Dynamic Regulatory Price Setting with Partial Regulatory Commitment*

David Besanko and Shana Cui
Northwestern University and University of Florida

July 7, 2022

Abstract

This paper studies a stochastic dynamic game of regulatory price setting with partial commitment. A regulated firm makes investments each period that stochastically reduce its marginal costs. The infinite horizon over which the firm and the regulator interact is divided into regulatory cycles that last for T-periods. At the beginning of a cycle, the regulator commits to a T-period schedule of access prices and a lump-sum transfer to maximize a weighted welfare function, subject to the firm's participation constraint. It cannot, however, commit to schedules for subsequent regulatory cycles. Regulatory lag and access prices each affect the firm's investment but in different ways. Investment in any period of the cycle (except the last) decreases in the variable access charges in all subsequent periods of the cycle (the "Arrow effect"). Recognizing this, the regulator distorts the schedule of variable access charges downward relative to what it would choose if it maximized welfare within the cycle. The optimal variable access charge is less than the static-first price corresponding to the expected marginal cost for that period. Regulatory lag induces an additional source of static inefficiency besides the traditional one (prices within a cycle not tailored to changes in actual productivity). Computational analysis reveals that as inducements to investment, regulatory lag and access prices are usually substitutes rather than complements. The regulator prefers a longer lag to a shorter lag except for extreme parameterizations. However, there are diminishing returns to regulatory lag, with most of the gains achieved with lags no greater than 5 to 6 years.

1 Introduction

In the 1980s and 1990s, as countries around the world privatized or deregulated vital infrastructure such as railroads, airports, and telecommunications systems, an elegant and powerful mechanism for regulating natural monopoly firms emerged: price cap regulation (Littlechild 1983, Beesley and Littlechild 1989, Cabral and Riordan, 1989, Brennan, 1989, Linhart and Radner 1992, Armstrong, Cowan, and Vickers 1994). Price cap regulation has two distinct economic features: a fixed length of time between regulatory reviews—regulatory

*We thank David Sappington for his very helpful comments and suggestions.

lag—and a schedule of price caps that decline in real terms—the X-factor.[1] Regulatory lag changes the nature of the regulatory interaction from a cost-plus contract to a fixed-price price contract, and in so doing it strengthens incentives for cost efficiency, an insight that emerges from both the literature on rate of return regulation (Baumol and Klevorick 1970, Bailey 1974) and the mechanism design literature (e.g., Laffont and Tirole, 1984, 1993, and Cowan 2002). The X-factor can ensure that the firm shares some or all of the benefits from improved efficiency with consumers, an insight comes from the literature on non-Bayesian regulatory mechanisms (Vogelsang and Finsinger 1979, Vogelsang, 1989).

This paper studies a stochastic dynamic game of regulatory price setting in which regulatory lag and the price schedule play prominent roles in motivating a firm to make productivity-enhancing investments. Our model consists of a regulated network firm that provides access to bottleneck infrastructure and a regulator that sets the price of access, which is paid by downstream operating firms that provide a service to end consumers.[2] The regulated firm makes investments each period that stochastically improve the productivity of its network and reduce its marginal costs. The regulator perceives that it is in a continuing relationship with the firm over the infinite horizon. It is an active player in the game that seeks to maximize expected social welfare, subject to satisfying the network firm's participation constraint. The horizon over which the firm and the regulator interact is divided into regulatory cycles that last for T-periods. Prior to the beginning of a new regulatory cycle, the regulator reviews the firm's operating capabilities and verifies its productivity. Based on this verified productivity, the regulator commits to a T-period schedule that specifies a price for each period of the cycle. This schedule is an analogous to the choice of an X-factor in price cap regulation.[3]

Our paper makes four contributions to the literature on price cap regulation. First, we highlight that while regulatory lag and the schedule of access prices both create incentives for more investment, they work differently. For the first $T-1$ periods of the regulatory cycle, the firm effectively faces a fixed-price contract, while in period T it faces a cost-plus contract. In the first $T-1$ periods, therefore, the interests of the regulator and the regulated firm are fully aligned when it comes to investment: marginal social welfare from increasing investment is the same as the marginal expected profit of the firm.[4] A longer regulatory lag is valuable to the regulator because it lengthens the period of full alignment.

The access charges, by contrast, directly increase the return on productivity-enhancing effort because the benefits of cost-reducing investment in period t are magnified when the firm produces more in the remaining periods $t+1, \ldots, T$ of the regulatory cycle. However, cost-reduction in the current regulatory cycle is *not* the reason the regulator uses the access charges as an incentive device. If the regulator's horizon was limited to the current cycle, it

[1] One could argue that price cap regulation had a third distinctive feature: delegation of pricing decisions to the regulated firm so long as it adheres to the cap, i.e., pricing discretion. However, pricing discretion is not necessarily a feature of all applications of price cap regulation, and it is not necessarily the case that it is welfare-optimal for the regulator, even when the firm has hidden information. See Armstrong and Vickers (2000) or Armstrong and Sappington (2007) for extensive discussion of pricing discretion.

[2] A special case of our model is a traditional regulatory pricing setting in which the network firm is vertically integrated and sells to end consumers directly.

[3] The price schedule is not *precisely* the same as an X-factor because the schedule specifies the access prices themselves, not caps. Moreover, the prices in the schedule need not change at a constant percentage rate. Still, the price adjustments in the tariff schedule partly reflect expected reductions in marginal costs over time due to the firm's investments, and thus, like X-factors in price cap regulation, they share prospective gains in productivity with consumers.

[4] However, they are not aligned when it comes to the access tariff because the regulator's objective function is social surplus, while the regulated firm cares only about its own discounted profit.

would not need to distort the access prices to create additional investment incentives. Nor would it need to distort prices if regulatory lag was infinite. In both instances—which have the common feature that the regulator and the firm operate with the same time horizon—the marginal investment incentives of the firm and regulator would be fully aligned. The use of the access prices as an incentive device arises when the regulator recognizes that the benefits of cost-reducing investment in the current cycle accrue to future cycles, while the firm (because the expected present value of its profit is reset to zero at the beginning of each new cycle) internalizes investment benefits only within a cycle. Our model thus highlights an aspect of regulation that has been underemphasized in the literature on price caps: linkages across regulatory reviews and the benefits of designing current policies (including X-factors) that can improve society's position for future reviews.

Second, we show that the use of a price schedule as an incentive device worsens the traditional trade-off between static efficiency and dynamic efficiency that arises under price cap regulation. (That is, regulatory lag encourages investment, but it prevents price from being tailored to the changes in productivity that result from greater investment.) That trade-off arises in our model, but we show that there is an additional source of static inefficiency. Even if the firm's actual productivity equaled its expected productivity in a particular period of the regulatory cycle, the price, which is distorted to encourage more investment, would still be less than the one that maximizes welfare in that period. This worsens static inefficiency.

Third, while regulatory lag and the price schedule could be substitutes or complements when it comes to creating investment incentives, in computing equilibria over a wide swath of parameter space, we find that they are much more likely to be substitutes than complements. This implies with a longer regulatory lag the changes in prices within a cycle are more likely to purely reflect productivity changes than would be the case with a shorter lag.

Fourth, our analysis highlights that regulatory lag is an enormously and robustly powerful incentive device. We show that a longer lag does not *necessarily* increase welfare. However, in our computational analysis, we find that it takes extreme parameter values for shorter lag to increase social welfare (e.g., extremely high marginal investment costs and virtually no value created in the vertical structure), and even then, the gain from shorter lags is minimal. When when we compute equilibria over a wide range of plausible parameter values and for regulatory review periods ranging from one year to eight years, we find that expected social welfare is nearly always higher for a longer regulatory lag.

We also endogenize the lag by allowing the regulator to choose its most preferred lag at the beginning of a regulatory cycle. In other words, the regulator no longer commits to fixed lag but instead chooses a lag that optimizes discounted welfare for any particular productivity level that is revealed during the regulatory review. When we compute the regulatory equilibrium for our baseline parameterization, we find that the regulator chooses the longest feasible lag whatever the review reveals about the firm's productivity.

Our paper adds to that portion of the literature on regulatory price setting in which the regulated firm is modeled as a dynamic optimizer who anticipates how its current decisions will shape its future costs as well as future regulatory decisions. In these models, productivity or costs evolve stochastically and endogenously through time, which in turn induces endogenous evolution of regulated prices. Related papers include Lima and Gómez-Lobo (2010), Pint (1992), and Biglaiser and Riordan (2000). Our paper is most closely related to those by Linhart, Radner, and Sinden (1991) (hereafter LRS) and Armstrong, Rees, and Vickers (1995) (hereafter ARV). LRS model a dynamic game between a regulator the manager of a regulated firm. Similar to our model, the manager in LRS makes non-contractable decisions over time that can stochastically increase the firm's productivity. In contrast to our model, the regulator does not seek to maximize an explicit objective but instead follows a rule that entails decreased prices over time and which motivates the manager to increase productivity. ARV also formulate an explicitly dynamic game between the regulator and the regulated firm, and like our model, they assume that the regulator maximizes welfare over the infinite horizon. In ARV, the regulated firm's marginal cost in any period can either be high or low, and the firm makes cost-reducing investments that can reduce the firm's marginal cost if it is high or keep it low if it is low. The regulator commits to a regulatory review period and sets a fixed price over this period. Through time, the regulator adjusts its review period and fixed price as circumstances evolve. Our model differs from ARV's in three respects. The first is minor: ARV assume a one-part tariff, while we assume that the firm can receive a lump-sum subsidy.[5] Second, ARV assume that the regulator sets a uniform price over the entire regulatory horizon, while we allow the regulator to set a time-varying schedule of prices over the review period. This enables us to study the benefits of using prices as an incentive mechanism.[6] Finally, we have a richer state structure than ARV, with multiple cost states, instead of two. This enables us to do computational analysis with empirical plausible parameter values to gain insight into practical policy design issue such as the relative magnitude of gains from increasing regulatory lag.

The remainder of the paper is organized as follows. Section 2 describes the model. Section 3 states a series of general results on the Markov perfect equilibrium. Section 4 reports the results from our computational analysis. Section 5 considers an extension of the model in which we consider the case in which the regulator cannot commit over time to a fixed value of T, i.e., T can vary from regulatory cycle to regulatory cycle. Section 6 summarizes and concludes.

2 Model

We consider a vertical structure consisting of end consumers who purchase a service from N downstream operating firms, who in turn rely on a regulated network firm for access to bottleneck infrastructure. For example, as in the U.K., the operating firms could be freight railway undertakings in a vertically separated rail system, and the network firm could be the railway infrastructure manager. We normalize so that one unit of downstream service requires one unit of infrastructure access.

2.1 Downstream surplus

When operating firm i pays the regulated price c to access the network, its total costs are $(c + \mu_i)\, Q_i + f_i$, where μ_i is a marginal cost of operating, and f_i is a fixed operating cost. Downstream operating firms are assumed to attain a Nash equilibrium in prices, which in turn implies an equilibrium quantity $D_i^*(c)$ for each firm i and a market demand for network access given by $D(c) = \sum_{j=1}^{N} D_j^*(c)$.[7]

Downstream social surplus is the sum of end consumer surplus and downstream firm profit, $\Psi(c)$, less the social cost $\Omega = (1 + \lambda) \sum_{i=1}^{N} f$ subsidies provided to downstream firms

[5] Later in the paper, we discuss how our results would change if we had one-part pricing.

[6] ARV briefly consider the possibility of time varying prices at the end of their paper.

[7] Letting $Q_i(P_1, \ldots, P_N)$ denote firm i's demand function, the Nash equilibrium prices $P_1^*(c), \ldots, P_N^*(c)$ are given by $P_i^*(c) = \mu_i + c - \frac{D_i^*(c)}{\frac{\partial Q_i(P^*(c),\ldots,P^*(c))}{\partial P_i}}$, $i = 1, \ldots, N$, and $D_i^*(c) = Q_i(P_1^*(c), \ldots, P_N^*(c))$.

to cover their fixed costs, where $\lambda \geq 0$ is the marginal cost of public funds.[8] Consumer surplus plus downstream profit is

$$\Psi(c) \equiv V\left(D_1^*(c), \ldots, D_N^*(c)\right) - \sum_{i=1}^{N} \mu_i D_i^*(c) - cD(c),$$

where $V(\cdot)$ is the gross benefit function of a representative end consumer.[9] Utility maximization by the representative consumer and profit maximization by downstream firms can be shown to imply

$$\Psi'(c) = -D(c) + \left[\overline{P}^*(c) - (\overline{\mu}(c) + c)\right] D'(c), \tag{1}$$

where $\overline{P}^*(c) = \sum_{i=1}^{N} \frac{D_i^{*\prime}(c)}{D'(c)} P_i^*(c)$ and $\overline{\mu}(c) = \sum_{i=1}^{N} \frac{D_i^{*\prime}(c)}{D'(c)} \mu_i$ are weighted averages of the equilibrium prices of the operating firms and their marginal costs, respectively.[10]

2.2 Productivity and investment

The network firm's profit in period τ is $(c_\tau - \widetilde{\eta}_\tau)D(c_\tau) - F - I_\tau$, where $\widetilde{\eta}_\tau \equiv \eta(\theta_{\tau-1})$ is period-τ marginal cost which depends on the realization of a productivity draw $\widetilde{\theta}_{\tau-1}$. The firm's fixed cost consists of an exogenous component F and endogenous productivity-enhancing investment I_τ in period τ. We let θ denote a particular realization of productivity, where $\theta \in [\underline{\theta}, \overline{\theta}]$, and $\overline{\theta} >> \underline{\theta}$. We refer to θ as the firm's state. Increases in productivity reduce marginal cost, i.e., $\eta'(\cdot) < 0$. Let $\Delta\eta(\theta) \equiv \eta(\theta) - \eta(\theta + 1)$ denote the cost saving from a one-unit improvement in productivity.

Following Doraszelski and Besanko (2004), the productivity draw in period τ is given by $\widetilde{\theta}_\tau = \theta_{\tau-1} + \widetilde{\rho}_\tau$, where $\widetilde{\rho}_\tau \in \{-1, 0, 1\}$ is a productivity shock. Productivity can thus increase or decrease by at most one unit per period from its current level. This Markovian productivity process implies that the past history of productivity change is summarized in current productivity. This has a plausible implication: while the current period's investment in productivity enhancement can potentially shape the trajectory of future productivity realizations, its effect will not be extreme. Thus, productivity will evolve gradually, a sensible characteristic, we believe, for regulated infrastructure networks.

The probability distribution of the productivity shock is given by

$$\begin{aligned}
\Pr(\widetilde{\rho}_\tau &= 1) = (1-\delta)G(I_\tau) \\
\Pr(\widetilde{\rho}_\tau &= 0) = 1 - (1-\delta)G(I_\tau) - \delta(1 - G(I_\tau), \\
\Pr(\widetilde{\rho}_\tau &= -1) = \delta(1 - G(I_\tau)),
\end{aligned}$$

where δ is an exogenous parameter we refer to as depreciation, and $G(\cdot)$ is a strictly increas-

[8]The assumption that downstream firms are subsidized to cover their fixed costs is not essential to our analysis, but it simplifies notation a little.

[9]We assume that end consumers have a quasi-linear utility function $V(\mathbf{Q}) + Z$ where $\mathbf{Q}$ is the vector of quantities provided by downstream firms, and Z is a numeraire. The benefit function $V(\cdot)$ is assumed to be increasing in the quantities Q_i and strictly concave.

[10]An alternative formulation of our model would be to assume that the regulated monopolist sells directly to end consumers. In that case, $\mu_i = 0$, and $\overline{P}^*(c) = c$. We thus have the familar result that $\Psi'(c) = -D(c)$. In addition, because there is no subsidy to downsteam firms, $\lambda = 0$.

ing, strictly concave function with $G(0) = 0$ and $\lim_{I\to\infty} G(I) = 1$.[11] We further assume that for all $I \in [0, \infty)$

$$G'(I) = \frac{1}{\gamma}\frac{1 - G(I)}{G(I)}, \tag{2}$$

where $\gamma > 0$ is a parameter. This formulation satisfies the Inada conditions $\lim_{I\to 0} G'(I) = \infty$ and $\lim_{I\to\infty} G'(I) = 0$.

It is convenient to think of the firm directly determining $q_\tau = G(I_\tau)$—which hereafter we refer to as investment—and in doing so, it incurs a total cost $I(q_\tau) = G^{-1}(q_\tau)$. The associated marginal cost of investment is thus $I'(q_\tau) = \frac{1}{G'(I(q_\tau))}$. Given (2), $I'(q_\tau) = \gamma\frac{q_\tau}{1-q_\tau}$, and the associated total investment cost is $I(q_\tau) = \gamma\left[-q_\tau - \ln(1 - q_\tau)\right]$.

2.3 Static optimum

As a benchmark, consider a static model in which marginal cost equals η, and the regulator determines a price c. We assume that the regulator also provides the firm with a lump-sum subsidy A. In practice, infrastructure network providers are often subsidized by government. For example, the U.K. rail network provider, Network Rail, receives 70 percent of its funding from government grants (Network Rail 2021). Similarly, in the Dutch railway system, the network firm ProRail railroad receives about 58 percent of its revenue from government allocations (Statistica, 2018). An alternative interpretation of the subsidy applicable to airports is that the regulator uses single-till regulation in which a forecast of revenue from an airport's non-aeronautical activities such as retail establishments and parking is included along with revenue received from landing charges to airlines to determine the airport's overall revenue requirement. Under this interpretation, A would be the commercial revenue forecast approved by the regulator.

The firm is assumed to have the option to exit the market. If it does, it receives a per-period profit normalized to zero. Exit is assumed to be sufficiently costly to the regulator that it will choose the price and subsidy to satisfy the network firm's participation constraint.

The network firm's profit is $(c - \eta)D(c) - F + A$, where F is its fixed operating cost. The regulator's objective is a weighted sum of downstream surplus (less the social cost of the network firm's subsidy) and network firm profit, with the welfare weight on the latter denoted by ψ, and $\psi \in [0, 1)$:

$$\Psi(c) - \Omega - (1 + \lambda)A + \psi\left[(c - \eta)D(c) - F + A\right]. \tag{3}$$

[11] For the boundary cases in which $\theta_{\tau-1} = \underline{\theta}$ and $\theta_{\tau-1} = \overline{\theta}$, the probability distribution of $\widetilde{\rho}_\tau$ extends in the natural way. When $\theta_{\tau-1} = \underline{\theta}$

$$\begin{aligned}\Pr(\widetilde{\rho}_\tau &= 1) = (1-\delta)G(I_\tau)\\ \Pr(\widetilde{\rho}_\tau &= 0) = 1-(1-\delta)G(I_\tau),\\ \Pr(\widetilde{\rho}_\tau &= -1) = 0,\end{aligned}$$

and when $\theta_{\tau-1} = \overline{\theta}$,

$$\begin{aligned}\Pr(\widetilde{\rho}_\tau &= 1) = 0\\ \Pr(\widetilde{\rho}_\tau &= 0) = 1-\delta(1-G(I_\tau)\\ \Pr(\widetilde{\rho}_\tau &= -1) = \delta(1-G(I_\tau)).\end{aligned}$$

The regulator chooses c, A to maximize (3), subject to the network firm's participation constraint, $(c-\eta)D(c)-F+A \geq 0$. Because the weighted welfare function strictly decreases in A, the participation constraint must bind. The static welfare optimum $c^0(\eta)$ thus solves

$$\max_c \omega(c,\eta) \equiv \Psi(c) - \Omega + (1+\lambda)\left[(c-\eta)D(c) - F\right]. \tag{4}$$

Using (1), the solution is characterized by the condition[12].

$$c^0(\eta) = \eta - \frac{\lambda}{1+\lambda}\frac{D(c^0(\eta))}{D'(c^0(\eta))} - \frac{\overline{P}^*(c^0(\eta)) - \overline{\mu}(c^0(\eta)) - c^0(\eta)}{1+\lambda}, \tag{5}$$

The optimal $\frac{\beta}{\gamma}\left[\begin{array}{c}(1-\delta)\Delta u_{t+1}(\theta, \mathbf{c}_{t+1}) \\ +\delta\Delta u_{t+1}(\theta-1, \mathbf{c}_{t+1})\end{array}\right]$price $c^0(\eta)$ must increase in η.[13] As is standard in full-information models of optimal regulatory pricing, the solution is independent of the welfare weight ψ but not the social cost of public funds λ.[14] The static optimal price could be less than marginal cost to counteract market power exercised by the downstream operating firms (Laffont and Tirole 2000).[15] Throughout, we let $\omega^0(\eta) \equiv \omega(c^0(\eta), \eta)$ denote maximum static welfare when marginal cost is η.

2.4 Regulatory dynamics

Figure 1 illustrates the timing of the game between the firm and the regulator. The firm makes investment decisions every period. Every T periods (the regulatory lag) the regulator observes the firm's productivity and then commits to a schedule of access charges $\mathbf{c} = (c_1, \ldots, c_T)$ and a lump-sum subsidy A.[16] This process repeats itself over the infinite horizon. For now the regulatory lag T is exogenous; later we relax this assumption and allow the regulator to determine T. While τ indexes time more generally in our model, we let $t = 1, \ldots, T$ denote a typical time period within the interval between reviews—what we call the *regulatory cycle*.

The firm's productivity in period t of a cycle is $\widetilde{\theta}_{t-1}$. The firm's investments q_t are unobservable to the regulator, and except for $\widetilde{\theta}_0$, so too are the realizations of $\{\widetilde{\theta}_{t-1}\}_{t=1}^{T}$.

---

[12] The second-order condition for the static welfare maximization problem, $\omega''(c^0(\eta)) < 0$, can be shown to hold if and only if

$$m^*(c^0(\eta)) + 2\lambda - \lambda\frac{D(c^0(\eta))D''(c^0(\eta))}{\left[D'(c^0(\eta))\right]^2} > 0,$$

where $m^*(c^0(\eta)) \equiv \frac{d\overline{P}^*(c^0(\eta))}{dc} - \frac{d\mu(c^0(\eta))}{dc}$ is the net pass-through rate for downstream firms. This condition holds if the demand functon is concave, linear, or "not too" convex.

[13] Differentiating the first-order condition $\omega'(c^0(\eta)) = 0$, with respect to η and rearranging terms gives us $\frac{dc^0(\eta)}{d\eta} = -\frac{(1+\lambda)D(c^0(\eta))}{\omega''(c^0(\eta))} > 0$, since the second-order condition of the static welfare maximization problem implies $\omega''(c^0(\eta)) < 0$.

[14] This is because when $\lambda = 0$ and $\psi \in [0, 1)$, the regulator is indifferent between increasing the network firm's overall profit via an increase in its operating profit $(c-\eta)D(c) - F$ (through c) or by increasing its subsidy A. By contrast, when $\lambda > 0$, the regulator strictly prefers to increase the network firm's profits by increasing its operating profits rather than increasing its subsidy.

[15] In an alternative formulation of our model in which the regulated monopolist sells directly to end consumers (which as noted above entails $\lambda = 0$) and the regulator maximizes a weighted sum of consumer surplus and the regulated firm's profit, with a smaller weight on the latter, then the participation constraint will bind, and condition 5 reduces to the traditional efficiency condition $c^0(\eta) = \eta$.

[16] We show in the Online Appendix that there is no loss of generality in assuming the firm receives a single subsidy every T periods as opposed to a sequence of subsidies A_t for $t = 1, \ldots, T$.

Before a new cycle begins, the regulator reviews the firm's assets and operating capabilities, enabling it to observe the productivity realization $\widetilde{\theta}_T = \theta$ emerging from the last period of the just-ended regulatory cycle. (This realization of $\widetilde{\theta}_T$ becomes the (singleton) support of the firm's productivity $\widetilde{\theta}_0$ in the first period of the next cycle.) Based on this observed productivity, the regulator sets the schedule of access charges for the upcoming cycle and determines the subsidy. The regulator's objective is to maximize discounted expected social welfare over the infinite horizon. The regulator and the firm use a common discount factor $\beta \in (0, 1)$.

As in the static benchmark, the network firm has the option to exit the market, and receives a discounted stream of profits equal to zero if it does. Using Baron and Besanko's (1987) concept of fairness, we assume that as long as the access pricing schedule and subsidy provide the firm with an expected present value of profit over the horizon of the regulatory cycle at least as large as the firm's outside option, the firm is legally bound to remain in the market for each period $1, \ldots, T$ following a regulatory review.[17]

The regulator is a strategic player in this game, and as such, it behaves optimally given the information it observes. The regulator cannot commit to future schedules access tariffs beyond period T, and because it cannot observe the realizations of $\{\widetilde{\theta}_{t-1}\}_{t=2}^T$, the regulator cannot commit to a schedule $(\mathbf{c}(\{\widetilde{\theta}_{t-1}\}_{t=2}^T), A(\{\widetilde{\theta}_{t-1}\}_{t=2}^T)$ of state-contingent access prices and subsidies.[18] The upshot is that the regulator has partial, but not complete commitment ability. The longer the regulatory lag T, the greater the degree of the regulator's commitment.

2.5 Regulated firm's investment problem and Bellman equation

Let $\mathbf{c}_t = (c_t, \ldots, c_T)$ denote the access charges that remain to be implemented in any period $t = 1, \ldots, T-1$ of the regulatory cycle prior to the last period. If realized productivity in period t is $\theta \in (\underline{\theta}, \overline{\theta})$ (i.e., $\widetilde{\theta}_{t-1} = \theta$), the firm's optimal investment program can be summarized by a Bellman equation

$$\begin{aligned} u_t(\theta, \mathbf{c}_t) \quad = \quad & \max_{q_t \in [0,1]} (c_t - \eta(\theta))D(c_t) - F - I(q_t) + \beta u_{t+1}(\theta, \mathbf{c}_{t+1}) \\ & + \beta \left\{ (1-\delta) q_t \Delta u_{t+1}(\theta, \mathbf{c}_{t+1}) - \delta(1-q_t) \Delta u_{t+1}(\theta - 1, \mathbf{c}_{t+1}) \right\}, \end{aligned} \tag{6}$$

where $u_t(\theta, \mathbf{c}_t)$ is the present value of the network firm's operating profits (exclusive of the monetary transfer A) in periods $t, \ldots, T-1$ of the regulatory cycle, and $\Delta u_{t+1}(\theta', \mathbf{c}_{t+1}) = u_{t+1}(\theta'+1, \mathbf{c}_{t+1}) - u_{t+1}(\theta', \mathbf{c}_{t+1})$, $\theta' = \theta, \theta - 1$.[19] Further note that (6) implies that investment in period t of a regulatory cycle depends only on the access prices $\mathbf{c}_{t+1}$ in the remaining periods $t+1, \ldots, T$. It follows that the access prices in later periods affect more periods of investment than the access prices in earlier periods. (Thus, for example, the price c_2 in

[17] Fairness corresponds to a setting in which the network firm cannot wthdraw its assets without regulatory approval.

[18] We rule out the possibility that the regulator can commit to a T-period menu of contracts $(\mathbf{c}(\{\widetilde{\theta}_{t-1}\}_{t=2}^T), \mathbf{A}(\{\widetilde{\theta}_{t-1}\}_{t=2}^T)$ that induces the firm to self select based on its productivity $\widetilde{\theta}_{t-1}$ (or equivalently, to report its productivity truthfully every period.) Our focus here is on a relatively simple mechanism is broadly similar to price cap regulation used in practice.

[19] For $\theta = \underline{\theta}$ and $\theta = \overline{\theta}$, the recursion in (6) is modified using the transition probabilities in (**??**) and (**??**). Throughout the rest of this paper, it will be understood that analytical expressions pertain to interior realizations $\theta \in (\underline{\theta}, \overline{\theta})$ of the productivity process and that expressions for boundary realizations $\theta = \underline{\theta}$ and $\theta = \overline{\theta}$ can be obtained by using the transition probabilities in (**??**) and (**??**).

period 2 affects only the investment in period 1, while the price c_T in the last period of the cycle affects investments in periods $1, \ldots, T-1$.)

In period T of the regulatory cycle, the network firm anticipates that at the end of that period the regulator will observe the realization of its productivity $\widetilde{\theta}_0$ for the first period of a new cycle. The firm expects that the regulator will set a tariff schedule $\mathbf{c}(\widetilde{\theta}_0)$ and subsidy $A(\widetilde{\theta}_0)$ for that upcoming cycle contingent on the realization of $\widetilde{\theta}_0$. (In equilibrium, the firm's expectation of the regulator's strategy will be correct.) If the realization of $\widetilde{\theta}_{T-1}$ was θ (so that the firm's marginal cost in in period T is $\eta(\theta)$), the possible realizations of $\widetilde{\theta}_0$ are $\theta-1, \theta$, or $\theta+1$. The network firm's period-T Bellman equation is then

$$\begin{aligned} u_T(\theta, c_T) &= \max_{q_T \geq 0} (c_T - \eta(\theta))D(c_T) - F - I(q_T) + \beta\left[u_1(\theta, \mathbf{c}(\theta)) + A(\theta)\right] \\ &+\beta \left\{ \begin{array}{c} (1-\delta)q_T\left\{[u_1(\theta+1, \mathbf{c}(\theta+1)) + A(\theta+1)] - [u_1(\theta, \mathbf{c}(\theta)) + A(\theta)]\right\} \\ -\delta(1-q_T)\left\{[u_1(\theta, \mathbf{c}(\theta)) + A(\theta)]) - [u_1(\theta-1, \mathbf{c}(\theta-1)) + A(\theta-1)]\right\} \end{array} \right\} \end{aligned} \tag{7}$$

where $u_1(\theta, \mathbf{c}_1) + A$ is the firm's value at the beginning of period 1. Unlike the expressions for $u_t(\cdot)$ for $t < T$, here the network firm anticipates that a change in its productivity level will affect its access prices.

2.6 Equilibrium

We study the Markov perfect equilibrium (MPE) of the game between the regulator and the network firm, which consists of objects $\{q_t^*(\cdot, \mathbf{c}_t)\}_{t=1}^T$, $\{u_t^*(\cdot, \mathbf{c}_t\}_{t=1}^T$, $\mathbf{c}^*(\cdot), A^*(\cdot)$ satisfying the following conditions:

- For any θ and any schedule of access prices and subsidy, $\mathbf{c}$ and A, the network firm's value functions $u_t^*(\theta, \mathbf{c}_t), t = 1, \ldots, T-1$ satisfy (6). The network firm's value function, $u_T^*(\theta, c_T)$ satisfies (7) given the regulator's equilibrium strategy $\mathbf{c}^*(\theta), A^*(\theta)$. The firm's equilibrium investment strategies $\{q_t^*(\theta, \mathbf{c}_t)\}_{t=1}^T$ solve the corresponding optimization problems in (6) and (7).
- For any θ, the equilibrium access prices and subsidy $\mathbf{c}^*(\theta), A^*(\theta)$ maximizes expected social welfare, subject to the network firm's participation constraint, given the firm's equilibrium investment strategy.[20]

3 Characterization of the MPE

3.1 Preliminaries[21]

Because the expected social cost of subsidizing the network firm is $(1+\lambda)A$, and expected welfare strictly decreases in A, the participation constraint in the regulator's optimization problem will bind, i.e., $u_1(\theta, \mathbf{c}^*(\theta)) + A^*(\theta) = 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. From (7) we can then obtain

Proposition 1 *If the firm anticipates that the regulator will choose an optimal regulatory policy* $\mathbf{c}^*(\cdot), A^*(\cdot)$ *in the next cycle, then for any policy* $\mathbf{c}, A$ *the firm faces in the current*

[20] We formally state the regulator's optimization problem in the next subsection.

[21] All expression and results stated in this section, including Proposition 1, are derived or proven in the Online Appendix.

regulatory cycle, (a) the firm does not invest in the last period T of the cycle, i.e., $q_T^*(\theta) = 0$; (b) for any realization θ of $\widetilde{\theta}_{T-1}$,

$$u_T(\theta, c_T) = (c_T - \eta(\theta))D(c_T) - F. \tag{8}$$

Proposition 1, part (a) is a standard result. In the terminal period, as it looks ahead to the next regulatory review, the firm faces cost-plus regulation, squashing its incentive to enhance its productivity. Part (b) of the result then follows directly: the present value of the network firm's operating profit at the beginning of period T of the regulatory cycle is simply its static operating profit in that period.

The recursion in (6) implies that for any realization $\theta \in (\underline{\theta}, \overline{\theta})$ of the firm's productivity $\widetilde{\theta}_{t-1}$ in any period of the regulatory cycle prior to the last, the firm's optimal investment is given by[22]

$$\begin{aligned} q_t^*(\theta, \mathbf{c}_{t+1}) &= \frac{\beta \begin{bmatrix} (1-\delta)\Delta u_{t+1}(\theta, \mathbf{c}_{t+1}) \\ +\delta \Delta u_{t+1}(\theta - 1, \mathbf{c}_{t+1}) \end{bmatrix}}{\gamma + \beta \begin{bmatrix} (1-\delta)\Delta u_{t+1}(\theta, \mathbf{c}_{t+1}) \\ +\delta \Delta u_{t+1}(\theta - 1, \mathbf{c}_{t+1}) \end{bmatrix}} \in (0,1), \\ t &= 1, \ldots, T-1. \end{aligned} \tag{9}$$

The present value of the network firm's operating profit at the beginning of the regulatory cycle can be expressed as

$$u_1(\theta, \mathbf{c}) = \sum_{t=1}^{T} \beta^{t-1} \left\{ \begin{array}{c} (c_t - \widehat{\eta}_t(\theta, \mathbf{c}_2))D(c_t) - F \\ -E_{\widetilde{\theta}_{t-1}} \left[I(q_t^*(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1})) | \theta, \mathbf{c}_2 \right] \end{array} \right\}, \tag{10}$$

where $\widehat{\eta}_t(\theta, \mathbf{c}_2) \equiv E_{\widetilde{\theta}_{t-1}} \left[\eta(\widetilde{\theta}_{t-1}) | \widetilde{\theta}_0 = \theta, \mathbf{c}_2 \right]$ is the expected marginal cost in period t of the regulatory cycle when the network firm's productivity at the beginning of the regulatory cycle is θ, and the firm follows its optimal investment program throughout the cycle.[23] (Note that $\widehat{\eta}_1(\theta, \mathbf{c}_2) = \eta(\theta)$.) The discounted present value of the regulator's welfare within a regulatory cycle is $\sum_{t=1}^{T} \beta^{t-1} [\Psi(c_t) - \Omega] + \psi [u_1(\theta, \mathbf{c}) + A] - (1+\lambda)A$. Given that the participation constraint must bind, $-A = u_1(\theta, \mathbf{c})$, and with (10), we can then write the regulator's problem as

$$\begin{aligned} W(\theta) &= \max_{\mathbf{c}} \sum_{t=1}^{T} \beta^{t-1} \left[\Psi(c_t) - \Omega + (1+\lambda) \left\{ \begin{array}{c} (c_t - \widehat{\eta}_t(\theta, \mathbf{c}_2))D(c_t) - F \\ -E_{\widetilde{\theta}_{t-1}} \left[I(q_t^*(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1})) | \theta, \mathbf{c}_2 \right] \end{array} \right\} \right] \\ &\quad + \beta^T E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2 \right], \end{aligned} \tag{11}$$

where $W(\theta)$ is the regulator's value function. As in the static model, once we account for the binding participation constraint, the regulator's welfare depends on λ but not the welfare weight ψ. The regulator's problem is akin to a Ramsey pricing problem, with each period in the regulatory cycle akin to a distinct "good" with different marginal costs.

[22] For $\theta = \underline{\theta}$ and $\theta = \overline{\theta}$, this expression is modified in the natural way given the corner transition probabilities in (**??**) and (**??**).

[23] The expectation $\widehat{\eta}_t(\theta, \mathbf{c}_2) \equiv E_{\widetilde{\theta}_{t-1}} \left[\eta(\widetilde{\theta}_{t-1}) | \theta, \mathbf{c}_2 \right]$ is a function $\mathbf{c}_2$ because it depends on the probability distribution for the sequence of random variables $\{\widetilde{\theta}_t\}_{t=1}^T$ induced by the network firm's optimal investment program $\left\{ q_t(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1}) \right\}_{t=1}^T$, which, in turn, depends on the access prices $\mathbf{c}_{t+1}$ in future periods of the cycle.

3.2 Characterization of the solution to the regulator's problem

The characterization of the solution to the regulator's problem depends on how $E_{\widetilde{\theta}_T}\left[W(\widetilde{\theta}_T)|\theta,\mathbf{c}_2\right]$ depends on the access charges, and that in turn depends on how they affect investment. We begin with investment.

Proposition 2 *For any $t=1,\ldots,T-1$, realization θ of $\widetilde{\theta}_{t-1}$, and values of $\mathbf{c}_{t+1}$, $\frac{\partial q_t^*(\theta,\mathbf{c}_{t+1})}{\partial c_{t+s}}<0, s=1,\ldots,T-t$. That is, when the firm's investment in period t of regulatory cycle is an interior solution, a lower price in any subsequent period of the cycle increases investment in period t.*

Proposition 2 formalizes the linkage between the access prices and investment within a regulatory cycle. In a model of price cap regulation in which the regulated firm makes a one-time investment in cost reduction, Cabral and Riordan (1989) find that decreasing the price cap increases the level of the one-time investment. As they explain, this is essentially the "Arrow effect": a monopolist has a weaker incentive than a firm operating in a perfectly competitive market to pursue a nondrastic process innovation because the monopolist restricts output (Arrow 1962). In our model, this Arrow effect operates intertemporally. By lowering the price in period $t+s$, the regulator gives the firm the prospect of a higher quantity demanded in that period. This magnifies the benefit to the network firm of reducing marginal cost in that period, which in turn increases the incentive for investment in a prior period t. This is because more investment in period t increases the likelihood that the stochastic process of marginal cost will evolve toward lower costs from period t to period $t+s$.

To highlight how the regulator could potentially use the relationship between $q_t^*(\cdot,\mathbf{c}_{t+1})$ and $\mathbf{c}_{t+1}$ to increase expected welfare, we note that a finite regulatory lag (i.e., $2\leq T<\infty$) creates both a partial alignment of interests between the regulator's welfare objective and the firm's expected profit objective and a partial conflict of interests. Regulatory lag ensures that within a regulatory cycle, the firm's within-cycle cost reduction behavior is fully aligned with the within-cycle portion of the regulator's welfare objective, i.e., everything in (11) except the continuation value $\beta^T E_{\widetilde{\theta}_T}\left[W(\widetilde{\theta}_T)|\theta,\mathbf{c}_2\right]$.[24] In other words, if all that mattered to the regulator was within-cycle discounted welfare, it could delegate investment to the firm, knowing that the program $\{q_t^*(\widetilde{\theta}_{t-1},\mathbf{c}_{t+1})\}_{t=1}^{T-1}$ the firm chooses would be what the regulator itself would choose. This is analogous to a static model of fixed-price regulation under complete information in which the regulator can delegate the choice of input mix to the firm knowing that the firm's cost function reflects a socially efficient choice of inputs conditional on output.[25]

However, the regulator also cares about costs in future regulatory cycles, while the firm, by contrast, cares only about the current cycle because at the beginning of each cycle its expected discounted profits are reset to zero. (In effect, the firm perceives itself facing a cost-plus contract *between* cycles.) The divergence between the regulator's investment goals and

[24] Formally, the alignment of interests between the firm and the regulator when it comes to reducing costs in the current regulatory cycle is reflected in absence of terms involving $\frac{\partial\widetilde{\eta}_s(\theta,c2)}{\partial c_t}$ and $\frac{\partial E_{\widetilde{\theta}_{s-1}}\left[q_s(\widetilde{\theta}_{s-1})^2|\theta,\mathbf{c}_2\right]}{\partial c_t}$ in (12). The envelope theorem applied to the firm's optimization problem embedded in $u_1(\theta,c)$ implies that they disappear when we differentiate the objective in (11) with respect to c_t.

[25] Of course, the alignment of interests is not the case when it comes to *access prices* within a cycle. The firm prefers access prices that maximize $u_1(\theta,\mathbf{c})$, while a regulator prefers access prices that maximize discounted social welfare.

the firm's manifest itself in two distinct ways. First, and most obviously, from Proposition 1, the firm does not invest in the last period of the cycle. Second, the firm's investments in all other periods $t = 1, \ldots, T-1$ of a regulatory cycle are socially inefficient because they do not take into account the benefit that productivity enhancement today has on marginal costs beyond the current cycle. Lacking direct control over investment, the regulator cannot do much about the first inefficiency, but by Proposition 2 it can potentially do something about the second inefficiency: it can reduce access prices in the current cycle to "turbocharge" the firm's investment incentives so that the firm acts "as if" it cared about the impact of its investment on marginal costs in future cycles.

The regulator's benefit from using the price to offset the firm's comparatively myopic investment behavior is reflected in the impact of $\mathbf{c}_2$ on the regulator's continuation value $E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\theta, \mathbf{c}_2]$. We can establish that the regulator can increase its future expected discounted welfare by lowering the prices in periods $t = 2, \ldots, T$ of the current regulatory cycle—and by Proposition 2—increasing the firm's investment—in periods $t = 1, \ldots, T-1$ of the current cycle.

Proposition 3 $\frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\theta, \mathbf{c}_2]}{\partial c_t} < 0$, $t = 2, \ldots, T$.

Proposition 3 then enables us to establish that in an important special case—interior investment levels—the firm has a tendency to underinvest in cost reduction in some periods prior to the final period of a regulatory cycle.

Proposition 4 *Suppose for access prices* $\mathbf{c}$ *the firm's optimal investments in all periods of the regulatory cycle except the last are interior, i.e.,* $q_t^*(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1}) < 1$ *for* $t = 1, \ldots, T-1$ *and all realizations of* $\widetilde{\theta}_{t-1}$. *There exists at least one period* t *prior to the last period and some realizations of productivity* $\widetilde{\theta}_{t-1}$ *in that period in which the regulator would prefer that the firm invests more than it actually does. That is, in addition to underinvestment in the last period of the regulatory cycle, there is also some underinvestment in earlier periods of the cycle.*

The tendency for underinvestment has a direct implication for how the regulator chooses the access pricing schedule. The first-order conditions for the schedule of access charges $\mathbf{c}^*(\theta)$ that solves the problem in (11) can be written as[26]

$$c_t^*(\theta) = \widehat{\xi}_t(\theta, \mathbf{c}_2^*(\theta)) - \frac{\lambda}{1+\lambda}\frac{D(c_t^*(\theta))}{D'(c_t^*(\theta))} - \frac{\overline{P}^*(c_t^*(\theta)) - (\mu(c_t^*(\theta)) + c_t^*(\theta))}{1+\lambda}, \tag{12}$$

where

$$\widehat{\xi}_t(\theta, \mathbf{c}_2) \equiv \widehat{\eta}_t(\theta, \mathbf{c}_2) - \beta^{T-t+1}\frac{\frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\theta, \mathbf{c}_2]}{\partial c_t}}{(1+\lambda)D'(c_t)} < \widehat{\eta}_t(\theta, \mathbf{c}_2), \tag{13}$$

is the *social marginal cost* of increasing output in period t. It is the period-t expected marginal cost $\widehat{\eta}_t(\theta, \mathbf{c}_2)$ minus a long-run "incentive adjustment" that equals the rate of change of the regulator's future discounted expected welfare with respect to a one-unit change in output in period t of the regulatory cycle.

In period 1, $\frac{\partial E_{\widetilde{\theta}_T}\left[W(\widetilde{\theta}_T)|\theta, \mathbf{c}_2\right]}{\partial c_1} = 0$, so $\widehat{\xi}_1(\theta, \mathbf{c}_2) = \widehat{\eta}_1(\theta, \mathbf{c}_2) = \eta(\theta)$ and $c_1^*(\theta) = c^0(\eta(\theta))$, i.e., the price in the first period of the regulatory cycle is the first-best price. For periods $t = 2, \ldots, T$, social marginal cost $\widehat{\xi}_t(\theta, \mathbf{c}_2)$ is less than expected marginal cost $\widehat{\eta}_t(\theta, \mathbf{c}_2)$, and the regulator thus distorts the regulated prices for incentive purposes.

[26] Details can be found in the Online Appendix.

Proposition 5 *Under an optimal pricing schedule, in all but the first period of the regulatory cycle, the regulator chooses a price schedule less than the first-best schedule corresponding to the expected marginal costs, i.e.,* $c_t^*(\theta) < c^0(\widehat{\eta}_t(\theta, \mathbf{c}_2^*(\theta)))$, $t = 2, \ldots, T$.

The regulator thus counters the firm's tendency to underinvest by committing to a more aggressive schedule of access prices than it would have if, like the firm, it made decisions solely on the basis of the current regulatory horizon. This predisposes the regulator to chose access prices in periods $t = 2, \ldots, T$ that are lower than the price $c_1^*(\theta) = c^0(\eta(\theta))$ in period 1. Indeed, when there is no depreciation—when expected marginal costs will stochastically decline over time—some access prices later in the cycle will be strictly less than the price in the initial period.

Proposition 6 *If there is no depreciation, i.e.,* $\delta = 0$, *the regulator commits to access prices in later periods of the regulatory cycle that are less than the price at the beginning of the regulatory cycle, i.e.,* $c_t^*(\theta) < c_1^*(\theta)$ *for all* $t = 2, \ldots, T$.

Summing up, there are two related but distinct investment distortions that arise in our model. First, anticipating a new regulatory review, the firm chooses no investment in the terminal period of the regulatory cycle. Second, the firm has a tendency toward underinvestment in all periods of the regulatory cycle. Regulatory lag can reduce the first distortion by decreasing the frequency of reviews. It can reduce the second distortion by making the horizon over which the firm makes its investment decisions more closely approximate the regulator's infinite horizon. The access pricing schedule cannot affect the first distortion, but it can reduce the second, since lower prices tend to increase investment in all periods but the last.

This discussion raises the question of whether access pricing and regulatory lag are complements or substitutes. Specifically, does a longer lag make it less attractive or more attractive to lower access prices to increase investment? If, as expected, a longer lag increases investment for any given schedule of access prices, a longer lag can mechanically lead to lower access prices. We thus want to remove this mechanical effect to assess complementarity or substitution. In our computational analysis below, we do so by examining the difference $\widehat{\eta}_t(\theta, \mathbf{c}_2) - \widehat{\xi}_t(\theta, \mathbf{c}_2)$ between the expected marginal cost and social marginal cost. If this difference becomes smaller (in any period t) as the regulatory lag T increases, the access prices and regulatory lag are substitutes. To the extent that the difference becomes larger, they are complements.

The access prices in our model are not, strictly speaking, price caps because our model does not give the firm flexibility to set a price less than the cap. Still, Proposition 6 hints at a rationale for why a regulator might want a price cap to decline over time (an X-factor) that goes beyond the traditional rationale grounded in both efficiency and distributional considerations. Decreases in price are an efficient response to decreases in expected marginal costs. Required decreases in price also provide a way for the regulator to force the firm to share some of its gains in productivity with consumers between regulatory reviews. Proposition 6 suggests a second efficiency rationale for an X-factor: a decrease in the price cap at some point beyond the first year of a regulatory cycle provides an additional incentive for investment beyond that provided by regulatory lag. This additional efficiency rationale suggests the use of X-factors that exceed the rate of decrease in marginal cost due to productivity growth.[27]

[27] There are other reasons why X-factors might differ from expected rates of cost reduction. For example, Brennan and Crew (2016) point out that when the regulated service faces declining demand, the X factor

3.3 The trade-off between static and dynamic efficiency and the deadweight loss under the regulatory equilibrium

As has long been understood, price cap regulation entails a trade-off between static efficiency and dynamic efficiency. The regulatory lag in our model gives rise to this traditional trade-off, but it involves an additional element due to the use of the regulated price as an incentive device

To see why, we derive an expression for the deadweight loss, which requires characterization of the first-best solution. In the first-best solution, the regulator chooses prices and investment levels contingent on realized productivity, subject to satisfying the individual rationality constraint for the firm each period, $NA + (c - \eta(\theta))D(c) - F - I(q) \geq 0$.[28] (Under this solution, it is as if there is a regulatory review each period, with the regulator, not the firm, deciding the investment level.) The first-best price is static first-best price $c^0(\eta(\theta))$ tailored to realized productivity θ observed in a period. First-best welfare can be written as

$$\begin{aligned} W^0(\theta) &= \frac{\omega^0(\eta(\theta))}{1-\beta} + \sum_{t=1}^{T} \frac{\beta^{t-1}}{1-\beta^T} \left\{ E_{\widetilde{\theta}_{t-1}} \left[\omega^0(\eta(\widetilde{\theta}_{t-1})) | \mathbf{q}^0(\theta) \right] - \omega^0(\eta(\theta)) \right\} \\ &+ \frac{1}{1-\beta^T} \left\{ \begin{array}{c} \beta^T \left[E_{\widetilde{\theta}_T} \left[W^0(\widetilde{\theta}_T) | \mathbf{q}^0(\theta) \right] - W^0(\theta) \right] \\ - \sum_{t=1}^{T} \beta^{t-1} (1+\lambda) E_{\widetilde{\theta}_{t-1}} \left[I(q^0(\widetilde{\theta}_{t-1})) | \mathbf{q}^0(\theta) \right] \end{array} \right\}. \end{aligned} \tag{14}$$

where $\mathbf{q}^0(\theta) = \left\{ q^0(\widetilde{\theta}_{t-1}) | \theta \right\}_{t=1}^{T}$ denotes the set of state-contingent first-best investment levels over a T-period horizon when the initial productivity is θ.

To derive an analogous expression for the regulator's welfare $W^T(\theta)$ under a T-period regulatory cycle, we can rewrite the regulator's objective function in (11) as

$$\begin{aligned} W^T(\theta) &= \frac{\omega^0(\eta(\theta))}{1-\beta} + \sum_{t=1}^{T} \frac{\beta^{t-1}}{1-\beta^T} \left\{ \omega(c_t^*(\theta), \widehat{\eta}_t(\theta, \mathbf{c}_2^*(\theta))) - \omega^0(\eta(\theta)) \right\} \\ &+ \frac{1}{1-\beta^T} \left\{ \begin{array}{c} \beta^T \left[E_{\widetilde{\theta}_T} \left[W^T(\widetilde{\theta}_T) | \mathbf{q}^*(\theta) \right] - W^T(\theta) \right] \\ - \sum_{t=1}^{T} \beta^{t-1} (1+\lambda) E_{\widetilde{\theta}_{t-1}} \left[I(q_t^*(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1}^*(\theta)) | \mathbf{q}^*(\theta) \right] \end{array} \right\}, \end{aligned} \tag{15}$$

where $\mathbf{q}^*(\theta) \equiv \{ q_t^*(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1}^*(\theta))^2 | \theta, \mathbf{c}_2^*(\theta) \}_{t=1}^{T}$ denotes the firm's state-contingent investment levels over the T-period regulatory cycle when initial productivity is θ.

The deadweight loss under the regulatory equilibrium is $W^0(\theta) - W^T(\theta)$. Subtracting (15) from (14), the deadweight loss $DWL^T(\theta)$ can be written recursively:

$$DWL^T(\theta) = W^0(\theta) - W^T(\theta) = SE^T(\theta) + DE^T(\theta) + \beta^T E_{\widetilde{\theta}_T} \left[DWL(\widetilde{\theta}_T) | \mathbf{q}^0(\theta) \right], \tag{16}$$

should be corrected to reflect the rate of demand decline, the rate at which average cost increases as quantity falls due to the demand decrease, and the rate at which the change in price itself further induces a change in quantity.

[28] Alternatively, the individual rationality constraint could be that the firm's expected profit over the infinite horizon is non-negative. This would only affect the profile of fixed access charges, not the variable access charge.

where

$$SE^T(\theta) = \sum_{t=1}^{T} \beta^{t-1} \left\{ E_{\widetilde{\theta}_{t-1}} \left[\omega^0(\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right] - \omega(c_t^*(\theta), E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right]) \right\} \tag{17}$$

$$DE^T(\theta) = \left[\begin{array}{c} \sum_{t=1}^{T} \beta^{t-1} \left(\begin{array}{c} E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^*(\theta) \right] \\ -E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right] \end{array} \right) D(c_t^*(\theta)) \\ \left\{ \begin{array}{c} \beta^T E_{\widetilde{\theta}_T} \left[W^T(\widetilde{\theta}_T)|\mathbf{q}^{\mathbf{0}}(\theta) \right] \\ - \sum_{t=1}^{T} \beta^{t-1}(1+\lambda) E_{\widetilde{\theta}_{t-1}} \left[I(q^0(\widetilde{\theta}_{t-1}))|\mathbf{q}^0(\theta) \right] \end{array} \right\} \\ - \left\{ \begin{array}{c} \beta^T E_{\widetilde{\theta}_T} \left[W^T(\widetilde{\theta}_T)|\theta, \mathbf{q}^*(\theta) \right] \\ - \sum_{t=1}^{T} \beta^{t-1}(1+\lambda) E_{\widetilde{\theta}_{t-1}} \left[I(q_t^*(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1}^*(\theta))|\mathbf{q}^*(\theta) \right] \end{array} \right\} \end{array} \right]. \tag{18}$$

The first component $SE^T(\theta)$ is static inefficiency within the T-period regulatory cycle. The second component $DE^T(\theta)$ is dynamic inefficiency. It consists of two parts. The first is the reduction in expected costs within the regulatory cycle when the investment schedule is $\mathbf{q^0}(\theta)$ rather than $\mathbf{q}^*(\theta)$, holding output fixed at $D(c_t^*(\theta))$ in each period. The second part holds the continuation value fixed at the level achieved in the regulatory equilibrium, $W^T(\cdot)$ and is the *net* increase in this expected continuation value (after deducting expected investment costs) when the investment schedule is $\mathbf{q^0}(\theta)$ rather than $\mathbf{q}^*(\theta)$. This portion of the deadweight loss arises because the equilibrium investment profile induced by the regulatory equilibrium will not necessarily correspond to the first-best level. The recursion in (16) implies that $DWL^T(\theta)$ is a (complicated) weighted sum of the $SE^T(\cdot)$ and $DE^T(\cdot)$ for all θ.[29]

Static inefficiency $SE^T(\theta)$ is unambiguously positive for two reasons, represented by the inequalities in the following expression.

$$\begin{aligned} E_{\widetilde{\theta}_{t-1}} \left[\omega^0(\eta(\widetilde{\theta}_{t-1}))|\mathbf{q}^0(\theta) \right] &> E_{\widetilde{\theta}_{t-1}} \left[\omega(c^0(E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right]), \eta(\widetilde{\theta}_{t-1}))|\mathbf{q}^0(\theta) \right] \\ &= \omega(c^0(E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right], E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right]) \\ &= \omega^0(E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right]) \\ &\geq \omega(c_t^*(\theta), E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right]),\ t = 2, \ldots, T. \end{aligned} \tag{19}$$

The first inequality reflects the traditional source of static inefficiency under price cap regulation: there is a welfare loss because the access prices are not tailored to actual marginal cost.[30] The second inequality arises because the optimal price is not necessarily the static optimal price when marginal cost in period t is $E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right]$. This is true for two reasons. First, the expected marginal cost in period t under the regulator's optimal pricing schedule is $\widehat{\eta}_t(\theta, \mathbf{c}_2^*(\theta)) = E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^*(\theta) \right] \neq E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right]$ because investment is

[29] The weights depend on β and $q^0(\cdot)$.

[30] Recall that, $\omega^0(\eta) = \max_c \omega(c, \eta)$, so $\omega^0(\eta(\widetilde{\theta}_{t-1})) > \omega(c^0(E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right]), \eta(\widetilde{\theta}_{t-1}))$ since $c^0(E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right])$ is a feasible but not optimal access price when $\eta(\widetilde{\theta}_{t-1}) \neq E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right]$. Taking expectations with respect to the probability distribution $\mathbf{q}^0(\theta)$ yields the first inequality.

not first best. Secondly, even if $\widehat{\eta}_t(\theta, \mathbf{c}_2^*(\theta)) = E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta)\right]$, in periods $t = 2, \ldots, T$, $c_t^*(\theta) < c^0(\widehat{\eta}_t(\theta, \mathbf{c}_2^*(\theta))$ because the price is based on social marginal cost $\widehat{\xi}_t(\theta, \mathbf{c}_2^*(\theta)$, not expected marginal cost $\widehat{\eta}_t(\theta, \mathbf{c}_2^*(\theta))$.

As noted above, the regulator creates investment incentives in two ways: increasing regulatory lag and lowering access prices. Both incentive devices reduce static efficiency. If these devices are substitutes, the regulator would rely less on distortions in the price as T increases and would rely on regulatory lag to increase investment incentives. If so, it seems possible that increases in T would not necessarily increase the static inefficiency component of the deadweight loss. If, by contrast, these two incentive devices are complements, distortions in the price would increase as T increases, and $SE^T(\theta)$ would presumably increase in T.

From (19) a lower bound on $SE^T(\theta)$ will be

$$\sum_{t=1}^{T} \beta^{t-1} \left\{ E_{\widetilde{\theta}_{t-1}} \left[\omega^0(\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right] - \omega^0(E\left[\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta)\right]) \right\}.$$

For the specification in Table 1 below this has a simple form:

$$\frac{N}{2} \frac{(N-1)b + 1 - b}{(N-1)b + 2(1-b)} \sum_{t=1}^{T} \beta^{t-1} Var_{\widetilde{\theta}_{t-1}} \left[(\eta(\widetilde{\theta}_{t-1}))|\theta, \mathbf{q}^0(\theta) \right],$$

where $Var_{\widetilde{\theta}_{t-1}}(\cdot)$ denotes the variance with respect to $\widetilde{\theta}_{t-1}$. The incremental change in static inefficiency from increasing regulatory lag from $T-1$ to T is thus proportional to $\beta^{T-1} Var_{\widetilde{\theta}_{T-1}} \left[(\eta(\widetilde{\theta}_{T-1}))|\theta, \mathbf{q}^0(\theta) \right]$. If $\eta(\theta)$ is convex and most productivity gains have been exhausted, then we would expect $Var_{\widetilde{\theta}_{T-1}} \left[(\eta(\widetilde{\theta}_{T-1}))|\theta, \mathbf{q}^0(\theta) \right]$ to be small. This suggests that there are circumstances under which increasing regulatory lag might entail only a modest loss of static efficiency.

To further study the determinants of the deadweight loss, we compute the deadweight loss and the component of the decomposition for a wide range of parameterizations. We report those results below.

3.4 Is some regulatory lag better than no lag?

The trade-off between static and dynamic efficiency raises the possibility that no regulatory lag ($T = 1$) could be better than some lag ($T > 1$). However, for the special case in which there is no depreciation, we can show that some commitment always benefits the regulator.

Proposition 7 *Suppose there is no depreciation of productivity, i.e.,* $\delta = 0$*,* $W^T(\theta) - W^1(\theta) > 0$ *for all* θ *and* $T > 1$*, i.e., some regulatory lag (*$T > 1$*) always results in higher expected social welfare than continuous regulatory reviews (*$T = 1$*).*

The key insight used to prove this result is that when it comes to investment in the first $T-1$ periods of a T-period cycle, the regulator's interests and the firm's interests are not "too far" out of alignment. As discussed above, the firm's optimal investments when faced with a schedule of access tariffs maximize discounted social welfare within the regulatory cycle. Though myopic (because the firm does not consider the benefits of its investments beyond the current cycle), the firm's behavior is still preferable to there being no investment at all, as is the case when $T = 1$.

It does not follow that a longer regulatory lag is always better for the regulator. If the marginal cost of investment is sufficiently high and there is positive depreciation, a longer lag can make the regulator worse off.

Proposition 8 *If $\delta > 0$ and the slope of the marginal investment cost function becomes arbitrarily large, i.e., $\gamma \to \infty$, then $W^T(\theta) - W^1(\theta) < 0$, i.e., a longer regulatory lag ($T > 1$) results in lower social welfare than continuous regulatory reviews ($T = 1$).*

Proposition 8 implies there is nothing inherent in the structure of regulation in our model that makes a longer regulatory lag better. But Proposition 8 is merely a possibility result. The question we ask in the next section is whether there is an interesting trade-off between shorter and longer lags in plausible economic environments. We explore that question next.

4 Computational Analysis

To further explore properties of optimal regulatory pricing schedule and the trade-offs it entails, we compute equilibria of the model.

4.1 Specification and parameter values

Table 1 presents expressions for the economic objects underlying our computations. The parameter a scales the magnitude of downstream demand, while b is the degree of horizontal differentiation among downstream firms, with $b = 0$, corresponding to the case in which the downstream firms are independent, and as $b \to 1$, the services are seen by end consumers as perfect substitutes. Marginal and fixed costs are assumed to be the same for all operating firms and equal to μ and f, respectively. With this specification, the downstream price-setting game has a symmetric Bertrand-Nash equilibrium. In the relationship between productivity and marginal cost, η_1 is the annual percentage decrease if productivity increased by one unit each year.

Table 2 shows the parameter grid $\mathcal{G}$ we use in our computations. Occasionally, we focus on a baseline parameterization in which parameters take on the bold-faced value in the table. We take a period to be a year, so the parameter values are chosen relative to that yardstick. For example, the values of η_1 imply that the maximum rate of potential cost decrease of the network firm ranges from 0.5 percent per year to 15 percent per year.

The grid $\mathcal{G}$ consists of 31,250,000 distinct parameter combinations. The parameter values are not intended to represent any particular setting, but they encompass, we believe, plausible economic environments. For example, the values of a imply that the price elasticity of market demand for access at the static welfare optimum ranges between about -0.49 to -2.65. Upper and lower bounds for some parameters, such as b and β, are implied by theory. For most other parameters, we selected values that were empirically plausible. An exception was γ. Because it is difficult to identify a plausible value of γ, we let it vary within a wide range. The values of the parameters in the last fives rows of Table 2 μ, f, F, $\underline{\theta}$, and $\overline{\theta}$ are fixed throughout, with the value of μ based on empirical estimates of cost conditions for freight railroads, as discussed in Besanko and Cui (2016).

All computations are done in MATLAB 9.2 (R2017a) using the University of Florida's supercomputer cluster. In each parameterization, we use Gauss-Jacobi iterative method to compute the MPE prices, investments, and welfare. Our calculations achieved convergence for 86.83 percent of the parameterizations in $\mathcal{G}$. The results of our computations over $\mathcal{G}$

Object	Expression
Consumer benefit function	$V(Q_1,\dots,Q_N) = a\sum_{i=1}^{N} Q_i - \frac{1}{2[1+(N-1)b]}\left\{\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}Q_iQ_j\right\}$
	$b_{ij} = 1$ for $i = j,\ i,j \in \{1,\dots,N\}$ and $b_{ij} = b_{ji} = b \in [0,1)$ for $i \neq j$.
Market demand for access	$D(c) = N\frac{(N-1)b+1-b}{(N-1)b+2(1-b)}[a-\mu-c]$
Downstream social surplus	$\Psi(c) - \Omega = \frac{1}{2N}\left[\frac{[(N-1)b+3(1-b)]}{[(N-1)b+(1-b)]}\right][D(c)]^2 - \Omega$
Static welfare function	$\omega(c,\eta) = \frac{1}{2N}\left[\frac{[(N-1)b+3(1-b)]}{[(N-1)b+(1-b)]}\right][D(c)]^2 + (1+\lambda)[(c-\eta)D(c) - F] - \Omega$
Static optimal price	$c^0(\eta) = (1-\phi)(a-\mu) + \phi\eta$
	$\phi = \frac{1+\lambda}{\left[-\frac{(N-1)b+3(1-b)}{(N-1)b+2(1-b)}+2(1+\lambda)\right]} \in (\frac{1}{2}, 2)$
Maximal static welfare	$\omega^0(\eta) = \frac{N}{2}\left[\frac{[(N-1)b+1-b]}{[(N-1)b+2(1-b)]}\right](1+\lambda)\phi(a-\mu-\eta)^2$
Marginal cost-productivity mapping	$\eta(\theta) = \eta_0(1-\eta_1)^{\theta-1},\ \eta_0 > 0, \eta_1 > 0$

Table 1: Expressions for economic objects underlying the computational analysis

Parameter	Economic interpretation	Parameter grid
a	Magnitude of potential market demand	$\{8, \mathbf{10}, 13, 15, 16\}$
b	Product differentiation parameter	$\{0.10, 0.50, \mathbf{0.80}, 0.90, 0.99\}$
N	Number of operating firms	$\{\mathbf{2}, 3, 4, 5, 6\}$
β	Discount factor	$\{0.91, 0.93, \mathbf{0.95}, 0.97, 0.98\}$
η_1	Maximum potential rate of cost reduction	$\{0.005, 0.02, \mathbf{0.05}, 0.075, 0.10.0\}$
η_0	Marginal cost network firm, with no productivity increase	$\{0.5, 1, 1.5, \mathbf{2}, 2.5\}$
γ	Slope of the marginal investment cost function	$\{1, \mathbf{3}, 20, 500, 5000\}$
λ	Marginal cost of public funds	$\{0.01, 0.10, \mathbf{0.20}, 0.30, 0.50\}$
δ	Depreciation rate	$\{0, 0.10, \mathbf{0.20}, \ldots, 0.90\}$
T	Regulatory lag	$\{1, \ldots, 8\}$
μ	Marginal cost operating firms	2.7
f	Fixed cost operating firms	2.5
F	Fixed cost network firm	5
$\underline{\theta}, \overline{\theta}$	Minimum and maximum levels of productivity	$1, 30$

Table 2: Parameter grids used in computations.

reported below exclude all parameterizations for which we had non-convergence for some T. For example, if we have convergence for a particular parameterization for $T = 1, \ldots, 7$ but not $T = 8$, we exclude this case.

4.2 Computational results

In what follows, we characterize regularities established by our numerical calculations as "Results." Results are, of course, distinguished from the propositions above established by formal arguments.

Figure 2 presents our findings for state $\theta = 15$, a state in which the network firm has achieved some productivity gains but has not fully exhausted its potential for even more.[31] We present analogous figures for states $\theta = 5, 10, 20, 25$ in the Online Appendix. The results displayed in those figures are broadly similar with those presented here.

We begin by characterizing equilibrium investment for a fixed productivity state.

Result 1 *For 100% of the parameterizations in $\mathcal{G}$, a longer regulatory lag induces more investment in equilibrium at any period within the regulatory cycle prior to the final period, i.e., $q_t^{T+1*}(\theta) > q_t^{T*}(\theta)$, $t = 1, \ldots, T-1, T = 1, \ldots 7$, for productivity state $\theta = 15$.*

Result 2 *For 100% of the parameterizations in $\mathcal{G}$, for any regulatory lag, investment decreases monotonically throughout the regulatory cycle, i.e., $q_{t-1}^{T*}(\theta) > q_t^{T*}(\theta)$, $t = 2, \ldots, T, T = 1, \ldots 8$, for productivity state $\theta = 15$.*

These results, illustrated in the two upper panels on the right hand-side of Figure 2, highlight the power of regulatory lag in motivating investment. Result 1 shows that for swath of parameter space for which we computed equilibria, a longer lag always results in more investment than a shorter lag at a given point in a regulatory cycle. Result 2 shows that as the firm's decision horizon diminishes before its next review, investment goes down (eventually reaching zero by the terminal period in a cycle, as established in Proposition 1).

The next result summarizes our findings on the impact of regulatory lag on welfare illustrated by the bottom panel the right-hand side of Figure 2.

Result 3 *For more than 97% of the parameterizations in $\mathcal{G}$, the regulator's long-run welfare was higher the longer the regulatory lag, i.e., $W^{T+1}(\theta) > W^T(\theta), T = 1. \ldots, 7$, for productivity state $\theta = 15$..*[32]

The percentage of parameterizations in which the regulator prefers a longer lag to a shorter one ranges from 0.9764 when $T = 7$ (so $W^8(\theta) > W^7(\theta)$ in 97.64 percent of the parameterizations in $\mathcal{G}$) to 0.9937 when $T = 2$. These proportions are noteworthy for two related reasons. They show, consistent with Proposition 8, that within our parameter grid the regulator might not *always* prefer some commitment to no commitment. They also indicate that the set of parameterizations in which the regulator would prefer not to commit is extremely narrow. The small sliver of cases in which the regulator prefers a shorter lag

[31] In Figure 2, the superscript T denotes the dependence of equilibrium objects on the length of regulatory lag. Further, the term $q_t^{T*}(\theta)$ denotes equilibrium investment in period t of a T-period cycle, i.e., $q_t^{T*}(\theta) = q_t^*(\theta, \mathbf{c}_{t+1}^*(\theta))$ for $t = 1, \ldots, T$.

[32] It should be noted that in slightly over one percent of parameterizations in Figure 2, $W^T(15) < 0$, indicating that the regulator would prefer, if it could, to have the first salvage its assets rather than produce. We tried to avoid this possibility through our choices of a, f, and F. However, if we exclude the cases in whcih $W^T(15) < 0$, the percentages and patterns are virtually identical to those in Figure 2.

tend to occur for extreme parameterizations, primarily when the slope γ of the marginal investment cost function is so large, or the market demand parameter a is so small, that there is practically no investment. This can be seen in Figure 3 which shows the percentage of parameterizations for which $W^{T+1}(15) > W^T(15)$ with a single parameter fixed at a particular value. (Thus, the lowest line in the top left-hand panel of Figure 3 is the percentage of equilibria for which $W^{T+1}(15) > W^T(15)$ with γ fixed at 5,000 and all other parameters varying in $\mathcal{G}$.). The results on $W^{T+1}(15) > W^T(15)$ in Figures 2 and 3 further reinforce the theme of how potent regulatory lag can be despite its adverse impact on static efficiency.

Even though a longer lag generally increases the regulator's long-run welfare, the returns to a longer lag generally diminish. We illustrate this in the upper-left panel of Figure 4, which plots $W(\theta)$ for the baseline parameterization. We see in this figure clear diminishing marginal returns to increasing regulatory lag. For example, at a low level of productivity such as $\theta = 5$, where opportunities for future cost reduction are abundant, there are meaningful gains in expected welfare in moving from no lag to a two-year lag (about 9.4 percent) or from a two-year lag to a three-year lag (about 6.2 percent). However, these gains are much smaller in moving beyond a six-year lag (e.g., the gain from $T = 6$ to $T = 7$ is 1.7 percent). At an intermediate level of productivity such as $\theta = 15$, the gains from increasing regulatory lag are more modest, about 3 percent when we move from $T = 1$ to $T = 2$, and 2.3 percent from $T = 2$ to $T = 3$. Beyond $T = 6$, the gain from increasing regulatory lag are less than one percent per additional year.

Returning to Figure 2 and moving from the bottom to the top of the left-hand panels, we have several results on the equilibrium access prices. The first pertains to the percentage decrease $\frac{c_{t-1}^{T*}(\theta)-c_t^{T*}(\theta)}{\left|c_{t-1}^{T*}\theta\right|}$ in prices.

Result 4 *For all $T = 1, \ldots, 8$, the percentage decrease $\frac{c_{t-1}^{T*}(\theta)-c_t^{T*}(\theta)}{\left|c_{t-1}^{T*}\theta\right|}$ in price is greater than the percentage decrease $\frac{\widehat{\eta}_{t-1}(\theta,\mathbf{c}_2^*(\theta))-\widehat{\eta}_t(\theta,\mathbf{c}_2^*(\theta))}{\left|\widehat{\eta}_{t-1}(\theta,\mathbf{c}_2^*(\theta))\right|}$ in expected marginal cost in more than 60% of parameterizations in productivity state $\theta = 15$.*[33] *Beyond the second period of a regulatory cycle (i.e., $t \geq 3$) the percentage decrease in the price exceeds the percentage change decrease in expected marginal cost in more than 80% of parameterizations for all $T = 1, \ldots, 8$.*

The percentage decreases in prices, $\left\{\frac{c_{t-1}^{T*}(\theta)-c_t^{T*}(\theta)}{\left|c_{t-1}^{T*}\theta\right|}\right\}_{t=2}^{T}$, along the price schedule can be thought of as analogous to a commitment to a set of time-varying X-factors. In practice, the X-factors in price cap regulation have two rationales—they more closely match prices to marginal costs that fall over time due to improved productivity (an efficiency rationale), and they share the benefits of productivity improvements with consumers (a distributional rationale). Our analysis has highlighted that the price schedule to which the regulator commits has an additional efficiency rationale: motivating the firm to internalize the benefits of investment in the current cycle for cost reduction beyond the current cycle. Result 4

[33] When the access price *increases* from one period to the next (as is possible when depreciation is sufficiently positive), $\frac{c_{t-1}^T-c_t^T}{\left|c_{t-1}^T\right|}$ will be negative. A tendency for $\frac{c_{t-1}^T-c_t^T}{\left|c_{t-1}^T\right|} > \frac{\eta_{t-1}^T-\eta_t^T}{\left|\eta_{t-1}^T\right|}$ indicates that percentage increases in access prices tend to be less than percentage increases in expected marginal cost. Our model would thus suggest that when marginal costs are expected to rise, price escalation should be less than expected cost escalation.

the bottom panel on the left-hand side of tells us that this increase tends to be less (i.e., more negative) than the increase suggests that this additional role would often call for X-factors that are larger than those based solely on traditional efficiency and distributional considerations. The incentive role of access prices is particularly large for prices later in the cycle. The price in any period affects investment in all prior periods of a regulatory cycle. Later-period prices thus play an outside role in shaping investment incentives.

Result 4 speaks only to the direction of the inequality between $\frac{c_{t-1}^{T*}(\theta)-c_t^{T*}(\theta)}{\left|c_{t-1}^{T*}\theta\right|}$ and $\frac{\widehat{\eta}_{t-1}(\theta,\mathbf{c}_2^*(\theta))-\widehat{\eta}_t(\theta,\mathbf{c}_2^*(\theta))}{\left|\widehat{\eta}_{t-1}(\theta,\mathbf{c}_2^*(\theta))\right|}$, but we find that the difference in magnitudes can be quite large. For example, when $T = 2$, for the baseline parameterization the percentage decrease in expected marginal cost between periods 1 and 2 when $\theta = 9$ is very small, about 0.02%. However, the percentage decrease in prices between periods 1 and 2 is 13.8%.

We also investigated how the incentive adjustment term $\widehat{\eta}_t^T(\theta, \mathbf{c}_2^{T*}(\theta)) - \widehat{\xi}_t^T(\theta, \mathbf{c}_2^{T*}(\theta)$ is affected by regulatory lag.

Result 5 *For any t, the incentive adjustment term $\widehat{\eta}_t^T(\theta, \mathbf{c}_2^{T*}(\theta)) - \widehat{\xi}_t^T(\theta, \mathbf{c}_2^{T*}(\theta))$ decreases as T increases in more than 85% of parameterizations in productivity state $\theta = 15$.*

This result indicates that there is a strong tendency for the distortion to access pricing to decrease as regulatory lag increases. Regulatory lag and the price schedule are thus much more likely to be substitute incentive mechanisms than complementary ones.

The remaining panel on the left-hand side of Figure 2 pertains to the access prices themselves.

Result 6 *For any given period t of the regulatory cycle, the price $c_t^{T*}(\theta)$ increases in the length T of the regulatory cycle in less than 25 percent of parameterizations when the productivity $\theta = 15$.*

This result reflects the beneficial impact regulatory lag has on investment and thus the social marginal cost on which access prices are based.

4.3 Deadweight loss decomposition

The decomposition of the deadweight loss in (16) illustrates the source of the welfare gains from a higher T. Panel 1 in Figure 5 shows the deadweight loss $DWL^T(\theta)$ in the baseline parameterization for regulatory lags $T = 1, \ldots, 10$, and panel 2 shows the corresponding continuation value $\beta^T E_{\widetilde{\theta}_T}\left[DWL(\widetilde{\theta}_T)|\theta, \mathbf{q}^0(\theta)\right]$. For this particular parameterization, the deadweight loss goes down as T increases.

As noted above, the deadweight loss for any θ will be a weighted sum of the $SE^T(\cdot)$ and $DE^T(\cdot)$ for all value of θ. Panel 3 in Figure 5 shows the static inefficiency $SE^T(\theta)$ for regulatory cycles ranging from $T = 1$ to $T = 10$. While $SE^T(\theta)$ increases in T when productivity is sufficiently low, it does not necessarily increase in T at higher levels of productivity. This is consistent with regulatory lag and access pricing being substitutes rather than complements when it comes to motivating investment.

Panel 4 in Figure 5 shows the dynamic inefficiency component of deadweight loss $DE^T(\theta)$. For $T \geq 4$, the dynamic inefficiency decreases as T increases. However, monotonicity does not hold for $T = 1, 2, 3$.

The key insight here is that while longer regulatory lags significantly reduce the deadweight loss from dynamic efficiency that can arise with a shorter lag, they entail just a modest sacrifice of static inefficiency. This is because with a shorter lag, the regulator relies heavily on reduction in the price to increase investment incentives, which, as we have seen, reduces static efficiency. With a longer lag, the regulator does not need to lean into these distortions quite as much. The longer lag by itself helps boost investment.

4.4 How important is it for the regulator to be foresighted?

In our model, the regulator is foresighted, taking into account how the price schedule in the current regulatory cycle affects decision making by itself and the firm in all future cycles.

5 Endogenizing the length of the regulatory cycle

So far we have assumed that the regulator could commit *ex ante* to the length T of the regulatory cycle. In this section, following ARV, we relax this assumption and allow the regulator to determine T at the beginning of a new cycle. The lag T thus becomes an equilibrium choice along with $\mathbf{c}$ and A.

The regulator's problem is

$$\begin{aligned} W(\theta) &= \max_{c_1,\dots c_T, T\in\mathcal{T}} \sum_{t=1}^{T} \beta^{t-1}\left[\Psi(c_t)-\Omega\right] \\ &+ (1+\lambda)u_1(\theta,c_1,\dots c_T)+\beta^T E_{\widetilde{\theta}_T}\left[W(\widetilde{\theta}_T)|\theta,c_2,\dots c_T\right], \end{aligned} \tag{20}$$

where $\mathcal{T}$ is the set of possible cycle lengths. The solution to this problem involves a regulatory cycle length $T(\theta)$ that will depend on the firm's productivity coming out of the previous regulatory cycle.

5.1 Is a fixed cycle length a Markov perfect equilibrium?

This framework allows for the possibility that the regulator chooses a cycle of a fixed length irrespective of the firm's productivity, i.e., $T(\theta) = T'$ for all θ for some particular value of T, namely T'. Because such an outcome would replicate the full-commitment case studied in the previous section, we refer to this as the *full-commitment equilibrium*.

A necessary condition for a full-commitment equilibrium of length T is that the regulator has no incentive for a "one-shot deviation" to another cycle length $\widehat{T}$ and followed by a return to cycle length T when the deviation is over. That is, if $W^T(\theta)$ is the value function corresponding to the full-commitment solution with a regulatory cycle of length T, then for any other cycle length $\widehat{T}$.

$$\begin{aligned} W^T(\theta) &\geq W^{T,\widehat{T}}(\theta) \equiv \max_{c_1,\dots c_{\widehat{T}}} \sum_{t=1}^{\widehat{T}} \beta^{t-1}\left[\Theta(c_t)-\Omega\right] \\ &+ (1+\lambda)u_1(\theta,c_1,\dots c_{\widehat{T}})+\beta^{\widehat{T}} E_{\widetilde{\theta}_{\widehat{T}}}\left[W^T(\widetilde{\theta}_{\widehat{T}})|\theta,c_2,\dots c_{\widehat{T}}\right], \end{aligned} \tag{21}$$

Figure 6 shows computations of $W^T(\theta)$ and $W^{T,\widehat{T}}(\theta)$ for the baseline parameterization and $T = 1,\dots,10$. We compute $W^{T,\widehat{T}}(\theta) - W^T(\theta)$ (shown as $W\widehat{T}T - WT$ in the figure

6).[34] For $T < 10$, the regulator has an incentive to deviate to a longer cycle for all θ, i.e., $W^{T,10}(\theta) - W^T(\theta) > 0$ for all θ.[35]

Figure 6 reveals two regularities. First, the gain from a one-shot deviation becomes smaller as the firm's productivity increases. A commitment to a regulatory lag is less vulnerable to opportunistic deviation by the regulator when potential gains from enhanced dynamic efficiency are smallest. Second, the gain from a one-shot deviation is greater for shorter regulatory cycles than for longer ones. For example, when $T = 2$ and $\theta = 10$, deviating to $T = 10$ increases the regulator's discounted welfare from 494.4577 to 539.0846 (about 9 percent). When $T = 8$ and $\theta = 10$, deviating to $T = 10$ increases the regulator's discounted welfare from 560.1224 to 563.9871 (about 0.69 percent). Thus, shorter regulatory cycles are more vulnerable to opportunistic deviation by the regulator. For regulatory lags used in practice such as $T = 4$ or $T = 5$, the gains from one-shot deviations in our baseline parameterization are relatively small for all θ.

5.2 Markov perfect equilibrium regulatory cycle

The analysis of one-shot deviations rule out the possibility that the equilibrium regulatory lag coincides with the full-commitment equilibrium for $T < 10$. However, it does not tell us what the Markov equilibrium policy $T(\theta)$ is.

To determine the equilibrium, we solve the regulator's optimization in (20) assuming regulator has a choice among ten different regulatory lags. We use an iterative algorithm to solve this problem computationally. At the beginning of an iteration, we conjecture a single value function $\widehat{W}(\theta)$. Along with that we conjecture sets of investment and access pricing policy functions for each T. We use the common value function and the T-specific conjectures of the investment and pricing policy functions to compute an expectation $E_{\widetilde{\theta}_T}\left[\widehat{W}(\widetilde{\theta}_T)|\theta, c_2, \ldots c_T\right]$ for each candidate $T \in \{1, \ldots, T\}$. This, in turn, enables us to compute equilibrium access prices, investments, and firm discounted profits for each possible T, which then become our conjectures for the next iteration. This step also implies value function candidates, $\{W^1(\theta), W^2(\theta), ..., W^{10}(\theta)\}$. Our conjectured value function for the next iteration is $\max_{T\in\{1,...,T\}}\{W^T(\theta)\}$, which then becomes the basis a revised expectation on the right-hand side of (20). If this process converges, the equilibrium lag $T^*(\theta) = \arg\max_{T\in\{1,...,T\}}\{W^T(\theta)\}$, and the regulator's equilibrium welfare $W(\theta) = W^{T^*}(\theta)$. By construction, $T^*(\theta)$ and $W^{T^*}(\theta)$ satisfy (20) and constitute a Markov perfect equilibrium.

When we compute the equilibrium for the baseline parameterization, we find $T^*(\theta) = 10$. Figure 7 shows the regulator's value functions for each of the candidate regulatory lag. The figure illustrates that if the regulator "believes" that $T^*(\theta) = 10$, it will in fact have an incentive to select $T = 10$.

It is well understood by economists and policy scholars that regulatory lag is a valuable incentive tool when it comes to motivating regulatory firms to make efficient investment decisions. We see the analysis in this section as showing just how powerful regulatory lag can be. Recall that, in principle, the regulator could motivate the firm to invest with a short regulatory lag and access prices that sharply decline beyond the first period, and in doing so, it preserves some degree of static efficiency. The fact that the regulator, even when it cannot commit to a lag *ex ante*, will nevertheless choose the longest feasible regulatory lag shows that adjusting the access pricing schedule (or, equivalently, adjusting the X-factors)

[34] Note that $W10\# - W\#$ refers to $W^{T,\widehat{T}}(\theta) - W^T(\theta)$ when $T = \#$ and $\widehat{T} = 10$.

[35] Note that for $T = 10$, $W^{T,\widehat{T}}(\theta) - W^T(\theta) < 0$ for all θ and $\widehat{T} < 10$.

goes only so far in promoting dynamic efficiency. Ultimately, the shorter regulatory lag has a serious bug: in the last period of every cycle, firm will behave opportunistically and not invest in anticipation of the next review. With a shorter lag, this opportunistic behavior recurs more frequently, which takes its toll on expected social welfare.

6 Summary and conclusions

This paper studied a stochastic dynamic game of regulatory price setting. In our model, a regulated firm makes investments each period that stochastically improves its productivity and reduces its marginal costs. The infinite horizon over which the firm and the regulator interact is divided into regulatory cycles that last for T-periods. At the beginning of a regulatory cycle, the regulator—whose objective is to maximize expected welfare, subject to satisfying the network firm's participation constraint—commits to a T-period schedule of access charges in each period of the cycle. It cannot, however, commit to subsequent schedules for future regulatory cycles.

It is not a general property of our model that a longer regulatory lag is always better for a welfare-maximizing regulator. Our computational analysis illustrates that a longer lag can decrease expected welfare. For parameterizations that predispose investment to be low (e.g., when the marginal investment cost function is steeply sloped or when market size is small), the loss in static efficiency from a longer lag can exceed the gains in dynamic efficiency from a longer lag. This occurs not only because of the traditional source of static inefficiency with regulatory lag—the regulator not tailoring prices to within-cycle changes in the firm's productivity—but also because the regulator uses prices as incentive device to induce more investment in productivity enhancement.

Still, we find that a longer lag is better for expected welfare than a shorter lag for the overwhelming number of parameterizations for which we computed equilibria. This comes, though, with an important caveat: we find strong diminishing returns to longer lags. Most of the additional benefit of a longer cycle is captured with regulatory lags of lengths no greater than the longest lags used in practice, five to six years.

A key force in our model is that reducing the regulated price in any given period of the regulatory cycle beyond the first period increases the marginal product of investment in any prior period through an intertemporal Arrow effect. The prospect of a higher quantity demanded in period $t + s$ of a regulatory cycle magnifies the benefits of reducing marginal cost in that period, which in turn increases the incentive for investment in a prior period t because such investment increases the likelihood that the stochastic process of marginal cost will evolve toward lower costs from period t to period $t + s$. But the regulator does not use the price to motivate more investment to reduce expected marginal costs *within* the regulatory cycle. This is because within a cycle the regulator and the firm's incentives for cost reduction are aligned. Instead, the regulator uses the price schedule to offset the firm's tendency to ignore the impact of current-cycle investment on marginal costs in future cycles. With the firm's discounted expected profit reset to satisfy its participation constraint at the beginning of each period, the firm's perceives no benefits in future cycles from enhancing its productivity within the current cycle. The regulator, however, does perceive a future benefit from productivity increases within the current cycle. By distorting the price schedule downward, the regulator "juices" the firm's investment incentives within the current cycle so that it acts "as if" it did care about the impact of those investments on future productivity. This underscores an important policy implication of this paper: in setting prices or caps on prices, regulators should think of their impact beyond the current regulatory cycle. Our analysis also suggests the use of X-factors in price cap regulation that are greater than expected decreases in costs based on anticipated increases in productivity. This latter conclusion is reinforced by our computational analysis finding that the percentage decrease in access prices within a regulatory cycle often exceeds the percentage decrease in expected marginal costs.

In general, lower access charges and regulatory lag could either be substitutes or complements. Our computational analysis illustrates that they are substitutes far more often than complements, i.e., a longer lag leads to less distortion in the price relative to the case in which the regulator does not use the price as an incentive device. This makes sense. With a longer lag, the firm's investment horizon expands, and its investment incentives are more aligned with the regulator's. The regulator's need to distort prices downward is accordingly attenuated.

There are number of extensions of the model presented here. For example, we have assumed that the network firm's outside option is invariant to its productivity level. While this seems natural in our context, where the firm's assets are highly specialized and have very limited redeployment value, it would be straightforward to extend our model so that the outside option depends on productivity. This would mitigate, somewhat, the firm's drop-off in investment in the last period of the regulatory cycle because the firm would now have a reason to care about the impact of its investment in that period on its future profitability. This would, one might conjecture, make it less necessary for the regulator to distort the access charges for incentive purposes. Another extension would be to introduce the possibility of large changes in economic circumstances that could exacerbate the static inefficiency from commitment to a price schedule. Such a possibility reduces the power of regulatory lag as an incentive device. Given that we find strong diminishing returns to regulatory lag when there is no possibility of drastic changes in economic fundamentals, we conjecture that the possibility of potentially large and hard-to-predict changes would result in optimal regulatory lags that are consistent with those we observe in practice, three to five years. Finally, investment in cost reduction is not the only nonverifiable investment a firm could make. A network infrastructure firm could also invest in network quality, as in Besanko and Cui (2016). If network quality reduces the marginal costs of operating firms or directly increases the demand of end consumers for the downstream service, then more investment in quality by the network firm would increase the demand for access to the bottleneck infrastructure. Given this, the marginal benefit of the network firm's investment in quality would depend partly on the margin between the network firm's price and its marginal cost. Unlike the cost-reducing investment featured here, a higher price might stimulate more investment in quality. Distortions to access pricing by the regulator to stimulate both cost-reducing investment and investment in network quality may offset each other. As a consequence, the optimal price schedule may result in changes in price during the regulatory cycle that more closely track changes in expected marginal costs than is the case in the model in this paper.

7 Appendix

Proof of Proposition 2:
Preliminaries: Let

$$z_t^*(\theta, \mathbf{c}_{t+1}) \equiv \frac{q_t^*(\theta, \mathbf{c}_{t+1})}{1 - q_t^*(\theta, \mathbf{c}_{t+1})} = \frac{\beta}{\gamma} \left[(1-\delta)\Delta u_{t+1}(\theta, \mathbf{c}_{t+1}) + \delta \Delta u_{t+1}(\theta - 1, \mathbf{c}_{t+1}) \right] \tag{22}$$

Because the function $z(q) = \frac{q}{1-q}$ is strictly increasing in q, to prove that $\frac{\partial q_t^*(\theta,\mathbf{c}_{t+1})}{\partial c_{t+s}} < 0, s = 1,\ldots,T-t$, it suffices to prove that $\frac{\partial z_t^*(\theta,\mathbf{c}_{t+1})}{\partial c_{t+s}} < 0, s = 1,\ldots,T-t$

Part 1: Derivation of $\frac{\partial z_t^*(\theta,\mathbf{c}_{t+1})}{\partial c_{t+s}}$**:**

We begin by establishing that for any $t = 1,\ldots,T-1$, realization θ of $\widetilde{\theta}_{t-1}$, and values of $\mathbf{c}_{t+1}$,

$$\begin{aligned} \frac{\partial z_t^*(\theta,\mathbf{c}_{t+1})}{\partial c_{t+s}} &= \frac{\beta^s}{\gamma}\left\{ \begin{array}{l} (1-\delta)\Delta E_{\widetilde{\theta}_{t+s-2}}\left[\eta(\widetilde{\theta}_{t+s-2})|\theta,\mathbf{c}_{t+2}\right] \\ +\delta\Delta E_{\widetilde{\theta}_{t+s-2}}\left[\eta(\widetilde{\theta}_{t+s-2})|\theta-1,\mathbf{c}_{t+2}\right] \end{array}\right\} D'(c_{t+s}) \leq 0 \\ s &= 1,\ldots,T-t, \end{aligned} \tag{23}$$

where

$$\begin{aligned} \Delta E_{\widetilde{\theta}_{t+s-2}}\left[\eta(\widetilde{\theta}_{t+s-2})|\theta,\mathbf{c}_{t+2}\right] &\equiv E_{\widetilde{\theta}_{t+s-2}}\left[\eta(\widetilde{\theta}_{t+s-2})|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right] \\ &\quad -E_{\widetilde{\theta}_{t+s-2}}\left[\eta(\widetilde{\theta}_{t+s-2})|\widetilde{\theta}_{t-1}=\theta+1,\mathbf{c}_{t+2}\right]. \end{aligned}$$

First consider $s = 1$. Using (22 we have

$$\frac{\partial z_t^*(\theta,\mathbf{c}_{t+1})}{\partial c_{t+1}} = \frac{\beta}{\gamma}\left[(1-\delta)\frac{\partial\Delta u_{t+1}(\theta,\mathbf{c}_{t+1})}{\partial c_{t+1}} + \delta\Delta\frac{\partial\Delta u_{t+1}(\theta-1,\mathbf{c}_{t+1})}{\partial c_{t+1}}\right].$$

To evaluate this term, we begin with the version of (6) for period $t+1$:

$$\begin{aligned} u_{t+1}(\theta,\mathbf{c}_{t+1}) &= \max_{q_{t+1}\in[0,1]} (c_{t+1}-\eta(\theta))D(c_{t+1}) - F - I(q_{t+1}) + \beta u_{t+2}(\theta,\mathbf{c}_{t+2}) \\ &\quad +\beta\left\{(1-\delta)q_{t+1}\Delta u_{t+2}(\theta,\mathbf{c}_{t+2}) - \delta(1-q_{t+1})\Delta u_{t+1}(\theta-1,\mathbf{c}_{t+2})\right\} \end{aligned} \tag{24}$$

Applying the envelope theorem to (24)

$$\frac{\partial u_{t+1}(\theta,\mathbf{c}_{t+1})}{\partial c_{t+1}} = (c_{t+1}-\eta(\theta))D'(c_{t+1}) + D(c_{t+1})$$

Thus

$$\frac{\partial\Delta u_{t+1}(\theta,\mathbf{c}_{t+1})}{\partial c_{t+1}} = \frac{\partial u_{t+1}(\theta+1,\mathbf{c}_{t+1})}{\partial c_{t+1}} - \frac{\partial u_{t+1}(\theta,\mathbf{c}_{t+1})}{\partial c_{t+1}} = \Delta\eta(\theta)D'(c_{t+1}) < 0,$$

Similarly, $\frac{\partial\Delta u_{t+1}(\theta-1,\mathbf{c}_{t+1})}{\partial c_{t+1}} = \Delta\eta(\theta-1)D'(c_{t+1}) < 0$. Thus,

$$\frac{\partial z_t^*(\theta,\mathbf{c}_{t+1})}{\partial c_{t+1}} = \frac{\beta}{\gamma}\left\{(1-\delta)\Delta\eta(\theta) + \delta\Delta\eta(\theta-1)\right\}D'(c_{t+1}) < 0. \tag{25}$$

Now, consider $s = 2$, i.e., $\frac{\partial z_t^*(\theta,\mathbf{c}_{t+1})}{\partial c_{t+2}}$.[36] Using (22),

$$\frac{\partial z_t^*(\theta,\mathbf{c}_{t+1})}{\partial c_{t+2}} = \frac{\beta}{\gamma}\left[(1-\delta)\frac{\partial\Delta u_{t+1}(\theta,\mathbf{c}_{t+1})}{\partial c_{t+2}} + \delta\frac{\partial\Delta u_{t+1}(\theta-1,\mathbf{c}_{t+1})}{\partial c_{t+2}}\right]D'(c_{t+2}) \tag{26}$$

[36] This is provided $t+2 \leq T$, or $t \leq T-2$.

To evaluate this, we need to determine the expressions for $\frac{\partial \Delta u_{t+1}(\theta, \mathbf{c}_{t+1})}{\partial c_{t+2}}$ and $\frac{\partial \Delta u_{t+1}(\theta-1, \mathbf{c}_{t+1})}{\partial c_{t+2}}$. Applying the envelope theorem to (24) and rearranging terms in the resulting expression gives us

$$\frac{\partial u_{t+1}(\theta, \mathbf{c}_{t+1})}{\partial c_{t+2}} = \beta \left\{ \begin{array}{c} (1-\delta) q_{t+1}^* \frac{\partial u_{t+2}(\theta+1, \mathbf{c}_{t+2})}{\partial c_{t+2}} + \\ \left[1 - (1-\delta) q_{t+1}^* - \delta(1 - q_{t+1}^*)\right] \frac{\partial u_{t+2}(\theta, \mathbf{c}_{t+2})}{\partial c_{t+2}} \\ + \delta(1 - q_{t+1}^*) \frac{\partial u_{t+2}(\theta-1, \mathbf{c}_{t+2})}{\partial c_{t+2}} \end{array} \right\},$$

where it is understood that $q_{t+1}^* = q_{t+1}^*(\theta, \mathbf{c}_{t+2})$. Using the logic of our analysis above, $\frac{\partial u_{t+2}(\theta, \mathbf{c}_{t+2})}{\partial c_{t+2}} = (c_{t+2} - \eta(\theta)) D'(c_{t+2}) + D(c_{t+2})$, so

$$\begin{aligned} \frac{\partial u_{t+1}(\theta, \mathbf{c}_{t+1})}{\partial c_{t+2}} &= \beta \left\{ \begin{array}{c} (1-\delta) q_{t+1}^* \left[(c_{t+2} - \eta(\theta+1)) D'(c_{t+2}) + D(c_{t+2})\right] + \\ \left[1 - (1-\delta) q_{t+1}^* - \delta(1 - q_{t+1}^*)\right] \left[(c_{t+2} - \eta(\theta)) D'(c_{t+2}) + D(c_{t+2})\right] \\ + \delta(1 - q_{t+1}^*) \left[(c_{t+2} - \eta(\theta-1)) D'(c_{t+2}) + D(c_{t+2})\right] \end{array} \right\} \\ &= \beta \left[(c_{t+2} - E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \widetilde{\theta}_{t-1} = \theta, \mathbf{c}_{t+2} \right] D'(c_{t+2}) + D(c_{t+2}) \right]. \end{aligned}$$

where,

$$E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \widetilde{\theta}_{t-1} = \theta, \mathbf{c}_{t+2} \right] = \left\{ \begin{array}{c} (1-\delta) q_{t+1}^* \eta(\theta+1) \\ + \left[1 - (1-\delta) q_{t+1}^* - \delta(1 - q_{t+1}^*)\right] \eta(\theta) \\ + \delta(1 - q_{t+1}^*) \eta(\theta - 1) \end{array} \right\}.$$

Similarly,

$$\begin{aligned} \frac{\partial u_{t+1}(\theta+1, \mathbf{c}_{t+1})}{\partial c_{t+2}} &= \beta \left[(c_{t+2} - E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \widetilde{\theta}_{t-1} = \theta+1, \mathbf{c}_{t+2} \right] D'(c_{t+2}) + D(c_{t+2}) \right], \\ \frac{\partial u_{t+1}(\theta-1, \mathbf{c}_{t+1})}{\partial c_{t+2}} &= \beta \left[(c_{t+2} - E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \widetilde{\theta}_{t-1} = \theta-1, \mathbf{c}_{t+2} \right] D'(c_{t+2}) + D(c_{t+2}) \right] \end{aligned}$$

Thus

$$\begin{aligned} \frac{\partial \Delta u_{t+1}(\theta, \mathbf{c}_{t+1})}{\partial c_{t+2}} &= \frac{\partial u_{t+1}(\theta+1, \mathbf{c}_{t+1})}{\partial c_{t+2}} - \frac{\partial u_{t+1}(\theta, \mathbf{c}_{t+1})}{\partial c_{t+2}} \\ &= \beta \Delta E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \theta, \mathbf{c}_{t+2} \right] D'(c_{t+2}), \end{aligned}$$

where

$$\Delta E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \theta, \mathbf{c}_{t+2} \right] = E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \widetilde{\theta}_{t-1} = \theta, \mathbf{c}_{t+2} \right] - E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \widetilde{\theta}_{t-1} = \theta+1, \mathbf{c}_{t+2} \right].$$

Likewise

$$\frac{\partial \Delta u_{t+1}(\theta-1, \mathbf{c}_{t+1})}{\partial c_{t+2}} = \beta \Delta E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \theta-1, \mathbf{c}_{t+2} \right] D'(c_{t+2}).$$

Putting these pieces together implies

$$\frac{\partial z_t^*(\theta, \mathbf{c}_{t+1})}{\partial c_{t+2}} = \frac{\beta^2}{\gamma} \left\{ (1-\delta) \Delta E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \theta, \mathbf{c}_{t+2} \right] + \delta \Delta E_{\widetilde{\theta}_t} \left[\eta(\widetilde{\theta}_t) | \theta-1, \mathbf{c}_{t+2} \right] \right\} D'(c_{t+2}). \quad (27)$$

for all $t \in 1, 2, \ldots, T-1$.

Consider, next, $s = 3$, i.e., $\frac{\partial z_t^*(\theta, \mathbf{c}_{t+1})}{\partial c_{t+3}}$.[37] Using (22)

$$\frac{\partial z_t^*(\theta, \mathbf{c}_{t+1})}{\partial c_{t+3}} = \frac{\beta}{\gamma}\left[(1-\delta)\frac{\partial \Delta u_{t+1}(\theta, \mathbf{c}_{t+1})}{\partial c_{t+3}} + \delta \frac{\partial \Delta u_{t+1}(\theta - 1, \mathbf{c}_{t+1})}{\partial c_{t+3}}\right]. \tag{28}$$

To evaluate this, we need to determine the expressions for $\frac{\partial \Delta u_{t+1}(\theta, \mathbf{c}_{t+1})}{\partial c_{t+3}}$ and $\frac{\partial \Delta u_{t+1}(\theta-1, \mathbf{c}_{t+1})}{\partial c_{t+3}}$. Applying the envelope theorem to (24)

$$\frac{\partial u_{t+1}(\theta, \mathbf{c}_{t+1})}{\partial c_{t+3}} = \beta \left\{ \begin{array}{c} (1-\delta) q_{t+1}^* \frac{\partial u_{t+2}(\theta+1, \mathbf{c}_{t+2})}{\partial c_{t+3}} \\ + \left[1 - (1-\delta) q_{t+1}^* - \delta(1 - q_{t+1}^*)\right] \frac{\partial u_{t+2}(\theta, \mathbf{c}_{t+2})}{\partial c_{t+3}} \\ + \delta(1 - q_{t+1}^*) \frac{\partial u_{t+2}(\theta-1, \mathbf{c}_{t+2})}{\partial c_{t+3}} \end{array} \right\}. \tag{29}$$

where it is understood, as above, that $q_{t+1}^* = q_{t+1}^*(\theta, \mathbf{c}_{t+2})$. To determine $\frac{\partial u_{t+2}(\theta, \mathbf{c}_{t+2})}{\partial c_{t+3}}$, use (6), we can write the expression for $u_{t+2}(\theta, \mathbf{c}_{t+2})$ as

$$\begin{aligned} u_{t+2}(\theta, \mathbf{c}_{t+2}) &= \max_{q_{t+2} \in [0,1]} (c_{t+2} - \eta(\theta)) D(c_{t+1}) - F - I(q_{t+2}) + \beta u_{t+3}(\theta, \mathbf{c}_{t+3}) \\ &\quad + \beta \left\{(1-\delta) q_{t+2} \Delta u_{t+3}(\theta, \mathbf{c}_{t+3}) - \delta(1 - q_{t+2}) \Delta u_{t+3}(\theta - 1, \mathbf{c}_{t+3})\right\}. \end{aligned} \tag{30}$$

Applying the envelope theorem to (30) and rearranging terms gives us

$$\frac{\partial u_{t+2}(\theta, \mathbf{c}_{t+2})}{\partial c_{t+3}} = \beta \left\{ \begin{array}{c} (1-\delta) q_{t+2}^* \frac{\partial u_{t+3}(\theta+1, \mathbf{c}_{t+3})}{\partial c_{t+3}} \\ + \left[1 - (1-\delta) q_{t+2}^* - \delta(1 - q_{t+2}^*)\right] \frac{\partial u_{t+3}(\theta, \mathbf{c}_{t+3})}{\partial c_{t+3}} \\ + \delta(1 - q_{t+2}^*) \frac{\partial u_{t+3}(\theta-1, \mathbf{c}_{t+3})}{\partial c_{t+3}} \end{array} \right\}, \tag{31}$$

where it is understood that $q_{t+2}^* = q_{t+2}(\theta, \mathbf{c}_{t+3})$. Now, from (6) applied to $t + 3$

$$\frac{\partial u_{t+3}(\theta, \mathbf{c}_{t+3})}{\partial c_{t+3}} = (c_{t+3} - \eta(\theta)) D'(c_{t+3}) + D(c_{t+3}).$$

Substituting this expression (and the corresponding ones for $\frac{\partial u_{t+3}(\theta+1, \mathbf{c}_{t+3})}{\partial c_{t+3}}$ and $\frac{\partial u_{t+3}(\theta-1, \mathbf{c}_{t+3})}{\partial c_{t+3}}$) into (31) gives us

$$\frac{\partial u_{t+2}(\theta, \mathbf{c}_{t+2})}{\partial c_{t+3}} = \beta \left[(c_{t+3} - E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1}) | \widetilde{\theta}_t = \theta, \mathbf{c}_{t+3}\right]) D'(c_{t+3}) + D(c_{t+3})\right]. \tag{32}$$

and therefore

$$\frac{\partial \Delta u_{t+2}(\theta, \mathbf{c}_{t+2})}{\partial c_{t+3}} = \beta \Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1}) | \theta, \mathbf{c}_{t+3}\right] D'(c_{t+3}) \tag{33}$$

$$\frac{\partial \Delta u_{t+2}(\theta - 1, \mathbf{c}_{t+2})}{\partial c_{t+3}} = \beta \Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1}) | \theta - 1, \mathbf{c}_{t+3}\right] D'(c_{t+3}), \tag{34}$$

where

$$\Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1}) | \theta, \mathbf{c}_{t+3}\right] = \left\{ \begin{array}{c} E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1}) | \widetilde{\theta}_t = \theta, \mathbf{c}_{t+3}\right] \\ -E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1}) | \widetilde{\theta}_t = \theta + 1, \mathbf{c}_{t+3}\right] \end{array} \right\}$$

$$\Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1}) | \theta - 1, \mathbf{c}_{t+3}\right] = \left\{ \begin{array}{c} E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1}) | \widetilde{\theta}_t = \theta - 1, \mathbf{c}_{t+3}\right] \\ -E_{\widetilde{\theta}_{t+2}}\left[\eta(\widetilde{\theta}_{t+1}) | \widetilde{\theta}_t = \theta, \mathbf{c}_{t+3}\right] \end{array} \right\}$$

[37] This is provided $t + 3 \leq T$, or $t \leq T - 3$.

Substituting (32),(33), and (34) into (29) and simplifying gives us

$$\frac{\partial u_{t+1}(\theta,\mathbf{c}_{t+1})}{\partial c_{t+3}}=\beta^2\left[\begin{array}{c}(c_{t+3}-\left\{\begin{array}{c}(1-\delta)q^*_{t+1}E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right]\\+\left[\begin{array}{c}1-(1-\delta)q^*_{t+1}\\-\delta(1-q_{t+1})\end{array}\right]E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right]\\+\delta(1-q^*_{t+1})E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta-1,\mathbf{c}_{t+3}\right]\end{array}\right\})D'(c_{t+3})\\+D(c_{t+3})\end{array}\right], \tag{35}$$

where again, it is understood that $q^*_{t+1}=q^*_{t+1}(\theta,\mathbf{c}_{t+2})$. Now, note that

$$\left\{\begin{array}{c}(1-\delta)q^*_{t+1}E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right]\\+\left[\begin{array}{c}1-(1-\delta)q^*_{t+1}\\-\delta(1-q_{t+1})\end{array}\right]E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right]\\+\delta(1-q^*_{t+1})E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta-1,\mathbf{c}_{t+3}\right]\end{array}\right\}=E_{\widetilde{\theta}_t}\left[E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t,\mathbf{c}_{t+3}\right]|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right] \tag{36}$$

Substituting (36) into (35) gives us

$$\frac{\partial u_{t+1}(\theta,\mathbf{c}_{t+1})}{\partial c_{t+3}}=\beta^2\left[\begin{array}{c}(c_{t+3}-\left\{E_{\widetilde{\theta}_t}\left[E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t,\mathbf{c}_{t+3}\right]|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right]\right\})D'(c_{t+3})\\+D(c_{t+3})\end{array}\right],$$

By the general properties of conditional expectations

$$E_{\widetilde{\theta}_t}\left[E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t,\mathbf{c}_{t+3}\right]|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right]=E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right]. \tag{37}$$

Thus

$$\frac{\partial u_{t+1}(\theta,\mathbf{c}_{t+1})}{\partial c_{t+3}}=\beta^2\left[(c_{t+3}-E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right])D'(c_{t+3})+D(c_{t+3})\right], \tag{38}$$

It follows from (38)

$$\frac{\partial\Delta u_{t+1}(\theta,\mathbf{c}_{t+1})}{\partial c_{t+3}}=\beta^2\Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\theta,\mathbf{c}_{t+2}\right]D'(c_{t+3}) \tag{39}$$

$$\frac{\partial\Delta u_{t+1}(\theta-1,\mathbf{c}_{t+1})}{\partial c_{t+3}}=\beta^2\Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\theta-1,\mathbf{c}_{t+2}\right]D'(c_{t+3}) \tag{40}$$

where

$$\Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\theta,\mathbf{c}_{t+2}\right]\equiv\left\{\begin{array}{c}E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right]\\-E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta+1,\mathbf{c}_{t+2}\right]\end{array}\right\}.$$

$$\Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\theta-1,\mathbf{c}_{t+2}\right]\equiv\left\{\begin{array}{c}E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta-1,\mathbf{c}_{t+2}\right]\\-E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right]\end{array}\right\}.$$

Substituting (39) and (40) into (28) gives us

$$\frac{\partial z^*_t(\theta,\mathbf{c}_{t+1})}{\partial c_{t+3}}=\frac{\beta^3}{\gamma}\left\{(1-\delta)\Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\theta,\mathbf{c}_{t+2}\right]+\delta\Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\theta-1,\mathbf{c}_{t+2}\right]\right\}D'(c_{t+3}). \tag{41}$$

for all $t \in 1, 2, \ldots, T-1$.

Repeating the logic applied above to the cases of $c_{t+4}, \ldots, c_T$, yields the expression (23).

Part 2: Proof that $\frac{\partial z_t^*(\theta, \mathbf{c}_{t+1})}{\partial c_{t+1}} < 0$: This is established in (25).

Part 3: Proof that $\frac{\partial z_t^*(\theta, \mathbf{c}_{t+1})}{\partial c_{t+s}} < 0, s = 2, \ldots T-t$:

In light of (23), it suffices to show that $\Delta E_{\widetilde{\theta}_{t+s-2}}\left[\eta(\widetilde{\theta}_{t+s-2})|\theta, \mathbf{c}_{t+2}\right] > 0$ (since it would immediately follow that $\Delta E_{\widetilde{\theta}_{t+s-2}}\left[\eta(\widetilde{\theta}_{t+s-2})|\theta-1, \mathbf{c}_{t+2}\right] > 0$ as well). Let us first consider $\frac{\partial q_t^*(\theta, \mathbf{c}_{t+1})}{\partial c_{t+2}}$ (i.e., $s = 2$) and note that

$$\begin{aligned} E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1} = \theta, \mathbf{c}_{t+2}\right] &= \eta(\theta) - (1-\delta)q_{t+1}^*(\theta, \mathbf{c}_{t+2})\Delta\eta(\theta) + \delta(1 - q_{t+1}^*(\theta, \mathbf{c}_{t+2}))\Delta\eta(\theta - 1) \\ &> \eta(\theta) - (1-\delta)\Delta\eta(\theta) \\ &= \delta\eta(\theta) + (1-\delta)\eta(\theta+1), \end{aligned}$$

where the inequality follows because $E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1} = \theta, \mathbf{c}_{t+2}\right]$ decreases in $q_{t+1}^*(\theta, \mathbf{c}_{t+2})$ and thus attains its lowest value when $q_{t+1}^*(\theta, \mathbf{c}_{t+2}) = 1$. Also note that

$$\begin{aligned} E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1} = \theta+1, \mathbf{c}_{t+2}\right] &= \eta(\theta+1) - (1-\delta)q_{t+1}^*(\theta+1, \mathbf{c}_{t+2})\Delta\eta(\theta+1) + \delta(1 - q_{t+1}^*(\theta, \mathbf{c}_{t+2}))\Delta\eta(\theta) \\ &< \eta(\theta+1) + \delta\Delta\eta(\theta) \\ &= \delta\eta(\theta) + (1-\delta)\eta(\theta+1), \end{aligned}$$

where the inequality follows because $E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1} = \theta+1, \mathbf{c}_{t+2}\right]$ decreases in $q_{t+1}^*(\theta+1, \mathbf{c}_{t+2})$ and thus attains its highest value when $q_{t+1}^*(\theta+1, \mathbf{c}_{t+2}) = 0$. These chains of inequalities imply $\Delta E_{\widetilde{\theta}_{t+s-2}}\left[\eta(\widetilde{\theta}_{t+s-2})|\theta, \mathbf{c}_{t+2}\right] = E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1} = \theta, \mathbf{c}_{t+2}\right] - E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1} = \theta+1, \mathbf{c}_{t+2}\right] > 0$.

Next consider $\frac{\partial z_t^*(\theta, \mathbf{c}_{t+1})}{\partial c_{t+3}}$ (i.e., $s = 3$). To show $\frac{\partial z_t^*(\theta, \mathbf{c}_{t+1})}{\partial c_{t+3}} < 0$, it suffices to show $\Delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\theta, \mathbf{c}_{t+2}\right] = E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1} = \theta, \mathbf{c}_{t+2}\right] - E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1} = \theta, \mathbf{c}_{t+2}\right] > 0$.

Note that by the properties of conditional expectations,

$$
\begin{aligned}
E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right] &= E_{\widetilde{\theta}_t}\left[E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t,\mathbf{c}_{t+3}\right]|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right] \\
&= E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right] \\
&\quad + \left\{\begin{array}{c} -(1-\delta)q^*_{t+1}(\theta,\mathbf{c}_{t+2}) \\ \times\left\{\begin{array}{c} E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right] \\ -E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right] \end{array}\right\} \\ +\delta(1-q^*_{t+1}(\theta,\mathbf{c}_{t+2})) \\ \times\left\{\begin{array}{c} E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta-1,\mathbf{c}_{t+3}\right] \\ -E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right] \end{array}\right\} \end{array}\right\} \\
&> E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right] \qquad (42) \\
&\quad -(1-\delta)\left\{\begin{array}{c} E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right] \\ -E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right] \end{array}\right\} \\
&= \delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right] \qquad (43) \\
&\quad +(1-\delta)E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right],
\end{aligned}
$$

where the inequality sign in (42) arises for the following reason; (i) $E_{\widetilde{\theta}_t}\left[E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t,\mathbf{c}_{t+3}\right]|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right]$ decreases in $q^*_{t+1}(\theta,\mathbf{c}_{t+2})$. This is a consequence of having established, just above, that $E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right]-E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1}=\theta+1,\mathbf{c}_{t+2}\right]>0$ and $E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1}=\theta-1,\mathbf{c}_{t+2}\right]-E_{\widetilde{\theta}_t}\left[\eta(\widetilde{\theta}_t)|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right]>0$ (and thus, replacing $t+1$ with t, $E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right]-E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right]>0$ and $E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta-1,\mathbf{c}_{t+3}\right]-E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right]$); (ii) consequently, $E_{\widetilde{\theta}_t}\left[E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t,\mathbf{c}_{t+3}\right]|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right]$ attains its lowest value when $q^*_{t+1}(\theta,\mathbf{c}_{t+2})=1$.

Also by the properties of conditional expectations, we can similarly derive the following chain of relations:

$$
\begin{aligned}
E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta+1,\mathbf{c}_{t+2}\right] &= E_{\widetilde{\theta}_t}\left[E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t,\mathbf{c}_{t+3}\right]|\widetilde{\theta}_{t-1}=\theta+1,\mathbf{c}_{t+2}\right] \\
&= E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right] \\
&\quad + \left\{\begin{array}{c} -(1-\delta)q^*_{t+1}(\theta+1,\mathbf{c}_{t+2}) \\ \times\left\{\begin{array}{c} E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right] \\ -E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+2,\mathbf{c}_{t+3}\right]\end{array}\right\} \\ +\delta(1-q^*_{t+1}(\theta+1,\mathbf{c}_{t+2})) \\ \times\left\{\begin{array}{c} E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right] \\ -E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right]\end{array}\right\}\end{array}\right\} \\
&< E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right] \\
&\quad +\delta\left\{\begin{array}{c} E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right] \\ -E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right]\end{array}\right\} \\
&= \delta E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta,\mathbf{c}_{t+3}\right] \qquad (44) \\
&\quad +(1-\delta)E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_t=\theta+1,\mathbf{c}_{t+3}\right].
\end{aligned}
$$

Comparing (43) and (44), we see that

$$
E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta,\mathbf{c}_{t+2}\right] > E_{\widetilde{\theta}_{t+1}}\left[\eta(\widetilde{\theta}_{t+1})|\widetilde{\theta}_{t-1}=\theta+1,\mathbf{c}_{t+2}\right],
$$

which establishes $\frac{\partial z^*_t(\theta,\mathbf{c}_{t+1})}{\partial c_{t+3}} < 0$.

We have established the result for $s = 1, 2, 3$. Repeated application of the logic used in the case of $s = 3$ establishes that when $\delta = 0$, $\frac{\partial z^*_t(\theta,\mathbf{c}_{t+1})}{\partial c_{t+s}} < 0$ for all $s = 1, \dots T - t$. ■

Proof of Proposition 3:
We begin by establishing the following claim.

Claim 1 *For any access prices* $\mathbf{c} = (c_1, \dots c_T)$*,* $u_t(\theta, \mathbf{c}_t)$ *strictly increases in* θ *for* $t = 1, \dots, T$*, and in particular* $u_1(\theta, \mathbf{c})$ *is strictly increasing in* θ*.*

Proof. We will prove the result by induction. We first show the result holds for $t = T$ and $t = T - 1$. Then we show that if the result holds for any $t = T - 2, \dots, 2$, it must then hold for $t - 1$. From (8) we have $\frac{\partial u_T(\theta,c_T)}{\partial\theta} = -\eta'(\theta)D(c_T) > 0$ since $\eta'(\theta) < 0$. Thus $u_T(\theta, c_T)$ is strictly increasing in θ. Now, consider $t = T - 1$. From (6)

$$
\begin{aligned}
u_{T-1}(\theta,\mathbf{c}_{T-1}) &= \max_{q_{T-1}\in[0,1]}(c_{T-1}-\eta(\theta))D(c_{T-1})-F-I(q_{T-1})+\beta u_T(\theta,c_T) \\
&\quad +\beta\left\{(1-\delta)q_{T-1}\Delta u_T(\theta,c_T)-\delta(1-q_{T-1})\Delta u_T(\theta-1,c_T)\right\} \\
&= \max_{q_{T-1}\in[0,1]}(c_{T-1}-\eta(\theta))D(c_{T-1})-F-I(q_{T-1}) \\
&\quad +\beta E_{\widetilde{\theta}_{T-1}}\left[u_T(\widetilde{\theta}_{T-1},c_T)|\mathbf{c}_{T-1},q_{T-1},\widetilde{\theta}_{T-2}=\theta\right]
\end{aligned}
$$

By the envelope theorem

$$\frac{\partial u_{T-1}(\theta,\mathbf{c}_{T-1})}{\partial\theta}=-\eta'(\theta)D(c_{T-1})+\beta\frac{\partial E_{\widetilde{\theta}_{T-1}}\left[u_T(\widetilde{\theta}_{T-1},c_T)|\mathbf{c}_{T-1},q_{T-1},\widetilde{\theta}_{T-2}=\theta\right]}{\partial\theta}.$$

Now note that given that $\widetilde{\theta}_{T-2}=\theta$ is the state in period $T-1$, then $\widetilde{\theta}_{T-1}$ has one of three possible realizations: $\theta+1,\theta$, and $\theta-1$. Thus, an increase in θ shifts the distribution of $\widetilde{\theta}_{T-1}$ in the sense of first-order stochastic dominance. Because we just established that $u_T(\theta,c_T)$ is increasing in θ, it follows that $\frac{\partial E_{\widetilde{\theta}_{T-1}}\left[u_T(\widetilde{\theta}_{T-1},c_T)|\mathbf{c}_{T-1},q_{T-1},\widetilde{\theta}_{T-2}=\theta\right]}{\partial\theta}>0$, establishing that $\frac{\partial u_{T-1}(\theta,\mathbf{c}_{T-1})}{\partial\theta}>0$. Thus, we have established the result holds for $t=T$ and $t=T-1$. To show it holds generally—and for $u_1(\theta,\mathbf{c})$ in particular— let's assume that $u_t(\theta,\mathbf{c}_t)$ is strictly increasing in θ, and we will now show that $u_{t-1}(\theta,\mathbf{c}_{t-1})$ is also strictly increasing in θ. From (6)

$$\begin{aligned}u_{t-1}(\theta,\mathbf{c}_t) &= \max_{q_{t-1}\in[0,1]}(c_{t-1}-\eta(\theta))D(c_t)-F-I(q_{t-1})+\beta u_t(\theta,\mathbf{c}_t)\\ &\quad+\beta\left\{(1-\delta)q_{t-1}\Delta u_t(\theta,\mathbf{c}_t)-\delta(1-q_{t-1})\Delta u_t(\theta-1,\mathbf{c}_{t+1})\right\}\\ &= \max_{q_{t-1}\in[0,1]}(c_{t-1}-\eta(\theta))D(c_t)-F-I(q_{t-1})+\beta E_{\widetilde{\theta}_{t-1}}\left[u_t(\widetilde{\theta}_{t-1},c_t)|\mathbf{c}_{t-1},q_{t-1},\widetilde{\theta}_{t-2}=\theta\right].\end{aligned}$$

By the envelope theorem

$$\frac{\partial u_{t-1}(\theta,\mathbf{c}_t)}{\partial\theta}=-\eta'(\theta))D(c_t)+\beta\frac{\partial E_{\widetilde{\theta}_{t-1}}\left[u_t(\widetilde{\theta}_{t-1},c_t)|\mathbf{c}_{t-1},q_{t-1},\widetilde{\theta}_{t-2}=\theta\right]}{\partial\theta}.$$

Note that $\widetilde{\theta}_{t-2}=\theta$ is the state in period $t-1$, then $\widetilde{\theta}_{t-1}$ has one of three possible realizations: $\theta+1,\theta$, and $\theta-1$. Thus, an increase in θ shifts the distribution of $\widetilde{\theta}_{t-1}$ in the sense of first-order stochastic dominance. By the induction hypothesis $u_t(\theta,\mathbf{c}_t)$ is strictly increasing in θ, it follows that $\frac{\partial E_{\widetilde{\theta}_{t-1}}\left[u_t(\widetilde{\theta}_{t-1},\mathbf{c}_t)|\mathbf{c}_{T-1},q_{T-1},\widetilde{\theta}_{t-2}=\theta\right]}{\partial\theta}>0$, establishing that $\frac{\partial u_{t-1}(\theta,\mathbf{c}_{t-1})}{\partial\theta}>0$. By induction, $u_t(\theta,\mathbf{c}_t)$ is strictly increasing in θ for $t=T-2,\ldots,1$, and in particular, $u_1(\theta,\mathbf{c})$ is increasing in θ.■ ■

We next establish:

Claim 2 *$W(\theta)$ is strictly increasing in θ, i.e., $W(\theta+1)>W(\theta)$ for all θ.*

Proof. From (11)

$$W(\theta)=\max_{\mathbf{c}}\sum_{t=1}^{T}\beta^{t-1}\left[\Psi(c_t)-\Omega\right]+(1+\lambda)u_1(\theta,\mathbf{c})+\beta^T E_{\widetilde{\theta}_T}\left[W(\widetilde{\theta}_T)|\theta,\mathbf{c}_2\right]. \tag{45}$$

For any $\mathbf{c}_2$ and θ, the investments $\{q_t(\theta,\mathbf{c}_2)\}_{t=1}^{T}$ (coupled with the depreciation rate δ) gives rise to a Markov process $M(\theta,\mathbf{c}_2)$ that begins in state θ and generates random variables $\widetilde{\theta}_1,\ldots,\widetilde{\theta}_T$. Let $\mathbf{c}^*(\theta)=(c_1^*(\theta),\ldots,c_T^*(\theta))$ and $\mathbf{c}^*(\theta+1)=(c_1^*(\theta+1),\ldots,c_T^*(\theta+1))$ solve the optimization problem in (45) for a given θ and corresponding $\theta+1$. Because $\mathbf{c}^*(\theta)$ is a feasible but not necessarily optimal solution to the regulator's problem when the initial state is $\theta+1$ then necessarily

$$W(\theta+1)\geq\sum_{t=1}^{T}\beta^{t-1}\left[\Psi(c_t)-\Omega\right]+(1+\lambda)u_1(\theta+1,\mathbf{c}^*(\theta))+\beta^T E_{\widetilde{\theta}_T}\left[W(\widetilde{\theta}_T)|\theta+1,\mathbf{c}_2^*(\theta)\right].$$

Subtracting $W(\theta)$ from each side of the inequality above and using the expression for $W(\theta)$ and simplifying implies

$$W(\theta+1) - W(\theta) \geq (1+\lambda) \left[\begin{array}{c} u_1(\theta+1, \mathbf{c}^*(\theta)) \\ -u_1(\theta, \mathbf{c}^*(\theta)) \end{array} \right] + \beta^T \left[\begin{array}{c} E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T) | \theta+1, \mathbf{c}_2^*(\theta) \right] \\ -E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right] \end{array} \right].$$

As we have just shown in the previous claim, $u_1(\theta+1, \mathbf{c}^*(\theta)) - u_1(\theta, \mathbf{c}^*(\theta)) > 0$. Thus

$$W(\theta+1) - W(\theta) > \beta^T \left[\begin{array}{c} E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T) | \theta+1, \mathbf{c}_2^*(\theta) \right] \\ -E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right] \end{array} \right].$$

For a given sequence of transition probabilities, the stochastic process governing $\widetilde{\theta}_T$ given that we start at $\theta+1$ is the same as the stochastic process governing $\widetilde{\theta}_T + 1$ given that we start at θ. Thus

$$E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T) | \theta+1, \mathbf{c}_2^*(\theta) \right] - E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right] = E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T + 1) - W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right].$$

Thus, the inequality above can be written as

$$W(\theta+1) - W(\theta) > \beta^T E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T + 1) - W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right] \tag{46}$$

Now, this inequality holds for all θ and in particular it holds for all realizations of $\widetilde{\theta}_T$ that could arise conditional on starting at θ. This implies

$$E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T + 1) - W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right] > \beta^T \left[E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T + 1) | \theta, \mathbf{c}_2^*(\theta) \right] - E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right] \right]$$

Since $\beta^T \in (0,1)$, this necessarily implies that $E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T + 1) - W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right] > 0$. This is because the inequality could clearly not hold if $E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T + 1) - W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right] = 0$, and if $E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T + 1) - W(\widetilde{\theta}_T) | \theta, \mathbf{c}_2^*(\theta) \right] < 0$, then dividing each side by $E_{\widetilde{\theta}_T} \left[W(\widetilde{\theta}_T + 1) - W(\widetilde{\theta}_T) || \theta, \mathbf{c}_2^*(\theta) \right]$ implies $1 < \beta^T$, which cannot hold. Thus (46 implies $W(\theta+1) - W(\theta) > 0$.■ ■

To complete the proof, recall that Proposition 2 established that investment $q_t^*(\cdot, \mathbf{c}_{t+1})$ in each period t of the regulatory cycle (except the final period T when investment is zero) decreases in the access prices $\mathbf{c}_{t+1} = (c_{t+1}, \ldots, c_T)$ in the subsequent periods. Thus an increase in c_t decreases $q_1^*(\cdot, \mathbf{c}_2), q_2^*(\cdot, \mathbf{c}_3), \ldots, q_{T-1}^*(\cdot, \mathbf{c}_t)$ for any realization of the sequence of random variables $\{\widetilde{\theta}_t\}_{t=1}^{T-1}$. A decrease in c_t thus increases the likelihood that productivity increases between any two periods prior to period T. A decrease in c_t thus causes an upward shift the distribution of the random variable $\widetilde{\theta}_T$ (conditional on initial productivity $\widetilde{\theta}_0 = \theta$) in the sense of first-order stochastic dominance. Because $W(\cdot)$ increases in its argument, it follows that $\frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\theta, \mathbf{c}_2]}{\partial c_t} < 0$, $t = 2, \ldots, T$. ■

Proof of Proposition 4: Let $\mathbf{q}(\theta) = \{q_t(\widetilde{\theta}_{t-1})|\theta\}_{t=1}^T$ denote the Markov process over investment levels when the realization of initial productivity $\widetilde{\theta}_0$ is θ. In other words, $\mathbf{q}(\theta)$ is a vector of time-contingent, state-contingent, investment levels. A particular such vector is the set of investment levels chosen by the firm when the access prices are $\mathbf{c}$: $\mathbf{q}^*(\theta, \mathbf{c}) \equiv \{q_t^*(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1})\}_{t=1}^{T-1}$, and thus the firm's problem in a regulatory cycle can be written as a choice of $\mathbf{q}(\theta)$. By (6) and the Principle of Optimality, $\mathbf{q}^*(\theta, \mathbf{c})$ maximizes $u_1(\theta, \mathbf{q}(\theta))$, and thus if we have interior investments

$$\left.\frac{\partial u_1(\theta, \mathbf{q}(\theta))}{q_t(\widetilde{\theta}_{t-1})}\right|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} = 0, \quad \text{for all } t = 1, \ldots, T-1, \text{all realizations of } \widetilde{\theta}_{t-1}. \tag{47}$$

Now, the distribution of $\widetilde{\theta}_T$ thus depends on $\mathbf{q}(\theta)$ and we can rewrite $E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\theta, \mathbf{c}_2]$ as

$$E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})}$$

Fix a particular t, and write $\frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\theta,\mathbf{c}_2]}{\partial c_t}$ as follows:

$$\frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\theta, \mathbf{c}_2]}{\partial c_t} = \begin{bmatrix} \left\{ \frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_1}|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} \right\} \frac{\partial q_1^*(\theta,\mathbf{c}_2)}{\partial c_t} \\ +\sum_{\widetilde{\theta}_1} \left\{ \frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_2(\widetilde{\theta}_1)}|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} \right\} \frac{\partial q_2^*(\widetilde{\theta}_1,\mathbf{c}_3)}{\partial c_t} \\ +\ldots \\ +\sum_{\widetilde{\theta}_{t-2}} \left\{ \frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_{t-1}(\widetilde{\theta}_{t-2})}|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} \right\} \frac{\partial q_{t-1}^*(\widetilde{\theta}_{t-2},\mathbf{c}_t)}{\partial c_t} \end{bmatrix}, \tag{48}$$

where the summations indicate summations over states. For example, assuming $\theta \in (\underline{\theta}, \overline{\theta})$, $\widetilde{\theta}_1$ takes on possible values, $\max\{\theta-1, \underline{\theta}\}, \theta, \min\{\theta+1, \overline{\theta}\}$, and the second line in (48) would be

$$\begin{bmatrix} \left\{ \frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_2(\max\{\theta-1,\underline{\theta}\})}|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} \right\} \frac{\partial q_2^*(\max\{\theta-1,\underline{\theta}\},\mathbf{c}_3)}{\partial c_t} \\ + \left\{ \frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_2(\theta)}|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} \right\} \frac{\partial q_2^*(\theta,\mathbf{c}_3)}{\partial c_t} \\ + \left\{ \frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_2(\min\{\theta+1,\overline{\theta}\})}|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} \right\} \frac{\partial q_2^*(\min\{\theta+1,\overline{\theta}\},\mathbf{c}_3)}{\partial c_t} \end{bmatrix}$$

Given Propositions 2 and 3, at least one of the terms in curly brackets in (48) must be positive, indicating that for at least one period $s \in \{1, \ldots, t\}$, $\frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_s(\widetilde{\theta}_{s-1})}|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} > 0$. Now, let us use the formulation in which the regulator's discounted welfare and the firm's discounted profits depend on the vector $\mathbf{q}(\theta)$ of state-contingent investments. We can write the regulator's discounted expected welfare as

$$W(\theta, \mathbf{q}(\theta)) = \sum_{t=1}^{T} \beta^{t-1} [\Psi(c_t) - \Omega] + u_1(\theta, \mathbf{q}(\theta)) + \beta E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)].,$$

The derivative of the regulator's welfare with respect to an investment in a particular time $s < T$ and state $\widetilde{\theta}_{s-1}$ is

$$\frac{\partial W(\theta, \mathbf{q}(\theta))}{q_s(\widetilde{\theta}_{s-1})} = \frac{\partial u_1(\theta, \mathbf{q}(\theta))}{q_s(\widetilde{\theta}_{s-1})} + \beta \frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_s(\widetilde{\theta}_{s-1})}.$$

Evaluating the above derivative at $\mathbf{q}(\theta) = \mathbf{q}^*(\theta, \mathbf{c})$ and using (47) gives us

$$\left.\frac{\partial W(\theta, \mathbf{q}(\theta))}{q_s(\widetilde{\theta}_{s-1})}\right|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} = \beta \left.\frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_s(\widetilde{\theta}_{s-1})}\right|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})}.$$

By the result just established for any there must be some s and realizations of $\widetilde{\theta}_{s-1}$ for which $\left.\frac{\partial E_{\widetilde{\theta}_T}[W(\widetilde{\theta}_T)|\mathbf{q}(\theta)]}{\partial q_s(\widetilde{\theta}_{s-1})}\right|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})}$ and hence $\left.\frac{\partial W(\theta,\mathbf{q}(\theta))}{q_s(\widetilde{\theta}_{s-1})}\right|_{\mathbf{q}(\theta)=\mathbf{q}^*(\theta,\mathbf{c})} > 0$. Hence, in at least one period prior to the and one state, the regulator prefers more investment than the firm actually makes.■

Proof of Proposition 5:

The result follows immediately from part (a) of Proposition **3** (which implies $\widehat{\xi}_t(\theta, c_2) < \widehat{\eta}_t(\theta, c_2)$), equations (5) and (12), and the property that $c^0(\eta)$ decreases in η.■

Proof of Proposition 6:

Since $c^0(\cdot)$ is strictly increasing, it follows that

$$\begin{aligned} c_t^*(\theta) &= c^0(\widehat{\xi}_t(\theta, \mathbf{c}_2)) \\ &< c^0(\widehat{\eta}_t(\theta, \mathbf{c}_2)) \\ &< c^0(\eta(\theta)) \\ &= c_1^*(\theta) \end{aligned}$$

where (a) the first equality follows directly from (5) and (12); (b) the first inequality follows because, as established in Proposition 3, $\widehat{\xi}_t(\theta, \mathbf{c}_2) < \widehat{\eta}_t(\theta, \mathbf{c}_2)$; (c) the second inequality follows because when because when $\delta = 0$, the expectations of marginal cost must decline, i.e., so $\widehat{\eta}_t(\theta, \mathbf{c}_2) < \eta(\theta)$.■

Proof of Proposition 7:

If $T = 1$, Proposition 1 implies that the firm does not invest. With no depreciation, productivity never changes, and since $c_1^*(\theta) = c^0(\eta(\theta))$, we have $W^1(\theta) = \frac{\omega^0(\eta(\theta))}{1-\beta}$. Using (15), we have

$$W^T(\theta) - W^1(\theta) = \max_{c_1,\ldots,c_T} \frac{1}{1-\beta^T} \left\{ \begin{array}{c} \sum_{t=1}^T \beta^{t-1} \left\{\omega(c_t, \widehat{\eta}_t(\theta, \mathbf{c}_2)) - \omega^0(\eta(\theta))\right\} \\ -\sum_{t=1}^T \beta^{t-1}(1+\lambda) E_{\widetilde{\theta}_{t-1}} \left[I(q_t(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1}))|\theta, \mathbf{c}_2 \right] \\ +\beta^T \left[E_{\widetilde{\theta}_T} \left[W^T(\widetilde{\theta}_T)|\theta, \mathbf{c}_2 \right] - W^T(\theta) \right] \end{array} \right\}. \tag{49}$$

To proceed, we establish the following result:

Claim 3 *Consider two multivariate functions $f(\mathbf{x})$ and $g(\mathbf{x})$ such that $g(\mathbf{x}) \geq 0$ for all $\mathbf{x}$. Then*

$$\max_{\mathbf{x}} [f(x) + g(x)] \geq \max_{\mathbf{x}} f(x).$$

Proof. Let

$$H(\varphi) = \max_{\mathbf{x}} [f(\mathbf{x}) + \varphi g(\mathbf{x})]$$

where $\varphi \in [0, 1]$, and let $\mathbf{x}^*(\varphi)$ be the optimal solution. By the envelope theorem $H'(\varphi) = g(\mathbf{x}^*(\varphi))$. Since $g(\cdot) \geq 0$ we have $H'(\varphi) \geq 0$, and thus

$$H(1) = \max_{\mathbf{x}} [f(\mathbf{x}) + g(\mathbf{x})] \geq \max_{\mathbf{x}} f(\mathbf{x}) = H(0).$$

■

Applying this result to (49), we have

$$\begin{aligned} W^T(\theta) - W^1(\theta) &= \max_{c_1,\ldots,c_T} \frac{1}{1-\beta^T} \left\{ \begin{array}{c} \sum_{t=1}^T \beta^{t-1} \left\{\omega(c_t, \widehat{\eta}_t(\theta, \mathbf{c}_2)) - \omega^0(\eta(\theta))\right\} \\ -\sum_{t=1}^T \beta^{t-1}(1+\lambda) E_{\widetilde{\theta}_{t-1}} \left[I(q_t(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1})) | \theta, \mathbf{c}_2 \right] \\ +\beta^T \left[E_{\widetilde{\theta}_T} \left[W^T(\widetilde{\theta}_T) | \theta, \mathbf{c}_2 \right] - W^T(\theta) \right] \end{array} \right\} \\ &\geq \max_{c_1,\ldots,c_T} \frac{1}{1-\beta^T} \left\{ \begin{array}{c} \sum_{t=1}^T \beta^{t-1} \left\{\omega(c_t, \widehat{\eta}_t(\theta, \mathbf{c}_2)) - \omega^0(\eta(\theta))\right\} \\ -\sum_{t=1}^T \beta^{t-1}(1+\lambda) E_{\widetilde{\theta}_{t-1}} \left[I(q_t(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1})) | \theta, \mathbf{c}_2 \right] \end{array} \right\} \end{aligned} \tag{50}$$

The inequality follows from the claim just proven and $E_{\widetilde{\theta}_T}\left[W^T(\widetilde{\theta}_T)|\theta, \mathbf{c}_2\right] - W^T(\theta) \geq 0$. This latter inequality holds because: (i) as established shown in the proof of Lemma 3, the regulator's value function is strictly increasing in θ, and (ii) by assumption there is no depreciation, so any realization of $\widetilde{\theta}_T$ conditional on the initial state being θ, must be at least as large as θ.

Now an implication of (6) is that for any state θ at the start of a regulatory cycle, the firm can be thought of as choosing a set of state-contingent investments $\mathbf{q}(\theta) = \{q_1(\theta), q_2(\theta), q_2(\theta+1), \ldots, q_{T-1}(\theta), q_{T-1}(\theta+1), \ldots, q_{T-1}(\theta+T-2)\}$ that maximize the discounted present value of its expected profit for any schedule $\mathbf{c}_2$ of access prices it faces. (Recall that in the terminal period T, there is no investment and also recall that the price c_1 in the first period of the cycle does not affect investment decisions.) Any such choice induces a probability distribution over the the sequence of random variables $\{\widetilde{\theta}_0, \widetilde{\theta}_1, \ldots \widetilde{\theta}_{T-1}\}$ conditional on $\widetilde{\theta}_0 = \theta$, and thus determine the expectations of marginal cost $\left\{E_{\widetilde{\theta}_{t-1}}\left[\eta(\widetilde{\theta}_{t-1})|\theta, \mathbf{q}(\theta)\right]\right\}_{t=1}^T$. When $\mathbf{q}(\theta)$ equals firm's optimal investment strategies $\{q_t(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1})\}$, the expectations of marginal cost are the objects $\{\widehat{\eta}_t(\theta, \mathbf{c}_2)\}_{t=1}^T$. Now, because $\sum_{t=1}^T \beta^{t-1}\left\{\omega(c_t, \widehat{\eta}_t(\theta, \mathbf{c}_2)) - (1+\lambda)E_{\widetilde{\theta}_{t-1}}\left[I(q_t(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1}))|\theta, \mathbf{c}_2\right]\right\}$ is discounted downstream surplus (which does not depend on investment) plus $(1+\lambda)$ times discounted expected profit evaluated at the firm's optimal investment strategies $\{q_t(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1})\}$, it follows that the firm's optimal investments $\{q_t(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1})\}$ in (50) are the maximizers of the objective function in ((50). We can thus rewrite (50) as

$$W^T(\theta) - W^1(\theta) \geq \max_{c_1,\ldots,c_T,\mathbf{q}(\theta)} \frac{1}{1-\beta^T} \left\{ \begin{array}{c} \sum_{t=1}^T \beta^{t-1} \left\{\omega(c_t, E_{\widetilde{\theta}_{t-1}}\left[\eta(\widetilde{\theta}_{t-1})|\theta, \mathbf{q}(\theta)\right]) - \omega^0(\eta(\theta))\right\} \\ -\sum_{t=1}^T \beta^{t-1}(1+\lambda) E_{\widetilde{\theta}_{t-1}} \left[I(q_t(\widetilde{\theta}_{t-1})) | \theta, \mathbf{q}(\theta) \right] \end{array} \right\}. \tag{51}$$

A feasible but not optimal solution to the optimization problem in (51) is $c_t = c^0(\eta(\theta))$, $t = 1, ..., T$ and $\mathbf{q}(\theta) = \{q_1(\theta), q_2(\theta), q_2(\theta+1), \ldots, q_{T-1}(\theta), q_{T-1}(\theta+1), \ldots, q_{T-1}(\theta+T-2)\} = \{0, \ldots, 0\}$, i.e., set all prices to the first-best level given the initial marginal cost and set all investments to zero. This would imply $E_{\widetilde{\theta}_{t-1}}\left[\eta(\widetilde{\theta}_{t-1})|\theta, \mathbf{q}(\theta)\right] = \eta(\theta)$ for all t, and since by definition, since by definition, $\omega^0(\eta(\theta)) = \omega(c^0(\eta(\theta))$, with this solution, the objective function in (51) attains a value of zero. Thus,

$$\max_{c_1,\ldots,c_T,\mathbf{q}(\theta)} \frac{1}{1-\beta^T} \left\{ \begin{array}{c} \sum_{t=1}^T \beta^{t-1} \left\{\omega(c_t, E_{\widetilde{\theta}_{t-1}}\left[\eta(\widetilde{\theta}_{t-1})|\theta, \mathbf{q}(\theta)\right]) - \omega^0(\eta(\theta))\right\} \\ -\sum_{t=1}^T \beta^{t-1}(1+\lambda) E_{\widetilde{\theta}_{t-1}} \left[I(q_t(\widetilde{\theta}_{t-1})) | \theta, \mathbf{q}(\theta) \right] \end{array} \right\} > 0,$$

so $W^T(\theta) - W^1(\theta) > 0$.■

Proof of Proposition 8:

When $\lambda \to \infty$, investment becomes arbitrarily small for any access prices and realizations of productivity, i.e., $q_t(\widetilde{\theta}_{t-1}, \mathbf{c}_{t+1}) \to 0$ for any regulatory lag. In addition, for any productivity realization, $q^0(\widetilde{\theta}_{t-1}) \to 0$. This implies that a continuous regulatory review, $T = 1$, implements the first-best solution, i.e., $W^1(\theta) = W^0(\theta)$.[38] Thus, $W^1(\theta) - W^T(\theta) = DWL^T(\theta)$. Moreover, from (16) and (18)

$$DWL^T(\theta) = SE^T(\theta) + \beta^T E_{\widetilde{\theta}_T}\left[DWL(\widetilde{\theta}_T)|\widetilde{\theta}_0 = \theta, \mathbf{q}^0(\theta) = \mathbf{0}\right]. \tag{52}$$

Even though investment is zero, because $\delta > 0$, the productivity process $\left\{\widetilde{\theta}_{t-1}\right\}_{t=1}^{T}$ is still stochastic (with productivity destined to decrease over time). Thus, there is a positive probability in periods $t = 2, \ldots, T$ that there are some realizations of $\widetilde{\theta}_{t-1}$ for which $\eta(\widetilde{\theta}_{t-1}) \neq \widehat{\eta}_t(\theta) = E_{\widetilde{\theta}_{t-1}}\left[\eta(\widetilde{\theta}_{t-1}))|\widetilde{\theta}_0 = \theta, \mathbf{q}^*(\theta) = \mathbf{0}\right]$.[39] For such realizations,

$$\omega^0(\eta(\widetilde{\theta}_{t-1}))|\mathbf{q}^0(\theta) = \mathbf{0}) = \max_c \omega(c, \eta(\widetilde{\theta}_{t-1})) > \omega(c^0(\widehat{\eta}_t(\theta)), \eta(\widetilde{\theta}_{t-1})).$$

Since $c_t^*(\theta) = c^0(\widehat{\eta}_t(\theta))$, it follows that for periods $t = 2, \ldots, T$,

$$\begin{aligned} E_{\widetilde{\theta}_{t-1}}\left[\omega^0(\eta(\widetilde{\theta}_{t-1}))|\widetilde{\theta}_0 = \theta, \mathbf{q}^0(\theta) = \mathbf{0}\right] &> E_{\widetilde{\theta}_{t-1}}\left[\omega(c_t^*(\widehat{\eta}_t(\theta)), \eta(\widetilde{\theta}_{t-1}))||\widetilde{\theta}_0 = \theta, \mathbf{q}^*(\theta) = \mathbf{0}\right] \\ &= \omega(c_t^*(\widehat{\eta}_t(\theta)), \widehat{\eta}_t(\theta)), \end{aligned}$$

which from (17) implies $SE^T(\theta) > 0$. To complete the proof we need to show that this implies $DWL^T(\theta) > 0$. This follows from an induction argument. Specifically, for $\theta = \underline{\theta}$, there is zero probability that productivity decreases, and $DWL^T(\underline{\theta}) = SE^T(\underline{\theta}) + \beta^T DWL(\underline{\theta})$,so $DWL^T(\underline{\theta}) = \frac{SE^T(\underline{\theta})}{1-\beta^T} > 0$. Similarly,

$$DWL^T(\underline{\theta}+1) = SE^T(\underline{\theta}+1) + \beta^T\left[(1-\delta)DWL(\underline{\theta}+1) + \delta DWL(\underline{\theta})\right],$$

or

$$DWL^T(\underline{\theta}+1) = \frac{SE^T(\underline{\theta}+1)}{1-(1-\delta)\beta^T} + \frac{\delta DWL(\underline{\theta})}{1-(1-\delta)\beta^T} > 0.$$

This establishes that $DWL^T(\underline{\theta})$ and $DWL^T(\underline{\theta}+1)$ are positive. The induction hypothesis is that $DWL^T(\theta-1), \ldots, DWL^T(\underline{\theta})$ are positive. This then implies that $DWL^T(\theta)$ is positive because

$$DWL^T(\theta) = SE^T(\theta) + \beta^T(1-\delta)DWL^T(\theta) + \sum_{i=1}^{\theta-\underline{\theta}} a_i DWL^T(\theta - i),$$

where a_i are probabilities that are between zero and one. It follows that $W^1(\theta) - W^T(\theta) = DWL^T(\theta) > 0$.■

[38] Hereafter in the proof, any statements of equality refer to equality in the limit as $\lambda \to \infty$.

[39] Because investment is zero in the limit, the object $\widehat{\eta}_t(\theta, \mathbf{c}_2^*(\theta))$—the expectation of marginal cost in period t conditional on the productivity at the beginning of the regulatory cycle being θ—no longer depends on the access prices, so we write it without the $\mathbf{c}_2^*(\theta)$ argument.

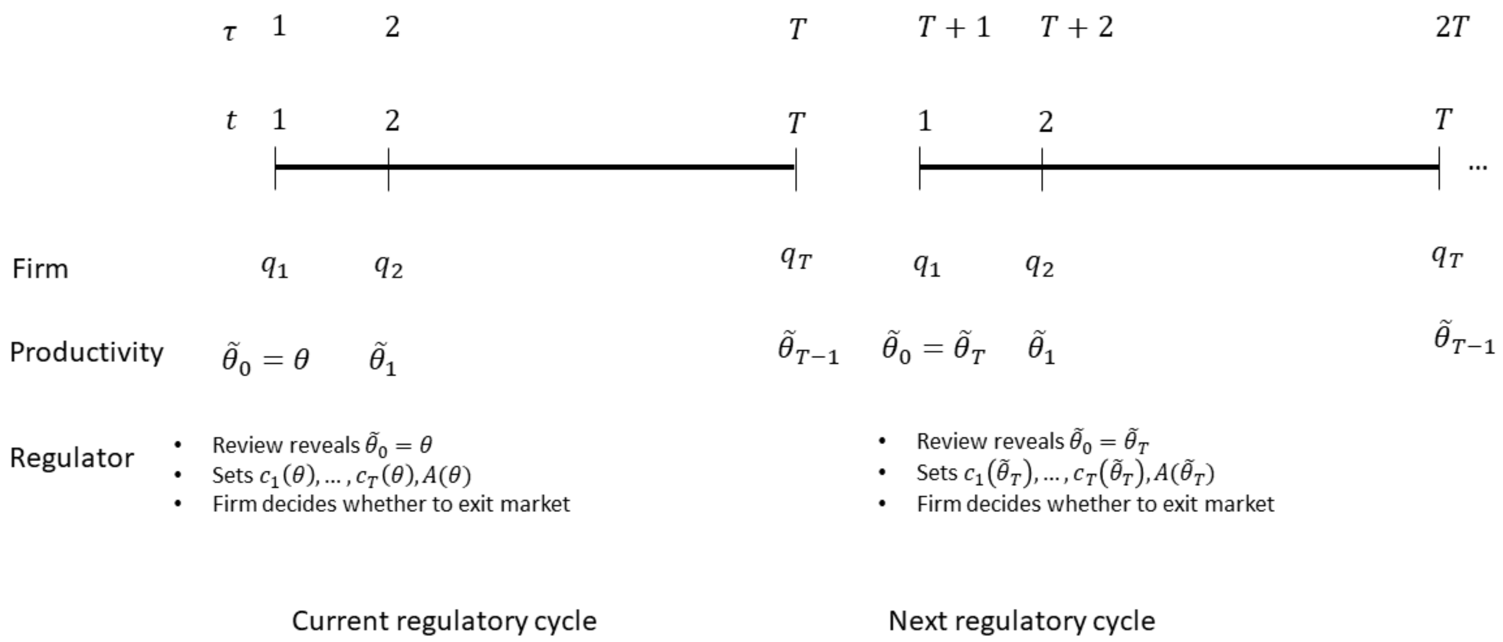

Figure 1: Timing of actions and events.

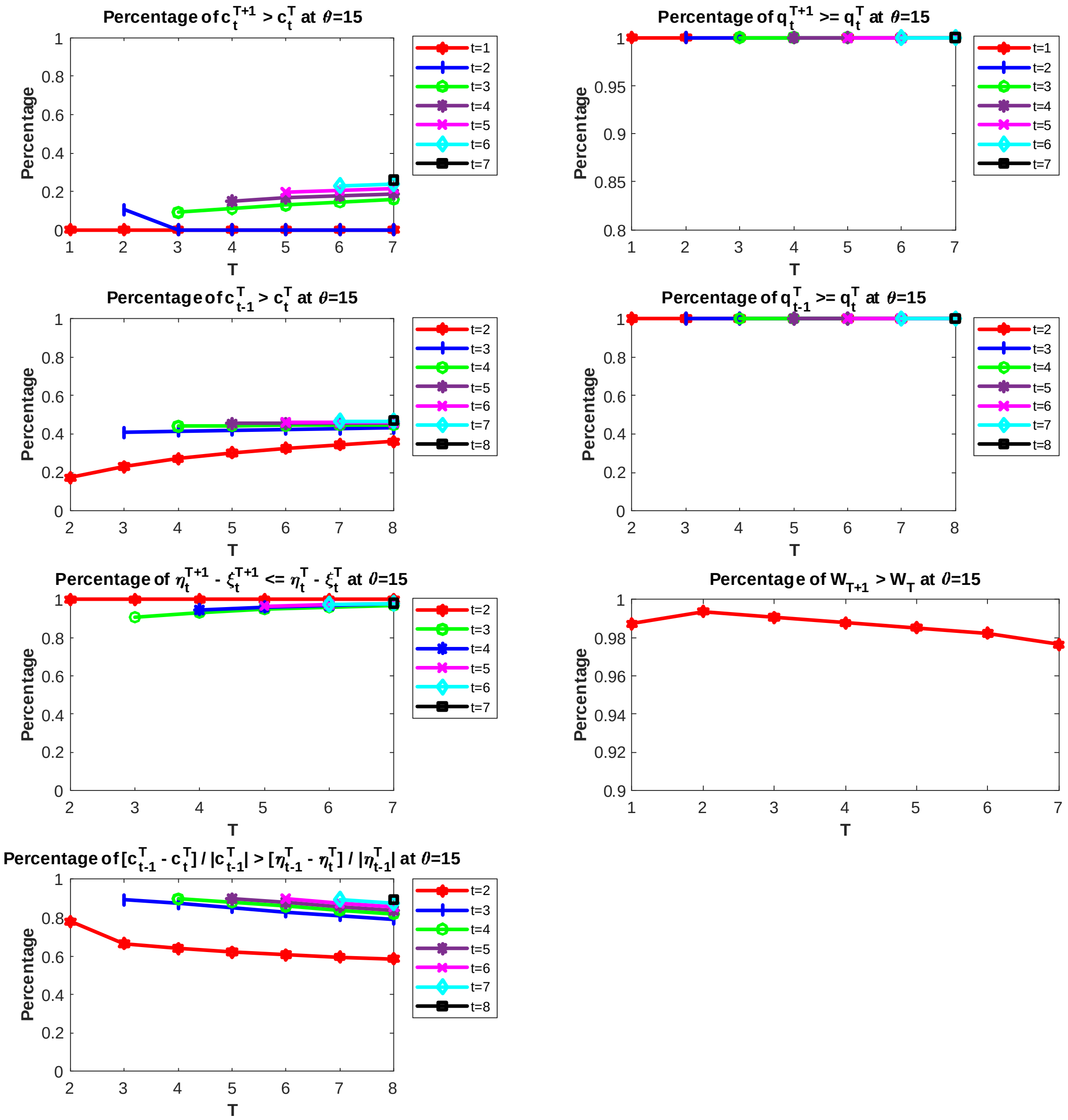

Figure 2: Percentage of parameterizations satisfying various properties across the paramter grid $\mathcal{G}$.

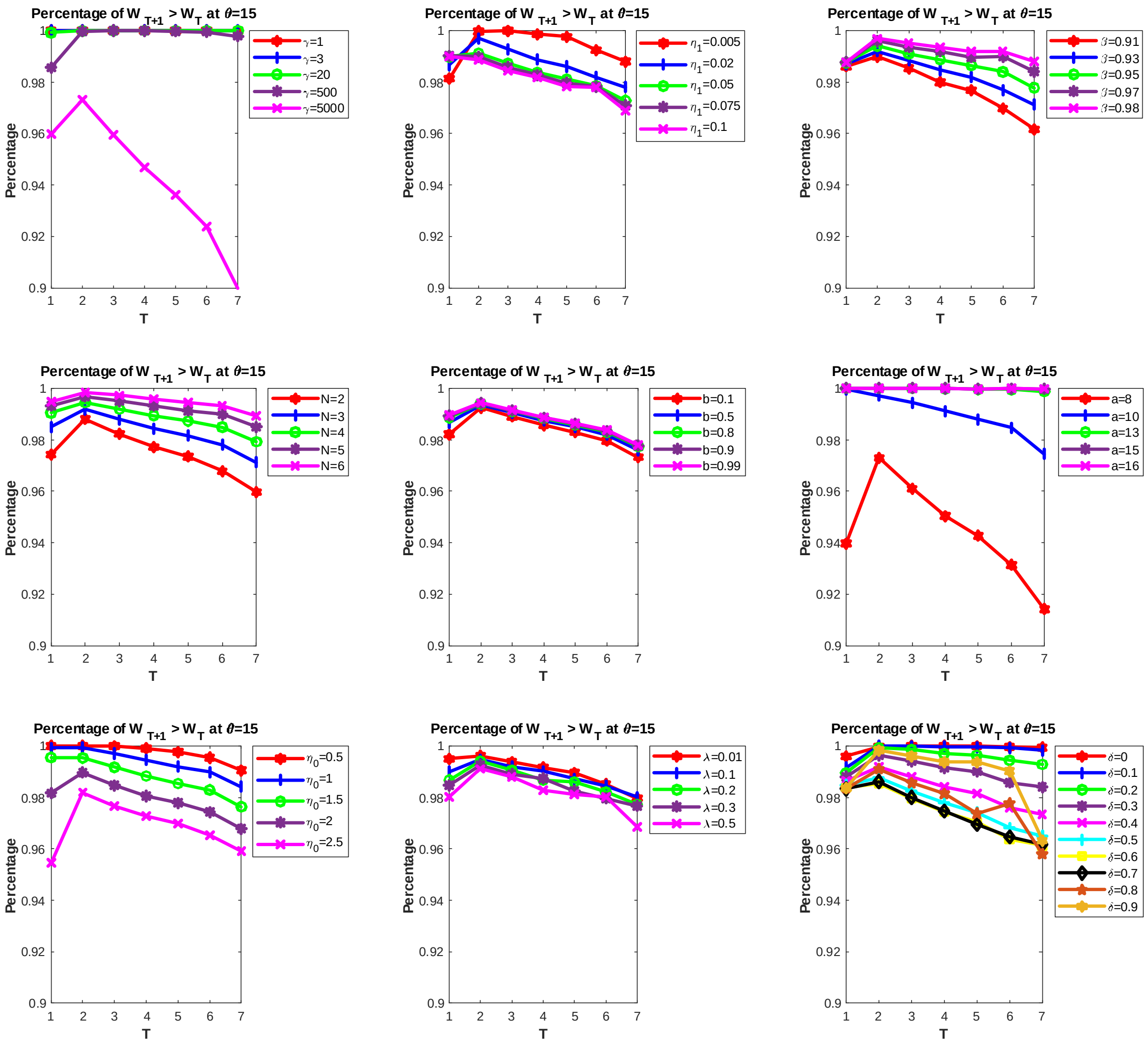

Figure 3: Percentage of parameterizations for which $W^{T+1} > W^T$ (at $\theta = 15$) across the paramter grid $\mathcal{G}$, holding a single parameter fixed at a particular value.

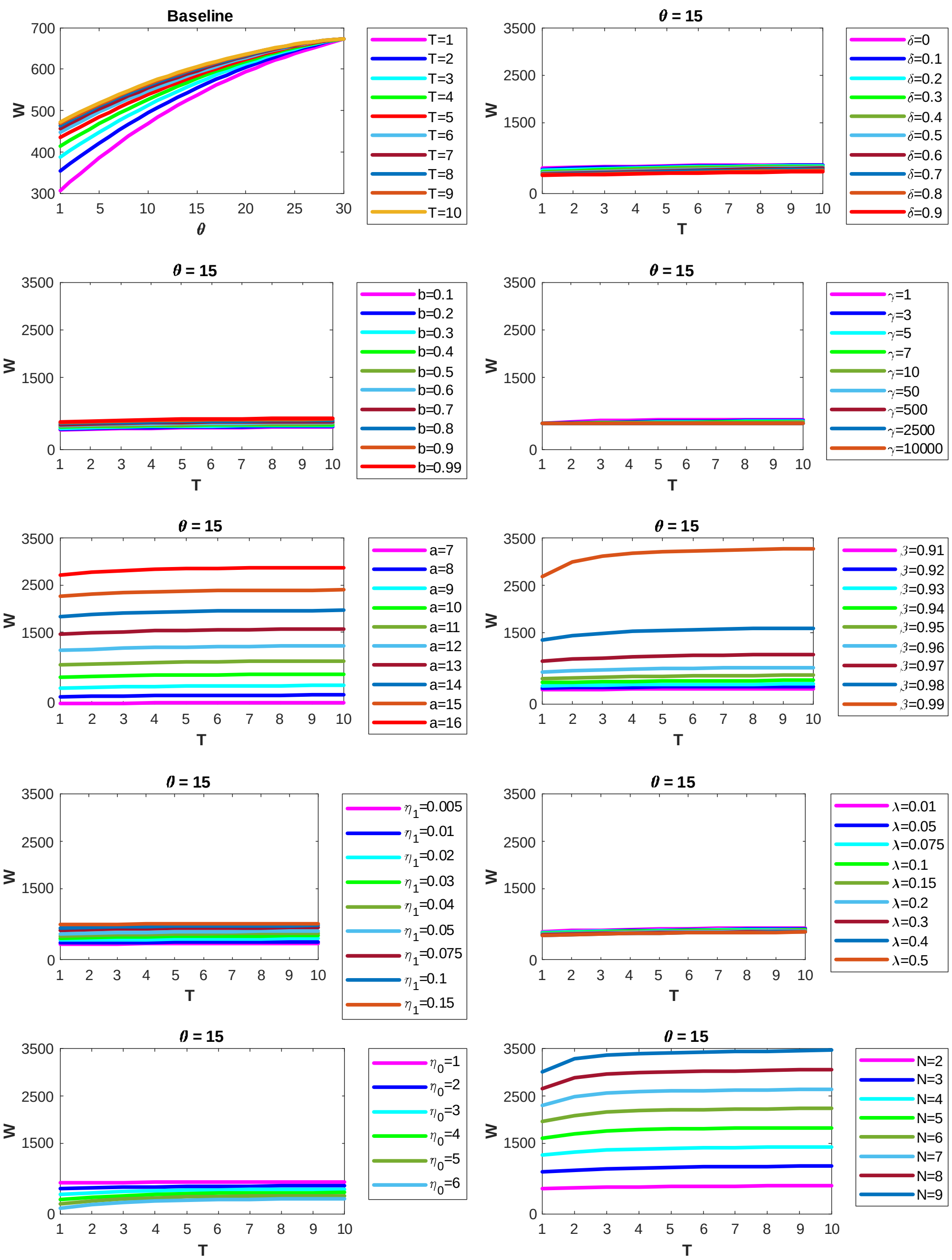

Figure 4: The regulator's welfare $W^T(\theta)$ for $T = 1, \ldots, 10$.

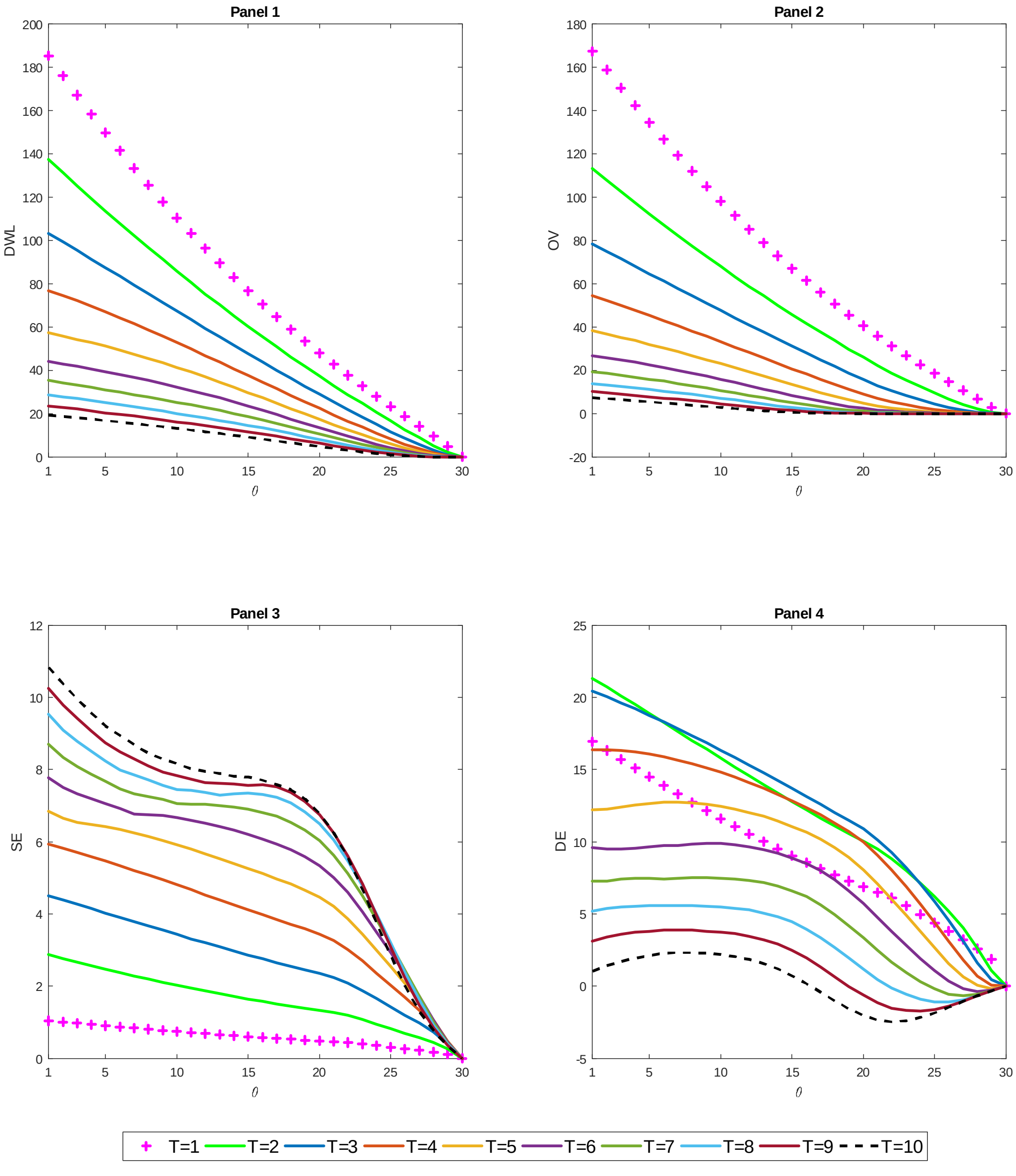

Figure 5: Deadweight loss decomposition

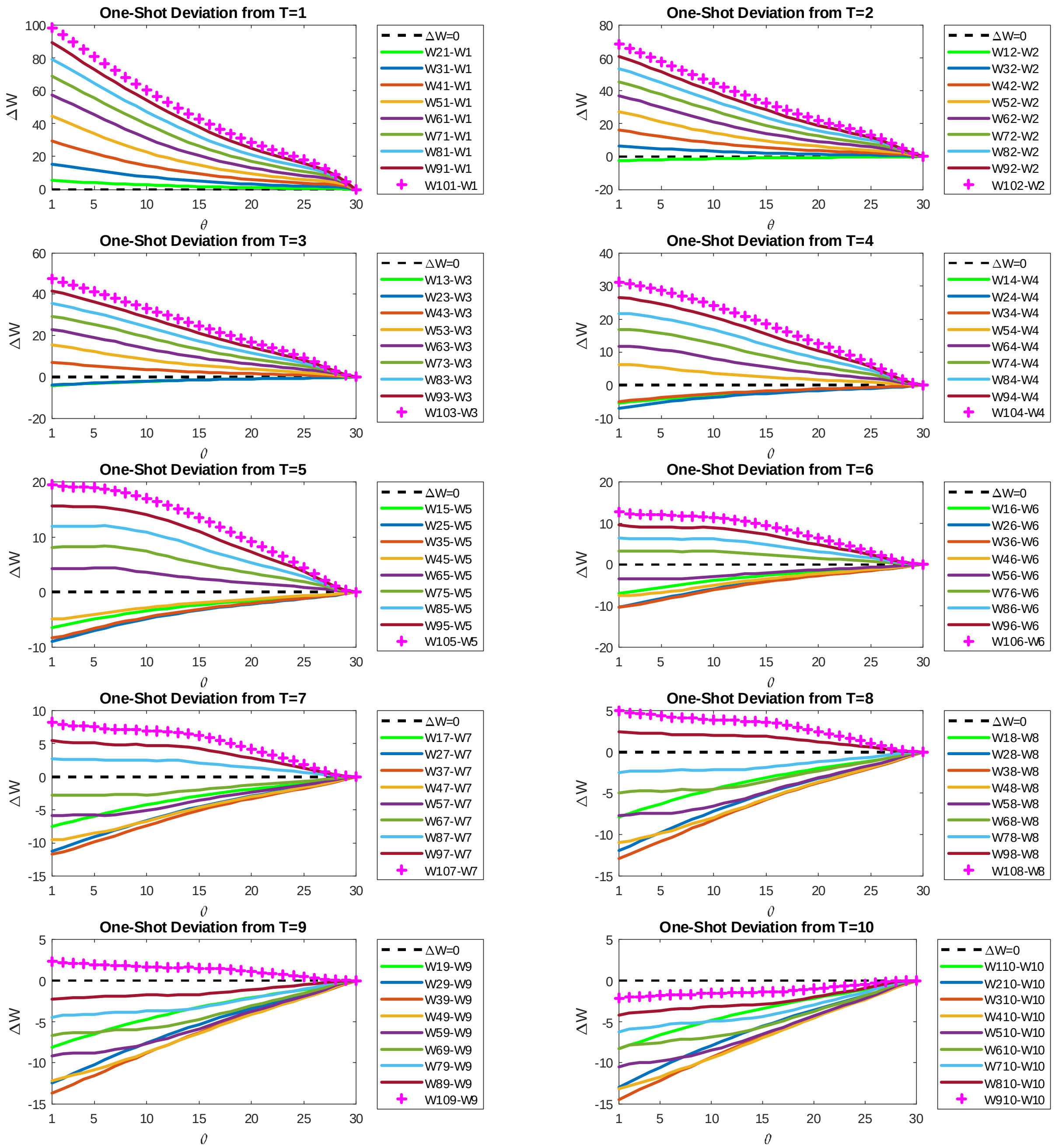

Figure 6: Each panel shows $W^T(\theta)$—the regulator's value function with a cycle length T—and $W^{T,\widehat{T}}(\theta)$—the regulator's value function when for a one-shot deviation to a cycle length $\widehat{T}$, followed by a return to cycle length T in the future.

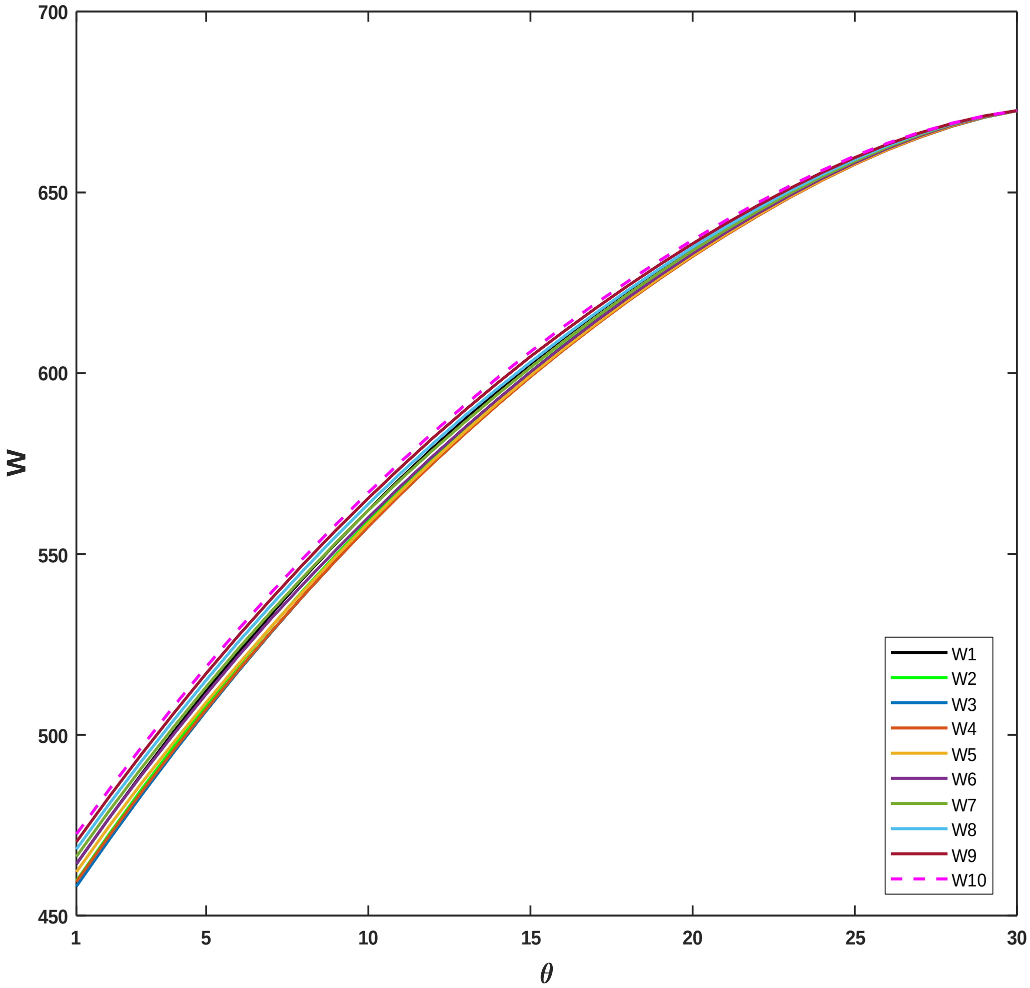

Figure 7: The regulator's welfare $W^T(\theta)$ for selected values of T and θ when T is endogenous

References

[1] Ai, Chunrong and David Sappington (2002). "The Impact of State Incentive Regulation on the U.S. Telecommunications Industry," *Journal of Regulatory Economics*, Vol. 22, No. 2, pp. 132-160.

[2] Armstrong, Mark, Simon Cowan, and John Vickers (1994). *Regulatory Reform: Economic Analysis and British Experience*, (Cambridge, MA: MIT Press).

[3] Armstrong, Mark and David Sappington (2007). "Recent Developments in the Theory of Regulation," Chapter 27 in *Handbook of Industrial Organization*, Volume 3, M. Armstrong and R. Porter, eds., (Amsterdam: Elsevier), pp. 1557-1700.

[4] Armstrong, Mark, Ray Rees, and John Vickers (1995). "Optimal Regulatory Lag under Price Cap Regulation," *Revista Espanola De Economia*, Vol. 12, pp. 93–116.

[5] Arrow, Kenneth (1962). "Economic Welfare and the Allocation of Resources for Invention," In National Bureau of Economic Research, *The Rate and Direction of lnventive Activity*, (Princeton: Princeton University Press).

[6] Baron, David and David Besanko (1987). "Monitoring, Moral Hazard, Asymmetric Information, and Risk Sharing in Procurement Contracting," *RAND Journal of Economics,* Vol. 18, No. 4, pp. 509–532.

[7] Baron, David and David Besanko (1987). "Commitment and Fairness in a Dynamic Regulatory Relationship," *Review of Economic Studies*, Vol. 54, No. 3, pp. 413-436.

[8] Baron, David and Roger Myerson (1982). "Regulating a Monopolist with Unknown Costs," *Econometrica*, Vol. 50, No. 4, pp. 911–930.

[9] Beesley, Michael and Stephen Littlechild (1989). "The Regulation of Privatized Monopolies in the United Kingdom," *RAND Journal of Economics*, Vol. 20, No. 3, pp. 454-472

[10] Bernstein, Jeffrey and David Sappington (1999). "Setting the X factor in Price-Cap Regulation Plans," *Journal of Regulatory Economics*, Vol. 16, No. 1 pp. 5-25.

[11] Bernstein, Jeffrey and David Sappington (2000). "How to Determine the X in RPI-X Regulation: a User's Guide," *Telecommunications Policy*, Vol. 24, pp. 63-68.

[12] Besanko, David and Ulrich Doraszelski (2004). "Capacity Dynamics and Endogenous Asymmetries in Firm Size," *RAND Journal of Economics*, 2004, Vol. 35, issue 1, 23-49

[13] Biglaiser, Gary and Michael Riordan, M. (2000). "Dynamics of price regulation," *RAND Journal of Economics*, Vol. 31, No. 4, pp. 744-767.

[14] Bourreau, Marc, Carlo Cambini, Steffen Hoernig (2020). "Fiber Investment and Access Under Uncertainty: Long-term Contracts, Risk Premia, and Access Options," *Journal of Regulatory Economics*, 57(2), April.

[15] Brennan Timothy (1989) "Regulating by 'Capping' Prices." Journal of Regulatory Economics, Vol. 1, pp. 133–147.

[16] Cabral, Luis and Michael Riordan, M.H. (1989), "Incentives for Cost Reduction Under Price Cap Regulation," *Journal of Regulatory Economic*s, Vol. 1, pp. 93–102.

[17] Cantos, Pedro and Javier Campos (2005). "Recent Changes in the Global Rail Industry: Evaluating the New Regulatory Instruments," *European Transport*, Vol. 29, pp. 22-45.

[18] Cowan, Simon (2002). "Price-cap Regulation," *Swedish Economic Policy Review*, Vol. 9, pp. 167-188.

[19] Crew, Michael and Timothy Brennan (2016). "Price Cap Regulation and Declining Demand," Chapter 1 in *The Future of the Postal Sector in a Digital World*, T. Brennan and M. Crew, eds., (New York: Springer).

[20] Decker, Christopher (2015), *Modern Economic Regulation: An Introduction to Theory and Practice* (Cambridge: Cambridge University Press).

[21] DeHouse, Franklin and Benedetta Mariscola (2015). "The EU'S Fourth Railway Package: a New Stop in a Long Regulatory Journey," Egmont Paper 76.

[22] de Rus, G. and M.P. Socorro (2014). "Access Pricing, Infrastructure Investment and Intermodal Competition," *Transportation Research Part E: Logistics and Transportation Review*, Vol. 70, pp. 374–387.

[23] Evans, Lewis, and Graeme Guthrie (2012). "Price-cap Regulation and the Scale and Timing of Investment." *The RAND Journal of Economics,* Vol. 43, No. 3, pp. 537-561.

[24] Giannakis, Dimitrios, Tooraj Jamas, and Michael Pollitt (2005). "Benchmarking and Incentive Regulation of Quality of Service: an Application to the UK Electricity Distribution Networks," *Energy Policy*, Vol. 33, pp. 2256–2271

[25] Guthrie, Graeme (2006). "Regulating Infrastructure: The Impact of Risk and Investment," *Journal of Economic Literature*, Vol. 44, No. 4, pp. 925-972.

[26] Joskow, Paul (2014). "Incentive Regulation in Theory and Practice: Electricity Distribution and Transmission Networks," in *Economic Regulation and Its Reform: What Have We Learned?,* N. Rose, ed. (Chicago: University of Chicago Press), pp. 291-344.

[27] Laffont, Jean-Jacques and Jean Tirole (1986). "Using Cost Observation to Regulate Firms," *Journal of Political Economy,* Vol. 94, No. 3, pp. 614–641.

[28] Laffont, Jean-Jacques and Jean Tirole (2000). *Competition in Telecommunications*, (Cambridge, MA: MIT Press).

[29] Lewis, Tracy and David Sappington (1989), "Regulatory Options and Price-Cap Regulation," *The RAND Journal of Economics*, Vol. 20, No. 3, pp. 405-416.

[30] Littlechild, Steven (1983). "Regulation of British Telecommunications' Profitability," Department of Industry, Report to the Secretary of State, https://www.eprg.group.cam.ac.uk/wp-content/uploads/2019/10/S.-Littlechild_1983-report.pdf.

[31] Lima, José Luis and Andrés Gómez-Lobo (2010), "Good Regulatory Lags," available at SSRN: https://ssrn.com/abstract=657481 or http://dx.doi.org/10.2139/ssrn.657481.

[32] Linhart, Peter and Roy Radner (1992). "Price Caps," *Annals of Operations Research*, Vol. 36, pp. 17-32.

[33] Linhart, Peter, Roy Radner, and Frank Sinden (1991), "A Sequential Mechanism for Direct Price Regulation," In *Price Caps and Incentive Regulation in Telecommunications*, (Boston: Springer), pp. 127-153.

[34] Link, Heike (2018), "Track Access Charges: Reconciling Conflicting Objectives: Case Study—Germany," Centre on Regulation in Europe, working paper, https://cerre.eu/sites/cerre/files/180509_CERRE_TrackAccessCharges_CaseStudy_Germany_Fin (accessed April 13, 2021).

[35] Network rail (2021). "How we are funded," https://www.networkrail.co.uk/who-we-are/how-we-work/how-were-governed-and-managed/how-were-funded/ (accessed October 18, 2021).

[36] Pint, Ellen (1992). "Price-cap versus Rate-of-Return Regulation in a Stochastic-Cost Model," *RAND Journal of Economics*, Vol. 23, No. 4, pp. 564-578.

[37] Sappington, David (1983). "Limited Liability Contracts Between Principal and Agent," *Journal of Economic Theory*, Vol. 29, No. 1, pp. 1-21.

[38] Sappington, David and David Sibley (1988). "Regulating Without Cost Information: The Incremental Surplus Subsidy Scheme," *International Economic Review*, Vol. 29, No. 2, pp. 297–306.

[39] Statistica (2018). "Annual Revenue of ProRail in the Netherlands in 2018, by Type of Income," https://www.statista.com/statistics/657119/annual-revenue-of-prorail-in-the-netherlands-by-type-of-income/ (accessed November 22, 2021).

[40] Vogelsang, Ingo and Jorg Finsinger (1979). "A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms," *Bell Journal of Economics*, Vol. 10, No. 1, pp. 151–171.