DRAFT
COMMENTS WELCOME
April 2003

Good Regulatory Lags in Price Cap and Rolling Cap contracts[1]

Andrés Gómez-Lobo
Department of Economics
University of Chile

and

José Luis Lima
Department of Economics
University of Chile

Abstract

Price caps are a popular form of monopoly price regulation. One of its disadvantages is the perverse incentives that regulated firms might have to scamp on cost reducing effort during the last years before a price review. In order to avoid this problem a "rolling cap" contract was introduced in the United Kingdom that overcomes this last problem. In spite of their popularity, there is scant research on the optimal regulatory lag (number of years between price reviews) of a price cap or rolling cap contract. In practice, around the world most price cap or rolling cap contracts have a lag of 4 to 5 years, but this is not based on any optimality consideration. As is well known, the regulatory lag determines the power of an incentive contract and thus the incentives to undertake cost reducing effort.

Schmalensee (1989) studied the optimal power of regulatory contracts in a static model with uncertainty and asymmetric information. She finds that medium powered contracts are generally superior to the polar cases of high or low powered contracts. In this paper, we extend Schmalensee (1989) model used to study the optimal power of regulatory contracts to a dynamic framework. We use numerical simulation to study the optimal regulatory lag for different combinations of demand and cost parameters under a particular linear quadratic structure. We find that in general a 2 year lag is optimal under both a price cap and rolling cap contracts and that a benevolent regulator prefers the rolling cap over the price cap contract in almost all the cases.

Key Words: Price Cap, Rolling Cap, Regulatory Lag, Dynamic Programming.
JEL Classification: C61, D81, L50, L52

[1] We thank comments by Dante Contreras. The usual disclaimers apply.

1. Introduction

This paper analyzes the optimal time lag between price reviews in natural monopoly regulation. Most countries that adopted incentive based regulation during the liberalization and privatization wave of the eighties and nineties applied a fixed regulatory lag of 4 to 5 years between price reviews. In countries such as Chile, a pioneer in privatization of public utilities, tariffs are set in real terms for a five year period in the water, electricity and telecoms sector. In Argentina, price caps were adopted in almost all concession contracts with price reviews every five years (except for the first period in some sectors). In the United Kingdom's RPI-X price cap system, price reviews for the gas, electricity and water sectors occur every five years while in telecoms and rail it is four years. In all of these cases, the length of the period between price reviews seems to have been adopted more by convention and administrative convenience rather than careful consideration of the economic costs and benefits of different lag periods.

The idea behind price cap regulation is that by fixing prices for a period of time, firms would be residual claimants to profits generated by cost reducing effort (or would suffer the losses from cost increases) and thus would have strong incentives to increase productive efficiency (Littlechild, 1983). A price cap regime with a fixed lag between reviews would overcome some of the inefficiencies purported to characterize traditional regulatory schemes such as rate of return regulation (sometimes also called cost plus regulation). In the parlance of incentive theory a pure cap regime would be a 'high powered' mechanism while rate of return regulation is a 'low powered' scheme.

In spite of the clear incentive properties of a price cap regime, under asymmetric information there are also disadvantages to this high powered regulatory contract. A regulator may not know the cost reduction potential of a company and may set prices too high.[2] This will harm consumers and generate allocative inefficiencies since prices would

[2] If a regulator has a prior belief regarding the cost potential of the firm and has to guarantee a non-negative profit for all types of firms, he then has to set tariffs according to the upper bound of the distribution of beliefs. If the firm turns out to be more efficient than this pessimistic assumption then it will earn an above average profit rate.

be above costs until the next price review. A similar phenomenon occurs if there are unexpected cost shocks that affect the firm's costs after tariffs are set. Prices will again be out of line with costs until the next price review.

That regulators often underestimate the capacity of firms to reduce costs can be shown by a series of examples. In the first price review after privatization of the United Kingdom water sector, the regulator believed companies' costs would rise during the next five years as a consequence of new environmental regulations imposed by the European Union. In accordance, the regulator set the X factor in the RPI-X formula to be -2%. Thus, water charges *increased* by 2% in real terms during the period 1995-1999. Ex-post it turned out that the regulator had underestimated how much companies could cut costs. Operating costs were reduced by 12% during the period. Users suffered the consequences of higher than necessary charges for an extended period of time. The high rates of return earned by privatized Chilean public utilities during the 90's also attest to the difficulty regulators have in setting prices at their efficient level.[3]

The optimal regulatory mechanism will trade-off the incentive properties of a high-powered scheme with the allocative, distributive and rent extracting properties of a lowered powered one. Theory suggests that the optimal regulatory contract under asymmetric information is to offer regulated firms a menu of contracts, the majority of which are medium powered (Laffont and Tirole, 1993).[4] If only one contract can be offered, the results of Schmalensee (1989) suggest that a medium tolow contract is optimal. In these models, the power of the contract is given by the parameters of a profit sharing rule. Price-cap and cost-plus regulation are polar cases of this profit sharing rule.

One reason why the above theoretical results have not had much eco in regulatory practice stems from the administrative difficulties associated with the implementation of a profit

[3] See the articles in Meller (2002) for a review of the Chilean regulatory experience during the last decade. Rates of return in the electricity and the regulated telecom sectors were very high, reaching 50% in some cases and with 30% being common for some companies. The water sector was privatized during the late nineties so it cannot be compared.

sharing rule. Regulators would have to monitor profits on a constant basis, rather than once during the price review period. There are many ambiguities in the practical measure of profits and ample accounting discretion can be used to manipulate these figures. Under a profit sharing scheme there would be strong incentives for regulated companies to use profit accounts as a strategic variable. Regulators would need stronger auditing capacities than under price-cap regulation. The latter is supposed to be more forward looking (projected future profits matter more than past profits in setting prices) in comparison to a profit sharing scheme which is dependent on past profits.

Thus, it would seem that only the polar cases of a high powered price-cap contract or a low power rate of return scheme are relevant for practical applications and the results of the theory of regulation would be irrelevant. However, this is incorrect. It is well known that one can alter the power of a price-cap contract by changing the regulatory lag period between price reviews. In the limiting case where price reviews are undertaken on a continuous basis, the price-cap regime collapses to a pure cost-plus regime. A longer regulatory lag increases the power of a regulatory scheme, providing more incentives to firms to undertake cost reducing effort. In a pure price-cap regime, where regulated firms face maximum incentives for cost reductions, the regulatory lag is infinite. Existing price-cap contracts are essentially of intermediate power since the regulatory lag is finite (4 to 5 years).

Rolling cap contracts are a variant of price caps introduced recently in the United Kingdom. Under a price cap regime the observed costs of a firm during the last years before a price review have a strong influence on the prices set for the next period. Therefore the firm's incentives to reduce costs are weakened as the price review period approaches. In order to eliminate this bias, under a rolling price regime a firm is allowed to keep any cost reduction for a fixed number of years irrespective of whether there is a price review in between. For example, if the regulatory lag is five years and a firm reduces its costs on the third year into

[4] The model proposed by Baron and Myerson (1982) also results in a menu of contracts offered. However, in this case the regulator is assumed not to observe costs and therefore the only mechanism to extract rents are through price distortions.

the price period, then this cost reduction will not be reflected in tariffs until the third year of the next pricing period.

Notice that the change from a price cap to a rolling cap implies an increase in the power of the regulatory scheme. In view of the trade-off between incentives and allocative efficiency, if the regulatory lag before this change was optimal, then the lag should be shortened when a rolling cap is introduced. This was not done in the UK. Therefore, either the lag was not considered to be optimal initially or it is currently suboptimal. This paper will shed some light as to the quantitative importance of shortening the regulatory lag when a rolling cap contract is introduced.

To date little research has been undertaken to determine whether the regulatory lags of existing regulatory contracts are optimal not in light of the economic trade-offs emphasized in the theoretical literature. Besides Armstrong, Rees and Vickers (1995) not much has been written explicitly on this topic. However, all of the literature on the optimal power of incentive contracts bears on this issue. Schmalensee (1989) is noteworthy in this sense. She examined, using numerical methods for a matrix of parameter values, the optimal cost sharing rule for a simple linear contract. As mentioned above his result was that the optimal contract was in most cases of medium power (neither a pure cost-plus nor a pure price-cap). However the static nature of his model makes it difficult to inform real world policy questions.

In this paper we extend Schmalensee's model to a dynamic setting to study 'good' regulatory lags using the same parameter matrix as in his paper. We are thus able to obtain conclusions that may be more readily applied to real policy questions. Changing the regulatory lag of a price-cap contract or even offering a menu of price-cap contracts with different regulatory lags is probably easier than introducing a profit-sharing regulatory contract.

Even the simple static model presented in Schmalensee (1989) is computationally demanding. Its extension to a dynamic setting increases the number of parameters of the model and raises a series of technical difficulties. We solve these difficulties using recursive methods to determine the optimal behavioral variables of the model and use numerical techniques to find quantitative solutions.

The paper is organized as follows: Section 2 introduces the general framework we use to study the monopoly and regulator problems under different regulatory regimes. In section 3 we made some particular assumptions about the functional form of demand, disutility of effort, costs, law of motion of the shock and regulator's beliefs so the problem meet the conditions to be analized as a Linear Quadratic Problem. We also show some interesting results about the observed behaviour of the effort exerted under different regimes. In section 4 we define the parameter values and the algorithms to be used on the simulations. In section 5 we place some results obtained under the last section particular conditions and conclude.

2. The model

Time is discreet. The Regulator and the monopolist sign a contract at t = 0. The contract lasts for infinite periods with price revisions every J periods, where J is the "regulatory lag". The Regulator can also opt to not intervene in the industry and let the monopoly free to fix prices. From t = 1 the Monopoly starts production and sells, taking as given the parameters fixed by the regulatory contract.

<u>Monopoly's Problem:</u>

A risk neutral monopolist with constant returns to scale faces the dynamic problem:

$$\Pi(\varphi, J) \equiv \underset{\{e_t\}}{Max} \quad E_1\left[\sum_{t=1}^{\infty} \delta^{t-1}\{(p_t - c_t)Q(p_t) - \psi(e_t;\varphi)\}/\varepsilon_0\right]$$

$$\begin{aligned} s.t.\quad & c_t = c_t(c_{t-1}, e_t, \varepsilon_t) \qquad , \quad c_{t1} > 0 \quad c_{t2} < 0 \quad , \quad c_t \geq 0\ \forall t \\ & \varepsilon_t \ \textit{follows a 1st. order stationary Markov process} \\ & c_0\ , \quad \{p_t\}(\textit{if regulated}) \quad \textit{and} \quad J \ \textit{given} \\ & Q' < 0; \quad \psi' > 0, \quad \psi'' > 0, \quad \psi(0;\varphi) = 0 \end{aligned}$$

where p_t *is* the price fixed by the regulator (if she lets the monopoly fixes the price, this is also part of its decision variables); *J is* the regulatory lag of the contract (if the monopoly fixes the price, *J=0*); e_t *is* the monopoly's effort level in reducing unit costs at t; c_t *is* the unitary cost of producing *Q* units of a *homogeneous no storable good*, which depends on the last period unit cost, on the current period effort level and on a stochastic shock ε_t that is materialized immediately after e_t is exerted: this means that the decision of the current period effort level has to be made in terms of the expected value of the unit cost at t, e.g. $E_t(c_t / \varepsilon_{t-1})$, that makes sense if there exits some kind of short term planning.

Because of the constant returns to scale the unit cost equals both marginal and mean production costs. We absorb from the existence of sunk costs that generates natural barriers to entry in the industry. The stochastic shock follows a stationary 1^{st} order markov process. $Q(p_t)$ is the demand function the monopoly faces at the market, whose functional form is constant and depends negatively on the current period price. $\psi(e_t; \varphi)$ is a concave function that represents the pecuniary cost or contemporary disutility of effort that the manager of the monopoly suffers when carrying out e_t; the intensity of the disutility depends on $\varphi \in \Re_+$ that determines the different types of monopolies the regulator may face. $\Pi(\varphi, J)$ is the discounted present value of the net benefits that a monopoly type φ with price revisions every J periods expects.

The timing of the problem at each period is the following:

- The firm decides and executes an effort level (and decides the price when it keeps unregulated) before ε_t is observed.
- Once the effort is carried out (and the price is chosen), ε_t is materialized and observed.
- The monopoly produces at the current cost and sells in the market all the quantity that is needed to satisfy current demand, and obtains profits net of the disutility of effort.

<u>Regulator's Problem:</u>

There is a risk neutral benevolent regulator with a priori beliefs about the efficiency of the firm (φ) summarized by a continuous distribution function *f(φ).* The Regulator may offer to the monopoly two contractual forms: a Price Cap and a Rolling Cap contract (Cost Plus is possible when J = 1).

A Price Cap contract establishes price revisions every J periods: at the revision period the regulator can observe the previous periods total costs and production levels and use that information to determine the unit cost that will be established as the new price that will prevail from that period until the next revision.

In a Rolling Cap contract, the regulator fixes the current period price equals to the unit cost realized J periods before (during the first J periods immediately after the sign of the contract, the price equals c_0).

Like Armstrong et. al. (1995) we implicitly assume that the regulator can commit to respect the price fixed at each revision until the next one. We also assume that *before the sign of the contract the Regulator chooses a J for each contractual form and commits to respect it forever.* There are two reasons for this and to not find a sequence $\{J_t\}$ contingent to the last observed and relevant unit cost:

- The Regulator is usually a governmental agency which may have different objectives than that of a benevolent regulator. It could also happen that the monopoly "captures" the agency and make her fixes a J that permits the former to obtain higher benefits[6]. Avoiding high discretion in regulator's decisions may result in better results. It would also prevent the waste of scant resources the monopoly may be interested to spend to obtain higher lags and softer control (rent-seeking and even corrupt practices).

---

[5] That makes sense if there exists a legal norm demanding that the regulated firm self finances its operations at every period (mean cost tariffication).

[6] It can be observed continuously higher $J_t s$ that allows the monopoly to be the residual plaintiff of its cost savings for more time than the socially optimal one, maybe without a significant descent in observed prices

- ♦ The second reason is that although having a single J can generate potential dynamic inconsistency problems, because a benevolent regulator may wish to diminish or increase J in face of new information about costs and shock, the quality of the new information about costs (and shock) may be bad, such that it can be better to have ex ante a fixed rule that obtains on average an acceptable reduction in costs.

If for these or other reasons a fixed rule for J is preferred, then it will still be important that the regulator commits her to respect it for the whole relationship in a credible way if she wants to keep reputation when fixing regulatory lags for future contracts in other industries. This doesn't mean that, according to the characteristics of an industry, the Regulatory lag differs from one industry to another; what really matters is for the fixed J to be respected.

The regulator doesn't know in advance the monopoly's type (φ) at the time she chooses the regulatory lag that optimizes her expected value function, so she has to do it according to the expected value on the distribution of types too. Her objective value function, called *W*, is composed by the weighted sum of the Expected Discounted Present Value of the Consumer's Net Surplus with the Expected Discounted Present Value of the Monopoly's Net Benefits:

$$W = \max_{J} \int_{\varphi \in \Re_+} \left[\, E.\,P.\,V.\,Net\ Cons.\,Surplus + \lambda\, \Pi(\varphi, J) \,\right] f(\varphi) d\varphi \quad , \quad 0 \le \lambda \le 1$$

$$s.t. \quad \Pi(\varphi, J) \ge 0 \quad \forall \varphi$$

$$E.\,P.\,V.\,Net\ Cons.\,Surplus = E_1\left[\sum_{t=1}^{\infty} \delta^{t-1} \left\{ \int_0^{p_t} Q(\widetilde{p}_t)\, d\widetilde{p}_t - p_t Q(p_t) \right\} / \,\varepsilon_0 \right]$$

where λ measures the degree of importance that the regulator gives to the firm relative to the consumers. The regulator wishes the monopoly wants to participate, but because she doesn't know a priori the monopoly's type she should satisfy the Participation Constraint of every possible type.

The weight λ indicates that there exist reasons so that the regulator worries more about consumer's surplus than for the firm's benefits. Some of them can be subjective or political but others can have economic meaning: a) allocative efficiency requires prices equal to marginal cost at every period, hence, fixing prices for J periods can improve only productive efficiency if demand is not perfectly inelastic (Armstrong et. al., 1995). b) The weight can reflect the existence of hidden costs in obtaining and processing the necessary information to fix prices at each revision (audits and the maintenance of a regulatory agency are costly) that has to be covered by means of distortionary taxes.

The regulator doesn't know in advance which type of monopoly she'll have in front, as well as she cannot identify it after because she will not be able to observe the effort and shock composition of the cost level at any revision. This is due because both price cap and rolling cap contracts, even though they give strong incentives to carry out effort they are not designed as truthful revealing mechanisms (that is evident with the absence of an incentive compatibility restriction in the regulator's problem). Hence, the regulator will never be sure about the monopoly's type and as a consequence her beliefs, in an extreme case used here, won't be revised.

It is nevertheless assumed that the resulting unitary cost can be fully identified at each revision period as well as corroborated and audited. Even though Baron and Myerson (1982) suggest that there may exist some degree of asymmetry even in the cost information that manages the regulator and the monopoly that favours the later, we follow Laffont and Tirole (1986) and assume that costs are observable.

The regulator also calculates the *W* of not regulating the monopoly and compares it with the *W* of regulating the monopoly with Price Cap or Rolling Cap. As the functional forms are the same ones at each case, direct comparison of the *Ws* from each regulatory regime will show us which option is better for the regulator.

The initial cost and shock, the functional forms of the (no stochastic) demand, disutility of effort, unitary costs and the stochastic process for ε_t are common knowledge at the time of the sign of the contract. The Regulator and the monopoly share the same discount rate.

2.1 Unregulated Monopoly's Problem

If the monopoly was not regulated it could exercise market power through the election of the price at every period. Supposing that the regulator studies the possibility of not regulates the monopoly, the problem that a monopoly type φ has to solve is:

$$\underset{\{e_t, p_t\}}{Max} \quad E_1\left[\sum_{t=1}^{\infty}\delta^{t-1}\{(p_t - c_t)Q(p_t) - \psi(e_t;\varphi)\}/\varepsilon_0\right]$$

$$\begin{aligned} s.t. \quad & c_t = c_t(c_{t-1}, e_t, \varepsilon_t) \quad , \quad c_{t1} > 0 \quad c_{t2} < 0 \quad , \quad c_t \geq 0 \; \forall t \\ & \varepsilon_t \textit{ follows a stationary } 1st \textit{ order Markov process} \\ & c_0 \textit{ and } \varepsilon_0 \textit{ given} \\ & Q' < 0; \quad \psi' > 0, \quad \psi'' > 0, \quad \psi(0;\varphi) = 0 \end{aligned}$$

We have supposed that both effort and price are chosen before the contemporary shock is materialized. Given the monopoly's objective function, the price decision can be obtained by solving the problem at every period, e.g.:

$$\underset{\{e_t, p_t\}}{Max} \quad E_t\left[(p_t - c_t)Q(p_t) - \psi(e_t;\varphi)/\varepsilon_{t-1}\right]$$

$$\begin{aligned} s.t. \quad & c_t = c_t(c_{t-1}, e_t, \varepsilon_t) \quad , \quad c_{t1} > 0 \quad c_{t2} < 0 \quad , \quad c_t \geq 0 \; \forall t \\ & \varepsilon_t \textit{ follows a stationary } 1st \textit{ order Markov process} \\ & c_{t-1} \textit{ and } \varepsilon_{t-1} \textit{ given} \\ & Q' < 0; \quad \psi' > 0, \quad \psi'' > 0, \quad \psi(0;\varphi) = 0 \end{aligned}$$

that for p_t results in the Lerner rule for a single product monopoly (defining $\hat{c}_t = E_t(c_t / \varepsilon_{t-1}) = \hat{c}_t(c_{t-1}, e_t, \varepsilon_{t-1})$):

$$\frac{p_t^M - \hat{c}_t}{p_t^M} = \frac{1}{\left|\eta_{P_t^M}\right|} \quad , \quad \eta_{P_t^M} = \frac{p_t^M Q'(p_t^M)}{Q(p_t^M)}$$

From this rule we can determine that the monopoly price will be a function of the level of the expected cost at every period ($p_t^M = p_t^M(\hat{c}_t)$) that is at the same time a function of the level of contemporary effort.

Having found the way in which the monopoly fixes prices, we now solve for its level of effort. In this case, the monopoly's problem can be expressed by means of dynamic programming. Let the state variables at each period be the previous period cost and shock, and let the effort at each period be the decision variable. Then we can define the functional $V(c_{t-1},\varepsilon_{t-1})$ that summarizes the monopoly's optimized problem at t, as:

$$V(c_{t-1},\varepsilon_{t-1}) = \max_{e_t} E_t\left[\left\{(p_t^M - c_t)Q(p_t^M) - \psi(e_t;\varphi)\right\}/\varepsilon_{t-1}\right] + \delta E_t\left[V(c_t,\varepsilon_t)/\varepsilon_{t-1}\right]$$

$$\begin{aligned} s.t.\quad & c_t = c_t(c_{t-1},e_t,\varepsilon_t) \quad , \quad c_{t1} > 0 \quad c_{t2} < 0 \quad , \quad c_t \geq 0\ \forall t \\ & \varepsilon_t \textit{ follows a stationary } 1st \textit{ order Markov process} \\ & c_{t-1} \textit{ given} \quad , \quad Q' < 0; \quad \psi' > 0, \quad \psi'' > 0, \quad \psi(0;\varphi) = 0 \end{aligned}$$

The FOC of this problem together with the Benveniste-Sheinkman conditions, gives rise to the Euler equation that implicitly determines the effort that the monopoly will carry out at every period (we define $\hat{c}_{t+i} \equiv E_{t+i}(c_{t+i}/\varepsilon_{t+i-1}) = \hat{c}_{t+i}(c_{t+i-1},e_{t+i},\varepsilon_{t+i-1})$, $i \geq 0$):

$$E_t\left[\sum_{i=0}^{\infty}\delta^i\left\{\left(\frac{\partial p_{t+i}^M}{\partial \hat{c}_{t+i}} - 1\right)Q(p_{t+i}^M) + (p_{t+i}^M - \hat{c}_{t+i})Q'(p_{t+i}^M)\frac{\partial p_{t+i}^M}{\partial \hat{c}_{t+i}}\right\}\frac{\partial \hat{c}_{t+i}}{\partial e_t}\right] = \psi'(e_t)$$

$$iff \quad \lim_{j\to\infty}\delta^{j-1}E_t\left[\frac{\partial V(c_{t+j},\varepsilon_{t+j})}{\partial c_{t+j}}\frac{\partial c_{t+j}}{\partial c_{t-1}}\right] = 0$$

the first equation establishes the equality among the Present Value of the net marginal benefit of an increase in effort at period t, and the marginal cost incurred by the manager to exercise it. The effort carried out today influences marginal benefits both today and in the future by affecting present and future unitary costs levels that influence present and future price decisions, also affecting present and future delivered quantities through its effect on prices. The limit condition establishes that the monopoly doesn't have to expect extraordinary earnings or losses if there is a variation in the level of costs it begins with at every period (it must not discontinuously rise or low the levels of future costs).

2.2 *Monopoly's Problem under Price Cap*

Under Price Cap the monopoly calculates its benefit based on the announcement of J from the regulator at the moment of the sign of the contract. We will assume that at each revision the regulator will fix the price that the monopoly will charge to the public until the next revision equals to the last unitary cost.[7]

To model this problem it is convenient to use a special notation: let x_t^τ be the value that takes the variable x at period t after having passed τ price revisions; if the lapse of time between revisions lasts for J periods, t can take values from 1 to J; τ takes values from 1 to infinite (it is considered that the first price revision happens at the moment of the sign of the contract when the regulator fixes the initial price equals to c_0). Let also E_t^τ be the mathematical operator of the expected value at period t after τ revisions.

With this notation we can define the monopoly's problem like this:

$$\underset{\{e_t^\tau\}}{Max} \quad E_1^1\left[\sum_{\tau=1}^{\infty}\delta^{J(\tau-1)}\sum_{t=1}^{J}\delta^{t-1}\left\{(c_0^\tau - c_t^\tau)Q(c_0^\tau) - \psi(e_t^\tau;\varphi)\right\}/\,\varepsilon_0^1\right]$$

$$\begin{aligned} s.t.\quad & c_t^\tau = c_t^\tau(c_{t-1}^\tau, e_t^\tau, \varepsilon_t^\tau) \quad , \quad c_{t1} > 0 \quad c_{t2} < 0 \quad , \quad c_t^\tau \geq 0 \ \forall t,\tau \\ & \varepsilon_t^\tau \ \textit{follows a stationary 1st order Markov process} \\ & c_0^1 \ \textit{and} \ \varepsilon_0^1 \ \textit{given} \\ & Q' < 0; \quad \psi' > 0, \quad \psi'' > 0, \quad \psi(0;\varphi) = 0 \end{aligned}$$

We also make the following definitions that complete the transition of costs and technological shocks from one revision period to another: $c_J^\tau \equiv c_0^{\tau+1}$, $\varepsilon_J^\tau \equiv \varepsilon_0^{\tau+1}$. In order to simplify the complexity of the problem we also make the following assumption: the monopoly decides or plans the sequence of effort to carry out for the following J periods after each revision, using only the information contained in c_0^τ and ε_0^τ (that is reasonable if there is some kind of medium term planning). The problem of the monopoly can be expressed through dynamic programming, defining the functional that summarizes the present value of the benefits in the following way:

$$V(c_0^\tau, \varepsilon_0^\tau) = \max_{\{e_t\}_{t=1}^J} E_1^\tau \left[\sum_{t=1}^{J} \delta^{t-1} \left\{ (c_0^\tau - c_t^\tau) Q(c_0^\tau) - \psi(e_t^\tau; \varphi) \right\} / \varepsilon_0^\tau \right] + \delta^J E_1^\tau \left[V(c_0^{\tau+1}, \varepsilon_0^{\tau+1}) / \varepsilon_0^\tau \right]$$

$$\begin{aligned} s.t.\quad & c_t^\tau = c_t^\tau(c_{t-1}^\tau, e_t^\tau, \varepsilon_t^\tau) \quad , \quad c_{t1} > 0 \quad c_{t2} < 0 \quad , \quad c_t^\tau \geq 0\ \forall t \\ & \varepsilon_t^\tau \text{ follows a stationary 1st order Markov process} \\ & c_0^\tau \text{ and } \varepsilon_0^\tau \text{ given}; \quad Q' < 0; \quad \psi' > 0, \psi'' > 0, \psi(0;\varphi) = 0; \quad c_J^\tau \equiv c_0^{\tau+1}; \quad \varepsilon_J^\tau \equiv \varepsilon_0^{\tau+1} \end{aligned}$$

The FOC originates a system of equations for the levels of effort in every period until the next revision, summarized by (define $\hat{c}_t^\tau \equiv E_1^\tau(c_t^\tau / \varepsilon_0^\tau) = \hat{c}_t^\tau(c_{t-1}^\tau, e_t^\tau, \varepsilon_0^\tau)$):

$$e_i^\tau : \quad -\sum_{t=i}^{J} \delta^{t-i} \frac{\partial \hat{c}_t^\tau}{\partial e_i^\tau} Q(c_0^\tau) - \psi'(e_i^\tau) + \delta^{J+1-i} E_1^\tau \left[\frac{\partial V(c_0^{\tau+1}, \varepsilon_0^{\tau+1})}{\partial c_0^{\tau+1}} \frac{\partial c_0^{\tau+1}}{\partial e_i^\tau} \right] = 0; \quad i = 1, \ldots, J$$

and using the conditions of Benveniste-Sheinkman we can obtain the following expression that implicitly defines the effort at every period:

$$\begin{aligned} e_i^\tau : \quad & -\sum_{t=i}^{J} \delta^{t-i} \frac{\partial \hat{c}_t^\tau}{\partial e_i^\tau} Q(c_0^\tau) + \delta^{J+1-i} E_1^\tau \Bigg[\sum_{m=1}^{\infty} \delta^{J(m-1)} \Bigg\{ \sum_{t=1}^{J} \delta^{t-1} \\ & \left((1 - \frac{\partial \hat{c}_t^{\tau+m}}{\partial c_0^{\tau+m}}) Q(c_0^{\tau+m}) + (c_0^{\tau+m} - \hat{c}_t^{\tau+m}) Q'(c_0^{\tau+m}) \right) \frac{\partial c_0^{\tau+m}}{\partial e_i^\tau} \Bigg\} / \varepsilon_0^\tau \Bigg] = \psi'(e_i^\tau) \\ & i = 1, \ldots, J. \end{aligned}$$

$$iff \lim_{m \to \infty} \delta^{Jm} E_1^\tau \left[\frac{\partial V(c_0^{\tau+m}, \varepsilon_0^{\tau+m})}{\partial c_0^{\tau+m}} \frac{\partial c_0^{\tau+m}}{\partial c_0^\tau} / \varepsilon_0^\tau \right] = 0$$

---

7 As in Schmalensee (1989) we assume that the fixed price serves as a roof as well as a floor.

The first equation establishes the equality among the Expected Present Value of the Net Marginal Benefit of exercising additional effort with the marginal pecuniary cost it causes. The limit condition establishes that the monopoly doesn't expect extraordinary future earnings or losses for a small variation in the level of costs at the beginning of any price revision.

2.3 *Monopoly's Problem under Rolling Cap*

Under a Rolling Cap contract the monopoly takes as given the announcement of J made by the regulator at the moment of the sign of the contract. The regulator also fixes the price for each period equals to the unitary cost obtained by the monopoly J periods back. As there is no information of the corresponding past costs during the first J periods of the contract, the regulator fixes the prices for those periods equal to c_0.

We can define the problem of the monopoly as:

$$\begin{aligned}
&\underset{\{e_t\}}{Max} \quad E_1\left[\sum_{t=1}^{\infty}\delta^{t-1}\left\{(c_{t-J}-c_t)Q(c_{t-J})-\psi(e_t;\varphi)\right\}/\varepsilon_0\right] \\
&\quad s.t. \quad c_t = c_t(c_{t-1},e_t,\varepsilon_t) \quad , \quad c_{t1}>0 \quad c_{t2}<0 \quad , \quad c_t \geq 0 \ \forall t \\
&\qquad\quad \varepsilon_t \ \textit{follows a stationary } 1st \textit{ order Markov process} \\
&\qquad\quad c_m \textit{ and } \varepsilon_m \textit{ given for } m \leq 0 \\
&\qquad\quad Q'<0; \quad \psi'>0, \quad \psi''>0, \quad \psi(0;\varphi)=0
\end{aligned}$$

This problem can also be expressed with dynamic programming:

$$\begin{aligned}
&V(c_{t-J},c_{t-1},\varepsilon_{t-1}) = \max_{e_t} E_t\left[\left\{(c_{t-J}-c_t)Q(c_{t-J})-\psi(e_t;\varphi)\right\}/\varepsilon_{t-1}\right]+\delta E_t\left[V(c_{t-J},c_{t-1},\varepsilon_t)/\varepsilon_{t-1}\right] \\
&\quad s.t. \quad c_t = c_t(c_{t-1},e_t,\varepsilon_t) \quad , \quad c_{t1}>0 \quad c_{t2}<0 \quad , \quad c_t \geq 0 \ \forall t \\
&\qquad\quad \varepsilon_t \ \textit{follows a stationary } 1st \textit{ order Markov process} \\
&\qquad\quad c_{t-m} \textit{ and } \varepsilon_{t-m} \textit{ given for } m \leq 1 \quad , \quad Q'<0, \ \psi'>0, \psi''>0
\end{aligned}$$

The FOC of this problem together with the envelope conditions originates the expression that implicitly determines the optimal level of effort the monopoly carries out at every period, noticing that the value function has as arguments two different period levels of costs (define $\hat{c}_{t+i} = E_{t+i}(c_{t+i} / \varepsilon_{t+i-1}) = \hat{c}_{t+i}(c_{t+i-1}, e_{t+i}, \varepsilon_{t+i-1})$, for $i \geq 0$):

$$E_t\left[\quad -\sum_{i=0}^{J-1} \delta^i \left(\frac{\partial \hat{c}_{t+i}}{\partial e_t} \right) Q(c_{t+i-J}) + \right.$$
$$\left. + \sum_{i=J}^{\infty} \delta^i \left((1 - \frac{\partial \hat{c}_{t+i}}{\partial c_{t+i-J}}) Q(c_{t+i-J}) + (c_{t+i-J} - \hat{c}_{t+i}) Q'(c_{t+i-J}) \right) \frac{\partial c_{t+i-J}}{\partial e_t} \quad / \varepsilon_{t-1} \right] = \psi'(e_t)$$

$$Iff \ \lim_{m \to \infty} \delta^{J+m} \ E_t \left[\frac{dV(c_{t+m}, c_{t+J-1+m}, \varepsilon_{t+J-1+m})}{dc_{t-1}} / \varepsilon_{t-1} \right] = 0$$

The first equation equals the Present Value of the net Marginal Benefits of exercising additional effort in the current period, and the marginal pecuniary cost of carrying out this effort. The transversality condition has the same spirit as the first two cases.

3. Particular Case: A Linear Quadratic Approach

In order to get some results about the election of J, it is necessary to assume some functional forms for demand, disutility of the effort and unitary costs, as well as for the law of motion of the shock and the regulator's beliefs over types; it will also be necessary to use numerical methods to obtain them.

3.1 Functional form for demand.

A more general lineal structure than Schmalensee (1989) is assumed, that will allow us to better isolate the effect of the price elasticity of demand:

$$Q(p_t) = a - b\, p_t$$

If E_{p_t} is the demand elasticity when the price is p_t, then: $E_{p_t} = \frac{-b p_t}{a - b p_t}$.

3.2 *Functional form for the Disutility of effort.*

We assume a quadratic form for the disutility of the effort, a little different to that used in Schmalensee:

$$\psi(e_t, \varphi) = \frac{e_t^2}{2\varphi}$$

and we don't discard the possibility that the monopoly decides to exercise a negative level of effort in some period. Due to its quadratic form it implies that a given level of effort causes the manager the same disutility level even if it is positive or negative.

Intuition and literature suggest an asymmetric treatment depending on the sign of effort, giving it a higher weight when positive but a smaller or null weight when negative. However we have three reasons to prefer this functional form just as it is: The first one is that if the manager chooses to carry out negative effort at any period, in spite of the fact that it is also expensive for him, then it becomes clear that the incentives to do it are quite strong and we should see an even higher negative level when using asymmetric functional forms. The second reason is that negative effort can be interpreted as a deliberate decision of the monopoly's manager to make the costs go up, which implies that she should also make effort to obtain this with its respective objective and subjective costs.

The third reason is rather a numerical one, as it will be seen later on, so that the numerical solution of this particular case, using standard techniques, needs the matrix that accompanies the effort to be negative definite, and we make it sure with this functional form. In any case, trying to use some asymmetric functional form would make the problem unnecessarily complex for the effects of the question to be responded in this paper.

However it could be reasonable to expect that the monopoly's cost and its investments decitions are under tight regulatory control, so she cannot deliberatedly try to rise her costs.

As will become clear later this may affect the election of the regulatory lag under a Price Cap contract, so we expect to meet this posible limitation in subsequent work under a different particular structure.

3.3 Functional form for the Unitary Costs.

The unitary cost should commit at least 2 reasonable conditions: it must respond negatively to effort, and it should always be positive.

Because the general cost function assumes that the effort exercised at one period affects also the level of costs in subsequent periods, we can take two ways to model it (suppose for the moment that there is no random shocks): we can assume that the effect of effort is permanent, so a unit of effort exercised at one period diminishes proportionally both the contemporary and future levels of costs without having its effect disappear over time; the second way is to assume that as time goes the effect of a unit of effort carried out today will eventually disappear over time, like a sort of "depreciation" of effort.

As the common way of reducing costs is related with investments in new technology and equipment and/or with more efficient ways of resource administration, it should be expected that as time passes the firm incurs in additional costs of maintenance (the equipment may need a specialized and expensive technical body) and quick depreciation of high tech equipments; or in the case of using more efficient ways of administration, these cannot be exempt of continous surveillance, control, preparation and motivation of the company's human resources under the new outlines that maintains the efficiency gains.

The case of a permanent effect of effort is not intuitively reasonable, therefore we will use the focus of effort depreciation through time. This implies that the monopoly's effort will be split between two ends: a part that will maintain the level of cost reached in previous periods, and a part that will obtain a new cost reduction within the period.

Hence, we will use the following functional form for the unitary costs:

$$c_t = c_0(1-\theta_t) \quad , \quad \theta_t \leq 1 \quad \forall t$$

where c_0 *is* the initial cost level the monopoly begins with before the signature of the regulatory contract, and θ_t is the percentage cost reduction (increase) at t in respect of c_0. Notice that although $\theta_t \leq 1$ so the cost is positive, there is no restriction on θ_t to be nonnegative: the monopoly can exert negative effort or receive a shock such that the cost in t can overcome the initial cost. The dynamics of θ_t are given by:

$$\theta_t = \rho\theta_{t-1} + e_t + \varepsilon_t \quad , \quad 0 \leq \rho < 1 \quad , \quad \theta_0 = 0$$

where ρ picks the idea that only a fraction of the cost reduction gained at previous periods spends to the following (when $\rho = 0$ then each period effort affects costs just on that period, with a total reversion to c_0 at the beginning of the next one); e_t *is* the monopoly's effort level exerted at t; ε_t is the random shock materialized immediately after the effort is carried out at the corresponding period.

Note three important details in this specification:

1) $\rho < 1$ implies that when $e_t = 0$ then c_t will ascend towards c_0, and when $e_t < 0$ the cost will go up towards c_0 faster and more permanently;

[8] We can complicate a little bit the analysis by assuming that the cost reducing effort has a certain probability of success. Armstrong et. al. (1995) assumes that the probability in which cost decreases from one period to another is a function of the effort carried out at every period, making endogenous the probability of success. Another way to do it would be to assume an exogenous probability of success (independent of effort), as for example:

$$\theta_t = \rho\theta_{t-1} + k_t e_t + \varepsilon_t \quad , \quad 0 \leq \rho < 1 \quad , \quad \theta_0 = 0 \quad , \quad k_t = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

that would introduce an additional parameter (p) to include in the simulations. In this work we will assume that the effort is 100% effective ($p = 1$).

2) when placing ε_t directly on the θ_t's equation, we are implicitly assuming that it follows the same logic than effort; hence, $\varepsilon_t > 0$ diminishes the cost at t but its effect disappears in later periods, and $\varepsilon_t < 0$ raises costs faster. We can think on $\varepsilon_t > 0$ as a technological improvement in equipments that also depreciates, or a non waited increment in the efficiency of the administration (hiring a more laborious group of workers than expected, for example) that should also be fomented and controlled later on.

3) even though the former two don't represent a serious drawback for our purposes, there is another detail that is important and is given by the linearity of θ_t on ε_t , which is unfortunately necessary to apply the linear quadratic framework: depending on a particular history and realization of ε_t, θ_t may be greater than 1. Hence, we will only warrant for θ_t to be less than 1 in steady state when $\varepsilon_t = 0$ for all t.[9]

3.4 Law of motion for the shock in costs.

The shock follows a 1st order stationary Markov process that for continuous states can be represented by a 1st order stationary autoregressive process like this:

$$\varepsilon_t = \beta\,\varepsilon_{t-1} + \mu_t \quad , \quad 0 < \beta < 1 \quad , \quad \mu_t \sim N(0, \sigma^2)$$

where μ_t is an i.i.d. innovation occurred at t. The reason for the positive autocorrelation assumed in the process is that it is plausible that a persistent technological shock will keep its sign at every t.

3.5 Analytical expression for effort at each regulatory regime.

[9] We plan to meet this caveat on subsequent work, using a more general framework than the linear quadratic, but many interesting results can still be obtained.

Given the previous functional forms, the equation that determines the effort level for each regulatory regime takes the following form (let it be $\hat{c}_{t+i} \equiv E_{t+i}(c_{t+i} / \varepsilon_{t+i-1})$ and $\hat{c}_t^{\tau+m} \equiv E_t^{\tau+m}(c_t^{\tau+m} / \varepsilon_{t-1}^{\tau+m})$):

UNREGULATED MONOPOLY:

$$e_t = \varphi c_0 \, E_t\left[\sum_{i=0}^{\infty} (\delta\rho)^i \left(\frac{a - b\hat{c}_{t+i}}{2} \right) / \, \varepsilon_{t-1} \right] \qquad , \qquad \left(p_t^M = \frac{a + b\hat{c}_t}{2b} \right)$$

PRICE CAP:

$$e_i^\tau = \varphi c_0^1 \left\{ (a - b c_0^\tau) \frac{1 - (\delta\rho)^{J-i+1}}{1 - \delta\rho} + \delta(\delta\rho)^{J-i} E_1^\tau \left[\sum_{m=1}^{\infty} (\delta\rho)^{J(m-1)} \left\{ -(a - b c_0^{\tau+m}) \left(\frac{1-\delta^J}{1-\delta} - \right. \right. \right. \right.$$

$$\left. \left. \left. \left. - \rho \frac{1 - (\delta\rho)^J}{1 - \delta\rho} \right) + b \sum_{t=1}^{J} \delta^{t-1} (c_0^{\tau+m} - \hat{c}_t^{\tau+m}) \right\} / \, \varepsilon_0^\tau \right] \right\} \qquad , \qquad i = 1, 2, \ldots.., J$$

ROLLING CAP:

$$e_t = \varphi c_0 \, E_t\left[\sum_{i=0}^{J-1} (\delta\rho)^i \, (a - b c_{t+i-J}) + \delta^J \sum_{i=J}^{\infty} (\delta\rho)^{i-J} \left\{ -(1-\rho^J)(a - b c_{t+i-J}) + b(c_{t+i-J} - \hat{c}_{t+i}) \right\} / \, \varepsilon_{t-1} \right]$$

Even in this particular case it cannot be settle down analytically which regulatory regime generates a higher level of effort at every period. To understand the complexity of this task we can analyze the steady state of the non stochastic problem and prove that the effort under Rolling Cap is higher than that of an unregulated monopoly if $1 - \delta^J > 0.5$. Even in this case it is not possible to rank Price Cap with the others.

When we made the simulations we observed that the effort exerted by an unregulated monopoly and under Rolling Cap is always positive and quite stable at any period. Even though it is not evident in the effort expression for Price Cap, it is observed on the simulations that the expectational term is negative, which indicates that *in the expected effect of today's effort it weighs more the reduction in benefits due to the fall in fixed future prices than the increase in benefits due to increments in demanded quantity (because of littler fixed prices), and the former effect becomes stronger as the next revision approaches*: it makes the sequence of effort to decline between revisions, consistent with Armstrong et. al.

The effort level under Price Cap can be negative in those periods just before the next revision if the expectation term is sufficiently large: intuitively, because the cost used to fix prices is the last one obtained just before the revision, the monopoly (including stockholders as subsequent cost reduction becomes harder and requires the use of profits) has strong incentives to make it as high as possible so being less demanded for the following periods. The effort level is reduced below the minimum necessary to maintain the gain of previous period and even more so as to increase the level of costs.

Figures 3 and ***4*** illustrate the main characteristics of the effort that are consistently observed on all Price Cap parameterizations: the effort sequence always falls as the next revision period becomes closer *at an increasing rate*, and what is interesting to notice, complementary to Armstrong et. al. (1995), is that *as J becomes higher the absolute level of effort at each period increases*. Hence, the total surplus generated by the monopoly's effort will be smaller the lower is the time left to the monopoly to enjoy its cost reductions, a result similar in spirit to that of Williamson (1997).[10]

Notice that *the level of the negative effort is also encouraged by higher lags*. This is because higher lags result in effort sequences which are higher at initial periods, such that the monopoly enjoys the generated surplus for more periods before the revision; as the revision comes closer the best thing for him to do is to deliberately try to rise its costs: the higher the effort at initial periods the lower or even more negative is the effort needed to overcome it, such that the regulator fixes an starting high price level in the next revision.

[10] His work looks for the best proportion of surplus to be passed over to consumers, fixing J = 5, and finds that a higher proportion discourages monopoly's effort and reduces total surplus.

Figures 5 and ***6*** present the main differences between effort levels of the regulatory regimes that are consistently observed in the simulations. The lowest average level of effort is the one of the unregulated monopoly, even though it could be higher than that of a Price Cap before price revisions. Immediately after each revision, the Price Cap level of effort is the highest but quickly falls below the Rolling Cap level in subsequent periods. The Rolling Cap effort is higher on average than that of an unregulated monopoly at every period. We can also observe that the unregulated monopoly and Rolling Cap's effort are both positive and stable (lightly growing after the sign of the contract but stabilizing few periods ahead).

We can rationalize the lower average effort level of the unregulated Monopoly as follows: we have assumed that effort reduces the monopoly's unitary cost, and under constant returns it implies that it reduces its marginal cost too. The unregulated monopoly maximizes benefits at every period by equalizing expected marginal cost with marginal income, choosing quantity and sales price. Given an initial marginal cost the monopoly sets the price p_1^M and obtains benefits π_1^M; if marginal income is a decreasing function the effort will not only reduce marginal cost but also p_1^M. At one hand π_1^M increases because the smaller marginal cost allows the monopoly to sell to more consumers and get some surplus; and on the other hand π_1^M diminishes because total revenues fall because p_1^M falls. The net effect on π_1^M depends on the elasticity of demand. Under Price Cap and Rolling Cap the second effect is not immediate but can take some periods until the next price revision.

Also, under p_1^M the optimal quantity is smaller than under marginal cost pricing, so a cost reduction generates a proportionally smaller surplus for an unregulated monopoly than under Price Cap or Rolling Cap regimes. The sum of all these effects implies that the monopoly will have greater incentives to make effort under a Rolling Cap and at least during the first periods of a Price Cap regime.

Figure 3

First Sequence of Effort after the sign of a Price Cap contract under Parameter Values: $E = 1.8$, $\delta = 0.7$, $\lambda = 1$, $\sigma = 0.04$, $\beta = 0.95$, $d = 0.9$, $\rho = 0.7$, $D = 0.3$. Initial Shock $\varepsilon_0 = 0$. Regulatory Lag J = 4, 6 and 8

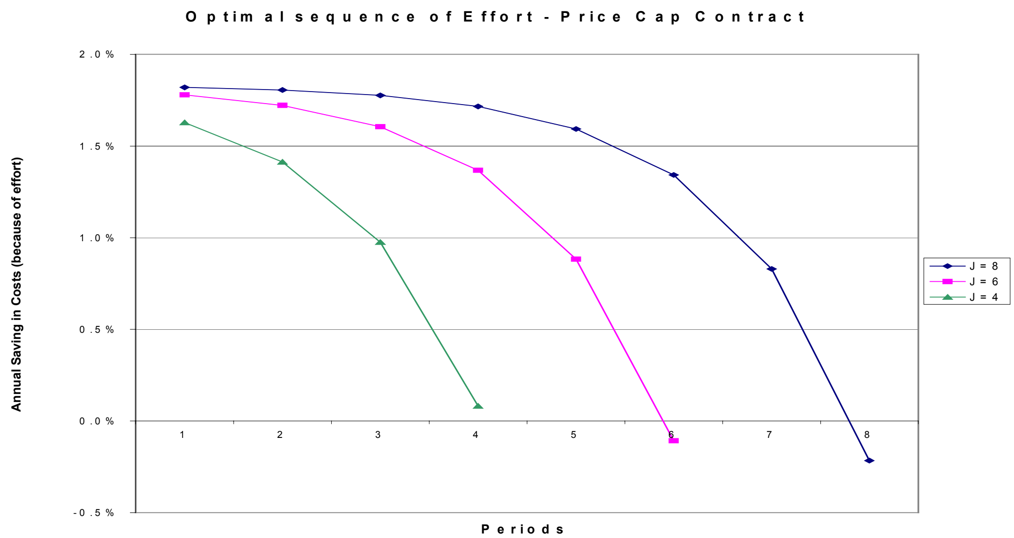


Figure 4

First Sequence of Effort after the sign of a Price Cap contract under parameter Values of: $E = 0.2$, $\delta = 0.5$, $\lambda = 0.25$, $\sigma = 0.04$, $\beta = 0.95$, $d = 0.9$, $\rho = 0.9$, $D = 0.05$. Initial Shock $\varepsilon_0 = 0$. Regulatory Lag J = 4, 6 and 8

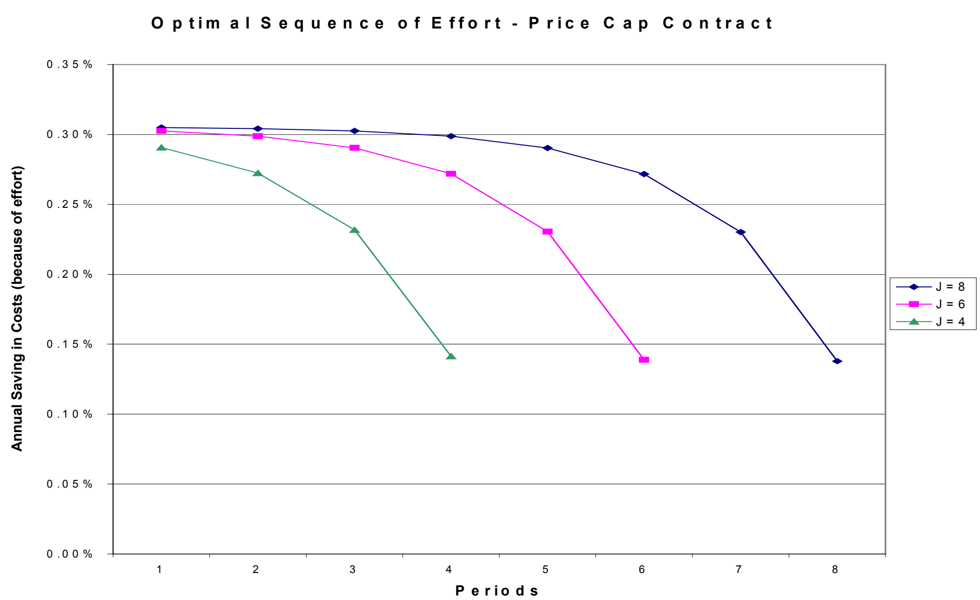


Figure 5

First Sequence of effort after the sign of a Price Cap and Rolling Cap contracts, and Unregulated Monopoly for Parameter Values: $E = 1.8$, $\delta = 0.7$, $\lambda = 1$, $\sigma = 0.04$, $\beta = 0.95$, $d = 0.9$, $\rho = 0.7$, $D = 0.3$. $\varepsilon_0 = 0$. J = 8.

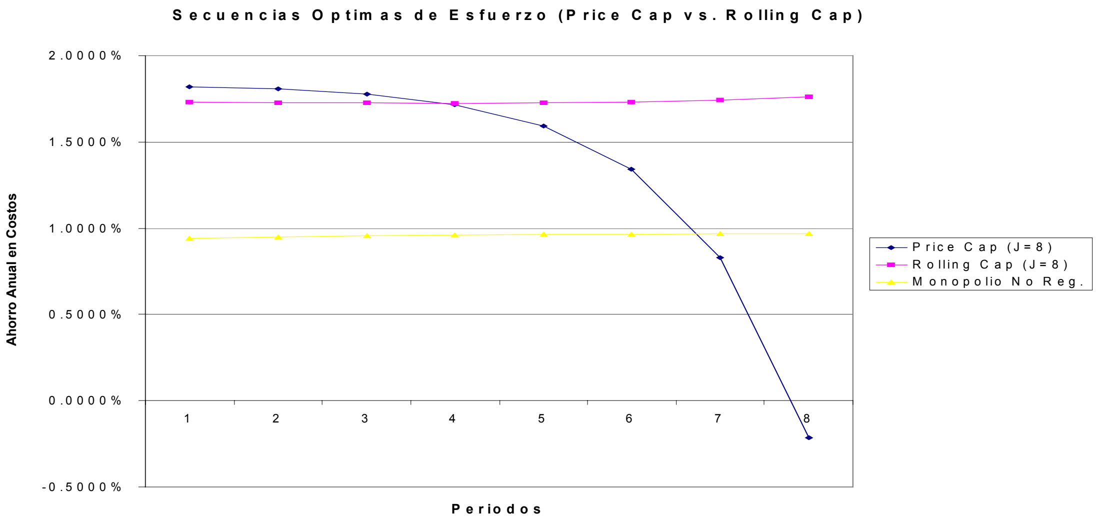
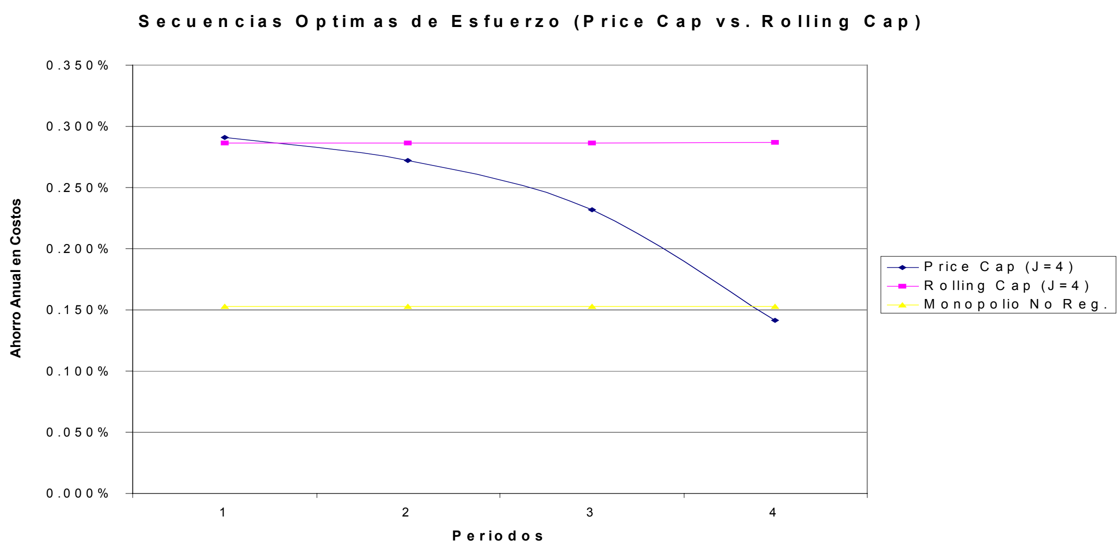


Figure 6

First Sequence of effort after the sign of a Price Cap and Rolling Cap contracts, and Unregulated Monopoly for Parameter Values: $E = 0.2$, $\delta = 0.5$, $\lambda = 0.25$, $\sigma = 0.04$, $\beta = 0.95$, $d = 0.9$, $\rho = 0.9$, $D = 0.05$. $\varepsilon_0 = 0$. J = 4.



---

11 ***The initial sequences of effort in Rolling Cap and Unregulated Monopoly are obtained assuming that $\varepsilon_t = 0$ for the corresponding periods, to make them comparable with those of the Price Cap.***

3.6 *Regulator's beliefs about the monopoly's efficiency types.*

As in Schmalensee (1989) we suppose the Regulator has beliefs over possible cost savings. In particular she thinks that the maximum percentage saving in costs that a monopoly can meet per year has a uniform distribution between D(1-d) and D(1+d), e.g. on average the monopoly can make a maximum percentage cost saving per annum of D%, with a minimum of D(1-d)% and a maximum of D(1+d)%, where d measures the uncertainty the regulator has about it.

To associate this belief with the possible monopoly's types, the regulator mentally solves a Pure Price Cap (or Pure Rolling Cap) placing $J \to \infty$, obtaining the effort level carried out on this situation and comparing it directly with his a priori distribution of maximum percentage cost saving.

Under Pure Price Cap (or Pure Rolling Cap), the resulting effort level for every period is constant and independent of the random shock, and is given by:

$$e_\infty = \frac{\varphi c_0 (a - b c_0)}{(1 - \delta\rho)}$$

The regulator uses this expression an her beliefs on the maximum saving in costs to obtain the implied distribution of types $f(\varphi)$, which is also uniform between $\varphi_{\min} = \omega D(1\text{-}d)$ and $\varphi_{\max} = \omega D(1+d)$, with $\omega = \dfrac{(1-\delta\rho)}{c_0 (a - b c_0)}$, that define the most inefficient and efficient type of monopoly respectively.

3.7 *Regulator's Problem.*

As we have already seen, when deciding the regulatory lag (*J*) the benevolent regulator has to maximize a social welfare function that may give different weights to consumer and monopoly's surplus (by means of λ). Using the assumed particular functions we can rewrite the problem of the regulator in the following way:

$$W = \max_{J} \int_{\varphi_{\min}}^{\varphi_{\max}} \left[E.\,P.\,V.\ Net\ Cons.\ Surplus\ (\varphi) + \lambda V(x_0\,;\varphi, J)\ \right] f(\varphi) d\varphi \quad , \quad 0 \leq \lambda \leq 1$$

$$s.t.\quad V(x_0\,;\varphi, J) \geq 0 \quad \forall \varphi$$

$$E.\,P.\,V.\ Net\ Cons.\ Surplus(\varphi) = E_1 \left[\sum_{t=1}^{\infty} \delta^{t-1} \left\{ \frac{(a - b\, p_t(\varphi))^2}{2b} \right\} / \varepsilon_0 \right]$$

where $V(x_0\,;\varphi, J)$ is the Value function of a monopoly type φ with price revisions every J periods at the moment of the signing of the contract. The vector x_0 contains the monopoly's state variables initial values, which are common knowledge; which variable is considered as a state will depend on the particular contractual relationship. Note that the expected Present Value of the Consumer's net surplus also depends on the price fixed by the regulator and so indirectly on the monopoly's type.

With regard to the participation constraints for all types of monopoly, because we are using the same functional forms as Schmalensee (1989) we know that *V(.)* grows with φ, so higher levels of efficiency are accompanied by higher benefits for the same level of the state variables. We know from the past expressions that at the optimum a higher φ implies a higher level of effort at every period and therefore a smaller unitary cost; the disutility of effort increases because of effort but also diminishes because of the increase in φ. That *V(.)* is growing in φ means that the smaller expected unitary cost at every period generates an increase of 1st. order in benefits, and that the disutility of the effort causes a 2nd order decrease, so the net effect is positive.

The previously exposed justifies the following objective function for the Regulator, used in the simulations:

$$W = \max_J \int_{\varphi_{\min}}^{\varphi_{\max}} [\; E.\,P.\,V.\; Net\; Cons.\; Surplus\,(\varphi) + \lambda V(x_0\,;\varphi, J)\;]\, f(\varphi)d\varphi \quad , \quad 0 \le \lambda \le 1$$

$$s.t. \quad V(x_0\,;\varphi_{\min}, J) \ge 0$$

$$E.\,P.\,V.\; Net\; Cons.\; Surplus\,(\varphi) = E_1 \left[\; \sum_{t=1}^{\infty} \delta^{t-1} \left\{ \frac{(a - b\,p_t(\varphi))^2}{2b} \right\} / \, \varepsilon_0 \;\right]$$

The integral that defines the Expected Value has to be approximated using a Gauss – Legendre Quadrature Rule whose motivation is in **Appendix 3**.

4. Starting simulations: Election of Parameters and Algorithms

In our model one period of time represents one year. It was chosen the following parameter values for the simulation exercise:

$$E \in \{ 0.2 , 0.6 , 1 , 1.4 , 1.8 \}$$
$$\delta \in \{ 0.1 , 0.3 , 0.5 , 0.7 , 0.9 \}$$
$$\lambda \in \{ 0 , 0.25 , 0.5 , 0.75 , 1 \}$$
$$\beta \in \{ 0.05 , 0.25 , 0.5 , 0.75 , 0.95 \}$$
$$d \in \{ 0.1 , 0.3 , 0.5 , 0.7 , 0.9 \}$$
$$D \in \{ 0.008 , 0.018 , 0.028 , 0.038 \}$$
$$\rho \in \{ 0.6 , 0.7 , 0.8 , 0.9 \}$$
$$\sigma \in \{ 0.008 , 0.018 , 0.028 , 0.038 \}$$

The values for *E*, *d*, δ and λ are the same as in Schmalensee (1989) and it seems reasonable to explore them also in this context. For the case of *D* we may think that the percentage saving in costs due to incentives has an average annual maximum of 3.8%.[12] A

[12] Schmalensee (1989) suggests a maximum of 20% cost saving as directly attributable to incentives. At that time it was frequently observed Price Cap contracts that lasts between 10 to 15 years, implying an average cost reduction of 2% to 1.3% per year. The water sector in UK, regulated with RPI-X, gained a 12% reduction in operating costs from 1995 to 1999, averaging a 2% cost reduction per year. Hence, expecting a

depreciation of effort beyond 40-50% per annum doesn't seem defendable (even if there are investments in high-tech technology), so the minimum value we choose for ρ is 0.6. As there is no information about the possible values the persistence of the technological shock β may take (which can also vary from an industry to another) we investigate in the whole possible range from 0 to 1. With regard to the standard deviation of the innovation σ we look into a dispersion that allows a 2 standard deviations from 1.6% to 7.6% per year.

The initial cost c_0 is normalized to 1. The values of the demand parameters a and b are fixed following 2 rules: 1) the elasticity of demand at t = 0 is the one at c_0, and 2) the one period consumer surplus when $p_t = c_0$ equals to 1 (fixing the surplus at any value will allow us to isolate the real effect of E).[13]

We look for the (locally) optimal regulatory lag between 0 and 40 years: it doesn't seem reasonable to look further since in practice we observe concession contracts of as a maximum of 40 years long.

Not all the combinations of parameters are feasible. As we noted earlier we will at least warrant that the steady state level of costs, when $\varepsilon_t = 0$ for all t, is nonnegative. There are two groups of combinations that don't meet this requirement: { $\rho = 0.9$, $D = 0.028$ } and { $\rho = 0.9$, $D = 0.038$ }. We can justify their elimination on the grounds that it is generally true that the highest reductions in costs are due to investments in frontier technology that suffers a fairly quick depreciation.

We have 175.000 possible combinations of parameters. The solution algorithms were programmed in Gauss.

---

maximum of 3.8% per annum in Pure Price Cap contracts seems reasonable (notice that the upper limit of the distribution of possible savings in costs can be as large as 7.2% per year, when $d = 0.9$).

[13] On this framework fixing the level of surplus has no impact on the resulting price elasticity: $E_p = E_{c_0=1} p / (E_{c_0=1} + 1 - E_{c_0=1} p)$, for any $p \neq c0$.

The problem is first solved as an unrestricted one, then it is verified that the resulting J satisfies the participation constraint of the less efficient type; if not, we find the optimum among the cases that permit its participation.

The solution algorithm for Price Cap and Rolling Cap is the following one:

1. N quadrature points and weights $\{\varphi_i, w_i\}_{i=1}^{N}$ are obtained.
2. We fix a value for J between 0 and 41.
3. It is assumed that $\varepsilon_0 = 0$, so both the regulator and the monopoly don't observe or have information about previous technological shocks.
4. With a fixed value of J we find the monopoly's optimal effort sequence and the value of $V(x_0; \varphi_i, J)$ using Linear-Quadratic numerical solution techniques, and the value of the consumer's surplus for each φ_i.
5. Once obtained all N values of $V(x_0; \varphi_i, J)$ and consumer's surplus we calculate the regulator's W associated to J, using a quadrature rule to solve the integral over φ_i.
6. We repeat steps 2 to 5 to obtain all 42 values of W.
7. Once those are obtained, the optimal J is chosen as the one associated with the maximum value of W (unrestricted maximization). If exists more than one J that meets this requirement, we choose the smallest (if the regulator is indifferent among several values of J then prevails her subjective – maybe political – desire to pass over costs to prices as soon as possible).
8. We check if the participation constraint of the most inefficient type is met. If it is not, we choose the minimum value of J that maximizes W among the cases where $V(x_0; \varphi_{\min}, J) \geq 0$. If there are not such cases, then J is placed equal to 0. [14]

The solution algorithm for the unregulated Monopoly, that also chooses prices and is not subject to price revisions (J = 0), is the following one:

1. N quadrature points and weights $\{\varphi_i, w_i\}_{i=1}^{N}$ are obtained. It is assumed that $\varepsilon_0 = 0$.

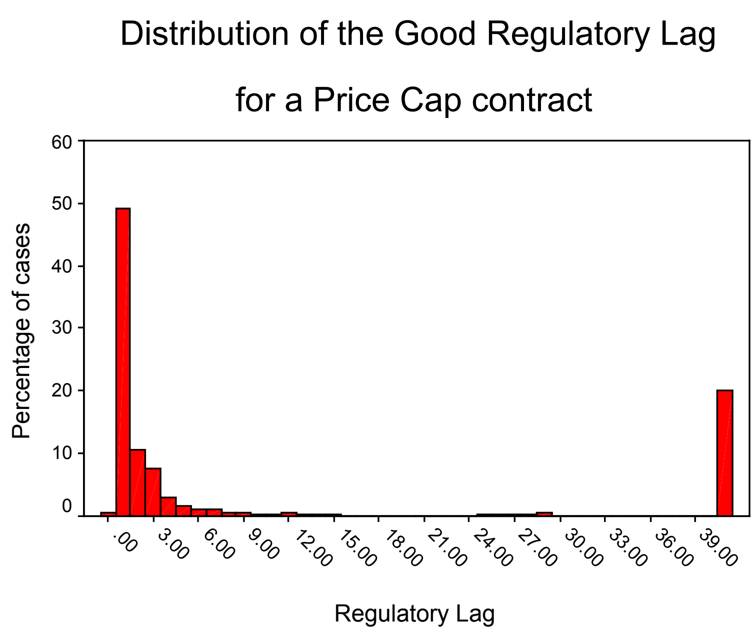
2. We find the monopoly's optimal effort sequence and the value of $V(x_0 ; \varphi_i)$ through Linear-Quadratic numerical solution techniques, and the value of the consumer's surplus for each φ_i.

3. Once obtained the N values of $V(x_0 ; \varphi_i)$ and consumer's surplus we proceed to calculate W, using the quadrature rule.

The detailed form of how to solve this problem by means of Linear-Quadratic numerical solution is on ***Appendix 1.*** The analytic form used to compute the consumer's surplus that enters in the objective function of the Regulator is detailed on ***Appendix 2.***

5. Simulation results and Conclusions

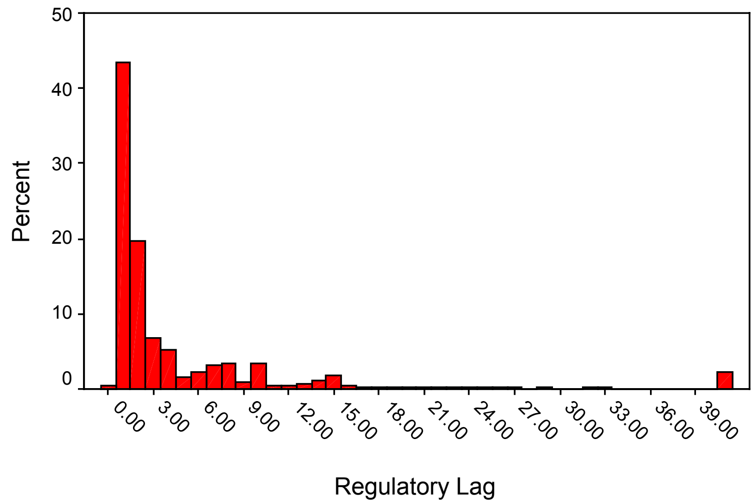
The simulations result in the following distributions for the Regulatory Lags under Price Cap and Rolling Cap contracts:



[14] Fortunately there were always such cases on the simulations, however some optimal lags resulted to be 0.

Distribution of the Good Regulatory Lag

for a Rolling Cap Contract



The median of the distribution for both regimes is a lag of 2 periods. However, note that the distribution of the Price Cap regime has more polar cases than that of the Rolling Cap (under Price Cap more contracts result in a short lag of 1 period and a higher lag of 41 periods or more).

These can be explained because of the monopoly's incentives to reduce effort (even to negative levels) across periods at an increasingly rate under a Price Cap contract with intermediate lags, overcoming its benefits and favouring shorter lags (with no enough periods to expect a significant reduction in effort) and larger ones (with a high and more constant sequence of effort). This incentives are not present in a Rolling Cap contract, so it was expected to have more intermediate lags.

There exists some marginal cases when a lag of 0 is the best for the regulator (the case of a Cost Plus contract; we don't take into account that this case is also accompanied by higher costs of auditing and control, giving it the best chance to succed but it didn't). A detailed analysis at the parameter level, not included, shows that the distribution of lags under both regimes are highly sensitive to the discount factor and less sensitive to the other parameters.

When we make direct comparisons of regulator's welfare between regimes we found that a Rolling Cap is superior to a Price Cap: only in 0.3% of the cases Price Cap is a better option than Rolling Cap, and in 25.4% of the cases the regulator is indifferent among them. In none case the unregulated monopoly is a better option (under this work assumptions).

This adverse result for the Price Cap contract can also be justified on the grounds of the relatively high incentives for the monopoly to diminish and even exert negative effort for intermediate lags, that favoured both shorter lags (with less surplus to pass to consumers) and larger lags (with a higher surplus that is however enjoyed by the monopoly alone) and diminishes the desirability of this kind of contract for the Regulator.

Some preliminar conclusions

The previous general results suggest that when there is no control on the level of costs, a Rolling Cap is a better way to regulate than a Price Cap because of the strong incentives to diminish cost reducing effort across periods or even deliberately rise the costs before a price revision under the last one. An interesting framework where this result can work could be the chilean regulatory scheme where a monopoly is regulated on the grounds of a "competitive model firm", that uses the best technology subject to demand and other demographic and geographical considerations (Galetovic and Bustos, 2000), whose costs are used to fix the price every 4 to 5 years as in a Price Cap. However in practice the regulator always has to look at the actual costs information of the operating monopoly, so the later may have the same incentives we study here under a Price Cap regime.

We are aware of two problems with the specification used here that we want to meet in subsequent work under a more general structure than the linear quadratic one: the possibility of negative costs and effort. The last one may have important effects on the desirability of Price Cap contracts.

6. References

1. Armstrong, Rees and Vickers (1995). "Optimal Regulatory Lag under Price Cap Regulation". *Revista Española de Economía. Monográfico "Regulación". Edición Especial de 1995. pp. 93-116.*

2. Baron, D. and Myerson, R. (1982). "Regulating a monopolist with unknown costs". *Econometrica*, vol. 50, pp. 911-929.

3. Braeutigam, R. and Panzar, J. (1993). "Effects of the Change from Rate of Return to Price Cap Regulation". *The American Economic Review, Vol. 83, No. 2, pp. 191-198.*

4. Clemenz, G. (1991). "Optimal Price-Cap Regulation". *Journal of Industrial Economics, 39, pp. 391-408.*

5. Dobbs, R. and Elson M. (1999). "Regulating utilities: Have we got the formula right?". *Mimeo. The McKinsey Quarterly, No. 1.*

6. Galetovic, A. and Bustos, A. (2000). "Regulación por empresa eficiente: ¿quién es realmente usted?". C.E.P.

7. Gasmi, Laffont and Sharkey (1994). "Rent Extraction and Incentives for Efficiency in Recent Regulatory Proposals". *Journal of Regulatory Economics Vol. 12, pp. 5-25.*

8. Gasmi, Laffont and Sharkey (1997). "Empirical Evaluation of Regulatory Regimes in Local Telecommunications Markets". *Journal of Regulatory Economics Vol. 12, pp. 5-25.*

9. Hansen, L. and Sargent, T. (1998). "Recursive Linear Models of Dynamic Economies". *Manuscript.*

10. Judd, K. (1998). "Numerical Methods in Economics". *MIT Press*.

11. Laffont, J. and Tirole, J. (1986). "Using cost observation to regulate firms". *Journal of Political Economy, No. 94, pp. 614-641*.

12. Laffont, J. and Tirole, J. (1993). "A Theory of Incentives in Procurement and Regulation". *MIT Press*.

13. Littlechild, S. (1983). "Regulation of British Telecommunications Profitability". London, HMSO.

14. Ljungqvist, L. and Sargent, T. (2000). "Recursive Macroeconomic Theory". *MIT Press.*

15. Marimon, R. and Scott, A. (1999). "Computational Methods for the Study o Dynamic Economies". *Oxford University Press.*

16. Mayer, C. and Vickers J. (1996). "Profit Sharing Regulation: An Economic Appraisal". *Fiscal Studies, Vol. 17, No. 2, pp. 83-101*.

17. Meller, P. (2002). "Regulación y Competencia en los Servicios de Utilidad Pública". *Perspectivas, Vol. 6, No. 1, pp. 3-10. Edited by the Department of Industrial Engineering, University of Chile.*

18. Press,Teukolsky, Vetterling and Flannery (1988). "Numerical Recipes in C. The Art of Scientific Computing". *Second Edition. Cambridge University Press.*

19. Sargent, T. (1987). "Dynamic Macroeconomic Theory". *Harvard University Press.*

20. Schmalensee, R. (1989). "Good Regulatory Regimes". *Rand Journal of Economics, Vol. 20, No. 3, pp. 417-436.*

21. Tauchen, G. and Hussey, R. (1991). "Quadrature-Based Methods or Obtaining Approximate Solutions to Nonlinear Asset Pricing Models". *Econometrica, Vol. 59, No. 2, pp. 371-398.*

22. Williamson, B. (1997) "Incentives and Commitment in RPI – X Regulation". *N.E.R.A. Topics No. 20.*

APPENDIX I

THE LINEAR-QUADRATIC FRAMEWORK.

Given the assumptions of risk neutrality, linear demand and cost, and quadratic disutility of effort the dynamic problem of the monopoly can be expressed as a Linear Quadratic Problem that is summarized in general as follows: let x_t be an Nx1 vector of state variables, u_t a Kx1 vector of control variables and w_t an Nx1 vector of i.i.d. innovations, such that $E(w_t w_t') = I$ and $E(w_s w_r') = 0$ for $s \neq r$. Hence, a monopoly should find a contingent plan $\{u_t\}_{t=0}^{\infty}$ that maximizes:

$$E_0 \sum_{t=0}^{\infty} \delta^t \left[x_t' R x_t + u_t' Q u_t + 2 x_t' W u_t \right] \quad , \quad 0 < \delta < 1$$

$$s.t. \quad x_{t+1} = A x_t + B u_t + C w_{t+1} \quad , t \geq 0 \tag{1}$$

$$x_0 \ given$$

where R is a symmetric negative semidefinite matrix, Q is a negative definite matrix, W doesn't have any restriction, A and B defines the law of motion of the state variables and C relates the innovations to the system. An additional condition is needed to find a numerical solution,

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(|x_t|^2 + |u_t|^2 \right) < \infty \tag{2}$$

so it must not be expected that both states and controls jump to infinity (see Hansen and Sargent, 1998). One way to solve this problem is using dynamic programming. Let $V(x_t)$ be the value function when the current state is x_t, hence the Bellman equation is:

$$V(x_t) = \max_{u_t} \left\{ x_t' R x_t + u_t' Q u_t + 2 x_t' W u_t + \delta \ E_t V(x_{t+1}) \right\}$$

$$s.t. \quad x_{t+1} = A x_t + B u_t + C w_{t+1} \quad , t \geq 0 \tag{3}$$

$$x_t \ given$$

The conventional way to solve this problem is through iteration on $V(.)$, that is to build a sequence $V_j(x_t)$ that converges to $V(xt)$. In particular let's define:

$$V_{j+1}(x_t) = \max_{u_t} \left\{ x_t' R x_t + u_t' Q u_t + 2 x_t' W u_t + \delta \ E_t V_j(x_{t+1}) \right\} \tag{4}$$

and suppose that we begin the iterations at $j=0$ from any concave $V_0(x)$ ($V_0(x) = 0 \ \forall x$, may work). It has been demonstrated (see Sargent, 1987) that the iterations on (4) take the quadratic form:

$$V_j(x_t) = x_t' P_j x_t + \rho_j \tag{5}$$

where P_j and ρ_j satisfy the differential equations:

$$P_{j+1} = R + \delta A' P_j A - (\delta A' P_j B + W')(Q + \delta B' P_j B)^{-1}(\delta B' P_j A + W) \qquad (6)$$

$$\rho_{j+1} = \delta \rho_j + \delta\, traza(P_j CC') \qquad (7)$$

Equation (6) is called the Riccati Differential Equation, and the resulting P_j is a symmetric matrix. Notice that the iterations on (6) are independent of ρ_j and that the C matrix only affects the sequence of ρ_j but not that of P_j. Hence, the sequence of P_j is independent of the innovations of the system and coincides with that of the non stochastic problem.

This is known as the "certainty equivalence" result that establishes that the solution of the stochastic problem is the same to that of the non stochastic one as consequence of the linear-quadratic structure of the problem. This result doesn't hold for other nonlinear structures, or when w_t is not i.i.d. (when w_t presents some persistence over time, for example, we can still define conveniently the state variables in order to obtain i.i.d. shocks but this makes clear that the results will differ from the nonstochastic problem). This feature allows to search for the control variable policy function without considering the stochastic component of the problem.

Let P and ρ be the convergence limits of (6) and (7) respectively, then the value function at the limit can be written as:

$$V(x_t) = x_t' P x_t + \rho$$

$$P = R + \delta A' PA - (\delta A' PB + W')(Q + \delta B' PB)^{-1}(\delta B' PA + W)$$

$$\rho = \frac{\delta}{1-\delta} trace(PCC').$$

Using the F.O.C., the policy function for the control variables is given by:[15]

$$u_t = -Fx_t$$

where $F = (Q + \delta B' PB)^{-1}(\delta B' PA + W)$. Notice that F is also independent of C, and therefore of the innovations. To solve the whole system we iterate directly on (6), and then applying P to find F, ρ and $V(x)$.

It is useful to define $r(x_t, u_t) = x_t' R x_t + u_t' Q u_t + 2 x_t' W u_t$. It is left to establish the particular form of the states, controls and of *R, Q, W, A, B* and *C* for each regulatory regime[16].

15 The S.O.C. of the problem is $(Q + \delta B' PB)$: if the resulting matriz are negative definite then the solution is a local maximum, that is also global because of the concavity of the problem. It is also important to check that all the Eigen Values of the $(A - BF)$ matrix lies inside the unitary circle so that the transition dynamics of states has a limiting stationary distribution (it is necessary to check the eigen values associated to those state variables that are not constants).

UNREGULATED MONOPOLY'S PROBLEM

Let's define $x_t = \{ \theta_{t-1}, \varepsilon_{t-1}, 1 \}$, $u_t = e_t$ and $w_t = \xi_t \sim N(0,1)$ (hence $\varepsilon_t = \beta\varepsilon_{t-1} + \sigma\xi_t$). The expression for $r(x_t, u_t)$ is:

$$r(\theta_{t-1}, \varepsilon_{t-1}, 1 \; ; \; e_t) = \frac{1}{4b}\Big[(a-bc_0)^2 + 2bc_0(a-bc_0)\rho\theta_{t-1} + 2bc_0(a-bc_0)e_t + \\ + 2bc_0(a-bc_0)\beta\varepsilon_{t-1} + (bc_0\rho)^2\theta_{t-1}^2 + (b\alpha)^2 e_t^2 + (bc_0\beta)^2\varepsilon_{t-1}^2 + 2(bc_0)^2\rho\theta_{t-1}e_t + \\ + 2(bc_0)^2\rho\beta\theta_{t-1}\varepsilon_{t-1} + 2(bc_0)^2\beta e_t\varepsilon_{t-1}\Big] - \frac{e_t^2}{2\varphi}$$

The matrices for *r(.)* and the law of transition of the state variables are:

$$R_{(3x3)} = \frac{1}{4b}\begin{bmatrix} (bc_0\rho)^2 & (bc_0)^2\rho\beta & (a-bc_0)bc_0\rho \\ (bc_0)^2\rho\beta & (bc_0\beta)^2 & (a-bc_0)bc_0\beta \\ (a-bc_0)bc_0\rho & (a-bc_0)bc_0\beta & (a-bc_0)^2 \end{bmatrix} \quad ;$$

$$Q = \left(-\frac{1}{2\varphi} + \frac{(bc_0)^2}{4b}\right); \qquad W_{(3x1)} = \frac{1}{4b}\begin{bmatrix} (bc_0)^2\rho \\ (bc_0)^2\beta \\ (a-bc_0)bc_0 \end{bmatrix} \quad ;$$

$$A_{(3x3)} = \begin{bmatrix} \rho & \beta & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad ; \qquad B_{(3x1)} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad ; \qquad C_{(3x1)} = \begin{bmatrix} \sigma \\ \sigma \\ 0 \end{bmatrix}$$

MONOPOLY'S PROBLEM UNDER A PRICE CAP CONTRACT

Following the special nomenclature and assumptions in the paper, we define $x_t = \{\theta_0^\tau, \varepsilon_0^\tau, 1\}$, $u_t = \{e_1^\tau, e_2^\tau, ..., e_J^\tau\}$ and $w_t = \{\xi_1^\tau, \xi_2^\tau, ..., \xi_J^\tau\}$ (with $\xi_i^\tau \sim N(0,1)$, $I = 1,..., J$). After some tedious algebra, $r(x_t, u_t)$ can be expressed as:

16 For computational effects, we defined the state and control variables for the monopoly's problem under distinct regimes in a different way than that of the theoretical part where we needed to obtain analytic expressions for effort and make the recursive nature of each problem clear.

$$r(\theta_0^\tau, \varepsilon_0^\tau, 1 \; ; \; e_1^\tau, e_2^\tau, ..., e_J^\tau) =$$

$$c_0^1(a - bc_0^1)\theta_0^\tau y \; + \; b(c_0^1\theta_0^\tau)^2 y + \; c_0^1(a - bc_0^1)\sum_{t=1}^{J}\delta^{t-1}\left(\frac{1-(\delta\rho)^{J+1-t}}{1-\delta\rho}\right)e_t^\tau \; +$$

$$+ \, b(c_0^1)^2\theta_0^\tau\sum_{t=1}^{J}\delta^{t-1}\left(\frac{1-(\delta\rho)^{J+1-t}}{1-\delta\rho}\right)e_t^\tau \; + \frac{c_0^1(a-bc_0^1)}{\rho-\beta}\beta\varepsilon_0^\tau z + \frac{(c_0)^2 b}{\rho-\beta}\beta\theta_0^\tau\varepsilon_0^\tau z \; -$$

$$- \; \sum_{t=1}^{J}\delta^{t-1}\frac{(e_t^\tau)^2}{2\varphi}$$

where $y = \left[\rho\frac{1-(\delta\rho)^J}{1-\delta\rho} - \frac{1-\delta^J}{1-\delta}\right]$, and $z = \left[\rho\frac{1-(\delta\rho)^J}{1-\delta\rho} - \beta\frac{1-(\delta\beta)^J}{1-\delta\beta}\right]$. The matrices of the objective function are:

$$R_{(3x3)} = \begin{bmatrix} b(c_0^1)^2 y & \frac{b(c_0^1)^2\beta}{2(\rho-\beta)}z & \frac{c_0^1(a-bc_0^1)}{2}y \\ \frac{b(c_0^1)^2\beta}{2(\rho-\beta)}z & 0 & \frac{c_0^1(a-bc_0^1)\beta}{2(\rho-\beta)}z \\ \frac{c_0^1(a-bc_0^1)}{2}y & \frac{c_0^1(a-bc_0^1)\beta}{2(\rho-\beta)}z & 0 \end{bmatrix};$$

$$Q_{(JxJ)} = -\frac{1}{2\varphi}\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \delta & 0 & \cdots & 0 \\ 0 & 0 & \delta^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \delta^{J-1} \end{bmatrix};$$

$$W_{(3xJ)} = \frac{1}{2}\begin{bmatrix} b(c_0^1)^2\frac{1-(\delta\rho)^J}{1-\delta\rho} & \delta b(c_0^1)^2\frac{1-(\delta\rho)^{J-1}}{1-\delta\rho} & \cdots \; \cdots & \delta^{J-1}b(c_0^1)^2 \\ 0 & 0 & \cdots \; \cdots & 0 \\ c_0^1(a-bc_0^1)\frac{1-(\delta\rho)^J}{1-\delta\rho} & \delta c_0^1(a-bc_0^1)\frac{1-(\delta\rho)^{J-1}}{1-\delta\rho} & \cdots \; \cdots & \delta^{J-1}c_0^1(a-bc_0^1) \end{bmatrix}$$

Some additional calculations give us the equation of the law of movement for each state (using the fact that $\theta_J^\tau \equiv \theta_0^{\tau+1}$ and $\varepsilon_J^\tau \equiv \varepsilon_0^{\tau+1}$):

$$\theta_0^{\tau+1} \equiv \theta_J^\tau = \rho^J\theta_0^\tau + \sum_{i=1}^{J}\rho^{J-i}e_i^\tau + \beta\frac{\rho^J - \beta^J}{\rho-\beta}\varepsilon_0^\tau + \sum_{i=1}^{J}\left(\frac{\rho^{J+1-i} - \beta^{J+1-i}}{\rho-\beta}\right)\sigma\xi_i^\tau$$

$$\varepsilon_0^{\tau+1} \equiv \varepsilon_J^{\tau} = \beta^J \varepsilon_0^{\tau} + \sum_{i=1}^{J} \beta^{J-i} \sigma \xi_i^{\tau}$$

hence, the matrices of the state variable transition equation are:

$$A_{(3x3)} = \begin{bmatrix} \rho^J & \beta \dfrac{\rho^J - \beta^J}{\rho - \beta} & 0 \\ 0 & \beta^J & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad ; \qquad B_{(3xJ)} = \begin{bmatrix} \rho^{J-1} & \rho^{J-2} & \cdots & \rho & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \quad ;$$

$$C_{(3xJ)} = \begin{bmatrix} \sigma \dfrac{\rho^J - \beta^J}{\rho - \beta} & \sigma \dfrac{\rho^{J-1} - \beta^{J-1}}{\rho - \beta} & \cdots & \sigma \dfrac{\rho^2 - \beta^2}{\rho - \beta} & \sigma \\ \sigma\beta^{J-1} & \sigma\beta^{J-2} & \cdots & \sigma\beta & \sigma \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

MONOPOLY'S PROBLEM UNDER A ROLLING CAP CONTRACT

Let's define $x_t = \{\theta_{t-J}, e_{t-J+1} + \varepsilon_{t-J+1}, e_{t-J+2} + \varepsilon_{t-J+2}, ..., e_{t-1} + \varepsilon_{t-1}, \varepsilon_{t-1}, 1\}$, $u_t = e_t$ y $w_t = \xi_t$, (with $\xi_t \sim N(0,1)$). After some calculations, taking into account that ε_t is not observed until e_t is decided, $r(x_t, u_t)$ can be expressed as:

$$\begin{aligned} r(\theta_{t-J}, e_{t-J+1} + \varepsilon_{t-J+1}, e_{t-J+2} + \varepsilon_{t-J+2}, ..., e_{t-1} + \varepsilon_{t-1}, \varepsilon_{t-1}, 1 \;;\; e_t) = \\ c_0(a - bc_0)e_t + b(c_0)^2 \theta_{t-J}\, e_t + c_0(a - bc_0) \sum_{i=1}^{J-1} \rho^{J-i} (e_{t-J+i} + \varepsilon_{t-J+i}) + \\ + b(c_0)^2 \theta_{t-J} \sum_{i=1}^{J-1} \rho^{J-i} (e_{t-J+i} + \varepsilon_{t-J+i}) + c_0(a - bc_0)\beta\, \varepsilon_{t-1} + b(c_0)^2 \beta\, \theta_{t-J} \varepsilon_{t-1} - \\ - c_0(a - bc_0)(1 - \rho^J)\theta_{t-J} - b(c_0)^2 (1 - \rho^J)\theta_{t-J}^2 - \frac{e_t^2}{2\varphi} \end{aligned}$$

defining the matrices of the value function as:

$$R = \frac{c_0}{2}\begin{bmatrix} -2bc_0(1-\rho^J) & bc_0\rho^{J-1} & \cdots & bc_0\rho & bc_0\beta & -(1-\rho^J)(a-bc_0) \\ bc_0\rho^{J-1} & 0 & \cdots & 0 & 0 & (a-bc_0)\rho^{J-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ bc_0\rho & 0 & \cdots & 0 & 0 & (a-bc_0)\rho \\ bc_0\beta & 0 & \cdots & 0 & 0 & (a-bc_0)\beta \\ -(1-\rho^J)(a-bc_0) & (a-bc_0)\rho^{J-1} & \cdots & (a-bc_0)\rho & (a-bc_0)\beta & 0 \end{bmatrix}$$

$(J+2)x(J+2)$

$$Q = -\frac{1}{2\varphi} \quad ; \quad W_{(J+2)x1} = \frac{1}{2}\begin{bmatrix} b(c_0)^2 \\ 0 \\ \vdots \\ 0 \\ c_0(a-bc_0) \end{bmatrix}$$

Given that $\theta_t = \rho\theta_{t-1} + e_t + \varepsilon_t$ y $\varepsilon_t = \beta\varepsilon_{t-1} + \xi_t$, the matrices of the states law of movement are:

$$A_{(J+2)x(J+2)} = \begin{bmatrix} \rho & 1 & 0 & \cdots & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & \beta & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & \beta & 0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 1 \end{bmatrix} \quad ; \quad B_{(J+2)x1} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad ; \quad C_{(J+2)x1} = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ \sigma \\ \sigma \\ 0 \end{bmatrix}$$

APPENDIX II

THE CONSUMER SURPLUS

*The Expected Present Value of the Consumer's Surplus (*E.V.C.S. *), given the assumed functional forms, takes the following form:*

$$E.V.C.S. = E_1\left[\sum_{t=1}^{\infty}\delta^{t-1}\left\{\frac{(a-bp_t)^2}{2b}\right\} / \varepsilon_0\right]$$

Consumer's Surplus generated by an unregulated Monopoly.

In this case the monopoly also decides the price for each period before the corresponding technological shock is realized. Define $\tilde{c}_t = E_t(c_t / \varepsilon_{t-1})$, then:

$$E.V.C.S. = E_1\left[\frac{1}{2b}\sum_{t=1}^{\infty}\delta^{t-1}\left(\frac{a-b\tilde{c}_t}{2}\right)^2 / \varepsilon_0\right]$$

The cost function at t can be rewritten in the following way:

$$\tilde{c}_t = c_0(1-\tilde{\theta}_t) = c_0(1-\rho\theta_{t-1}-\tilde{e}_t-\beta\varepsilon_{t-1}) = c_0(1-\rho^t\theta_0-\sum_{i=1}^{t-1}\rho^{t-i}(e_i+\varepsilon_i)-\tilde{e}_t-\beta\varepsilon_{t-1})$$

defining $\bar{c}_t = c_0(1-\rho^t\theta_0-\sum_{i=1}^{t}\rho^{t-i}e_i) = c_0(1-\sum_{i=1}^{t}\rho^{t-i}e_i)$, with $\theta_0 = 0$, as the part of $\tilde{c}_t$ that doesn't depend on the shock[17], then $\tilde{c}_t = \bar{c}_t - c_0\sum_{i=1}^{t-1}\rho^{t-i}\varepsilon_i - c_0\beta\varepsilon_{t-1}$, and:

$$V.P.E.C. = E_1\left[\frac{1}{8b}\sum_{t=1}^{\infty}\delta^{t-1}\left((a-b\bar{c}_t)+bc_0\sum_{i=1}^{t-1}\rho^{t-i}\varepsilon_i+bc_0\beta\,\varepsilon_{t-1}\right)^2 / \varepsilon_0\right]$$

$$=\sum_{t=1}^{\infty}\delta^{t-1}\frac{(a-b\bar{c}_t)^2}{8b}+\frac{c_0\rho\beta\varepsilon_0}{4(\rho-\beta)}\sum_{t=1}^{\infty}\left[(\delta\rho)^{t-1}-(\delta\beta)^{t-1}\right](a-b\bar{c}_t)+\frac{c_0\beta\varepsilon_0}{4}\sum_{t=1}^{\infty}(\delta\beta)^{t-1}(a-b\bar{c}_t)+$$

17 This is only a practical simplification, since the optimal effort is a function of the last technological shock. We approximate the Consumer's Surplus for all the cases in the paper, to make them comparable, using a single path of effort that assumes $\varepsilon_t = 0$ for all t but respecting the value of ε_0. This simplification is very useful for the simulations and it approximates very well the consumer's surplus, without altering the results. The simplification respects the fact that as the regulatory lag increases in both Price Cap and Rolling Cap contracts their surpluses must converge to that of a Pure Price Cap.

$$+\frac{bc_0^2\beta^2}{8}\left[\frac{2\delta\beta\rho}{(1-\delta\beta\rho)(1-\delta\beta^2)}+\frac{\rho^2\delta(1+\delta\beta\rho)}{(1-\delta\beta\rho)(1-\delta\beta^2)(1-\delta\rho^2)}+\frac{1}{1-\delta\rho^2}\right]\varepsilon_0^2 +$$
$$+\frac{bc_0^2\delta}{8(1-\delta\beta^2)(1-\delta)}\left[\frac{2\beta\rho}{1-\delta\beta\rho}+\frac{\rho^2(1+\delta\beta\rho)}{(1-\delta\beta\rho)(1-\delta\rho^2)}+\beta^2\right]\sigma^2$$

in the simulations we use a single and sufficiently long serie of effort and costs (250 are enough) to obtain this.

Consumer's Surplus for the Monopoly under a Price Cap Contract

Because the regulator fixes the monopoly's price using information about past costs, the price is known at the beginning of every period. The Consumer's Surplus takes the following form:

$$V.P.E.C.=E_1^1\left[\sum_{\tau=1}^{\infty}\delta^{J(\tau-1)}\sum_{t=1}^{J}\delta^{t-1}\frac{(a-bc_0^\tau)^2}{2b}/\varepsilon_0^1\right]=\frac{1-\delta^J}{1-\delta}E_1^1\left[\sum_{\tau=1}^{\infty}\delta^{J(\tau-1)}\frac{(a-bc_0^\tau)^2}{2b}/\varepsilon_0^1\right]$$

defining $c_0^{\tau+1}\equiv c_J^\tau=c_0^1(1-\theta_J^\tau)$ (notice that $\theta_0^{\tau+1}=\theta_J^\tau$).

We can write the value that the percentage saving in costs takes at every period as follows: $\theta_0^{\tau+1}\equiv\theta_J^\tau=\rho^{J\tau}\theta_0^1+\sum_{m=1}^{\tau}\rho^{J(\tau-m)}\sum_{i=0}^{J-1}\rho^i(e_{J-i}^m+\varepsilon_{J-i}^m)$.

Let $\bar{c}_0^{\tau+1}\equiv\bar{c}_J^\tau=c_0^1\left(1-\rho^{J\tau}\theta_0^1-\sum_{m=1}^{\tau}\rho^{J(\tau-m)}\sum_{i=0}^{J-1}\rho^i e_{J-i}^m\right)$ be the part of the unitary cost the monopoly arrives with to the next price revision and that doesn't depend directly on the innovations, then $c_0^{\tau+1}=\bar{c}_0^{\tau+1}-c_0^1\sum_{m=1}^{\tau}\rho^{J(\tau-m)}\sum_{i=0}^{J-1}\rho^i\varepsilon_{J-i}^m$.

Using these definitions we can express the consumer's surplus as:

$$V.P.E.C.=E_1^1\left[\frac{1-\delta^J}{1-\delta}\sum_{\tau=1}^{\infty}\delta^{J(\tau-1)}\frac{(a-bc_0^\tau)^2}{2b}/\varepsilon_0^1\right]$$
$$=\frac{1-\delta^J}{1-\delta}\left\{\frac{(a-bc_0^1)^2}{2b}+E_1^1\left[\sum_{\tau=1}^{\infty}\delta^{J\tau}\frac{(a-bc_0^{\tau+1})^2}{2b}/\varepsilon_0^1\right]\right\}$$
$$=\frac{1-\delta^J}{1-\delta}\left\{\frac{(a-bc_0^1)^2}{2b}+\frac{1}{2b}E_1^1\left[\sum_{\tau=1}^{\infty}\delta^{J\tau}\left((a-b\bar{c}_0^{\tau+1})+c_0^1\sum_{m=1}^{\tau}\rho^{J(\tau-m)}\sum_{i=0}^{J-1}\rho^i\varepsilon_{J-i}^m\right)^2/\varepsilon_0^1\right]\right\}$$

$$= \frac{1-\delta^J}{1-\delta}\left\{ \sum_{\tau=1}^{\infty} \delta^{J(\tau-1)} \frac{(a-b\overline{c}_0^{\,\tau})^2}{2b} + \frac{c_0^1 \beta \varepsilon_0^1}{\rho-\beta} \sum_{\tau=1}^{\infty} \left[(\delta\rho)^{J\tau} - (\delta\beta)^{J\tau}\right](a-b\overline{c}_0^{\,\tau+1}) + \right.$$

$$+ \frac{\delta^J b (c_0^1 \beta)^2 (1+(\delta\beta\rho)^J)(\varepsilon_0^1)^2}{2(1-(\delta\beta^2)^J)(1-(\delta\beta\rho)^J)(1-(\delta\rho^2)^J)}\left(\frac{\rho^J-\beta^J}{\rho-\beta}\right) +$$

$$+ \frac{\delta^J b (c_0^1)^2 \sigma^2}{2(\rho-\beta)^2(1-\delta^J)(1-(\delta\rho^2)^J)}\left[\rho^2 \frac{1-\rho^{2J}}{1-\rho^2} - 2\rho\beta\frac{1-(\rho\beta)^J}{1-\rho\beta} + \beta^2\frac{1-\beta^{2J}}{1-\beta^2}\right] +$$

$$+ \frac{\delta^{2J} b (c_0^1 \beta)^2 (1+(\delta\beta\rho)^J)\sigma^2}{2(1-\delta^J)(1-(\delta\beta^2)^J)(1-(\delta\beta\rho)^J)(1-(\delta\rho^2)^J)}\left(\frac{\rho^J-\beta^J}{\rho-\beta}\right)^2\left(\frac{1-\beta^{2J}}{1-\beta^2}\right) +$$

$$+ \frac{\delta^{2J} b (c_0^1)^2 \beta\sigma^2}{(1-\delta^J)(\rho-\beta)^2}\left[\frac{\rho(1-(\beta\rho)^J(1-\beta^2)) - \beta(1-\beta^{2J}(1-\beta\rho))}{(1-\beta\rho)(1-\beta^2)}\right]\left[\frac{\rho^{2J}-(\rho\beta)^J}{(1-(\delta\beta\rho)^J)(1-(\delta\rho^2)^J)}\right]$$

as in the unregulated monopoly case we use a single sufficiently long serie (250) of unitary costs to calculate this expression (note that $c_0^1 = \overline{c}_0^1$ because it is the initial cost).

Consumer's Surplus for the Monopoly under a Rolling Cap Contract

As in Price Cap, the price is known at the beginning of every period. The Consumer's Surplus takes the following form:

$$V.P.E.C. = E_1\left[\sum_{t=1}^{\infty} \delta^{t-1}\left(\frac{(a-bc_{t-J})^2}{2b}\right) / \varepsilon_0\right]$$

Define $c_t = c_0(1-\theta_t) = c_0(1-\rho^t\theta_0 - \sum_{i=0}^{t-1}\rho^i(e_{t-i}-\varepsilon_{t-i}))$ and $\overline{c}_t = c_0(1-\rho^t\theta_0 - \sum_{i=0}^{t-1}\rho^i e_{t-i})$ $= c_0(1-\sum_{i=0}^{t-1}\rho^i e_{t-i})$ (the part of the unitary cost that doesn't depend on the shock, with $\theta_0 = 0$). Hence, $c_t = \overline{c}_t - c_0\sum_{i=0}^{t-1}\rho^i\varepsilon_{t-i}$, for $t > 0$ ($\overline{c}_t = c_t$ for $t \le 0$). The consumer's surplus can be expressed in the following way:

$$V.P.E.C. = \sum_{t=1}^{\infty}\delta^{t-1}\frac{(a-b\overline{c}_{t-J})^2}{2b} + \frac{\delta^J\beta\varepsilon_0 c_0}{\rho-\beta}\sum_{t=1}^{\infty}\left[\rho(\delta\rho)^{t-1} - \beta(\delta\beta)^{t-1}\right](a-b\overline{c}_t)^2$$

$$+ \frac{\delta^J b(c_0\beta\varepsilon_0)^2}{2}\left[\frac{1+\delta\rho\beta}{(1-\delta\rho^2)(1-\delta\rho\beta)(1-\delta\beta^2)}\right] + \frac{\delta^J b(c_0)^2(1+\delta\rho\beta)\sigma^2}{2(1-\delta)(1-\delta\rho^2)(1-\delta\rho\beta)(1-\delta\beta^2)}$$

Once again, we use a single sufficiently long serie (250) of unitary costs to calculate this expression. The reader can verify that at J = 1 both expressions for consumer's surplus in Price Cap and Rolling Cap coincides.

For a Pure Price Cap (J = ∞, with $p_t = c_0 \ \forall t > 0$), the consumer's surplus is:

$$V.P.E.C. = \frac{1}{1-\delta}\left(\frac{(a-bc_0)^2}{2b}\right)$$

notice that it is independent of the monopoly's type because the consumers never benefit from the savings in monopoly's costs. We also calculate the Producer's Surplus for this case. Let $e_\infty = \dfrac{\varphi c_0 (a-bc_0)}{1-\delta\rho}$ be the effort carried out at every period. The Expected Present Value of the Producer's Surplus is:

$$\Pi(\varphi,\infty) = \frac{\varphi}{2(1-\delta)}\left(\frac{c_0(a-bc_0)}{1-\delta\rho}\right)^2 + \frac{c_0(a-bc_0)\beta}{(1-\delta\rho)(1-\delta\beta)}\varepsilon_0$$

APPENDIX III

THE Gauss-Legendre quadrature rule

The Gaussian Quadrature Rules are extensively used in mathematics to approximate numerically the value of complex integrals. Many economic problems require some decisions to be based on the expected value (integral) of certain variables. Who initially introduced Quadrature Rules techniques in economics were Tauchen and Hussey (1991). In general, we want to solve:

$$\int_a^b f(x)\,dx \quad , \quad \textit{for some a and b}$$

A quadrature rule approximates the integral by means of a weighted sum of values for *f(x),* evaluated at some selected points. Both the points and weights are selected using the quadrature rule. The quadrature rules allow for high order integration, that is not the same as high precision: high order is accompanied by high precision when *f(x)* is very "soft", in the sense of being "very well approximated for a polynomial" (Press et. al., 1988).

The rule replaces *f(x)* with *f(x)* times some function *W(x)*. Given *W(x)* and an integer *N*, the weights w_j and points (abscissas) x_j are found such that:

$$\int_a^b W(x) f(x)\,dx \quad \approx \quad \sum_{j=1}^{N} w_j f(x_j)$$

which is exact when *f(x)* is a polynomial with degree between *N* and *2N - 1*. The abscissas correspond to the roots of the N - degree polynomial associated to *W(x)*. This polynomial is orthogonal to any other associated to *W(x)* with degree different to *N*. Hence, if $p_N(x)$ is polynomial of degree N, then $\int_a^b p_N(x)\, p_J(x) W(x)\,dx = 0$, *for* $J \neq N$ and equal to a constant when $J = N$. The resulting weights *wj* are functions of these polynomials.[18]

When *W(x) = 1*, *a = -1* and *b = 1*, the quadrature rule is a Gauss-Legendre one whose orthogonal polynomials follow the iterative process:

$$P_{-1}(x) = 0 \quad , \quad P_0(x) = 1$$
$$(N+1)\, P_{N+1}(x) = (2N+1)\, x\, P_N(x) - N P_{N-1}(x) \quad \textit{para } N \geq 1$$

With this process we can obtain the polynomial of degree *N* whose roots in (*-1,1*) will serve as the abscissas $\tilde{x}_j$. The weights $\tilde{w}_j$ are calculated using this formula:

[18] For further reading about Quadrature Rules and orthogonal polynomials see Press et. al. (1988), Judd (1998) and Marimon et. al. (1999).

$$\tilde{w}_j = \frac{2}{(1-\tilde{x}_j^2)P_N'(\tilde{x}_j)}$$

where P_N' is the derivative of P_N. In this case it can demonstrated that the abscissas and weights of the orthogonal polynomials coincide with those of the ortho-normal polynomials (that are orthogonals with $\int_a^b [p_N(x)]^2 W(x)\,dx = 1$, $for\ N \geq 0$).

To integrate a function in the interval (*a,b*) we have to adjust the abscissas and the weights in the following way: let $x_m = 0.5\ (a + b)$ be the midpoint of the interval, and $x_l = 0.5\ (b\text{-}a)$ a half of its longitude, then:

$$x_i = x_m + x_l \tilde{x}_i \quad , \quad w_i = x_l \tilde{w}_i$$

Our original problem is to approximate the following integral:

$$\int_{\varphi_{\min}}^{\varphi_{\max}} \left[\, V.\,P.\,E.\,Exc.\,Neto\ Cons.(\varphi) + \lambda V(x_0\,;\varphi, J)\,\right] f(\varphi)d\varphi \quad , \quad 0 \leq \lambda \leq 1$$

$$where\ f(\varphi) = \frac{1}{2\omega Dd}\ ,\ \ \omega = \frac{1-\delta\rho}{(a-bc_0)\,c_0}. \quad Hence:$$

$$\frac{1}{2\omega Dd}\int_{\varphi_{\min}}^{\varphi_{\max}} \left[\, V.\,P.\,E.\,Exc.\,Neto\ Cons.(\varphi) + \lambda V(x_0\,;\varphi, J)\,\right] d\varphi \quad , \quad 0 \leq \lambda \leq 1$$

defining $g(\varphi) = \left[\, V.\,P.\,E.\,Exc.\,Neto\ Cons.(\varphi) + \lambda V(x_0\,;\varphi, J)\,\right]\ with\ \ 0 \leq \lambda \leq 1$, our problem is to approximate $\frac{1}{2\omega Dd}$ times $\int_{\varphi_{\min}}^{\varphi_{\max}} g(\varphi)\, d\varphi$. We have already pointed out that *g(.)* for all the different combinations of parameters is almost exactly approximated with a polynomial of degree 9. Hence, the Gauss-Legendre rule allows us to use al least 5 points.