A Critique of the Theory of Incentive Regulation: 
Implications for the Design of Performance Based Regulation for Postal Service\textsuperscript{1}

Michael A. Crew  
*Rutgers University*

Paul R. Kleindorfer  
*University of Pennsylvania*

The main purpose of this paper is to examine developments in the regulation of natural monopoly over the last twenty years and draw some implications for the regulation of postal service. While price regulation has traditionally taken a back seat under the traditional public enterprise monopolies that were almost universal in postal service this is changing with the privatization of the Dutch and German PO's, the increased liberalization of markets, the creation of specialized regulators most recently in the United Kingdom and similar developments.\textsuperscript{2} These developments make the form of regulation to be employed an important and topical problem. In this paper we examine what lessons can be drawn for postal regulation from the theory and the practice in other industries where regulation has existed for a considerable time.

Over the last twenty years regulated monopolies have undergone significant changes not just in the United States but also in Europe, Australia and New Zealand. Over this period public enterprises have been privatized, notably in the United Kingdom, and a wave of divestitures and mergers has been evident, especially in the U.S. Regulatory economists have often played important roles in promoting these developments and in legitimating them. However, in the final analysis, despite a great deal of activity in economic theory and in policy, not much progress has been evident in the development of new theories or frameworks for regulation. The action has been in price cap regulation (PCR) or "incentive regulation". The first major application of PCR followed the proposals of Littlechild (1983, 1986) for the regulation of telecommunications and water in the U.K. But developments in PCR have arguably been few and far between since then.

We assess the reasons for this slow progress and argue that they are at least in part the result of a failure by economists to model properly the nature of the context in which regulation takes place. We will evaluate theory and practice particularly as it affects the electric utility industry. We will be concerned primarily with the role of the mechanism design theory of regulation, notably, Baron and Myerson (1981) and Laffont and Tirole (1993). This theory is concerned with the regulation of an old-style regulated monopolist. This contrasts with the kind of regulated monopolies that we face today where competitive entry is a key issue.

\textsuperscript{1} A version of this paper was prepared for presentation at the Rutgers University 13\textsuperscript{th} Annual Western Conference on Competitive Challenges in Network Industries, Monterey, CA, July 5-7, 2000. We would like to than our discussants, Fred Curry, Jutta Shaw and Yasugi Otsuka for helpful comments. In addition, we benefitted from discussions with Eric Ackerman, Carl Danner, Ahmad Faruqui and Dennis Weisman.

\textsuperscript{2} We are ignoring the problem of ownership. For purposes of this paper we consider the national post office to be privately owned. In this way we can consider it to be similar to other kinds of private companies in its response to regulation.
We will not focus on the entry issue in applying mechanism design theory but rather on the oft-cited problem in incentive regulation of assuring what economic theorists refer to as regulatory commitment.

In section 1 we provide a critical review of developments in incentive regulation in theory and practice over the last two decades. We argue that economic design theory underlying regulation has failed to address properly the constraints inherent in the regulatory process. In section 2 we develop a simple model of the regulatory process, which addresses one of the major perceived problems with PCR, namely that of commitment. Commitment in this context means that the regulator allows the PCR to continue to operate according to its design parameters, irrespective of the outcome in terms of profits, etc. We argue there that the usual claims of theoreticians about the “commitment problem” are fundamentally the result of misconstruing the nature of the regulatory process. In a word, problems of excess profits and potential bankruptcy are essential features of regulation and cannot be avoided by regulators, even though these constraints have typically been neglected by regulatory theorists. However, in practice, such constraints place natural bounds on the ability of regulators to commit to long-term PCR regimes; adjustment of such regimes is therefore to be expected and reflects an essential reality of the regulatory environment. Understanding what is likely to drive the adjustment of the PCR regime in various contexts is a more fruitful line of research and policy than decrying the failure of regulators to commit to something that is not in their power to control. Section 3 explores some of the implications of this analysis for the implementation of price caps in postal service. Section 4 is by way of summary, conclusions and implications for future research and policy. These final two sections provide our assessment of what distribution companies can and should do in working with their regulators to design incentive regulatory schemes in practice.

1. Review of Incentive Regulation in Theory and Practice

There have been several significant developments in the governance of natural monopolies over the last twenty years. These include the AT&T Divestiture in 1984 of its local telephone companies into seven regional companies, RBOCs (regional bell operating companies), and the subsequent consolidation into four companies with further consolidation in progress. This occurred at a time of very rapid technological change in microelectronics, optical fiber and wireless telephony to say nothing of the Internet revolution. A second major development was the privatization of public enterprise natural monopolies in Europe, Australia and New Zealand. A third development was the opening up of gas purchases to competition. Finally, the Energy Policy Act of 1992 was a major attempt at opening up electricity generation to competition. This created pressure for divestiture of generation by vertically integrated utilities and the creation of distribution companies in the United States, which was preceded by similar developments in the UK and has been followed by similar legislation in the European Union in 1999.

Despite this flurry of activity on the regulatory front, both in theory and in practice, actual progress has been rather slight. Indeed, there has been no shortage of contributions by both theoretical and applied economists claiming to expand the frontiers of regulatory economics. It might be concluded from some contributors to the literature and practice that a new age had dawned in regulation in theory and in practice.

---

No longer would economists have to suffer the inefficiencies of regulation as exemplified in Rate-of-Return regulation (ROR). Incentive regulation, in the form of PCR, was the way forward and with it would come the efficiency gains that economists had long sought and that were presumably intended from the major changes in the structure of industries that were taking place.

However, not much of significance has happened in the structure or form of PCR since its original inception. Our critical review of the highlights in regulatory developments will argue that the lack of progress arises from a failure in economic design. In general terms, the failure arose from ignoring essential institutional constraints in the underlying economic models. To provide a basis for understanding our argument, we review briefly some of the highlights of theoretical contributions from the past several decades.

Let us begin with Littlechild’s analysis of the rationale and problems with PCR. Recall that PCR was introduced at a time when economists were highly critical of previous regulation in theory and practice. Littlechild (1986, 26) argued: “The problems of ‘cost plus’ contracts are well known in the UK. Likewise, a company that is regulated under the US system has reduced incentive to cut its costs and seek efficiency, insofar as any savings must immediately be passed on. (This is offset to some extent by ‘regulatory lag’.) ... To counter these tactics, the regulatory authority needs to determine in some detail which expenses are allowable, what depreciation policy is appropriate, and what rate of return is ‘fair’. ” Thus, Littlechild’s basic problem with ROR concerned its failure to provide incentives for efficiency and its high transactions costs.

In proposing PCR for the UK, Littlechild sought to provide incentives for efficiency while reducing transactions cost and micro management. Bear in mind that he was writing in the context of industries that were about to be privatized, industries that he probably believed were grossly inefficient and needed the spur of RPI – X to have them improve efficiency. In addition, he was probably aware that this would not last but that the one-time opportunity to secure increased internal efficiency (also known as X-efficiency) should be taken. Indeed, he showed a considerable awareness of the problems ex ante that he was about to face ex post as the regulator of the UK electric utility industry. “The longer the period that RPI – X control is expected to remain in operation, the greater the uncertainty about the location of the ‘feasible range’ for X... But if X is able to be revised, the regulated company will consider what effects its actions are likely to have on future levels of X. It will realize that greater cost reductions today will lead to pressure for greater price reductions when X is reset in the future. Higher Profits may even lead to a public demand to revise X prematurely. To this extent, incentives to maximum efficiency are blunted. When RPI – X is seen as a permanent regulation, this is potentially a serious problem.” (Littlechild 1986, p. 29)

Theorists, both preceding and following the introduction of PCR in the UK in the 1980s, were no less critical of previous regulatory theory and practice than Littlechild, but their aspirations were much higher. We illustrate with a few highlights from the theory of the past three decades. Loeb and Magat (1974) proposed rewarding the regulated firm based in part upon the consumer surplus that their prices generated. The inexorable consequence of this would be a rapid convergence to first-best prices albeit with some mildly disgruntled consumers who would be watching their surplus flow through the regulator to the monopolist. Aside from this unsettling feature, the common knowledge of the demand curve assumed was raised as an insurmountable barrier to implementation and, indeed, there have been no takers for this scheme in practice.

Loeb and Magat were by no means the only contributors pointing regulatory economists, companies and regulators to the Promised Land of increased efficiency through “better” economic design of regulatory institutions. Vogelsang and Finsinger (1979) avoided the problem of common knowledge of an entire demand or cost function, but their proposal had problems of its own. The V-F scheme proposed a quasi-dynamic
regulatory process under which the regulated firm would be confined to prices that would be a function of their previous levels of revealed costs and demands. Assuming both accurate reporting and stationarity of cost and demand functions, the V-F scheme can be shown to converge in some circumstances to Ramsey prices and outputs. One of the assumptions criticized that gave rise to its own literature was that the regulated firm would optimize its current price myopically in each period, without regard for the effects of the current pricing decision on constraints that would be imposed on the firm in the future via the operation of the V-F scheme. This gave rise to an interchange between V-F, Sappington (1980) and others. We are not going to track the outcome of this discussion here, noting only that there have not yet been any takers in practice for the V-F theory.

Dissatisfaction with these “old style” approaches to incentive regulation underlay Laffont and Tirole’s (1993) *tour de force* on regulation, which illustrates well the approach taken and the high expectations generated.4 “In the policy arena discontent was expressed with the price, quality, and cost performance of regulated firms and government contractors...More powerful incentive schemes were proposed and implemented, deregulation was encouraged... [but] regulation theory largely ignored incentive issues.” (Laffont and Tirole 1993, xvi) Previous regulatory theory, they argued, “…did not meet the standards of newly developed principal-agent theory, whose aim is to highlight the information limitations that impair agency relationships. Furthermore the considerably simplified formal models that assumed away imperfect information were less realistic in that they implied policy recommendations that require information not available to regulators in practice.” While we accept that these criticisms certainly have some validity, we argue that the contributions that replaced them were at least as limited in their applicability and fell far short of the expectations created by their authors. Ironically, a principal reason for this is precisely the reason raised above by Laffont and Tirole in ushering in the new theory, namely, a heavy reliance by such schemes on information that is not available to regulators. Indeed, the entire mechanism design literature, beginning with Baron and Myerson (1981) and strongly promoted by Laffont and Tirole, is based in one way or another on assumptions like common knowledge that endow the regulator with information that he cannot have without a contested discovery process that always leaves him in a state far short of the level of information assumed in these theories.

The promise of these mechanism-design-style theories was ostensibly considerable. They promised none other than the holy grail of X-efficiency, something previous regulation had manifestly failed to deliver. X-efficiency, however, was only achieved if two conditions - aside from the basic assumptions criticized above - were met. The first condition was that achievement of the promised X-efficiency required that the regulator concede some information rents to the firm.5 The second condition was what is referred to in mechanism design theory as commitment. This is the notion that the presence of information rents would not present a problem to the regulator and that, as a result, he was committed to his original agreement with the firm. In other words the *ex post* appearance of excess profits would not cause the regulator to renege on his commitment to the original incentive scheme. Why this would not be a fatal flaw in the whole scheme was never considered. The new theory promised efficiency as long as the regulator is prepared to allow

---

4 Although we were critical of the “new style” regulatory economics, Crew and Kleindorfer (1986), we also joined the party on incentive regulation, Crew and Kleindorfer (1996). It should be noted that several papers have raised questions about the “new style” economics, including Schmalensee (1989), Braeutigam and Panzar (1992) and Lyon (1994), each of whom raise some doubts on the applicability of pure price caps versus hybrid mechanisms including deadbands and sharing.

5 These rents arose from the information advantages of the firm relative to the regulator.
information rents. How do these rents differ so much from the old style monopoly rents that would make them acceptable to the regulator when it was monopoly rents that were the principal motivation of regulation in the first place? Thus, the promise of the holy grail of X-efficiency was hedged with conditions which, we argue, make the theory of little significance for real world regulation, as subsequent events have shown.

Of course, all theory makes simplifying assumptions, which depart from reality in one way or another. The key, at least for normative economic models, is that the assumptions made should not give rise to fundamental infeasibilities when implementing the results of the theory in practice. An apparent case in which this has not been true is the case of regulatory theory derived from the mechanism design literature. Other than being a rich source of classroom exercises, this theory seems to have found no takers in practice. A consideration for the reasons underlying this failure may provide useful insights for the future innovations needed in development and application of regulatory theory. This is, indeed, one aim of this paper.

While further elaborations of these theories of incentive regulation may lead to innovations for practice, at present the only innovation of significance in the past thirty years remains PCR.\(^6\) Notwithstanding some of the gains that have accompanied the implementation of PCR, particularly making good on Littlechild's original bet, the limitations of PCR have remained center stage. Indeed, a chorus of economists has continued to voice the concerns of Littlechild and more. The central theme recently has focused on regulatory commitment as the Achilles heel of PCR. This issue is our major concern here.

The mechanism design theory of regulation was careful to show that to achieve X-efficiency on the part of the regulated firm that it would be necessary for the Regulator to leave some of the rents to the firm arising from asymmetric information on the table for the firm. Theorists, however, never understood the impossibility of this in practice. No regulator can even admit that it allows the firm to retain information rents let alone commit to such a practice. For the regulator this is a congenital problem of far greater magnitude than has been recognized in economic theory.\(^7\) This is a constraint that must be recognized in developing models in regulatory economics. Like the mechanism design theorists, we recognize the importance of commitment and its effect on incentives. The difference between our approach and the mechanism design approach is that we argue for a more realistic view of the constraints on regulatory commitment that are likely to be fundamental to the very nature of regulation. As we will see, taking such a realistic view gives rise to a rather different approach to the design of incentive regulation than the pure view of previous theory.

Recognizing the importance of institutional constraints, our approach sees commitment as a two-sided process requiring commitment not just on the part of the regulator, which is the case in the mechanism design approach, but also on the part of the Firm. Approaches that start with the economist's ideal view of the world, which then attempt to impose it on the real world of regulation, create unrealistic aspirations and may lead to greater inefficiencies than the original economic institutions and models which the mechanism design theorists criticized.

---

\(^6\) Of course, there have been significant developments in regulatory economics outside the realm of incentive regulation. These include developments in the understanding of costs, in access and peak-load pricing, in the application of auction theory and in models of competitive entry in markets with dominant firms. These developments can be traced in the *Journal of Regulatory Economics* since its inception.

\(^7\) Loeb and Magat, and Vogelsang and Finsinger implicitly rely on this same notion of commitment.
2. A Stylized Model of the Regulatory Process

We consider a two-period model of the regulatory process. Each of the two periods may be thought of as a period of normal price cap operation, of duration, say, 4-5 years. Thus, the price cap is intended to operate uninterrupted during each of the two periods in the sense that the parameters of the price cap process are unchanged during each of these two periods. We envisage that the Regulator would see no need to adjust the price cap parameters if the operation of the price cap during the first period yields “expected results” in terms of the level of profits attained by the Firm. In the event that profits are outside these expected bounds, the Regulator is assumed to respond between period 1 and period 2 by adjusting the parameters of the price cap. The intuitive interpretation of the lower bound on profits is that of impending bankruptcy of the firm, which we assume is not an allowable outcome, i.e., the threat of bankruptcy will definitely lead to a regulatory hearing to determine appropriate solutions to avoid bankruptcy. On the other end of the spectrum, we assume that profits sufficiently above the normal or expected level, even if they are the result of efficiency enhancements that will be long-lasting and therefore shared eventually with the consumer, are not politically acceptable. The appearance of such profits will thus also lead to regulatory intervention.

We assume the following sequence of events:

Prior to Period 1: At an omnibus rate hearing, the Regulator sets the parameters of the PCR regime (e.g., the initial price cap and the “X factor”) based in part on the Firm’s existing rate base. Thereafter, the regulated Firm determines its incremental investment in cost-reducing innovation.

Period 1: The Firm sets price subject to the price cap and meets all demand arising from its price. The Firm may engage in pure waste to avoid regulatory review at the end of the period, and while the Firm’s total cost, revenue and profit results are perfectly observable ex post, the Regulator is unable to observe whether the Firm has engaged in pure waste or not; only total expenses are observable.

Prior to Period 2: If the Firm’s profits are excessively high or excessively low as judged by its achieved rate of return in Period 1, the Regulator and the firm undertake a full rate hearing and reset rates to achieve a benchmark return. If the Firm’s returns on invested capital are within a pre-specified band, the Firm is allowed to proceed with a minimal review to a second period of PCR regulation, (which we assume is) subject to the same parameters as for Period 1.

Period 2: The Firm sets price subject to the price cap and produces output as in Period 1.

We use the following notation.

\[ H = H_R + H_F = \] Cost of a full rate hearing prior to period 1 and prior to period 2 if needed, where \( H_R \) are the costs borne by the Regulator and \( H_F \) are the costs borne by the Firm.

\[ P_C = \] Level of the price cap in period \( t = 1, 2 \), set by the Regulator.

\[ D(P, \omega_t) = \] Demand function in period \( t = 1, 2 \), where \( \omega_t \) is a real-valued random variable, with cumulative distribution function \( G(x) = Pr(\omega_t \leq x) \), and where \( D(P, \omega_t) \) is an increasing function of \( \omega_t \). One source of this uncertainty in the electric context would be weather. We assume \( \omega_1 \) and \( \omega_2 \) are statistically independent with the same distribution \( G(x) \).
\( F(K, L, M) \) = Production function for the Firm, a function of a long-term factor \( K \), a short-term factor \( L \) (set after observing the state of the world \( \omega_t \)), and \( M \) is a capital expenditure on productivity enhancements. We assume that \( F \) satisfies the usual properties of a neoclassical production function.

\[ N_t = \] Pure waste, in dollars, in period \( t = 1, 2 \).

\[ rK + wL + rM = \] Total production-related costs, with \( r \) the period cost of capital and \( w \) the cost of the short-term factor \( L \).

\[ S_{D}, s, S_{E} = \] Levels of return on capital per period, where \( S_{D} \) is the “distress level”, going below which is assumed to trigger a full rate hearing before period 2, \( s \) is the normal or expected rate of return, and \( S_{E} \) is the “excessive level” of return, going above which is assumed to trigger a full rate hearing before period 2.

Assuming the short-run factor \( L \) is adjusted to meet demand \( D_t = D(P_t, \omega_t) \) in each period \( t \), we can define the Firm’s period \( t \) cost as:

\[
C(K, M, D_t) = r(K + M) + wL o(D_t, K, M) ; \quad t = 1, 2
\]  

(1)

where the short-term factor input decision \( L^o \) solves \( F(K, L, M) = D \) for given \( K, M, D \), i.e.

\[
F(K, L o(D_t, K, M), M) = D_t; \quad t = 1, 2
\]  

(2)

Ignoring discounting, the expected profits of the Firm across the two periods are:

\[
E \Pi(P, K, M, N) = E \left\{ \sum_{t=1}^{2} \left[ P_t D(P_t, \omega_t) - C(K, M, D(P_t, \omega_t)) - N_t \right] \right\}
\]  

(3)

Next we formulate the regulatory constraints the Firm faces, including the requirements for a rate hearing prior to period 2. These are expressed in terms of trigger events related to the observed returns on capital, where period \( t \) returns \( R_t \) are given by:

\[
R_t = R(P_t, D_t, K, M, N_t, \omega_t)
\]

\[
= \frac{P_t D(P_t, \omega_t) - wL o(D(P_t, \omega_t), K, M) - N_t(\omega_t)}{K + M}
\]  

(4)

Note that we assume that both \( K \) and \( M \) are capitalized in the Firm’s rate base and we ignore depreciation here. Note also that \( N_t \) is shown as a function of \( \omega_t \), reflecting our assumption that wasteful expenditures, if any, are decided upon after the state of the world is known in each period.

We are not concerned here with the determination or effects of the X-factor and we assume it is equal to zero. The Firm therefore faces a constant price cap; prices are required to be set such that \( P_t \leq P_{Ct} \), the price cap in period \( t \). We assume that the price cap is set at the initial rate hearing so that the return level \( s \) will be attained if the Firm continues its current/historical productivity performance, i.e., it invests nothing to enhance productivity and expected demand conditions obtain. Thus, it is assumed that the price cap is set
initially through the following expected revenue-expected cost equality:

\[ P_{CI} ED_1(P_{CI}) = sK + wEL(P_{CI}, 0) \]  

where

\[ ED(P) = E_\omega \{ D(P, \omega) \}; \quad EL(P, M) = E_\omega \{ L^\omega(D(P, \omega), K, M) \}; \quad t = 1, 2 \]  

The target return level \( s \) is determined at this initial rate hearing following established financial procedures, possibly adjusted to reflect the Regulator’s opinion that the Firm can achieve significant cost reductions through investments in productivity and innovations. We assume that the return levels \( S_D \) and \( S_E \) are determined exogenously, and both are known to the Regulator and the Firm before the fact. While (5) may have the appearance of a traditional RoR-like revenue requirement, its only similarity is that a benchmark rate of return is envisaged in setting the initial price cap.\(^8\) The Firm can (and is expected to) achieve something in excess of this through increased efficiencies during the first price cap period.

Note that (5) determines only the first-period price cap level \( P_{CI} \). We now turn to the determination of the second-period price cap level. This will depend on the outcomes achieved at the end of period 1. The following represents the possible outcomes at the end of Period 1:

- \( S_D \leq R_t \leq S_E \): No rate hearing is scheduled; the Firm faces the price cap \( P_{C2} \) determined by

\[ P_{C2} ED(P_{C2}) = sK + wEL(P_{C2}, 0) \]  

Comparing (5) and (7), we see that in this case \( P_{C2} = P_{CI} \), i.e. no regulatory review is triggered; therefore there is no change in the price cap. Note that the Firm obtains the full benefits of any innovations it makes in the first period throughout the second period. Adding an \( X \) factor to this expression would not change any of the logic to the discussion below. The point is simply that the Regulator and the Firm understand prior to period 1 that returns in the indicated, normal range will give rise to a second period of price caps with no detailed regulatory hearing and with parameters basically the same as those prevailing during the first period. This constancy of regulatory regime is implicitly or explicitly assumed in extant theory (e.g. Laffont and Tirole 1993, Loeb and Magat 1974, Vogelsang and Finsinger 1979) and is essentially what we believe most authors have in mind as “regulatory commitment”.

- \( R_t < S_D \): A rate hearing is triggered, and the price cap is adjusted upwards to reflect a sharing of the Firm’s “losses” relative to the benchmark level of return \( s \). It is assumed that the Regulator revises the price cap, setting the price cap for period 2 as the solution \( P_{D2} \) to the following relationship:

\[ R_t < S_D \]

---

\(^8\) ROR considerations clearly played a role in PCR in the United Kingdom, despite Littlechild (1983, 1986) excoriating remarks on ROR. For example, Beesley and Littlechild (1989) clearly recognized the role of ROR in British PCR. Referring to the Monopolies and Mergers Commission’s approach in the Manchester Airport case they state, “The rate of return on (historic) book capital was only one of nine financial projections and ratios that the MMC looked at.” (p 459) On the British Telecommunications’ Regulator they say, “He gave no detailed explanation for his choice of \( X \), beyond indicating that rate of return was the most important criterion, but not the only one.” (P460)
\[ P_{D_2} ED(P_{D_2}) = \left[ s(K + M) + wEL(P_{D_2}, M) \right] \]
\[ + \zeta \left[ S_D(K + M) - \left( P_1 D(P_1, \omega_1) - wL^0(D(P_1, \omega_1), K, M) - N_1(\omega_1) \right) \right] \]  

(8)

where \( \zeta \in [0, 1] \) is a “sharing parameter”. The first term on the r.h.s. of (8) reflects revenue requirements for the Firm to achieve returns \( R_2 = s \) if expected demand occurs in period 2 (recall that \( E(\omega_i) = 1 \)) and the second term reflects that a share of the difference between achieved period 1 returns and distress-level returns \( S_D \) will be passed on to consumers through an increase in the price cap in period 2. Note that the second expression in brackets is positive by assumption, so that (8) does reflect a positive adjustment to the price cap in the event of distress-level profits. Note also that it is assumed that the rate hearing is able to ascertain accurately the level of the total investments \( K + M \) as well as the required factor payments for the short-term factor \( L \) going forward.

\( R_1 > S_E \): A rate hearing is triggered, and the price cap is adjusted downwards to reflect a sharing of the Firm’s “excess profits” relative to the benchmark level of return \( s \). It is assumed that the Regulator revises the price cap, setting the price cap for period 2 as the solution \( P_{E_2} \) to the following relationship:

\[ P_{E_2} ED(P_{E_2}) = \left[ s(K + M) + wEL(P_{E_2}, M) \right] \]
\[ + \zeta \left[ S_E(K + M) - \left( P_1 D(P_1, \omega_1) - wL^0(D(P_1, \omega_1), K, M) - N_1(\omega_1) \right) \right] \]  

(9)

where \( \zeta \in [0, 1] \) is a “sharing parameter”, which is assumed to be the same as for distress-level returns. Note that the second expression in brackets is negative by assumption, so that (9) does reflect a downward adjustment to the price cap in the event of excess profits. As in the case of distressed profits, we assume that the regulatory hearing accurately determines the level of the rate base and the magnitude of short-term factor costs required at the expected demand level going forward.

We are assuming here that the rate hearing, made necessary by period 1 below distress-level or above excess-profit-level returns, uses only the observed total costs (including waste if any) incurred by the Firm. The key to this model is the descriptive model used by the Firm to predict the outcome of the rate hearing at the end of period 1 if it occurs. Different such models will give rise to different first-period pricing and investment behavior by the Firm. The model in (6)-(8) is but one such model that the Firm might use. For example, the Regulator could establish precedents for how wasteful spending would be determined and what (dis-)allowance rules would be used.\(^9\) In general, the Regulator could, through its behavior over time, establish precedents as to how it would behave in estimating sharing results for either distress levels of excess levels and these precedents would clearly affect the predicted model a rational Firm would use to predict the outcomes of future regulatory hearings should they be triggered by events.

We now revisit the expected profit expression (3), with the aim of determining the optimal prices and waste decisions, given the above anticipated regulatory actions. We begin with an assumption to assure that

---

\(^9\) Regulators may not distinguish between different kinds of waste. In some cases waste may take the form of rents in the form of excessive bonuses to top management. In other cases it may involve inefficient input combinations. The former may be easier to detect than the latter although the same dollar amount of waste causes greater inefficiency losses in the latter case.
returns outside of the expected zone \([S_D, S_E]\) are in fact possible.

**Assumption (Possibility of Unacceptable Returns):** We assume that the random variable \(\omega\) has compact support, \(\Pr\{0 < \omega_L < \omega < \omega_U < \infty\}\), and that \(R(P_{C2}, D_0, K, M, 0, \omega_L) < S_D\) and \(R(P_{C2}, D_2, K, M, 0, \omega_U) > S_E\), so that if price in period 1 is set at the benchmark price cap level in (5), \(P_1 = P_{C1}\), and if the Firm does not engage in waste, \(N_i(\omega_i) = 0\), then distress-level returns obtain at the minimum possible state of the world \(\omega_L\) and excess profits obtain at the maximum possible state of the world \(\omega_U\). [Recall that \(D(P, \omega)\) is increasing in \(\omega\).]

Given this assumption, there is some non-empty interval \(\Omega(P_1) = \{\omega_D(P_1), \omega_E(P_1)\}\) over which regulatory review at the end of period 1 takes place, i.e., such that:

\[
R(P_1, D_1, K, M, N_1(\omega_1), \omega_1) < S_D, \text{ for } \omega_1 < \omega_D(P_1)
\]

\[
S_D \leq R(P_1, D_1, K, M, N_1(\omega_1), \omega_1) \leq S_E, \text{ for } \omega_D(P_1) \leq \omega_1 \leq \omega_E(P_1) \tag{10}
\]

\[
R(P_1, D_1, K, M, N_1(\omega_1), \omega_1) > S_E, \text{ for } \omega_1 > \omega_E(P_1)
\]

Now we note the following facts concerning optimal period prices and wastes.

**Proposition 1:** Price in period 1 is always equal to the price cap \(P_{C1}\). Assuming that the Firm operates in the inelastic region of expected demand, and that the benchmark return \(s > r\) and less than the monopoly return, optimal prices are the maximum of those allowed. Thus, for period 1, the optimal price is set to the price cap level determined by (5). For period 2, price is also equal to the allowed price cap \(P_{C2}, P_{C2}\) or \(P_{C2}\) depending on which of the three possible cases in (10) occurs at the end of period 1). Optimal wastes \(N_i(\omega_2) = 0\), for all \(\omega_2\) while \(N_i(\omega_i) = 0\) whenever \(\omega_1 < \omega_D(P_1)\), and in particular whenever \(\omega_1\) either leads to distressed level profits or expected profits. When \(\omega_1\) leads to excess returns, \(N_i(\omega_i) = 0\) if and only if:

\[
(1 - \zeta)[P_1D(P_1, \omega_1) - wL^o(D(P_1, \omega_1), K, M) - S_E(K + M)]
\]

\[
> H_F + w[EL(P_{C2}, 0) - EL(P_{C2}, M)] - sM \tag{11}
\]

Thus, denoting by \(\omega_w(P, M)\) the value of \(\omega_1\) for which equality obtains in (11), the Firm engages in pure waste whenever \(\omega_1 \in (\omega_E(P_1), \omega_w(P, M)]\), with the level of waste \(N_i(\omega_i) > 0\) then determined by:

\[
N_i(\omega_1) = P_1D(P_1, \omega_1) - wL^o(D(P_1, \omega_1), K, M) - S_E(K + M)
\]

\[
where \quad \omega_E(P_1) < \omega_1 \leq \omega_w(P_1, M). \tag{12}
\]

The proof of this proposition is given in the Appendix. The intuition behind the proof is straightforward. The fact that optimal prices are as high as allowed follows from our assumption that demand is inelastic and
that profits are positive for the maximum possible $\omega$.\textsuperscript{10} No pure waste is optimal in the second period for the same reason; nothing would be gained by the Firm by “throwing money out” in the final period. In the first period, the Firm must reason as follows, noting that $N_\rho$ is determined after the state of the world $\omega_1$ is observed. If the Firm is in the expected zone of profits, clearly there is nothing to be gained by pure waste. If the Firm is in the depressed profit zone even when zero waste is undertaken, then engaging in positive waste would simply depress profits further and the ensuing rate case would only return a fraction $\zeta < 1$ of this waste to the Firm.\textsuperscript{11} However, when returns are only slightly above the excess level the Firm would avoid a rate hearing by engaging in some pure waste. But if returns are significantly in excess of the ceiling $S_E$, then the Firm would have to sacrifice too much to avoid the fine-tuning of the rate hearing. The precise break-even point between these two regimes can be computed by comparing the two decisions (engage in just enough pure waste to avoid a hearing versus engage in no waste and go through a hearing and resetting of the price cap according to the rule specified in (9)). The result of this comparison is summarized in (11).

Concerning the interpretation of (11), the basic intuition is that if excess profits are not too high, the Firm finds engaging in some pure waste to avoid a rate hearing. The l.h.s. of (11) is the incremental profit above the trigger return level $S\rho(K + M)$ that the Firm would be allowed to keep if a rate hearing were triggered (the fraction $\zeta$ of the excess would be returned to customers in the event of a rate hearing). The r.h.s. reflects the change in expected short-term factor costs plus investment costs that would be recognized if a rate hearing were to take place. When investments $M$ in productivity are (weakly) efficient, then the r.h.s. of (11) will be (non-negative) positive. Recall that if no hearing takes place, then the Regulator continues to set price caps as if productivity in the Firm had remained unchanged. The intuition behind (11) is that if the realized excess profits is small, then the Firm will find it profitable to avoid a rate hearing by engaging pure waste. If, however, the realized excess profits are large, then the Firm will find it more profitable to recognize these profits on its books and share them with ratepayers. In doing so, the Firm would forego the additional revenue it would receive for its investments in productivity improvements, such revenue being only determinable on the basis of a rate hearing. The set of $\omega_1$ for which (11) is satisfied and pure waste occurs will be the larger, the larger $\zeta$ is, the smaller $S_E$ is, and the larger the productivity gains from investments in M are. If the Firm also had to bear some part of the fixed costs $H$ of the regulatory process triggered by excess profits, then these costs would be added to the r.h.s. of (11) and the set of $\omega_1$ for which pure waste.

Given the above Proposition, there are essentially four regions of outcomes of interest, below the distress level (region D), expected or normal levels (region N), above the excess level by less than the l.h.s. of (11), giving rise to waste (region W), or above the excess level by at least the l.h.s. of (11) (region X). Treating $K$ as fixed, we can express the expected profit function (3) as a function of two decision variables $P_1$ and $M$

\textsuperscript{10} This is obvious for period 2 price, but requires a little bit of algebra for the first-period price. It results from the fact that the FOC w.r.t. $P_1$ of the expected profit function (13) is increasing. Intuitively, no matter what state of the world occurs, the options available to the Firm are better at the end of period 1 if Revenue is higher (i.e., if price $P_1$ is higher).

\textsuperscript{11} This result follows because, in the regulatory review process modeled here (see (8)), we assume that there is no discontinuity in sharing payments to the Firm when its performance goes below the distressed profit level. As Yasugi Otsuka pointed out, if the Firm were to receive a lump-sum payment, discontinuous in observed profits, then the Firm might well waste money also on the downside. For example, if the Firm were to be given rate relief through the second-period price cap that made the Firm whole relative to, say, the benchmark return level "s" in the event that its realized returns were below the level $S_0 < s$, then the Firm would find it optimal to engage in pure waste when observed returns in period 1 were anywhere in the open interval $(S_0, s)$. 

and the random state of the world $\omega_1$. We use these four regions of interest to compute the expected profits for the Firm over both periods as follows:

$$ E\Pi(P, K, M) = \int_{\omega_L} \Pi_D(P_1, M, \omega_1) dG(\omega_1) $$

$$ + \int_{\omega_L} \Pi_H(P_1, M, \omega_1) dG(\omega_1) $$

$$ + \int_{\omega_L} \Pi_w(P_1, M, \omega_1) dG(\omega_1) $$

$$ + \int_{\omega_L} \Pi_x(P_1, M, \omega_1) dG(\omega_1) $$

(13)

where the profit functions for each of these regions are computed from (7)-(11) (see the Appendix for details) as follows:

$$ \Pi_D(P_1, M, \omega_1) = $$

$$ (1 - \zeta) \left[ P_1 D(P_1, \omega_1) - wL_0(D(P_1, \omega_1), K, M) \right] + (s + \zeta S_D - 2r)(K + M) $$

$$ + P_{C_2} E(D(P_{C_2}) - wE(P_{C_2}, M) $$

(14)

$$ \Pi_H(P_1, M, \omega_1) = P_1 D(P_1, \omega_1) - wL_0(D(P_1, \omega_1), K, M) - 2r(K + M) $$

$$ + P_{C_2} E(D(P_{C_2}) - wE(P_{C_2}, M) $$

(15)

$$ \Pi_w(P_1, M, \omega_1) = $$

$$ (S_F - 2r)(K + M) + sK + w \left[ E(P_{C_2}, 0) - E(P_{C_2}, M) \right] $$

(16)

$$ \Pi_x(P_1, M, \omega_1) = P_1 D(P_1, \omega_1) - wL_0(D(P_1, \omega_1), K, M) - 2r(K + M) - H_F $$

$$ + s(K + M) + \zeta \left[ S_F (K + M) - \left( P_1 D(P_1, \omega_1) - wL_0(D(P_1, \omega_1), K, M) \right) \right] $$

(17)

Substitution of (14)-(17) into (13) yields the desired expected 2-period profit, as a function of the decision.
variables $P$, and $M$. We are not going to go through the detailed analysis of first-order conditions for these decisions. We wish to note only briefly the impact of the constraints in this problem on the decision variables. We note from Proposition 1 that prices are determined entirely by the Regulator in this simple model (i.e., the Firm's decisions always lead to setting prices at their regulatory upper bounds). As is clear from (12), wastes are determined in part by regulatory policy (the setting of the price caps, sharing parameters and policies affecting the fixed costs of regulatory hearings on the Firm). We return to this point below.

Concerning the value of productivity investments $M$, the situation is quite complex. The efficient choice of productivity investment $M$ is easily seen to be the solution to the following problem:

$$
\text{Minimize} \left[ 2rM + wEL(P_1, M) + wEL(P_2, M) \right] \quad (18)
$$

The actual incentives faced by the Firm under the above regulatory regime, as embodied in (13)-(17) are rather different from those embodied in (18). Indeed, considering the four profit regions described in (14)-(17), we see that incentives in the region D are enhanced for investment, since the regulator will pick up part of the tab in this region. In region N, the incentives are the normal incentives for an expected profit-maximizing firm, i.e. the incentives in this region would reinforce efficiency since the Firm bears the full consequences of its actions in this region. In region W, the incentives for efficiency are weakened since any benefits achieved through such productivity enhancements will (rationally) be discarded. Finally, in region X, incentives are also weakened since the Regulator will insist on sharing the benefits of productivity enhancements (and other factors leading to excess profits) with consumers. Exactly, how these different incentives play out will clearly depend on the probabilities of being in each of these regions, together with the impact of regulatory parameters that influence the magnitude of these effects in each region. General results (such as over- or under-investment) are not likely to be available. What can be said is that regulatory commitment is not the issue here; regulatory judgment is.

3. Implications for Incentive Regulation in Postal Service

Price regulation has traditionally not been a major issue in postal service. However, with liberalization of markets and the creation of formal and independent regulatory institutions this is changing and the nature and role of regulation has become an important issue. Where regulation has been proposed it has not been the product of profound analysis. For example, in Crew and Kleindorfer (1994) we proposed price cap regulation for the United States Postal Service (USPS) based upon the notion of pure price caps. Pure PCR, in addition to the gains in X-efficiency examined in section 1, has a number of other advantages that make it ostensibly attractive to postal service. Not least is the way PCR resolves the cross subsidy issue. As Braeutigam and Panzar (1989) showed PCR does not have the incentives that ROR has to misrepresent the costs of the regulated product to make it possible to cross subsidize competitive products. This is a particularly important problem in postal service as, under liberalization, a PO would find it advantageous to be in both monopoly and competitive markets. In the Postal Reform Bill (H.R. 22) currently before the U.S. Congress similar price cap proposals to the ones we made are proposed. Unfortunately, this approach runs into a number of problems.12

---

12 Braeutigam and Panzar were appropriately circumspect in terms of this apparently extremely attractive feature of PCR. They conclude: “...price-cap regulation will not perform as well as it does in a static completer information model with only one core service.” (1989, 390)
The first problem arises from the application of PCR to a public enterprise. To achieve the claimed efficiency of PCR requires residual claimants, which are so weak as effectively do not exist under public ownership, as we argued in testimony before the House Sub Committee on Postal Service on April 16, 1997 and in Crew and Kleindorfer (2000). This problem we will not address here but rather be concerned with the second problem. If postal service is provided by a regulated privately owned monopoly, then what form of regulation is appropriate? What lessons can be drawn from the analysis in sections 1 and 2 and experience of other industries? In addressing these questions we will assume that postal service is a regulated privately owned company, thus making it comparable with the economic theory of regulated monopoly and the practice in other industries.

The concern about pure price caps is that the unrealistic notion of commitment assumed in the theory cannot be achieved in practice. This has been the case in other industries where regulatory arrangements have been developed to enable both the firm and the regulator to address the commitment issue. Indeed, we have argued that achieving commitment in regulated industries is feasible, as long as commitment is understood in realistic terms. Our model has provided an illustration depicting commitment as a two-sided process, which is the product of negotiations between the firm and the regulator while recognizing the constraints that are essential to the context of these negotiations. In contrast to the mechanism design theory where commitment is seen as a factor effectively exogenous to the firm and where there is nothing that the firm can do except rely on the regulator’s commitment, the necessary commitment in this case is something that is worked out by the parties. The device used in our model to reflect two essential constraints (bankruptcy and no excess profits) is triggered regulatory hearings and sharing. In particular, if the firm’s earnings fall outside a certain range, sharing on the upside and downside is triggered. This, at least partially, addresses a major problem faced by real world regulators of how to address high profits as there is always the concern that they arose not so much from efficient operation or good fortune but from monopoly exploitation. The regulator clearly has to be sensitive to such issues.

There is some evidence to support the argument that in the real world of regulated industries the firms and the regulators develop methods of addressing the commitment problem and arrangements that reflect the characteristics of the industry. If, indeed, firms and regulators are able to address the problem of commitment in the way we argue then we would expect to find relatively few instances of pure price cap regulation and relatively many instances of sharing. As partial support for this argument, consider the data from a recent survey of incentive regulation by the Edison Electric Institute\(^{13}\) that reviewed incentive regulation plans for 32 plans recently implemented for electric utilities and distribution companies throughout the U.S. In 14 instances revenue sharing, which would be consistent with our general model, was explicitly featured as an integral part of the plan. In 16 instances benchmarks for e.g. service quality were featured as part of the plan. There were 6 instances of price cap plans unqualified by sharing or other similar devices. However, in all these 6 cases it is not clear that the price caps involved would qualify as “pure price caps” as envisaged in the mechanism design theory. They all seem to have been associated with some kind of “rate freeze”, which was likely part of a deal for stranded cost recovery.\(^{14}\) Thus, in one way or another, these plans reflect the fact that companies and regulators were highly cognizant of the need to be concerned about excessive profits. Revenue sharing clearly addresses this issue. Similarly, stranded cost recovery performs a similar function. Any excess “profits” in this case are deemed stranded cost recovery,

\(^{13}\) PBR Survey (EEI Member Survey), conducted September, 1998; updated March, 2000.

\(^{14}\) A notable case of this was California, where the PBR plans included price caps and frozen rates with stranded cost recovery explicitly featured as part of the plan.
letting the regulator off the hook for allowing the company to make what would otherwise be excessive profits, but are strategically presented as allowing the company to recover its previously stranded costs.

While it could be argued that these considerations do not apply to the postal sector and the postal sector is “different” we disagree. It is true that because postal service has little sunk capital the issue of stranded costs is not a major one. However, all postal services face transition problems as more of their markets are opened up to competition. So while they may not have stranded costs of the magnitude seen in fixed network industries they may have non-trivial transition costs. Setting aside this problem the real issue is whether regulators of postal service are likely to be any different from regulators of other industries. In particular, are postal regulators going to behave in the manner prescribed by the mechanism design theorists? Are they going to have the neutral attitude required toward excess profits and are they going to be unconcerned about the threat of bankruptcy? In our view this is unlikely to be the case. With this in mind we are going to examine briefly how incentive regulation might operate in postal service.

In postal service, just as in telecommunications and electricity there are many opportunities for regulators to renege on their commitments. The familiar means include: price reductions, increasing the X factor, and allowing additional entry into the regulated firm’s monopoly markets. Price reductions are completely transparent. From a regulator’s point of view this could, under different circumstances be advantageous or not. If the regulator wishes to punish the firm for what it considers some misconduct or breach of faith, transparency might be attractive. However, not all issues can be categorized as black or white in which case a highly transparent action might make the regulator vulnerable to being overruled by appeal to the next level. The regulator may therefore prefer something less transparent but something that is well understood by himself and the regulated firm. Increasing the X factor and entry fall into this category.

Postal service’s situation is closer to that of electricity in the sense that its productivity and demand are slow growing or even potentially declining than it is to that of telecommunications where technological change is rapid and demand is growing quickly. In such a situation small changes in the X factor could have a significant effect on the profitability of a PO. An increase of 10 basis points in the X factor for postal may significantly hurt profits whereas a full percentage point in telecommunications may be taken in stride with profits still increasing at a good clip. Postal service is also very vulnerable to entry. If the regulator enforces a tough universal service obligation yet at the same time allows increased entry into the PO’s reserved area the profit consequences could be serious.

The implication of these considerations would appear to be that it is perhaps even more important for regulators and POs to develop arrangements to achieve a mutually sustainable commitment than in other industries because the margin for error in the postal sector seems to be smaller. Thus, the lessons of other industries in achieving commitment would appear to be worthy of consideration in the postal sector. We consider briefly the issues of sharing and quality monitoring.

Given the likely sensitivity of a PO’s earnings to small changes in the X factor, sharing would seem to hold out considerable benefits to both the regulator and the PO. From the regulator’s point of view if he set the X factor to low and earnings are higher than expected the sharing mechanism will trigger. This makes the regulator less vulnerable to charges that he allowed excess profits by being to easy on the firm. In other words it shields him from the charge that he did not do his job properly. The PO will also see some benefits in earnings sharing. It will be particularly concerned with the problem of the regulator setting the X factor too high with the result that it makes a low return. Here earnings sharing will trigger automatic rate relief.
The issue of quality is particularly important to POs and their regulators. Traditionally, POs have been vulnerable to criticism of low quality. Under private ownership this would continue and enforcing quality standards is an important way that a postal regulator will make his presence felt. In addition, to careful monitoring and reporting of standards achieved the regulator is likely to find beneficial the use of incentives to attain and exceed minimum quality standards. Tying the PO's earning to achieving quality standards is likely to be an important feature of postal regulation. The lessons learned in electricity and telecommunications may be of interest here. These indicate the need for the design of very careful incentives. Performance based incentives by their very nature are tricky. PG&E responded very effectively to performance based incentives for Diablo Canyon nuclear plant, achieving plant factors and reliability levels far in excess of previously achieved. They were rewarded handsomely in terms of earnings, so handsomely in fact, that the California Public Utilities Commission revised the terms of the agreement after a several years of operation. It is quite difficult to design quality incentives. If they are set too high then excess profits and excess quality are triggered and vice versa. This provides another reason why sharing may be beneficial because it will attenuate the effects of excess profits from overly generous performance incentives.

Clearly, it is not possible to state that the commitment issue can be resolved in practice. However, what is exceedingly clear from the evidence is that expectations both on the part of regulators and firms are limited as defined in the PBR contracts. Both have a reasonably clear idea of what is expected of either side. Given the complexity of the processes it is not possible ex ante to state every contingency in a PBR agreement. Things have to be worked out as unforeseen situations emerge. As Goldberg (1976, 427) stated "...if one is interested in seriously analyzing regulation and other institutional alternatives, it is necessary to open up the 'black box' of contract." Goldberg (1976, 427) criticized economists "...for failure to appreciate the complexity of contractual arrangements in the private sector... [making]... the analyst unduly sanguine as to the efficiency of private market solutions to problems in the regulatory sector." We would consider this criticism apt when it comes to mechanism design theory and its implication for pure price caps. Evidence of the practice of regulation seems to show otherwise. Regulators and firms seem to be acutely aware that regulation is what Goldberg refers to as an administered contract, requiring the exercise of judgment in the design and operation of PBR plans. In particular, this design and operation are not once-and-for-all, but they adapt to the changing circumstances and outcomes of the industry and company in question. This is an important lesson to consider going forward when it comes to the design and operation of regulation in the postal sector.

4. Summary, Conclusions and Implications for Future Research and Policy

In this paper we have attempted to address a puzzle in the recent work in regulatory economics. The mechanism design literature has led to the notion of pure price cap regulation (PCR). However, in practice, PCR does not take this pure form but rather takes the form of PBR where the high powered incentives of PCR are dulled in various ways. It is prima facie somewhat puzzling that PCR, which offers greater apparent benefits to the Firm and greater X-efficiency, does not drive out the apparently inferior PBR. We provide a partial explanation of this puzzle.

Our explanation stems largely from the failure of the theoretical underpinnings of PCR. The mechanism design literature promises efficiency based on a notion of commitment that is not attainable in the real world of regulation. In effect, the theory proceeds by ignoring an immutable institutional constraint. Such a failure of economic design is likely to lead to less than optimal outcomes and is misleading. It is misleading in that
it may result in attempts to pursue PCR that are bound to failure. The alternative is to base a theory on an economic design that is not flawed in this way, and this is the approach we adopt in this paper. When faced with failure of the theory, the response of the mechanism design theory is to bemoan the failure of regulatory commitment. As we show in this paper, regulatory commitment is possible and does exist. However, it exists in a very different form from that required in the mechanism design literature. As a result some efficiencies are inevitably lost in the process. This is regrettable but seems to be a fact of regulatory life. We contend that it is much better to understand the process as it is than mislead ourselves into believing that efficiencies can be attained that are just unattainable because of institutional constraints. We summarize the results of our efforts below and then proceed to discuss briefly some implications and extensions.

The basic logic of the above critique in Section 1 and model development in Section 2 is that normative analysis of regulatory policies and institutions should reflect constraints that are likely to be present in reality and that may have a significant impact on the Firm’s behavior. Such constraints will reasonably be anticipated by rational firms. Neglecting these constraints in the design of regulatory policies will lead to erroneous conclusions. In the case at hand, the erroneous conclusions that have been derived for PCR are those that relate to regulatory commitment. We have argued that financial distress and excess profit constraints are likely to be an essential aspect of PCR and PBR in practice. Neglecting them in the analysis of PCR and PBR, either in comparing it to traditional cost-of-service regulation or in setting PBR parameters, will lead to flawed conclusions. As we and others have noted, PBR does have promise for improving regulation over traditional cost-of-service regulation, but it is not perfect by any means. The imperfections in PBR are inherent in the problem of monopoly regulation and the market and political constraints that are part of this setting. It is pointless to define these “imperfections” in the context of monopoly regulation as a problem of flawed regulatory commitment.

In the stylized model we used to illustrate this basic conclusion, we posed a problem in which the constraints associated with bankruptcy and excess profits were explicitly integrated with the model of the regulatory process. We did not derive optimal PCR parameters for this model and contrast them with the parameters that would have derived if these constraints were missing. But it should be fairly obvious that these constraints, in some contexts, would have a fundamental impact on the welfare-optimal values of the parameters of the PCR model.

One strong indication of this is the fact that in our model, incorporating such constraints, there are clear incentives for the Firm to engage in pure waste, something that is not even a possibility in the standard PCR model neglecting the effects of such constraints. The Firm behaves differently depending on its profitability during the price-cap regime and its anticipation as to how such profitability will lead to adjustments in its price cap going forward. When the Firm’s ROR is not above “normal” levels, it has the incentive to operate efficiently and not distort its costs. As the Firm begins to exceed levels at which it expects increased regulatory scrutiny and resetting of its price-cap parameters, it has an incentive to inflate its costs by incurring waste. Where it is below the minimum level it does not inflate its costs by further waste but seeks relief from the regulator. At levels sufficiently above the maximum it will avoid pure waste and share the benefits with ratepayers according the sharing parameter $\zeta$. Of course, $\zeta$ can be considered to be similar to an income tax on marginal income and will itself be distorting in terms of the Firm’s incentives to invest in productivity improvements (via $M$). The magnitude of such incentives will depend on the nature of uncertainty and a host of other technology and demand characteristics of the Firm’s environment.

The level at which the various parameters are set will affect welfare significantly. For example, if the fixed cost of rate hearings is large, and is entirely borne by the Regulator, then the Regulator will want to
make the sharing parameter $\zeta$ fairly large so as to encourage the Firm to engage in pure waste to avoid rate hearings on the upper end.\textsuperscript{15} This will also have the "ex ante appearance" of taking away most of the excess profits of the Firm if they are revealed. But, of course, the higher the level of $\zeta$ (i.e., sharing), the greater will be the incentives for the Firm to engage in pure waste to avoid "true-up" at the end of period 1, i.e., the lower will be the probability of a rate hearing in the first place, whatever the level of potential excess profits are. Similarly, if the fixed costs of regulatory hearings are borne by the Firm, the Firm will also want to engage in waste to avoid the hearings. So fixed costs of regulation actually end up causing wastes, if properly anticipated by either the Regulator or the Firm.

There are several further open questions which we have not addressed. Most notably, the problem of quality standards was raised in the early days of incentive regulation but has received little attention since then, although such standards are being increasingly raised in electric power, as our discussion of the recent EEI Survey results noted. Generally, the problems of incentives discussed here have received very little attention up to now. In the U.S., one can trace this lack of attention partly to the fact that the experience with incentive regulation has been primarily with telecommunications. Here technological change and the accompanying rapid growth of the industry has made it possible to mask these fundamental problems that exist with the practical application of incentive regulation. Now that incentive regulation is beginning to spread to electricity and postal service, these problems are surfacing. In these industries commitment, suppression of rents and quality are on their way to becoming major problems.

Our analysis has only embodied profit constraints. In practice there may be other constraints of a social nature that will trigger action on the part of the regulator. Examples are on the service dimension or bad faith on the part of the firm. Bad service could trigger a review. Actions that appear to be in bad faith could also trigger review, such as laying off thousands of employees just after an initial price cap had been allowed that was based in part on higher employment levels. Other examples would be indicating ex ante a large expense, tree trimming, and then spending much less. Numerous instances of bad faith on the part of the firm could trigger review, which in the context of mechanism design theory could be interpreted as a failure of commitment. The reader will have anticipated by now what our response would be to such an interpretation.

When it comes to designing regulation for the postal sector the lessons of other industries should not be ignored. Pure PCR is no more likely to work in the postal sector than it is in other industries where it has been shown to have serious problems, so serious as to render it inoperable. The postal sector should look to the experience of other industries in designing its regulatory mechanisms. While postal regulation will have to be designed to meet some of the unique characteristics of the postal sector it seems clear that many of the lessons learned in other industries carry over. POs and their regulators will be able to learn much from other regulated industries in designing and operating their own regulatory schemes.

\textsuperscript{15} For notational convenience, we assumed the same sharing parameter in case of both distress and excess, but the logic only applies to the excess conditions. There is no analytical or regulatory reason that the sharing parameter need be identical for both distress and excess.
Appendix

Here we prove Proposition 1. The assertions made concerning price and waste in period 2 are obvious, as noted in the text. Similarly, it is clear that if profits are in the normal zone, nothing is to be gained by engaging in pure waste. We therefore only need to prove the assertions made concerning profits below distressed level or above the excess return level.

Suppose that $\omega_i$ is observed and the Firm is in the depressed profit zone for $N_i(\omega_i) = 0$. The Firm's price in the second period will therefore be $P_2 = P_{D_2}$, and expected profits for both periods will be determined from (3) as:

$$
\Pi_D(P_1, M, \omega_1) = P_1D(P_1, \omega_1) - w L^o(D(P_1, \omega_1), K, M) - N_1(\omega_1) - H_F
+ P_{D_2}ED(P_{D_2}) - EL(P_{D_2}, M) - 2r(K + M)
$$

(A1)

where we note that $H_F$ is incurred by the Firm because a rate hearing is triggered at the end of period 1 (whatever the level of $N_i(\omega_i)$ incurred). Thus, using (7), we obtain

$$
\Pi_D(P_1, M, \omega_1) = P_1D(P_1, \omega_1) - w L^o(D(P_1, \omega_1), K, M) - N_1(\omega_1)
- 2r(K + M) - H_F + s(K + M)
+ \zeta\left[S_D(K + M) - \left(P_1D(P_1, \omega_1) - w L^o(D(P_1, \omega_1), K, M) - N_1(\omega_1)\right)\right]
$$

(A2)

Since the coefficient on $N_i(\omega_i)$ in (A2) is $-(1-\zeta) < 0$, it is clear from this that $\Pi_D(\omega_i)$ will only be depressed further if the Firm sets $N_i(\omega_i) > 0$. Thus, $N_i(\omega_i) = 0$ is optimal in this case as well.

Now consider the case where the Firm is in the excess profit zone when $N_i(\omega_i) = 0$. Logic similar to that leading to (11)-(12) shows that if the Firm increases $N_i(\omega_i)$ but remains in the excess profit zone, then it simply depresses profits. Suppose, however, that the Firm were to increase $N_i(\omega_i)$ just enough to avoid a rate hearing. Doing so would require that first-period returns $R_1$ just equaled $S_E(K + M)$, i.e., from (4), $N_i(\omega_i)$ would be set to achieve the following equality:

$$
P_1D(P_1, \omega_1) - w L^o(D(P_1, \omega_1), K, M) - N_1(\omega_1) = S_E(K + M)
$$

(A3)

Given this scenario of using just enough waste to avoid a rate hearing, we can compute the Firm's joint profits over both periods (conditional on having observed $\omega_i$ and having set $N_i(\omega_i)$ as in (A3)) as follows:

$$
\Pi_W(P_1, M, \omega_1) = P_1D(P_1, \omega_1) - w L^o(D(P_1, \omega_1), K, M) - N_1(\omega_1)
+ P_{C_2}ED(P_{C_2}) - EL(P_{C_2}, M) - 2r(K + M)
$$

(A4)

where the subscript "W" indicates that these profits include some positive waste. Using (7) and (A3) we
obtain

$$\Pi_w(P_1, M, \omega_1) = \left( S_E - 2r \right) (K + M) + sK + w\left( EL(P_{C2}, 0) - EL(P_{C2}, M) \right)$$ (A5)

These joint profits are those which would obtain if the firm were not to allow the excess profit condition to be revealed, setting $N_i(\omega_i)$ just large enough to avoid a hearing. We wish to compare these profits with those which would obtain if the Firm were to allow its excess profits to be observed, setting $N_i(\omega_i) = 0$. As in the logic leading to (A1)-(A2), this scenario would yield joint profits of:

$$\Pi_x(P_1, M, \omega_1) = P_1D(P_1, \omega_1) - wL^o(D(P_1, \omega_1), K, M)$$

$$- 2r(K + M) - H_F + s(K + M)$$

$$+ \left[ S_E(K + M) - \left( P_1D(P_1, \omega_1) - wL^o(D(P_1, \omega_1), K, M) \right) \right]$$ (A6)

where the “X” subscript indicates these profits contain no positive waste and are in the excess profit zone. Comparing these two expected profit streams, we see that it is optimal for the Firm to engage in pure waste, as embodied in (A3), precisely when $\Pi_w(\omega_i) \geq \Pi_x(\omega_i)$, which, with a little algebra, is seen to be the complement of the asserted condition (11) for no positive waste. The Proposition therefore follows.

References


Crew, Michael A. and Paul R. Kleindorfer.1996. “Incentive Regulation in the United Kingdom and the


