ , then the present value of building a network for x customers equals $\beta cx / (\beta - 1)$. The entries in the tables report the value of $\beta / (\beta - 1)$ corresponding to the indicated values of μ , σ and λ . In all cases $r = 0.05$.

extra costs do not arise just from expected growth in customer numbers, as $C(x, 0) > cx$ even when $\mu = 0$. They also arise due to volatility in customer numbers. To see why, note that any increases in customer numbers beyond x , even short-term increases, trigger investment in new connections. Because of the irreversible nature of this investment, reductions in customer numbers below x do not result in off-setting cash inflows. The asymmetry means that the present value of investment outlays exceeds the initial outlay cx . For reasonable levels of volatility, this can increase the apparent cost of building the network by one third.¹⁹ This means that the firm

¹⁹A positive drift in customer numbers raises the multiplier further, due to the greater sequence of future investment outlays. When customer numbers have positive systematic risk, future expansion costs are discounted more, so that the present value of future costs falls, as is evident from comparing the two panels of the table. The key point is that it is not sufficient to look at expected growth in, and systematic risk of, customer numbers;

must be allowed to collect revenue with present value greater than cx if it is to be willing to build the network in the first place.

There is thus a potentially significant cost differential between the post-investment cost $C(x_t, s_t)$ that regards investment to time t as sunk, and the ex ante costs $cx_t + C(x_t, x_t)$ or $cs_t + C(x_t, s_t)$. Because $C(x, s)$ is declining in s , a larger network requires less revenue to carry on in business than a smaller network. A firm with an existing network with capacity s , and only $x < s$ customers, will choose to stay in business if the regulator offers a revenue stream with a present value exceeding $C(x_t, s_t)$. However, the firm would not construct the network at the outset if it had good reason to anticipate that, once built, the network revenues would simply be sufficient to cover $C(x, s)$. The regulator's rules must solve this potential time-inconsistency issue and thereby induce the firm's participation under regulation.²⁰

2.3 What level of revenue is 'reasonable'?

We suppose that the regulator forces the firm to provide services to all those customers who want them, and that it allows the firm to earn a revenue flow of $\Pi_t dt$ at date t . This revenue can depend on the number of customers connected to the network, as well as the network's capacity; that is, $\Pi_t = \Pi(x_t, s_t)$ for some function Π . We let $F(x_t, s_t)$ denote the value of the regulated firm at time t .

The revenue function is chosen such that the regulated firm earns a 'reasonable' return. We impose two conditions which the revenue function must satisfy. Firstly, the regulated firm must be financially viable. Secondly, it should not be possible for another firm to profitably build a network from scratch and replace the incumbent. The first condition imposes a lower bound on revenue, while the second condition imposes an upper bound. We discuss these two conditions in turn.

Assuming the regulated firm remains in the industry indefinitely, its value at time t is $F(x_t, s_t)$. Since the firm's investments are irreversible, its value will be zero if it exits the industry. Therefore, remaining in the industry is value-maximizing if and only if $F(x_t, s_t) \geq 0$. Therefore, we impose the following firm participation condition on the revenue function:

Condition C1 (Financial viability) *The revenue function must lead to a firm value function which satisfies $F(x, s) \geq 0$ for all $x \geq 0$ and $s \geq 0$.*

This condition says that the regulated firm can expect to recover the costs of any future investment in additional capacity, measured at the time the investment takes place. It does not

volatility is crucial.

²⁰See Fudenberg and Tirole (1991, pp. 74–76) for an elaboration of time inconsistency.

guarantee that the firm will recover the cost of its past investment. Neither does it guarantee that the (stochastic) future revenues will actually cover the cost of future investment.

Any potential replacement for the regulated firm is required to build a network capable of servicing the regulated firm's customers. If the incumbent's network has excess capacity, the replacement firm would build a smaller network: it replaces, rather than replicates, the existing network. Thus, if the existing network has x customers and capacity of s , the replacement firm only needs to build a network of size x , which it could do profitably if $F(x, 0) > 0$. Thus we impose the following additional condition on the revenue function:

Condition C2 (Lowest cost provider) *The revenue function must lead to a firm value function which satisfies $F(x, 0) \leq 0$ whenever $x \geq 0$.*

This condition exposes the regulated firm to the risk of asset stranding. It may be stuck with the cost of capital invested in a network which is too big (that is, $x_t < s_t$), while the hypothetical replacement firm, which effectively caps the incumbent's allowed revenue, only faces the cost of building a smaller network.

The present value of the firm's future investment outlays is a decreasing function of s . Therefore, provided the allowed revenue is increasing in s , the value of the firm must be increasing in s .²¹ In this case, the firm is financially viable if and only if $F(x, 0) \geq 0$, proving the following useful result:

Lemma 1 *If the allowed revenue function $\Pi(x, s)$ is increasing in s , then conditions C1 and C2 are equivalent to $F(x, 0) = 0$ for all $x \geq 0$.*

We consider two forms of revenue function. In the first case, which corresponds to traditional rate of return regulation, the regulator allows the firm a revenue flow $\Pi_t = \Pi(s_t)$ which is a function of the size of the network, and not the number of customers. The following proposition shows that there is a unique reasonable revenue function of this form.

Proposition 2 (Traditional regulation) *Suppose that the allowed revenue flow is a function only of the network's capacity: $\Pi_t = \Pi(s_t)$ for some function Π . The only such function satisfying conditions C1 and C2 is $\Pi_{TR}(s) = rc s$, implying a value of the regulated firm of $F(x, s) = cs$.*

Once the firm with this revenue function invests in a new connection, costing c , it is guaranteed an incremental revenue flow of rc in perpetuity. In particular, the firm is certain to recover the

²¹By 'increasing', we mean that the relevant partial derivative is greater than or equal to zero. Thus, this result also holds if allowed revenue is independent of the network's existing capacity.

cost of all of its investment. Since its value equals cs_t , the firm is not exposed to the risk of asset stranding under traditional regulation. Instead, this risk is borne by the customers who remain connected to the network, since each customer connected at time t must contribute $rcs_t dt/x_t$. This price is bounded below by $rc dt$ (when $x_t = s_t$), but is unbounded above — as customer numbers fall, the price for connecting to the network rises without bound.

The second form of regulation which we consider, in which the regulator allows the firm a total revenue flow that depends only on the number of connected customers, corresponds to incentive regulation. The following proposition shows that there is a unique reasonable revenue function from this family.

Proposition 3 (Incentive regulation) *Suppose that the allowed revenue flow is a function only of the number of customers connected to the network: $\Pi_t = \Pi(x_t)$ for some function Π . The only such function satisfying conditions C1 and C2 is*

$$\Pi_{IR}(x) = \left(\frac{\beta c}{\beta - 1} \right) (r + \lambda - \mu)x,$$

where β is given by (1). The value of the regulated firm is

$$F(x, s) = \begin{cases} \frac{cx}{\beta-1} \left(\beta - \left(\frac{x}{s}\right)^{\beta-1} \right), & 0 \leq x \leq s, \\ cs, & s < x. \end{cases}$$

The firm now bears the risk of asset stranding. This is clear from Figure 3, where the solid curves plot the firm's value when capacity is $s = 100$, and the dashed curves plot the corresponding values for the higher capacity of $s = 130$. The horizontal curves plot the firm's value under traditional regulation, and the upward-sloping curves its value under incentive regulation. When operating at full capacity, the value of the firm equals $F(s_t, s_t) = cs_t$, but it falls below this level when customer numbers drop below capacity. The firm receives compensation for bearing stranding risk in the form of greater allowed revenue, at least as long as the network operates at close to full capacity. In fact, as long as customer numbers satisfy

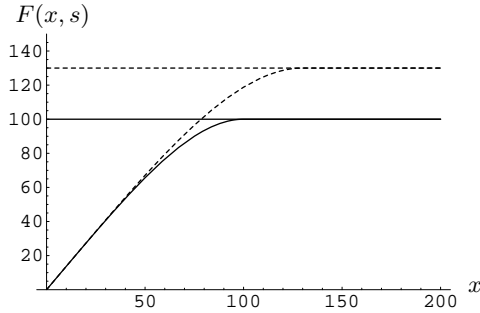
$$\frac{x_t}{s_t} > \frac{\beta - 1}{\beta} \cdot \frac{r}{r + \lambda - \mu}$$

allowed revenue is higher under incentive regulation.²²

The market value of the regulated firm equals $F(x, s)$, but if the firm was replaced by a hypothetical firm which built its network in order to minimize cost, the replacement firm would

²²Evidence presented in Appendix B suggests that volatility of $\sigma = 0.1$ is realistic. For this level of volatility, when customer numbers have zero systematic risk and the riskfree rate is $r = 0.05$, allowed revenue is higher under incentive regulation whenever $x_t/s_t > 0.724$ when $\mu = 0$. If customer numbers have drift $\mu = 0.02$, the threshold is $x_t/s_t > 0.833$, and if drift is $\mu = -0.02$, the threshold is $x_t/s_t > 0.602$.

Figure 3: Value of the incumbent network

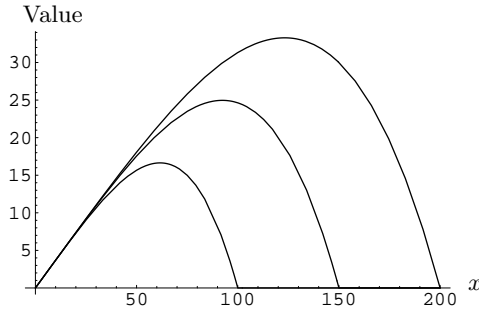


Notes. The solid horizontal line plots the value of the firm (as a function of the number of customers) under traditional regulation, for the revenue function in Proposition 2 and capacity of $s = 100$. The solid curve plots the firm's value under incentive regulation, for the revenue function in Proposition 3 and capacity of $s = 100$. The dotted curves plot the corresponding functions when capacity is $s = 130$. In all cases, $c = 1$, $\mu = 0$, $\sigma = 0.1$, $r = 0.05$ and $\lambda = 0$.

spend just cx on its network. The difference, $F(x, s) - cx$, is the amount by which the value of the regulated firm is allowed to exceed the replacement cost of its assets. As long as the regulated firm has excess capacity (that is, $s > x$), then this difference is positive. In other words, it is simply not reasonable to drive the value of the regulated firm down to the value of a hypothetical efficient provider. Figure 4 plots the (allowed) value of the firm's excess capacity as a function of x . The top curve corresponds to existing capacity of $s = 200$, the middle curve to $s = 150$, and the bottom curve to $s = 100$. The height of each curve equals the amount by which the market value of the regulated firm exceeds the cost of the assets of a hypothetical efficient replacement firm.

In contrast to traditional regulation, under incentive regulation customers face a constant price for connecting to the network. In effect, customers pay for their option to abandon the network while they are connected to it, whereas under traditional regulation those customers who remain connected to the network pay for the abandonment options of those customers who have left. Traditional regulation and incentive regulation thus differ in how they allocate the risk of asset stranding. In order to analyze the welfare implications of this difference, we suppose that each customer derives a flow of utility $u(p_t)dt$ from connecting to the network, where p_t is the amount they pay for network access and the function u is decreasing and strictly concave; that is, customers are risk-averse with respect to fluctuations in the cost of network access.

Figure 4: Allowed value of excess capacity



Notes. The bottom solid curve plots $F(x, s) - cx$, the amount by which the market value of the regulated firm, $F(x, s)$, exceeds the cost of the assets of the hypothetical efficient replacement firm, cx . In all cases, $c = 1$, $\mu = 0$, $\sigma = 0.1$, $r = 0.05$ and $\lambda = 0$.

Aggregating over all customers, the total flow of utility is

$$x_t u \left(\frac{rcs_t}{x_t} \right) dt$$

under traditional regulation, and

$$x_t u \left(\left(\frac{\beta c}{\beta - 1} \right) (r + \lambda - \mu) \right) dt$$

under incentive regulation. We denote the present values of these flows, measured at time 0 when $s_0 = 0$, by U_{TR} and U_{IR} respectively. The following proposition shows that customers are better off under incentive regulation.

Proposition 4 (Welfare comparison) *The present value of aggregate utility satisfies*

$$U_{TR} < U_{IR}.$$

Since the present value of producer surplus is identical under the two regimes, it follows that welfare is higher under incentive regulation than under traditional regulation.

The present value of customers' aggregate contribution to the cost of the network is the same under each regime. The key difference between the two regimes is that each customer's contribution is volatile under traditional regulation (although the total contribution is deterministic) and constant under incentive regulation (although the total contribution is volatile). With traditional regulation, connected customers benefit when demand is high and the network is operating at full capacity since this is when their individual contribution is low. However, when demand for the network falls, the remaining customers must each bear a greater burden of the

network’s cost. Proposition 4 shows that the reduced utility resulting from negative demand shocks outweighs the increased utility resulting from positive demand shocks.²³

In practice, allowed revenue is often expressed in terms of a so-called ‘reasonable rate of return’ on some asset base. The choice of asset base affects the reasonable rate of return. Under traditional regulation, the future revenue flows depend on the actual amount invested in the network, so it is natural to use cs_t as the asset base. In this case the reasonable rate of return is the riskless rate: $\hat{r}_{\text{TR}} = r$. The situation is not quite so simple with incentive regulation, as an adjustment reflecting the firm’s risk of asset stranding needs to be made somewhere, either to the asset base or to the rate of return. For example, Propositions 1 and 3 imply that the growth- and risk-adjusted discount rate $r + \lambda - \mu$ is a reasonable rate of return provided it is applied to the asset base comprising the present value of all current and future costs of replacing the network; that is,

$$\Pi_{\text{IR}}(x) = (r + \lambda - \mu)C(x, 0).$$

In contrast, if only the cost of replacing the assets actually “in the ground” is included in the asset base, then the reasonable rate of return is

$$\hat{r}_{\text{IR}} = \left(\frac{\beta}{\beta - 1} \right) (r + \lambda - \mu),$$

since then $\Pi_{\text{IR}}(x) = \hat{r}_{\text{IR}}cx$. This reasonable rate of return \hat{r}_{IR} is reported in Table 2 for a range of values of the key parameters. The first column of returns shows that if there is no uncertainty in customer numbers, and positive growth, so that irreversibility is not a problem (the firm is unlikely to want to reverse its investment in new connections), then the firm will break even if it can earn the riskfree rate. However, irreversibility is a very real problem if customer numbers are falling, in which case an upward adjustment in the return is required if the firm is to break even. The table also shows that we must adjust for volatility in customer numbers. This adds an additional premium, which is an increasing function of the volatility in future customer numbers.²⁴

Finally, Figure 3 illustrates an important implication for the firm’s investment behavior.

²³It is straightforward to calculate the present value of all future revenues, which measures the cost of the network to consumers. We find that under traditional regulation the present value of all current and future revenues equals $cs + C(x, s)$; that is, the cost of replicating the network. In contrast, with incentive regulation the present value equals $cx + C(x, x)$; that is, the cost of replacing the network with one which is efficiently-configured. These are the top two curves in Figure 2.

²⁴The empirical example contained in Appendix B.1 shows that parameter values of $\mu = 0.026$ and $\sigma = 0.163$ are realistic, implying a reasonable rate of return of $r_{\text{IR}} = 0.070$ when customer numbers have zero systematic risk and the riskfree rate is $r = 0.05$. Thus, when volatility is considered, the reasonable rate of return can exceed the riskfree rate by two percentage points, even when there is a significant upward trend in customer numbers.

Table 2: Reasonable rates of return with incentive regulation and initial outlay as asset base (\hat{r}_{IR})

$\mu \backslash \sigma$	$\lambda = 0.00$				$\lambda = 0.04$		
	0.000	0.050	0.100	0.150	0.050	0.100	0.150
-0.020	0.070	0.074	0.083	0.094	0.112	0.119	0.128
0.000	0.050	0.059	0.069	0.080	0.093	0.100	0.111
0.020	0.050	0.053	0.060	0.070	0.074	0.083	0.094
0.040	0.050	0.052	0.056	0.063	0.059	0.069	0.080

Notes. If customer numbers follow a geometric Brownian motion with drift μ and volatility σ , and attract a (systematic) risk premium of λ , then the reasonable return on the initial outlay required to replace the existing network is \hat{r}_{IR} . The entries in the tables report the value of \hat{r}_{IR} corresponding to the indicated values of μ , σ and λ . In all cases $r = 0.05$.

The solid curves plot the firm's value when capacity is $s = 100$, and the dashed curves plot the corresponding values for the higher capacity of $s = 130$. Inspection of the two horizontal lines shows that the firm's value increases by the cost of the extra capacity. Thus, under traditional regulation, the firm breaks even when undertaking any investment (even unnecessary investment). In contrast, while increasing capacity raises the value of the firm under incentive regulation, the increase is less than the cost of the extra connections when the investment is unnecessary (that is, when current capacity exceeds current customer numbers). Only when investment is required to connect some customers does the firm break even. Thus, when the firm bears the risk of stranding, the regulator can delegate investment decisions to the firm. This is not the case with traditional regulation. We investigate investment incentives in more detail in the following section.

3 Combining reversible and irreversible technologies

In this section we generalize the model in Section 2 by supposing that alternative technologies are available which vary according to their cost and their flexibility. In order to keep the analysis as simple as possible we assume that the network operator has the choice of two (not mutually-exclusive) technologies, a *rigid* one and a *flexible* one. The rigid technology is cheaper, but investment is irreversible when using this technology, while it is perfectly reversible using the flexible technology (that is, all costs can be recovered). While obviously highly stylized, this set-up captures the essence of the choices facing network owners. For example, expanding telecommunications firms can choose between relatively cheap copper networks with costs which are largely sunk, or more expensive cellular networks which offer greater flexibility; electricity

distribution networks can choose between upgrading capacity or installing temporary diesel generators when relieving a transmission constraint.²⁵

The network owner is able to mix-and-match the two technologies. Thus, part of the network can be built using the cheap but inflexible technology, while the rest can use the more expensive flexible one. When investment is necessary, but the long term viability of expansion is uncertain, it may be optimal to use the expensive technology since this policy avoids irreversible investments that are likely to soon be unnecessary — many times small increases in potential customer numbers will be undone by themselves, and delaying irreversible investment in favor of more expensive (in the short-run) reversible investment preserves the valuable option to wait and see what happens. It is only sensible to take advantage of the cheap technology when customer numbers are so high that even if numbers subsequently fall, the network will be operating at full capacity for some time, allowing the firm to earn some sort of return on its (irreversible) investment.²⁶ Appropriate use of the two technologies can greatly reduce the risk of asset stranding.

3.1 Setting up the model

As in Section 2, the firm must connect all potential customers to its network. Once again x_t equals the number of customers, but now s_t equals the number of connections built using the rigid technology. Therefore if $x_t \leq s_t$ then all customers connect to the rigid part of the network (which thus has excess capacity), while if $x_t > s_t$ then $x_t - s_t$ additional connections, built using the flexible technology, are needed. The network composition is summarized in Table 3. The construction of each connection built using the rigid technology incurs an immediate sunk cost of c_1 , while each connection built using the flexible technology incurs an immediate cost of c_2 . As in Section 2, these costs are assumed constant across time and across firms. Investment using the flexible technology can be costlessly reversed at any time; that is, each flexible connection has a scrap value of c_2 . We assume that $c_1 < c_2$, so that the rigid technology is cheaper than the flexible technology.²⁷

²⁵Another, less obvious, interpretation of the flexible technology is increased maintenance. For example, instead of renewing a wire, maintenance patches can be put in place each time it breaks (that is, effectively converting capital to labor).

²⁶This behavior is typical of firms which have to make irreversible investment decisions in an uncertain environment. Waiting can have value; investment is only optimal when the value of the profit flow from investing immediately exceeds the value of the option to delay. See Dixit and Pindyck (1994) for an excellent treatment of this ‘real options’ approach to investment.

²⁷Because each flexible connection can be sold at a cost of c_2 at any time in the future, we can interpret each flexible connection as incurring only a flow cost of $rc_2 dt$. (Since c_2 is constant over time, the riskfree rate is

Table 3: Network composition

		Rigid provision	Flexible provision	Total
Connections	$x_t < s_t$	s_t	0	s_t
	$x_t \geq s_t$	s_t	$x_t - s_t$	x_t
Customers	$x_t < s_t$	x_t	0	x_t
	$x_t \geq s_t$	s_t	$x_t - s_t$	x_t
Excess capacity	$x_t < s_t$	$s_t - x_t$	0	$s_t - x_t$
	$x_t \geq s_t$	0	0	0

Notes. The table describes the composition of the network when it has excess capacity (that is, when $x_t < s_t$), and when it does not (that is, when $x_t \geq s_t$).

There are many investment rules which the firm could adopt. Three particularly important ones are described below.

Flexible technology. The firm only uses the flexible technology when building its network.

Thus $s_t = 0$ for all t , ensuring that there is never any excess capacity.

Rigid technology. The firm only uses the rigid technology. The capacity of the rigid part of the network therefore increases over time according to

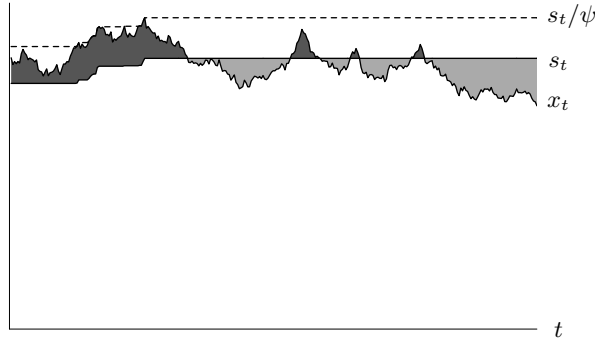
$$ds_t = \begin{cases} dx_t & \text{if } s_t = x_t \text{ and } dx_t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

This is the investment strategy analyzed in Section 2.

Instantaneous control. The firm only adds to its stock of rigid connections when this is required to keep the proportion of connections which are of the rigid type above some threshold ψ , and uses the flexible technology to connect any customers who cannot be connected using the available rigid technology. Thus, as long as s_t/x_t is greater than ψ , the firm holds the level of s_t constant; changes in demand are met by changes in the number of flexible connections. However, whenever s_t/x_t drops below ψ the firm builds just enough connections of the rigid type to restore s_t/x_t to ψ . Thus, investment in rigid connections is triggered when x_t climbs above s_t/ψ . The situation is shown in Figure 5. The curve labelled x_t plots the number of potential customers as a function of time, while the curve labelled s_t plots the number of connections built using the rigid technology. As

appropriate here.) This is the approach taken in proving the results in this section.

Figure 5: Adjusting network composition over time



Notes. The curve labelled x_t plots the number of potential customers as a function of time, while the curve labelled s_t plots the number of connections built using the rigid technology. As long as x_t is less than s_t/ψ the firm holds the level of s_t constant. Whenever x_t is greater than s_t/ψ the firm builds just enough new connections using the rigid technology to keep s_t/ψ equal to x_t . When the number of customers exceeds the size of the rigid part of the network, the firm builds the required additional connections using the flexible technology. The height of the darkly-shaded region shows the number of flexible connections needed, and the lightly-shaded region shows the number of rigid connections which are stranded when customer numbers fall below the network's rigid capacity.

long as x_t is less than s_t/ψ the firm holds the level of s_t constant. Whenever x_t climbs above the threshold s_t/ψ the firm adds just enough rigid connections to keep s_t/ψ equal to x_t . When the number of customers exceeds the size of the rigid part of the network, the firm builds the required additional connections using the flexible technology. The height of the darkly-shaded region shows the number of flexible connections needed when the rigid part of the network has insufficient capacity, and the lightly-shaded region shows the number of rigid connections which are stranded when customer numbers fall below the network's rigid capacity. The number of connections built using the rigid technology therefore increases over time according to

$$ds_t = \begin{cases} \psi dx_t & \text{if } s_t = \psi x_t \text{ and } dx_t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

In fact, the first two rules are special cases of the third. When $\psi = 0$ the firm only uses the flexible technology, and when $\psi = 1$ it only uses the rigid technology.

3.2 Measuring the network's cost

We suppose that the firm adopts an investment rule of the instantaneous control type and that it chooses the level of ψ .²⁸ In fact, because the rigid technology has no fixed costs, only variable ones, the instantaneous control policy will be optimal for some value of ψ . (See, for example, Dixit, 1993, Chapter 4.) Comparison of the lightly-shaded regions in Figures 1 and 5 shows how the firm can use the flexible technology to reduce the risk of asset stranding. The firm essentially trades off the greater costs associated with the flexible technology (represented by the darkly-shaded region in Figure 5) against the benefits of a reduced risk of stranding (represented by the lightly-shaded region).

The following proposition gives the present value of all future expansion costs, assuming that the network is configured using some form of instantaneous control policy.

Proposition 5 (Cost function) *Suppose that the network only invests using the rigid technology when $x_t > s_t/\psi$, and then it only increases capacity to ψx_t . At all other times, it holds the minimum number of flexible connections needed to ensure all customers can connect to the network. Then the present value of all future expenditures is*

$$C(x, s; \psi) = \begin{cases} f(\psi)s \left(\frac{x}{s}\right)^\beta, & 0 \leq x \leq s, \\ \frac{rc_2x}{r+\lambda-\mu} - c_2s + g(\psi)s \left(\frac{x}{s}\right)^\beta + h(\psi)s \left(\frac{x}{s}\right)^{-\gamma}, & s < x \leq s/\psi, \\ C(x, \psi x; \psi) + c_1(\psi x - s), & s/\psi < x, \end{cases}$$

where β is given by (1),

$$\gamma = -\frac{1}{2} - \frac{\lambda - \mu}{\sigma^2} + \sqrt{\frac{2r}{\sigma^2} + \left(\frac{1}{2} + \frac{\lambda - \mu}{\sigma^2}\right)^2},$$

and

$$\begin{aligned} f(\psi) &= -\frac{\psi^\beta(c_2 - c_1)}{\beta - 1} + \frac{\psi^{\beta+\gamma}c_2}{\beta + \gamma} \cdot \frac{\gamma + 1}{\beta - 1} \cdot \frac{r + \beta(\lambda - \mu)}{r + \lambda - \mu} + \frac{c_2}{\beta + \gamma} \cdot \frac{r - \gamma(\lambda - \mu)}{r + \lambda - \mu}, \\ g(\psi) &= -\frac{\psi^\beta(c_2 - c_1)}{\beta - 1} + \frac{\psi^{\beta+\gamma}c_2}{\beta + \gamma} \cdot \frac{\gamma + 1}{\beta - 1} \cdot \frac{r + \beta(\lambda - \mu)}{r + \lambda - \mu}, \\ h(\psi) &= \frac{c_2}{\beta + \gamma} \cdot \frac{r + \beta(\lambda - \mu)}{r + \lambda - \mu}. \end{aligned}$$

Consider two special cases. When $\psi = 0$ and $s_t = 0$ for all t , the present value of all future costs is $C(x, 0; 0) = rc_2x/(r + \lambda - \mu)$. In contrast, when $\psi = 1$, we get the cost function from Proposition 1, so that the present value of building the network from scratch is $C(x, 0; 1) =$

²⁸In practice, technological limitations might impose some restrictions on the firm's choice of ψ . For example, perhaps the network simply cannot function if it has too little of the rigid technology, imposing a lower bound on the firm's choice of ψ . We ignore this added complication in this paper.

Table 4: Efficient network configuration (ψ^*)

$\mu \backslash \sigma$	$\lambda = 0.00$				$\lambda = 0.04$		
	0.000	0.050	0.100	0.150	0.050	0.100	0.150
-0.020	0.758	0.718	0.636	0.543	0.422	0.386	0.337
0.000	1.000	0.888	0.774	0.660	0.553	0.500	0.432
0.020	1.000	0.961	0.871	0.761	0.718	0.636	0.543

Notes. Customer numbers follow a geometric Brownian motion with drift μ and volatility σ , and attract a (systematic) risk premium of λ . The present value of building a network is minimized by ensuring that the number of connections built using the rigid technology never drops below ψ^* times the number of customers, where ψ^* is the number given in the body of the table. The entries in the tables report the value of ψ^* corresponding to the indicated values of μ , σ and λ . In all cases $r = 0.05$, $c_1 = 1$, $c_2 = 2$.

$\beta c_1 x / (\beta - 1)$. By selecting an intermediate value of ψ , the flexible technology can be used to mitigate the stranding risk associated with using the cheap but rigid technology. The optimal threshold is given in the following corollary.

Corollary 1 (Efficient network configuration) *The cost function $C(x, s; \psi)$ is minimized by setting $\psi = \psi^*$, where*

$$\psi^* = \left(\frac{c_2 - c_1}{c_2} \right)^{1/\gamma},$$

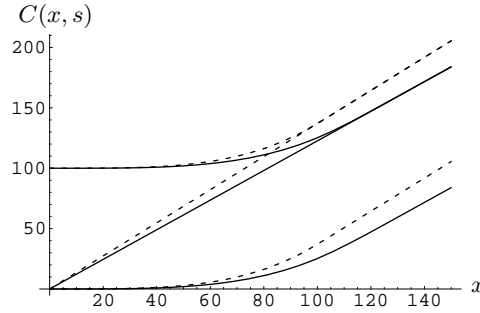
and

$$\frac{C(x, 0; \psi^*)}{x} = \frac{r(\psi^* c_1 + (1 - \psi^*) c_2)}{r + \lambda - \mu}. \quad (2)$$

Notice that the efficient network configuration depends (via γ) on the dynamics of customer numbers, as well as their systematic risk (λ), the riskfree interest rate, and the relative cost of the two technologies. Table 4 reports values of ψ^* for various combinations of the key parameters, showing that it is efficient to make greater use of the expensive flexible technology if customer numbers are more volatile or declining more rapidly. These are exactly the circumstances in which the risk of stranding is greatest.

As in the simple model presented in Section 2, there are several ways in which the cost of the network can be measured. The cost of most relevance to the incumbent is the present value of all future investment outlays, $C(x_t, s_t; \psi^*)$. In contrast, potential entrants are more interested in the cost of replicating or replacing the network. Replicating at time t requires the construction of s_t connections using the rigid technology, costing $c_1 s_t$. The present value of all subsequent outlays, measured immediately after this investment is completed, is $C(x_t, s_t; \psi^*)$. Thus the total cost of replicating the network is $c_1 s_t + C(x_t, s_t; \psi^*)$. On the other hand, replacing

Figure 6: Present value of all future expansion costs



Notes. The bottom solid curve plots $C(x, s; \psi^*)$, the present value of all future investment outlays incurred by the incumbent, when $s = 100$. The middle solid curve plots $c_1\psi^*x + C(x, \psi^*x; \psi^*)$, the present value of all costs incurred by an entrant which replaces the network with one which is efficiently-configured. The top solid curve plots $cs + C(x, s; \psi^*)$, the present value of all costs incurred by an entrant which replicates the network. The broken curves plot the corresponding cost measures, assuming that only the rigid technology is used. In all cases, $c_1 = 1$, $c_2 = 2$, $\mu = 0$, $\sigma = 0.1$, $r = 0.05$ and $\lambda = 0$.

the network at time t with one which is efficiently-configured requires the construction of ψ^*x_t connections using the rigid technology, costing $c_1\psi^*x_t$. The present value of all future outlays equals $C(x_t, \psi^*x_t; \psi^*)$. The total cost of replacing the network is thus

$$c_1\psi^*x_t + C(x_t, \psi^*x_t; \psi^*) = C(x_t, 0; \psi^*).$$

These three measures of cost are plotted in Figure 6. The solid curves show the three cost measures when the firm chooses the efficient policy (that is, with $\psi = \psi^*$), while the broken curves show the corresponding cost measures when the firm follows the policy described in Section 2 (that is, with $\psi = 1$). The most obvious feature is that the introduction of the more expensive (flexible) technology has lowered all three cost measures. At first glance, Figure 6 suggests that the revenue flow required for the firm to be financially viable must drop compared to the case in Section 2 — since the curve representing $C(x, s)$ is now lower, the present value of future revenues required for the firm to cover its future investment costs must also be lower. However, this comparison is made for common values of s . In practice, the existence of the flexible technology allows the firm to choose lower values of s , thereby raising this curve to $C(x, s')$ for some $s' < s$.

3.3 What level of revenue is ‘reasonable’?

Now consider what happens when the network is operated by a regulated firm. The sequence of events is as follows. Firstly, the regulator specifies the revenue which the firm is allowed to collect. Secondly, the firm chooses its investment policy (represented by the threshold ψ). Finally, nature chooses the evolution of customer numbers. We consider two forms of revenue function, corresponding to traditional and incentive regulation.

Under traditional regulation the firm is allowed revenue which depends on the size of its network, with a revenue flow of $\pi_1 dt$ per rigid connection, and of $\pi_2 dt$ per flexible connection. We analyze the situation by working backwards through the sequence above. Firstly, we calculate the value of the firm for arbitrary revenue flows and an arbitrary investment policy. Secondly, we calculate the firm’s optimal investment policy as a function of these revenue flows. Finally, we calculate the revenue flows which are reasonable. The results are reported in the following proposition.

Proposition 6 (Traditional regulation) *Suppose that the network is allowed a revenue flow of $\pi_1 dt$ per rigid connection, and of $\pi_2 dt$ per flexible connection, and that the firm chooses the investment threshold ψ which maximizes its value. The only revenue functions which are reasonable (that is, which satisfy conditions C1 and C2) are*

1. $\pi_1 = rc_1$ and $\pi_2 = rc_2$, in which case the firm is indifferent between all values of ψ and the value of the firm is $F(x, s) = c_1 s$;
2. $\pi_1 < rc_1$ and $\pi_2 = rc_2$, in which case the firm only uses the flexible technology (that is, $\psi = 0$) and the value of the firm is zero;
3. $\pi_1 = rc_1$ and $\pi_2 < rc_2$, in which case the firm only uses the rigid technology (that is, $\psi = 1$) and the value of the firm is $F(x, s) = c_1 s$.

Since returns on the the firm’s investment are guaranteed by the regulator, if returns are to be reasonable the regulator cannot allow the firm to earn a rate of return greater than the risk free rate: $\pi_i \leq rc_i$ for $i = 1, 2$. Moreover, at least one of these restrictions must hold with equality. However, because returns are guaranteed, the firm has no incentive to choose the efficient investment policy. At best, the firm will be indifferent between all investment policies; at worst, it will favor one or the other of the two technologies. This induced inefficiency means that consumers not only bear the risk of asset stranding, they actually bear an increased risk of asset stranding.

If the regulator abandons the objective of achieving a ‘reasonable’ revenue function, it can

give the firm the appropriate incentives to follow the efficient investment policy. However, as the following corollary shows, this comes at the cost of not allowing the firm to break even.

Corollary 2 *Suppose that the network is allowed a revenue flow of $\pi_1 dt$ per rigid connection, and of $\pi_2 dt$ per flexible connection, and that the firm chooses the investment threshold ψ which maximizes its value. The only revenue functions which induce the firm to adopt the efficient investment policy have $\pi_1 = \rho c_1$ and $\pi_2 = \rho c_2$ for any constant $\rho < r$. The value of the firm satisfies*

$$\frac{F(x, 0; \psi^*)}{x} = \frac{r(\rho - r)(\psi^* c_1 + (1 - \psi^*) c_2)}{r + \lambda - \mu} < 0.$$

To conclude our discussion about traditional regulation, we note that the information requirements are great if reasonable returns are the goal — from Proposition 6, the regulator needs to know the composition of the network, rather than just its total capacity, if traditional regulation is to allow the firm to earn reasonable revenue. The following corollary shows that if the allowed revenue function depends only on the network's total capacity, and not its composition, then the firm will choose $\hat{\psi} > \psi^*$. That is, the firm will invest in too much of the rigid technology.

Corollary 3 *Suppose that the network is allowed a revenue flow of πdt per connection, regardless of whether it uses the rigid or the flexible technology, and that the firm chooses the investment threshold ψ which maximizes its value. If $rc_1 < \pi < rc_2$, the firm will choose $\hat{\psi} > \psi^*$.*

The second regime we consider, which corresponds to incentive regulation, has relatively few information requirements, since the firm is allowed to collect a constant revenue of πdt from each connected customer, implying a total revenue flow at date t of $\pi x_t dt$. Since the firm cannot influence its total revenue flow, the only way for it to maximize its value is to minimize its costs. Thus, incentive regulation will lead to an efficiently-configured network. The following proposition states this result formally.

Proposition 7 (Incentive regulation) *Suppose that the network's allowed revenue function is $\Pi(x_t) = \pi x_t$ for some constant π , and that the firm chooses the investment threshold ψ which maximizes its value. The only revenue function which is reasonable (that is, which satisfies conditions C1 and C2) is*

$$\Pi_{IR}(x_t) = (r + \lambda - \mu)C(x_t, 0; \psi^*). \quad (3)$$

The firm chooses the efficient investment threshold ψ^ , and the value of the firm is*

$$F(x, s; \psi) = C(x, 0; \psi^*) - C(x, s; \psi^*).$$

Consider two special cases, beginning with the limiting case in which $c_1 \rightarrow c_2$. Since the flexible technology is no more expensive than the rigid one, the firm will only use the flexible technology to build its network. Thus $\psi^* \rightarrow 0$ and $\Pi_{\text{IR}} \rightarrow r c_2 x_t$; the reasonable rate of return on the firm's investment is the riskfree rate because investment using the flexible technology is costlessly reversible, and there is no risk associated with the resale value of the network's assets. The second special case, when $c_2 \rightarrow \infty$, is slightly more complicated. In this case, the flexible technology is so expensive that the firm only uses the rigid technology to build its network: $\psi^* = 1$. Application of l'Hôpital's rule shows that

$$\Pi_{\text{IR}} = \left(\frac{\beta c_1}{\beta - 1} \right) (r + \lambda - \mu) x_t,$$

which is the revenue function derived in Proposition 3.

As in Proposition 3, the reasonable rate of return depends on the choice of asset base. From equation (3), when the asset base is the present value of all current and future costs required to replace the network with one which is efficiently-configured, then the reasonable rate of return is just $r + \lambda - \mu$.²⁹ However, the revenue function in (3) can also be written as

$$\Pi_{\text{IR}} = r(c_1(\psi^* x) + c_2((1 - \psi^*)x)),$$

implying a reasonable rate of return equal to the risk free interest rate when the asset base is the initial outlay required to replace the existing network with one which is efficiently configured.³⁰ However, while this allowed rate of return might seem low, it is applied to an 'inflated' asset base. To understand the nature of this 'inflation', we must reexamine the notion of an efficient network configuration. It is a dynamic concept — an efficient configuration minimizes the present value of the cost of building a network capable of maintaining all current and future customers. That is, it minimizes the present value of all current and future outlays required to build the network. It does not minimize the immediate investment outlays; in fact, it will often be the case that the initial outlay required to replicate the network is lower than the initial outlay needed to replace it with one which is efficiently-configured. In other words, it is efficient to build flexibility into the network, and this is expensive.

²⁹It might appear that current customers are required to contribute towards the cost of expanding the network to meet the needs of future customers. However, a better interpretation of this result is that current customers are paying for the cost of their option to abandon the network at a time of their choosing. Recall that under traditional regulation, this cost is borne by the (other) customers who remain connected to the network.

³⁰Recall that if a new (efficiently-configured) network was built at time t to replace the existing one, it would comprise $\psi^* x_t$ connections using the rigid technology and $(1 - \psi^*) x_t$ connections using the flexible technology. The rigid connections cost $c_1 \psi^* x_t$, and the flexible ones cost $c_2 (1 - \psi^*) x_t$.

4 Concluding remarks

In this paper we developed a model of a regulated firm which incorporated time, uncertainty, and irreversibility in investment, with possible substitutable reversible technology.

The irreversible nature of investment, combined with increased competition and the emergence of new technologies, means that any investment in network infrastructure may be rendered uneconomic before the end of its physical life. The forms of regulation we analyzed allocate this risk in different ways. Under traditional regulation, the regulator, and ultimately consumers, bear the risk of asset stranding. Connected consumers benefit when demand is high and the network is operating at full capacity since this is when their individual contribution is low. However, when demand for the network falls, the remaining consumers must each bear a greater burden of the network's cost. Thus, the burden falls on the remaining customers when some customers abandon the network. In contrast, under incentive regulation customers pay for their option to abandon the network while they are still connected. The risk of asset stranding is borne by the firm's shareholders and, because shareholders are better able to diversify this risk, overall welfare is higher under incentive regulation than traditional regulation.

The two forms of regulation have different impacts on the regulated firm's investment incentives. With traditional regulation, the firm is guaranteed to recover the cost of any investment, even if the investment is unnecessary (in the case of the simple model) or is inconsistent with an efficiently-configured network (in the case of the dual-technology model). Reasonable returns delivered by traditional regulation thus result in inefficient investment; consumers not only bear the risk of asset stranding, but this risk is greater than it need be. In contrast, incentive regulation allows the regulator to delegate investment decisions to the firm: in the simple model, the firm will not invest in unnecessary connections; in the dual-technology model, the firm will configure its network efficiently. In addition, the information requirements are greater with traditional regulation. For traditional regulation to be reasonable, the regulator must be able to observe the network's composition. In contrast, under incentive regulation the regulator only needs to observe the number of customers.

While traditional regulation may be feasible when entry to the regulated market is prevented, serious problems arise in the now typical situation where a regulated incumbent competes with unregulated entrants. Such competition adds to the volatility of the incumbent's customer numbers, and therefore to the volatility of the prices faced by customers. Although we did not include it in our model, traditional regulation will accelerate the decline in customer numbers: if customer numbers fall, the burden of generating the regulated firm's guaranteed revenue falls on the remaining customers, who are thus more likely to abandon the network in favor of one of

its competitors. Furthermore, the regulator's commitment to guarantee the regulated firm the reasonable revenue flow is not credible: a regulator is unlikely to allow the incumbent to collect more revenue from each of its remaining customers after it loses market share to a competitor. If a regulator levies the entire industry to support an incumbent firm's loss of customers to competitors, it will inhibit the process of competition.³¹

A final factor favoring incentive regulation is its less demanding information requirements. For traditional regulation to generate reasonable revenues, the regulator must be able to observe the composition of the network, not just its total capacity. If the regulator can only observe the network's total capacity, traditional regulation induces the firm to invest in too much of the rigid technology. In contrast, incentive regulation requires only that the regulator can observe the number of customers connected to the network. One might argue that calculating reasonable revenue under incentive regulation is more complicated, since it requires knowledge of parameters such as the drift and volatility of customer numbers. However, the calculation of reasonable revenue only needs to be performed once, unlike traditional regulation which requires that revenues are revised frequently. Furthermore, while any errors in estimating these parameters will affect the value of the firm, they will not affect its investment policy: regardless of regulatory error, the firm configures the network efficiently.

Our approach treats the distribution of demand as exogenous to prices although welfare is affected by customer prices. It enables formal analysis of the implications for regulation of the potential for stranding in the context of irreversible investment. However, it would be desirable to extend the approach by rendering customer numbers endogenous. This would involve modeling competition, and therefore the strategic exercise of real options. Other possible extensions include allowing competitive firms a choice of technology, and permitting technological change in the network and thereby a further source of stranding.

References

- Asplund, Marcus (2000). "What fraction of a capital investment is sunk costs?" *Journal of Industrial Economics* 48(3), 287–304.
- Beard, T. Randolph, David L. Kaserman, and John W. Mayo (2003). "Regulation, competition, and the optimal recovery of stranded costs," *International Journal of Industrial Organization* 21(6), 831–848.

³¹Depending upon the design of the scheme, transfers among telecommunication-industry participants to support universal service may have this effect.

- Brennan, Michael J., and Eduardo S. Schwartz (1982a). “Consistent regulatory policy under uncertainty,” *Bell Journal of Economics* 13(2), 506–521.
- Brennan, Michael J., and Eduardo S. Schwartz (1982b). “Regulation and corporate investment policy,” *Journal of Finance* 37(2), 289–300.
- Davis, Blaine E., and F. T. Sparrow (1972). “Valuation models in regulation,” *Bell Journal of Economics and Management Science* 3(2), 544–567.
- Dixit, Avinash K. (1993). *The Art of Smooth Pasting*. Fundamentals of Pure and Applied Economics 55, Amsterdam: Harwood Academic Publishers.
- Dixit, Avinash K, and Robert S. Pindyck, R. (1994). *Investment Under Uncertainty*. Princeton: Princeton University Press.
- Economides, Nicholas (1999). “Real options and the costs of the local telecommunications network.” In: Alleman, James, and Eli Noam (eds.), *The New Investment Theory of Real Options and its Implications for the Cost Models in Telecommunications*, Kluwer Academic Publishers, Boston, MA.
- Evans, Lewis T., and Graeme A. Guthrie (2003). “Regulation, investment timing, and the choice of rate base,” NZISCR working paper.
- Fudenberg, Drew, and Jean Tirole (1991). *Game Theory*, Cambridge Mass.: MIT Press.
- Gunn, Calum (2003). “On intertemporal subsidy-free prices and economic depreciation: Constrained market pricing revisited,” *Journal of Regulatory Economics* 24(2), 135–159.
- Hausman, Jerry (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,” *Journal of Regulatory Economics* 22(3), 287–310.
- Kandel, Eugene, and Neil D. Pearson (2002). “Option value, uncertainty, and the investment decision,” *Journal of Financial and Quantitative Analysis* 37(3), 341–374.
- Leland, Hayne E. (1974). “Regulation of natural monopolies and the fair rate of return,” *Bell Journal of Economics and Management Science* 5(1), 3–15.
- Marshall, William J., Jess B. Yawitz, and Edward Greenberg. (1981). “Optimal regulation under uncertainty,” *Journal of Finance* 36(4), 909–921.

- Mayer, Colin, and John Vickers (1996). “Profit-sharing regulation: An economic appraisal,” *Fiscal Studies* 17(2), 83–101.
- Myers, Stewart C. (1972). “The application of finance theory to public utility rate cases,” *Bell Journal of Economics and Management Science* 3(1), 58–97.
- Panteghini, Paolo M., and Carlo Scarpa (2003). “Irreversible investments and regulatory risk,” CESifo Working paper No. 934,
- Ramey, Valerie A., and Matthew D. Shapiro (2001). “Displaced capital: A study of aerospace plant closings,” *Journal of Political Economy* 109(5), 958–992.
- Rodini, Mark, Michael R. Ward, and Glenn A. Woroch (2002). “Going mobile: Substitutability between fixed and mobile access,” forthcoming in *Telecommunications Policy*.
- Weisman, Dennis L. (2002). “Did the high court reach an economic low in Verizon v. FCC?” *Review of Network Economics* 1(2), 90–105.

A Proofs

A.1 Proofs of Propositions 1, 2, and 3

We start by proving the following lemma:

Lemma 2 *If the firm has the revenue function $\Pi(s_t)$ for some function Π , then the value of the firm at time t is*

$$F(x, s) = \begin{cases} \frac{\Pi(s)}{r} + A(s)x^\beta, & 0 \leq x < s, \\ \frac{\Pi(x)}{r} + A(x)x^\beta - c(x - s), & s < x, \end{cases}$$

where

$$A'(s) = cs^{-\beta} - \frac{1}{r}s^{-\beta}\Pi'(s).$$

Proof. Let $F(x_t, s_t)$ denote the value of the regulated network at time t . If $0 \leq x < s$, then capacity remains constant over the next short time interval of length dt , during which time the firm receives profit flow $\Pi(s)dt$. Thus F must satisfy

$$F(x, s) = \Pi(s)dt + e^{-r dt}E[F(x + dx, s)], \quad 0 \leq x < s,$$

where r is the riskless interest rate and the expected value is calculated using the risk-neutral process for x . This implies the differential equation

$$0 = \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 F}{\partial x^2} + (\mu - \lambda)x \frac{\partial F}{\partial x} - rF + \Pi(s), \quad 0 \leq x < s.$$

Since 0 is an absorbing boundary for x (that is, if x currently equals zero, then the network will never have any customers), the solution must satisfy the boundary condition $F(0, s) = 0$ for all s . It follows that

$$F(x, s) = \frac{\Pi(s)}{r} + A(s)x^\beta, \quad (\text{A-1})$$

where $\beta > 1$ is a solution to the quadratic equation

$$0 = \frac{1}{2}\sigma^2\beta(\beta - 1) + (\mu - \lambda)\beta - r \quad (\text{A-2})$$

and A is a function (to be determined) of the network's capacity.

If $x > s$, then the network's capacity is immediately increased by $x - s$, costing $c(x - s)$ in total. Therefore

$$F(x, s) = F(x, x) - c(x - s), \quad x > s.$$

Continuity of $\partial F/\partial s$ along the boundary $x = s$ implies that

$$F_s(x, s)\Big|_{x=s} = c.$$

Imposing this condition on the function in (A-1) implies that

$$A'(s) = cs^{-\beta} - \frac{1}{r}s^{-\beta}\Pi'(s).$$

Given the regulator's choice of $\Pi(s)$, this equation determines $A(s)$, implying that the value of the regulated firm is

$$F(x, s) = \begin{cases} \frac{\Pi(s)}{r} + A(s)x^\beta, & 0 \leq x < s, \\ \frac{\Pi(x)}{r} + A(x)x^\beta - c(x - s), & s < x. \end{cases}$$

Proof of Proposition 1

Set $\Pi(s) \equiv 0$ in Lemma 2, so that $-F(x, s)$ equals the present value of all future costs incurred in building the network. The function $A(s)$ in Lemma 2 is

$$A(s) = \frac{-cs^{1-\beta}}{\beta - 1}.$$

The result follows by noting that

$$F(x, s) = \begin{cases} \frac{-cx}{\beta-1} \left(\frac{x}{s}\right)^{\beta-1}, & 0 \leq x < s, \\ \frac{-cx}{\beta-1} - c(x - s), & s < x. \end{cases}$$

Proof of Proposition 2

The regulator chooses the function Π in Lemma 2 such that $F(x, 0) = 0 \forall x$. That is,

$$\frac{\Pi(x)}{r} + A(x)x^\beta - cx = 0 \quad \forall x.$$

Solving this equation for A and substituting the result into the above equation for A' shows that $\Pi(s) = rcs$. It follows that $A = 0$ and the value of the regulated firm is $F(x, s) = cs$ for all combinations of x and s .

Proof of Proposition 3

The value of the regulated network at time t is

$$F(x_t, s_t) = G(x_t) - C(x_t, s_t),$$

where $G(x_t)$ is the present value of the revenue flow $\Pi(x_t) dt$ and $C(x_t, s_t)$ is the cost function from Proposition 1. Since G satisfies

$$G(x) = \Pi(x)dt + e^{-r dt} E[G(x + dx)],$$

where the expected value is calculated using the risk-neutral process for x , it must satisfy the differential equation

$$0 = \frac{1}{2}\sigma^2 x^2 \frac{d^2 G}{dx^2} + (\mu - \lambda)x \frac{dG}{dx} - rG + \Pi(x). \quad (\text{A-3})$$

The regulator chooses the revenue function $\Pi(x)$ such that the value of a hypothetical rival firm is

$$0 = F(x, 0) = G(x) - C(x, 0) = G(x) - \frac{\beta cx}{\beta - 1}.$$

That is, it chooses $\Pi(x)$ in such a way that

$$G(x) = \frac{\beta cx}{\beta - 1}.$$

Substituting this expression for $G(x)$ into equation (A-3) shows that $\Pi(x)$ must satisfy

$$\Pi(x) = \left(\frac{\beta cx}{\beta - 1} \right) (r + \lambda - \mu).$$

With this revenue function, the value of the regulated firm is

$$F(x, s) = \begin{cases} \frac{cx}{\beta-1} \left(\beta - \left(\frac{x}{s}\right)^{\beta-1} \right), & 0 \leq x \leq s, \\ cs, & s < x. \end{cases}$$

A.2 Proof of Proposition 4

It is easily shown that the function

$$\Gamma(x, s) = xu \left(\frac{rcs}{x} \right)$$

is concave in (x, s) . We can therefore use the Taylor series expansion about an arbitrary point (\bar{x}, \bar{s}) to show that

$$\begin{aligned} \Gamma(x, s) &= \Gamma(\bar{x}, \bar{s}) + \Gamma_x(\bar{x}, \bar{s})(x - \bar{x}) + \Gamma_s(\bar{x}, \bar{s})(s - \bar{s}) + \frac{1}{2}\Gamma_{xx}(x^*(x), s^*(s))(x - \bar{x})^2 \\ &\quad + \Gamma_{xs}(x^*(x), s^*(s))(x - \bar{x})(s - \bar{s}) + \frac{1}{2}\Gamma_{ss}(x^*(x), s^*(s))(s - \bar{s})^2 \\ &\leq \Gamma(\bar{x}, \bar{s}) + \Gamma_x(\bar{x}, \bar{s})(x - \bar{x}) + \Gamma_s(\bar{x}, \bar{s})(s - \bar{s}) \end{aligned}$$

for some $x^*(x)$ and $s^*(s)$. The inequality is strict except when $(x - \bar{x})/x^*(x) = (s - \bar{s})/s^*(s)$.

We will use

$$\bar{x} = r \int_0^\infty e^{-rt} E_0[x_t] dt = \frac{rx_0}{r + \lambda - \mu}, \quad \bar{s} = r \int_0^\infty e^{-rt} E_0[s_t] dt = \frac{\beta x_0}{\beta - 1}.$$

Under traditional regulation, the present value of the flow of total utility is

$$\begin{aligned} U_{\text{TR}} &= \int_0^\infty e^{-rt} E_0[\Gamma(x_t, s_t)] dt \\ &< \int_0^\infty e^{-rt} E_0 [\Gamma(\bar{x}, \bar{s}) + \Gamma_x(\bar{x}, \bar{s})(x_t - \bar{x}) + \Gamma_s(\bar{x}, \bar{s})(s_t - \bar{s})] dt \\ &= \frac{\Gamma(\bar{x}, \bar{s})}{r} + \Gamma_x(\bar{x}, \bar{s}) \int_0^\infty e^{-rt} (E_0[x_t] - \bar{x}) dt + \Gamma_s(\bar{x}, \bar{s}) \int_0^\infty e^{-rt} (E_0[s_t] - \bar{s}) dt \\ &= \frac{\Gamma(\bar{x}, \bar{s})}{r} \\ &= \frac{x_0}{r + \lambda - \mu} u \left(\left(\frac{\beta c}{\beta - 1} \right) (r + \lambda - \mu) \right) \\ &= U_{\text{IR}}, \end{aligned}$$

completing the proof.

A.3 Proof of Propositions 5, 6, and 7

We start by proving the following lemma:

Lemma 3 *Suppose that the firm's allowed revenue is $\pi_1 dt$ from each rigid connection and $\pi_2 dt$ from each flexible one, and that the firm only invests in new rigid capacity when $x_t > s_t/\psi$ for some constant ψ , and then invests in enough new capacity to raise rigid capacity to ψx_t . Then the value of the firm at time t is*

$$F(x, s; \psi) = \begin{cases} \frac{\pi_1 s}{r} - \hat{f}(\psi) s \left(\frac{x}{s} \right)^\beta & 0 \leq x \leq s, \\ \frac{(\pi_1 - \pi_2 + rc_2)s}{r} + \frac{(\pi_2 - rc_2)x}{r + \lambda - \mu} + \hat{g}(\psi) s \left(\frac{x}{s} \right)^\beta + \hat{h} s \left(\frac{x}{s} \right)^{-\gamma}, & s < x \leq s/\psi, \\ F(x, \psi x; \psi) - c_1 (\psi x - s), & s/\psi < x, \end{cases}$$

where

$$\begin{aligned}\hat{f}(\psi) &= \frac{\psi^\beta}{\beta-1} \left(c_1 - c_2 - \frac{\pi_1 - \pi_2}{r} \right) - \frac{\pi_2 - rc_2}{(\beta+\gamma)(\beta-1)} \left(\frac{\beta-1}{r+\beta(\lambda-\mu)} + \frac{(\gamma+1)\psi^{\beta+\gamma}}{r-\gamma(\lambda-\mu)} \right), \\ \hat{g}(\psi) &= -\hat{f}(\psi) - \frac{(\pi_2 - rc_2)}{\beta+\gamma} \left(\frac{\gamma+1}{r+\lambda-\mu} - \frac{\gamma}{r} \right), \\ \hat{h} &= \frac{-(\pi_2 - rc_2)}{\beta+\gamma} \left(\frac{\beta-1}{r+\lambda-\mu} - \frac{\beta}{r} \right).\end{aligned}$$

Proof. We let $F(x_t, s_t)$ denote the value of the firm at date t .

- If $0 \leq x_t \leq s_t$, then the firm does not need to rent any flexible connections, total profit flow equals $\pi_1 s_t dt$, and $F(x, s)$ satisfies the ordinary differential equation

$$0 = \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 F}{\partial x^2} + (\mu - \lambda)x \frac{\partial F}{\partial x} - rF + \pi_1 s, \quad 0 \leq x < s.$$

This equation has solution

$$F(x, s) = \frac{\pi_1 s}{r} + A(s)x^\beta,$$

for some function $A(s)$.

- If $s_t < x_t \leq s_t/\psi$, then the firm must rent $x_t - s_t$ flexible connections, total profit flow equals $(\pi_1 s_t + (\pi_2 - rc_2)(x_t - s_t)) dt$, and $F(x, s)$ satisfies the ordinary differential equation

$$0 = \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 F}{\partial x^2} + (\mu - \lambda)x \frac{\partial F}{\partial x} - rF + (\pi_2 - rc_2)x + (\pi_1 - \pi_2 + rc_2)s, \quad s < x \leq \theta s.$$

This equation has solution

$$F(x, s) = \frac{(\pi_1 - \pi_2 + rc_2)s}{r} + \frac{(\pi_2 - rc_2)x}{r + \lambda - \mu} + B_1(s)x^\beta + B_2(s)x^{-\gamma},$$

for some functions $B_1(s)$ and $B_2(s)$.

- If $x_t > s_t/\psi$ then the firm will immediately increase s_t to the level ψx_t . Therefore

$$F(x, s) = F(x, \psi x) - c_1(\psi x - s).$$

We require F and $\partial F/\partial x$ to be continuous at $x = s$. These conditions lead to the equations

$$\frac{\pi_1 s}{r} + A(s)s^\beta = \frac{(\pi_1 - \pi_2 + rc_2)s}{r} + \frac{(\pi_2 - rc_2)s}{r + \lambda - \mu} + B_1(s)s^\beta + B_2(s)s^{-\gamma}, \quad (\text{A-4})$$

$$\beta A(s)s^{\beta-1} = \frac{\pi_2 - rc_2}{r + \lambda - \mu} + \beta B_1(s)s^{\beta-1} - \gamma B_2(s)s^{-\gamma-1}. \quad (\text{A-5})$$

Solving these equations for $B_1(s)$ and $B_2(s)$ gives

$$B_1(s) = A(s) - \frac{(\pi_2 - rc_2)s^{1-\beta}}{\beta + \gamma} \left(\frac{\gamma + 1}{r + \lambda - \mu} - \frac{\gamma}{r} \right) = A(s) - \frac{(\pi_2 - rc_2)s^{1-\beta}}{\beta + \gamma} \frac{1}{r + \beta(\lambda - \mu)}$$

and

$$B_2(s) = \frac{-(\pi_2 - rc_2)s^{1+\gamma}}{\beta + \gamma} \left(\frac{\beta - 1}{r + \lambda - \mu} - \frac{\beta}{r} \right).$$

We also require F and $\partial F/\partial x$ to be continuous at $x = s/\psi$. The first condition is automatically satisfied, but the second requires that

$$c_1 = \frac{\pi_1 - \pi_2 + rc_2}{r} + \psi^{-\beta} s^\beta B_1'(s) + \psi^\gamma s^{-\gamma} B_2'(s).$$

This reduces to

$$A'(s) = f(\psi)s^{-\beta},$$

where

$$f(\psi) = \psi^\beta \left(c_1 - c_2 - \frac{\pi_1 - \pi_2}{r} \right) - \frac{\pi_2 - rc_2}{\beta + \gamma} \left(\frac{\beta - 1}{r + \beta(\lambda - \mu)} + \frac{(\gamma + 1)\psi^{\beta+\gamma}}{r - \gamma(\lambda - \mu)} \right).$$

Integrating with respect to s gives

$$A(s) = -\frac{s^{1-\beta} f(\psi)}{\beta - 1}.$$

Substituting the solution for $A(s)$ into the expressions above for $B_1(s)$ and $B_2(s)$ allow us to calculate the value of the firm in all possible states of nature. The result is the function given in the statement of the lemma.

Proof of Proposition 5

The result follows if we set $\pi_1 = \pi_2 = 0$ in Lemma 3, so that $-F(x, s)$ equals the present value of all future costs incurred in building the network.

Proof of Proposition 6

We analyze the situation by working backwards through time. Given the value of the firm derived in Lemma 3, we first calculate the firm's optimal investment policy as a function of the revenue parameters π_1 and π_2 . We then calculate the revenue flows which are reasonable.

The firm will choose the expansion threshold ψ which maximizes $F(x, s; \psi)$. That is, it will choose ψ in order to minimize $\hat{f}(\psi)$. Notice that

$$\hat{f}(\psi) = a_0 + a_1\psi^\beta - a_2\psi^{\beta+\gamma},$$

where

$$\begin{aligned} a_1 &= \frac{(\pi_2 - rc_2) - (\pi_1 - rc_1)}{r(\beta - 1)}, \\ a_2 &= \frac{(\gamma + 1)(\pi_2 - rc_2)}{\gamma(\beta + \gamma)(r + \lambda - \mu)}, \end{aligned}$$

and the constant a_0 is unimportant. It follows that $f'(\psi) = 0$ if and only if $\psi = 0$ or $\psi = \hat{\psi}$, where

$$\hat{\psi}^\gamma = \frac{a_1\beta}{a_2(\beta + \gamma)} = \frac{\pi_2 - \pi_1 - rc_2 + rc_1}{\pi_2 - rc_2},$$

and that

$$f''(\hat{\psi}) = -a_1\beta\gamma\hat{\psi}^{\beta-2}.$$

We can classify the possible scenarios into five distinct cases.

Case 1 If $a_1, a_2 > 0$, then $\hat{\psi} > 0$ and $\hat{f}(\psi)$ is increasing on the interval $(0, \hat{\psi})$ and decreasing for larger values of ψ . The firm will therefore choose either 0 or 1 as its investment threshold, depending on which of $\hat{f}(0)$ or $\hat{f}(1)$ is smaller.

Case 2 If $a_1 > 0$ and $a_2 \leq 0$ (or $a_1 \geq 0$ and $a_2 < 0$), then \hat{f} is strictly increasing in ψ and the firm will choose $\psi = 0$ as its investment threshold.

Case 3 If $a_1 < 0$ and $a_2 \geq 0$ (or $a_1 \leq 0$ and $a_2 > 0$), then \hat{f} is strictly decreasing in ψ and the firm will choose $\psi = 1$ as its investment threshold.

Case 4 If $a_1, a_2 < 0$, then $\hat{\psi} > 0$ and $\hat{f}(\psi)$ is decreasing on the interval $(0, \hat{\psi})$ and increasing for larger values of ψ . The firm will therefore choose $\psi = \hat{\psi}$ as its investment threshold if $\hat{\psi} \leq 1$, and will choose $\psi = 1$ otherwise.

Case 5 If $a_1 = a_2 = 0$ then $\hat{f}(\psi)$ is constant and the firm is indifferent between all possible investment thresholds.

Finally, we calculate the revenue flows which are reasonable. Suppose the regulator chooses π_1 and π_2 . Then, from above, the firm responds by choosing the investment threshold 0, 1, or some intermediate value. Firstly, if the firm chooses $\psi = 0$, meaning that it only uses the flexible technology, then its value satisfies $F(x, 0; 0) = (\pi_2 - rc_2)/(r + \lambda - \mu)$. For the revenue function to be reasonable, the regulator must set $\pi_2 = rc_2$. Secondly, if the firm chooses $\psi = 1$, meaning that it only uses the rigid technology, then its value satisfies

$$F(x, 0; 1) = \frac{(\pi_1 - rc_1)x}{r} \cdot \frac{\beta}{\beta - 1}.$$

A reasonable revenue function must therefore have $\pi_1 = rc_1$. Thirdly,

$$F(x, 0; \hat{\psi}) = \frac{r(\hat{\psi}(\pi_1 - rc_1) + (1 - \hat{\psi})(\pi_2 - rc_2))x}{r + \lambda - \mu}.$$

It is straightforward to show that any combination of π_1 and π_2 for which $\hat{\psi} < 1$ implies that $F(x, 0; \hat{\psi}) < 0$. In particular, it is impossible to construct a reasonable revenue function which

induces the firm to invest in both technologies. Therefore, we only need to consider revenue functions which induce the firm to choose $\psi = 0$ or $\psi = 1$.

If the regulator induces the firm to choose $\psi = 0$, then it must set $\pi_2 = rc_2$, implying that $a_2 = 0$, if the revenue function is to be reasonable. The desired firm behavior can only be achieved in Case 2 (provided $a_1 > 0$), and Case 5 (provided $a_1 = 0$). Thus, the reasonable revenue function has $\pi_1 \leq rc_1$ and $\pi_2 = rc_2$.

If the regulator induces the firm to choose $\psi = 1$, then it must set $\pi_1 = rc_1$, implying that a_1 and a_2 have the same sign. The desired firm behavior can only be achieved in Case 1 (provided $\hat{f}(0) > \hat{f}(1)$), Case 4, and Case 5 (provided $a_1 = a_2 = 0$). Thus, the reasonable revenue function has $\pi_1 = rc_1$ and $\pi_2 \leq rc_2$.

Proof of Proposition 7

Unlike with traditional regulation, here the firm cannot influence its total revenue flow, so the only way for the firm to maximize its value is to minimize its costs. Thus, it will choose the efficient investment threshold ψ^* . The value of the regulated network at time t is therefore

$$F(x_t, s_t; \psi^*) = \frac{\pi x_t}{r + \lambda - \mu} - C(x_t, s_t; \psi^*),$$

since $\pi x_t / (r + \lambda - \mu)$ is the present value of the revenue flow and $C(x_t, s_t)$ is the cost function from Proposition 5. The regulator chooses π such that

$$0 = F(x, 0; \psi^*) = \frac{\pi x}{r + \lambda - \mu} - C(x, 0; \psi^*).$$

That is, it chooses

$$\pi x = (r + \lambda - \mu)C(x, 0; \psi^*).$$

B Calibration

In this appendix we demonstrate how the model in Section 2 can be calibrated to a telecommunications network and to electricity distribution networks.

B.1 Telecommunications networks

Telecom New Zealand Limited (TCNZ) has provided data on the number of customers connected to each of its 765 exchanges at 31 March 2000 and at 1 July 2002. All figures are broken down into business and residential customers. For the purposes of this exercise, we assume that (actual) growth rates in customer numbers are distributed independently across exchanges and across

time.³² Since the sample represents 27 months, we annualize means by multiplying by $12/27$, and standard deviations by multiplying by $(12/27)^{0.5}$. However, some small exchanges experience very large percentage changes in customer numbers, which could dominate sample means and standard deviations. We adopted several approaches to overcome this problem. Firstly, the data set is split into quartiles based on exchange size (measured by the total number of customers at the start date), and the annualized mean and standard deviation of returns are calculated for the exchanges in each quartile. Secondly, using the whole sample, when calculating the mean and standard deviation we weight each exchange by the proportion of TCNZ's total customers connected to it. For example, if one percent of TCNZ's residential customers connected to a particular exchange, then a weight of 0.01 is attached to its growth rate when calculating the mean and standard deviation of the growth in residential customers. This has the effect of reducing the importance of those very small exchanges which experienced explosive growth during this period. Finally, we calculate growth rates in customer numbers for each of TCNZ's 13 regions, and calculate the mean and standard deviation of this sample of 13 growth rates. This will underestimate the relevant volatility, and therefore underestimate the cost multiplier.

Table 5 reports the results of our analysis. For each of the separate cases described above, it shows the estimated mean and standard deviation of the growth rate in customer numbers (corresponding to μ and σ respectively), the cost multiplier $\beta/(\beta - 1)$, as well as the reasonable rate of return (\hat{r}_{IR}). The later two quantities are calculated using a riskless interest rate of five percent (that is, $r = 0.05$) and a risk premium of zero (that is, $\lambda = 0$). As predicted, the regional volatilities are very low, and therefore the reasonable rate of return is only slightly greater than the riskless rate in this case. The volatilities calculated using unweighted averages are relatively high, and lead to reasonable rates of return in the range 2–4 percent higher than the riskfree rate. When we weight the observations as described above, returns in the range 0.5–3.5 percent are reasonable. A more detailed picture emerges when we break the data into quartiles. The largest exchanges have fairly stable customer numbers, and therefore relatively low volatilities. Reasonable returns are not much greater than the riskfree rate. Inspection of Table 5 shows that as the exchange shrinks, customer numbers become more volatile. For smaller exchanges, a reasonable rate of return is between three and five percentage points higher than the riskfree rate.

³²Consistent with the model's assumption that customer numbers evolve according to geometric Brownian motion, we use continuously compounded growth rates. That is, if customer numbers change from x_0 to x_1 , the growth rate equals $\log(x_1/x_0)$.

Table 5: Parameterizing the model for a telecommunications network

Quartile	Drift (μ)			Volatility (σ)			$\beta/(\beta - 1)$			\hat{r}_{IR}		
	Bus.	Res.	Total	Bus.	Res.	Total	Bus.	Res.	Total	Bus.	Res.	Total
1	0.022	0.044	0.043	0.278	0.253	0.238	3.64	13.77	11.58	0.103	0.084	0.081
2	0.004	0.022	0.023	0.245	0.162	0.171	2.25	2.58	2.70	0.104	0.072	0.073
3	-0.003	0.015	0.012	0.084	0.071	0.066	1.27	1.62	1.50	0.067	0.057	0.056
4	-0.014	0.014	0.003	0.059	0.045	0.041	1.10	1.46	1.18	0.070	0.053	0.055
Regions	-0.010	0.015	0.007	0.034	0.019	0.021	1.05	1.46	1.18	0.063	0.051	0.051
Unweighted	0.002	0.026	0.022	0.199	0.163	0.160	1.91	2.89	2.55	0.092	0.070	0.071
Weighted	-0.030	0.015	0.000	0.069	0.061	0.059	1.07	1.58	1.21	0.086	0.055	0.060

Notes. The first six columns report annualized mean and standard deviations of growth rates in customer numbers at exchanges operated by Telecom New Zealand Ltd using data covering the period from 31 March 2000 to 1 July 2002. The next three columns report the cost multiplier $\beta/(\beta - 1)$ from Proposition 1. The final three columns report reasonable rates of return if the firm is compensated ex ante for the risk of asset stranding. Results are decomposed into business and residential customers, and grouped according to the size of each exchange, and the weighting scheme used to calculate the mean and standard deviations.

B.2 Electricity distribution networks

We calculated continuously compounded annual growth rates in the total electricity supplied for each of New Zealand’s 29 electricity distribution companies during the period 1995-2002, giving a crude measure of the annual change in the capacities required by these networks. The average annual growth rate is $\mu = 0.0196$ and the standard deviation is $\sigma = 0.0446$. These parameters imply $\beta = 2.384$, a cost multiplier of $\beta/(\beta - 1) = 1.723$ and, assuming $r = 0.05$ and $\lambda = 0$, a reasonable rate of return of $\hat{r}_{\text{IR}} = 0.0524$.