Risk, Price Regulation, and Irreversible Investment

Lewis T. Evans

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

Graeme A. Guthrie

Victoria University of Wellington, Wellington, New Zealand

December 22, 2003
Risk, Price Regulation, and Irreversible Investment

Lewis T. Evans∗
New Zealand Institute for the Study of Competition and Regulation,
and Victoria University of Wellington, Wellington, New Zealand

Graeme A. Guthrie
Victoria University of Wellington, Wellington, New Zealand

December 22, 2003

Abstract

We show that regulators’ price-setting, rate base, and allowed rate of return decisions are inextricably linked. Once regulators switch from traditional rate of return regulation, the irreversibility of much infrastructure investment significantly alters the results of the usual approach to price-setting, as exemplified by Marshall, Yawitz and Greenberg (1981). In particular, the practice of ‘optimizing’ inefficient assets out of the regulated firm’s rate base, as in total element long-run incremental cost (TELRIC) calculations in telecommunications, exposes the firm to demand risk. The firm requires an economically-significant premium for bearing this risk, and this premium is an increasing function of the unsystematic risk of demand shocks. In addition, we argue that if the firm is to break even under incentive regulation then the level of the rate base will not generally equal the optimized replacement cost.

JEL Classification code: G31, L5

Keywords: Regulation, Cost of capital, Rate base, Sunk costs

1 Introduction

A regulator needs to make three decisions when setting the prices which a regulated utility may charge. It needs to choose the appropriate cost of the firm’s assets (the rate base), the rate of return the firm is allowed to earn on this rate base, and the prices the firm is allowed to charge. Marshall, Yawitz and Greenberg (1981) show how the last two decisions are inter-related. However, because they focus on traditional rate of return regulation, they do not discuss the effect of the choice of rate base. Nor do they consider the implications of irreversible investment, which characterizes most industries subject to price regulation.1 In this paper we show that the choice of rate base can have a crucial impact on the other two decisions, and that the reason for this is the irreversibility of investment. In particular, we demonstrate that the regulator’s choice of rate base and the form of regulation have a profound effect on the risks which the regulated firm faces, and thus on the rate of return it should be allowed to earn.

∗Corresponding author. Address: ISCR, PO Box 600, Victoria University of Wellington, Wellington, New Zealand. Ph: 64-4-4635562. Fax: 64-4-4635566. Email: lew.evans@vuw.ac.nz

1Irreversibility is a widespread phenomenon, even in industries where physical capital is not especially industry-specific. For example, between 50 and 80 percent of the cost of machine tools in Sweden is sunk (Asplund, 2000), and the market value of physical capital in the U.S. aerospace industry is just 28 percent of its replacement cost on average (Ramey and Shapiro, 2001). Irreversibility is likely to be even greater in most infrastructure networks. Hausman (1999) and Economides (1999) debate the extent of irreversibility in the context of telecommunications.
There are two widely-applied rate bases. Traditional rate of return regulation uses the depreciated historical installation cost of existing assets as the rate base. When combined with forecast operating expenditure, this yields the revenue requirement that, along with forecast demand, is used to set prices.\(^2\) The historical cost rate base continues to be used in some situations: for example, for elements of the electricity transmission system in the U.S. Historical cost rate of return regulation was widely used until the 1980s when it was gradually replaced with incentive regulation, where prices are set in ways that seek to mimic competitive markets. There are two common approaches. In the first, RPI-X regulation, a price path is specified by allowing a starting price to grow at the rate of inflation, adjusted for industry-specific factors such as relative productivity growth and input price changes. We consider the second approach to incentive regulation, in which prices are set using a rate base that is calculated at either the optimized replacement cost (ORC) of the assets, or the optimized deprival value (ODV), a closely related concept.\(^3\) Under the ORC rate base, the firm’s prices are set periodically using the least cost bundle of assets required to service existing customers. The ODV of an individual asset is typically calculated as the minimum of its replacement cost and the present value of the maximum net revenue which the asset can generate.\(^4\) Newbery (1999, Chapter 7) argues that in telecommunications, where ORC is the total element long-run incremental cost (TELRIC) of a service, use of an ORC rate base will yield price paths that approximate, as well as is possible, those of competitive markets. In the TELRIC calculation, costs are based on the elements of the system needed to provide the service, including the total attributable costs of that element calculated as the incremental cost required to produce an extra unit of that service over the long run (where all elements of the system can be varied). The TELRIC approach in telecommunications has been applied widely in the U.K. and the U.S., and is recommended by the European Commission (Newbery, 1999, p. 339).

The revenue which the regulated firm requires if it is to break even equals the sum of two components: expected economic depreciation and a reasonable rate of return earned on the rate base.\(^5\) We show that if the regulator imposes a historical cost rate base, then the only risk which the firm must bear is the risk that demand and operating cost experience shocks after the price-setting process is complete (and before prices are reset in the future). If, instead, the regulator adopts a replacement cost rate base, then the firm is also exposed to the risk of capital price shocks on its existing assets. The addition of the optimized rate base adds demand shocks at the time prices are reset to the list of risks faced by the regulated firm. Since fluctuations in demand affect the capacity of the (hypothetical) assets on which the optimization calculation is based, they affect the rate base. That is, when the regulator chooses ORC or ODV as the firm’s rate base, the firm is exposed to the risk of future capital price and demand shocks at the time prices are reset.

Despite the fact that our analysis is performed with the CAPM as our valuation model, the irreversible nature of the investment we consider means that unsystematic demand risk, as well as its systematic counterpart, affects the required rate of return when the rate base is subject to optimization. In fact, greater unsystematic risk compounds the impact of the systematic component of demand risk. For example, the firm’s ORC falls if demand falls, since falling demand means that even more units of capacity are under-utilized. In contrast, rising demand only raises replacement cost to the point where all existing assets are fully utilized; larger increases have no further impact. This asymmetry means that increased unsystematic demand risk, which (by definition) has no effect on the covariance of demand with market

\(^2\) Typically, approved investment plans also affect the revenue requirement under rate of return regulation. For a discussion of the process of rate of return regulation see Spulber (1989, pp. 270–279).

\(^3\) Our analysis of this form of incentive regulation has direct implications for settings under RPI-X regulation.

\(^4\) See Clarke (1998) for the application of the optimized deprival value concept in Australia and New Zealand.

\(^5\) We ignore any timing options which the regulated firm might have (in fact, we assume them away), as well as any capital market frictions which the firm might face. Both of these would be expected to further raise the allowed rate of return for the regulated firm.
returns, increases the covariance between the firm’s ORC and market returns. This, in turn, raises the systematic risk of the firm’s cash flows. Using simulations, we show that the effect on the firm’s allowed rate of return is economically significant for reasonable parameter values.

Drawing on the analysis of TELRIC of Mandy and Sharkey (2003), Littlechild (2003) argues that optimized replacement cost regulation shifts the risk of forecast errors to the regulated firm and raises its cost of capital because it is impossible to predict with accuracy the future path of cost, technology, and demand. He goes on to say that acceptance of this risk by the firm may improve the prospects of competition because prices are much more stable where the firm takes on the regulatory-price setting risk. Hausmann (1997) uses the option to invest to argue that the use of TELRIC to price elements of a network underprices the economic cost of the services provided and will adversely affect investment. Jorde, Sidak, and Teece (2000) argue that the common practice of using TELRIC in pricing elements of telecommunications networks that are unbundled by mandate raises the cost of equity of firms that own the networks, and consequently reduces investment in these networks. Ingraham and Sidak (2003) find that for their sample of firms the cost of equity is indeed raised by mandatory unbundling, particularly in recession periods. This is a prediction of our model where equipment prices are less sensitive than demand to recessions.

Rate of return regulation is most plausible where entry is prevented and technologies are changing slowly, whereas incentive regulation more readily allows price regulation to co-exist with some entry and decentralized decisions about investment. Recently, incentive regulation has been preferred over rate of return regulation because it de-links, to an extent that depends on the particulars of the regime, firm-specific costs and profits from prices and thereby provides incentives for firms to behave efficiently. The widespread adoption of incentive regulation renders it extremely important that its implementation does actually promote efficiency over other forms of regulation. The effect on investment is particularly important through its effect on dynamic efficiency, and the number of works that examine the implementation of incentive regulation in this context is growing. Our model addresses the dynamic efficiency over time issue discussed by Littlechild (2003, pp. 304–306). We show that if the firm is forced to supply and is expected to break-even on new investment, then the rate base should equal the ORC of fully utilized assets plus an amount representing the present value of the expected future cost savings resulting from the availability of any excess capacity at the time prices are (re)set.

In the next section we outline the various regulatory possibilities we consider. Section 3 develops the analysis of Marshall et al. (1981) for irreversibility and applies it to the different regulatory regimes. We examine what our results mean for regulatory policy in Section 4, where we present some numerical measures of the implications of our analysis. We conclude in Section 5.

2 The model

We consider a regulated firm which faces uncertain future demand and uncertain future capital prices. In year $t$ the capacity of the firm’s assets equals $S_t$. Each new unit of capacity built by the firm in year $t$ costs $P_t$. Investment in capacity is irreversible and the firm’s assets are infinitely-lived. The firm requires an asset with capacity $X_t$ to meet the needs of all its potential

---

6Mandy and Sharkey (2003) explain the effect on cost recovery of the regulated firm when prices are set on the basis of TELRIC at shorter intervals than asset lives in a world of certainty.

7The distinction between rate of return and incentive regulation and the critical issues relating to incentive regulation are pointed out by Newbery (1999, Chapter 2) and Laffont and Tirole (1993, Chapters 1 and 2).

8Baumol (2002) argues that investment yielding dynamic efficiency is the dominant factor affecting economic welfare.

9Evans and Guthrie (2003) show that irreversibility and the requirement to supply imply that incentive regulation is efficient relative to rate of return regulation.
customers. It has total revenue and operating cost in year \( t \) of \( R_t \) and \( C_t \) respectively.\(^{10}\)

The firm is regulated, and this regulation takes two forms. Firstly, the firm is required to meet the demand of all customers who wish to trade with the firm; that is, its capacity must satisfy \( S_t \geq X_t \) at all times. We call this a universal service obligation. Secondly, the regulator restricts the firm’s output price. The precise form of this restriction is not specified — for example, the regulator might set the firm’s output prices, or it might impose a cap which the firm cannot breach. The important thing is that the regulator is able to control the year-ahead expected revenue; that is, in year \( t \) the regulator controls \( E_t[R_{t+1}] \).\(^{11}\)

We assume that the firm has no flexibility in the timing of its investment — in particular, it does not invest in capacity which is not needed to meet demand. In practice, the uncertainty surrounding future capital prices means that there may be instances in which the firm would choose to build excess capacity, such as when capital prices are expected to rise in the future. However, we eliminate such flexibility for several reasons. First, this assumption means that excess capacity really is “excess” — that is, if a hypothetical firm invested in assets to replace the regulated firm, it would not build the excess capacity. Second, there has been vigorous debate about the effect of investment flexibility (and other real options) on the revenue which regulated firms should be allowed to collect, with some authors arguing that a ‘real option’ premium should be added to the weighted-average cost of capital when calculating such firms’ costs of capital. In this paper we show that some sort of premium is appropriate even when the regulated firm has no real options. Last, and by no means least, this assumption keeps the model tractable.

The combination of irreversibility, the universal service obligation imposed on the firm, and its lack of investment timing flexibility has important implications for investment behavior. If the firm has capacity \( S_t \geq X_t \) in year \( t \), it will only have to invest in year \( t+1 \) if \( X_{t+1} > S_t \); in which case investment expenditure equals \( I_{t+1} = P_{t+1}(X_{t+1} - S_t) \) and the assets’ capacity will rise to \( S_{t+1} = X_{t+1} \). If \( X_{t+1} \leq S_t \), then the firm’s investment is zero and the capacity of its assets remains at \( S_{t+1} = S_t \). Thus, the capacity of its assets in year \( t+i \) will equal

\[
S_{t+i} = \max\{S_t, X_{t+1}, X_{t+2}, \ldots, X_{t+i}\}
\]

and investment expenditure in that year will be

\[
I_{t+i} = P_{t+i} \max\{X_{t+i} - \max\{S_t, X_{t+1}, X_{t+2}, \ldots, X_{t+i-1}\}, 0\}.
\]

Regardless of the future path of demand, investment in each future year is a non-increasing function of current capacity.

Although a firm building a network from scratch in year \( t \) would not build any excess capacity, it does not follow that excess capacity held by an existing firm has no value. In fact, if the regulated firm has excess capacity in year \( t \), its future investment expenditure will sometimes be lower, and will never be greater, than that of a firm which starts business in year \( t \). The latter firm incurs investment expenditure of

\[
I_{t+i}^{(t)} = P_{t+i} \max\{X_{t+i} - \max\{X_t, X_{t+1}, \ldots, X_{t+i-1}\}, 0\}
\]

in year \( t+i \), so that the regulated firm’s excess capacity reduces its investment expenditure in year \( t+i \) by the amount

\[
I_{t+i} - I_{t+i}^{(t)} = P_{t+i} \max\{X_{t+i} - \max\{X_t, X_{t+1}, \ldots, X_{t+i-1}\}, 0\} - P_{t+i} \max\{X_{t+i} - \max\{S_t, X_{t+1}, \ldots, X_{t+i-1}\}, 0\}.
\]

\(^{10}\)We are keeping our model deliberately abstract in order to illustrate the importance of irreversibility, and not the particular industry, on regulation. In particular, the precise interpretation of capacity will vary with the industry considered. For a telecommunications firm, capacity might represent the number of connections to the network. For an electricity distribution network, capacity might reflect the peak load carried over the network.

\(^{11}\)We assume the regulator resets prices annually. Although the calculations would become more complicated, and the expressions for expected revenue and the allowed rates of return would change, we could relax this assumption without changing our main results.
This is zero if $X_{t+i} \leq \max\{X_t, X_{t+1}, \ldots, X_{t+i-1}\}$ and positive otherwise. The value of the regulated firm’s excess capacity is the market value of this stream of future cash flows. We will see that successful implementation of incentive regulation requires an understanding of the value excess capacity can generate.

The market value of the regulated firm equals the market value of the net cash flows received by the firm’s owners. These depend, in part, on the revenue which the regulator allows the firm to collect from its customers. Allowed revenue typically includes compensation for the firm’s investment in physical capital, and this generally takes the form of the product of the “value” of the firm’s assets and some regulated rate of return. Clearly this value cannot equal the market value of the regulated firm, as then a circularity results: the firm’s market value depends on its allowed revenue, which depends on its market value. What is needed is some exogenous measure of the value of the firm’s assets, known as the firm’s rate base. Several possibilities have been suggested.\footnote{Our assumption that the firm’s assets have infinite lives means that the issue of physical depreciation does not arise. The following definitions of rate bases need to be modified when assets have finite physical lives. To keep the analysis as simple as possible, we do not consider physical depreciation, focusing instead on the consequences of irreversibility and uncertainty.}

**Historical cost (HC).** From year $t - 1$ to year $t$, the firm invests $I_t$ in new assets. Thus, the historical cost of the firm’s assets evolves according to $HC_t = HC_{t-1} + I_t$.

**Replacement cost (RC).** In year $t$, if the firm was to replace its assets it would have to acquire $S_t$ units of capacity at a price of $P_t$. Thus, replacement cost in year $t$ is $RC_t = P_t S_t$.

**Optimized replacement cost (ORC).** If the firm was to replace its assets in year $t$ with an optimal configuration, it would have to acquire $\min\{S_t, X_t\}$ units of capacity at a price of $P_t$, since only $X_t$ units would be required if there was excess capacity, while all $S_t$ would have to be replaced if there was no excess capacity. Therefore, optimized replacement cost in year $t$ is $ORC_t = P_t \min\{S_t, X_t\}$.

**Optimized deprival value (ODV).** This asset value measures the reduction in the value of the firm if it was deprived of ownership of the assets.\footnote{Regulators in some jurisdictions use a valuation methodology which they term ‘optimized deprival value’. However, as we explain in Section 4.1, their approach has more in common with ORC than with our ODV concept.} The impact of such an event depends on whether or not the assets are currently being used to full capacity. If the firm is currently operating at full capacity, it would immediately rebuild $S_t$ units of capacity if it lost ownership of its assets, and future investment expenditures would be unaffected by the loss. Thus, losing ownership of the assets would cost the firm $ODV_t = P_t S_t$. On the other hand, if the firm currently has excess capacity, then it would immediately rebuild just $X_t$ units of capacity, costing $P_t X_t$, if it lost ownership of its assets. However, future investment expenditure would rise by $I_{t+i} - I_{t+i} \geq 0$ in year $t+i$ for all $i \geq 1$. Thus, losing ownership of the assets would cost the firm

$$ODV_t = P_t X_t + U_t^{(t)} - M_t,$$

where $U_t^{(t)}$ denotes the value in year $t$ of all future investment expenditure, from year $t+1$ onwards, of a network built in year $t$, and $M_t$ denotes the value in year $t$ of all of the firm’s future investment expenditure, from year $t+1$ onwards. This is the sum of the optimized replacement cost of the firm’s assets and the value of the firm’s excess capacity. Combining these two cases, we see that the optimized deprival value of the firm’s assets is

$$ODV_t = ORC_t + (\text{value of excess capacity at } t).$$
We now turn to the problem of determining a “reasonable” level of revenue which the regulated firm should be allowed to earn. We will see that the solution depends on the chosen rate base.

3 The approach of Marshall, et al.

In our model the regulator sets prices (or, at least, sets parameters which influence the prices which the regulated firm can set) in order that the regulated firm can achieve some desired level of revenue. This level of revenue must be sufficient to compensate the firm for the costs which it incurs, comprising operating costs and the cost of capital. The regulator determines the firm’s cost of capital by specifying the rate base (essentially the ‘cost’ of the firm’s assets) as well as the rate of return which the firm is allowed to earn on this rate base. Thus, the regulator must make three decisions: (1) What is the appropriate cost of the firm’s assets? (2) What rate of return is the firm allowed to earn on this rate base? (3) What set of prices is the firm allowed to charge?

Marshall, et al. (1981) point out that the prices set by a regulator affect the risk borne by the firm, and therefore the rate of return which the firm should be allowed to earn on its rate base. In particular, the three questions above cannot be considered separately. Marshall, et al. use the CAPM to integrate the rate of return determination and price setting decisions (but do not consider the choice of rate base). They argue that the regulator should set output prices in such a way that the market value of the firm equals the cost of the firm’s assets. In this section we describe how Marshall, et al.’s approach would be implemented for the types of firms we consider in this paper, firstly for an arbitrary choice of rate base, and then for the four rate bases described in Section 2.

If the market value of the firm is $F_{t+1}$ at the end of year $t+1$, then the value of investors’ stake in the firm at the start of year $t+1$ is

$$V_{t+1} = R_{t+1} - C_{t+1} - I_{t+1} + F_{t+1}.$$  

From the certainty equivalent form of the CAPM, the value of their stake in year $t$ is

$$F_t = \frac{E_t[V_{t+1}] - \lambda_t \text{Cov}_t[V_{t+1}, r_{m,t}]}{1 + r_{f,t}},$$

where $\lambda_t = (E_t[r_{m,t}] - r_{f,t})/\text{Var}_t[r_{m,t}]$ is the market price of risk, $r_{f,t}$ is the risk-free interest rate over the year $[t, t + 1]$, and $r_{m,t}$ is the (risky) rate of return on the market portfolio over the same period. The regulator’s objective is to set prices such that the market value of the firm is always equal to some exogenous rate base. For the rate base $B_t$, this can be achieved provided that

$$B_t = \frac{E_t[R_{t+1} - C_{t+1} - I_{t+1} + B_{t+1}] - \lambda_t \text{Cov}_t[R_{t+1} - C_{t+1} - I_{t+1} + B_{t+1}, r_{m,t}]}{1 + r_{f,t}}.$$  

(1)

Solving for $E_t[R_{t+1}]$ shows that expected revenue must equal

$$E_t[R_{t+1}] = (1 + r_{f,t}) B_t + \lambda_t \text{Cov}_t[R_{t+1} - C_{t+1} - I_{t+1} + B_{t+1}, r_{m,t}] + E_t[C_{t+1}] + E_t[I_{t+1}] - E_t[B_{t+1}].$$  

(2)

By choosing the prices which the firm is allowed to charge, the regulator effectively chooses the distribution from which revenue $R_{t+1}$ will be drawn. We will not consider the price-setting decision explicitly, but rather assume that the resulting distribution of revenue is consistent with equation (2).

---

$^{14}$The appearance of investment expenditure in equation (1) ensures that the firm will always be able to attract capital when investment is required.
It is convenient to rewrite (2) in the form

\[ E_t[R_{t+1}] = E_t[C_{t+1}] + r_tB_t + (B_t + E_t[I_{t+1}] - E_t[B_{t+1}]) \]  

(3)

and

\[ r_t = r_{f,t} + \lambda_t \text{Cov}_t \left[ \frac{R_{t+1} - C_{t+1} - I_{t+1} + B_{t+1}}{B_t}, r_{m,t} \right]. \]  

(4)

Equation (4) gives the rate of return, \( r_t \), which the firm is allowed to earn on its rate base. Marshall, et al.’s observation, that the regulator’s choice of prices affects \( r_t \), is illustrated by the appearance of \( R_{t+1} \) in the right hand side of (4). The expected revenue given by equation (3) can be naturally decomposed into three terms:

**Operating cost.** At the end of the year the firm expects to incur operating costs of \( E_t[C_{t+1}] \).

**Return on capital.** The second term, \( r_tB_t \), equals the amount which the firm’s investors are allowed to earn on the rate base. From (4), the allowed rate of return is the sum of the risk free interest rate and a risk premium which compensates investors for the systematic risk of shocks to the future value of their stake in the firm.

**Return of capital.** In year \( t \), the value of the firm equals \( B_t \). After one year, this changes to \( B_{t+1} \). However, in the meantime the firm’s investors have invested a further \( I_{t+1} \). Provided investment raises the rate base dollar-for-dollar, then the market value of their year \( t \) investment equals \( B'_{t+1} = B_{t+1} - I_{t+1} \) in year \( t+1 \). Thus the term \( E_t[B_t - B'_{t+1}] \) can be interpreted as the expected reduction in value of the investors’ asset.

Especially in the case of infrastructure assets, the level of the rate base and the allowed rate of return on the rate base are crucial determinants of the regulated firm’s allowed revenue.

We now consider four possible implementations of this approach which differ in their choice of rate base.

### 3.1 Historical cost

Under traditional rate of return regulation, the rate base rises each year by the amount of investment made by the regulated firm. That is, if the rate base in year \( t \) equals \( B_t \), then the rate base in year \( t+1 \) equals \( B_{t+1} = B_t + I_{t+1} \), the historical cost of the firm’s assets. From equation (3), the firm’s expected revenue equals

\[ E_t[R_{t+1}] = E_t[C_{t+1}] + r_tB_t. \]

The firm does not receive any compensation for expected economic depreciation because no such depreciation can occur. From equation (4), the appropriate risk-adjusted discount rate is

\[ r_t = r_{f,t} + \lambda_t \text{Cov}_t \left[ \frac{R_{t+1} - C_{t+1}}{B_t}, r_{m,t} \right], \]

since the only risk the firm is exposed to is that shocks to revenue and operating costs might occur after the regulator sets prices (and before it resets them in year \( t+1 \)).

### 3.2 Replacement cost

When the replacement cost of the firm’s assets is chosen as the rate base, substituting \( B_t = P_tS_t \), \( B_{t+1} = P_{t+1}S_{t+1} \), and \( I_{t+1} = P_{t+1}(S_{t+1} - S_t) \) into equations (3) and (4) shows that the firm’s expected revenue must equal

\[ E_t[R_{t+1}] = E_t[C_{t+1}] + r_tB_t + (P_t - E_t[P_{t+1}])S_t \]  

(5)
and the allowed rate of return must be

\[ r_t = r_{f,t} + \lambda_t \text{Cov}_t \left[ \frac{R_{t+1} - C_{t+1}}{B_t} + \frac{P_{t+1}}{P_t}, r_{m,t} \right]. \]  

(6)

Compared to the case of an historical cost rate base, additional terms, reflecting the possibility of fluctuations in capital prices, appear in both the allowed revenue and the allowed rate of return. The new component of the firm’s allowed revenue, \((P_t - E_t[P_{t+1}]) S_t\), adjusts for expected economic depreciation — that is, the firm is allowed to collect just enough additional revenue to compensate it for any expected decline in the level of its rate base. The new component in the firm’s allowed rate of return,

\[ \lambda_t \text{Cov}_t \left[ \frac{P_{t+1}}{P_t}, r_{m,t} \right], \]

compensates the firm for the systematic risk of shocks to capital prices (and hence to the firm’s rate base).

### 3.3 Optimized replacement cost

The regulated firm is exposed to additional risk when the optimized replacement cost of its assets is chosen as the rate base. To see why, note that after any required investment has occurred in year \( t \) the capacity of the firm’s assets satisfies \( S_t \geq X_t \), so that the ORC of its assets is \( B_t = P_t X_t \). If demand in year \( t + 1 \) satisfies \( X_{t+1} \leq S_t \), then no investment is necessary; \( I_{t+1} = 0 \), \( S_{t+1} = S_t \), and \( B_{t+1} = P_{t+1} X_{t+1} \). On the other hand, if demand satisfies \( X_{t+1} > S_t \), then the firm will have to expand capacity; investment expenditure will equal \( I_{t+1} = P_{t+1} (X_{t+1} - S_t) \), capacity will rise to \( S_{t+1} = X_{t+1} \), and the closing rate base will equal \( B_{t+1} = P_{t+1} X_{t+1} \). Combining these two cases, we have that

\[ B_{t+1} - I_{t+1} = P_{t+1} \text{min}\{S_t, X_{t+1}\}. \]

Substituting this value into equations (3) and (4) shows that the firm’s expected revenue must equal

\[ E_t[R_{t+1}] = E_t[C_{t+1}] + r_t B_t + (P_t X_t - E_t[P_{t+1} \text{min}\{S_t, X_{t+1}\}]) \]

and its allowed rate of return is

\[ r_t = r_{f,t} + \lambda_t \text{Cov}_t \left[ \frac{R_{t+1} - C_{t+1}}{B_t} + \frac{P_{t+1}}{P_t} \text{min}\left\{ \frac{S_t}{X_t}, \frac{X_{t+1}}{X_t} \right\}, r_{m,t} \right]. \]

Compared to the case with simply a replacement cost rate base, now the distribution of future demand \( X_{t+1} \) affects both expected revenue and the allowed rate of return.

As was the case with a replacement cost rate base, the firm is exposed to the risk of capital price shocks, but now it also faces demand risk: holding the capital price constant, if demand rises then the value of the rate base increases, while if demand falls the value of the rate base falls. However, the risk is asymmetric, since increases in demand beyond the assets’ current capacity have no additional impact, while there is unlimited downside risk from negative demand shocks. We will see in Section 4 that this asymmetry means that the total risk, and not just its systematic component, affects both expected economic depreciation and the allowed rate of return.

The value of the firm in year \( t \), \( P_t X_t \), does not depend on the capacity of its assets. But we know that having excess capacity reduces the future investment expenditure of the firm and,

---

\(^{15}\) Notice that when investment is reversible, we obtain the same expressions for expected revenue and allowed rate of return as in Section 3.2. The reason is that with reversible investment the firm will never carry excess capacity, and \( S_t = X_t \) at all times. If demand rises from year \( t \) to year \( t + 1 \), the firm will have to invest \( P_{t+1} (X_{t+1} - X_t) \) in new capacity; if demand falls from year \( t \) to year \( t + 1 \), the firm will raise \( P_{t+1} (X_t - X_{t+1}) \) from selling excess capacity. Overall, investment is \( I_{t+1} = P_{t+1} (X_{t+1} - X_t) \). Substituting these values into equations (3) and (4), together with the new rate base \( B_t = P_t X_t \), results in equations (5) and (6).
holding all else equal, raises the value of the firm. Here, however, all else is not held equal — the greater the firm’s excess capacity, the smaller its allowed revenue. In effect, the firm is punished for having excess capacity. This is inconsistent with modern incentive regulation, which requires that the firm’s allowed revenues depend only on the cost structure of an efficient alternative provider. Thus incentive regulation is unable to drive the market value of the regulated firm to the replacement cost of an efficiently-configured firm. We close this section by discussing one choice of rate base which is consistent with incentive regulation.

3.4 Optimized deprival value

In the fourth and final possibility we consider, the regulated firm receives the level of revenue that would be required for a hypothetical firm (which can invest in assets configured to meet current and future demand) to replace the incumbent and just break even. If such a replacement firm invests in year \( t \) and undertakes all future investment required to meet the universal service obligation, then its net cash flow is \(-P_t X_t\) in year \( t \), and \( R_{t+i} - C_{t+i} - \hat{I}_{t+i}^{(t)}\) in year \( t+i \) for all \( i \geq 1 \). The regulator sets prices such that the market value of all cash flows from year \( t+1 \) onwards equals \( P_t X_t \), so that the hypothetical efficient replacement firm just breaks even. Relative to this firm, the regulated firm receives cash flows of \( \hat{I}_{t+i}^{(t)} - I_{t+i} \) in year \( t+i \) for all \( i \geq 1 \). Thus, the market value of the regulated firm exceeds the market value of the stream of incremental cash flows. It follows that the market value of the regulated firm equals the sum of \( P_t X_t \), the ORC of its assets, and the value of its excess capacity; that is, the market value of the regulated firm equals the ODV of its assets.

Substituting the firm’s ODV as the rate base in equations (3) and (4) shows that expected revenue equals

\[
E_t[R_{t+1}] = E_t[C_{t+1}] + r_t ODV_t + (ODV_t + E_t[I_{t+1}] - E_t[ODV_{t+1}])),
\]

where the rate of return allowed on the ODV rate base equals

\[
r_t = r_{f,t} + \lambda_t \text{Cov}_t \left[ \frac{R_{t+1} - C_{t+1} - I_{t+1} + ODV_{t+1}}{ODV_t}, r_{m,t} \right].
\]

Since the firm’s ODV depends on the actual capacity of its assets, and ODV appears in both equations above, it might seem that the firm’s allowed revenue depends on this capacity. However, this is not the case — the firm’s expected revenue is independent of \( S_t \).

**Proof.** Recall that \( M_t \) denotes the value in year \( t \) of all of the firm’s future investment expenditure, from year \( t+1 \) onwards. Therefore \( M_t \) satisfies

\[
M_t = \frac{E_t[I_{t+1} + M_{t+1}]}{1 + r_{f,t}}.
\]

Recall also that \( U_i^{(t)} \) denotes the value in year \( t \) of all future investment expenditure, from year \( t+1 \) onwards, of a replacement firm that started business in year \( t \), so that

\[
ODV_t = P_t X_t + U_i^{(t)} - M_t.
\]

Substituting this expression for the ODV asset base into equation (1) gives

\[
P_t X_t + U_i^{(t)} - M_t = \frac{E_t[R_{t+1} - C_{t+1} - I_{t+1} + P_t X_{t+1} + U_i^{(t+1)} - M_{t+1}]}{1 + r_{f,t}} - \lambda_t \text{Cov}_t \left[ R_{t+1} - C_{t+1} - I_{t+1} + P_t X_{t+1} + U_i^{(t+1)} - M_{t+1}, r_{m,t} \right].
\]

Using equation (7) to eliminate \( M_t \) allows us to simplify this to

\[
P_t X_t + U_i^{(t)} = \frac{E_t[R_{t+1} - C_{t+1} + P_t X_{t+1} + U_i^{(t+1)}]}{1 + r_{f,t}} - \lambda_t \text{Cov}_t \left[ R_{t+1} - C_{t+1} + P_t X_{t+1} + U_i^{(t+1)}, r_{m,t} \right].
\]
This equation does not depend on the capacity \( (S_t) \) of the firm’s assets, but rather on current \( (X_t) \) and future \( (X_{t+1}) \) demand. Thus, when ODV is used as the rate base, Marshall, et al.’s approach results in a form of incentive regulation.

If the firm’s allowed revenue does not depend on the capacity of its assets, then what does it depend on? We find that the regulated firm’s allowed revenue is

\[
E_t[R_{t+1}] = E_t[C_{t+1}] + r_t P_t X_t + \left( P_t X_t - E_t[P_t \min\{X_t, X_{t+1}\} + U_{t+1}^{(t+1)} - U_{t+1}^{(t)}]\right),
\]

where the allowed rate of return \( r_t \) equals

\[
r_t = r_{f,t} + \lambda_t \text{Cov}_t \left\{ \frac{R_{t+1} - C_{t+1}}{P_t X_t} + \frac{P_{t+1}}{P_t} \min\left\{ 1, \frac{X_{t+1}}{X_t}\right\} + \frac{U_{t+1}^{(t+1)} - U_{t+1}^{(t)}}{P_t X_t}, r_{m,t} \right\}. \tag{10}
\]

**Proof.** In keeping with the definition of \( U_t^{(i)} \) in Section 2, we denote by \( U_t^{(i)} \), the market value, measured in year \( t+i \), of all investment expenditure from year \( t+i+1 \) onwards of an efficient hypothetical firm which replaced the regulated firm in year \( t \). Therefore

\[
U_t^{(i)} = \frac{E_t[\hat{P}_{t+i}^{(t)} + U_{t+i}^{(i)}] - \lambda_t \text{Cov}_t [\hat{P}_{t+i}^{(t)} + U_{t+i}^{(i)}, r_{m,t}]}{1 + r_{f,t}}.
\]

We use this to eliminate \( U_t^{(i)} \) from equation (8), obtaining

\[
P_t X_t = \frac{-U_t^{(t)} + E_t[R_{t+i} - C_{t+i} + P_{t+i} X_{t+i} + U_{t+i}^{(t+1)}]}{1 + r_{f,t}} - \frac{\lambda_t \text{Cov}_t [R_{t+i} - C_{t+i} + P_{t+i} X_{t+i} + U_{t+i}^{(t+1)}, r_{m,t}]}{1 + r_{f,t}}
\]

\[
= \frac{E_t[R_{t+i} - C_{t+i} + P_{t+i} X_{t+i} - \hat{j}_{t+i}^{(t+1)} - U_{t+i}^{(t+1)}]}{1 + r_{f,t}} - \frac{\lambda_t \text{Cov}_t [R_{t+i} - C_{t+i} + P_{t+i} X_{t+i} - \hat{j}_{t+i}^{(t+1)} - U_{t+i}^{(t+1)}, r_{m,t}]}{1 + r_{f,t}}
\]

\[
= \frac{E_t[R_{t+i} - C_{t+i} + P_{t+i} \min\{X_t, X_{t+i}\} + \hat{U}_{t+i}^{(t+1)} - \hat{U}_{t+i}^{(t)}]}{1 + r_{f,t}} - \frac{\lambda_t \text{Cov}_t [R_{t+i} - C_{t+i} + P_{t+i} \min\{X_t, X_{t+i}\} + \hat{U}_{t+i}^{(t+1)} - \hat{U}_{t+i}^{(t)}, r_{m,t}]}{1 + r_{f,t}},
\]

where we have used the fact that \( P_{t+i} X_{t+i} - \hat{j}_{t+i}^{(t+1)} = P_{t+i} \min\{X_t, X_{t+i}\} \) in the final step. Solving this equation for \( E_t[R_{t+1}] \) results in equations (9) and (10). Together they describe the expected revenue which the regulated firm should be allowed to earn if its value is always to equal the ODV of its assets.

The term in large brackets in equation (9) can be interpreted as the expected economic depreciation for the hypothetical replacement firm: The first term, \( P_t X_t \), is the cost of building the efficiently-configured asset in year \( t \); the second term, \( E_t[P_t \min\{X_t, X_{t+1}\}] \), is its expected optimized replacement cost in the following year; the third term, \( E_t[U_{t+1}^{(t+1)} - U_{t+1}^{(t)}] \), is the expected value of its excess capacity in year \( t+1 \). Thus, we can interpret the term in large brackets as the expected change in the ODV of the hypothetical replacement firm over the first year of its life.

This suggests one possible interpretation of the regulation which results when ODV is adopted as the firm’s rate base: the regulated firm is allowed to collect the same expected revenue as a hypothetical replacement firm which is allowed to keep the value of any capacity which is not utilized after the first year of operation.

On a more practical note, we see that incentive regulation can be achieved using ORC as the rate base, but not if we use a naive implementation of Marshall, et al.’s approach. Instead, allowance must be made for the possibility that some current, efficiently-configured, capacity
becomes temporarily under-utilized in the future. The value of this future excess capacity affects both expected economic depreciation and the allowed rate of return. This is important, since, as we discuss in the following section, regulators do not seem to include the value of excess capacity in ODV calculations.

4 Policy implications

4.1 Calculating ODV

Regulators in various jurisdictions have considered a rate base which they term optimized depreciable value.\(^\text{16}\) However, their rate base has more in common with our ORC measure than ODV. In this section we discuss the differences between ODV as it is used in practice and ODV as we advocate it in this paper.

Current regulatory practice is to set the ODV of an asset equal to its replacement cost, when the firm would replace the asset if it was deprived of the asset’s use, and the so-called “economic value” of the asset otherwise. Economic value is defined to be the present value of the profit-maximizing revenue which could be generated from the asset. Since the asset would not be replaced if this present value is less than the asset’s replacement cost, this rule leads to the following expression for the ODV of an individual asset:

\[
ODV' = \min\{RC_t, \text{economic value at } t\}.
\]

In our model, assets are either fully utilized or are not utilized at all, so that the economic value term in the expression above is zero. Therefore, if \(X_t \leq S_t\), then only \(X_t\) units of capacity would be replaced (costing \(P_t X_t\)), and the remaining \(S_t - X_t\) units of capacity have zero economic value, implying an ODV of \(P_t X_t\) for the firm as a whole. In contrast, if \(X_t > S_t\), then all \(S_t\) units of capacity would be replaced, costing \(P_t S_t\), which is the implied ODV for the firm. That is, if the regulators’ ODV calculation is applied to the firm in our model, the result would be an ODV of \(P_t \min\{X_t, S_t\}\), which is just the ORC of the firm’s assets.

The key omission from regulators’ ODV calculations is any allowance for the value of excess capacity. Our ODV equals the sum of the replacement cost of those assets which are currently in use and the value of the excess capacity provided by those assets which are not in use. To highlight the differences from current practice, we note that our expression for ODV equals the replacement cost of an asset, if the asset is currently in use, and the value of the excess capacity it provides otherwise. That is, the ODV of an individual asset is

\[
ODV'' = \min\{RC_t, \text{value of excess capacity at } t\}.
\]

Thus, the key difference is the approach to calculating economic value. Whereas current practice is to use the present value of profit-maximizing revenue which the asset can generate, our approach effectively adds in the present value of the investment expenditures which will be avoided in the future because of the firm’s ownership of the asset. In other words, current practice does not measure the true effect on the value of a firm if it was deprived of the ownership of its assets.

4.2 The role of demand risk

Sections 3.3 and 3.4 show that as soon as the firm’s asset base is subject to optimization by the regulator, the firm is exposed to the risk of demand shocks.\(^\text{17}\) In this section we discuss the

\(^{16}\)Most notably, electricity transmission in New Zealand uses such a rate base (Ministry of Economic Development, 2000). Although considered favorably by regulators in Australia, ODV was rejected in favor of an ORC regime (Clarke, 1998; Johnstone, 2003, p. 3).

\(^{17}\)This would not be the case if investment was reversible, for then the firm can simply sell (or redeploy) any excess capacity following a negative demand shock — the firm is exposed to the risk of fluctuations in the capital price, but not in the quantity of capital required.
Table 1: The effect of demand risk on the allowed rate of return

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\sigma$</th>
<th>0.0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td></td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td>0.0055</td>
<td>0.0079</td>
<td>0.0089</td>
<td>0.0093</td>
</tr>
<tr>
<td>0.50</td>
<td></td>
<td>0.0110</td>
<td>0.0134</td>
<td>0.0158</td>
<td>0.0170</td>
</tr>
<tr>
<td>0.75</td>
<td></td>
<td>0.0165</td>
<td>0.0186</td>
<td>0.0216</td>
<td>0.0237</td>
</tr>
</tbody>
</table>

Notes. The entries in the table report the risk premium resulting from the risk of demand shocks. The return on the market portfolio is normally distributed with mean $\bar{r}_m = 0.12$ and standard deviation $\sigma_m = 0.2$, and the growth in demand equals $X_{t+1}/X_t - 1 = \beta r_{m,t} + \epsilon_t$, where the noise term is normally distributed with mean zero and standard deviation $\sigma_\epsilon$. Demand shocks thus have systematic risk of $\beta$. The riskless interest rate is $r_f = 0.04$. The covariances required to calculate the risk premia are each estimated using simulations with 100,000 draws.

Implications of this result for the calculation of the firm’s allowed rate of return.

Because the precise level of the firm’s ODV depends on the distribution of all future levels of demand, we are prevented from deriving simple expressions involving $(X_t, X_{t+1})$ for the firm’s expected revenue and allowed rate of return under ODV-regulation. However, such simple functional forms are possible in a special case of our model in which required capacity may change from year $t$ to year $t + 1$ but remains constant at its new level after year $t + 1$. This means that no investment is required from year $t + 2$ onwards, implying that $U_{t+1}^{(t+1)} = U_{t+1}^{(t)}$. Substituting this into equation (10) shows that ODV-regulation can result if the regulated firm is allowed to earn the rate of return

$$r_t = r_{f,t} + \lambda_t \text{Cov}_t \left[ \frac{R_{t+1} - C_{t+1}}{P_t X_t} + \frac{P_{t+1}}{P_t} \min \left\{ 1, \frac{X_{t+1}}{X_t} \right\}, r_{m,t} \right]$$

on the ORC of its assets.

We use numerical simulations to show the significance of demand risk (both systematic and unsystematic) on the allowed rate of return, simplifying our analysis by assuming that the capital price is constant. For a range of parameter values, we estimate the component of the risk premium due to demand shocks. That is, we estimate

$$\lambda_t \text{Cov}_t \left[ \min \left\{ 1, \frac{X_{t+1}}{X_t} \right\}, r_{m,t} \right]. \quad (11)$$

The return on the market portfolio is normally distributed with mean $\bar{r}_m = 0.12$ and standard deviation $\sigma_m = 0.2$, and the growth in demand equals

$$\frac{X_{t+1}}{X_t} - 1 = \beta r_{m,t} + \epsilon_t, \quad (12)$$

where the noise term $\epsilon_t$ is normally distributed with mean zero and standard deviation $\sigma_\epsilon$. Demand shocks thus have systematic risk of $\beta$. The riskless interest rate is $r_f = 0.04$. The covariances required to calculate the risk premia are each estimated using simulations with 100,000 draws, and the results of this exercise are reported in Table 1. For example, if demand shocks have systematic risk of $\beta = 0.5$ and the unsystematic component of demand shocks has a standard deviation of $\sigma_\epsilon = 0.1$, the appropriate risk premium is 1.34 percent.

The first point to note is that for reasonable parameter values the risk premium for demand risk can be economically significant. Secondly, the risk premium is an increasing function of systematic demand risk. Thirdly, if demand has positive systematic risk, then any additional unsystematic demand risk adds to the risk premium because of the nonlinearity introduced by
the irreversibility of investment and the requirement to supply. To illustrate this, consider the demand growth equation (12). It includes both (i) the elements of demand correlated with $r_{m,t}$ (that is, $E[X_{t+1}/X_t - 1|r_{m,t}] = \beta r_{m,t}$) and (ii) the uncorrelated elements represented by $\epsilon_t$. The otherwise unsystematic elements, $\epsilon_t$, enter systematic risk under regulation because they affect the truncation point thus, using equation (10),

$$\text{Cov}_t \left[ \min \left\{ 1, \frac{X_{t+1}}{X_t} \right\}, r_{m,t} \right] = \text{Cov}_t [\beta r_{m,t} + \epsilon_t, r_{m,t} | \beta r_{m,t} + \epsilon_t < 0].$$

This covariance is clearly a function of the unsystematic risk component of demand, $\epsilon_t$, providing $\beta \neq 0$.

In short, when installed assets are sunk and the firm is required to supply, regulatory regimes that use ORC or ODV to set prices impart additional volatility to the firm’s cash flow. This raises the revenue that the firm must be allowed to earn if it is to break even on new investment. The higher revenue is to cover the expected cost of assets under-utilized in the future, as well as the higher returns needed to compensate for the increased risk due to capital price and demand uncertainty. In addition, the truncation of revenues implied by irreversibility means that both systematic and unsystematic demand risk affect the reasonable allowable rate of return.

5 Concluding remarks

Our work suggests that under incentive regulation the rate base should not be optimized replacement cost, but rather optimized deprival value calculated using the present value of expected cost savings as the value placed upon excess capacity. Further, if firms are to break even on new investments then the allowed rate of return should be sufficient to compensate for the extra risks implied by incentive regulation. Also, where regulation requires firms to undertake irreversible investments, systematic risk may be increased by the introduction of specific unsystematic risks because of the truncation resulting from irreversibility and the requirement to supply. Not using this rate base specification or allowing for these risks will incorrectly compensate the firm for its investment and thereby adversely affect investment and dynamic efficiency.

There are two sets of measurement issues suggested by our analysis. The first is that it is not sufficient to draw comparator firms from the same industry where benchmark comparisons are made in the process of setting regulatory parameters — they must also be subject to the same regulatory environment. For example, European and U.S. firms’ rates of return under regulation, even if drawn from the same industry, will likely be systematically different as a result of different regulatory regimes. Our results suggest that significant differences may occur solely from variations in regulation. The second measurement issue relates to assessing the profitability of regulated industries. Those industries that are subject to rate of return regulation can be expected to have lower ex ante and ex post rates of return than those subject to incentive regulation. Under incentive regulation, the risks of stranding and capital price fluctuations are borne by the firm rather than consumers; the reverse is true for rate of return regulation. This difference between the regimes is exacerbated where regulation seeks to enhance competition and thereby introduces further risk to the incumbent’s business.

We have not evaluated RPI-X regulation, but we note that the issues dealt with in this paper are directly relevant to the choice of X and the starting price, and to evaluating violations of such a regime. The factors of irreversibility, growth in demand, supply and capital prices that are important to these investigations will enter the analysis as described in this paper in (9) and (10), for example, if the regulated firm is to just cover the costs of its investments.

---

18 We thus disagree with those who would set prices based on TELRIC calculations at intervals on the grounds of dynamic efficiency.

19 Fisher and McGowan (1983) point out the significant difficulties of any profitability assessment.
Although it is a policy issue for the firm and regulators, and a topic deserving further research, we have not investigated the decision to invest in advance of demand. Where there are economies to be achieved by investing in asset configurations for which capacity is not expected to be attained until some future date, such investment may be economically desirable. We conjecture that use of the ODV rate base would allow the firm to just recover its investment if it could include assets in our concept of ODV no matter whether the assets had been used then stranded or were unused and installed in anticipation of demand. Assets installed in advance would have valuation equal to the potential cost savings that they implied, leaving the firm the cost saving as profit inducement for cost-efficient investment prior to demand; assets for which there is no prospective use would have no future cost savings, and hence a zero ODV. The systematic allowance of future cost savings associated with the calculated ODV is critical to specifying the rate base that allows the expected recovery of investment, and it has a significant speculative component that the regulator will generally want to assess. In this respect, and in the verification of ODV more generally, these sorts of incentive regimes may require detailed monitoring and regulator-decision-making that approach those of rate of return regulation.

References


