The Benefits and Costs of Rate of Return Regulation

By JEFFREY CALLEN, G. FRANK MATHEWSON, AND HERBERT MOHRING*

The seminal article "Behavior of the Firm Under Regulation" by Harvey Averch and Leland Johnson has evoked extensive comment during the years since it appeared. Only in recent articles by Eytan Sheshinski and Alvin Klevorick, however, has the Averch-Johnson analytical framework been employed to deal with the question: If, for whatever reason, a rate of return constraint is the device that must be employed to control a monopolist's profits, at what level should the allowable rate of return be set? The answer to this question and pseudoempirical analysis of its implications are the subjects of this paper.

Specifically, Section I provides a simpler development of Sheshinski's and Klevorick's main formal conclusions than was contained in their papers.² Section II analyzes, as did Klevorick, the special case of constant elasticity demand and Cobb-Douglas production functions in a world in which income distribution considerations can be ignored and in which the regulatory process involves no resource costs. Expressions are derived and tabulated for the optimum allowable rate of return and for the effects on "social welfare" (i.e., the sum of consumers' and producers' surpluses), output, costs, and capital intensity of imposing a rate of return constraint on a formerly unregulated monopoly.

If this simple example sustains generalization, an allowable rate of return in excess of a monopoly's cost of capital would be desirable only if some combination prevails of a low demand elasticity, modest scale economies, and a small exponent on capital in the production function.² Furthermore, the gains from rate of return regulation (gross of the costs of the regulatory process) are substantial even if the allowed return differs considerably from its optimum value. Finally, at least for combinations of parameter values which reflect capital intensive production processes with substantial scale economies, rate of return regulation could capture a substantial share of the benefits derivable from setting price equal to average or (except when the elasticity of demand is large) even marginal cost.

I. The Optimum Allowable Rate of Return

As preparation for dealing with the optimization problem involved in specifying the rate of return allowed a regulated monopoly, consider first the optimization problem faced by the monopoly itself. Its profits can be written:

\[ \Pi = R(Q) - wL(Q, K) - rK \]

where \( R(Q) \) is the revenue derived from selling \( Q \) units of product; \( L(Q, K) \)—obtained by inverting the production function—is the amount of variable input required.

* Assistant professor of business economics, University of Manitoba; associate professor of economics, University of Toronto; and professor of economics, University of Minnesota, respectively. Most of the work described in this paper was undertaken during 1973 while we were all associated with the University of Toronto's Faculty of Management Studies and department of political economy. We are indebted to Bela Balassa and Carl Christ for helpful suggestions.

² Ignoring the motivational and indeterminacy problems that arise when the allowable rate of return is set equal to the cost of capital.
quired to produce \( Q \) units of output if a capital stock valued at \( K \) dollars is employed; \( w \) is the wage rate of the variable input; and \( r \) is the cost of capital. Depreciation expenses are assumed away.

The constraint that the monopoly is allowed to earn a rate of return \( s \), which is less than the rate \( s^* \) it would earn in the absence of regulation, can be written

\[
(2) \quad sK = R(Q) - wL(Q, K)
\]

This constraint can be interpreted as specifying the required capital stock \( K \) as an implicit function of the output level \( Q \). Using this implicit function to eliminate \( K \) in equation (1) reduces the set of decision variables to \( Q \) alone. Proceeding in this fashion avoids the difficulties involved in interpreting auxiliary variables. Differentiating the amended equation (1) with respect to \( Q \) yields

\[
(3) \quad R' - w(L_Q + L_K K') - rK' = 0
\]

where, from (2)

\[
(4) \quad K' = \frac{R' - wL_Q}{s + wL_K}
\]

Substituting (4) into (3) and rearranging terms yields:

\[
(5) \quad (s - r)(R' - wL_Q) = 0.
\]

If \( r \) is less than \( s \), this reduces to:

\[
(6) \quad R'(Q) = \frac{wL_Q}{Q, K}
\]

As several writers have noted, if \( s = r \), the behavior of the monopoly is indeterminate since each of the continuum of \( K \) values which satisfies equation (2) would yield the same (zero) economic profit.

Equation (6) is the analogue for a regulated monopoly of the standard condition for profit maximization, marginal revenue equals marginal cost: \( R' \) is marginal revenue while \( wL_Q \) is regulated or pseudomarginal cost. That is, \( wL_Q \) is the cost of the variable inputs required to produce an additional unit of output when \( K \) is set at the level required to satisfy the rate of return constraint, equation (2). In the range \( r < s < s^* \), using equation (2) to eliminate \( K \) from equation (6) yields \( Q^{**} = Q^{**}(s, w) \), the monopoly's profit-maximizing output as a function of the allowable rate of return and the wage rate of the variable input. Substituting \( Q^{**} \) for \( Q \) in equation (2) yields the profit-maximizing capital input level \( K^{**} = K^{**}(s, w) \).

Turning to the regulatory authority's optimization problem, we assume, as did Kleverick, the social objective to be maximization of the sum of consumers' and producers' surpluses:

\[
(7) \quad W = \int_0^{Q^{**}} P(q)dq - wL(Q^{**}, K^{**})
\]

\[
- rK^{**}
\]

Differentiating with respect to \( s \) yields:

\[
(8) \quad [P - wL_Q]Q^{**}_s = [r + wL_K]K^{**}_s
\]

as the necessary condition for maximizing \( W \). The left-hand side of this equation is the difference between price and (pseudo) marginal cost times the change in output resulting from a change in the allowable rate of return—the marginal net benefit of a change in \( s \). The right-hand side of equation (8) is the change in capital employed by the monopoly as a result of a change in \( s \) times the difference between the annual cost of a dollar's worth of capital plant and the reduction in variable input costs resulting from employment of that additional unit of capital. That is, the right-hand side of (8) equals the marginal

\[\text{Cost minimization would dictate employing that amount of capital for which } r + wL_K = 0. \text{ If the monopolist is to stay in business, } s \text{ must be greater than or equal to } r. \text{ As is well known from the Averch-Johnson literature, a rate of return constraint induces the employment of an inefficiently large amount of capital, i.e., a level for which } r + wL_K > 0. \text{ The denominator on the right-hand side of (4) would therefore be positive.} \]
net increase in costs resulting from the excessive substitution of capital for labor that rate of return regulation induces.

An unconstrained profit-maximizing monopoly would produce at minimum cost. Therefore, with \( s = s^* \), \( wLQ \) would equal the marginal cost of its product while \( r \) would equal \( -wL_k \). This being the case, when evaluated at \( s = s^* \), the left-hand side of (8) would be positive while the right-hand side would be zero. Reducing \( s \) to a value less than \( s^* \) could be expected to lower the difference between price and (pseudo) marginal cost. Also, by inducing the substitution of capital for labor, reducing \( s \) would lead \( -wL_k \) to fall short of \( r \). These considerations suggest that if administration costs could be ignored, reducing \( s \) below \( s^* \) would increase our measure of social welfare. That is, they suggest that regardless of the specific natures of the production and demand functions involved, the optimum allowable rate of return is less than that which an unconstrained monopolist would earn. It also seemed plausible to conjecture that the optimum value of \( s \) could be shown to lie in the range \( r < s < s^* \). We were unable to demonstrate the general validity of this conjecture, however. Indeed, the specific example described in the following section provides a counterexample to it, at least when increasing returns to scale are involved.

II. The Benefits and Costs of Rate of Return Regulation: An Example

We suppose, as did Klevorick, that the monopoly is faced by a constant elasticity demand schedule and that its production function is Cobb-Douglas. Specifically, the demand function is:

\[
P = AQ^{-\varepsilon}
\]

where \( A \) is a constant and \( 1/\varepsilon \) is the elasticity of demand. By appropriately specifying the units in which output is measured, the production function can be written without further loss of generality as

\[
Q^\beta = K^\alpha L^{1-\alpha}
\]

where \( 0 < \alpha < 1 \), and \( \beta \leq 1 \) is the reciprocal of the order of homogeneity of \( Q \). For notational convenience, it proves useful to define the following combinations of parameters:

\[
\begin{align*}
\gamma &= \beta + \varepsilon - 1 \\
\delta &= \omega\beta/[A(1 - \varepsilon)(1 - \alpha)] \\
\eta &= \gamma + \alpha(1 - \varepsilon) \\
\varrho &= \beta - (1 - \varepsilon)(1 - \alpha)
\end{align*}
\]

In the absence of constraints, the monopoly’s profit function would be

\[
\Pi = AQ^{1-\varepsilon} - wL - rK
\]

Using (10) to eliminate \( Q \) in equation (11) and differentiating with respect to \( L \) and \( K \) yields\(^5\) the profit-maximizing values of these variables. Substituting the results into equation (10) yields the profit-maximizing output level:

\[
Q^* = [A(1 - \varepsilon)(\alpha/\gamma)^{((1 - \alpha)/\omega - 1 - \varepsilon)}]^{1/\gamma}
\]

The difference between the revenue derived from this output level and the cost \( wL^* \) of the variable input used in producing it is the monopoly’s accounting profit. This difference divided by its profit-maximizing capital stock yields the ratio of its rate of return on invested capital as conventionally measured to its cost of capital:

\[
s^*/r = [\beta/(1 - \varepsilon) - (1 - \alpha)]/\alpha
\]

The profit-maximizing capital stock itself can be written:

\[
K^* = [\alpha w/((1 - \alpha)r)]^{1-\alpha}Q^\varrho
\]

\(^4\) Roger Sherman (pp. 390-93) presents a geometric analysis suggesting the validity of this conjecture when the monopoly operates under constant returns to scale.

\(^5\) The algebra involved is straightforward if tedious. Only the final results and a sketchy outline of the steps involved will therefore be given.
Substituting $K^*$ and $L^*$ into $C = wL + rK$ yields the monopoly's total cost:

$$C^* = \frac{(r/\alpha)w}{1 - \alpha}Q^{\frac{1}{1-\alpha}}$$

(15)

Turning to the regulated monopoly, substituting the demand and production functions into equation (2) and using the result to eliminate $K$ in equation (6) yields its profit-maximizing output:

$$Q^{**} = \left(\frac{w}{(1 - \alpha)\alpha}\right)^{1/\alpha}$$

(16)

Inserting (16) into equation (2) yields the profit-maximizing capital stock:

$$K^{**} = \delta^{(1 - \alpha)/\alpha}Q^{**\frac{1}{(1 - \alpha)}}$$

(17)

Introducing these last two equations into $L = L(Q, K)$ and the result into $C = wL + rK$ yields the constrained monopoly's total cost function:

$$C^{**} = wQ^{**\frac{1}{1 - \alpha}} + \frac{r}{\alpha}Q^{**\frac{1}{1 - \alpha}}$$

(18)

For the Cobb-Douglas production function under examination, $L(Q, K)$ is:

$$L = \left(\frac{Q^{\alpha}}{K^{\alpha}}\right)^{1/1-\alpha}$$

(19)

Inserting equation (9), the derivatives of $L$ with respect to $K$ and $Q$, and the derivatives of equations (16) and (17) with respect to $s$ into equation (8) yields the optimum allowable rate of return $s^{opt}$ if a value of $s > r$ satisfies equation (8):

$$s^{opt} = \frac{\eta}{\alpha(\beta - (1 - \alpha)(1 - \epsilon))}$$

(20)

If this equation is satisfied only for a value of $s \leq r$, welfare maximization would involve a corner solution. That is, it would require establishing an $s$ sufficiently in excess of $r$ to avoid the indeterminacy (see equation (5)) which arises when $s$ equals $r$ and the incentive to withdraw from production associated with an $s < r$. Note that the ratio $s^{opt}/r$ derivable from equation (20) depends only on parameters having straightforward economic interpretations: the elasticity of demand ($\epsilon$), the exponent of capital in the production function ($\alpha$), and the order of homogeneity of that function ($1/\beta$).

Division of equation (16) by equation (12) gives the relative increase in output effected by establishing a ratio $r/s$ of the cost of capital to the allowable rate of return as a function of these same parameter values:

$$Q^{**}/Q^* = \left(\frac{r/s}{\eta}\alpha(1 - \epsilon)\right)^{1/\gamma}$$

(21)

Substitution of $Q^{**}$ for $Q^*$ in equation (14) yields the amount of capital that would be required for production of $Q^{**}$ at minimum cost. Division of the resulting expression into equation (17) then yields a measure of the extent to which rate of return regulation leads to excessive employment of capital:

$$K^{**}(Q^{**})/K^*(Q^{**}) = \left(\frac{r/s}{\eta}\alpha(1 - \epsilon)\right)^{1/\gamma}$$

(22)

A similar substitution in equation (15) and division into equation (18) yields a measure of the cost increasing effect of rate of return regulation:

$$C^{**}(Q^{**})/C^*(Q^{**}) = \left(\frac{r/s}{\eta}\alpha(1 - \epsilon)\right)^{1/\gamma}$$

(23)

Given the way the analysis has been set up, the logical measure of the social benefits attributable to rate of return regulation might seem to be $W^{**}/W^*$, “net social welfare” contributed by the monopoly's operations under regulation divided by that associated with its unregulated behavior. However, as the absolute value of the elasticity of demand approaches one from above, the area under a constant elasticity demand schedule approaches infinity. This being the case, the percentage gain in $W$ resulting from a given absolute gain in the sum of consumers' and producers' surpluses diminishes with reductions in the elasticity of demand. With low elasticity values, small percentage
Table 1 gives the ratios of \( s^* \), the unconstrained monopoly rate of return, and \( s^{opt} \), the optimum allowable rate of return (more accurately, the value of \( s \) which satisfies equation (20)) to \( r \) the cost of capital for alternative values of \( \alpha, \beta, \) and \( \epsilon \) that yield stable equilibria. In addition, it provides values of equations (21)-(23) and (25) with \( s/r \) equal to \( s^{opt}/r \) when \( s^{opt} > r \) and \( s/r \) equal to one when a value of \( s \leq r \) satisfies equation (20). The equation values associated with \( s/r = 1 \) would not, of course, be observed in the real world. Still, it seems better to tabulate them than values associated with one plus some arbitrarily chosen increment.

If \( \beta + \epsilon < 1 \), the demand schedule intersects the marginal cost schedule from below. Entries in this table therefore involve only \( \beta + \epsilon \) values greater than one.
Perhaps the most important conclusion to be drawn from Table 1 is that the optimum allowable rate of return exceeds the cost of capital only when some combination of modest scale economies, an elasticity of demand close to one, and a small exponent of capital in the production function are in effect. If the $\Delta W^r/R^*$ benefit measure is accepted as reasonable, the potential gains (again, gross of administrative costs) from rate of return regulation are substantial. This form of regulation does produce substantial increases in capital inputs and, to a lesser extent, production costs. However, these cost increases are far more than offset by the additional consumers' plus producers' surpluses generated by the increased output. While the smallest value of the $\Delta W^r/R^*$ index tabulated (for $\alpha=0.33$, constant returns to scale, and a demand elasticity of four) is 11 percent, one index value in excess of 200 percent appears while values in excess of 100 percent are common.

Showing that rate of return regulation could make substantial contributions to social welfare given appropriate parameter combinations is not, of course, the end of the story. If raising subsidies involves no resource misallocation, the sum of consumers' and producers' surpluses would be maximized by establishing that output level for which price equals long-run marginal cost. If, for whatever reason, subsidies cannot be paid, the output level at which price equals long-run average cost is the best that can be hoped for. If the benefits derivable from rate of return regulation are close to those attainable with average or marginal cost pricing, the search for alternative regulatory procedures would be pointless.

To explore the potential benefits of alternative regulatory procedures, define the increases in net social welfare that would result from replacing unconstrained monopoly pricing by price equal to average and marginal cost, respectively, as $\Delta W^a=W^a-W^*$ and $\Delta W^m=W^m-W^*$ where $W^a$ and $W^m$ are the respective values of net social welfare resulting from setting price equal to marginal and to average cost. Using equations (9) and (15) to determine the output levels that would result from average and marginal cost pricing, making appropriate substitutions in equation (24), and dividing by $R^*$ yields:

\[
\Delta W^m/R^* = (1 - \epsilon)^{\beta/\gamma} - (1 - \epsilon)^{(-1-\epsilon)/\gamma - \gamma}
\]

Dividing equations (26) and (27) into equation (25) yields $\Delta W^r/W^a$ and $\Delta W^r/\Delta W^m$—the ratios of benefits derivable from rate of return regulation to those associated with average and marginal cost pricing. These ratios are given in the final two columns of Table 1.

The values of these ratios differ considerably from one set of parameter values to another. The $\Delta W^r/\Delta W^a$ and $\Delta W^r/\Delta W^m$ measures range from respective lows of 0.32 and 0.07 to respective highs of 0.94 and 0.82. Both of these ratios increase with increases in the exponent of capital in the production function. Also, $\Delta W^r/\Delta W^a$ increases with increases in the elasticity of demand and in scale economies whereas $\Delta W^r/\Delta W^m$ exhibits an irregular pattern of change with increases in these variables.

It seems highly unlikely that any regulatory authority anywhere has the information at its disposal that would be necessary for precise determination of optimum allowable rates of return for the activities it controls. Indeed, with such complete data, rate of return regulation would be pointless. No less information is necessary to determine $s^{opt}$ than to dictate the inputs necessary to produce efficiently that output for which demand price equals average or marginal cost. That is, the information necessary to determine $s^{opt}$ would permit
complete elimination of Averch and Johnson-type distortions.

For this reason, it is of interest to determine how the benefits suggested by Table 1 would be affected by specification of a nonoptimum allowable rate of return. It is this issue to which Table 2 is addressed. It lists $\Delta W^*/R^*$ for values of $s/r$ different from the optimum. In the first group of columns, ratios are calculated for allowable rates of return equal, alternatively, to 1.1, 1.25, 1.5, and 2 times the cost of capital when these values do not exceed the rate of return an unconstrained monopolist would earn. In the second group of columns, ratios are calculated for $s/r$ values equivalent, alternatively, to $a=(s-r)/(s^*—r)$ values of 0, 0.25, 0.5, and 0.75. The ratio $a$ is the amount by which the allowable rate of return exceeds the cost of capital divided by the amount by which the unconstrained monopoly rate of return exceeds the cost of capital.

Deviations from optimum allowable rates of return do lead to reduced benefits, of course. Reductions are particularly great when allowable rates are close to the cost of capital under circumstances in which $s'/r$ is substantially greater than one. But the benefits listed in Table 2 are positive and typically substantial even when there is a considerable difference between the allowable and the optimum rate of return. If—a very big if—generalization is warranted on the basis of these tables, regardless of the values of the technological and demand parameters under which it operates, subjecting a

### Table 2—$\Delta W^*/R^*$ Welfare Gain Ratios for Alternative Nonoptimal Ratios of the Allowable Rate of Return to the Cost of Capital

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$1/e$</th>
<th>$s^{opt}/r$</th>
<th>$\Delta W^<em>/R^</em>$ when $s/r$ equals:</th>
<th>$\Delta W^<em>/R^</em>$ when $a$ equals:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.1</td>
<td>1.25</td>
<td>1.5</td>
</tr>
<tr>
<td>.33</td>
<td>.6</td>
<td>1.1</td>
<td>2.06</td>
<td>.58</td>
<td>.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0</td>
<td>1.03</td>
<td>.78</td>
<td>.48</td>
</tr>
<tr>
<td>.8</td>
<td>.6</td>
<td>1.1</td>
<td>2.06</td>
<td>.63</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0</td>
<td>2.80</td>
<td>.39</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0</td>
<td>1.20</td>
<td>.17</td>
<td>*</td>
</tr>
<tr>
<td>1.0</td>
<td>.6</td>
<td>1.1</td>
<td>2.66</td>
<td>.36</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0</td>
<td>1.60</td>
<td>.13</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0</td>
<td>1.20</td>
<td>.11</td>
<td>.11</td>
</tr>
<tr>
<td>.67</td>
<td>.6</td>
<td>1.1</td>
<td>.82</td>
<td>2.17</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0</td>
<td>.55</td>
<td>.69</td>
<td>.14</td>
</tr>
<tr>
<td>.8</td>
<td>.6</td>
<td>1.1</td>
<td>1.11</td>
<td>1.55</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0</td>
<td>.84</td>
<td>.59</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0</td>
<td>.74</td>
<td>.11</td>
<td>*</td>
</tr>
<tr>
<td>1.0</td>
<td>.6</td>
<td>1.1</td>
<td>1.41</td>
<td>1.15</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0</td>
<td>1.14</td>
<td>.36</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.0</td>
<td>1.04</td>
<td>.16</td>
<td>.11</td>
</tr>
</tbody>
</table>

Note: $\alpha =$ capital exponent; $\beta =$ scale economy coefficient; $1/e =$ demand elasticity; $s^{opt}/r =$ optimum allowable rate of return $/$ cost of capital; $s/r =$ allowable rate of return $/$ cost of capital; $a =$ $(s-r)/(s^*—r) =$ deviation of allowable rate of return from cost of capital $/$ deviation of monopoly rate of return from cost of capital.

- $s/r > s^*/r$.
- Equation (20) satisfied for $s < r$. 

296 THE AMERICAN ECONOMIC REVIEW     JUNE 1976
formerly unregulated monopoly to the constraint that its rate of return on invested capital not exceed, say, 1.1 times its cost of capital would generate social gains substantially in excess of those which would accrue in the absence of regulation. Indeed, for some combinations of values of the scale economy, capital intensity, and demand elasticity variables, this form of regulation could yield benefits almost as great as those derivable from cost minimizing production coupled with average or even marginal cost pricing.

The literature contains a number of reports on attempts to fit Cobb-Douglas production functions to regulated utilities. Unfortunately, however, we have found only one such study which provides estimates of all three of the parameters that enter into our model. In this study, A. Rodney Dobell et al. derived estimates of our $\alpha$ and $\beta$ parameters equal to 0.36 and 0.90 respectively from an analysis of Bell Canada data for the period 1952–67. In addition, when evaluated for 1967 conditions, their demand relationship yields a long-run price elasticity of 2.01.

From equation (13), an unconstrained profit-maximizing monopolist subject to these conditions would earn a rate of return equal to 3.2 times its cost of capital. From equation (20), the optimum ratio of the allowable rate of return to the cost of capital would be 1.26. Imposing this constraint would generate a $\Delta W'/R^*$ benefit (equation (25)) of 0.28. This benefit is 45 and 43 percent respectively of those that would result from average and marginal cost pricing (equations (26) and (27)). Setting the allowable rate of return equal to 1.1 times the cost of capital would only lower the $\Delta W'/R^*$ benefit to 0.27, 43, and 41 percent respectively of the average and marginal cost pricing benefits. Even allowing a rate of return equal to twice the cost of capital would generate benefits equal to 30 and 29 percent respectively of those associated with average and marginal cost pricing.

REFERENCES
