Dynamically-Efficient Incentive Regulation of Networks with Sunk Costs

Lewis Evans*

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

Graeme Guthrie

Victoria University of Wellington, Wellington, New Zealand

September 16, 2002

*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

Dynamically-Efficient Incentive Regulation of Networks with Sunk Costs

Abstract

Incentive regulation allows decentralised decision-making under regulatory parameters set on the basis of industry characteristics. When there is uncertainty, sunk costs, and flexibility in the timing of investment a monopoly will invest later than is socially desirable because it garners only a fraction of the benefits. This study considers the design of regulatory profit caps based on a measure of cost, either historical or replacement cost, to which a regulatory rate of return is applied. It demonstrates that the sources and extent of supply and demand uncertainties, and thereby characteristics of the industry, determine whether historical or replacement cost regulation is desirable. The welfare optimising level of the regulatory rate of return differs between historical and replacement cost regulation, this return is generally higher than the weighted average cost of capital, and welfare is degraded much more if it is set below, as opposed to above, the optimal regulatory return.

JEL Classification code: G18, G31, L5

Keywords: Uncertainty, Sunk Costs, Incentive Regulation

Dynamically-Efficient Incentive Regulation of Networks with Sunk Costs

1 Introduction

This paper examines welfare produced by regulated firms that have a significant proportion of their costs generated by irreversible (sunk) investment. The extent to which investment in networks is subject to uncertainty and is sunk is arguable (see Hausman (1999) and Economides (1999) for the debate).[1] However, it is widely accepted that there is significant uncertainty attached to network costs and demand, and that much investment in physical networks is sunk because it has a specific use and because, once in place, it is very expensive to recover for use elsewhere. The irreversibility implies that the value attached to waiting is a critical determinant of the optimal time to invest. Because even a monopoly does not garner all the surplus from investment, it may delay the act of investing beyond that desired by a welfare-maximizing social planner.[2]

In this paper we examine the effect on the timing of investment of a regulation-imposed profit cap that is composed of a cost-base and a regulatory rate of return. We consider both the level of the regulatory rate of return (the "cap-return") and whether it is applied to historical cost or future best-practice replacement cost of the network. The regulation is dynamically efficient if it produces the largest (present value) of the sum of producer and consumer gain to the indefinite future.

Our setting is regulation within the context of decentralised investor decision-making. It presumes that the only instrument of regulatory policy is the profit cap. There are three points to be made about such an environment. First, the approach is in the spirit of most forms of incentive price regulation and rate of return regulation.[3] Incentive regulation can be viewed as a weak form of rate of return regulation. In the latter, prices are set on the basis of a revenue requirement generated by actual, or estimated, production costs and a regulatory-chosen rate of return. Incentive regulation is weaker in that profits are allowed to exist without the regulator adjusting the cap over intervals of time, but ultimately the determination of the level of the cap is revised and typically a significant input to the revision is some indicator of profitability. We apply the pure form of incentive regulation

[1]Hausman (1999) argues that, because of their sunk costs, telecommunications networks cannot mimic a contestable market but must be competitively imperfect. He criticizes the ECPR rule of Baumol on that basis. Hausman considers regulatory schema with forward-looking caps. The rule is agnostic as to whether the investments should be made by an incumbent or the entrant: the same criterion applies.

[2]The notion that investment has significant beneficial externalities extends at least as far back as Sidgwick (1887) as cited in Baumol (p. 122, 2000).

[3]The profit cap could be replaced with a revenue, even price, cap without altering the qualitative features of our results under certain conditions.

where the cap is entirely determined by fundamental demand and supply characteristics of the industry.

Second, our approach enables decentralised decision-making subject only to a profit cap that is designed to maximise welfare by the level of the regulatory rate of return and by its application to historical cost or future best-practice costs. In fact, under many forms of incentive regulation, and more so under standard rate of return regulation, investment is significantly affected by the regulator. Regulatory prescription of investment that firms may not otherwise have carried out has precipitated the phenomenon of stranded assets that have constrained deregulation and the adoption of technological advance in electricity and telecommunications in the USA since the 1970s (Spulber, 1989; Sidak and Spulber, 1997). In our model the profit cap allocates the total surplus generated by the relevant market between the firm's shareholders and consumers, but it does not affect the size of the surplus. Social welfare is affected by the timing of investment. We do not specifically consider the issues of divergence of firm and regulator actions made possible by asymmetric information holdings between the firm and the regulator, or allow the regulator to dictate investment. The effect on consumer and economic welfare of mis-timed investment or product adoption is often very substantial and we examine this independently of other regulatory issues.[4]

Thirdly, our approach treats uncertainty explicitly. Uncertainty stemming from both systematic risk and industry-specific idiosyncratic risk properly affects investment decisions and these are jointly considered in our model. The standard approach incorporates systematic risk in the calculation of a firm's weighted average cost of capital (WACC). However, systematic risk is but one element of risk that investment decisions must consider if they are to enhance economic efficiency. To ignore industry-specific risks is to ignore firms' inputs to investment decision-making and thereby important factors determining investment. It is rational and socially desirable for investment to recognise these sources of uncertainty because of the economic costs of getting the timing of investment wrong and the potential downside of investments (in particular, the costs of bankruptcy). Our model suggests that demand and supply uncertainty are critically important in setting the level of the cap-return and in the choice of historical or future cost base. Indeed, these industry-specific characteristics determine which should be chosen, and they affect the extent to which systematic risk should be incorporated in the design of the cap.

We find that if a regulated firm has flexibility in timing of investment, then regulating it utilising future replacement (forward looking) cost as an input induces some delay in investment. For some levels of the cap-return, utilising historical (backward looking) cost

[4]Goolsbee (2000) makes the point that the welfare cost of delayed investment may be very high especially where new products are being introduced. The cost arises because, in contradistinction to welfare analysis of existing products, delayed investment in new products results in a "missing market" where all consumer and producer surplus is "missing".

as an input can lead to earlier investment than the unregulated case, but only if the network's replacement cost is expected to fall and/or is significantly negatively correlated with changes in the welfare surplus produced. Otherwise, backward-looking regulation also leads to later investment. We conjecture that telecommunications — with rapid technical change producing future cost uncertainty and cost reductions correlated with gains in consumer welfare — is more a candidate for historical cost regulation than are more technologically stable industries such as gas and electricity transmission. We show that the systematic risk component of the allowed rate of return differs between backward and forward looking regulation and that the welfare effect of the cap-return is strongly asymmetric: setting the cap-return below the optimal level produces a much greater degradation of consumer and total welfare than does setting the cap-return too high. We also note that the optimal level of the cap-returns is higher than the firms' WACCs in our examples, and that this is consistent with evidence about investment hurdle rates that firms have been observed to adopt.[5]

Our approach is closest to Guthrie, Small and Wright (2001), which is the only other formal analysis of this issue that we are aware of. That paper considers setting cost-based access charges for a network whose cost is uncertain. The profit and consumer surplus flow from that access charge is known. Guthrie, Small and Wright's broad conclusion is that historical cost is preferred on welfare grounds to replacement cost in most situations. Only when replacement cost is expected to rise with little uncertainty is it likely to be preferred as it induces earlier investment. Our paper differs from Guthrie, Small and Wright in respect of its incorporation of uncertainty in both network cost and economic surpluses. Our different results stem in large part from the importance of the correlation between these surpluses and costs, but they are affected by other characteristics of the market as well.

2 Model set-up

A firm has the perpetual right to invest in a network. If investment occurs at date T, the firm incurs a lump sum (real) cost of k_T which evolves according to the geometric Brownian motion

$$dk_t = \mu_k k_t dt + \sigma_k k_t d\zeta_t,$$

where μ_k and σ_k are constants and ζ_t is a Wiener process. We assume that at any time at which the network is operating, it will die during the next short interval of length dt with probability $\phi\, dt$, for some constant ϕ. Thus, the network's lifetime is drawn from an exponential distribution, and the expected lifetime of each incarnation of the network is $1/\phi$ years. Immediately after the network dies, the firm has the right to rebuild. If

[5] Poterba and Summers (1995) provide evidence of investment hurdle rates for Fortune 1000 firms.

reconstruction occurs at date T', the firm must pay the lump sum $k_{T'}$ at that date.[6] Investment is irreversible.

If the network is in place at date t, it generates a flow of total surplus θ_t. This evolves according to the geometric Brownian motion

$$d\theta_t = \mu_\theta \theta_t dt + \sigma_\theta \theta_t d\xi_t,$$

where μ_θ and σ_θ are constants and ξ_t is a Wiener process. The two shocks can be correlated: $d\zeta_t d\xi_t = \rho_{k\theta} dt$ for some constant $\rho_{k\theta} \in [-1, 1]$. We assume that the firm can extract a constant proportion γ of the total surplus if the network is not regulated. That is, the unregulated network generates a flow of profit equal to $\gamma\theta_t$ and of consumers' surplus equal to $(1-\gamma)\theta_t$.

The stochastic process for the replacement cost of the network can reflect the adoption of best-practice techniques, which may also produce enhanced services that are reflected in the total surplus generated by the network, as well as capital price changes. The correlation between shocks to economic surplus and capital costs can reflect this. For example, a negative correlation could reflect network cost reductions that are associated with service enhancements that produce increases in economic surpluses. In this model shocks arrive independently over time, but the use of geometric Brownian motion implies that they will have long term effects. The parameters of these processes determine the firm's behavior and thereby affect the optimal regulatory profit cap.

In our model a regulator acts as a von Stackelberg leader. It imposes a cap on the firm's profit flow from the network, based either on the replacement or historical cost of the network.[7] The firm responds by choosing its investment policy. We evaluate the various regulatory policies by valuing the flow of total surplus from the network assuming the firm follows the profit-maximizing investment policy.

We consider two different forms of regulation. In the first case, the level of the profit cap is proportional to the cost of replacing the network. Thus, if investment occurred at time T, profit at date $T+t$ is capped at $\rho_{rc}k_{T+t}$, where ρ_{rc} is the maximum allowable rate of return (the cap-return). Actual profit at date $T+t$ is

$$\hat{\pi}_{T+t} = \min\{\gamma\theta_{T+t}, \rho_{rc}k_{T+t}\}. \tag{1}$$

The second case caps the profit according to the cost of building the network, measured at the time investment took place. Thus, if investment occurred at date T, and the network

[6]Dixit and Pindyck (pp. 202–204, 1994) make this assumption when modelling depreciation. We mimic their procedure for determining the firm's optimal investment policy, although the regulation introduced below makes the analysis much more complicated.

[7]This could be implemented by taxing network profits above this cap. Alternatively, knowledge of the cap may constrain the behavior of the network in such a way that it always keeps its realised profit at or below the level of the cap.

is still operating t years later, profit at date $T + t$ is capped at $\rho_{hc} k_T$, where ρ_{hc} is the cap-return. The actual profit at date $T + t$ therefore equals

$$\hat{\pi}_{T+t} = \min \{\gamma \theta_{T+t}, \rho_{hc} k_T\} . \tag{2}$$

We assume that regulation has no effect on the total surplus. Therefore the flow of consumers' surplus at date $T + t$ equals $\theta_{T+t} - \hat{\pi}_{T+t}$.[8]

We value these future flows of profit and consumers' surplus using contingent claims analysis. Since k_t and θ_t are not the prices of traded assets we make the standard assumption that traded assets exist with returns which are perfectly correlated with dk_t and $d\theta_t$. Suppose that the expected rates of return for these assets are $r + \lambda_k$ and $r + \lambda_\theta$ respectively, where r is the real riskless interest rate and the λs are risk premia which can be derived from an equilibrium asset pricing model such as the CAPM. Any future cash flow can then be valued by discounting the expected cash flow at the real riskless interest rate, provided that the expected value is calculated using the following 'risk-neutral' process for the state variables[9]:

$$\begin{aligned} dk_t &= (\mu_k - \lambda_k) k_t dt + \sigma_k k_t d\zeta_t, \\ d\theta_t &= (\mu_\theta - \lambda_\theta) \theta_t dt + \sigma_\theta \theta_t d\xi_t. \end{aligned}$$

We close this section by defining some constants which will appear frequently in the rest of the paper:

$$\begin{aligned} \mu &= (\mu_\theta - \lambda_\theta) - (\mu_k - \lambda_k), \\ \sigma^2 &= \sigma_k^2 - 2\rho_{k\theta}\sigma_k\sigma_\theta + \sigma_\theta^2, \\ \beta_1 &= \left(\frac{1}{2} - \frac{\mu}{\sigma^2}\right) + \sqrt{\frac{2(r - \mu_k + \lambda_k)}{\sigma^2} + \left(\frac{1}{2} - \frac{\mu}{\sigma^2}\right)^2}, \\ \beta_2 &= \left(\frac{1}{2} - \frac{\mu}{\sigma^2}\right) - \sqrt{\frac{2(r - \mu_k + \lambda_k)}{\sigma^2} + \left(\frac{1}{2} - \frac{\mu}{\sigma^2}\right)^2}, \\ \beta_3 &= \left(\frac{1}{2} - \frac{\mu}{\sigma^2}\right) + \sqrt{\frac{2(r + \phi - \mu_k + \lambda_k)}{\sigma^2} + \left(\frac{1}{2} - \frac{\mu}{\sigma^2}\right)^2}. \end{aligned}$$

3 Welfare and investment behavior

This section is the most technical part of the paper. Section 3.1 derives overall welfare as a function of the firm's chosen investment policy. Sections 3.2 and 3.3 then derive the firm's optimal investment behavior under the two forms of regulation. We compare the two forms of regulation in Section 4.

[8]Since $\hat{\pi}_{T+t} \leq \gamma\theta_{T+t}$, the flow of consumers' surplus will never fall below $(1-\gamma)\theta_{T+t}$. In particular, an operating network always generates a positive flow of consumers' surplus.

[9]See Cochrane (Section 3.2, 2001) for a recent textbook treatment of risk-neutral pricing.

3.1 Measuring welfare

The assumption that both k_t and θ_t evolve according to geometric Brownian motion greatly simplifies the model. It means that most of the analysis can be done in terms of the variable $y_t = \theta_t/k_t$, which evolves according to the geometric Brownian motion

$$dy_t = \mu y_t dt + \sigma y_t d\eta_t,$$

where μ and σ are constants defined at the end of Section 2 and η_t is a Wiener process.[10] In particular, we can consider investment policies of the type "invest if θ/k exceeds a given threshold, otherwise delay investment." The timing of investment thus depends on both the cost of building the network and the flow of surplus it can generate. Consider what happens when an existing network breaks down. If the network's replacement cost is sufficiently low, relative to the total surplus, then it will be rebuilt immediately and the flow of total surplus will be uninterrupted. However, if the replacement cost is relatively high, (re)investment may be delayed. The total surplus will not begin to flow until the relative cost has fallen sufficiently.

Our first result calculates the value of the future flow of total surplus assuming the firm adopts an investment policy of this form.[11]

Proposition 1 *Suppose the firm invests as soon as $\theta/k \geq y^*$ for some constant y^*. Then at any time prior to investment, the value of the flow of total surplus equals*

$$F(k, \theta; y^*) = \theta^{\beta_1} k^{1-\beta_1} \frac{U(y^*)}{(y^*)^{\beta_1}},$$

where

$$\begin{aligned} U(y^*) \quad = \quad & \left(\frac{1-\beta_2}{\beta_1-\beta_2}\right) \frac{y^*}{r+\lambda_\theta-\mu_\theta} + \left(\frac{\beta_3-1}{\beta_1-\beta_2}\right) \frac{y^*}{r+\lambda_\theta+\phi-\mu_\theta} \\ & + \left(\frac{\beta_2}{\beta_1-\beta_2}\right) \frac{\phi}{r+\lambda_k-\mu_k} - \left(\frac{\beta_3-\beta_2}{\beta_1-\beta_2}\right). \end{aligned} \quad \blacksquare$$

Consider the special case where, once built, the network will last forever. In this case $\phi = 0$, $\beta_3 = \beta_1$ and

$$\frac{F(k, \theta; y^*)}{k} = \left(\frac{\theta}{ky^*}\right)^{\beta_1} \left(\frac{y^*}{r+\lambda_\theta-\mu_\theta} - 1\right).$$

That is, taking the network's replacement cost as the numeraire, the present value equals the net present (social) value of building the network, scaled by a factor less than one

[10] Exploiting the problem's homogeneity in this way can be traced back at least to Margrabe (1978) and his analysis of the option to exchange one asset for another, which was greatly simplified by taking one asset as numeraire. Here all prices are expressed relative to the replacement cost of the network.

[11] Proofs for all results can be found in the appendix.

which accounts for the delay until investment occurs. If y^* is high, the investment payoff will be high. However, we may have to wait a long time for θ/k to reach this threshold, in which case the high investment payoff will be heavily discounted.

Proposition 1 gives the dependence of overall welfare on the firm's investment threshold. The firm will decide the timing of investment in practice. However, we can determine the investment decision rule which a welfare-maximizing social planner would choose. If investment in the network could not be delayed and the network could not be rebuilt, the present value of the future flow of total surplus would equal $\theta/(r+\lambda_\theta-\mu_\theta+\phi)$. Investment would therefore be socially optimal if $\theta/k \geq r+\lambda_\theta-\mu_\theta+\phi$. The irreversibility of investing in the network makes the investment timing option valuable, raising the investment threshold above $r+\lambda_\theta-\mu_\theta+\phi$. However, as Dixit and Pindyck (p. 204, 1994) point out, the ability not to rebuild the network in the future reduces the effect of irreversibility — the original investment decision can effectively be reversed by not rebuilding the network when it dies. This has the effect of lowering the threshold. The precise investment threshold chosen by a social planner is given in Corollary 1.

Corollary 1 *A social planner would invest as soon as*

$$\frac{\theta}{k} \;\geq\; \frac{\beta_3}{\beta_3-1}(r+\lambda_\theta+\phi-\mu_\theta).$$

■

It is straightforward to show that provided $\mu_k < r+\lambda_k+\phi$ and $\mu_\theta < r+\lambda_\theta+\phi$ then the 'option multiplier' $\beta_3/(\beta_3-1)$ is greater than 1.[12] That is, the social planner will not invest unless the present value of the flow of total surplus is significantly greater than the cost of building the network.

The cost of delay is that society does not receive the flow of total surplus until some later date. This opportunity cost must be balanced against the value of waiting. If investment is delayed and either the network's construction cost falls or the flow of total surplus rises then investment can occur on more favourable terms; that is, the investment's (social) payoff will be higher. On the other hand, if the cost of building the network rises or the flow of total surplus falls, investment can be further delayed. This ability to further delay investment introduces an asymmetry into the payoff from delaying investment. There is unlimited upside potential, but limited downside potential. As a result, the expected payoff from delaying investment will be positive. Investment will only occur when the opportunity cost of delaying investment exceeds this expected payoff from delay.

If the profit of the firm is unregulated, it receives a profit flow $\pi_t = \gamma\theta_t$ which is proportional to the total surplus. When deciding when to invest, the firm measures the cost of building the network against the present value of this flow. Both the firm and the hypothetical social planner face the same lump sum cost, but the firm receives only

[12]In plausible circumstances the multiplier may be greater than 2. See Dixit and Pindyck (p. 204, 1994).

a fraction (γ) of the flow. Thus, the firm will wait longer before investing. The following corollary of Proposition 1 describes the exact investment behavior of the unregulated firm.

Corollary 2 *If the firm is not regulated in any way, so that its profit flow is $\pi_t = \gamma\theta_t$, it will invest whenever*

$$\frac{\gamma\theta}{k} \geq \frac{\beta_3}{\beta_3 - 1}(r + \lambda_\theta + \phi - \mu_\theta). \qquad \blacksquare$$

The firm waits for profit to exceed the threshold, whereas the social planner waits for surplus to exceed it. When both investment rules are expressed in terms of a threshold for θ/k it is clear that the unregulated firm sets a threshold which exceeds the social planner's by a factor of $1/\gamma > 1$.[13]

One possible response to this problem is the introduction of some form of regulation which alters the firm's investment incentives. We consider profit caps in the remainder of this section.

3.2 Replacement cost

We start by supposing that the firm's maximum allowable profit is proportional to the replacement cost of its network, as in equation (1). The next proposition describes the investment policy chosen by the firm when faced with this form of regulation.

Proposition 2 *Suppose that the firm adopts a policy of investing whenever $\theta/k \geq y^*$ for some constant y^*, and that the regulator imposes a cap on the firm's profit flow which is proportional to the network's replacement cost. Then the firm's optimal investment threshold is defined implicitly by*

$$\beta_3 = \frac{y^* g'_{rc}(y^*)}{g_{rc}(y^*) - 1}, \tag{3}$$

where

$$g_{rc}(y) = E\left[\int_0^\infty e^{-(r+\lambda_k+\phi-\mu_k)t} \min\{\gamma y_t, \rho_{rc}\} dt \,\middle|\, y_0 = y,\ dy_t = \mu y_t dt + \sigma y_t d\eta_t\right] \tag{4}$$

and ρ_{rc} is the cap-return. At any time prior to investment, the value of the firm equals

$$\theta^{\beta_1} k^{1-\beta_1} \left(\frac{y^* h'_{rc}(y^*) - \beta_2 \left(h_{rc}(y^*) - 1 - \frac{\phi}{r + \lambda_k - \mu_k} \right)}{(\beta_1 - \beta_2)(y^*)^{\beta_1}} \right),$$

where

$$h_{rc}(y) = E\left[\int_0^\infty e^{-(r+\lambda_k-\mu_k)t} \min\{\gamma y_t, \rho_{rc}\} dt \,\middle|\, y_0 = y,\ dy_t = \mu y_t dt + \sigma y_t d\eta_t\right]. \qquad \blacksquare$$

[13] This presumes that the firm has no competition. As Dixit and Pindyck (Ch. 8 and 9, 1994) have shown, feasible competition will generally bring forward investment, in many circumstances to a date approximating that desired by the social planner.

We can use Lemma 1 in the appendix to rewrite the expression for $g_{rc}(y)$:

$$g_{rc}(y) = \int_0^\infty \left(e^{-(r+\lambda_\theta+\phi-\mu_\theta)t} \gamma y N(d_1) + e^{-(r+\lambda_k+\phi-\mu_k)t} \rho_{rc} N(d_2) \right) dt,$$

where

$$d_1 = \frac{-\log(\gamma y/\rho_{rc}) - (\mu + \frac{1}{2}\sigma^2)t}{\sigma t^{1/2}}, \quad d_2 = \frac{\log(\gamma y/\rho_{rc}) + (\mu - \frac{1}{2}\sigma^2)t}{\sigma t^{1/2}}.$$

Multiplying the resulting expression for $g_{rc}(y)$ through by k gives

$$k\, g_{rc}(\theta/k) = \int_0^\infty \left(e^{-(r+\lambda_\theta+\phi)t} (\pi e^{\mu_\theta t}) N(d_1) + e^{-(r+\lambda_k+\phi)t} (\rho_{rc} k e^{\mu_k t}) N(d_2) \right) dt.$$

The term $\pi e^{\mu_\theta t}$ is the expected unregulated profit t years after the network is built, and this is discounted at rate $r + \lambda_\theta + \phi$. The second component in the discount rate incorporates the systematic risk of surplus shocks (and hence of unregulated profit shocks), while the third adjusts for depreciation. The term $\rho_{rc} k e^{\mu_k t}$ is the expected level of the cap t years after the network is built, and this is discounted at rate $r + \lambda_k + \phi$. Since the cap is proportional to the network's replacement cost, it is subject to replacement cost shocks, explaining the second component of the discount rate — it adjusts for the systematic risk of shocks to the cap. Once more, the third component adjusts for depreciation. The integrand is (almost) a weighted average of two terms: the present value of profit, assuming the cap is non-binding, and the present value of profit, assuming the cap is binding. The 'weights' reflect the probabilities that unregulated profit will exceed the cap.[14]

The network lasts forever in the special case where $\phi = 0$. Then $\beta_3 = \beta_1$ and $h_{rc}(y) = g_{rc}(y)$ for all y. This implies that the value of the firm prior to investment is

$$\theta^{\beta_1} k^{1-\beta_1} \left(\frac{g_{rc}(y^*) - 1}{(y^*)^{\beta_1}} \right) = \left(\frac{\theta}{ky^*} \right)^{\beta_1} (kg_{rc}(y^*) - k).$$

The market value of the future profit flow at the time of investment is $kg_{rc}(y^*)$, so that the value of the firm equals the net present value at investment time, $kg_{rc}(y^*) - k$, multiplied by a factor $(y/y^*)^{\beta_1}$ to account for the delay until investment occurs.

3.3 Historical cost

In the second form of regulation we consider, the firm's maximum allowable profit is proportional to the actual cost of building its network, as in equation (2). The next proposition describes the investment policy chosen by the firm when faced with this form of regulation.

Proposition 3 *Suppose that the firm adopts a policy of investing whenever $\theta/k \geq y^*$ for some constant y^*, and that the regulator imposes a cap on the firm's profit flow which*

[14] It is not actually a weighted average because the 'weights', $N(d_1)$ and $N(d_2)$, do not sum to 1.

is proportional to the actual cost of building the network. Then the optimal investment threshold is defined implicitly by

$$\beta_3 = \frac{y^* g_{hc}'(y^*)}{g_{hc}(y^*) - 1}, \tag{5}$$

where

$$g_{hc}(z) = E\left[\int_0^\infty e^{-(r+\phi)t} \min\{\gamma z_t, \rho_{hc}\} dt \middle| z_0 = z,\ dz_t = (\mu_\theta - \lambda_\theta) z_t dt + \sigma_\theta z_t d\eta_t\right] \tag{6}$$

and ρ_{hc} is the cap-return. At any time prior to investment, the value of the firm equals

$$\theta^{\beta_1} k^{1-\beta_1} \left(\frac{y^* h_{hc}'(y^*) - \beta_2 \left(h_{hc}(y^*) - \frac{\phi}{r + \lambda_k - \mu_k} \right) + (\beta_3 - \beta_2)(g_{hc}(y^*) - 1)}{(\beta_1 - \beta_2)(y^*)^{\beta_1}} \right),$$

where

$$h_{hc}(y) = \phi E\left[\int_0^\infty e^{-(r+\lambda_k-\mu_k)t} g_{hc}(y_t) dt \middle| y_0 = y,\ dy_t = \mu y_t dt + \sigma y_t d\eta_t\right].$$

We can use Lemma 1 to rewrite the expression for $g_{hc}(y)$:

$$g_{hc}(y) = \int_0^\infty \left(e^{-(r+\lambda_\theta+\phi-\mu_\theta)t} \gamma y N(d_3) + e^{-(r+\phi)t} \rho_{hc} N(d_4) \right) dt,$$

where

$$d_3 = \frac{-\log(\gamma y/\rho_{hc}) - (\mu_\theta - \lambda_\theta + \frac{1}{2}\sigma_\theta^2)t}{\sigma_\theta t^{1/2}}, \quad d_4 = \frac{\log(\gamma y/\rho_{hc}) + (\mu_\theta - \lambda_\theta - \frac{1}{2}\sigma_\theta^2)t}{\sigma_\theta t^{1/2}}.$$

Multiplying the resulting expression for $g_{hc}(y)$ through by k gives

$$k\, g_{hc}(\theta/k) = \int_0^\infty \left(e^{-(r+\lambda_\theta+\phi)t} (\pi e^{\mu_\theta t}) N(d_3) + e^{-(r+\phi)t} (\rho_{hc} k) N(d_4) \right) dt.$$

The term $\pi e^{\mu_\theta t}$ is the expected unregulated profit t years after the network is built, and this is discounted at rate $r + \lambda_\theta + \phi$. The explanation for this discount rate is identical to that for the corresponding term in Section 3.2. The term $\rho_{hc} k$ is the (certain) level of the cap t years after the network is built, and this is discounted at rate $r + \phi$. Since the cap is based on the historical cost of building the network, it is not subject to replacement cost shocks, so there is no adjustment for the systematic risk of shocks to the cap. Once more, there is an adjustment for depreciation, and the integrand is (almost) a weighted average of two terms: the present value of profit, assuming the cap is non-binding, and the present value of profit, assuming the cap is binding. The 'weights' reflect the probabilities that unregulated profit will exceed the cap.

Consider the special case where $\phi = 0$ so that, once built, the network lasts forever. Then $\beta_3 = \beta_1$ and $h_{hc}(y) = 0$ for all y. This implies that the value of the firm prior to investment is

$$\theta^{\beta_1} k^{1-\beta_1} \left(\frac{g_{hc}(y^*) - 1}{(y^*)^{\beta_1}} \right) = \left(\frac{\theta}{k y^*} \right)^{\beta_1} (k g_{hc}(y^*) - k).$$

Notice that $kg_{hc}(y^*)$ equals the market value of the future profit flow at the time of investment, so that the value of the firm equals the net present value at investment time, $kg_{hc}(y^*) - k$, multiplied by a factor $(y/y^*)^{\beta_1}$ to account for the delay until investment occurs.

4 Comparing the two forms of regulation

This section begins with a non-technical discussion of the two forms of regulation and the intuition behind their relative performance. We use numerical analysis to obtain a deeper understanding in Section 4.2.

4.1 The bad news principle

As we have shown, in the absence of competition the unregulated firm will delay investment longer than the social planner. The purpose of regulation is to induce investment timing that more accords with that of the social planner. In evaluating profit cap regulation under historical and replacement cost, the key insight is provided by the so-called bad news principle first discussed by Bernanke (1983)[15]: the profit flow at the time of investment must be sufficient to compensate the firm not only for its investment of capital in the network, but also for any bad news which may arise following the (irreversible) investment decision. In our model, such bad news can take two forms: a fall in the network's replacement cost and a fall in profit. In the first case, if the reduction is large enough the firm will wish it had delayed investment and thus benefited from the lower construction cost; in the second case, if the reduction is large enough the firm will wish it had not invested at all. The situation is summarized in the first panel of Table 1.

Capping profit according to the network's replacement cost alters the potential for bad news to affect a firm which invested early. As in the unregulated case, a fall in replacement cost is bad news for a regulated firm which invested early — if it had delayed investment, it could have paid less for the network — as is a reduction in the total surplus. However, a reduction in the replacement cost of the network also tightens the profit cap and therefore potentially lowers the profit flow. Furthermore, because profit is capped, in any given state the present value of future profits under regulation will be less than the unregulated case. Thus the set of states in which the network becomes unprofitable grows when profits are capped according to the network's replacement cost. Some good news becomes bad news, and all bad news becomes worse. On average, the potential for bad news is made worse by this form of regulation, so that the network's initial profit flow will have to exceed its opportunity cost of capital by an even greater margin before investment occurs. Thus, this form of regulation delays investment.

[15]The ideas can also be found in Cukierman (1980).

Table 1: Profit cap regulation and the bad news principle

	Shock to	Event	Cost of project	Profit flow
Unregulated firm	Cost	$dk_t < 0$	Bad news	–
	Surplus	$d\theta_t < 0$	–	Bad news
Replacement cost	Cost	$dk_t < 0$	Bad news	Bad news
	Surplus	$d\theta_t < 0$	–	Bad news
Historical cost	Cost	$dk_t < 0$	Bad news	Good news
	Surplus	$d\theta_t < 0$	–	Bad news

Notes. The table classifies changes in the two state variables as either good news or bad news for a firm which has already invested in the network. News can impact on the cost of capital invested in the network and on the present value of the profit flow generated by the network.

The situation is more complicated when profits are capped according to the historical cost of building the network. As in the unregulated case, a fall in replacement cost is bad news for a regulated firm which invested early. However, this is tempered by the realization that if the firm had delayed investment, the fall in the network's replacement cost would have resulted in a tightening of the cap on its future profits. By investing early, the firm was able to lock in a permanently high cap. This reduces the extent of the bad news from downward movements in the network's replacement cost.[16] Since imposing a cap accentuates the effects of negative shocks to total surplus, the effect on investment timing is ambiguous. We find that this form of regulation promotes investment in some circumstances, and delays it in others, relative to the unregulated state. Two factors are relevant.

The first factor is the drift in the network's replacement cost. A downward trend in replacement cost means that a regulated firm which invests early is more likely to get the good news that by investing early it locked in a high level of the profit cap. The downward trend in replacement cost makes it especially important for the firm to invest early, and lock in the profit cap before it moves even lower. Thus, appropriate historical cost-based regulation can promote investment.

The correlation between changes in the two state variables is the second factor affecting the success of this form of regulation. If the two shocks are negatively correlated, any

[16] The effect is likely to be greatest for intermediate values of ρ_{hc}. When ρ_{hc} is very large, the cap will almost never be binding, and the reduction in present value from lowering the cap slightly will be negligible. When ρ_{hc} is small, the cap will be binding more often, but profits are capped so tightly that the reduction in present value will be small. However, for intermediate values of ρ_{hc}, the cap will bind often enough that any tightening will have a noticeable effect, while ρ_{hc} is large enough that the reduction in the cap ($\rho_{hc}\, dk_t$) will be significant.

reduction in replacement cost is likely to be accompanied by higher flows of total surplus, and therefore higher flows of potential profit — investing early locks in a higher profit cap when the firm experiences an increase in demand. In contrast, if the two shocks are positively correlated, locking in a higher profit cap is of little use, since this usually occurs when demand is falling. Thus, when shocks are negatively correlated, historical cost-based regulation has the effect of reducing the potential for bad profit news. If the allowed rate of return is chosen correctly, investment is promoted.

4.2 Numerical analysis

Because of the complexity of the model, we have to resort to numerical analysis to examine the relative performance of the two forms of regulation in more detail. Our baseline parameters are chosen in such a way that the two forms of regulation lead to identical investment behavior by the firm. We then vary crucial parameters and see how the performance of the two regimes is affected.

Careful comparison of the pair of equations (3) and (4), which describe optimal investment under replacement cost-based regulation, with equations (5) and (6), which apply to historical cost-based regulation, reveals strong similarities. In fact, if the cap-returns are the same (that is, $\rho_{rc} = \rho_{hc}$), and the risk-neutral processes for θ_t and θ_t/k_t have the same drifts and volatilities, then the two forms of regulation lead to exactly the same investment threshold. These three conditions reduce to[17]

$$\rho_{rc} = \rho_{hc}, \quad \mu_k = \lambda_k, \quad \beta_{k\theta} \equiv \frac{\rho_{k\theta}\sigma_\theta}{\sigma_k} = \frac{1}{2}. \tag{7}$$

While the historical cost and replacement cost regulatory profit caps appear similar, we have shown that they are fundamentally different at least in respect of accounting for systematic risk. As we shall demonstrate, there is no reason why the cap-returns should be the same even if they maximise welfare under the respective schemes.

Our baseline parameter values are shown in Table 2. Both conditions in (7) are satisfied. We first let $\mu_k \in \{0.01, -0.01\}$ to examine the role of drift in the network's replacement cost; all other parameters take their baseline values. Next, we choose $\rho_{k\theta} \in \{0.25, 0.75\}$, holding all other parameters equal to their baseline values.[18]

For each set of parameter values we calculate the firm's profit-maximizing investment threshold y^* as a function of the cap-return (ρ_{rc} or ρ_{hc}). Once we know the investment threshold we can calculate the level of welfare. From Proposition 1, the present value of the flow of total surplus is

$$W = \theta^{\beta_1} k^{1-\beta_1} \frac{U(y^*)}{(y^*)^{\beta_1}}$$

[17] $\beta_{k\theta}$ is the slope coefficient in the regression of changes in log θ on changes in log k.

[18] We have checked that our results hold for a wide range of baseline parameter values.

Table 2: Baseline parameters

	Cost	Surplus
Drift	$\mu_k = 0$	$\mu_\theta = 0$
Volatility	$\sigma_k = 0.05$	$\sigma_\theta = 0.05$
Correlation	$\rho_{k\theta} = 0.5$	
Risk premium	$\lambda_k = 0$	$\lambda_\theta = 0.02$
Riskfree rate	$r = 0.04$	
Depreciation	$\phi = 0$	
Unregulated profit	$\gamma = 0.3$	

Notes. The table shows the parameter values underlying the graphs in Figures 1 to 4.

in any state prior to investment. From Propositions 2 and 3 respectively, the value of the firm equals

$$\Pi_{rc} = \theta^{\beta_1} k^{1-\beta_1} \left(\frac{y^* h'_{rc}(y^*) - \beta_2 \left(h_{rc}(y^*) - 1 - \frac{\phi}{r + \lambda_k - \mu_k} \right)}{(\beta_1 - \beta_2)(y^*)^{\beta_1}} \right)$$

when its profits are capped according to the network's replacement cost, and

$$\Pi_{hc} = \theta^{\beta_1} k^{1-\beta_1} \left(\frac{y^* h'_{hc}(y^*) - \beta_2 \left(h_{hc}(y^*) - \frac{\phi}{r + \lambda_k - \mu_k} \right) + (\beta_3 - \beta_2)(g_{hc}(y^*) - 1)}{(\beta_1 - \beta_2)(y^*)^{\beta_1}} \right),$$

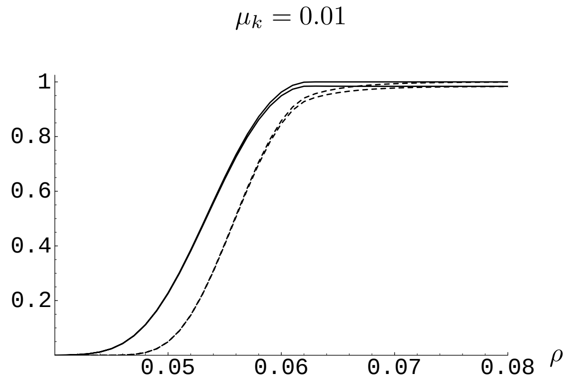
when the profit cap is proportional to the historical cost of the network. Notice that $W\theta^{-\beta_1}k^{\beta_1-1}$, $\Pi_{rc}\theta^{-\beta_1}k^{\beta_1-1}$ and $\Pi_{hc}\theta^{-\beta_1}k^{\beta_1-1}$ depend only on the investment threshold y^* and not on the state variables (k, θ). Therefore in any given state it is appropriate to rank different forms of regulation according to the value of $W\theta^{-\beta_1}k^{\beta_1-1}$ (if we are measuring overall welfare) and $\Pi\theta^{-\beta_1}k^{\beta_1-1}$ (if we are measuring profitability). We report the results of our numerical analysis in the following form:

$$\begin{aligned} \text{Welfare} &= W\theta^{-\beta_1}k^{\beta_1-1}, \\ \text{Profitability} &= \Pi\theta^{-\beta_1}k^{\beta_1-1}, \\ \text{Consumers' surplus} &= W\theta^{-\beta_1}k^{\beta_1-1} - \Pi\theta^{-\beta_1}k^{\beta_1-1}. \end{aligned}$$

These are all functions of y^*, ρ_{rc} (or ρ_{hc} where appropriate), and the parameters listed in Table 2.

Figure 1 compares the two regulatory regimes when $\mu_k = 0.01$. The top solid curve plots our welfare measure as a function of ρ_{rc}, assuming that profit is capped using the network's replacement cost; the bottom solid curve plots our measure of consumers'

Figure 1: Positive trend in replacement cost



Notes. The solid curves correspond to regulation based on replacement cost, the dashed curve to historical cost. For each pair, the upper curve plots overall welfare, and the lower curve consumers' surplus. With the exception of μ_k, all parameter values are given in Table 2.

surplus as a function of ρ_{rc}. The broken curves plot the corresponding functions for the case where profit is capped using the historical cost of building the network. The vertical axis is scaled so that welfare and consumers' surplus are expressed relative to the total welfare in the unregulated case. Thus both welfare measures converge to 1 as $\rho \to \infty$.[19]

Close inspection of the solid curves shows that capping profit based on replacement cost can make consumers better off (relative to the unregulated case), provided that the cap-return is not set too low. However, overall welfare will not be improved. Unless the cap-return is already very low, lowering it further simply results in a transfer from the firm to consumers — overall welfare is not affected. However, if the cap-return is set too low, overall welfare can be substantially lower. Regulation based on historical cost performs even worse in this case, lowering both consumers' surplus and overall welfare.

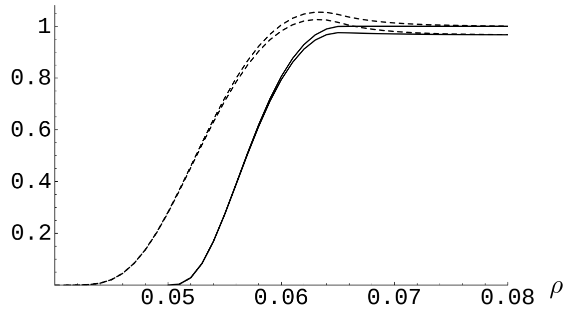
In contrast, when $\mu_k = -0.01$, historical cost is a better basis for regulation than replacement cost. The results are shown in Figure 2. The figure demonstrates that an appropriately chosen backward-looking profit cap can actually improve on the unregulated state, both in terms of consumers' surplus and overall welfare. This is consistent with our discussion in Section 4.1; the downward drift in the network's replacement cost makes it less likely that the firm would regret investing early and thereby locking in a high cap on profits.

Our final two figures concentrate on the effect of the correlation of shocks to total surplus and replacement cost. We first reset μ_k to its baseline value of zero, and set $\rho_{k\theta} =$

[19]As the cap-return is lifted, the outcome converges to the unregulated case. In contrast, if the cap-return is set too low, the firm will never invest in the network, explaining why overall welfare is zero in this region.

Figure 2: Negative trend in replacement cost

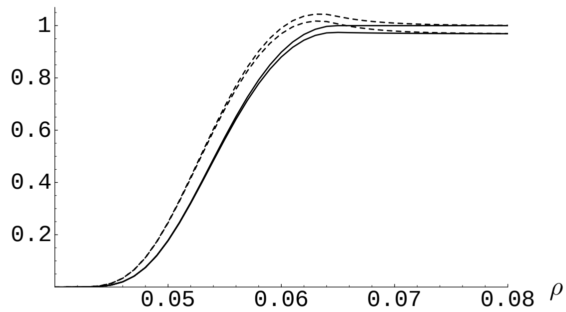
$\mu_k = -0.01$



Notes. With the exception of μ_k, all parameter values are given in Table 2.

Figure 3: Low correlation

$\rho_{k\theta} = 0.25$

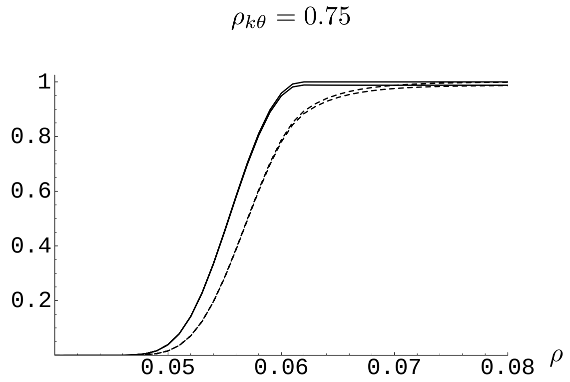


Notes. With the exception of $\rho_{k\theta}$, all parameter values are given in Table 2.

0.25. Figure 3 shows the results of our numerical analysis. The results are qualitatively similar to those in Figure 2, with historical cost outperforming replacement cost. If the backward-looking profit cap is chosen appropriately, consumers' surplus and total welfare are both higher than in the unregulated case.

Finally, we choose $\rho_{k\theta} = 0.75$. The results, reported in Figure 4, are qualitatively similar to those in Figure 1, with historical cost performing poorly. This outcome stems from the fact that under historical cost regulation and positive corrrelation between network costs and economic surplus, locking in a cap early does not protect the firm from declines in network costs and access to market surplus. The incentive to invest in early high cost periods is reduced because the effect of the cap will be moot. In these circumstances, replacement cost regulation induces relatively earlier investment because the cap on profit will adjust with economic surplus via the correlation with replacement cost.

Figure 4: High correlation



Notes. With the exception of $\rho_{k\theta}$, all parameter values are given in Table 2.

The results of our numerical analysis for the optimal cap-return are summarised in Table 3. The table reports the optimal cap-return ρ^* and a measure of its performance for the five different scenarios. For some the optimal cap-return is infinitely high and these are marked n/a in the table. We consider the caps which maximize consumers' surplus and those which maximize overall welfare. We report the level of consumers' surplus (overall welfare) as a proportion of unregulated consumers' surplus (overall welfare) in the rows labeled 'Perf.' In all the cases we consider, there exists a forward-looking profit cap which makes consumers better off than in the unregulated situation. When the drift in replacement cost is negative, and/or the correlation between the two shocks is low, a backward-looking profit cap also dominates the unregulated case, both in terms of consumers' surplus and overall welfare. The results suggest that historical cost-based regulation performs better than replacement cost when $\mu_k < \lambda_k$ and $\beta_{\theta k}$ is less than 0.5; that is, in circumstances where the risk-adjusted drift of the costs process is negative and where the correlation between economic surplus and costs is small or even negative.

5 Concluding remarks

In this paper we presented a model of a firm with monopoly rights to invest in a network. A regulator caps the firm's profit, with the maximum allowable profit based on either the actual cost of building the network or its replacement cost. Once the policy is in place, the firm, which faces uncertainty about future profit flows and the cost of replacing the network in the future, decides when to invest. The bad news principle provides an explanation of the circumstances in which alternative regulatory policies will be optimal.

To simplify the analysis, we assumed that the form of regulation has no effect on the level of total surplus, only on its distribution between producer and consumers. This

Table 3: Optimal backward and forward looking profit caps

		Consumers' surplus		Welfare
Parameters		FL	BL	BL
Baseline	ρ^*	0.0633	0.0633	n/a
		1.003	1.003	n/a
$\mu_k = 0.01$	ρ^*	0.0624	n/a	n/a
	Perf.	1.001	n/a	n/a
$\mu_k = -0.01$	ρ^*	0.0652	0.0632	0.0634
	Perf.	1.010	1.062	1.055
$\rho_{k\theta} = 0.25$	ρ^*	0.0647	0.0632	0.0634
	Perf.	1.006	1.051	1.045
$\rho_{k\theta} = 0.75$	ρ^*	0.0618	n/a	n/a
	Perf.	1.001	n/a	n/a

Notes. The table reports the replacement cost-based cap-return which maximizes consumers' surplus, and the historical cost-based cap-returns which maximize consumers' surplus and total welfare. Five different scenarios are considered. ρ^* is the optimal cap-return, and 'Perf' gives the level of consumers' surplus (overall welfare) as a proportion of unregulated consumers' surplus (overall welfare). All parameter values are given in Table 2. In particular, the WACC of an unregulated firm is $r + \lambda_\theta = 0.06$.

allowed us to concentrate on regulation's impact on investment timing. From the regulator's point of view, an unregulated firm will wait too long before investing — the firm bears the full cost of building the network, but must share some of the benefit with consumers.[20] Thus, any form of regulation which further delays investment will lower welfare even further. In contrast, regulation which leads the firm to invest sooner will improve overall welfare.

The regulated firm will delay investment longer (compared to the unregulated case) when the cap-return is applied to the network's replacement cost. However, for some levels of the cap-return, applying it to the historical cost of the network can lead to earlier investment than the unregulated case. This is possible if the risk-adjusted growth rate in replacement cost is less than zero, and the (positive) correlation between this cost and economic surplus is not too strong. Otherwise, historical cost-based regulation also leads to later investment.

A possible example where historical cost-based regulation could raise welfare is a telecommunications network, where technology shocks might move the network's replacement cost and total surplus in opposite directions (so that $\rho_{k\theta}$ is actually negative), and the cost of replacing the network is expected to fall rapidly over time. In contrast, the replacement cost of a network, such as gas, for which innovation is less prominent might actually increase over time and be positively correlated with economic surplus. The best way to regulate a gas network would then be to use its replacement cost. In many industries, only one of the conditions (drift in cost, or correlation) might be met, in which case detailed numerical analysis would be required to determine which form of regulation is appropriate.

In those situations where historical cost-based regulation can raise welfare relative to the unregulated state, such an improvement is only possible for some values of the cap-return. Furthermore, the welfare-maximizing value of the cap-return is sensitive to the properties of the network's replacement cost and the flow of surplus. Our numerical analysis makes it clear that welfare can drop dramatically if the cap-return is set even slightly too low. It is generally better to set the cap-return too high than too low. This is a consequence of the feature that delayed investment delays the market for consumers and producers and missing markets carry very high welfare costs (Goolsbee, 2000).

There are potential extensions to our work. One is to examine the situation where the regulated firm is forced to invest by the regulator. The analysis of this case will rest on what options are left to the firm and the specific characteristics of the regulatory regime.

Another possibility is to evaluate the role which geometric Brownian motion plays in our results. With historical cost regulation, if the cost of the network falls while the firm

[20]Our analysis has to be modified if entry to the industry is possible. In this circumstance the unregulated firm may not invest too late from society's point of view. The effect of, even problematic, entry has the effect of a profit cap and renders a different objective for any regulation.

is waiting to invest, the profit cap is permanently reduced. Our adoption of geometric Brownian motion means that a similar outcome results with replacement cost regulation — since shocks to cost are (stochastically) permanent, the profit cap is permanently lowered under replacement cost regulation as well. If cost was mean-reverting, the profit cap effects of cost shocks would be greater under historical cost than replacement cost regulation. This might imply that a historical cost cap is more effective at promoting investment than a replacement cost cap. We have not formally examined this issue.

Finally, we note that the profit caps differ between the two forms of regulation, partly reflecting the absence of systematic risk for historical cost regulation — risk that remains with replacement cost regulation. Both cap-returns are independent of time. While this may reflect the form of depreciation assumed, it does mean that under historical cost regulation the cap-return is applied to the full historical cost, and not the depreciated cost. This issue deserves further attention.

References

Baumol, William J., 2000. *The Free-Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*, Princeton University Press, Princeton.

Bernanke, B., 1983. "Irreversibility, uncertainty, and cyclical investment." *Quarterly Journal of Economics* 98, 85–106.

Cochrane, J., 2001. *Asset Pricing.* Princeton University Press, Princeton.

Cukierman, A., 1980. "The effects of uncertainty on investment under risk neutrality with endogenous information." *Journal of Political Economy* 88, 462–475.

Dixit, A., Pindyck, R., 1994. *Investment Under Uncertainty.* Princeton University Press, Princeton.

Economides, N., 1999. "Real options and the costs of the local telecommunications network." In: Alleman, J., Noam, E. (Eds.), *The New Investment Theory of Real Options and its Implications for the Cost Models in Telecommunications*, Kluwer Academic Publishers, Boston, MA.

Goolsbee, A., 2000. "The value of broadband and the deadweight loss of taxing new technology." Mimeo., University of Chicago.

Guthrie, G., Small, J., Wright, J., 2000. "Pricing access: forward versus backward looking cost rules." Mimeo., Department of Economics, University of Auckland.

Hausman, J., 1999. "The effect of sunk costs in telecommunications regulation," In: Alleman, J., Noam, E. (Eds). *The New Investment Theory of Real Options and its Implications for Telecommunications Economics*, Kluwer Academic Publishers, Boston, MA.

Margrabe, W., 1978. "The value of an option to exchange one asset for another." *Journal of Finance* 33, 177–186.

Poterba, J., Summers, L., 1995. "A CEO survey of U.S. companies' time horizons and hurdle rates." *Sloan Management Review*, Fall, 43–53.

Sidak, J., Spulber, D., 1997. *De-regulatory Takings and the Regulatory Contract: The Competitive Transformation of Network Industries in the United States.* Cambridge University Press, New York.

Spulber, D., 1989. *Regulation and Markets.* MIT Press, Cambridge MA.

Proofs

The following technical lemma will be useful.

Lemma 1 *Suppose that x_t evolves according to the geometric Brownian motion*

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

where μ and σ are constants. Then

$$\begin{aligned} E_T[\min\{x_{T+t}, A\}] &= x_T e^{\mu t} N\left(\frac{\log A - (\log x_T + \mu t) - \frac{1}{2}\sigma^2 t}{\sigma t^{1/2}}\right) \\ &\quad + AN\left(\frac{(\log x_T + \mu t) - \log A - \frac{1}{2}\sigma^2 t}{\sigma t^{1/2}}\right), \end{aligned}$$

where A is a constant and N is the cumulative distribution function for the standard normal distribution. ∎

Proof of Proposition 1

Let $F(k, \theta)$ denote the value of the flow of surplus if the network is not currently in place, and let $G(k, \theta)$ denote its value if the network is currently operating.

Whenever $\theta < y^* k$, F must satisfy

$$F(k, \theta) = e^{-r\,dt} E[F(k + dk, \theta + d\theta)].$$

Along the investment boundary (that is, whenever $\theta = y^* k$) it must satisfy

$$F(k, y^* k) = G(k, y^* k) - k.$$

We look for a solution of the form $F(k,\theta) = kU(y)$ where $y = \theta/k$ and find that U must satisfy the ordinary differential equation

$$0 = \frac{1}{2}\sigma^2 y^2 U''(y) + \mu y U'(y) - (r - \mu_k + \lambda_k)U(y).$$

We impose the extra condition $U(0) = 0$ so that the firm has zero value when the flow of surplus (and hence the flow of profit) is zero. The solution to this system is $U(y) = A_1 y^{\beta_1}$ for some constant A_1.

Now we turn to the situation where the network is in operation. Whenever $\theta < y^* k$, G must satisfy

$$G(k,\theta) = \theta dt + (1 - \phi dt)e^{-r\,dt}E[G(k + dk, \theta + d\theta)] + \phi dt e^{-r\,dt}E[F(k + dk, \theta + d\theta)].$$

We look for a solution of the form $G(k,\theta) = kV(y)$ where $y = \theta/k$ and find that V must satisfy the ordinary differential equation

$$0 = \frac{1}{2}\sigma^2 y^2 V''(y) + \mu y V'(y) - (r + \phi - \mu_k + \lambda_k)V(y) + y + \phi U(y).$$

If we impose the additional condition $V(0) = 0$, we find that the solution to this system is

$$V(y) = U(y) + \frac{y}{r + \lambda_\theta + \phi - \mu_\theta} + B_1 y^{\beta_3}$$

where B_1 is an arbitrary constant.

Returning to the investment boundary, recall that we require $F(k, y^*k) = G(k, y^*k) - k$ for all k. When written in terms of the functions U and V, this becomes

$$U(y^*) = V(y^*) - 1.$$

Upon substituting in our solutions for U and V this becomes

$$0 = \frac{y^*}{r + \lambda_\theta + \phi - \mu_\theta} + B_1 (y^*)^{\beta_3} - 1.$$

Therefore

$$B_1 = \left(1 - \frac{y^*}{r + \lambda_\theta + \phi - \mu_\theta}\right)(y^*)^{-\beta_3}. \tag{A-1}$$

We still need to consider the solution for $G(k,\theta)$ in the region where $\theta > y^* k$. Here it must satisfy

$$G(k,\theta) = \theta dt + (1 - \phi dt)e^{-r\,dt}E[G(k + dk, \theta + d\theta)] + \phi dt e^{-r\,dt}(E[G(k + dk, \theta + d\theta) - (k + dk)]).$$

We look for a solution of the form $G(k,\theta) = kV(y)$ where $y = \theta/k$ and find that V must satisfy the ordinary differential equation

$$0 = \frac{1}{2}\sigma^2 y^2 V''(y) + \mu y V'(y) - (r - \mu_k + \lambda_k)V(y) + y - \phi.$$

The solution to this system is

$$V(y) = \frac{y}{r+\lambda_\theta-\mu_\theta} - \frac{\phi}{r+\lambda_k-\mu_k} + B_2 y^{\beta_2}$$

where B_2 is an arbitrary constant. The function G must be continuous and differentiable everywhere. It follows that V must be continuous and differentiable at $y = y^*$. Continuity requires that $V(y^*-) = V(y^*+)$, so that

$$A_1(y^*)^{\beta_1} + \frac{y^*}{r+\lambda_\theta+\phi-\mu_\theta} + B_1(y^*)^{\beta_3} = \frac{y^*}{r+\lambda_\theta-\mu_\theta} - \frac{\phi}{r+\lambda_k-\mu_k} + B_2(y^*)^{\beta_2},$$

and differentiability requires that $V'(y^*-) = V'(y^*+)$, so that

$$\beta_1 A_1(y^*)^{\beta_1-1} + \frac{1}{r+\lambda_\theta+\phi-\mu_\theta} + \beta_3 B_1(y^*)^{\beta_3-1} = \frac{1}{r+\lambda_\theta-\mu_\theta} + \beta_2 B_2(y^*)^{\beta_2-1}.$$

Eliminating B_2 between these two equations, and using equation (A-1), allows us to show that

$$\begin{aligned} U(y^*) = A_1(y^*)^{\beta_1} &= \left(\frac{1-\beta_2}{\beta_1-\beta_2}\right)\frac{y^*}{r+\lambda_\theta-\mu_\theta} + \left(\frac{\beta_3-1}{\beta_1-\beta_2}\right)\frac{y^*}{r+\lambda_\theta+\phi-\mu_\theta} \\ &\quad + \left(\frac{\beta_2}{\beta_1-\beta_2}\right)\frac{\phi}{r+\lambda_k-\mu_k} - \left(\frac{\beta_3-\beta_2}{\beta_1-\beta_2}\right). \end{aligned}$$

The value of the surplus at the time the firm invests is

$$F(k, y^*k) = kU(y^*).$$

Proof of Proposition 2

Let $F(k,\theta)$ denote the value of the firm if the network is not currently in place, and let $G(k,\theta)$ denote its value if the network is currently operating.

Whenever $\theta < y^*k$, F must satisfy

$$F(k,\theta) = e^{-r\,dt}E[F(k+dk, \theta+d\theta)].$$

Along the investment boundary (that is, whenever $\theta = y^*k$) it must satisfy

$$F(k, y^*k) = G(k, y^*k) - k.$$

We look for a solution of the form $F(k,\theta) = kU(y)$ where $y = \theta/k$ and find that U must satisfy the ordinary differential equation

$$0 = \frac{1}{2}\sigma^2 y^2 U''(y) + \mu y U'(y) - (r - \mu_k + \lambda_k)U(y).$$

We impose the extra condition $U(0) = 0$ so that the firm has zero value when the flow of surplus (and hence the flow of profit) is zero. The solution to this system is $U(y) = A_1 y^{\beta_1}$ for some constant A_1.

Now we turn to the situation where the network is in operation. Whenever $\theta < y^*k$, G must satisfy

$$G(k,\theta) = \min\{\gamma\theta, \rho_{rc}k\}dt + (1-\phi dt)e^{-r\,dt}E[G(k+dk,\theta+d\theta)] + \phi dt e^{-r\,dt}E[F(k+dk,\theta+d\theta)].$$

We look for a solution of the form $G(k,\theta) = kV(y)$ where $y = \theta/k$ and find that V must satisfy the ordinary differential equation

$$0 = \frac{1}{2}\sigma^2 y^2 V''(y) + \mu y V'(y) - (r + \phi - \mu_k + \lambda_k)V(y) + \min\{\gamma y, \rho_{rc}\} + \phi U(y).$$

If we impose the additional condition $V(0) = 0$, we find that the solution to this system is

$$V(y) = U(y) + g_{rc}(y) + B_1 y^{\beta_3}$$

where

$$g_{rc}(y) = E\left[\int_0^\infty e^{-(r+\lambda_k+\phi-\mu_k)t}\min\{\gamma y_t, \rho_{rc}\}dt \middle| y_0 = y,\ dy_t = \mu y_t dt + \sigma y_t d\eta_t\right]$$

and B_1 is an arbitrary constant.

Returning to the investment boundary, recall that we require $F(k, y^*k) = G(k, y^*k) - k$ for all k. When written in terms of the functions U and V, this becomes

$$U(y^*) = V(y^*) - 1.$$

In addition, we impose the smooth-pasting condition

$$U'(y^*) = V'(y^*).$$

Upon substituting in our solutions for U and V these become

$$0 = g_{rc}(y^*) + B_1(y^*)^{\beta_3} - 1, \quad 0 = g'_{rc}(y^*) + \beta_3 B_1 (y^*)^{\beta_3 - 1}. \tag{A-2}$$

Eliminating B_1 gives us

$$\beta_3 = \frac{y^* g'_{rc}(y^*)}{g_{rc}(y^*) - 1}$$

which defines the firm's investment threshold.

We still need to consider the solution for $G(k,\theta)$ in the region where $\theta > y^*k$. Here it must satisfy

$$\begin{aligned} G(k,\theta) &= \min\{\gamma\theta, \rho_{rc}k\}dt + (1-\phi dt)e^{-r\,dt}E[G(k+dk,\theta+d\theta)] \\ &\quad + \phi dt e^{-r\,dt}(E[G(k+dk,\theta+d\theta) - (k+dk)]). \end{aligned}$$

We look for a solution of the form $G(k,\theta) = kV(y)$ where $y = \theta/k$ and find that V must satisfy the ordinary differential equation

$$0 = \frac{1}{2}\sigma^2 y^2 V''(y) + \mu y V'(y) - (r - \mu_k + \lambda_k)V(y) + \min\{\gamma y, \rho_{rc}\} - \phi.$$

The solution to this system is

$$V(y) = h_{rc}(y) - \frac{\phi}{r + \lambda_k - \mu_k} + B_2 y^{\beta_2}$$

where

$$h_{rc}(y) = E\left[\int_0^\infty e^{-(r+\lambda_k-\mu_k)t} \min\{\gamma y_t, \rho_{rc}\} dt \middle| y_0 = y,\ dy_t = \mu y_t dt + \sigma y_t d\eta_t\right]$$

and B_2 is an arbitrary constant.

The function G must be continuous and differentiable everywhere. It follows that V must be continuous and differentiable at $y = y^*$. Continuity requires that $V(y^*-) = V(y^*+)$, so that

$$A_1(y^*)^{\beta_1} + g_{rc}(y^*) + B_1(y^*)^{\beta_3} = h_{rc}(y^*) - \frac{\phi}{r + \lambda_k - \mu_k} + B_2(y^*)^{\beta_2},$$

and differentiability requires that $V'(y^*-) = V'(y^*+)$, so that

$$\beta_1 A_1(y^*)^{\beta_1 - 1} + g'_{rc}(y^*) + \beta_3 B_1(y^*)^{\beta_3 - 1} = h'_{rc}(y^*) + \beta_2 B_2(y^*)^{\beta_2 - 1}.$$

Eliminating B_2 between these two equations, and using equations (A-2), allows us to show that

$$U(y^*) = A_1(y^*)^{\beta_1} = \frac{y^* h'_{rc}(y^*) - \beta_2\left(h_{rc}(y^*) - 1 - \frac{\phi}{r + \lambda_k - \mu_k}\right)}{\beta_1 - \beta_2}.$$

The value of the firm at the time it invests is

$$F(k, y^* k) = kU(y^*).$$

Proof of Proposition 3

Let $F(k, \theta)$ denote the value of the firm if the network is not currently in place, and let $G(k, \theta; \bar{k})$ denote its value if the network is currently operating, where $\bar{k}$ is the historical cost of building the network.

Whenever $\theta < y^* k$, F must satisfy

$$F(k, \theta) = e^{-r\,dt} E[F(k + dk, \theta + d\theta)].$$

Along the investment boundary (that is, whenever $\theta = y^* k$) it must satisfy

$$F(k, y^* k) = G(k, y^* k; k) - k.$$

We look for a solution of the form $F(k, \theta) = kU(y)$ where $y = \theta / k$ and find that U must satisfy the ordinary differential equation

$$0 = \frac{1}{2}\sigma^2 y^2 U''(y) + \mu y U'(y) - (r - \mu_k + \lambda_k) U(y).$$

We impose the extra condition $U(0) = 0$ so that the firm has zero value when the flow of surplus (and hence the flow of profit) is zero. The solution to this system is $U(y) = A_1 y^{\beta_1}$ for some constant A_1.

Now we turn to the situation where the network is in operation. Whenever $\theta < y^* k$, G must satisfy

$$G(k, \theta; \bar{k}) = \min\{\gamma\theta, \rho_{hc}\bar{k}\}dt + (1-\phi dt)e^{-r\,dt}E[G(k+dk, \theta+d\theta; \bar{k})] + \phi dt e^{-r\,dt}E[F(k+dk, \theta+d\theta)].$$

We look for a solution of the form

$$G(k, \theta; \bar{k}) = kV(y) + \bar{k}W(z)$$

where $y = \theta/k$ and $z = \theta/\bar{k}$. We find that V and W must satisfy the ordinary differential equations

$$0 = \frac{1}{2}\sigma^2 y^2 V''(y) + \mu y V'(y) - (r + \phi - \mu_k + \lambda_k)V(y) + \phi U(y)$$

and

$$0 = \frac{1}{2}\sigma_\theta^2 z^2 W''(z) + (\mu_\theta - \lambda_\theta) z W'(z) - (r + \phi)W(z) + \min\{\gamma z, \rho_{hc}\}.$$

Solutions to these equations are

$$V(y) = U(y) + B_1 y^{\beta_3}$$

and

$$W(z) = g_{hc}(z),$$

where

$$g_{hc}(z) = E\left[\int_0^\infty e^{-(r+\phi)t} \min\{\gamma z_t, \rho_{hc}\}dt \middle| z_0 = z,\ dz_t = (\mu_\theta - \lambda_\theta)z_t dt + \sigma_\theta z_t d\eta_t\right]$$

and B_1 is an arbitrary constant.

Recall that we require $F(k, y^* k) = G(k, y^* k; k) - k$ for all k. When written in terms of the functions U, V and W, this becomes

$$U(y^*) = V(y^*) + W(y^*) - 1.$$

In addition, we impose the smooth-pasting condition

$$U'(y^*) = V'(y^*) + W'(y^*).$$

Upon substituting in our solutions for U, V and W these become

$$0 = g_{hc}(y^*) + B_1(y^*)^{\beta_3} - 1, \quad 0 = g'_{hc}(y^*) + \beta_3 B_1 (y^*)^{\beta_3 - 1}. \qquad \text{(A-3)}$$

Eliminating B_1 gives us

$$\beta_3 = \frac{y^* g'_{hc}(y^*)}{g_{hc}(y^*) - 1}$$

which defines the firm's investment threshold.

We still need to consider the solution for $G(k,\theta;\bar{k})$ in the region where $\theta > y^*k$. Here it must satisfy[21]

$$\begin{aligned} G(k,\theta;\bar{k}) &= \min\{\gamma\theta, \rho_{hc}\bar{k}\}dt + (1-\phi dt)e^{-r\,dt}E[G(k+dk,\theta+d\theta;\bar{k})] \\ &\quad + \phi dte^{-r\,dt}(E[G(k+dk,\theta+d\theta;k+dk)-(k+dk)]). \end{aligned}$$

We look for a solution of the form

$$G(k,\theta;\bar{k}) = kV(y) + \bar{k}W(z)$$

where $y=\theta/k$ and $z=\theta/\bar{k}$. We find that V and W must satisfy the ordinary differential equations

$$0 = \frac{1}{2}\sigma^2 y^2 V''(y) + \mu y V'(y) - (r-\mu_k+\lambda_k)V(y) + \phi W(y) - \phi.$$

and

$$0 = \frac{1}{2}\sigma_\theta^2 z^2 W''(z) + (\mu_\theta-\lambda_\theta)zW'(z) - (r+\phi)W(z) + \min\{\gamma z, \rho_{hc}\}.$$

Solutions to these equations are

$$V(y) = h_{hc}(y) - \frac{\phi}{r+\lambda_k-\mu_k} + B_2 y^{\beta_2}$$

and

$$W(z) = g_{hc}(z),$$

where

$$h_{hc}(y) = E\left[\int_0^\infty e^{-(r+\lambda_k-\mu_k)t}\phi g_{hc}(y_t)dt\,\middle|\, y_0=y,\ dy_t = \mu y_t dt + \sigma y_t d\eta_t\right]$$

and B_2 is an arbitrary constant.

The function G must be continuous and differentiable everywhere. It follows that V must be continuous and differentiable at $y=y^*$. Continuity requires that $V(y^*-) = V(y^*+)$, so that

$$A_1(y^*)^{\beta_1} + B_1(y^*)^{\beta_3} = h_{hc}(y^*) - \frac{\phi}{r+\lambda_k-\mu_k} + B_2(y^*)^{\beta_2},$$

and differentiability requires that $V'(y^*-) = V'(y^*+)$, so that

$$\beta_1 A_1(y^*)^{\beta_1-1} + \beta_3 B_1(y^*)^{\beta_3-1} = h'_{hc}(y^*) + \beta_2 B_2(y^*)^{\beta_2-1}.$$

Eliminating B_2 between these two equations, and using equations (A-3), allows us to show that

$$U(y^*) = A_1(y^*)^{\beta_1} = \frac{y^*h'_{hc}(y^*) - \beta_2\left(h_{hc}(y^*) - \frac{\phi}{r+\lambda_k-\mu_k}\right) + (\beta_3-\beta_2)(g_{hc}(y^*)-1)}{\beta_1-\beta_2}.$$

The value of the firm at the time it invests is

$$F(k, y^*k) = kU(y^*).$$

[21]Notice that the profit cap changes in the event that the network needs replacing. The cap is currently $\rho_{hc}\bar{k}$; if the network dies in the next instant, the new cap will be $\rho_{hc}(k+dk)$.