Dynamics of price regulation[¤]

Department of Economics Working Paper 99-07

Gary Biglaiser
University of North Carolina

Michael Riordan
Columbia University

August 1999

Abstract

We study the dynamics of price regulation for an industry adjusting to exogenous technological progress that lowers capital equipment and operating costs. First, we characterize the optimal capacity path and capacity replacement cycles in a neoclassical investment model with a ¯xed rate of technological progress. The corresponding optimal price path recovers operating costs, the user cost of capital, and a contribution to ¯xed costs based on Ramsey pricing principles. The user cost of capital includes economic depreciation re°ecting the declining replacement value of capacity and foregone operating cost improvements over the economic life of the investment. Second, we show that naive rate of return regulation, which ignores these components of economic depreciation, eventually results in a de¯cient level of capacity due to excessively high retail prices burdened by the need to recover the underdepreciated costs of historical investments. Naive rate of return regulation also lacks incentives for replacing economically obsolete capacity if accounting rates of depreciation are set too low. Third, we explain, on the one hand, how price cap regulation can incent more e±cient capital replacement decisions, and discuss the application of Ramsey principles for the recovery of stranded costs in a transition from rate of return regulation. On the other hand, ine±cient replacement of capital may arise depending on the length of the price-cap period. Finally, we use our results to interpret recent regulatory reforms in telecommunications markets.

1 Introduction

The price regulation of natural monopolies in the United States and elsewhere has gone through various phases of rate-of-return regulation, price-cap regula-

[¤]Early versions of this paper were presented at the ESSET Conference in Gerzensee, Switzerland in July 1998, at the TPRC conference in Washington D.C. in September 1998, at the Midwest Economic Theory Meetings in Ann Arbor in October 1998, at the Tel Aviv University Mini-Conference on Antitrust and Regulation in May 1999, and at seminars at BU, Columbia, LBS, LSE, MIT, NYU, Oxford, Toulouse, UCLA, and Yale. The authors thank numerous particpants for many helpful comments.

tion, and now, especially in telecommunications, new forms of wholesale price regulation.[1] These regulatory changes have occurred against a backdrop of rapid technological change. In telecommunications, capital equipment costs and operating costs have declined steadily with the development of microwave, fiber optic, and digital switching technologies, and are poised to drop further with packet-switching and internet-protocol technologies. Similar regulatory changes have occurred in electricity where costs have dropped with the development of new technologies for generation, including gas turbine technology.

We study price regulation in a model of industry structure emphasizing the dynamic paths of prices and investments induced by technological progress. The most important element of our model is that exogenous technological progress lowers both capital equipment and operating costs. Consequently, the optimal path of regulated prices reflects the value of waiting for further technological improvements. In particular, the optimal price at each date recovers operating costs, the user cost of capital, and a contribution to fixed costs. The user cost of capital includes economic depreciation resulting from the declining price of new capacity plus the value of operating cost improvements that are foregone by investing sooner rather than later. Non-recurring fixed costs are recovered over time in a manner that reflects Ramsey pricing principles for output sold at different dates. Investments in new capacity adjust to meet demand at these prices, and capacity becomes economically obsolete when the long-run marginal cost of replacement capacity drops below the short-run operating cost of the old capacity.

Rate-of-return regulation has been used to regulate public utilities for much of the century. Rate-of-return regulation is designed to recover historical investment costs. It is not designed to adjust prices to the changing long run marginal cost of a dynamic market structure. In particular, the regulated price recovers an allowed rate of return and an allowed depreciation of a rate base equaling cumulative undepreciated investments. Naive rate-of-return regulation, which under-depreciates assets by ignoring technological progress, distorts the path of prices relative to the optimal path. Initially the regulated prices are too low because these prices fail to recover economic depreciation. However, regulated prices eventually become too high due to the firm's entitlement to the allowed rate of return on a bloated rate base. Correspondingly, naive rate-of-return regulation eventually results in an economically deficient level of capacity. Moreover, naive rate-of-return regulation lacks incentives to retire old capacity and replace it with new capacity with lower operating costs.

Price-cap regulation provides better incentives for capital replacement, but still may still leave prices too high if the initial price cap is set to recover the

[1] The Federal Communications Commission adopted price cap regulation for AT&T in 1989, although retail long distance prices have since been deregulated. Approximately 30 states had adopted some form of price cap regulation as of 1996 (see Sappington and Weisman (1996) and Tardiff and Taylor (1996)). The FCC currently regulates interstate access fees for long distance service with price-caps. The regulation of long distance access fees is a form of wholesale price regulation. Wholesale price regulation is also being applied to unbundled elements of local telephone networks under the Telecommunications Act of 1996.

historical costs inherited from rate-of-return regulation. Moreover, there is a sense in which the regulator only gets one bite from the apple. Even in the case of deterministic technological progress, if the parameters of regulation are not set optimally initially, then a problem of stranded costs may prevent the subsequent achievement of full e±ciency by more enlightened regulation. Stranded costs arise when a change in regulatory structure prevents recovery of a fair rate of return on previous investments. A stranded-costs problem can be expected in a transition from a regime of naive rate-of-return regulation when the allowed depreciation rate has been set below economic depreciation.[2] We argue that with linear pricing it is optimal to recover stranded costs by setting a markup above the optimal price level according to Ramsey pricing principles. If the demand elasticity is constant over time, then the optimal markup is proportional to the declining retail price and a constant sales tax can implement the second-best solution.

In practice, price-cap regulation has been implemented only over ¯xed time horizons. Thus, regulators' commitments to replace rate-of-return regulation with price-cap regulation have been only temporary. We explain how a temporary price-cap regime provides the regulated ¯rm de¯cient incentives for capacity replacement. If the price-cap horizon is excessively short, then price-cap regulation provides no better incentives for capacity replacement than does rate-of-return regulation. For longer price-cap horizons, capacity replacements are concentrated at the start of the price-cap period, and the ¯rm has no incentive to replace capacity near the end of the price-cap period. Thus, repeated intervals of price-cap regulation generate cycles in capacity replacement.

There has been a great deal of research dealing with the properties of rate-of-return and price-cap regulation. For the most part, the research has dealt with static models. One set of models deals with the \Averch-Johnson e®ect"; see e.g. Averch and Johnson (1962) and Joskow (1972) where rate-of-return regulation distorts a ¯rms incentive to minimize costs because it receives a supracompetitive return on capital. A second line of research deals with the incentives for a ¯rm to produce in a market where it faces competition when it is regulated in another market; see e.g. see Braeutigam and Panzar (1989), and a third deals with the case where ¯rms have private information about costs or demand; see e.g. Cabral and Riordan (1989) and Lewis and Sappington (1989). There has been some work done on dynamic models. Bradley and Price (1988), Vogelsang (1988), and Brennan (1991) examine complete information models, while Sibley (1989) investigates a model where the ¯rm has private information about costs and demand. These papers ignore technological progress.

There are many dynamic regulatory models that examine investment issues by a regulated ¯rm in a political economy framework. In a symmetric information model, where regulators are short lived and only care about consumers during their term while the ¯rm is long lived, Lewis and Sappington (1991) show that on average investment is lower than the ¯rst best level. Besanko

[2]Sidak and Spulber (1996) argue forcefully that the ¯rm is legally entitled to recover stranded costs that result from a change of regulatory regime.

and Spulber (1992) and Laffont and Tirole (1992) examine how a regulator's commitment ability affects a firm's investment decisions when the firm has private information about its productivity. Gilbert and Newbery (1994) show in an infinite time horizon model under uncertainty, that rate of return regulation using the "used and useful criteria" is able to support efficient investment under a wider range of parameters than under standard rate of return regulation. Other papers with regulatory review include Lyon (1991) and Salant and Woroch (1992). Again, there is no technological progress in any of these models.

In section 2, we study optimal prices, investment, and capital replacement for the case of deterministic, exogenous technological progress and constant demand. We briefly discuss the implications of uncertain technological progress and growing and uncertain demand, and provide more detailed results in appendices. In sections 3 and 4, we examine the relationship between the optimal solution and rate-of-return and price-cap regulation, respectively. We discuss our results in relation to the telecommunications industry in section 5.

2 Optimal prices

Our model assumes the following market structure for a particular service. Demand is stable and described by a downward-sloping inverse demand curve $p = \phi(X)$. The corresponding revenue function is $R(X) = \phi(X)X$. For simplicity, we assume that the demand curve has a constant elasticity. Thus, marginal revenue is $R'(X) = [1 - \frac{1}{\epsilon}]\phi(X)$, where $\epsilon > 0$ is the price elasticity of demand. Consumers' willingness to pay for X units of output is equal to the area under the demand curve,

$$\Phi(X) = \int_0^X \phi(x)dx.$$

Our model allows several components of costs: capacity costs, operating costs, and fixed costs. The first two components decline with technological progress which is embodied in new capital equipment. In the short run, output is constrained by capacity. Thus, X units of output require at least X units of capacity. Capacity is a capital asset that depreciates at a physical rate $\delta \geq 0$.[3] The investment cost of a unit capacity is q, and declines at a deterministic rate $\mu \geq 0$. The operating costs for a unit of capacity of a given vintage is constant over the life of the asset. However, the operating cost of new (state of the art) capacity declines at rate μ along with its acquisition price. Thus, operating costs are proportional to the acquisition cost for each vintage of capacity, i.e. $c = \gamma q$; with $\gamma \geq 0$. There is a non-recurring fixed cost of operation, $F \geq 0$, which is paid up-front.

[3] This is akin to what Jorgenson (1963) calls the replacement ratio. It is the rate at which capacity of each vintage physically wears out. However, as we shall see, in our framework, capacity is also replaced due to economic obsolescence.

The optimal investment path maximizes the present discounted value of social welfare subject to a break-even constraint. The discount rate is equal to r. Social welfare depends on the path of investment in new capacity and the utilization of capacity over time. Investment in new capacity at date t is equal to $x(t)$, and the utilization rate of vintage τ capacity at date $t \geq \tau$ is $u(t;\tau)$. Obviously, $u(t;\tau)$ must lie between zero and one. If $e^{-\delta(t-\tau)}x(\tau)$ is the amount of vintage τ capacity available at date t, then the total amount of capacity utilized at date t, and hence total output is

$$X(t) = \int_0^t u(t;\tau)e^{-\delta(t-\tau)}x(\tau)d\tau + u(t;0)e^{-\delta t}X(0) \qquad (1)$$

and the market clearing price at date t is

$$p(t) = \phi(X(t))$$

Similarly, the firm's total operating cost at date t is

$$C(t) = \int_0^t u(t;\tau)e^{-\delta(t-\tau)}x(\tau)c(\tau)d\tau + u(t;0)e^{-\delta t}X(0)c(0) \qquad (2)$$

Thus, social welfare is

$$\int_0^\infty e^{-rt}[\Phi(X(t)) - C(t) - x(t)q(t)]dt - X(0)q(0) \qquad (3)$$

This social objective function is assumed to be well-defined and bounded in the relevant ranges, which will be the case if r is sufficiently large. The break-even constraint requires that the present discounted value of the stream of revenues resulting from investment and utilization paths covers the present discounted value of all costs. This constraint is

$$\int_0^\infty e^{-rt}\{R(X(t)) - C(t) - x(t)q(t)\}dt - q(0)X(0) \geq F \qquad (4)$$

Optimal investment and utilization paths maximize (3) subject to (4).[4]

The solution to the social planning problem can be derived by invoking the Kuhn-Tucker theorem to define a Lagrangian, by substituting the definition of output (1) and cost (2) and point-wise maximizing it with respect to $\{X(0); x(t); u(t;\tau)\}$. Let $\lambda \geq 0$ denote the Lagrangian multiplier for the break-even constraint (4). The following theorem characterizes the price path corresponding to the optimal investment and capital retirement plan, and characterizes the optimal economic life of capacity (Δ). It is proved in Appendix A.

[4] If demand is inelastic, i.e. $\varepsilon < 0$, then a policy satisfying the break-even constraint necessarily exists, as revenue can be made arbitrarily large by shrinking output. If demand is elastic, then it is assumed that F is not so large as to preclude a solution.

Theorem 1 The optimal price path satisfies

$$[1 - \frac{\lambda}{(1+\lambda)^2}]p(t) = [\gamma + r + \delta + \mu + \mu\gamma \int_0^{\Delta} e^{-(r+\delta)\tau} d\tau]q(t) \qquad (5)$$

where $\lambda \geq 0$ is a fixed constant and Δ uniquely solves

$$e^{\mu\Delta}\gamma = \gamma + r + \delta + \mu + \gamma\mu \int_0^{\Delta} e^{-(r+\delta)\tau} d\tau \qquad (6)$$

We interpret the theorem in steps. First, consider the base case in which there are no operating costs and no fixed costs, i.e. $\gamma = 0$ and $F = 0$. In this case, (6) does not apply. The break-even constraint is slack, i.e. $\lambda = 0$, it is never efficient to retire capacity, i.e. $\Delta(t) = \infty$, and the optimal pricing formula (20) becomes simply[5]

$$p(t) = (r + \delta + \mu)q(t) \qquad (7)$$

Along the efficient investment path, the retail price equals the user cost of capital, which equals a competitive return on investment plus economic depreciation. The rate of economic depreciation equals the rate of physical depreciation, δ, plus the rate of productivity improvement of new capacity, μ. It is easily seen that the efficient investment path can be implemented in a competitive equilibrium in which investments at each date are exactly recovered by the cash flow generated by the investments; (see, for example, Jorgenson (1967));[6]

$$\int_t^{\infty} e^{-(r+\delta)(s-t)} p(s) ds = q(t):$$

Thus, $p(t)$ satisfies the definition of forward-looking cost proposed by Salinger [24].

The next step is to consider the case in which both capital equipment and operating costs are falling at the same constant rate (μ), but there are no fixed costs ($F = 0$).[7] In this case, the output price falls at the rate of technological progress, and all units of capacity have the same economic life (Δ) defined by (6). This equation says that vintage τ capacity is optimally retired when price falls below its operating cost, i.e. $p(t) \leq c(\tau)$. Setting $\lambda = 0$; the optimal pricing formula of equation (20) becomes

$$p(t) = (\gamma + r + \delta + \mu)q(t) + \mu c(t) \int_0^{\Delta} e^{-(r+\delta)\tau} d\tau \qquad (8)$$

[5] See, for example, Nickell (1978, p. 10) for a derivation of a variation of this formula in a two-factor neoclassical investment model.

[6] From this it follows that the break-even constraint is slack in this case. In particular,

$$\int_0^{\infty} e^{-(r+\delta)t}\{p(t)dt - q(0)\}X(0) + \int_0^{\infty} e^{-rt}\{\int_t^{\infty} e^{-(r+\delta)(s-t)}p(s)ds - q(t)x(t)\}dt =$$
$$\int_0^{\infty} e^{-rt}\{p(t)X(t) - x(t)q(t)\}dt - q(0)X(0) = 0$$

[7] This is essentially the same problem solved by Malcomson (1975) who used optimal control techniques.

Equation (8) differs from equation (7), when operating costs were constant, by the term $\mu c(t)\int_0^{\Delta} e^{-(r+\delta)\tau}d\tau$. This term equals the presented discounted value of operating improvements that would be obtained by delaying investment slightly. This is properly considered an opportunity cost of current investment, and therefore a component of the user-cost of capital. As in the previous case, the optimal price path can be implemented in a competitive market, i.e.

$$\int_t^{t+\Delta} e^{-(r+\delta)(s-t)}[p(s)-c(t)]ds = q(t)$$

and again satisfies Salinger's (1998) definition of forward-looking cost.

This solution has a very natural interpretation in terms of Bellman's Principle of Optimality. Given the economic life of assets Δ, define the "state space" to be the stock of each vintage of unretired capacity, i.e. the state is $\{y(t;\tau) : t-\Delta \le \tau \le t\}$ with

$$y(t;\tau) = e^{-\delta(t-\tau)}x(\tau)$$

and

$$X(t) = \int_{t-\Delta}^{t} y(t;s)ds$$

The value function

$$\begin{aligned} V &= \int_t^{\infty} e^{-r(\tau-t)}[\Phi(X(\tau))-p(\tau)X(\tau)]d\tau + \\ &\int_{t-\Delta}^{t}\int_t^{\Delta+s} e^{-(r+\delta)(\tau-t)}[p(\tau)-c(s)]y(s;t)d\tau \end{aligned}$$

satisfies the Bellman equation[8],

$$rV = \max_x \{\Phi(X)-cX-qx+\int_{t-\Delta}^{t}\frac{\partial V}{\partial y}\frac{\partial y}{\partial t}ds+\frac{\partial V}{\partial t}\} \tag{9}$$

if $p(\tau)$ satisfies price path (5) defined by the theorem. The Bellman equation (9) says that the rate of return on the value function (rV) is equal to the flow of social surplus plus the rate of change of value resulting from depreciation of the assets ($\int_{t-\Delta}^{t}\frac{\partial V}{\partial y}\frac{\partial y}{\partial t}ds$) and exogenous technological progress ($\frac{\partial V}{\partial t}$). This value function itself has two components. The first component is the present discounted value of consumer surplus. The second component is the value of the stock of assets defining the state, i.e. the value of each unit of capacity is the present discounted value of the cash flow it generates over its remaining economic life.

[8] It is also straightforward to use Bellman's Principle to similarly characterize utilization rates, and hence the economic life of assets.

Finally, consider the general case with positive fixed costs. The optimal price path (5) can be rewritten as

$$\frac{p(t) - [\delta + r + \delta + \mu + \mu\gamma \int_0^{\Delta} e^{-(r+\delta)\tau} d\tau] q(t)}{p(t)} = \frac{\lambda}{(1+\lambda)^2}$$

From standard Ramsey principles, the optimal price-cost margin when the firm must recover fixed costs is inversely related to the elasticity of demand.[9][10] In calculating the Ramsey formula, the appropriate measure of long run marginal cost is operating cost plus the user-cost of capital, where the latter includes a term for foregone operating improvements as a component of economic depreciation. The optimal markup can be interpreted as a sales tax equal to $\psi = \frac{\lambda}{(1+\lambda)^2}$. The value of ψ is determined implicitly by substituting (5) into the binding constraint (4).[11] This outcome cannot be implemented competitively because of the scale economies arising from fixed costs.

This model of deterministic technological progress is highly tractable, but of course not realistic. Technological progress is uncertain and often quite discrete. Appendix B extends the basic model so that uncertain technological progress follows a discrete Poisson arrival process. In this case, the path of optimal prices reflects essentially the same opportunity costs as in the deterministic model, except that appropriate allowance must be made for the uncertain economic life of assets. In particular, the user cost of capital reflects the expected decline of capital equipment costs and the expected present discounted value of forgone operating improvements. To our knowledge, this extension to technological uncertainty is new to the neoclassical investment literature.

---

[9] The firm can be interpreted to be a multiproduct firm, producing different outputs at different dates. We have assumed for simplicity that the price elasticity of demand is the same at each date. If this were not the case, then the Ramsey markup would vary over time with the price elasticity of demand.

[10] A two-part tariff generally is more efficient than the Ramsey pricing rule. In particular, if the demand for the first unit of consumption by all consumers is quite inelastic, then recovery of fixed cost is best done by the fixed portion of the tariff. For various economic, distributional, and political reasons, this may not be possible to do. In that case, Ramsey pricing is the best way to recover the fixed costs.

[11] The exact condition determining ψ is that present value of receipts from the sales tax covers the fixed cost:

$$F = \psi \int_0^{\infty} e^{-rt} R(\phi(p(t)))dt$$

Since revenue declines at rate $(1-\epsilon)\mu$ along the optimal path, this condition can be rewritten as

$$F = \frac{\psi}{r + (1-\epsilon)\mu} R(\phi(p(0)))$$

where

$$p(0) = \frac{[\gamma + r + \delta + \mu + \mu\gamma \int_0^{\Delta} e^{-(r+\delta)\tau} d\tau] q(0)}{1-\psi}$$

from the proposition. The condition determining ψ has a unique solution, if $R(p)$ is increasing in p, which is true if $\epsilon \leq 1$. If $\epsilon > 1$, then a solution exists only if F is not too large.

The assumption of stable demand is easily generalized to allow for growing demand. The statement of Theorem 1 does not rely on the structure of demand, but only on the structure of costs. A growing demand, like increasing productivity, assures that nonnegativity constraints on new investments remain slack, which is what leads to the cost-based pricing rule of the theorem.[12] Uncertain demand is more problematic. If demand can shrink with some probability, which of course is not unrealistic given business cycles, then the characterization of optimal investment must take account of the option value of waiting to invest in the light of new demand information. Presumably, this requires an upward correct to the optimal price path. Appendix C draws on results from Dixit and Pyndyck (1994) to show how demand uncertainty is reflected in the optimal price path. This issue is potentially quite important, as is evident from large stranded costs in electricity markets that resulted from overly optimistic demand forecasts.

3 Rate-of-return regulation

Traditional rate-of-return regulation calculates a rate base in each period, and regulates the retail price to yield a fair return on the rate base. Operating costs and depreciation of the rate base are expensed. In this section, we consider a stylized form of continuous rate-of-return regulation that establishes an allowed rate of return on investment, $\hat{\rho}$, and an allowed rate of depreciation, $\hat{\delta}$, at each date, and a required time at which the asset must be removed from the rate base, $\bar{\Delta}$. For simplicity, we ignore the fixed cost F in this section; more generally any fixed non-recurring capital costs would be included in the rate base.

The rate base, $K(t)$, equals the undepreciated value of all previous investments. The initial rate base is $K(0) = q(0)X(0)$. The rate base subsequently decreases with depreciation and increases with new investment. Thus, the adjustment of the rate base is given by

$$\dot{K}(t) = -\hat{\delta}K(t) - e^{-\hat{\delta}\bar{\Delta}}q(t-\bar{\Delta})x(t-\bar{\Delta}) + q(t)x(t). \tag{10}$$

The first term in (10) represents the reduction in the rate base due to allowed depreciation. The second term are the assets that are taken out of the rate base that are $\bar{\Delta}$ periods old. The last term is the value of new investment. Thus, the value of the rate base at date t is

$$K(t) = \int_{t-\bar{\Delta}}^{t} e^{-\hat{\delta}(t-s)}q(s)x(s)ds \tag{11}$$

The revenue requirement of a regulated monopolist at date t is

$$R(X(t)) = C(t) + (\hat{\rho} + \hat{\delta})K(t), \tag{12}$$

[12]See Nickell (1978,pp.55-59) for a clear discussion of how a demand cycles influence the optimal (i.e. competitive) investment and price paths.

where $R(X(t)) = X\phi(X(t))$ is the revenue function. The revenue requirement covers operating costs, an allowed return on the rate base, and allowed depreciation of the rate base.[13]

Rate-of-return regulation can be interpreted as a series of loans made to consumers. The rate base is the outstanding loan balance. Each period consumers pay interest on the loan ($\hat{r}K(t)$) and repay part of the principal $\hat{\delta}K(t)$. These payments are in the form of allowed revenues above operating costs. New investments in capacity are financed by additional loans.

In theory, rate-of-return regulation could track optimal prices if the allowed rate of return is equal to the competitive rate ($\hat{r} = r$), and the allowed depreciation rate is equal to economic depreciation as characterized by Theorem 1 ($\hat{\delta} = \delta + \mu + \mu\gamma \int_0^{\Delta} e^{-(r+\delta)\tau} d\tau$), and if capacity were replaced whenever it became economically obsolete as characterized by the theorem ($\hat{\Delta} = \Delta$).[14] However, in practice, rate-of-return regulation generally has not taken proper account of technological progress in setting depreciation schedules (see Kahn (1988)).

To study the consequences of ignoring technological progress, we examine a model of naive rate-of-return regulation with $\hat{r} = r$, $\hat{\delta} = \delta$, and $\hat{\Delta} = \infty$.[15][16] Clearly, this form of rate-of-return regulation provides no incentive for the retirement of capacity. In this case, $C(t) = \gamma K(t)$ and

$$R(X(t)) = (\gamma + r + \delta)K(t), \tag{13}$$

The rate of change of the revenue requirement is

$$R'(X(t))\dot{X}(t) = (\gamma + r + \delta)\dot{K}(t). \tag{14}$$

where $R'(X(t)) \equiv \phi(X(t))[1 - \frac{1}{\rho}]$. Combining equations (10), (13) and (14), and rearranging gives

$$[R'(X(t)) - (\gamma + r + \delta)q(t)]\frac{\dot{X}(t)}{X(t)} = -\delta[\phi(X(t)) - (\gamma + r + \delta)q(t)] \tag{15}$$

[13] Strait line depreciation would have $\hat{\delta} = \frac{1}{\hat{\Delta}}$. This is the method used by the FCC and other regulatory bodies. The popular wisdom that depreciation schedules are too long amounts to saying that $\hat{\Delta}$ far exceeds the useful economic life of assets. We thank Ken Moran of the FCC for a useful discussion on depreciation.

[14] However, rate-of-return regulation does not appear to be flexible enough to mimick the efficient price path when there is technological uncertainty; (see Appendix B). One could obtain efficiency if the rate base were reduced at the actual rate of economic depreciation, while expected current economic depreciaition were treated as an expense. In this way, rate-of-return regulation with two different depreciation rates could mimick the efficient price path, but this is not how rate-of-return regulation is practiced in the United States.

[15] Schmalensee (1989) shows that a regulated firm prefers slow depreciation schedules if the allowed rate of return is above the competitive level. This preference is perhaps one reason why rate-of-return regulation may result in under-depreciation of capital assets. Another possible reason is that myopic regulators may be attracted to the temporarily lower prices that resulted from a low allowed depreciation rate.

[16] Admittedly, this is something of a caricature of modern rate-of-return regulation. There has been some recognition of economic obsolescence in setting the accounting lives of assets (Phillips,1993, p.275), although not much attention to other consequences of technological progress for economic depreciation. In any case, accounting lives have remained quite long.

Equation (15) is a first order differential equation in the quantity of output $X(t)$. The initial condition is $\phi(X(0)) = (\gamma + r + \delta)q(0)$. Therefore, $R'(X(0)) - (\gamma + r + \delta)q(0) < 0$ and $\dot{X}(0) = 0$. However, it is straightforward to show that $\ddot{X}(t) > 0$ because $\dot{q}(t) < 0$. It follows from equation (15) that, after the initial date, $X(t)$ evolves such that $\phi(X(t)) - (\gamma + r + \delta)q(t) > 0$ and $R'(X(t)) - (\gamma + r + \delta)q(t) < 0$. Thus, $\dot{X}(t) > 0$ after the initial date, i.e. price declines under naive rate-of-return regulation based on physical depreciation. This result is not surprising. The initial investment in capacity is financed by a markup of price above average operating cost. Some of the capacity depreciates and is replaced, but the replacement cost is lower because of technological progress. Therefore, the requisite markup is less.

In order to characterize the price distortions that arise under naive rate-of-return regulation, it is useful to rewrite the differential equation in terms of $Y(t) = \frac{p(t)}{q(t)}$. Since

$$\frac{\dot{Y}(t)}{Y(t)} = -\frac{1}{\epsilon}\frac{\dot{X}(t)}{X(t)} + \mu \tag{16}$$

by definition, substitution of equation (15) into equation (16) and simplification gives

$$\frac{\dot{Y}(t)}{Y(t)} = \frac{\delta(Y(t) - \gamma - r - \delta)}{(\epsilon - 1)Y(t) - \epsilon(\gamma + r + \delta)} + \mu$$

Since $Y(0) = \gamma + r + \delta$ at the initial condition, it follows that $\frac{\dot{Y}(t)}{Y(t)} > 0$ and $Y(t)$ converges to

$$\bar{Y} = \frac{\delta + \mu\epsilon}{\delta + \mu\epsilon - \mu}(\gamma + r + \delta)$$

$\overline{Y}$ is positive and greater than $[r + \delta + \gamma + \mu]$ if $\frac{\mu - \delta}{\mu} < \epsilon < \frac{\gamma + \mu + r}{\mu}$. Thus, we conclude that naive rate-of-return regulation eventually results in excessive prices and correspondingly deficient capacity under these conditions.

We summarize our results in the following theorem:

Theorem 2 Assume $\frac{\mu - \delta}{\mu} < \epsilon < \frac{\mu + \gamma + r}{\mu}$. Under naive rate of return there exists a t^* such that $p(t) < (\gamma + r + \delta + \mu)q(t)$ for $t < t^*$, $p(t) > (\gamma + r + \delta + \mu)q(t)$ for $t > t^*$, and $p(t) \to \frac{\delta + \mu\epsilon}{\delta + \mu\epsilon - \mu}(\gamma + r + \delta)q(t)$ as $t \to \infty$. Thus, investment in capacity is at first excessive, but eventually it becomes and stays deficient.

Theorem 2 rests on the parameter restriction $\frac{\mu - \delta}{\mu} < \epsilon < \frac{\gamma + \mu + r}{\mu}$. If demand elasticity is too low, then $Y(t)$ increases without bound, and the market collapses. If demand elasticity is too high then $p(t) < (\gamma + r + \delta + \mu)q(t)$ for all t. Since the firm breaks even in the long run under rate-of-return regulation, this means that consumers are better off than under the optimal investment path characterized by Theorem 1. This puzzle is aggravated by equation (13)

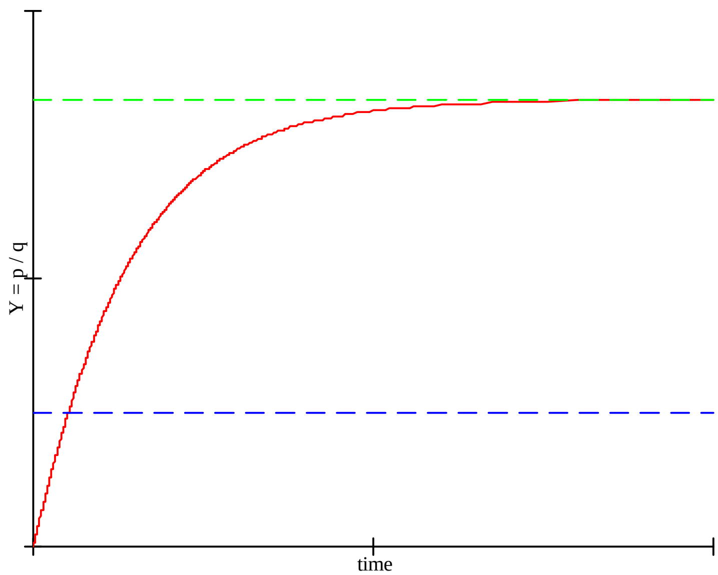

Figure 1: The path of $Y(t) = \frac{p(t)}{q(t)}$ under naive rate of return regulation

which is a break-even condition. If the price path is lower under rate-of-return regulation, then how does the firm break even? These anomalous properties stem from the fact that when $\epsilon > 1$ social surplus and revenue grow without bound as price goes to zero, since marginal operating costs go to zero. Discounting bounds the problem when demand is not too elastic, but otherwise, the problem of maximizing the present value of social surplus is not well defined for the elastic demand case unless the discount rate is sufficiently large.

The path of $Y(t)$ for illustrative parameters ($r = \delta = \rho = 0{:}05$, $\gamma = 0{:}1$, $\epsilon = 0{:}2$) is shown in Figure 1. In this diagram the lower dashed line corresponds to the constant optimal path $Y^* = (\gamma + r + \delta + \rho)$, while the upper dashed line is the limit of the price path under naive rate-of-return regulation, $\bar{Y}^\infty$.

4 Price-cap regulation

Rate-of-return regulation gradually has given way to price-cap regulation as the preferred method of regulating public utility prices. Under naive rate-of-return regulation, the incentive is for a firm is to keep old capital in its rate base until it is fully depreciated. Given excessively long depreciation schedules, this is potentially a huge inefficiency that price-cap regulation was designed to correct.

A price cap is a ceiling below which the regulated firm has price flexibility. Thus, price-cap regulation can be defined by a price path $\bar{p}(t)$ and a constraint

$p(t) \le \hat{p}(t)$. Since the price is generally below the monopoly price, the price-cap typically will be binding for a regulated monopolist. In our deterministic model, there is no loss of generality in assuming the price-cap is binding. Ignoring fixed costs, it is immediate from Theorem 1 that the optimal binding price-cap is equal to long run marginal cost, i.e.

$$\hat{p}(t) = c(t) + (r + \delta + \mu)q(t) + \mu c(t) \int_0^{\Delta} e^{-(r+\delta)\tau} d\tau$$

Given a regulatory commitment to this price path, the regulated firm maximizes profits by making cost-minimizing capacity replacement decisions, i.e. the firm will replace capacity after Δ periods, as specified in Theorem 1. Moreover, by construction, the price path is efficient and therefore induces optimal investments in new capacity.

Problems potentially arise in a transition from rate-of-return regulation to price-cap regulation. Suppose that the transition occurs at a date t^{δ} when price is above the optimal level and the rate base exceeds the value of capacity due to under-depreciation; i.e. $p(t^{\delta}) \equiv \phi(X(t^{\delta})) > (\gamma + r + \delta + \mu)q(t^{\delta})$ and $K(t^{\delta}) > q(t^{\delta})X(t^{\delta})$. In this circumstance, a move to optimal price-cap regulation would lead to stranded assets equal to $K(t^{\delta}) - q(t^{\delta})X(t^{\delta})$. That is, under the optimal price path the firm would earn a competitive return only on a rate base value of $q(t^*)X(t^*)$; the balance of the rate base would remain uncompensated.

A naive approach to price-cap regulation is to maintain the initial price level at the prevailing level, $\hat{p}(t^{\delta}) = c(t^{\delta}) + (r + \delta)\frac{K(t^{\delta})}{X(t^{\delta})}$, and to require the price-cap to decline at the rate of technological progress μ. However, this naive approach to price-cap regulation also potentially creates a stranded-costs problem. The problem is illustrated by a transition from naive rate-of-return regulation discussed above. In this case, the ratio of the retail price to the price of new capacity rises over time approaching an upper bound. A move to naive price-cap regulation essentially freezes the ratio of the retail price to the price of new capacity, thereby forcing retail price below the level it would have achieved under continued rate-of-return regulation. Consequently, in the absence of other benefits, the firm is deprived of the return on capital that it anticipated from its previous investment.

An important benefit to the firm under price-cap regulation is the returns to cost reduction from the replacement of obsolete capacity. As we show below, the value of this investment depends on the duration of the price-cap period and the age distribution of the firm's capital stock. To the extent that this benefit is insufficient to compensate the firm for past investments, the need to recover remaining stranded costs creates a second-best pricing problem. The optimal approach is to treat the stranded cost as a fixed cost and to apply Theorem 1. Thus, stranded costs should be recovered by a Ramsey markup on retail prices, or, equivalently, by an appropriate revenue tax. Obviously, any unrecovered fixed costs should be treated similarly.[17]

[17] Even with this correction for stranded-costs recovery, price-cap regulation is flawed in

Our discussion of price caps so far has presumed a permanent commitment to price caps. This is necessary if price-cap regulation is to provide optimal incentives for cost-minimizing capacity replacement decisions. Typically though, price-cap regulation has been implemented in a more limited way over a fixed planning horizon. At the end of the planning horizon, a new price-cap regime is established, or there could be a reversion to rate-of-return regulation. In the remainder of this section we discuss the consequences of a limited horizon for the regulated firm's capacity replacement decisions.

For purposes of this analysis, we assume that the current period is t, and the price-cap regime ends in future period T. Thus the length of the remaining price-cap horizon is $(T - t)$. At the end of the price-cap period we assume for simplicity a reversion to rate-of-return regulation, meaning that the rate-base is kept track of during the price-cap period, and that the firm is entitled to a competitive rate of return on the rate base at date T. Equivalently, we could assume that there is a new price-cap regime at the end of period, such that the firm is held indifferent between the new regime and renewed rate-of-return regulation.[18]

Consider in this scenario the firm's decision whether or not to scrap a unit of older capacity of vintage s. It is profitable to scrap this asset and replace it with state-of-the-art capacity if

$$\frac{1 - e^{-(r+\delta)(T-t)}}{r+\delta}[c(s) - c(t)] - e^{-(r+\delta)(T-t)}[q(s) - q(t)] \geq q(t)$$

The first term on the left-hand-side is the operating cost savings over the price-cap period from replacing the unit of capacity. The second term is the reduction in the value of the rate base at the end of the price-cap horizon. The net benefit must exceed the direct cost of the replacement. Taking account of the deterministic decline in operating and capital costs, this condition is equivalent to

$$\frac{\gamma}{r+\delta}[1 - e^{-(r+\delta)(T-t)}] - e^{-(r+\delta)(T-t)} - \frac{1}{e^{\mu(t-s)} - 1} \geq 0 \qquad (17)$$

Inequality (17) cannot be satisfied if $(T - t)$ is small, but for sufficiently large $(T - t)$, the equality defines a critical $(t - s)$ which is the age of capacity such that replacement is profitable for all older vintages. It is noteworthy that the

an environment with technological uncertaintly, as in Appendix B. As currently practiced, the price-cap is reduced at a fixed rate of productivity growth. However, to track efficient prices, a properly initialized price cap must fall at the rate of actual productivity growth. This could be accomplished by indexing the price cap for an individual firm to the rate of productivity growth of the whole industry. As long as the individual firm were small relative to the industry, this preserves the incentive properties of price cap regulation to efficiently scrap high cost capacity. The problem of adjusting price caps to uncertain costs has been studied by Armstrong and Vickers (1995) in a different context.

[18] The fact that the rate base is typically kept track of during the price cap period suggests that this may be a realistic assumption.

incentive to scrap capacity does not depend on the price path itself, but only on the price-cap horizon.

If the price-cap horizon is not long enough, then it will not be profitable to replace any capacity at all. Even the most ancient vintage of capacity will go not be replaced if

$$\frac{\gamma}{r+\delta}\left[1 - e^{-(r+\delta)(T-t)}\right] - e^{-(r+\delta)(T-t)} \le 0$$

or, equivalently,

$$T - t \le -\frac{1}{r+\delta}\ln\frac{\gamma}{\gamma+r+\delta} \tag{18}$$

Thus, a price-cap horizon that is too short will not have the intended effect of stimulating cost reduction through replacement of obsolete capacity.

Assuming the price-cap horizon $(T - t)$ is sufficiently large, making (17) an equality, defines a critical vintage that is profitable to scrap. Totally differentiating this equality gives

$$\frac{de^{\mu(t-s)}}{de^{-(r+\delta)(T-t)}} = \frac{\gamma+r+\delta}{r+\delta}\left[e^{\mu(t-s)} - 1\right]^2$$

or, equivalently,

$$\frac{d(t-s)}{d(T-t)} = -\frac{\gamma+r+\delta}{\mu}\left[e^{\mu(t-s)} - 1\right]^2\frac{e^{-(r+\delta)(T-t)}}{e^{\mu(t-s)}} < 0 \tag{19}$$

The expression (19) says that, as the price-cap horizon $(T - t)$ lengthens, it is profitable to replace increasingly younger vintages. The derivative itself is monotonically increasing, ranging from -1 as $T - t$ approaches the threshold $(-\frac{1}{r+\delta}\ln\frac{\gamma}{\gamma+r+\delta})$ when no replacement is profitable, and approaching 0 as $T - t$ goes to infinity. An implication is that, for moderately sized price-cap horizons, all capacity replacement is concentrated at the start of the price-cap period. For longer price-cap periods, capacity replacement may continue for a time within the price-cap period, but will cease once the remaining horizon becomes sufficiently short. The above analysis leads to the following theorem:

Theorem 3 Assume that the length of a price-cap period is T, and that the firm is guaranteed a competitive return on its rate base at the end of the price-cap period. Then the following holds: (i) The firm's incentive to buy new capacity and retire existing capacity does not depend on the path of prices, (ii) If T is sufficiently small, the firm has no incentive to retire capacity, (iii) If T is sufficiently large, then the firms' incentives to buy new capacity and retire old capacity are approximately efficient, (iv) For moderate lengths of T, capacity is retired only at the early stages of the price-cap period.

5 Discussion

In the United States, rate-of-return regulation has prevailed as a method of regulating telecommunications for much of the century. Beginning in the 1980's price-cap regulation gradually became ascendant. Now, particularly since the passage of the 1996 Telecommunications Act, there are emerging new forms of wholesale price regulation based on forward-looking cost concepts. It is reasonable to interpret forward-looking economic cost as long run marginal cost, including the user cost of capital, as interpreted in our models and consistent with Salinger's (1998) definition. It may also be reasonable, given the explosion of the internet, to assume (at least for now) growing demand and continuing technological progress, as we do in our basic models.

There seems to be a consensus that assets were underdepreciated under rate-of-return regulation. The conventional wisdom is that regulators felt short-term political pressures to keep rates low. Our models suggest that, as a consequence, with ongoing technological progress, historical costs, as reflected in the rate base, eventually became high relative to forward looking economic cost. This appears to be the case based on rate proceedings in various states. Thus, perhaps paradoxically, a political motivation to keep prices low may have resulted ultimately in excessively high prices.

As we argued section 3, rate-of-return regulation provides little incentive for plant and equipment replacement. Price-cap regulation was motivated largely by its incentive properties for cost reduction. When price is exogenous, the firm has better incentives to minimize cost by making timely capital replacement decisions. In the wake of price-cap regulation, there does appear to have been substantial new investment in capital equipment, as telecommunications firms adopted fiber optic and digital technologies (see Greenstein, McMaster and Spiller (1995)).

Price-cap regulation might have been an opportunity to establish efficient prices based on long-run marginal cost. However, price-cap and related incentive regimes generally were adopted voluntarily, and, therefore, were structured to recover historical costs. In many cases, price caps were set to decline faster than the estimated rate of productivity improvement or were initialized at somewhat below current prices. In these cases, it is likely that prices turned out to be lower than they would have been under rate-of-return regulation, but the regulated firms nevertheless were adequately compensated by the returns from the cost efficiencies that resulted from price-cap incentives.

The Telecommunications Act placed important new obligations on incumbent local exchange carriers (ILECs). In particular, ILECs are required to provide unbundled network elements at cost-based rates. In a controversial regulation, the FCC required the ILECs to provide the entire platform of network elements - essentially enabling the resale of existing services - at wholesale prices based on forward looking costs. Subsequently, the U.S. 8th Circuit Court determined that the Act did not require the wholesale provision of the entire platform of network elements, and that the FCC lacked jurisdiction over the pricing of unbundled elements. Recently, the Supreme Court has vacated the 8th Circuit decision, and restored the FCC's authority to establish a pricing methodology for network elements, but qualified that ILECs were only obligated to provide network elements that were essential for competition.

One possible interpretation of the FCC pricing rule is that it was an attempt to restore efficient pricing of retail services. In our model, this would appear to create a stranded-costs problem. In reality, this is not so clear. In addition to placing unbundling requirements on the ILECs, the Act also removed certain line-of-business restrictions on the Bell Operating Companies (BOCs). In particular, the BOCs will be able to enter long distance markets once they have opened their local markets to competition to the satisfaction of the FCC. It is possible that the expected profits from these new lines of business adequately compensate the ILECs for any stranded costs that might result in a transition from price-cap regulation to wholesale-price regulation based on a forward looking cost pricing methodology. To the extent that there are remaining stranded costs in need of recovery in the transition to a new regulatory regime, our analysis indicates that these should be recovered by sales taxes based on standard Ramsey principles.

References

[1] Armstrong, Mark, Ray Rees and John Vickers, \Optimal Regulatory Lag under Price Cap Regulation," Revista Espanola de Economia, 1995, pp. 93-116.

[2] Averch, Harold and Leland Johnson, \Behavior of the Firm Under Regulatory Constraint," American Economic Review, vol. 52, 1962, pp. 1052-1069.

[3] Besanko, David and Daniel Spulber, \Sequential-Equilibrium Investment by Regulated Firms," Rand Journal of Economics, 1992, pp. 153-170.

[4] Bradley, I. and C. Price, \The Economic Regulation of Privatized Monopolies in the U.K.," Journal of Industrial Economics, vol 37, 1988, pp. 99-106.

[5] Brautigam, Ronald and John Panzar, \Diversi¯cation Incentives under \price-based" and \cost-based" regulation," Rand Journal of Economics, vol. 20, 1989, pp. 373-391.

[6] Brennan, Timothy, \Depreciation, Investor Compensation, and Welfare Under Rate of Return Regulation," Review of Industrial Organization, vol. 6, 1991, pp.73-87.

[7] Cabral, Luis and Michael Riordan, \Incentives for Cost Reduction Under Price Cap Regulation," Journal of Regulatory Economics, vo1. 1, 1989, pp. 133-147.

[8] Dixit, Avinash and Robert Pyndyck, Investment under Uncertainty, Princeton University Press, 1994.

[9] Gilbert, Richard and David Newbery, \The Dynamic E±ciency of Regulatory Constitutions," Rand Journal of Economics, 1994, pp. 538-554.

[10] Greenstein, Shane, Susan McMaster and Pablo Spiller, \The E®ect of Incentive Regulation on Infrastructure Modernization: Local Exchange Companies' Deployment of Digital Technology," Journal of Economics, Management, and Strategy, vol. 4, 1995, pp.187-236.

[11] Hausman, Jerry, \Valuing the E®ect of Regulation on New Services in Telecommunications," Brookings Papers on Economic Activity: Microeconomics, Washington, D.C.: Brookings Institution, 1997.

[12] Jorgenson, Dale W., \Capital Theory and Investment Behavior," American Economic Review, Papers and Proceedings, vol. 53, 1963, pp. 247-259.

[13] Jorgenson, Dale W., \The Theory of Investment Behavior," in R. Ferber, ed., Determinants of Investment Behavior, NBER, 1967.

[14] Joskow, Paul, \The Determination of the Allowed Rate of Return in Formal Regulatory Hearings," Bell Journal of Economics, vol. 3, 1972, pp. 632-644.

[15] Kahn, Alfred, The Economics of Regulation, MIT Press, 1988.

[16] La®ont, Jean-Jacques and Jean Tirole, \Should Governments Commit?' European Economic Review, 1992, pp. 345-353.

[17] Lewis, Tracy and David E. M. Sappington, \Regulatory Options and Price-Cap Regulation," Rand Journal of Economics, vol. 20, 1989, pp. 405-416.

[18] ___ and ___, \Oversight of Long-Term Investment by Short-Lived Regulators," International Economic Review, 1991, pp. 579-600.

[19] Lyon, Thomas, \Regulation with 20-20 Hindsight: "Heads I Win, Tails You Lose," Rand Journal of Economics, 1991, pp. 581-595.

[20] Malcomson, James, \Replacement and the Rental Value of Capital Equipment Subject to Obsolescence," Journal of Economic Theory, vol. 10, 1975, pp. 24-41.

[21] Nickell, S.J., The Investment Decisions of Firms, Cambridge University Press, 1978.

[22] Phillips, Charles F. Jr., The Regulation of Public Utilities, Arlington Virginia: Public Utilities Reports, Inc. 1993.

[23] Salant, David and Glenn Woroch, \Trigger Price Regulation," Rand Journal of Economics, 1992, pp. 29-51.

[24] Salinger, Michael A., \Regulating Prices to Equal Forward-Looking Costs: Cost-Based Prices or Price-Based Costs," Journal of Regulatory Economics, vol. 14, 1998, pp. 149-164.

[25] Sappington, David E.M. and Dennis Weisman, \Seven Myths about Incentive Regulation," in Michael A. Crew (ed.) Pricing and Regulatory Innovations Under Increasing Competition. Boston/Dordrecth/London: Kluwer Academic Publishers, 1996.

[26] Schmalensee, Richard, \An Expository Note on Depreciation and Profitability under Rate-of-Return Regulation," Journal of Regulatory Economics, vol. 1, 1989, pp. 293-298.

[27] Sibley, David, \Asymmetric Information, Incentives, and Price-Cap Regulation," Rand Journal of Economics, vol. 20, 1989, pp. 392-404.

[28] Sidak, J. Gregory and Daniel F. Spulber, \Deregulatory Takings and Breach of the Regulatory Contract," New York University Law Review, vol. 71, 1996.

[29] Tardi®, Timothy J. and William E. Taylor, \Revising Price Caps: The Next Generation of Incentive Regulation Plans," in Michael A. Crew (ed.) Pricing and Regulatory Innovations Under Increasing Competition. Boston/Dordrecth/London: Kluwer Academic Publishers, 1996.

[30] Vogelsang, Ingo, \Price-Cap Regulation of Telecommunications Services: A Long-Run Approach," in Michael Crew (ed.) Deregulation and Diversi¯cation of Utilities. Boston/Dordrecth/London: Kluwer Academic Publishers, 1988.

6 Appendix A: Proof of Theorem 1

The first order conditions of the optimal investment problem with respect to $X(0)$ and $x(\tau)$ have an identical structure:

$$\int_\tau^1 e^{-r\tau-\delta(t-\tau)}u(t;\tau)\{[1+\lambda-\tfrac{\lambda}{2}]p(t)-(1+\lambda)c(\tau)\}dt-e^{-r\tau}(1+\lambda)q(\tau)=0$$

for all $\tau \geq 0$. Moreover, from the first order conditions for $u(\tau;t)$, it easy to show that an optimal utilization function takes on two values: either 0 or 1. In particular,

$$\begin{aligned} u(t;\tau) &= 1 \text{ if } [1+\lambda-\tfrac{\lambda}{2}]p(t)-(1+\lambda)c(\tau)\geq 0 \\ u(t;\tau) &= 0 \text{ otherwise} \end{aligned}$$

Assuming that $p(t)$ is decreasing, there exists a scrapping date for capital of vintage τ such that $u(t;\tau)=1$ for $\tau \leq t \leq \tau+\Delta(\tau)$ and $u(\tau;t)=0$ for $t>\tau+\Delta(\tau)$. Clearly, $\Delta(\tau)$ satisfies

$$[1-\frac{\lambda}{(1+\lambda)^2}]p(\tau+\Delta(\tau))=c(\tau)$$

and we can rewrite the first order conditions for investments as

$$\int_\tau^{\tau+\Delta(\tau)} e^{-rt-\delta(t-\tau)}\{[1+\lambda-\tfrac{\lambda}{2}]p(t)-(1+\lambda)c(\tau)\}dt-e^{-r\tau}(1+\lambda)q(\tau)=0$$

Interchanging t and τ, dividing by e^{-rt}, and differentiating this with respect to t gives

$$\begin{aligned} -\{[1+\lambda-\tfrac{\lambda}{2}]p(t)-(1+\lambda)c(t)\}+(1+\lambda)(r+\delta)q(t) \\ -(1+\lambda)\dot{c}(t)\int_t^{t+\Delta(t)} e^{-r\tau-\delta(\tau-t)}d\tau-(1+\lambda)\dot{q}(t) &= 0 \end{aligned}$$

and rearranging and substituting μ gives

$$[1-\frac{\lambda}{(1+\lambda)^2}]p(t)=c(t)+(r+\delta+\mu)q(t)-\mu c(t)\int_0^{t+\Delta(t)} e^{-(r+\delta)\tau}d\tau \qquad (20)$$

with $\Delta(t)$ satisfying

$$[1-\frac{\lambda}{(1+\lambda)^2}]p(t+\Delta(t))=c(t). \qquad (21)$$

Equations (20) and (21) are solved simultaneously to determine $p(t)$ and $\Delta(t)$, and it must be checked that $p(t)$ is decreasing at this solution.

Substitution of, $c(t) = \gamma q(t)$ into equations (20) and (21) gives (5) for a constant Δ satisfying

$$[1 - \frac{\lambda}{(1+\lambda)^2}]p(\Delta) = \gamma q(0). \tag{22}$$

Clearly, p is declining at the same rate as q. Since q is declining at rate μ, we have from (20):

$$\begin{aligned}
[1 - \frac{\lambda}{(1+\lambda)^2}]p(\Delta) &= [\gamma + r + \delta + \mu + \gamma\mu \int_0^{\Delta} e^{-(r+\delta)\tau} d\tau] e^{-\mu\Delta} q(0) \\
&= [\gamma + r + \delta + \mu + \gamma\mu \frac{1 - e^{-(r+\delta)\Delta}}{r+\delta}] e^{-\mu\Delta} q(0)
\end{aligned}$$

Substituting from (22) for $p(\Delta)$ gives condition (6),which has a unique solution in Δ.

7 Appendix B: Technological uncertainty

We examine the model when technological progress is uncertain and technological improvements arrive according to a Poisson distribution. For simplicity we maintain the assumption that fixed costs are zero. Ramsey principles would apply for the recovery of fixed costs analogously to the deterministic model. That is, the percentage price markup above the long run marginal is inversely proportional to the elasticity of demand. In the model below with no fixed cost, price is simply equal to long run marginal, for which we derive the formula.

Consider an infinite sequence of dates $\{t_i\}$ determined by a Poisson arrival process with a constant hazard rate $\lambda > 0$. The initial date is $t_0 = 0$. Corresponding to these dates is a decreasing sequence of operating costs and capital equipment costs, $\{c_i, q_i\}$, satisfying

$$(c_i, q_i) = ((1+\alpha)c_{i+1}, (1+\alpha)q_{i+1}) \tag{23}$$

for some constant $\alpha > 0$. Since both cost elements are declining at the same rate, we have $c_i = \gamma q_i$ for a constant $\gamma \geq 0$. We assume that the cost structure for new capacity acquired at a date $t \in [t_i, t_{i+1})$ is $(c(t), q(t)) = (c_i, q_i)$. Thus, the expected rate of productivity growth in this model is

$$\mu = \lambda \frac{\alpha}{1+\alpha}. \tag{24}$$

We refer to each t_i as an innovation date, and to the time between innovation dates as an innovation period.

The social welfare has the same structure as in the deterministic model. It is given by (3), with $X(t)$ and $C(t)$ defined by (1) and (2) respectively.

The difference here is that $c(t)$ and $q(t)$ decline stochastically. Therefore, the relevant objective function for the social planner is expected social welfare, with expectations taken over the sequence of innovation dates. Moreover, optimal investment and utilization paths are conditioned on the evolving history of innovations.

Theorem 4 When technological progress is uncertain, (having a Poisson distribution,) the optimal investment path implies that $p(t) = p_i$ for $t \in [t_i; t_{i+1})$ with

$$\begin{aligned} p_i - c_i \quad = \quad & (r + \delta + \mu)q_i \\ & + \mu c_i \left(\frac{1}{r+\delta+\lambda}\right) \sum_{m=0}^{n-2} \left(\frac{\lambda}{r+\delta+\lambda}\right)^m \\ & + \left(\frac{\lambda}{r+\delta+\lambda}\right)^n [p_{n+i} - c_{i+1}] \end{aligned} \tag{25}$$

and

$$p_{n-1+i} \geq c_i \geq p_{n+i}. \tag{26}$$

for some integer n. The optimal utilization path satisfies $u(t;\tau) = 1$ if $p(t) - c(\tau) \geq 0$ and $u(t;\tau) = 0$ otherwise.

Proof. Choosing an investment and capacity retirement path that maximizes expected social welfare implies a sequence of prices $\{p_i\}$ such that $p(t) = p_i$ for $t \in [t_i; t_{i+1})$, and

$$\sum_{m=0}^{n-1} E\{e^{-(r+\delta)t_m}\} \frac{E\{1 - e^{-(r+\delta)(t_{m+1}-t_m)}\}}{r+\delta}[p_{m+i} - c_i] = q_i \tag{27}$$

where expectations are over the sequence of innovation dates. The first order conditions show that price is constant within an innovation period, due to the stationarity of the Poisson process. The first order condition for optimal utilization at each date implies that the capacity replacement cycle is determined by (26). Capacity is retired after n innovation periods. Equation (27) is a competitive break-even condition stating that expected discounted cash flow over the life of an asset exactly covers its acquisition cost.

Substituting

$$E\{e^{-(r+\delta)t_m}\} = \left(\frac{\lambda}{r+\delta+\lambda}\right)^m$$

and

$$\frac{E\{1 - e^{-(r+\delta)(t_m - t_{m-1})}\}}{r+\delta} = \frac{1}{r+\delta+\lambda} \tag{28}$$

into (27) gives

$$\frac{1}{r+\delta+\lambda}\sum_{m=0}^{n-1}\left(\frac{\lambda}{r+\delta+\lambda}\right)^m[p_{m+i}-c_i]=q_i. \tag{29}$$

Subtracting

$$\sum_{m=1}\left(\frac{\lambda}{r+\delta+\lambda}\right)^m[p_{m+i+1}-c_{i+1}]=\lambda q_{i+1}$$

from each side of (29), and rearranging gives

$$\begin{aligned} p_i-c_i &= (r+\delta)q_i+\lambda(q_i-q_{i+1}) \\ &\quad +(c_i-c_{i+1})\sum_{m=1}^{n-1}\left(\frac{\lambda}{r+\delta+\lambda}\right)^m \\ &\quad +\left(\frac{\lambda}{r+\delta+\lambda}\right)^n[p_{n+i}-c_{i+1}]. \end{aligned} \tag{30}$$

Using (23) and (24) gives (25).[19] ■

The theorem has a straightforward interpretation which parallels closely that for the deterministic model (Theorem 1). As in the deterministic model, capacity is scrapped when price drops below its operating cost. In the Poisson model, each vintage of capacity is scrapped after n innovation periods. Price is constant between innovations and drops discretely with the arrival of each innovation. The term $(r+\delta+\mu)q_i$ in the pricing formula of the proposition is the counterpart of the user cost of capital for the basic deterministic model with neither operating costs nor fixed costs ($\gamma = 0$ and $F = 0$), except that here μ must be interpreted as the expected rate of productivity growth.

The term

$$\mu c_i\left(\frac{1}{r+\delta+\lambda}\right)\sum_{m=0}^{n-1}\left(\frac{\lambda}{r+\delta+\lambda}\right)^m=\mu c_i E_{t_{n-1}}\left\{\int_0^{t_{n-1}} e^{-(r+\delta)\tau}d\tau\right\}$$

is the counterpart of the operating improvement term of the deterministic model with declining operating costs. It equals the expected present value of operating improvements resulting from delaying the marginal investment, and reflects the

[19] Equation (30) is an $(n-1)$ order linear difference equation in p_i. The solution is obtained by assuming that $p_{m+i}=\frac{1}{(1+\alpha)^m}p_i$, and substituting and solving to obtain

$$\begin{aligned} p_i &= \frac{1}{\left[1-\frac{\lambda^n}{(r+\delta+\lambda)^n(1+\alpha)^n}\right]} \\ &\quad \times\left\{\left[1+\frac{\mu}{r+\delta}\right]\left[1-\frac{\lambda^n}{(r+\delta+\lambda)^n}\right]+r+\delta+\mu\right\}q_i \end{aligned} \tag{31}$$

which verifies that p_i indeed declines at rate $\frac{\alpha}{1+\alpha}$. This closed form solution is hard to interpret directly, but does enable numerical comparative statics.

fact that an operating improvement is an uncertain prospect that occurs only if delay results in a discrete improvement. The likelihood of such an improvement is λ, and the consequence is to reduce operating costs of new capacity by an amount $\frac{\alpha}{1+\alpha}c_i$; therefore, the expected reduction of operating costs over an interval of size $d\tau$ is μc_i. The term $E_{t_{n-1}}\{\int_0^{t_{n-1}} e^{-(r+\delta)\tau}d\tau\}$ recognizes that the value of the operating cost improvement accrues over a random period of length t_{n-1}, which is the time it takes for the remaining $(n-1)$ innovation periods of the asset's life.

The last term,

$$\left(\frac{\lambda}{r+\delta+\lambda}\right)^n[p_{n+i} - c_{i+1}]; \tag{32}$$

reflects the fact that, in the event of a technological advance, a new vintage of capacity will earn a return over an additional innovation period compared to the economic life of the previous vintage. This is something of a nuisance term. In view of (26) this term is bounded above by $\frac{\alpha}{1+\alpha}(\frac{\lambda}{r+\delta+\lambda})^n c_i$. Clearly, this bound goes to zero as α goes to zero. Holding μ constant, a small value of α corresponds to frequent small innovations. Indeed, for a given μ, the model converges to the deterministic case in the limit as α goes to zero.[20] However, even if α is not small, a numerical analysis shows that $\frac{\alpha}{1+\alpha}(\frac{\lambda}{r+\delta+\lambda})^n$ is approximately zero for reasonable parameter values. This is because, for a given rate of expected technological progress μ, the arrival rate of new technology λ becomes small as the size of the innovation α becomes large. Thus, we conclude that, typically,

$$\frac{p_i - c_i}{q_i} \approx r+\delta+\mu+\mu\gamma\left(\frac{1}{r+\delta+\lambda}\right)\sum_{m=0}^{n-1}\left(\frac{\lambda}{r+\delta+\lambda}\right)^m \tag{33}$$

Moreover, for given values of the other parameters, it can be shown analytically that there is an interval of values for γ that support a given replacement cycle of size n, and approximation (33) becomes exact as γ goes to the upper bound of this interval. Thus, for any positive integer n, there exist parameter values such that the optimal pricing formula under uncertainty is the exact analog of the pricing formula for the deterministic case.[21]

[20] We thank Roy Radner for pointing this out to us.

[21] This approximation indicates that technological uncertainty tends to lower the path of prices compared to deterministic progress at the same expected rate. This is due to a version of Jensen's inequality which states that the expected present value of foregone operating cost improvements in the Poisson model exceeds the present value of foregone operating cost evaluated at the expected life of assets, i.e. if Δ is the random life of assets then

$$E\left\{\int_0^{\Delta} e^{-(r+\delta)\tau}d\tau\right\} < \int_0^{E\{\Delta\}} e^{-(r+\delta)\tau}d\tau$$

This conclusion can be verified with numerical analyses.

8 Appendix C: Demand uncertainty

The following analysis is adapted from Dixit and Pyndyck (1994). To introduce demand uncertainty, we specify the inverse demand curve

$$p = YX^{-\theta}$$

where $\theta = \frac{1}{\epsilon}$ is the inverse demand elasticity, and Y is a geometric Brownian motion with drift parameter α and variance parameter σ. We assume that the only costs are the costs of acquiring new capacity, and that the price of new capacity, q, declines deterministically at rate μ. In particular, there are no additional operating costs ($\gamma = 0$) or fixed costs ($F = 0$), and no technological uncertainty. However, we do allow for positive depreciation ($\delta > 0$). It is assumed that the interest rate (r) is sufficiently large that the optimal investment problem is well-defined.

The solution to the optimal investment problem establishes a ceiling on the price ratio $y = \frac{p}{q}$. Positive investment keeps price from rising above the ceiling. When price drops below the ceiling, due to negative demand shocks, investment in new capacity is zero. More precisely, the optimal investment and price path is characterized by the following complementary slackness conditions

$$p \leq (r + \delta + \mu + \frac{\sigma^2\beta}{2})q$$

$$\dot{X} + \delta X \geq 0$$

$$[(r + \delta + \mu + \frac{\sigma^2\beta}{2})q - p][\dot{X} + \delta X] = 0$$

where β is the positive root of the quadratic equation

$$\frac{\sigma^2}{2}\beta(\beta - 1) + (\alpha - \mu + \delta\theta)\beta - (r + \delta + \mu) = 0$$

In other words, optimal investment equates prices to the user cost of capacity $(r + \delta + \mu + \frac{\sigma^2\beta}{2})q$. If price falls below the user cost of capacity, then it is optimal not to invest. The user cost of capacity has an upward adjustment to account for demand uncertainty. This additive adjustment is equal to $\frac{\sigma^2\beta}{2}$. The parameter β is itself a function of σ and other parameters. However, it is straightforward to show that $\sigma^2\beta$ is monotonically increasing in σ, is zero when $\sigma = 0$, and goes to infinity as σ goes to infinity. Thus, increased demand uncertainty raises the path of prices under an optimal investment plan, possibly substantially.

We sketch the argument for characterizing the threshold value of $y = \frac{p}{q}$ that triggers positive investment The state of the market at any date is a triple $(Y; X; q)$. There will be a region of the state space where it is optimal for no investment to take place. The value function over this region of inaction is denoted $V(Y;X;q)$. It makes sense that the value function is homogeneous of degree one in Y and q, in which case

$$V(Y;X;q) = qv(z;X)$$

where $Z = \frac{Y}{q}$. Bellman's Principle and Ito's Lemma imply

$$(r+\mu)v = \frac{ZX^{1-\theta}}{1-\theta} + \frac{Z^2\sigma^2}{2}v_{ZZ} + (\alpha-\mu)Zv_Z - \delta X v_X$$

over the interior of the zero-investment region. Now define $f = v_X$.and observe that

$$(r+\delta+\mu)f = ZX^{-\theta} + \frac{Z^2\sigma^2}{2}f_{ZZ} + (\alpha-\mu)Zf_Z - \delta X f_X$$

Suppose that

$$f(Z;X) = g(y)$$

where

$$y = ZX^{-\theta}$$

Then

$$(r+\delta+\mu)g = y + \frac{y^2\sigma^2}{2}g'' + (\alpha-\mu+\delta\theta)yg'$$

The solution to this differential equation has the form

$$g(y) = By^{\beta} + \frac{y}{r-\alpha+2\mu+(1-\theta)\delta}$$

where it is assumed that $r > \alpha - 2\mu - (1-\theta)\delta$. The value-matching and smooth-pasting conditions imply that the ceiling value of y at the boundary of the inactive region is

$$y^* = \frac{\beta}{\beta-1}[r-\alpha+2\mu+(1-\theta)\delta]$$

The quadratic equation for β can be rewritten as

$$\frac{\sigma^2}{2}\frac{\beta}{r+\delta+\mu} + 1 = \frac{r-\alpha+2\mu+(1-\theta)\delta}{r+\delta+\mu}\frac{\beta}{\beta-1}$$

whence

$$y^* = r+\delta+\mu + \frac{\sigma^2\beta}{2}$$

Thus, the optimal investment and price path is characterized by the complementary slackness conditions

$$p \quad y^* q$$

$$\dot{X} + \delta X \geq 0$$

$$[y^* q - p][\dot{X} + \delta X] = 0$$